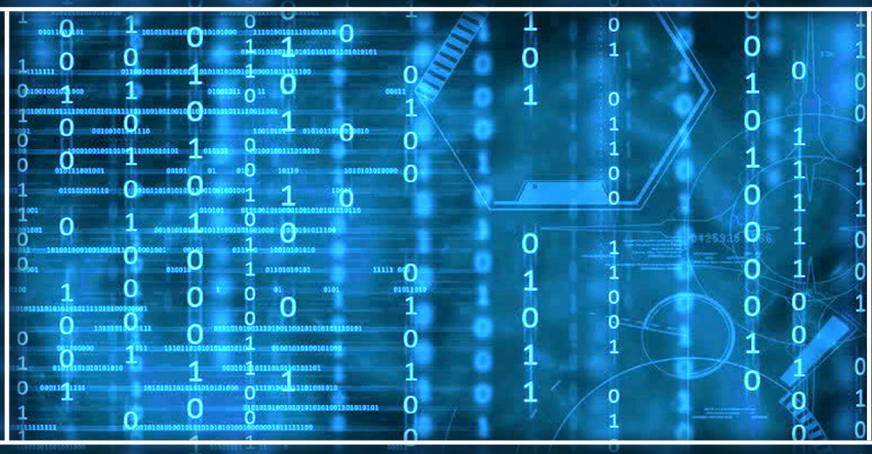
IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications

Volume 15 Issue 12
December 2024



ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

www.ijacsa.thesai.org

Editorial Preface

From the Desk of Managing Editor…

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 15 Issue 12 December 2024
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

CONTENTS

algoTRIC: Symmetric and Asymmetric Encryption Algorithms for Cryptography – A Comparative Analysis in AI Era

Naresh Kshetri[1], Mir Mehedi Rahman[2], Md Masud Rana[3], Omar Faruq Osama[4] and James Hutson[5]
Department of Cybersecurity, Rochester Institute of Technology, NY, USA[1]
School of Business and Technology, Emporia State University, KS, USA[2]
Department of Information Technology, San Juan College, NM, USA[3]
Dept. of System Science & Ind. Eng., Binghamton Univ., SUNY, NY, USA[4]
Department of Art History, AI & Visual Culture, Lindenwood University, MO, USA[5]

***Abstract*—The increasing integration of artificial intelligence (AI) within cybersecurity has necessitated stronger encryption methods to ensure data security. This paper presents a comparative analysis of symmetric (SE) and asymmetric encryption (AE) algorithms, focusing on their role in securing sensitive information in AI-driven environments. Through an in-depth study of various encryption algorithms such as AES, RSA, and others, this research evaluates the efficiency, complexity, and security of these algorithms within modern cybersecurity frameworks. Utilizing both qualitative and quantitative analysis, this research explores the historical evolution of encryption algorithms and their growing relevance in AI applications. The comparison of SE and AE algorithms focuses on key factors such as processing speed, scalability, and security resilience in the face of evolving threats. Special attention is given to how these algorithms are integrated into AI systems and how they manage the challenges posed by large-scale data processing in multi-agent environments. Our results highlight that while SE algorithms demonstrate high-speed performance and lower computational demands, AE algorithms provide superior security, particularly in scenarios requiring enhanced encryption for AI-based networks. The paper concludes by addressing the security concerns that encryption algorithms must tackle in the age of AI and outlines future research directions aimed at enhancing encryption techniques for cybersecurity.**

***Keywords*—Algorithms; analysis; artificial intelligence; asymmetric encryption; cryptography; cybersecurity; symmetric encryption**

I. Introduction

Algorithms are and were always the driving force behind cryptography and cybersecurity as we are marching towards the artificial intelligence (AI) and machine learning era. Several countermeasures, techniques, and cybersecurity practices are popular with the use of machine learning and deep learning algorithms apart from AI algorithms [1-2]. As we know cybersecurity combines information security and network security, the annual number of data breaches is growing every year. Loss of private information, malware attacks, use of smart gadgets, growing number of internet population, and several others are upcoming challenges for cryptography algorithms.

Vulnerabilities and attacks on ciphers, private keys, and algorithms are increasing as we are considering "Security for AI" and vise-versa [3-4]. New and unexpected attacks, development of several frameworks and tools are going on as we discuss various encryption algorithms. From the initial use of symmetric algorithms like Data Encryption Standard (DES), and their several weaknesses we tend to know that hackers are exploiting the powerful algorithms (like SHA3, MD5, up to CRYSTALS) today. The use of "secret key" in symmetric algorithms (although asymmetric works a little better as compared to symmetric) is no longer secret as attackers have successfully compromised the key in both symmetric and asymmetric algorithms.

The emergence of AI has revolutionized cybersecurity, providing adaptive and dynamic encryption techniques to combat swiftly changing cyber threats [1]. AI-driven methodologies have enhanced encryption systems' resilience, facilitating real-time identification of anomalies and threats that conventional methods find challenging to spot [2]. The use of AI, especially via machine learning (ML) and deep learning (DL) algorithms, has markedly improved the efficacy of encryption methods, rendering them more adept at managing the increasing complexity and volume of contemporary data environments.

AI is becoming more and more integrated into cybersecurity and encryption as technology advances. AI is essential for protecting AI systems from complex cyberattacks, in addition to fortifying encryption procedures by streamlining key generation and data security techniques [3]. In an increasingly linked and insecure digital world, the synergy between AI and encryption is essential because it allows more effective, scalable, and proactive security measures, guaranteeing the security of both data and AI systems [4].

This paper is structured as follows to explore the comparative analysis of encryption algorithms and their relevance in modern cybersecurity, particularly in the AI era. Section II provides a background study, outlining the historical evolution and challenges of cryptographic algorithms, and establishing the role of AI in enhancing these methods. Section

III examines the technical aspects of various encryption algorithms, focusing on their contributions to safeguarding sensitive data in complex systems. Section IV offers a comparative analysis of symmetric encryption (SE) and asymmetric encryption (AE), highlighting key differences in terms of efficiency, security, and scalability. Section V discusses the role of AI in transforming encryption practices, focusing on how AI enhances real-time adaptability, tackles emerging threats, and enables personalized encryption strategies. Section VI focuses on the security challenges encryption faces in modern society, particularly against emerging cyber threats. Section VII presents the discussion and conclusions, summarizing the insights gained from the comparative analysis and suggesting improvements for existing encryption techniques. Lastly, Section VIII outlines the future scope of research, discussing potential advancements in encryption algorithms and their application in AI-driven cybersecurity solutions.

II. Background Study

The foundation of contemporary encryption methodologies is well-examined in the literature, providing critical insights into their application in AI-driven contexts. Kapoor and Thakur [5] offer a thorough comparative analysis of symmetric and asymmetric key algorithms, underscoring the growing importance of encryption in safeguarding digital information in an increasingly networked environment. Their analysis highlights the superiority of asymmetric encryption, which employs two keys to enhance security through mathematical complexity. For symmetric algorithms, they emphasize the adaptability and efficiency of the Advanced Encryption Standard (AES), particularly its resilience against common attacks and its rapid execution speed. In contrast, the authors identify Elliptic Curve Cryptography (ECC) as the most secure asymmetric technique, noting its reliance on the algebraic properties of elliptic curves and finite fields. This detailed examination is a vital reference for algoTRIC, as it informs the optimization of AES and ECC within its architecture for large-scale, multi-agent systems. By focusing on trade-offs between speed and security resilience, the paper addresses critical challenges in mitigating emerging cyber threats.

Building on this foundation, Soomro et al. [6] conduct a comprehensive analysis of both symmetric and asymmetric cryptographic algorithms, focusing on their role in strengthening cybersecurity across diverse contexts. Their work identifies key cryptographic objectives—secrecy, integrity, authenticity, and non-repudiation—as essential for secure communication and data protection. They emphasize the speed and efficiency of symmetric algorithms, such as AES, making them suitable for high-throughput applications. Conversely, they highlight the robustness of asymmetric algorithms, notably RSA, for contexts requiring secure key management. This review contributes significantly to algoTRIC by elucidating how cryptographic strategies can be adapted to address the unique challenges of AI-driven systems. By balancing performance, scalability, and security resilience, this work helps frame the escalating need for robust data protection in modern cybersecurity frameworks.

Furthering these insights, Ustun et al. [7] introduce a machine learning-based intrusion detection system designed to address cybersecurity vulnerabilities in smart grids. Their approach leverages IEC 61850 Sampled Value (SV) messages to identify cyberattacks, particularly false data injection (FDI), within contemporary power system communication frameworks. By utilizing machine learning to distinguish between normal operations and cyberattacks, their system demonstrates high accuracy in identifying symmetrical and asymmetrical faults as well as FDI attacks. These findings are particularly relevant to algoTRIC's efforts to incorporate advanced intrusion detection algorithms in AI-driven environments. Specifically, their approach underscores the importance of integrating encryption techniques, such as AES and ECC, to secure communication streams while ensuring real-time intrusion detection in complex, multi-agent systems.

Similarly, Arora [8] examines the critical role of cryptographic methods in cybersecurity, emphasizing the importance of encryption and decryption for protecting digital data. The study underscores the efficiency of symmetric encryption techniques, such as AES, for managing large-scale data, alongside the superior security of asymmetric algorithms, like RSA, for secure key management. By addressing the fundamental principles of cryptography—confidentiality, integrity, and authenticity—Arora provides essential guidance for incorporating encryption into AI-driven systems. This analysis highlights the trade-offs between the high performance of symmetric encryption and the enhanced security of asymmetric approaches, offering a roadmap for optimizing encryption methods in AI systems that must balance computational demands with robust data protection.

Finally, Henriques and Vernekar [9] focus on the integration of symmetric and asymmetric cryptography to secure communication in Internet of Things (IoT) networks. They address the unique challenges posed by IoT systems, where sensitive data transmitted between devices demands heightened protection against cyberattacks. Their methodology combines the speed and efficiency of symmetric encryption, exemplified by AES, with the secure key management capabilities of asymmetric cryptography, such as RSA. This dual approach mitigates prevalent IoT vulnerabilities, including insecure network services and weak authentication mechanisms. Their work is particularly relevant to algoTRIC, as it explores how combining encryption algorithms can balance speed, scalability, and security in complex, large-scale AI-driven systems.

III. Encryption Algorithms for Cybersecurity

Building on the foundational insights from prior studies on the comparative strengths and applications of symmetric and asymmetric encryption algorithms, the next section delves into the practical implementation of these techniques within contemporary cybersecurity solutions. Encryption techniques and complex algorithms with respect to privacy preserving, wireless sensor networks (WSN), and AI are rapidly used in several system applications and solutions [10-11]. Applications like healthcare monitoring, smart cities, advertising, logistics with analysis of energy, overhead, speed are used for several AI-powered business models. Financial transactions (may consist of hash, public key, private key, and digital signature) today require high level security using Secure Hashing Algorithms

(SHA) and Message Digest (MD) algorithms used by distributed ledgers and blockchain technology (Table I).

Compressed sampling on encrypted images with a combined random Gaussian measurement matrix can also be used for AI based image encryption [12]. To resist several kinds of cyberattacks (primarily as primage attacks, collision attacks) that can pass plaintext sensitivity tests for successful communications. On the other hand, network security or endpoint security (of or partial of cryptography and/or cybersecurity), is fully achieved through data encryption using artificial intelligence [13]. Improving encryption speed, wireless sensors security, integrity of data proposed a proactive solution with remarkable performance as compared to static encryption methods.

Homomorphic encryption has arisen as a formidable method to bolster data security in AI-driven applications, facilitating computations on encrypted data without necessitating decryption. This capacity is essential for preserving data privacy in sensitive domains such as healthcare, banking, and smart city infrastructures, where AI is extensively employed for decision-making and data analysis. Homomorphic encryption encompasses several varieties, including fully homomorphic encryption (FHE), slightly homomorphic encryption (SWHE), and substantially homomorphic encryption (PHE), each presenting distinct trade-offs regarding computational complexity and efficiency [14]. Although Fully Homomorphic Encryption (FHE) permits infinite operations on encrypted data, its practical application is frequently constrained by substantial computing expenses and reduced processing velocities. Conversely, SWHE and PHE provide more efficient options by facilitating a limited range of actions, rendering them more appropriate for situations that emphasize performance while maintaining data security. In AI-driven contexts, including these encryption methods into machine learning models not only protects data during training and inference but also mitigates risks associated with emerging vulnerabilities such as data leakage and unauthorized access. As AI progresses, enhancing these encryption techniques will be essential for guaranteeing strong and scalable cybersecurity solutions.

Furthermore, the computational complexity of cryptographic algorithms emerges as a central concern, influencing not only the feasibility of deploying large-scale encryption solutions but also the security posture of data processing pipelines. Evaluations of complexity commonly employ Big-O notation, time-to-encrypt metrics, key-size scaling factors, and throughput measurements, all of which help determine the practical utility of a given cryptographic method. For symmetric encryption algorithms such as the Advanced Encryption Standard (AES), computational efficiency often proves to be one of their distinguishing strengths, as the complexity scales linearly with data input size, resulting in O(n) operations and predictable performance outcomes even as datasets grow larger. In contrast, asymmetric algorithms like RSA exhibit more pronounced complexity, commonly represented as $O(n^3)$ or higher when operations on large integers are involved, reflecting the significant computational overhead associated with public-key cryptography.

Moreover, the integration of AI-based threat detection and encryption acceleration further complicates these estimates, as machine learning heuristics and hardware-assisted optimizations can alter the baseline complexity by dynamically adjusting key distribution strategies, refining block-cipher rounds, or adopting hybrid encryption approaches. Evaluating complexity also demands close attention to scalability parameters in distributed environments, since multi-agent systems often require concurrent encryption-decryption operations across decentralized nodes, thereby magnifying the importance of parallelizable algorithms. Within this context, assessing complexity involves quantifying performance differentials over heterogeneous architectures, analyzing latency contributions from memory access patterns and cache line misses, and simulating the behavior of algorithms under diverse workload distributions.

As such, the integration of AI approaches with conventional encryption algorithms such as AES has demonstrated effectiveness in augmenting data security, especially in volatile threat landscapes. Recent research indicates that the integration of machine learning models, such as k-Nearest Neighbors (k-NN), with AES encryption markedly enhances the identification and mitigation of anomalies, facilitating real-time responses to new cyber threats. The k-NN's pattern recognition capabilities enhance the encryption process, adapting to emerging attack vectors and bolstering AES's resilience against advanced attacks [15]. This method enhances secure data transmission and bolsters the integrity of secret data storage. With the increasing volume and complexity of data in AI-driven systems, integrating machine learning with encryption methods such as AES is crucial for adopting a proactive approach to cybersecurity.

Chaotic algorithms have arisen as an effective solution for image encryption in AI-driven networking systems, owing to its intrinsic characteristics such as sensitive dependence on beginning conditions, topological mixing, and long-term unpredictability [16]. These qualities are utilized to generate intricate encryption patterns, where even minor alterations in the original settings result in completely distinct encrypted outputs, hence substantially improving data security. Recent implementations indicate that chaotic algorithms, along with sophisticated encryption techniques, can provide non-linear transformations that effectively rearrange and disperse pixel positions, rendering the image data into a highly randomized state. This method guarantees that the encryption process emulates a dynamical system, rendering the reversal of the process without precise system parameters computationally impractical [16]. Through the application of repeated chaotic functions, these encryption methodologies guarantee elevated entropy in the encrypted data, so successfully countering brute-force assaults and enhancing resilience against cryptographic scrutiny. In AI-driven environments, where data security is imperative against advanced threats, the amalgamation of chaotic systems with encryption enhances the security framework while preserving computational performance by reducing processing overhead.

TABLE I. INTUITIONS (UP TO THREE) OF SOME COMMON ADVANCED ENCRYPTION ALGORITHMS FOR SECURITY AND CRYPTOGRAPHY IN THE ARTIFICIAL INTELLIGENCE (AI)-DRIVEN SOCIETY

Ref	Encryption Type	Intuition I	Intuition II	Intuition III
[10]	Partial Homomorphic	Enables privacy-preserving computations on encrypted blockchain data	Mitigates risks from collision, preimage, and wallet attacks	Optimizes computational overhead for AI-integrated blockchain environments
[11]	AI-Driven Data Solutions	Adapts encryption parameters dynamically based on real-time network conditions	Integrates anomaly detection to proactively adjust encryption settings against threats	Optimizes computational and energy resources while maintaining high security levels
[12]	AI Image	Utilizes hyperchaotic sequences for robust pixel scrambling and diffusion	Enhances resistance against differential and brute-force attacks	Achieves high randomness and compression efficiency with compressed sensing
[13)	Innovative Data for WSANs	Adapts encryption parameters dynamically using AI for real-time threat response	Leverages LSTM networks to optimize encryption based on sequential data analysis	Employs Isolation Forests to enhance anomaly detection and network resilience
[14]	AI-based Homomorphic	Enables privacy-preserving computations on encrypted data without decryption	Mitigates data exposure risks in untrusted environments like cloud computing	Supports collaborative AI tasks with multi-key encryption across multiple parties
[15]	AI and AES	Combines AES's robust encryption with AI for adaptive threat detection	Utilizes AI-driven k-NN for real-time anomaly analysis in encrypted data	Enhances encryption efficiency through AI-optimized parameter selection
[16]	Image Transmission	Leverages chaotic mapping for high sensitivity and complex key generation	Enhances image confidentiality through pixel-level scrambling and diffusion	Mitigates brute-force attacks via topological chaos and statistical uniformity

IV. COMPARISONS OF ALGORITHMS W.R.T. SE AND AE

As encryption techniques continue to evolve, their applications in various domains underscore the need for a nuanced understanding of their operational strengths and limitations. The exploration of how these algorithms integrate into modern cybersecurity frameworks provides a foundation for deeper analysis. Thus, a focused comparison follows between symmetric encryption (SE) and asymmetric encryption (AE) will elucidate the key distinctions that influence their use. By examining differences in key management, scalability, performance, and reliability, this analysis aims to identify the most suitable encryption methods for specific applications and highlight the critical trade-offs involved in their deployment within contemporary cryptographic systems.

To evaluate the cryptographic algorithms, it is significant to contrast symmetric encryption algorithms with asymmetric encryption algorithms as modern cryptography is designed based on symmetric and asymmetric encryption which are two fundamental categories of encryption algorithms (Table II). The main purposes of both types of encryption are the same, that is to safeguard the data security and integrity over the diverse applications. Although their purposes are the same, they have significant differences based on the way of managing encryption keys, evaluating performance and functionality requirements. To identify the most effective encryption method for a particular scenario, it is essential to distinguish the strength, weakness, functionalities and other features of both types of encryption methods. This section of the paper distinguishes the fundamental types of encryption algorithm based on key management, scalability, swiftness and reliability.

One of the main differences of symmetric and asymmetric encryption is the number of keys used in the encryption process. There are two types of keys used in encryption and decryption processes which are known as public key and private key. In a symmetric algorithm, a private key is used alone to encrypt and decrypt data [17]. On the other hand, an asymmetric algorithm uses both the public key and private key where the public key is used to encrypt data and private key is used to decrypt data. Public key encryption is designed based on intensive computational mathematical functions; therefore, asymmetric algorithms are not very suitable or efficient for minor devices.

The second important term of differences between symmetric and asymmetric encryption is reliability. The encryption process of the symmetric method is simpler than the asymmetric method, however, in symmetric method both the sender and receiver share the common private key to encrypt and decrypt data which is a major concern about data security as eavesdropping can be conducted by attacker anytime in the channel of data exchange. Alternatively, in asymmetric encryption, the public key is used to encrypt the data while the private key is used to decrypt the data [18]. As the private key is secret and only the receiver knows the private key, it becomes very difficult for the attacker to decrypt the original data. As a result, asymmetric encryption is considered more reliable in comparison to symmetric encryption in case of data exchange.

Swiftness of encryption and decryption is also a very powerful component that can be considered to differentiate symmetric and asymmetric encryption. Al-Shabi, in his paper, conducted an analysis to compare the performance for identifying the strengths and weaknesses of different types of

symmetric and asymmetric encryption based on various factors such as battery consumption, block size, structure, time consumption and types of attacks. His result shows that based on real-time encryption, a symmetric algorithm is much faster than asymmetric encryption [18]. Similar kind of study was conducted by Panda in 2010. Her paper indicates that a symmetric algorithm is almost 1000 times faster than an asymmetric algorithm as an asymmetric algorithm needs more powerful computational resources. To compare different types of algorithm, 3 types of file such as text, image and binary were used in her analysis where the performance factors were decided considering Encryption Time, Decryption Time and Throughout. The result of her study found better performance from the AES algorithm, a subcategory of symmetric encryption, in comparison to other encryption algorithms based on Encryption Time, Decryption Time and Throughout [19].

Use of blocks is also a considerable component that can be used to distinguish between symmetric and asymmetric algorithms. There are mainly two important components considered in symmetric encryption known as block cipher and stream cipher, which are significantly crucial for confidentiality of data and integrity of cryptography [21]. AES, a subcategory of symmetric encryption, is operated on plaintext where the size of the block is 128 bits. This block cipher can also utilize different key lengths such as 128 bits, 192 bits or 256 bits of cipher secret [20]. On the other hand, asymmetric encryption does not require block size to encrypt data, rather this method leverages the idea of chunk data processing that is correspondent to the key size.

In the field of AI-driven cybersecurity, selecting between symmetric encryption (SE) and asymmetric encryption (AE) involves a thorough evaluation of performance, scalability, and security requirements. SE algorithms, such as AES, excel in real-time AI applications due to their high-speed encryption and low computational demands, which highlights as essential for AI tasks requiring rapid data processing [19]. However, AE algorithms like RSA provide enhanced security by leveraging public-private key pairs, a feature that underscores as crucial for maintaining confidentiality in sensitive data exchanges [18]. While SE is ideal for resource-constrained AI environments, such as IoT, due to its lower energy consumption, AE's computational intensity makes it better suited for secure initial key exchanges in distributed AI systems [20]. This difference in resource demands directly impacts scalability; SE supports continuous, high-throughput data streams often required in AI workflows, while AE's structure enables secure data sharing across complex, multi-agent networks through recent advances in secure communication protocols [21]. Effective cybersecurity in AI ultimately requires balancing SE's efficiency and AE's strong data protection, particularly in applications where threats to data integrity and confidentiality are significant [17].

TABLE II. COMPARISON OF SYMMETRIC ENCRYPTION AND ASYMMETRIC ENCRYPTION IN AI-DRIVEN CYBERSECURITY

Aspect	Symmetric Encryption (SE)	Asymmetric Encryption (AE)	Ref.
Integration with AI	SE algorithms like AES and Blowfish are efficient for real-time AI-driven data processing, supporting rapid encryption for high data volumes in AI workflows.	AE algorithms such as RSA and ECC are suitable for securely establishing initial connections in AI systems, though slower for real-time processing.	[19]
Data Throughput	High throughput makes SE ideal for handling large data in AI tasks (e.g., image processing or continuous data flows in AI-based IoT).	Lower throughput is better for secure, one-time exchanges rather than sustained high-speed AI-driven processing.	[21]
Resource Optimization	Low computational demands allow SE to support AI applications in resource-constrained environments, like mobile AI/IoT.	Higher resource needs make AE less suitable for low-power AI applications, though ideal for secure initial setup in complex AI networks.	[20]
Real-Time Efficiency	SE provides rapid encryption/decryption, enhancing real-time AI functions like anomaly detection in cybersecurity.	Slower speed limits AE in real-time AI scenarios; however, it provides robust security for secure data onboarding in AI systems.	[18]
Scalability in AI Systems	SE scales well within high-speed AI environments, enabling quick encryption across multi-agent or large data environments.	AE scales better for secure AI communications in distributed or cloud-based systems, especially for sensitive exchanges.	[21]
Battery and Power Use	Low power consumption suits AI-based mobile or IoT cybersecurity applications, allowing efficient continuous data encryption.	Higher power demand limits AE's suitability for battery-dependent AI devices, though it's viable for centralized secure key exchanges.	[20]
Security Strength	SE algorithms are faster but require secure key management in AI-driven environments to prevent compromise.	AE's public-private key pair provides greater security in AI-based networks with high confidentiality needs, particularly when securing data exchanges.	[18]
Complexity	Simpler structures in SE make it easier to embed into AI cybersecurity models needing rapid, low-latency responses.	AE's complexity is suitable for initial secure connections but can slow down ongoing high-volume AI data processing.	[19]
Use in AI Applications	Frequently applied in AI-driven real-time applications like intrusion detection, anomaly detection, and real-time threat monitoring.	Used to establish secure connections for sensitive AI operations, such as secure federated learning or distributed AI models.	[17]

V. Algorithms in the AI Era

The comparative analysis of symmetric and asymmetric encryption algorithms reveals critical insights into their respective strengths and limitations, offering a clear framework for selecting appropriate methods based on specific requirements. However, as the cybersecurity landscape continues to evolve, traditional encryption approaches must adapt to emerging challenges. The next section explores how AI is evolving encryption by introducing adaptive and dynamic capabilities. Through the integration of ML and DL models, encryption techniques are becoming more resilient, enabling real-time detection of threats and enhancement of key generation processes.

AI is increasingly integrated into encryption techniques, offering adaptive and dynamic solutions to address evolving cybersecurity threats. ML models play a pivotal role by analyzing large datasets to detect anomalies, making encryption protocols more resilient to cyberattacks [22] (Fig. 1). In recent years, there has been a surge in the application of deep learning to enhance cryptographic algorithms, particularly through convolutional neural networks (CNNs). These models help to create more robust key generation processes, as demonstrated in recent studies where CNNs were applied to Advanced Encryption Standard (AES) algorithms to improve encryption performance and security resilience [23]. Such AI-driven encryption systems are capable of continuously evolving, adapting to new security challenges, and countering sophisticated hacking attempts in real-time [24].

In addition to improving encryption processes, AI also aids in the proactive detection and mitigation of cyber threats. As Rangaraju [25] notes, through leveraging ML models, particularly deep learning algorithms, cybersecurity systems can predict potential vulnerabilities and strengthen encryption methods. These techniques not only enhance the overall security infrastructure but also allow for the development of intelligent, self-updating systems that can respond to newly emerging cyber threats. The real-time adaptability of AI in encryption is crucial, especially as traditional cryptography methods, such as RSA, become increasingly vulnerable to advanced cyberattacks [26]. This integration of AI into cryptography sets the stage for more secure communication and data protection in the AI era [27].

With the newly found ability to detect and mitigate cybersecurity threats, AI assists in offering advanced solutions that traditional encryption methods struggle to match. These solutions include CNNs, as noted, but also long short-term memory (LSTM), AI-driven systems that can analyze vast amounts of data in real-time, identifying patterns that signal potential threats. These AI-enhanced systems use data profiling techniques to categorize security events, enabling more accurate discrimination between legitimate threats and false positives [28]. For example, a study employing AI-based security information and event management (SIEM) demonstrated improved accuracy in detecting network intrusions by combining event profiling with various neural networks, outperforming traditional machine learning approaches [29] [30]. The ability to adapt to complex and evolving attack patterns makes these new technologies an essential tool for modern cybersecurity.

Such capacity to adapt and learn from emerging threats is critical as cybercriminals continuously develop more sophisticated attack methods. Deep learning models, especially when applied to real-time cybersecurity monitoring, can detect anomalies much faster than traditional methods, providing organizations with the agility to respond to cyberattacks proactively [31]. Recent advancements in deep learning-based intrusion detection systems (IDS) have shown promising results in identifying zero-day attacks, reducing detection time, and improving overall system security [32]. This proactive approach allows for not only quicker detection but also the anticipation of future attacks, helping organizations stay one step ahead of cybercriminals.

On the other hand, although integrating AI into encryption processes provides significant advancements and benefits, there are also numerous challenges and ethical concerns. One of the primary issues is the risk of over-reliance on AI-based systems, which could lead to complacency in monitoring and updating security protocols [33]. The dynamic nature of these tools can make encryption systems highly efficient, but this reliance also increases the risk that undetected vulnerabilities could be exploited by adversaries using AI for malicious purposes [34]. Furthermore, as AI-driven encryption systems become more widespread, the sheer volume of data processed raises concerns about privacy violations. AI models often require vast amounts of personal or sensitive information to function optimally, which can lead to unintended privacy breaches if not managed properly [35].

Another ethical concern involves the dual-use nature of AI technologies in encryption. While AI enhances security, it also opens avenues for adversaries to exploit AI systems to breach encrypted communications. AI-based algorithms could potentially be reverse-engineered or manipulated to bypass security protocols, creating a new type of cyber threat [36]. The sophistication of AI tools allows attackers to uncover hidden patterns or weaknesses in encryption systems, potentially leading to large-scale data breaches. This highlights the need for comprehensive governance frameworks that address not only the technical challenges but also the ethical risks associated with deploying AI in encryption and cybersecurity [37].

Looking ahead, ever-advancing AI tools are expected to play an increasingly central role in the future of encryption, evolving alongside the cyber threat landscape. The adaptability of AI to real-time data allows for personalized encryption solutions tailored to the behaviors and preferences of individuals, making it more difficult for cybercriminals to execute successful attacks [38]. Through learning from patterns in network traffic and user behavior, AI can continuously optimize encryption protocols, ensuring that they remain effective against emerging threats [39]. This ability to adapt to new challenges positions the technology as a vital tool in maintaining robust cybersecurity defenses in the coming years.

Moreover, integration into encryption technologies opens possibilities for more seamless and efficient security solutions. The use of AI to automate encryption processes could lead to faster, real-time encryption adjustments without human intervention. This is particularly valuable in dynamic environments, such as the Internet of Things (IoT), where

devices continuously communicate and exchange data [40]. The ability to monitor and respond to security threats in real-time ensures that encryption methods are always up to date, thus reducing the risk of breaches [41]. However, these advancements must be balanced with considerations for ethical use and the prevention of potential misuse of AI in malicious hacking activities.

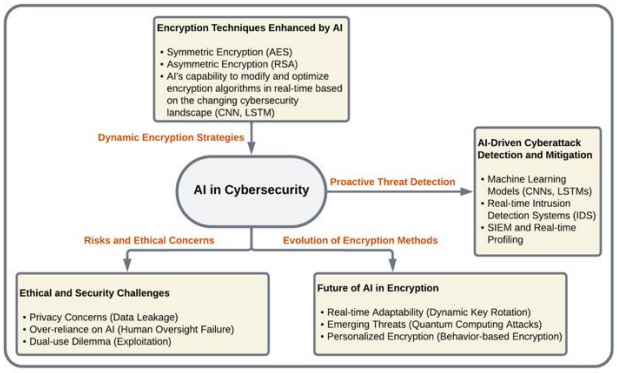


Fig. 1. AI-driven enhancements in encryption (including symmetric and asymmetric) and cybersecurity.

In the era of rapid technological progress, artificial intelligence has emerged as a revolutionary influence across several domains, including cybersecurity. As AI systems advance, the algorithms utilized for data protection as well as encryption must adapt to the intricacies of contemporary threats. The convergence of AI and encryption offers prospects for bolstering cybersecurity resilience via real-time monitoring, adaptive response strategies, and intelligent automation.

A. *Artificial Intelligence-Enhanced Encryption for Improved Cybersecurity*

As AI increasingly integrates with encryption, its transformative impact on cybersecurity becomes evident. The prior discussion outlined the potential of AI-driven methodologies in enhancing traditional encryption systems, offering adaptive and dynamic capabilities. This section delves deeper into the specific mechanisms by which AI enhances both symmetric and asymmetric encryption techniques, focusing on how AI-driven solutions address emerging cybersecurity threats through improved key generation, anomaly detection, and real-time responsiveness.

Conventional encryption techniques, such as the Advanced Encryption Standard (AES) in symmetric encryption and RSA in asymmetric encryption, have been significantly augmented by AI to boost their security and efficiency. AI's capacity to analyze vast datasets, identify trends, and adapt to evolving threats positions it as an ideal collaborator for cryptographic systems.

In symmetric encryption, AI-driven optimization strategies dynamically create, and update AES encryption keys based on real-time threat assessments. Machine learning (ML) algorithms now anticipate vulnerabilities and pre-empt brute-force attacks by identifying anomalous patterns across encrypted data. This dynamic methodology transforms AES into a more adaptable and resilient system, capable of addressing diverse threats without compromising operational speed [32].

For asymmetric encryption, RSA benefits from AI's ability to refine the key generation process. Genetic algorithms, a subset of AI methodologies, enhance the selection of prime numbers, ensuring that encryption keys are robust and less vulnerable to attacks [23]. These advancements reduce computational demands for both encryption and decryption processes while maintaining high levels of security, particularly in environments requiring secure communications.

Deep learning methodologies further expand the potential of AI-enhanced encryption. Techniques such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are now integrated into cryptographic frameworks to monitor encrypted communications in real-time. These algorithms detect irregularities in data streams, identify potential breaches, and enable pre-emptive responses to system intrusions [33]. By adding this layer of real-time detection, AI provides an additional safeguard that static encryption technologies cannot match.

Moreover, the incorporation of AI into cryptographic processes enhances both efficiency and effectiveness. For instance, CNNs within AES key generation operations not only improve security but also lower computational costs [28]. In resource-limited environments such as the Internet of Things (IoT), asymmetric cryptographic methods like RSA leverage AI-driven approaches to optimize encryption and decryption processes, ensuring secure communication without overburdening system resources [29].

B. *Homomorphic Encryption and Privacy-Enhancing Artificial Intelligence Methodologies*

The integration of AI into conventional encryption techniques highlights its transformative potential to enhance security, efficiency, and adaptability. While these advancements address many existing challenges, the need for encryption methods that maintain data confidentiality during processing is paramount, particularly in fields requiring large-scale data analysis. As such, one of the most exciting advancements in AI-driven encryption involves the progression of homomorphic encryption. Homomorphic encryption enables calculations upon encrypted data without necessitating decryption, so safeguarding sensitive information during processing. This is especially beneficial in AI applications requiring the analysis of large data sets, such as in finance, healthcare as well as cloud computing. Also, AI is significantly enhancing the efficiency and scalability of homomorphic encryption techniques. Utilizing AI methodologies might enhance the efficacy of homomorphic encryption by reducing the noise typically accumulated during calculations, hence making these techniques more appropriate for practical use [35].

This advancement is particularly significant for privacy-preserving AI applications, in which sensitive data, such as health-related records and financial information, must be safeguarded throughout the analytical process [22]. Homomorphic encryption, in conjunction with AI, allows businesses to cooperate upon encrypted data without disclosing the underlying knowledge. This privacy-preserving methodology has considerable ramifications for sectors such as healthcare, where patient information may be safely exchanged

and evaluated across institutions without jeopardizing privacy or regulatory adherence [39].

Moreover, AI-based methodologies have begun to influence the design, evaluation, and implementation of encryption algorithms, offering novel avenues for both enhancing and challenging traditional security paradigms. Such approaches incorporate machine learning-based techniques to identify patterns in cipher operations, anticipate potential vulnerabilities, and recommend key management strategies tailored to diverse computational contexts. By employing deep learning models trained on large-scale encryption datasets, researchers can detect subtle correlations in encrypted traffic and refine key scheduling protocols, leading to more resilient cryptographic schemes. In addition to bolstering algorithmic integrity, AI-driven methodologies assist in automating threat detection, as real-time analytics enable dynamic adjustments to key sizes, modes of operation, and encryption parameters based on evolving adversarial tactics. The infusion of AI elements further empowers hybrid encryption approaches where neural networks guide the selection between symmetric and asymmetric algorithms, optimizing both security and computational efficiency.

Lastly, reinforcement learning agents can adaptively determine when to apply advanced cryptographic primitives, such as fully homomorphic encryption, by weighing computational overhead against security gains. Beyond defensive capabilities, AI-based methodologies facilitate the detection and prevention of side-channel attacks, since carefully tuned machine learning classifiers recognize subtle anomalies in power consumption or electromagnetic emissions. Although these techniques hold immense promise, they also raise new ethical and regulatory questions regarding data privacy, algorithmic transparency, and model interpretability, necessitating continuous oversight and methodological rigor in future AI-cryptography research.

C. Blockchain and Artificial Intelligence: A Collaborative Strategy for Security

As homomorphic encryption exemplifies the potential of AI-driven methodologies for securing sensitive data during processing, the integration of AI with blockchain technology offers a complementary avenue for advancing cybersecurity. Blockchain, known for its decentralized and secure architecture, has emerged as a critical tool for safeguarding digital transactions across industries such as finance, healthcare, and supply chain management. However, the growing complexity of blockchain applications demands greater efficiency, scalability, and resilience. AI's integration with blockchain not only addresses these challenges but also enhances the foundational security and operational efficiency of blockchain networks.

Blockchain's inherent security lies in its decentralized structure, which distributes data across multiple nodes to prevent tampering and ensure transparency. When combined with AI, this architecture is further fortified by novel cryptographic techniques such as AI-driven homomorphic encryption. These advanced methods secure data transmission across blockchain networks, even as the volume and complexity of transactions increase. The incorporation of AI enhances blockchain's ability to handle sophisticated encryption requirements, making it a more robust framework for industries that rely on secure, high-throughput digital transactions.

AI also revolutionizes blockchain's consensus mechanisms, which are essential for verifying transactions and maintaining data integrity. Traditional methods like proof-of-work (PoW) and proof-of-stake (PoS) are often criticized for their high energy consumption and computational inefficiencies. AI-augmented consensus algorithms address these limitations by streamlining the validation process, significantly increasing transaction speed while reducing energy demands [40]. This optimization makes blockchain networks more sustainable and scalable, enabling their adoption in diverse and resource-intensive applications without compromising security.

Beyond efficiency, AI contributes to blockchain's real-time security capabilities by identifying and mitigating threats as they arise. Machine learning and anomaly detection algorithms enable blockchain networks to detect irregular transaction patterns, prevent unauthorized access, and counter distributed denial-of-service (DDoS) attacks. These proactive measures ensure that blockchain remains a reliable and resilient platform for secure digital transactions [37]. The fusion of AI's adaptive intelligence with blockchain's decentralized infrastructure not only addresses existing challenges but also sets new benchmarks for trust, scalability, and security in an evolving digital ecosystem.

D. Artificial Intelligence and Quantum-Resistant Cryptography

The integration of AI with blockchain technologies demonstrates its potential to address contemporary cybersecurity challenges, but the emergence of quantum computing introduces a new frontier of threats. Quantum computers, with their unparalleled ability to solve complex mathematical problems, threaten to undermine traditional cryptographic methods such as RSA and elliptic curve cryptography (ECC). As this technological shift looms, AI is playing a pivotal role in developing quantum-resistant cryptographic methods to ensure the continued security of digital communications.

One of the most promising approaches to quantum-resistant cryptography involves lattice-based algorithms, which rely on the computational difficulty of solving lattice problems—a complexity that remains formidable even for quantum computers. AI methodologies enhance the development and evaluation of these post-quantum cryptographic algorithms by identifying potential weaknesses and optimizing their implementation in practical systems [30]. By leveraging AI-driven simulations and predictive modelling, researchers can refine lattice-based encryption techniques to ensure their resilience against both theoretical and practical quantum attacks.

In addition to fortifying cryptographic algorithms, AI also contributes to preparing for the broader implications of quantum computing. Through the simulation of quantum assaults, AI enables the rigorous testing of existing encryption methods under quantum conditions. This proactive approach not only helps to identify vulnerabilities but also informs the creation of robust cryptographic standards designed to safeguard sensitive information in the quantum age [27]. Moreover, AI models are

used to predict the pace and direction of quantum computing advancements, enabling the development of encryption methods that stay ahead of potential threats [26]. The synergy between AI and quantum-resistant cryptography exemplifies the forward-thinking strategies required to navigate this impending technological shift. As quantum computing capabilities grow, the collaboration of AI and cryptography will be instrumental in ensuring that encryption techniques evolve to meet new challenges.

E. Ethical Implications in AI-Enhanced Cryptography

As advancements in AI-driven encryption and quantum-resistant cryptography push the boundaries of cybersecurity, they also introduce complex ethical considerations. The deployment of such powerful technologies raises critical questions about transparency, accountability, and equitable access, necessitating a careful examination of the broader societal implications of AI-enhanced cryptography. The incorporation of AI within encryption systems presents significant ethical dilemmas. As AI algorithms increase in complexity, the need for openness and accountability in their decision-making processes, especially in encryption and cybersecurity, is intensifying. It is essential to design AI-driven cryptography systems with ethical concerns to foster confidence and avoid abuse.

A primary worry is the dual-use characteristic of AI technology. Although AI may improve encryption as well as cybersecurity, it may also be utilized by nefarious individuals to develop more advanced assaults or to avoid detection. Developing AI-driven encryption systems with strong ethical standards is crucial to avoid their misuse for bad reasons [36]. Furthermore, as AI along with encryption technologies proliferate, it is essential to guarantee their accessibility and equity. It includes tackling the digital divide including guaranteeing that modern encryption technologies are accessible to all societal sectors, not just to those with the means to use them [24].

To get farther into the AI age, encryption algorithms must advance to match the increasing sophistication of cyber threats. Artificial intelligence is significantly transforming both asymmetric and symmetrical encryption systems, which renders them more adaptable, effective, and safe. The integration of AI in key generation and real-time threat detection is transforming cybersecurity methodologies. Nonetheless, the prospect of AI-driven cryptography has concerns as well. It is essential for these systems to be morally robust, transparent, and resilient against new dangers, including those from quantum computing, to ensure their success. Advancing and perfecting AI-driven encryption methods will enable the establishment of an increased secure digital future which safeguards sensitive information while promoting innovation.

VI. Algorithm Security in Modern Society

Encryption algorithms are essential tools in maintaining the confidentiality and integrity of digital communications in modern society (Fig. 2). With the increasing reliance on digital platforms for both personal and professional interactions, ensuring secure communication has become a priority [42]. Algorithms such as the AES and RSA are widely adopted to protect sensitive data, including emails, financial transactions, and other online communications. AES, a symmetric key algorithm, is favored for its speed and efficiency in encrypting large volumes of data, making it suitable for applications where rapid data processing is critical [43]. In contrast, RSA, an asymmetric key algorithm, is often used for secure key exchanges and digital signatures due to its robust security features, although it operates at a slower speed [44]. Together, these algorithms form the foundation of secure digital communications, providing the first line of defense against unauthorized access and cyberattacks.

As society becomes more dependent on digital communication, the application of encryption algorithms continues to expand. For instance, hybrid encryption schemes that combine the strengths of both AES and RSA are becoming more popular. These hybrid systems leverage the efficiency of AES in data encryption and the strength of RSA in secure key management, ensuring that both the data and the encryption keys are protected during transmission [45]. Such combined approaches offer enhanced security, particularly in environments where large volumes of sensitive information are frequently exchanged, such as in e-commerce or financial institutions [46]. As encryption technologies evolve, they continue to play a vital role in safeguarding digital communication, adapting to new threats and ensuring that sensitive information remains confidential and secure [47]. Thus, actionable risk assessment methodologies are particularly valuable for organizations that rely heavily on algorithms for their security, as they provide a clear framework to assess vulnerabilities, adapt to evolving threats, and reduce reliance on external vendors [48].

Yet, as noted, the rapid adoption of IoT devices and cloud computing has created new vulnerabilities in cybersecurity systems, particularly due to the limited computing capabilities of many IoT devices [34]. Many of these devices rely on lightweight encryption algorithms, such as the Data Encryption Standard (DES) or AES, which are efficient but may be more susceptible to attacks due to their reduced complexity [49]. Additionally, IoT devices often lack regular security updates, making them easy targets for cybercriminals. Cloud computing environments further complicate the situation, as data in transit and at rest in the cloud are vulnerable to interception, especially during migration between different cloud platforms [49]. This growing complexity necessitates the development of more robust encryption techniques tailored to the needs of both IoT and cloud environments [51].

Furthermore, the rise of supply chain attacks, where third-party software or hardware components are compromised, presents another significant challenge. As Hammi Zeadally and Nebhen (2023) point out, since many organizations rely on cloud services that integrate multiple external vendors, ensuring the security of every component is increasingly difficult [52]. In such environments, traditional encryption methods may not provide sufficient protection against sophisticated attacks. Emerging encryption models, such as lattice-based cryptography and hybrid encryption schemes, have been proposed as solutions to strengthen security, especially in resource-constrained IoT devices and cloud platforms [53]. As IoT and cloud ecosystems continue to expand, the demand for

an advanced encryption methods that can effectively address these new vulnerabilities [50] will only increase [54]. Also, the escalating sophistication of cryptojacking and ransomware highlights the importance of robust encryption algorithms to safeguard against unauthorized access and financial disruptions who are using blockchain technology for their security [55].

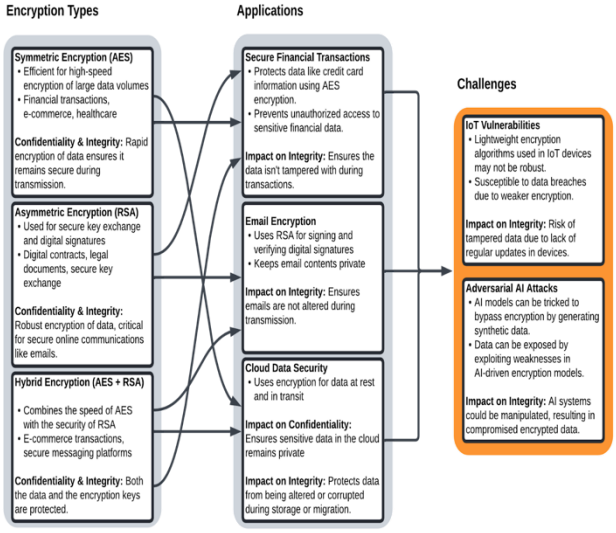


Fig. 2. Encryption algorithms (symmetric, asymmetric, and hybrid) securing digital communications.

While these tools can enhance encryption and cybersecurity, it also introduces new vulnerabilities, particularly through adversarial AI attacks. These attacks exploit the weaknesses in AI models by introducing adversarial inputs, causing the system to make incorrect decisions. In the context of encryption, adversaries can manipulate these models designed to detect anomalies in encrypted communications or tamper with ML algorithms that generate encryption keys [56]. For example, recent studies have shown that adversarial ML techniques can be used to bypass AI-driven encryption models by generating synthetic data that mimics normal traffic patterns, thereby fooling detection systems [57].

Moreover, adversarial attacks can target not just encryption algorithms but the entire AI-based cybersecurity framework. These attacks can render AI-based defenses ineffective by exploiting weaknesses in neural networks used for real-time threat detection [58]. For instance, Generative Adversarial Networks (GANs) have been employed to create realistic attack scenarios that deceive AI systems, making it harder for traditional encryption methods to safeguard data [59]. The increasing sophistication of adversarial AI raises the stakes for maintaining secure systems, requiring not only advancements in encryption but also in AI model robustness [60]. As these threats evolve, the integration of more secure AI models into encryption protocols will be vital for protecting sensitive information in the digital age.

Moreover, the widespread use of encryption technologies in sectors such as finance, healthcare, and national security brings with it significant ethical and legal challenges. Governments and regulatory bodies face the difficult task of balancing individual privacy rights with the need for surveillance to prevent criminal activities [61]. Encryption ensures that sensitive data remains confidential, but it also makes it harder for law enforcement agencies to access potentially crucial information [63]. As a result, there has been ongoing debate about the implementation of encryption backdoors, which would allow authorized entities to decrypt data under specific circumstances. However, these backdoors present a serious ethical dilemma, as they could be exploited by malicious actors if not properly secured [63]. As encryption continues to play a critical role in modern society, it will be essential for policymakers to develop clear, globally consistent frameworks that address both the ethical and legal challenges posed by these technologies [37]. In addition to the ethical concerns, encryption technologies also raise legal questions regarding jurisdiction and data ownership. As data crosses international borders, determining which country's laws apply to encrypted information becomes increasingly complicated [64]. For instance, different nations have varying regulations regarding data privacy and encryption standards, which can lead to conflicts when encrypted data is stored in one country but accessed or processed in another [65].

As encryption continues to play a critical role in modern society, it will be essential for policymakers to develop clear, globally consistent frameworks that address both the ethical and legal challenges posed by these technologies [37]. In addition to the ethical concerns, encryption technologies also raise legal questions regarding jurisdiction and data ownership. As data crosses international borders, determining which country's laws apply to encrypted information becomes increasingly complicated [62]. For instance, different nations have varying regulations regarding data privacy and encryption standards, which can lead to conflicts when encrypted data is stored in one country but accessed or processed in another [63].

VII. Discussion and Limitations

The integration of AI with encryption signifies a pivotal change in cybersecurity, offering both prospects and complex obstacles. The significance of AI in encryption has led to significant progress constantly in real-time threat detection as well as flexible security mechanisms, which are more vital in the contemporary linked and susceptible digital environment. This capacity allows encryption systems to promptly address abnormalities and emerging attack patterns, hence providing resilience unattainable by conventional static encryption approaches. Nonetheless, this progress entails an increasing dependence on machine learning as well as deep learning models, that, whilst augmenting encryption capabilities, can present weaknesses like adversarial assaults. These assaults target vulnerabilities in AI models using misleading inputs, compromising the precision and resilience of systems intended to identify and counter cyber threats. Thus, the dual-use characteristic of AI technology requires a measured and attentive strategy, especially in vital sectors such as healthcare, banking, and national security, wherein AI-driven encryption plays a crucial role in safeguarding extremely sensitive information.

Nonetheless, this progress is not without limitations. First, the over-reliance on AI systems for encryption may create blind spots, wherein undetected vulnerabilities can be exploited by

adversaries leveraging AI for malicious purposes. Second, the ethical and legal challenges surrounding AI-driven encryption, such as potential breaches of data privacy [62] and concerns about surveillance misuse, demand robust governance frameworks. These frameworks must include clear ethical guidelines and enforceable regulations to prevent the unintended misuse of AI-enhanced encryption technologies.

Additionally, accessibility disparities pose significant challenges. AI-driven encryption technologies, while offering scalable solutions, often require substantial technological resources and compliance capabilities. This raises questions about equity, as organizations with limited resources may struggle to implement these advanced systems effectively. Addressing such disparities is vital to ensuring the widespread and fair adoption of AI-powered encryption.

As such, the following validation approach should be used in future studies. In evaluating the proposed cryptographic solutions, employing a rigorous validation process establishes a credible foundation for comparative analysis and subsequent knowledge generation. This process begins with controlled laboratory testing, where encryption algorithms undergo quantitative benchmarking against standardized datasets, fixed key lengths, and pre-defined plaintext-ciphertext pairs to ensure reproducibility. By comparing time-to-encrypt, CPU utilization, memory usage, and latency across multiple cryptographic methods, researchers gain insights into both efficiency and scalability. The application of formal verification techniques, such as model checking and theorem proving, bolsters confidence in algorithmic correctness, ensuring that keys, modes of operation, and cipher primitives function as intended under a range of computational scenarios.

Beyond laboratory environments, field testing in distributed AI-driven systems delivers validation grounded in practical contexts, as real-time data streams reveal how well the chosen cryptographic methods withstand dynamic adversarial tactics. For comprehensive comparative analysis, conducting multi-criteria decision-making (MCDM) evaluations allows researchers to weigh performance metrics, security robustness, and resource overhead against one another. Statistical tests, including ANOVA or Wilcoxon signed-rank tests, further enhance credibility by confirming that observed differences in performance are significant and not attributable to random variation. Iterative refinement informed by validation feedback cycles contributes to continual improvement, bridging the gap between theoretical design and practical deployment. Through meticulous validation and comparison, the resulting cryptographic frameworks achieve a higher degree of reliability, fostering trust among stakeholders and ensuring that deployed solutions fulfill the intended security objectives in increasingly complex AI ecosystems.

Furthermore, the widespread use of AI-driven encryption systems raises urgent accessibility and ethical issues. Even while these technologies provide scalable solutions, their use in a variety of international businesses raises concerns about transparency and equality, particularly when firms have varying levels of technological resources as well as regulatory compliance skills. The fair distribution of these cutting-edge technologies must be given equal weight with technological resilience in the advancement of AI-powered encryption. These technologies also raise significant ethical and legal issues, including surveillance, data privacy, and the possible abuse of AI-enhanced encryption to provide vulnerabilities for illegal data access. To prevent AI-encrypted systems from unintentionally jeopardizing the same security and confidentiality they are meant to safeguard, strict governance structures and ethical standards must be established. Therefore, this open conversation covers both the enormous possibilities and the serious threats of AI-driven encryption, necessitating a thorough, interdisciplinary response to responsibly influence cybersecurity's future.

To sum up, the combination of encryption and artificial intelligence has brought about a new age in cybersecurity that offers increased resistance to a wide range of online dangers. AI-driven encryption is essential in today's fast-paced, data-intensive digital environment because of its adaptable, real-time features, which provide major benefits over conventional encryption methods. Homomorphic encryption and AI-enhanced algorithms are only two examples of the encryption techniques that have advanced because of this integration, strengthening data security and enabling sophisticated calculations on encrypted data. AI algorithms provide a strong defense against complex cyberattacks as they become better at managing the complexities of threat detection including adaptive encryption key management. However, this development raises fresh moral and legal issues. The ethical conundrums around privacy, transparency, and equality, together with the dual-purpose possibilities for AI technology, highlight the need for a concerted effort from all parties involved. To create moral guidelines including legal frameworks that encompass both the technical aspects of AI-enhanced encryption and its wider social ramifications, cooperation between government, business, and academia is crucial.

In the future, establishing a safe and flexible cybersecurity framework will require a proactive approach to the creation and management of AI-driven encryption systems. To reduce new dangers and protect sensitive data in a variety of industries, further research in fields like adversarial resilience, quantum-resistant encryption, and ethical AI will be essential. Through adopting this forward-thinking viewpoint, the cybersecurity industry can capitalize on AI's ability to develop encryption technologies while additionally making certain that those solutions are just, ethically appropriate, as well as resilient to the constantly changing cyberthreat scenario. In this sense, incorporating AI into encryption seems not just a technological development but also a step toward a digital future that is safe, sustainable, as well as considerate of privacy.

Although there are several issues in algorithms for both encryption and decryption, some of the major ones (in symmetric encryption and asymmetric encryption) are shown in Fig. 3 below. The challenges in algorithm generation, algorithm writing, and algorithm difficulty continues as the use of various language models including Artificial Intelligence (AI), Deep Learning (DL), and Machine Learning (ML) keeps growing.

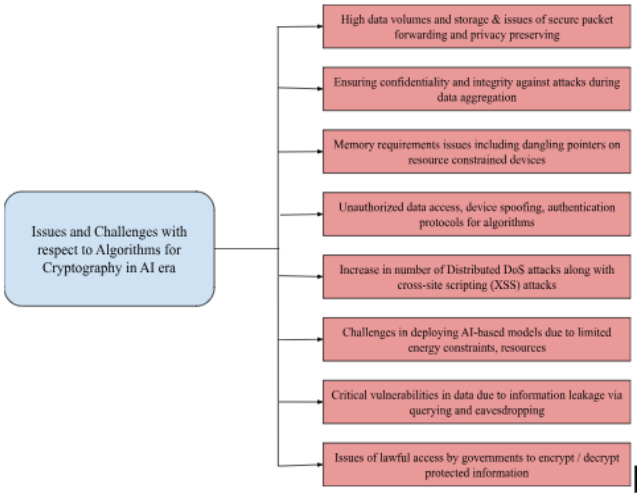


Fig. 3. Encryption algorithms (symmetric, asymmetric, and hybrid) securing digital communications.

VIII. Conclusion and Future Scope

We provided an in-depth study focusing on securing sensitive information with comparative analysis for symmetric encryption and asymmetric encryption algorithms for cryptography. The comparison on the study focuses on key factors like security resilience, scalability, speed in the light of evolving cyber threats. We have addressed the security concerns tackled by encryption algorithms in the Artificial Intelligence (AI) and Large Language Models (LLMs) age along with research directions to enhance overall cybersecurity and cryptography. The aspect comparison between symmetric encryption and asymmetric comparison allows us to decide the environments used including AI environments, key-pairs leveraging via secure key exchanges, and/or decision in secure protocols for secure data sharing.

Future scope of algorithms whether it be symmetric encryption and asymmetric encryption algorithms, largely rely upon use of AI models, reliability, scalability, and key management. Enabling machines to learn is always a future challenge that may require human intelligence in the next step for decision-making. As we progress into several AI algorithms in the future, all three types of learning (supervised learning, unsupervised learning, and reinforcement learning) we must be more intuitive in the future on how we process data and information.

References

[1] Thiyagarajan, P. (2020). A review on cyber security mechanisms using machine and deep learning algorithms. *Handbook of research on machine and deep learning applications for cyber security*, 23-41.

[2] Terumalasetti, S., & Reeja, S. R. (2022, August). A comprehensive study on review of AI techniques to provide security in the digital world. In *2022 third international conference on intelligent computing instrumentation and control technologies (ICICICT)* (pp. 407-416). IEEE.

[3] Al-Arjan, A., Rasmi, M., & AlZu'bi, S. (2021, July). Intelligent security in the era of AI: The key vulnerability of RC4 algorithm. In *2021 International Conference on Information Technology (ICIT)* (pp. 691-694). IEEE.

[4] Bertino, E., Kantarcioglu, M., Akcora, C. G., Samtani, S., Mittal, S., & Gupta, M. (2021, April). AI for Security and Security for AI. In *Proceedings of the Eleventh ACM Conf on Data and Appn Security and Privacy* (pp. 333-334).

[5] Kapoor, J., & Thakur, D. (2022). Analysis of symmetric and asymmetric key algorithms. In *ICT analysis and applications* (pp. 133-143). Springer.

[6] Soomro, S., Belgaum, M. R., Alansari, Z., & Jain, R. (2019, August). Review and open issues of cryptographic algorithms in cyber security. In *2019 Int Conf on Comp, Elect & Comm Engineering (iCCECE)* (pp. 158-162). IEEE.

[7] Ustun, T. S., Hussain, S. S., Yavuz, L., & Onen, A. (2021). Artificial intelligence-based intrusion detection system for IEC 61850 sampled values under symmetric and asymmetric faults. *Ieee Access*, *9*, 56486-56495.

[8] Arora, S. (2022). A review on various methods of cryptography for cyber security. *Journal of Algebraic Statistics*, *13*(3), 5016-5024.

[9] Henriques, M. S., & Vernekar, N. K. (2017, May). Using symmetric and asymmetric cryptography to secure communication between devices in IoT. In *2017 International Conf on IoT and Application (ICIOT)* (pp. 1-4). IEEE.

[10] Yaji, S., Bangera, K., & Neelima, B. (2018, December). Privacy preserving in blockchain based on partial homomorphic encryption system for AI applications. *25th Int Conf on HPC Workshops (HiPCW)* (pp. 81-85). IEEE.

[11] Arulmurugan, L., Thakur, S., Dayana, R., Thenappan, S., Nagesh, B., & Sri, R. K. (2024, May). Advancing Security: Exploring AI-driven Data Encryption Solutions for Wireless Sensor Networks. In *2024 Int Conf on Advances in Comp, Comm and Applied Informatics (ACCAI)* (pp. 1-6). IEEE.

[12] Xu, D., Li, G., Xu, W., & Wei, C. (2023). Design of artificial intelligence image encryption algorithm based on hyperchaos. *Ain Shams Engineering Journal*, *14*(3), 101891.

[13] Dharmateja, M., Rama, P. K., Asha, N., Nithya, P., Lalitha, S., & Manojkumar, P. (2024, March). Innovative Data Encryption Techniques using AI for Wireless Sensor Actuator Network Security. In *2024 Int Conf on Distributed Comp and Optimization Techniques (ICDCOT)* (pp. 1-6). IEEE.

[14] Hamza, R. (2023, October). Homomorphic Encryption for AI-Based Applications: Challenges and Opportunities. In *2023 15th International Conference on Knowledge and Systems Engineering (KSE)* (pp. 1-6). IEEE.

[15] Budhewar, A., Bhumgara, S., Tekavade, A., Nandkar, J., & Zanwar, A. (2024, April). Enhancing Data Security through the Synergy of AI and AES Encryption: A Comprehensive Study and Implementation. In *2024 MIT Art, Design and Tech Sch of Comp Int Conf (MITADTSoCiCon)* (pp. 1-5). IEEE.

[16] Tian, H., Yuan, Z., Zhou, J., & He, R. (2024). Application of Image Security Transmission Encryption Algorithm Based on Chaos Algorithm in Networking Systems of Artificial Intelligence. In *Image Processing, Electronics and Computers* (pp. 21-31). IOS Press.

[17] Abd Elminaam, D. S., Abdual-Kader, H. M., & Hadhoud, M. M. (2010). Evaluating the performance of symmetric encryption algorithms. *Int. J. Netw. Secur.*, *10*(3), 216-222.

[18] Al-Shabi, M. A. (2019). A survey on symmetric and asymmetric cryptography algorithms in information security. *International Journal of Scientific and Research Publications (IJSRP)*, *9*(3), 576-589.

[19] Panda, M. (2016, October). Performance analysis of encryption algorithms for security. In *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)* (pp. 278-284). IEEE.

[20] Hintaw, A. J., Manickam, S., Karuppayah, S., Aladaileh, M. A., Aboalmaaly, M. F., & Laghari, S. U. A. (2023). A robust security scheme based on enhanced symmetric algorithm for MQTT in the Internet of Things. *IEEE Access*, *11*, 43019-43040.

[21] Kuznetsov, O., Poluyanenko, N., Frontoni, E., & Kandiy, S. (2024). Enhancing Smart Communication Security: A Novel Cost Function for Efficient S-Box Generation in Symmetric Key Cryptography. *Cryptography*, *8*(2), 17.

[22] Halewa, A. S. (2024). Encrypted AI for Cyber security Threat Detection. *International Journal of Research and Review Techniques*, *3*(1), 104-111.

[23] Negabi, I., El Asri, S. A., El Adib, S., & Raissouni, N. (2023). Convolutional neural network based key generation for security of data through encryption with advanced encryption standard. *International Journal of Electrical & Computer Engineering* (2088-8708), 13(3).

[24] Rehan, H. (2024). AI-Driven Cloud Security: The Future of Safeguarding Sensitive Data in the Digital Age. *Journal of Artificial Intelligence General science (JAIGS)* ISSN: 3006-4023, 1(1), 132-151.

[25] Rangaraju, S. (2023). Ai sentry: Reinventing cybersecurity through intelligent threat detection. *EPH-International Journal of Science And Engineering*, *9*(3), 30-35.

[26] Saha, A., Pathak, C., & Saha, S. (2021). A Study of Machine Learning Techniques in Cryptography for Cybersecurity. *American Journal of Electronics & Communication*, *1*(4), 22-26.

[27] Yanamala, A. K. Y., & Suryadevara, S. (2023). Advances in Data Protection and Artificial Intelligence: Trends and Challenges. *International Journal of Advanced Eng Technologies and Innovations*, *1*(01), 294-319.

[28] Feisheng, L. (2024, April). Systematic Review of Sentiment Analysis: Insights Through CNN-LSTM Networks. In *2024 5th Int Conference on Industrial Engineering and Artificial Intelligence (IEAI)* (pp. 102-109). IEEE.

[29] Pacheco, J., Benitez, V. H., Felix-Herran, L. C., & Satam, P. (2020). Artificial neural networks-based intrusion detection system for internet of things fog nodes. *IEEE Access*, *8*, 73907-73918.

[30] Tashfeen, M. T. A. (2024). Intrusion detection system using AI and machine learning algorithms. In *Cyber security for next-generation computing technologies* (pp. 120-140). CRC Press.

[31] Mallick, M. A. I., & Nath, R. (2024). Navigating the Cyber security Landscape: A Comprehensive Review of Cyber-Attacks, Emerging Trends, and Recent Developments. *World Scientific News*, *190*(1), 1-69.

[32] Alionsi, D. D. D. (2023). AI-driven cybersecurity: Utilizing machine learning and deep learning techniques for real-time threat detection, analysis, and mitigation in complex IT networks. *Advances in Eng Innovation*, *3*, 27-31.

[33] Orner, C., & Chowdhury, M. M. (2024). AI and Cybersecurity: Collaborator or Confrontation. *Proceedings of 39th Int Confer*, *98*, 150-158.

[34] Jimmy, F. N. U. (2024). Cyber security Vulnerabilities and Remediation Through Cloud Security Tools. *JAIGS*, ISSN: 3006-4023, 2(1), 129-171.

[35] Gupta, A., Wright, C., Ganapini, M. B., Sweidan, M., & Butalid, R. (2022). State of AI ethics report (vol 6, feb 2022). *arXiv preprint arXiv:2202.07435*.

[36] Riebe, T. (2023). Dual-Use and Trustworthy? A Mixed Methods Analysis of AI Diffusion between Civilian and Defense R&D. In *Technology Assessment of Dual-Use ICTs: How to Assess Diffusion, Governance and Design* (pp. 93-110). Wiesbaden: Springer Fachmedien Wiesbaden.

[37] Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From chatgpt to threatgpt: Impact of generative ai in cybersecurity and privacy. *IEEE Access*.

[38] Evren, R., & Milson, S. (2024). The Cyber Threat Landscape: Understanding and Mitigating Risks. *Tech. rep. EasyChair*.

[39] Morley, J., & Floridi, L. (2020). An ethically mindful approach to AI for health care. *The Lancet*, *395*(10220), 254-255.

[40] Javadpour, A., Ja'fari, F., Taleb, T., Zhao, Y., Bin, Y., & Benzaïd, C. (2023). Encryption as a service for IoT: opportunities, challenges and solutions. *IEEE Internet of Things Journal*.

[41] Gupta, A., Royer, A., Heath, V., Wright, C., Lanteigne, C., Cohen, A., Ganapini, M., Fancy, M., Galinkin, E., Khurana, R., Akif, M., Butalid, R., Khan, F., Sweidan, M., (2020). The State of AI Ethics Report. *arXiv, abs/2011.02787*.

[42] Thabit, F., Can, O., Aljahdali, A. O., Al-Gaphari, G. H., & Alkhzaimi, H. A. (2023). Cryptography algorithms for enhancing IoT security. *Internet of Things*, *22*, 100759.

[43] Kuppuswamy, P., Al, S. Q. Y. A. K., John, R., Haseebuddin, M., & Meeran, A. A. S. (2023). A hybrid encryption system for communication and financial transactions using RSA and a novel symmetric key algorithm. *Bulletin of Electrical Engineering and Informatics*, *12*(2), 1148-1158.

[44] Pandey, P. K., Kansal, V., & Swaroop, A. (2023). Security challenges and solutions for next-generation VANETs: an exploratory study. In *Role of Data-Intensive Distributed Computing Systems in Designing Data Solutions* (pp. 183-201). Cham: Springer International Publishing.

[45] Akter, R. I. M. A., Khan, M. A. R., Rahman, F. A. R. D. O. W. S. I., Soheli, S. J., & Suha, N. J. (2023). RSA and AES based hybrid encryption technique for enhancing data security in cloud computing. Int. J. Comp. *Appl. Math. Comput. Sci*, *3*, 60-71.

[46] Liu, Y., Gong, W., & Fan, W. (2018). Application of AES and RSA Hybrid Algorithm in E-mail. *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, 701-703. https://doi.org/10.1109/ICIS.2018.8466380.

[47] Subramanian, A., Donta, L. S., & Supraja, P. (2024, May). Assessing the Strength of Hybrid Cryptographic Algorithms: A Comparative Study. In *2024 Int Conf on Intelligent Systems for Cybersecurity (ISCS)* (pp. 1-6). IEEE.

[48] Rahman, M. M., Kshetri, N., Sayeed, S. A., & Rana, M. M. (2024). AssessITS: Integrating procedural guidelines and practical evaluation metrics for organizational IT and cybersecurity risk assessment. Journal of Information Security, 15(4), 564–588. https://doi.org/10.4236/jis.2024.154032

[49] Singh, S., Sharma, P. K., Moon, S. Y., & Park, J. H. (2024). Advanced lightweight encryption algorithms for IoT devices: survey, challenges and solutions. Journal of Ambient Intelligence and Humanized Computing, 1-18.

[50] Siwakoti, Y. R., Bhurtel, M., Rawat, D. B., Oest, A., & Johnson, R. C. (2023). Advances in IoT security: Vulnerabilities, enabled criminal services, attacks, and countermeasures. *IEEE Int of Things Jou*, *10*(13), 11224-11239.

[51] Zhou, J., Cao, Z., Dong, X., & Vasilakos, A. (2017). Security and Privacy for Cloud-Based IoT: Challenges. *IEEE Communications Magazine*, *55*, 26-33. https://doi.org/10.1109/MCOM.2017.1600363CM

[52] Hammi, B., Zeadally, S., & Nebhen, J. (2023). Security threats, countermeasures, and challenges of digital supply chains. *ACM Computing Surveys*, *55*(14s), 1-40.

[53] Bagla, P., Sharma, R., Mishra, A., Tripathi, N., Dumka, A., & Pandey, N. (2023). An Efficient Security Solution for IoT and Cloud Security Using Lattice-Based Cryptography. *Int Conf on Eme Trends in Net and Comp Comm (ETNCC)*, 82-87. https://doi.org/10.1109/ETNCC59188.2023.10284931.

[54] Ahmed, S., & Khan, M. (2023). Securing the Internet of Things (IoT): A comprehensive study on the intersection of cybersecurity, privacy, and connectivity in the IoT ecosystem. *AI, IoT & the 4th Ind Rev Rev*, *13*(9), 1-17.

[55] Kshetri, N., Rahman, M. M., Sayeed, S. A., & Sultana, I. (2024). cryptoRAN: A review on cryptojacking and ransomware attacks w.r.t. banking industry - Threats, challenges, & problems. 2nd InCACCT (pp. 523–528). IEEE. https://doi.org/10.1109/InCACCT61598.2024.10550970

[56] Craighero, F., Angaroni, F., Stella, F., Damiani, C., Antoniotti, M., & Graudenzi, A. (2023). Unity is strength: Improving the detection of adversarial examples with ensemble approaches. *Neurocomputing*, *554*, 126576.

[57] Shroff, J., Walambe, R., Singh, S. K., & Kotecha, K. (2022). Enhanced security against volumetric DDoS attacks using adversarial machine learning. *Wireless Communications and Mobile Computing*, *2022*(1), 5757164.

[58] Sathupadi, K. (2023). Ai-based intrusion detection and ddos mitigation in fog computing: Addressing security threats in decentralized systems. *Sage Science Review of Applied Machine Learning*, *6*(11), 44-58.

[59] Zhang, C., Yu, S., Tian, Z., & Yu, J. J. (2023). Generative adversarial networks: A survey on attack and defense perspective. *ACM Computing Surveys*, *56*(4), 1-35.

[60] Fernando, P., & Wei-Kocsis, J. (2021). A Novel Data Encryption Method Inspired by Adversarial Attacks. *ArXiv, abs/2109.06634.*

[61] Allahrakha, N. (2023). Balancing cyber-security and privacy: legal and ethical considerations in the digital age. *Legal Iss in the Dig Age*, (2), 78-121.

[62] van Daalen, O. L. (2023). The right to encryption: Privacy as preventing unlawful access. *Computer Law & Security Review, 49*, 105804.

[63] Taddeo, M., McCutcheon, T., & Floridi, L. (2019). Trusting artificial intelligence in cybersecurity is a double-edged sword. *Nature Machine Intelligence*, 1-4. https://doi.org/10.1038/s42256-019-0109-1.

[64] Lubin, A. (2023). The prohibition on extraterritorial enforcement jurisdiction in the datasphere. In *Research Handbook on Extraterritoriality in International Law* (pp. 339-355). Edward Elgar Publishing.

[65] Nguyen, M. T., & Tran, M. Q. (2023). Balancing security and privacy in the digital age: an in-depth analysis of legal and regulatory frameworks impacting cybersecurity practices. *Int Jou of Int Auto & Comp*, *6*(5), 1-12.

A Framework for Privacy-Preserving Detection of Sickle Blood Cells Using Deep Learning and Cryptographic Techniques

Kholoud Alotaibi, Naser El-Bathy
Computer Science and Engineering, Oakland University Rochester, MI, USA

***Abstract*—Sickle cell anemia is a hereditary disorder where abnormal hemoglobin causes red blood cells to become rigid and crescent-shaped, obstructing blood flow and leading to severe health complications. Early detection of these abnormal cells is essential for timely treatment and reducing disease progression. Traditional screening methods, though effective, are time-intensive and require skilled technicians, making them less suitable for large-scale implementation. This paper presents a conceptual framework that integrates transfer learning, cryptographic algorithms, and service-oriented architecture to provide a secure and efficient solution for sickle cell detection. The framework uses MobileNet, a lightweight deep learning model, enhanced with transfer learning to identify sickle cells from medical images while operating on hardware-constrained environments. Advanced Encryption Standards (AES) ensure sensitive patient data remains secure during transmission and storage, while a service-oriented architecture facilitates seamless interaction between system components. Although not yet implemented, the framework serves as a foundation for future empirical testing, addressing the need for accurate detection, data privacy, and system efficiency in healthcare applications.**

Keywords*—Sickle cells; deep learning; transfer learning; encryption; AES; SOA

I. Introduction

Sickle cell disease (SCD) is a genetic disorder that significantly impacts the shape and function of red blood cells. Under normal conditions, red blood cells are round and flexible, allowing them to move smoothly through blood vessels but they become rigid and crescent-shaped due to defective hemoglobin, the protein responsible for carrying oxygen throughout the body, in individuals with sickle cell disease, [1, 2]. This abnormal shape causes the cells to get trapped in small blood vessels, blocking blood flow, which leads to a range of serious health complications such as pain, organ damage, and an increased risk of infections [3]. Early detection of sickle cells is very important because timely treatment can reduce these complications and greatly enhance the quality of life for those affected [4].

Even though early detection is important, traditional diagnostic methods are often complex and require highly skilled technicians, making them unsuitable for large-scale screening, particularly in low-resource settings [5]. These limitations slow down the ability to diagnose SCD sufficiently early to prevent severe health consequences. Moreover, depending on manual analysis in traditional diagnostics creates challenges in terms of speed and accessibility.

Recent advancements in artificial intelligence (AI), particularly in deep learning, have shown significant potential for automating medical image analysis with high accuracy [6]. These technologies offer faster and more efficient solutions, enabling early detection and scalable screening even in challenging settings. However, deploying AI in medical applications introduces unique challenges. For instance, ensuring patient data privacy is critical, given the sensitive nature of medical data and the strict regulations governing its use, such as HIPAA in the United States and GDPR in Europe [7]. Additionally, many deep learning models require substantial computational resources for training and inference, which limits their feasibility in environments with constrained hardware resources [8].

This paper proposes a novel framework that combines deep learning, cryptographic algorithms, and service-oriented architecture (SOA) to address these challenges. The framework uses MobileNet, a lightweight deep learning model optimized for efficient operation on hardware-constrained systems, to detect sickle cells in blood smear images. Transfer learning is employed to achieve high detection accuracy without requiring extensive computational resources. Advanced encryption techniques, such as AES, ensure patient data remains secure during transmission and storage, addressing critical privacy concerns. Furthermore, SOA enables seamless communication between system components, enhancing the system's scalability, modularity, and flexibility.

By integrating these components, the proposed framework offers a secure, efficient, and practical solution for sickle cell detection in diverse healthcare settings.

The remainder of this paper is organized as follows: Section II identifies the problem, and Section III reviews related work, focusing on previous efforts in using AI for sickle cell detection and employing cryptographic techniques to secure medical data. Section IV provides details on the proposed framework, explaining the roles of each component and how they are integrated. Section V illustrates the workflow of the system, showing how data is securely processed from start to finish. Section VI discusses the security and privacy considerations involved in the framework. Section VII outlines the feasibility and limitations of the approach, while Sections VIII and IX provide future directions and conclusions, respectively.

II. Problem Identification

A. Research Problem

Creating an automated detection system for sickle cells comes with several challenges, one of the most important being data privacy. Medical data is very sensitive and needs to be handled according to strict regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in Europe [7]. Ensuring patient privacy is very important when using AI models which require large amounts of data for training and validation. Additionally, deploying deep learning models can require more computational resources, which is challenging in environments where advanced hardware is not easily available [8]. Therefore, there is a need for a solution that balances accuracy, privacy, and efficiency to provide a secure and practical way to detect sickle cells.

B. Research Questions

How to build a system that can correctly detect sickle cells, keep patient data private, and work well with limited hardware? How to create a solution that meets the increasing need for automated medical testing, follows important privacy rules like HIPAA and GDPR, and performs well without needing expensive hardware? These are the key challenges to solve in creating a reliable and easy-to-use system for sickle cell detection.

C. Objective

The study objective is to develop a secure and efficient framework for detecting sickle cells by integrating deep learning, cryptographic algorithms, and an SOA. This framework uses the capabilities of deep learning for analyzing medical images while ensuring that patient data remains protected through cryptographic methods. The use of SOA makes the framework modular, meaning that each component—such as encryption, model inference, and data communication—can be flexible, scalable, and easily integrated into existing healthcare systems.

D. Significance

The significance of this proposed framework is its novel integration of AI with secure computation techniques designed to address healthcare needs. The framework handles the critical issue of privacy-preserving AI in healthcare by combining deep learning for medical image analysis with cryptographic methods for secure data handling [9]. Furthermore, the use of SOA allows for flexibility and scalability, making the solution adaptable to different clinical environments. This combination of accurate detection, data security, and scalability represents a unique approach to addressing the challenges of sickle cell detection to provide accessible high-quality care while protecting patient data.

III. Related Work

A. Existing Deep Learning Solutions for Sickle Cell Detection

Deep learning has brought remarkable improvements to medical imaging, especially in automating the detection of diseases like SCD. Traditionally, diagnosing SCD requires visually examining blood smears under a microscope, which is not only time-consuming but also prone to human error. To overcome these challenges, researchers have turned to deep learning models, which make the process faster and more reliable.

Goswami and colleagues used deep neural networks like ResNet50 and GoogleNet to classify sickle cells from digital blood smear images. They also applied explainable AI techniques, such as Grad-CAM to make the predictions easier for healthcare professionals to understand. This added transparency makes the diagnostic process more trustworthy. ResNet50 was the best-performing model, achieving an accuracy of 94.9%, which shows great potential for real-world clinical use [11].

Kawuma and his team compared different deep-learning techniques for detecting SCD, such as VGG16, VGG19, and Inception V3, demonstrating that Inception V3 achieved the highest accuracy at 97.3%, followed by VGG19 at 97.0%. These results highlight the effectiveness of pre-trained models for accurately identifying sickle cells [10].

Karunasena and colleagues took a different approach, using a region-based convolutional neural network (R-CNN) to detect sickle cells. Their model achieved over 90% accuracy and was particularly useful in segmenting and classifying specific regions within an image. This level of precision is important when dealing with complex blood smear images where cells may be crowded or overlap [12].

Another review by Balde et al. focused on recent advances in using AI to detect SCD, highlighting image segmentation and feature extraction techniques. These methods, particularly CNN, have been successful in analyzing microscopic images, even when they contain overlapping cells. However, there are still challenges in improving the robustness of these segmentation techniques to accurately distinguish between normal and sickled cells, especially in more complex or densely populated images [13].

Regardless of these advancements, one key issue remains largely unaddressed, data privacy. Most studies have focused primarily on improving detection accuracy but have not paid enough attention to the sensitive nature of medical data. This gap provides an opportunity for future research to explore privacy-preserving techniques, such as encryption, to protect patient information while maintaining the diagnostic accuracy of deep learning solutions.

In summary, existing research shows significant progress in using deep learning for sickle cell detection but there is still a need for better handling of overlapping cells, and a stronger focus on data privacy to create a comprehensive solution that can be widely adopted.

B. Cryptographic Techniques in Medical Data Security

Securing medical data is important, especially when using AI and machine learning (ML) models, as these often require sensitive patient information for processing. Cryptographic techniques have been a core part of ensuring that this data remains safe throughout its lifecycle—whether during transmission, storage, or even analysis.

One interesting approach is the MASS framework which uses blockchain technology to securely share medical data collected from wearable IoT devices. In this system, health information is encrypted using Ciphertext-Policy Attribute-Based Encryption (CP-ABE), which means that only authorized individuals can access specific parts of the data. This method not only keeps the data private but also ensures it cannot be altered, thus MASS is particularly effective for protecting data from wearable medical devices [14].

Amaithi Rajan and colleagues developed a secure way to retrieve medical images from encrypted cloud storage using a combination of deep learning and encryption techniques, ensuring that the images are not only safely stored but also easily retrievable when needed. This method relies on symmetric encryption and searchable encryption, allowing users to search through encrypted data and retrieve relevant images without compromising security [15].

Ahmad Al Badawi and Mohd Faizal Bin Yusof took another interesting approach by using fully homomorphic encryption (FHE). This allows medical data to be processed without ever decrypting it, meaning patient information remains secure even during analysis. They used this method for privacy-preserving pathological assessments using support vector machines (SVMs). This approach provides both strong data protection and effective analysis, making it highly suitable for privacy-sensitive medical diagnostics [16].

A different system developed by Kusum Lata and her team focused on detecting brain tumors using deep learning while ensuring privacy through encryption. They used the AES-128 algorithm to encrypt medical images before storage or transmission to keep patient data secure, even during the diagnosis. This system is a good example of balancing the need for secure data handling with the advanced diagnostic capabilities that AI offers [17].

Lastly, Runze Wu and colleagues developed a privacy-preserving system that used Gaussian kernel-based support vector machines (SVMs) and a simpler cryptographic method known as additive secret sharing. This approach is less computationally intensive compared to more advanced methods like homomorphic encryption, making it better suited for real-time applications. It ensures that both the patient's data and the healthcare provider's model are kept private during the diagnostic process, all while maintaining efficiency [18].

Overall, these cryptographic approaches highlight the importance of keeping patient data secure while using the power of AI in healthcare. While robust encryption methods like homomorphic encryption offer high security, they can be quite demanding in terms of computation. Hence, lightweight cryptography and blockchain-based methods are becoming attractive alternatives because they balance effective security with practical resource use—especially important in healthcare, where computational power is often limited.

C. Service-Oriented Architectures in Healthcare

The SOA has become popular in healthcare because it allows systems to be broken down into smaller, independent services that work together efficiently. This approach makes healthcare technology more flexible and scalable—perfect for dealing with the complexities of modern medical institutions.

Petrenko and Boloban explain how SOA can handle the increasing volume of healthcare data and provide efficient treatment by coordinating different services. By splitting a large healthcare system into smaller, interacting services, SOA improves how data is processed and how medical resources are managed, ultimately leading to better patient care. A major advantage is that new services can be easily added to the system without causing disruptions, making SOA both scalable and adaptable [19].

Liviu Ilie and colleagues looked at how SOA could be used to create a framework that connects electronic health records with other healthcare services to boost interoperability—the ability of different systems to exchange and use information. Interoperability is important in healthcare, especially for large medical institutions that need to integrate multiple systems. SOA provides a framework that helps different services communicate effectively, improving both the quality and efficiency of healthcare. This approach makes the system more flexible, easier to upgrade, and less expensive to maintain [20].

Similarly, Petrenko and Tsymbaliuk developed a cloud-based healthcare platform called "Clinic in Cloud" that uses SOA to bring together wearable sensors, data management systems, and user interfaces. This platform allows doctors to monitor patients remotely in real time, making diagnosis and treatment more accessible. The SOA approach ensures that all these different components, sensors, databases, and communication tools, work smoothly together, enabling timely and effective patient care. The modular design also makes it easy to add new features as healthcare needs develop [21].

In summary, SOA provides a strong foundation for building healthcare systems that are flexible and scalable. By breaking down complex systems into smaller, independent services, SOA makes integration easier, data management more efficient, and services more adaptable. This is especially valuable in healthcare, where technology is continuously changing and systems need to keep up.

D. Summary of Related Work

The research on deep learning, cryptographic techniques, and SOA has greatly advanced healthcare technology but there are still areas for improvement.

Deep learning models like ResNet50 and Inception V3 have made diagnosing SCD faster and more reliable compared to traditional manual methods. However, data privacy has not been addressed, a major issue when dealing with sensitive health information.

Methods like FHE and blockchain frameworks have been used to keep patient data secure and while they are highly effective, they can be computationally heavy, limiting their practicality in real-time applications or environments with limited computing resources.

The SOA has been used to make healthcare systems more modular and scalable. SOA improves integration between different healthcare tools and makes it easier to expand systems

when needed by breaking down large systems into smaller, easier-to-manage services.

The proposed framework builds on these existing solutions by combining the strengths of each approach while addressing their limitations. By integrating lightweight deep learning with efficient cryptographic techniques like AES and an SOA-based design, the framework prioritizes patient data privacy, operational efficiency, and system scalability. Unlike current solutions, this approach places patient privacy at the forefront while ensuring adaptability to evolving healthcare needs.

IV. Proposed Framework

A. Overview of the Framework

The proposed framework aims to provide a secure and efficient way to detect sickle cells in medical images. It consists of three main components that work together to ensure accuracy, privacy, and flexibility:

- A deep learning model for detection
- A cryptographic module
- A service-oriented architecture (SOA)

A pre-trained deep learning model, MobileNet [23], was employed to identify sickle cells in blood smear images. This model correctly detects sickle cells using transfer learning without requiring significant computational power, making it ideal for setups with limited hardware. The preprocessing steps include resizing the input images and normalizing pixel values to ensure compatibility with the MobileNet architecture. Performance evaluation metrics, such as detection accuracy and inference latency, will be used during the empirical validation phase to measure the effectiveness of the model.

Advanced Encryption Standards (AES) were used to encrypt the medical images before analysis to protect sensitive patient information. This ensures that the data remains confidential during both transmission and storage, providing strong security to prevent unauthorized access. Although AES introduces some computational overhead, its efficiency makes it a practical choice for environments with limited resources, balancing security, and performance. Beyond encryption, the framework is designed to align with data protection regulations such as HIPAA and GDPR, ensuring secure and compliant handling of patient data throughout the system.

SOA connects all the components, allowing perfect communication between the cryptographic module, the deep learning model, and the other system parts. Each function—such as encryption, analysis, and reporting—is treated as an independent service, thus the system is modular and scalable so components can be updated or replaced without disrupting the rest of the system, making it easy to expand or adapt as needed.

Fig. 1 illustrates how these components interact to visualize the flow of the system. Additionally, Table I provides a detailed overview of each system component, outlining its function and purpose within the framework.

B. Components

1) Deep learning model for detection: MobileNet was the transfer learning model selected for detecting sickle cells as it is specifically designed for environments with limited resources, making it ideal for training on a CPU. It employs a technique called depthwise separable convolutions [25], reducing the number of parameters and computational complexity without sacrificing accuracy.

MobileNet V2 is highly effective for different medical imaging tasks. For example, it achieved accuracy rates as high as 94% when used for brain tumor classification after being pre-trained and fine-tuned on relevant datasets [22].

This suggests that MobileNet can also perform well for sickle cell detection tasks.

MobileNet has also been applied for breast cancer classification, delivering fast execution times even on devices with limited computational resources without compromising accuracy [23]. This makes it an ideal choice for scenarios where only a CPU is available for training and inference.

MobileNet has also been successfully integrated into ensemble models for detecting conditions like cardiomegaly, showing that it is robust and works well in combination with other models [24].

TABLE I. System Components Overview

Component	Function	Purpose
Data Input and Encryption	Uploads and encrypts medical images using AES	Protect patient data during transmission
Service-Oriented Architecture (SOA)	Connects system components and manages secure data flow	Ensures modularity and scalability
Deep Learning (MobileNet)	Analyzes medical images to detect sickle cells	Provides accurate detection of sickle cells
Data Storage/ Transmission	Encrypts and securely stores or sends results to the user	Maintains data privacy throughout the process

This flexibility demonstrates its capability to handle complex medical imaging challenges effectively.

Its lightweight architecture makes it ideal for training and deploying on a CPU, which fits the hardware limitations of the proposed framework. Unlike heavier models like ResNet or VGG, MobileNet requires far less computational power while still delivering strong performance [25]. Additionally, pre-trained weights can be used and fine-tuned on the sickle cell dataset, allowing for efficient training even without high-end hardware.

2) Cryptographic module: AES was chosen for its well-known efficiency and security when encrypting large datasets like medical images. AES supports different key lengths (e.g., 128-bit or 256-bit) and provides an excellent balance between speed and security, making it suitable for both storing and transmitting medical data securely [33].

AES has been successfully used in cloud-based medical data systems to secure sensitive patient information. It ensures that data remains protected during transmission and storage, which is important for maintaining privacy in healthcare settings [26].

In comparison studies, AES was faster than other encryption methods for both encrypting and decrypting data, making it a reliable option for handling medical images without slowing down performance [27].

AES is also effectively used in combination with techniques like watermarking [26] to guarantee both the security and integrity of medical images. This dual functionality shows the flexibility of AES in ensuring data remains protected while preserving image quality, which is essential in healthcare.

AES offers strong encryption to protect patient information, ensuring confidentiality during transmission and storage. It efficiently secures large medical datasets without slowing down the system, thus is ideal for healthcare settings where both security and efficiency are important [26].

3) Service-oriented architecture (SOA): In this framework, SOA plays a key role in enabling the different components—such as the cryptographic module, deep learning model, and data management processes—to communicate smoothly. Each function, like encryption, model analysis, and data sharing, operates as an independent service, making the system more flexible and scalable, allowing each part to be developed, updated, and managed separately.

- Benefits of Using SOA

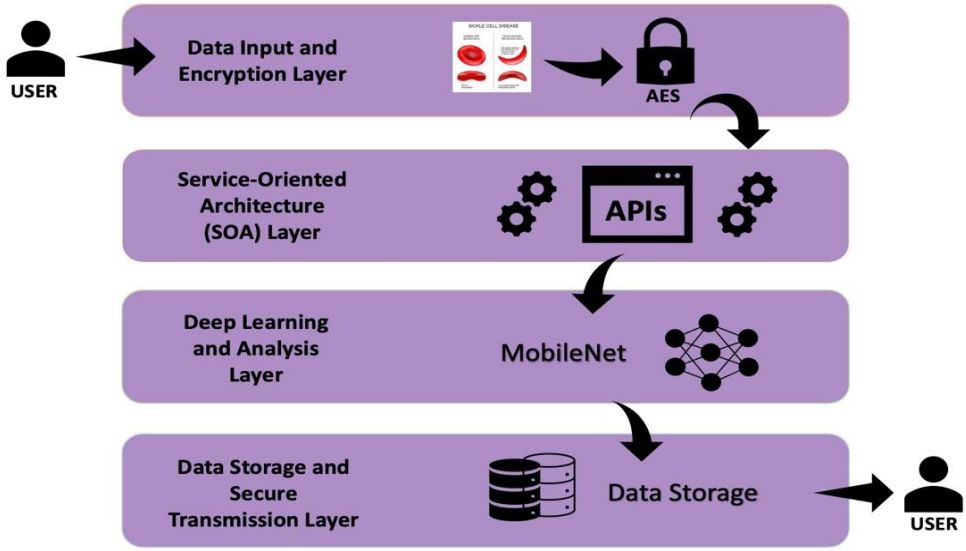


Fig. 1. High-level overview of the proposed framework showing the flow of data between components. Encrypted data is processed through the service-oriented architecture (SOA), analyzed by the Deep Learning Model (MobileNet), and securely transmitted and stored.

a) Modularity: SOA breaks down the system into smaller, reusable services so that each part, like data encryption, detection, or reporting, can be independently maintained. This modular approach makes it much easier to update or modify components without affecting the entire system. Essentially, the rest of the system can keep running smoothly if one part needs an upgrade [20, 28].

b) Interoperability: The framework includes different components such as cryptographic modules and AI models, and SOA makes sure they work well together. By using standardized protocols like SOAP and REST, SOA ensures that these services can easily communicate, even if they are built on different platforms or by different developers. This is important for smooth integration and effective communication between all system parts [28].

c) Scalability: Healthcare systems often need to handle growing amounts of data and SOA-based systems are inherently scalable, meaning new services can be added without disrupting the existing setup. This is particularly useful when integrating with cloud-based services or IoT devices for real-time monitoring, ensuring that the system can grow and adapt as needed [21].

- How SOA Enables Communication Between Modules:

SOA-based frameworks often use middleware and APIs to manage how different services interact. This not only helps standardize the flow of data but also supports plug-and-play integration, making it easy to add new components, such as an improved encryption method or an updated AI model, without causing disruptions to the rest of the system [28].

V. System Workflow

This section describes how data moves through each stage of the proposed framework, from the initial encryption of medical images to the secure return of prediction results. Each step is designed to ensure data privacy, accuracy, and efficiency.

The process starts by encrypting the medical images using AES [29] to ensure that sensitive patient data is protected from the beginning and remains confidential throughout transmission and processing.

Once encrypted, the data is securely transmitted to the deep learning service using an SOA. SOA allows the encrypted data to be passed easily between different services, making communication between the cryptographic module and the deep learning model smooth and secure [30]. This modular design ensures that the different system parts work together effectively without compromising sensitive information.

When the encrypted data reaches the deep learning model, it is first decrypted for analysis and then processed by the MobileNet [22] deep learning model to detect sickle cells in the medical images. This step generates predictions, helping identify any abnormalities in the blood smear images.

After the analysis is complete, the results are encrypted using AES to ensure privacy before they are sent back to the client. This step keeps the prediction results protected during transmission, maintaining the confidentiality of patient data at all times [30].

Fig. 2 illustrates the entire workflow, showing the interactions between each component:

1) Data encryption: Medical images are encrypted using AES for privacy.

2) Data transmission via SOA: Encrypted data is sent to the deep learning service.

3) Model inference: The deep learning model processes the data (after decryption) and generates predictions.

4) Result encryption and return: The analysis results are encrypted and returned securely to the client.

VI. Security and Privacy Analysis

The proposed framework includes different security and privacy protections to keep patient data safe and private. This section explains how privacy issues are managed, what security features are used, and how the framework balances being secure while still running efficiently.

A major concern with medical data is protecting patient privacy. The framework handles this by using AES encryption to secure medical images at every step. By locking the data before analysis and again when sending the results back, the system ensures that no one without permission can access sensitive patient information during transmission or storage. Using AES helps prevent data breaches that could compromise patient privacy in line with established privacy standards like HIPAA and GDPR which require strict protection for medical data to prevent unauthorized access [31].

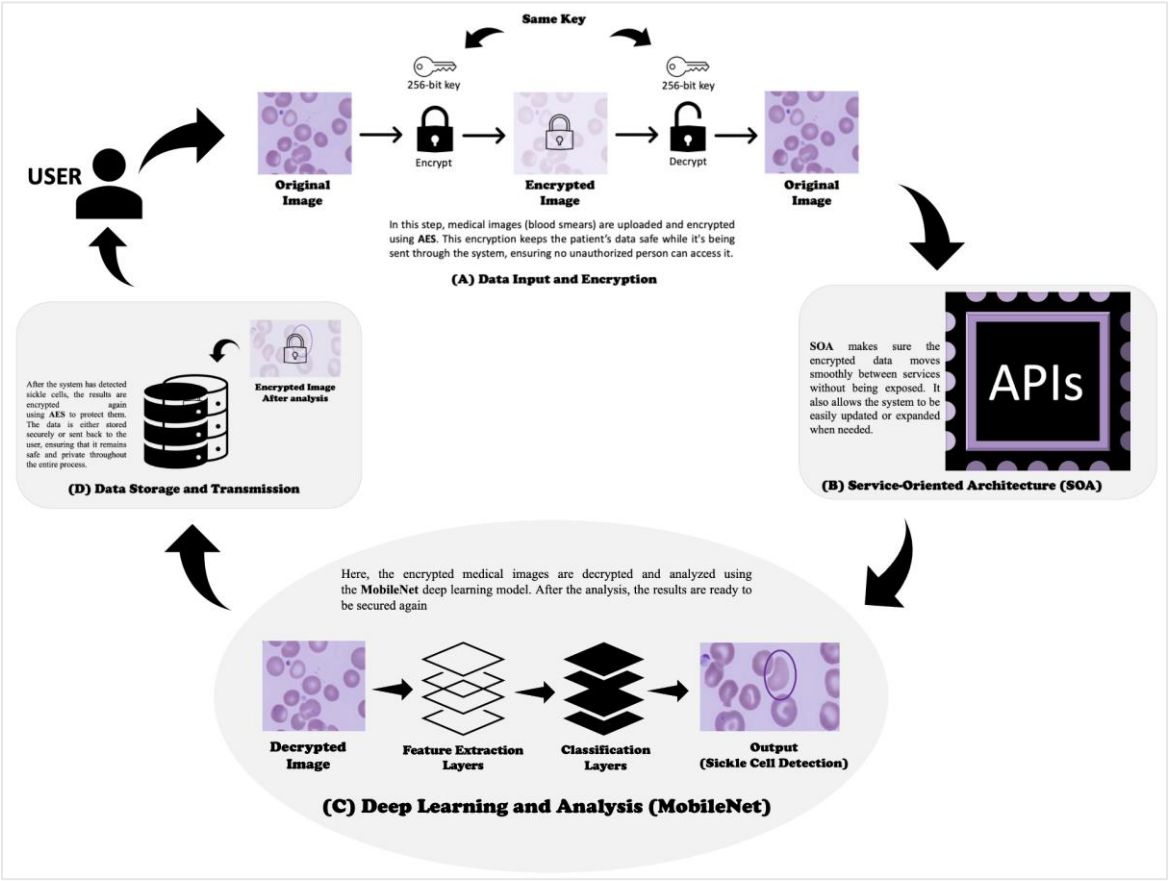


Fig. 2. System workflow diagram illustrating the flow of data within the proposed framework. The process begins with AES encryption of medical images, followed by secure transmission through SOA, analysis using the MobileNet model, and encryption of results before returning to the client. Each step ensures privacy, accuracy, and secure data handling.

The framework uses AES and other cryptographic algorithms to keep data secure and intact throughout the process. AES encrypts the data both when it is being sent and after analysis, providing:

- Confidentiality: AES keeps patient data private from the moment it is collected until the results are returned.
- Integrity: Encryption ensures that if anyone tries to tamper with the data during transmission, it becomes unreadable, keeping it accurate.

The framework also uses an SOA to safely transfer encrypted data between services, adding extra protection by reducing the risk of data exposure during transfers [30].

While security is very important, it is also necessary to consider how encryption affects system performance. AES encryption performs well at keeping data safe but can slow systems down, especially those with limited resources or using large amounts of data. Nonetheless, it is still more efficient than other options like RSA. AES finds a good middle ground between security and speed. It works well with large datasets, like medical images, because it provides strong protection without using too much computational power compared to heavier encryption methods [32].

1) Optimized workflow: Using the MobileNet model for analysis allows the framework to work efficiently, even on systems with just a CPU [23]. This helps lower the overall workload of both the encryption and the AI tasks, making the system both secure and practical for real-world healthcare use.

VII. Feasibility and Limitations

The framework is designed to work efficiently, even on systems with limited hardware like a standard CPU. This is possible through the use of MobileNet [24] with transfer learning. MobileNet is a lightweight deep learning model that reduces the computational workload by using fewer parameters compared to heavier models. By starting with a pre-trained model and only adjusting the final layers, the framework can achieve high accuracy without requiring advanced hardware, making it practical for environments with limited resources.

While the framework offers a balance between security and efficiency, a few challenges and limitations need to be addressed:

- Resource Limitations: Even though MobileNet is optimized for efficiency, using a CPU for processing may still be slower compared to a GPU, particularly when handling large datasets. This could limit the framework's scalability when analyzing a high volume of images.
- Balancing Privacy and Efficiency: There is a trade-off between ensuring data privacy and maintaining fast processing speeds. AES encryption provides strong data protection, but it adds steps to the workflow, which could slow down real-time performance.

This paper proposes a theoretical framework, and its implementation has not yet been carried out. While the framework integrates validated techniques such as MobileNet for deep learning, AES for encryption, and SOA for modularity, empirical testing will be conducted in future work. Metrics such as accuracy, latency, encryption strength, and scalability will be used to validate the framework's performance. The absence of implementation results reflects the current focus on the framework design, which serves as a foundation for future development and testing.

VIII. Future Work

The current goal is to design the framework and create a basic prototype to show how it works. In the future, the full system will be built, with all parts working together and tested in real-world situations. This full implementation will help identify practical limitations to improve the system's performance.

Once the system is fully built, a detailed evaluation of its performance will be performed, assessing the accuracy (how well it finds sickle cells) and speed (how long it takes to process each image) of the deep learning model, as well as the strength of the encryption (to keep patient data safe) and how efficiently the system runs (the encryption does not slow the system too much). The goal of this evaluation is to find the right balance between security, accuracy, and speed.

Another important future task is to determine how well the framework can handle larger datasets, like more medical images from different sources, to help understand how the system can be used in real-world healthcare, where securely and efficiently managing large amounts of data is essential.

IX. Conclusion

A. Summary

This paper presented a framework to help detect sickle cells in medical images securely and efficiently. The framework comprises three main parts: a deep learning model to identify sickle cells, a cryptographic module to protect patient data, and an SOA for efficient communication between all system parts. Together, these parts ensure that patient information stays safe during every step—from encryption to analysis and sending back the results.

B. Contribution

The main contribution of this work is creating a unique framework that combines deep learning and cryptographic techniques using an SOA-based structure. This design makes it possible to securely analyze medical images while maintaining patient confidentiality. Using MobileNet and transfer learning, the framework is also efficient for environments with limited hardware, making it useful in more healthcare settings.

C. Implications

This work is important for the future of secure AI in healthcare. This framework provides a way to develop AI solutions that not only work well but also protect patient privacy by combining strong encryption with deep learning in a scalable system. This approach can be applied to other types of medical imaging, improving diagnostics while keeping data safe—something that is becoming more important in today's digital healthcare world.

REFERENCES

[1] Ministry of Health Saudi Arabia, https://www.moh.gov.sa/en/Pages/Default.aspx.

[2] NIH National Heart, Lung, and Blood Institute https://www.nhlbi.nih.gov/health/sickle-cell-disease.

[3] Tanabe P, Spratling R, Smith D, Grissom P, Hulihan M. CE. (2019), Understanding the complications of sickle cell disease. Am J Nurs. 119(6):26-35. doi: 10.1097/01.NAJ.0000559779.40570.2c.

[4] Nationwide Children https://www.nationwidechildrens.org/family-resources-education/family-resources-library/early-diagnosis-key-to-dealing-with-sickle-cell-disease#:~:text=State%20laws%20require%20that%20babies,devastating%20complications%20of%20the%20disease.

[5] Dexter, D, McGann, PT (2022), saving lives through early diagnosis: the promise and role of point of care testing for sickle cell disease. Br J Haematol, 196: 63-69. https://doi.org/10.1111/bjh.17678.

[6] Bushra SN, Shobana, G (20121), Paediatric sickle cell detection using deep learning: A Review. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, pp. 177-183, doi: 10.1109/ICAIS50930.2021.9395756.

[7] Seun Solomon Bakare, Adekunle Oyeyemi Adeniyi, Chidiogo Uzoamaka Akpuokwe, & Nkechi Emmanuella Eneh. (2024). Data privacy laws and compliance: A comparative review of the EU GDPR and USA regulations. Computer Science & IT Research Journal, 5(3), 528-543. https://doi.org/10.51594/csitrj.v5i3.859.

[8] Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, Joel T Dudley (2018), Deep learning for healthcare: review, opportunities and challenges, Briefings in Bioinformatics, 19(6): 1236–1246, https://doi.org/10.1093/bib/bbx044.

[9] Priyanka, Singh, A.K (2023), A survey of image encryption for healthcare applications. Evol. Intel.16, 801–818. https://doi.org/10.1007/s12065-021-00683-x.

[10] Kawuma Simon, Mabirizi Vicent, Kyarisiima Addah, David Bamutura, Barnabas Atwiine, Deborah Nanjebe, Adolf Oyesigye Mukama, (2023), Comparison of deep learning techniques in detection of sickle cell disease. Artificial Intelligence and Applications,

[11] Neelankit Gautam Goswami, Anushree Goswami, Niranjana Sampathila, Muralidhar G. Bairy, Krishnaraj Chadaga, Sushma Belurkar. (2024), Detection of sickle cell disease using deep neural networks and explainable artificial intelligence. Journal of Intelligent Systems,

[12] G.M.K.B. Karunasena, H.M.K.K.M.B. Herath, H.D.N.S. Priyankara, B.G.D.A. Madusanka, (2023), Sickle cell disease identification by using region with convolutional neural networks (R-CNN) and digital image processing, Sri Lankan Journal of Applied Sciences,

[13] Abdourahmane Balde, Avewe Bassene, Lamine Faty, Mamadou Soumboundou, Ousmane Sall, Youssou Faye, (2023), Recent artificial intelligence advances in detection and diagnosis of sickle cell disease: A review, 2023 IEEE International Conference on Big Data (BigData).

[14] L. Chen et al., MASS: A multi-attribute sketch secure data sharing scheme for IoT wearable medical devices based on blockchain, IEEE Internet of Things Journal, doi: 10.1109/JIOT.2024.3468733.

[15] Amaithi Rajan, A., V, V., Raikwar, M. et al. (2024) SMedIR: secure medical image retrieval framework with ConvNeXt-based indexing and searchable encryption in the cloud. J Cloud Comp 13, 139. https://doi.org/10.1186/s13677-024-00702-z.

[16] Al-Badawi A, Faizal Bin Yusof M, (2024), Private pathological assessment via machine learning and homomorphic encryption. BioData Mining 17, 33 . https://doi.org/10.1186/s13040-024-00379-9.

[17] Lata K, Singh, P, Saini, S, Cenkeramaddi, LR, Deep learning-based brain tumor detection in privacy-preserving smart health care systems. IEEE Access, doi: 10.1109/ACCESS.2024.3456599.

[18] Wu R, Wang B, Zhao Z (2024), Privacy-preserving medical diagnosis system with Gaussian kernel-based support vector machine. Peer-to-Peer Netw. Appl.. https://doi.org/10.1007/s12083-024-01743-6.

[19] Petrenko, Anatolii/Boloban, Oleh (2023), Generalized information with examples on the possibility of using a service-oriented approach and artificial intelligence technologies in the industry of e-Health. Technology Audit and Production Reserves 4 (2/72), S. 10 - 17. https://journals.uran.ua/tarp/article/download/285935/280167/660189. doi:10.15587/2706-5448.2023.285935.

[20] Ilie L, Pop E, Caramihai SI, Moisescu MA, (2022), A SOA-based e-health services framework. E-Health and Bioengineering Conference (EHB), Iasi, Romania, 2022, pp. 1-4, doi: 10.1109/EHB55594.2022.9991608.

[21] Petrenko A, Tsymbaliuk R, (2024), A cloud-based platform ("Clinic in Cloud") as a significantly expanding the current capabilities of the Ukraine e-health system. EC Clinical and Medical Case Reports 7.9: 01-10.

[22] Arfan, TH, Hayaty M, Hadinegoro A (2021), Classification of brain tumours types based on MRI images using Mobilenet. 2nd International Conference on Innovative and Creative Information Technology (ICITech), Salatiga, Indonesia, 2021, pp. 69-73, doi: 10.1109/ICITech50181.2021.9590183.

[23] Ahmadi, Mahdie, Nader Karimi, and Shadrokh Samavi. "A lightweight deep learning pipeline with DRDA-Net and MobileNet for breast cancer classification." arXiv preprint arXiv:2403.11135 (2024).

[24] H. Atyam, S. C. Bachu, S. S. Kyatham and H. Satish, "Screening of Cardiomegaly using Ensemble Model of InceptionV3 and MobileNet," 2024 10th International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, India, 2024, pp. 1426-1431, doi: 10.1109/ICCSP60870.2024.10543609.

[25] Sarthak Joshi, Rachit Shah, Yashvi Chandola, Vivek Uniyal. "Classification of Brain MRI Images using End-to-End Trained AlexNet & End-to-End Pre-Trained MobileNet " International Journal of Research Publication and Reviews, Vol 5, no 7, pp 205-218 July 2024.

[26] Pendam T, Baig MM, Sonekar S, Sawwashere SS (2024), Enhancing security and privacy in cloud-based medical data systems using AES cryptography and digital envelopes. IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), Bhopal, India, pp. 1-8, doi: 10.1109/SCEECS61402.2024.10482070.

[27] Bowen Meng, Xiaochen Yuan, Qiyuan Zhang, Chan-Tong Lam, Guoheng Huang, (2024), Encryption-then-embedding-based hybrid data hiding scheme for medical images. Journal of King Saud University - Computer and Information Sciences, 36(1), 101932, ISSN 1319-1578, https://doi.org/10.1016/j.jksuci.2024.101932.

[28] Giao J, Nazarenko AA, Luis-Ferreira F, Gonçalves D, Sarraipa JA (2022), Framework for service-oriented architecture (SOA)-based IoT application development. Processes 2022, 10, 1782. https://doi.org/10.3390/pr10091782.

[29] Aravind Kumar Kalusivalingam, (2020), Advanced encryption standards for genomic data: evaluating the effectiveness of AES and RSA, AJST 3(1).

[30] Vasilescu E, Mun, SK (2006), Service Oriented Architecture (SOA) implications for large scale distributed health care enterprises. 1st Transdisciplinary Conference on Distributed Diagnosis and Home Healthcare, 2006. D2H2., Arlington, VA, USA, 2006, pp. 91-94, doi: 10.1109/DDHH.2006.1624805.

[31] Seh AH, Zarour M, Alenezi M, Sarkar AK, Agrawal A, Kumar R, Ahmad Khan R (2020). Healthcare Data Breaches: Insights and Implications. Healthcare 8, 133. https://doi.org/10.3390/healthcare8020133

[32] Hafsa A, Malek J, Machhout M (2021), Performance trade-offs of hybrid cryptosystem for medical images encryption – decryption. 18th International Multi-Conference on Systems, Signals & Devices (SSD), Monastir, Tunisia, pp. 1221-1229, doi: 10.1109/SSD52085.2021.9429477.

[33] Bhavitha M, Rakshitha K, Rajagopal SM (2024), Performance evaluation of AES, DES, RSA, and Paillier Homomorphic for image security. 9th International Conference for Convergence in Technology (I2CT), Pune, India, 2024, pp. 1-5, doi: 10.1109/I2CT61223.2024.10544282.

Trustworthiness in Conversational Agents: Patterns in User Personality-Based Behavior Towards Chatbots

Jieyu Wang[1]*, Merary Rangel[2], Mark Schmidt[3],Pavel Safonov[4]
Information Systems, Saint Cloud State University, St. Cloud, USA

***Abstract*—As artificial intelligence conversational agent (CA) usage is increasing, research has been done to explore how to improve chatbot user experience by focusing on user personality. This work aims to help designers and industrial professionals understand user trust related to personality in CAs for better human-centered AI design. To achieve this goal, the study investigates the interactions between users with diverse personalities and AI chatbots. We measured participant personalities with a Hogan and Champagnes (1980) typology assessment by categorizing personality dimensions into the extraversion vs. intuition (EN), extraversion vs. sensing (ES), introversion vs. intuition (IN), and introversion vs. sensing (IS) groups. Twenty-nine participants were assigned two tasks to engage with three different AI chatbots: Cleverbot, Kuki, and Replika. Their conversations with the chatbots were analyzed using the open-coding method. Coding schemes were developed to create frequency tables. Results of this study showed that EN personality participants had perceptions of high trustworthiness towards the chatbot, especially when the chatbot was helpful. The ES personality participants, on the other hand, often engaged in brief conversations regardless of whether the chatbot was helpful or not, leading to low trust levels towards the chatbot. The IN personality users experienced mixed outcomes; while some had perceived trusty-worthy conversations despite having unhelpful chatbot responses, others found helpful conversations, yet a perception of low trustworthiness. The IS personality participants typically had the longest conversations, often leading to high perceptions of high trust scores being given to the chatbots. This study indicates that users with diverse personalities have different perceptions of trust toward AI conversational agents. This research provides interpretations of different personality users' interaction patterns and trends with chatbots for designers as design guidelines to emphasize AI UX design.**

Keywords*—Trust; personality; human-centered AI design; user experience

I. Introduction

The use of Artificial Intelligence has increasingly been popular not only within the technological world but in business, health research, psychological aspects, supply-chain management, education, decision-making, science research, and financial aspects as well [1, 2]. Artificial Intelligence was popularized by Alan Turing, when he released a journal article proposing a question if Machines think, thus having people realize the difference in what machines can do when given instructions, versus making a decision based on the facts they are given [3]. McCarthy described artificial intelligence as "the science and engineering of making intelligent machines", popularizing the term and the widespread use of artificial intelligence [4]. Artificial intelligence has increased to mimic human tasks, such as conversations, information processing, educational assistants, computing, and more recently, predictive algorithm models [5, 6, 7].

Artificial Intelligence has been used by businesses to help automation and increase quality in customer satisfaction by personalizing experiences [8]. Among AI applications, CAs are becoming popular due to their purpose of serving customers. Recent studies have analyzed CAs' characters to categorize them [9]. Since CAs' main functions are to help users gather information and make decisions, designing how to better serve people with different personalities to enhance user experience is the key [10,11].

Thus, this work aims to explore the different trustworthiness in user behavior between the personalities of end users and AI chatbots within interactions. We also provide user behavior interpretations of patterns and trends for designers as design guidelines to better engage users with different personalities when they interact with AI chatbots.

II. Related Work

Artificial intelligence was first used to help humans, such as playing checkers or helping organize tasks for the ease of human labor and the human mind. Yet, due to it being highly used, there have been ethical questions between AI and the intersectionality of life, such as employment, politics, and educational aspects, with different personality types having different attitudes towards artificial intelligence. [12, 13 14].

According to a study done by Kaya et al. (2024), people who are less "computer-literate" tend to have negative attitudes toward artificial intelligence. Kaya explains that this population may not have the knowledge or the experience of using computing-based algorithms, and therefore, is worried that artificial intelligence can one day take tasks assigned to humans and have these tasks automated. People who have a higher education level and a higher use of computers tend to have more positive attitudes toward artificial intelligence. Those who have more positive attitudes towards artificial intelligence view it as a tool, rather than a burden. Those who also had positive attitudes were more open to new experiences. Yet, those with a higher education level, and higher computer literate level did report that users would have to keep up with the technology. This concept of "having to keep up" is known as self-actualization, a psychological theory of improving oneself, to become the better version of what the current mind stands once all other needs are met [15, 16]. By having a positive attitude, students can improve their technology literacy and continue self-improvement.

This aligns with another study conducted by Zhou et al. (2019) where they found that younger computer engineering students were able to feel comfortable using artificial intelligence to conversate in an interview. Not only were they comfortable, but they were able to trust the artificial chatbot to feel more assertive, outgoing, and being themselves rather than a human. The students' attitudes towards the artificially intelligent chatbot were seen as more positive rather than negative and were able to be the best version of themselves when they felt trust, and agreeableness with the chatbot [17].

Another study conducted by Heng Li (2023) showed similar results, when testing personality traits of intellectual humility and attitudes towards artificial intelligence. Li found that students in a Chinese university who scored higher on Intellectual Humility on a personality test tended to result in favorable use of artificial intelligence and accepted a form of generative AI called ChatGPT as an advantage. These students were also higher on an openness scale, meaning they were open to new experiences, and the use of newer artificial intelligence can be accommodating [18].

The Big Five is a psychological theory and model that models down all human personality traits into five categories, including Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism, or an acronym known as OCEAN. Stein et al. (2024) conducted research on the Dark Triad and the Conspiracy Mentality. The Dark Triad assessment focuses on Narcissism, Machiavellianism, and Psychopathy. The Conspiracy Mentality Questionnaire (CMQ), developed by Bruder et al., in 2013, is a questionnaire that measures attitudes in socio-political fields. In this research, Stein found that agreeableness was a key indicator in personality traits that had students being able to accept artificial intelligence as a positive tool. Yet, Stein found that those who have higher beliefs in conspiracy theories tended to have negative attitudes toward artificial intelligence [13].

The Big Five has been used widely in research to find trust attitudes towards artificial intelligence. Reidel et al. (2024) found that trust in artificial intelligence is related to the user's personality, with individuals who are more open to experience (Openness) are more likely to have a positive correlation with their experience in artificial intelligence. However, Sharan & Romano (2020) yielded results that indicate trust is a human factor, and trust towards artificial intelligent agents tends to be negatively viewed by individuals who are strongly associated with Neuroticism [19, 20].

Moreover, due to the rise of artificial intelligence, user experience in trust can be based on how reliable artificial intelligence is when interacting with humans. While personality is an influence on decision-making, this result indicates that artificial intelligent agents can be not a service to humans [21].

Personality varies from person to person, and it can be different in other cultures as well. In a recent study, researchers explored the relationship between trust in artificial intelligent agents and trust in humans. In this study, a culture that is advanced in technology has higher trust in artificial intelligence agents. Though this research study may not be applicable to all countries and generalized to every individual, it gives new insight into future research on how humans can engage more with artificial intelligent agents in trusting their algorithms [22].

Though the Big Five is commonly approached in studies, there is limited research on personality measured with Hogan and Champagne's (1980) Personal Style Inventory when measuring user experience with artificial intelligent agents [24]. The Hogan Assessment is used to predicate outcome of behavior, usually in a vocational setting. This work applies their Extroversion (E) v. Introversion (I) and Sensing (S) v. Intuition (N) personality dimensions to categorize the participants in the study. Extroversion is defined as individuals who are open to new ideas, see new opportunities, and are colorful in nature by engaging with meeting new people, and having new experiences. Extraversion counterpart is Introversion, where the individual feelings are from inward rather than outward [23, 24, 25, 26]. "Sensing" is a trait where the individual is focused on factual, and detailed oriented information, the counterpart is "Intuitive", where the individual trusts their instincts and is more abstract with their personality and thought process [27]. This paper aims to investigate the interaction between users' personalities on the Hogan and Champagne typology and the level of trust between three chatbots, Clever Bot, Kuki, and Replika.

III. METHOD

A. Participants

This study aims to explore diverse interaction patterns of users with different personalities when they communicate with AI chatbots. To achieve the research goal, we invited 29 participants who are college students to participate in the study. They are information systems major students who have a basic or moderate understanding of AI chatbots. Among them, 25 participants' data has been confirmed to be complete.

B. Procedures

As we previously described in our series of studies [28], the participants were recruited to chat with three CAs, Kuki, Replika, and Cleverbot in their own environments. These three chatbots were among the top ones that these participants preferred to interact with [29]. The participants were assigned the same two prompts for each of the three chatbots. One task was about travel planning and the other one concerned ordering food from restaurants.

Prompt 1: The spring break is coming. You are pretty interested in traveling. But you do not know where to travel. Please talk to each CA: 1) Kuki, 2) Replika, and 3) Cleverbot and gather enough information for you to create your travel itinerary (a detailed travel plan).

Prompt 2: Today you are too tired to cook. Also, you would like to explore restaurants. Talk to the three CAs and get your food.

The participants were required to record their conversation histories and rate each CA's response to their questions or interactions on a Word document within two weeks. We asked the participants to use a Likert scale of 7 (1=strongly untrustworthy, 2=untrustworthy, 3=moderately untrustworthy, 4=undecided, 5=moderately trustworthy, 6=trustworthy, 7=strongly trustworthy) to rate each CA response. The

participants were also required to provide reasons (written in text) for each response. This large amount of data allows us to explore the participants' extent of trust towards different chatbots' responses and the personality-based reasons explaining their behaviors.

C. Grounding Theory

Our previous study showed there were differences in the task accuracies of users with different personality dimensions when they interacted with a CA [29]. The study analysis is based on Hogan and Champagne's (1980) personality dimension matrix (Table I): introversion VS extroversion (IE), intuition VS sensing (NS) [24]. The dimensions provided the best middle ground as we categorized the participants into four major personality dimensions accordingly (EN, ES, IN, IS) while adding some detailed personality analysis of the 16 groups such as ISTJ, ENFJ, etc.).

D. Hierarchical Data Structure

The researchers of this study organized the data into a hierarchical data structure, where we analyze the dimensions of personalities and chatbots within a personality root (Fig. 1). EN represents the Extraversion and Intuition Dimensions, ES represents the Extraversion and Sensing Dimensions, IN represents the Introversion and Intuition Dimensions, and IS represents the Introversion and Sensing Dimensions.

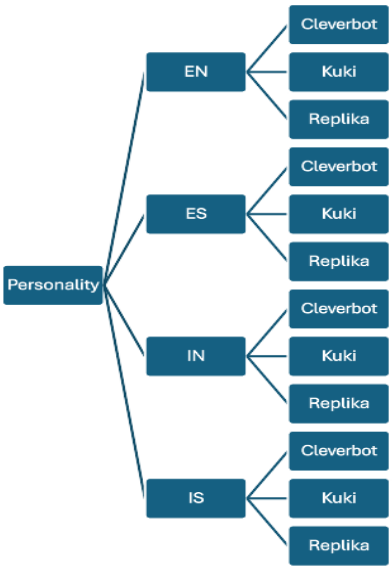


Fig. 1. Hierarchical data structure.

E. Coding Schemes

The researchers explored the data by conducting open coding and analyzing all conversations between participants and chatbots manually. We identified three coding categories for chatbot responses: Helpful, Unhelpful, and Neutral. These categories were derived from previous literature. 'Helpful' responses are those that facilitate meaningful and engaging exchanges [30, 37]. 'Unhelpful' responses occur when the conversation deviates from its purpose, such as when the chatbot fails to provide the desired information [31]. Lastly, 'Neutral' responses are neither helpful nor unhelpful from the chatbot, but the participant continued the conversation." The researchers went on to further explore the number of occurrences in exchanges between the participant and the chatbot, which led to the identification of a new coding category, labeled as Number of Interactions. In total, we were able to identify four coding categories, which are Helpful, Unhelpful, Neutral, and Number of Interactions. By coding the interactions, we were able to examine the perceptions of chatbots by the participants. Average of Trust is a variable in which trust levels were calculated by aggregating each user's ratings and averaging them by the number of interactions between the chatbots. This approach will allow for a comparison of perception influenced by the nature and frequency of interactions, providing insights into how Helpful, Unhelpful, and Neutral responses impact the overall user experience.

In the Helpful category, this code is exhibited when the chatbot is helpful, and interactive with the user, leading to a meaningful and pleasant experience with the user. Often, the chatbot gives the user helpful responses in their request or prompts. Below is an example of a "Helpful" Interaction in this study between a user and the Kuki Chatbot.

> **Participant Question:** "I want to travel next week but I am not sure where to travel to. Do you have any recommendations?"
> **Chatbot Kuki Response:** "I would advise you to go to Japan. It's so beautiful and there are so many amazing places to visit".

Within the "Unhelpful" category the chatbot is being unhelpful, such as being unable to help the user based on the prompt from the user, or being tangential, when the chatbot starts to be erratic and hostile towards the user. Below is an example of an "Unhelpful Interaction" in this study.

> **Participant Question**: "I am looking for a place to travel in United States in April. Can you recommend me some place?"
> **Chatbot Cleverbot Response**: "Lol"

"Neutral" exchanges occur when the chatbot is neither "Helpful" nor "Unhelpful", such as saying "Ok". These conversations don't contribute much but still have the user continue the conversation. Below is an example of a "Neutral Interaction" in this study.

> **Participant Question**: "I need it [the meal] here, not in Lasha"
> **Chatbot Replika:** "I understand".

"Number of Interactions" refers to how many exchanges took place between the user and the chatbot. It is the count of interactions that occurred and then summed. Below are three interactions between a participant and the Replika chatbot.

> **Participant Question**: "Hi! Who are you?"
> **Chatbot Replika:** "I am your personal AI companion.
> **Participant Question:** "Can you help me plan for spring break."

Chatbot Replika: "Sure, let me help you. I love this name! Let's take a moment to make your experience more personal."
Participant Question: "Maybe Later"
Chatbot Replika: "OK! Let's talk :) So, how are you doing?"

"Average Trust" represents the mean trust score calculated across all interactions between a user and the chatbot. For each interaction, a trust score is assigned based on the user's perception of the chatbot's response. These individual scores are then aggregated and averaged to provide a single "Average Trust" value that can represent the whole interaction. The formula is as follows:

$$\text{Average Trust (Mean)} = \frac{\text{Sum of Trust Scores for All Interactions}}{\text{Total Number of Interactions}}$$

Below is an example of an Average Trust of an interaction between a participant and Cleverbot.

Participant Question	Chatbot Cleverbot Response	Score Given
"I am looking for a place to travel in United States in April. Can you recommend me some place?"	"Lol"	1
"What's that place?"	"Its in the slender woods."	1
		Average Score = 1

F. Frequency Table with Average Trust

In this study, we tracked the frequency of each chatbot's interactions within each category and also put the average trust level score to indicate the average score of the whole conversation. There are two prompts, a meal prompt, indicated by "MP" and a travel prompt, indicated by "TP". This will lead to two occurrences of every personality in our graphs. Based on our hierarchical data structure, we split the frequencies into three tables for each dimension of personality. Table I shows an example of a frequency table from the EN dimension and Kuki chatbot. The frequency table exhibits how many occurrences of each category occurred during an interaction. For example, Table I indicates that within the "Helpful" row, there were six helpful exchanges, zero "Unhelpful" exchanges, and one "Neutral" exchange between the chatbot Replika, and the participant. In total, there were seven exchanged interactions, and the average Trust level was aggregated to be 6.3, which according to our Likert Scale, the participant perceived the chatbot Kuki to be trustworthy.

TABLE I. FREQUENCY AND AVERAGE TRUST TABLE

Participant, Personality Categories, and Prompt Type.	Helpful	Unhelpful	Neutral	Number of Interactions	Average Trust Level
P1 ENFJ TP	6	0	1	7	6.3
P2 ENFJ TP	3	3	3	9	4.5
P3 ENFP TP	9	2	0	11	6.9
P1 ENFJ MP	2	2	0	4	2.75
P2 ENFJ MP	5	2	0	7	5.1
P3 ENFP MP	5	2	1	8	6.5

IV. RESULTS

The results indicate mixed results between different personalities and the perception of chatbots. We will go over each personality and their results between the chatbot.

A. The EN Personality and Frequencies

Table II presents the descriptive statistics for EN personality users across the three chatbots. "N" represents the total number of conversations, and "n" denotes the number of participants. The mean trust level is 5.05 (SD = 1.36), while the mean number of interactions is 8.33 (SD = 2.54), indicating participants had about eight conversations on average. Helpful responses occurred 4-5 times on average (mean = 4.72, SD = 2.69). Unhelpful responses averaged 2.61 (SD = 2.73), showing moderate variation, while Neutral responses occurred about 1-2 times (mean = 1.17, SD = 1.34).

According to Fig. 2, the frequency data suggests that users with "ENFJ" personality gave low trust scores to Cleverbot when Cleverbot's responses were unhelpful, as identified through the researchers' coding schemes. Despite this, the participants with "ENFJ" personality continued to have the conversation even when it failed to provide the desired information. The participant with the personityal of "ENFP" continued to give Cleverbot chatbot moderately high trust scores, even when conversations were less meaningful. This aligns with Hogans and Champagne's Personal Style Inventory Typology (1980), where the personality type ENFPs are known to be high-spirited, extremely ingenious, more likely to have the ability to be imaginative, and often do whatever they feel like they want to do, [24]. In contrast, participants in this study with the personality type of ENFJ were more likely to find the conversation less meaningful if Cleverbot did not address their prompts effectively, reflecting their responsiveness to their environment and sense of responsibility [24].

TABLE II. DESCRIPTIVE STATISTICS FOR THE EN PERSONALITY DIMENSION

	Mean	Std. Deviation
Trust	5.0597	1.35981
Interaction	8.33	2.544
Helpful	4.72	2.697
Unhelpful	2.61	2.725
Neutral	1.17	1.339

a. *N =18
**n = 3

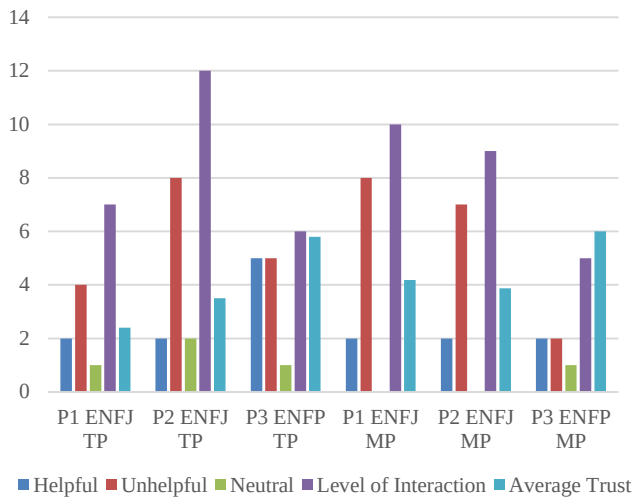


Fig. 2. Frequencies in cleverbot AI and EN personality.

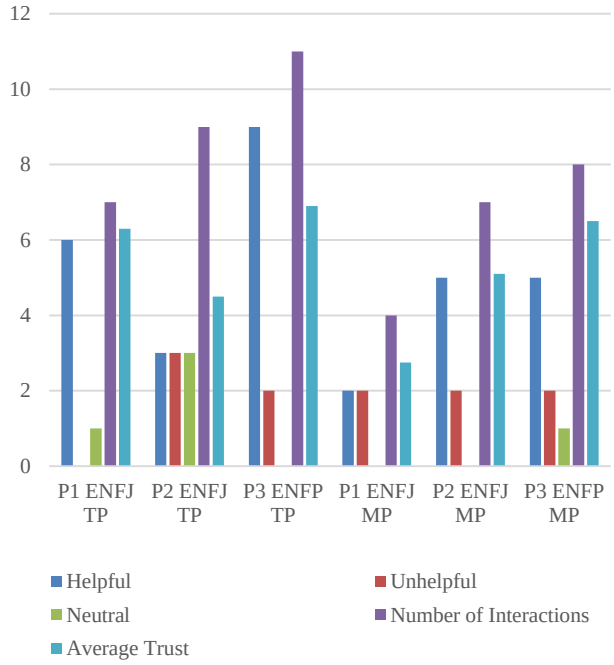


Fig. 3. Frequencies in kuki AI and EN personality.

According to Fig. 3, both participants with ENFJ personalities engaged in extended conversations with the Kuki chatbot. Through our careful analysis of their interactions, we determined that "Helpful" was the most frequently observed code across both conversations from ENFJ and the Kuki chatbot. The participant with an ENFP personality type also had meaningful conversations with Kuki and consistently gave it high scores, even when the chatbot's responses were neutral or unhelpful. This behavior aligns with Hogan and Champagne's Personal Style Inventory Typology (1980), which describes ENFPs as imaginative, adaptable, and optimistic in their interactions [24].

For ENFJ participants, conversations with Kuki involved fewer unhelpful interactions compared to Cleverbot. In the travel prompt scenario, Kuki AI received higher trust ratings, suggesting it was more effective in assisting users with their prompts. This indicates that Kuki's performance likely made it more appealing to users, earning significantly higher trust scores compared to Cleverbot.

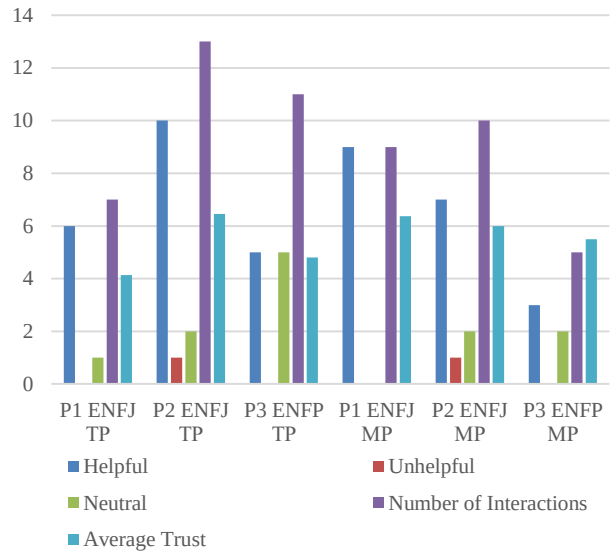


Fig. 4. Frequencies in replika AI and EN personality.

Fig. 4 illustrates the frequencies of each personality and the codes that were identified in their interactions. According to Fig. 4, Participant One (P1)-ENFJ showed fewer interactions and a moderately high average trust score during both prompts, indicating that the participant was able to obtain the information they needed, and ended the conversation when they were satisfied, leading to a moderate trust score given to the chatbot, deeming it trustworthy according to our Likert scale. P2-ENFJ yielded similar results. However, the participant with Personality ENFP had short conversations across both prompts and gave this chatbot a lower trust score compared to Cleverbot and Kuki. This pattern suggests that this participant may require more stimulating and engaging interactions to maintain longer conversations and find the chatbot trustworthy. By observing these behaviors, we can align it with Hogan and Champagne's Personal Style Inventory Typology (1980), where ENFJs are more responsible and ENFPs seek to be dynamic and explorative [24].

B. The ES Personality

Table III shows descriptive statistics for ES personality users across the three chatbots. The mean trust level is 4.06 (SD = 1.52), indicating some variation in trust among participants. The average number of interactions is 9.13 (SD = 6.97), suggesting high variability in how often participants engaged. Helpful responses occurred about four times on average (mean = 4.00, SD = 2.96), while unhelpful responses averaged 3.60 (SD = 4.02). Neutral responses were less frequent, averaging 1.73 (SD = 1.99), with some variation across participants. "N" represents 17 conversations, and n represents the number of participants within the ES personality dimension.

TABLE III. DESCRIPTIVE STATISTICS FOR ES PERSONALITY

	Mean	Std. Deviation
Trust	4.0642	1.52310
Interaction	9.13	6.966
Helpful	4.00	2.961
Unhelpful	3.60	4.018
Neutral	1.73	1.987

a.*N = 17

**n = 7

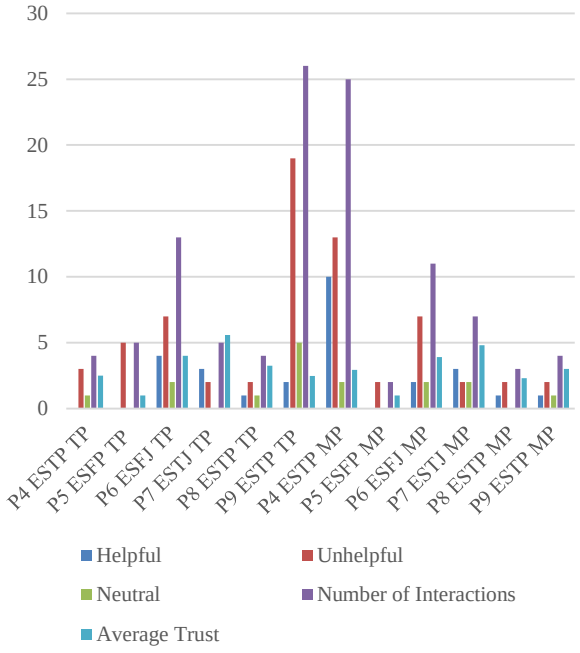


Fig. 5. Frequencies in cleverbot AI and ES personality.

Fig. 5 shows all ESTP participants rated Cleverbot AI poorly. This may be due to ESTPs needing to focus on practical tasks, so if this chatbot wasn't helpful, they ended conversations quickly to move on to more useful activities as they are sensitive to information and like to get explanations shorter, rather than a long conversation [24]. The participant with ESFP also had a low trustworthy experience with the chatbot, having small conversations but rating it a low trustworthy score. ESFPs, easily bored and less organized, likely disengaged if the chatbot wasn't engaging [38]. The participant with personality ESFJ had more unhelpful interactions but still gave a moderate trust score, likely due to their empathetic nature, born to be cooperative and need harmony to function [24]. When analyzing the participants with ESTJ personality, we found that their interactions with the chatbot were more helpful interactions, and can be the reason why it gave the chatbot a moderate average trustworthy score, which according to Hogan and Champagnes Personal Style Inventory Typology (1980), ESTJ's are known to be interested in subjects when they can be helpful and are often to see perspectives from another viewpoint [24].

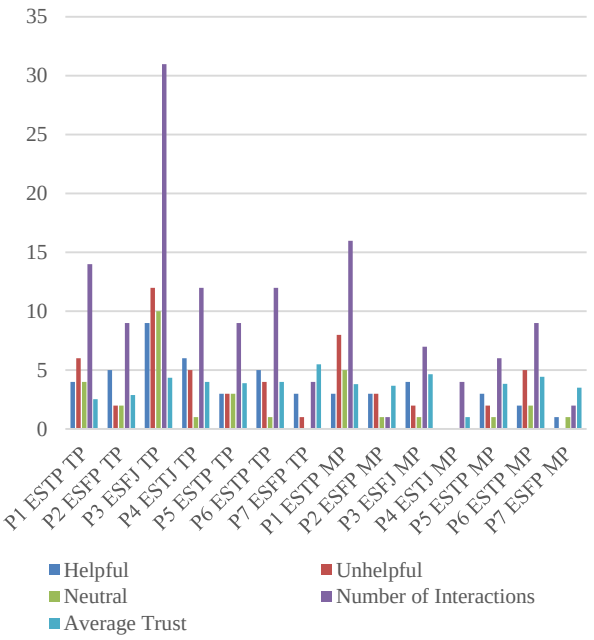


Fig. 6. Frequencies in kuki AI and ES personality.

According to Fig. 6, all three ESTP participants perceived Kuki as untrustworthy, as they all gave Kuki low trustworthy scores, even when responses were neutral. This can be due to their personality being blunt and may not have wanted to have long conversations if they were not straight to the point [24]. We had identified that one of the participants with ESFP (P1), personality had a handful of helpful interactions, as interpreted by our coding scheme, yet this participant rated Kuki lower, possibly due to its blunt, unengaging tone, as ESFPs like to enjoy things around them and enjoy entertainment from others [24, 38]. In contrast, the other participant with ESFP (P5) had fewer interactions but a moderate trust score, indicating that personality does vary from person to person. The participant with ESFJ had long conversations with Kuki, and we analyzed that most of the conversation was deemed to be unhelpful, but the participant still rated the conversation as moderately trustworthy, again relating to their personality where they want harmony and try to be nice [24]. The participant with personality ESTJ gave Kuki low trust scores even if the conversation was helpful.

Fig. 7 indicates that there are quite helpful interactions as observed by the researcher. Our findings suggest that ESTP participants engaged in more interactions overall when the conversation was meaningful and interactive, without having the chatbot to over-explain their responses. ESFJ users had high trust scores towards Replika and was observed by the researchers that during the meal prompt, there were some unhelpful interactions, but the participant still had viewed the chatbot as trustworthy, relating to their personality where they needed harmony. The participant with ESFP when interacting with Replika AI was seen to have conversations that were small, and neutral, yet still gave Replika an average trust score of 5, indicating moderate trustworthiness towards the chatbot. ESFP users are known to be outgoing, accepting, and make

their surroundings fun, thus if the participant with the personality of ESFP had meaningful interactions with Replika, they must have been accepting of it [24].

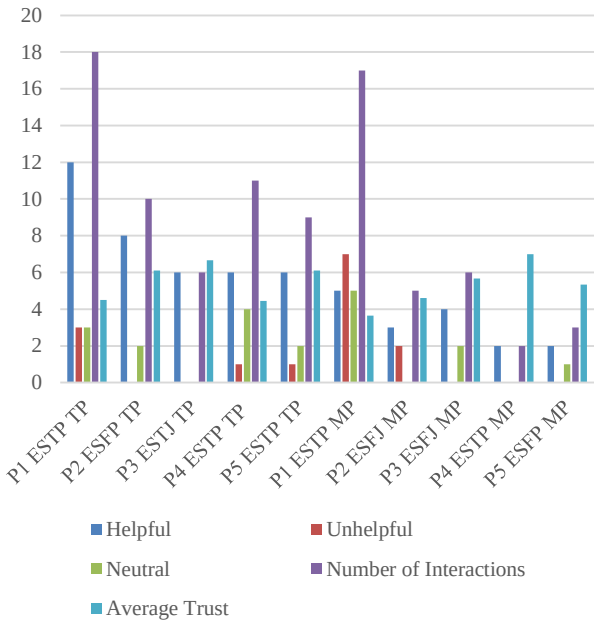


Fig. 7. Frequencies in replika AI and ES personality.

C. The IN Personality and Frequencies

Table IV shows descriptive statistics for IN personality users across the three chatbots. The mean trust level is 3.99 (SD = 1.49), the average number of interactions is 14.9 (SD = 10) suggesting high variability in how often participants engaged. Helpful responses averaged at 6.67 times (mean = 6.67, SD = 5.653), while unhelpful responses averaged a bit more with 6.80 average (SD = 4.02). Neutral responses were less frequent, averaging 1.70 (SD = 2.020), with some variation across participants.

TABLE IV. DESCRIPTIVE STATISTICS FOR IN PERSONALITY

	Mean	Std. Deviation
Trust	3.9908	1.49869
Interaction	14.93	10.007
Helpful	6.67	5.653
Unhelpful	6.80	6.18
Neutral	1.70	2.020

a.* N = 30

** n = 6

Fig. 8 indicates that the participant with personality INFJ had extremely long conversations in both prompts with Cleverbot. However, despite the long conversations, we observed that there were quite unhelpful interactions between the participant and the chatbot. This aligns with Champagnes *Personal Style Inventory Typology* (1980), where personality INFJ work hard for their needs met, and have a desire to get things done when they want it, they are hard workers and put a lot of effort into their work. It seems as if this participant with INFJ wanted Cleverbot to respond to the prompts accordingly.

The participants with INTJ users yielded similar results, long conversations, unhelpful responses frequently, and very low trustworthiness towards the chatbot. According to Hogan and Champagne, the INTJ personality type is often stubborn, skeptical, and critical [24]. It must have been that these participants were more likely to be critical of Cleverbot. The participant with INTP views trust towards Cleverbot as very untrustworthy, as Hogan and Champagne describe the INTP personality as extremely logical [24], so if the chatbot was not being logical with the user, the participant must have viewed it as untrustworthy.

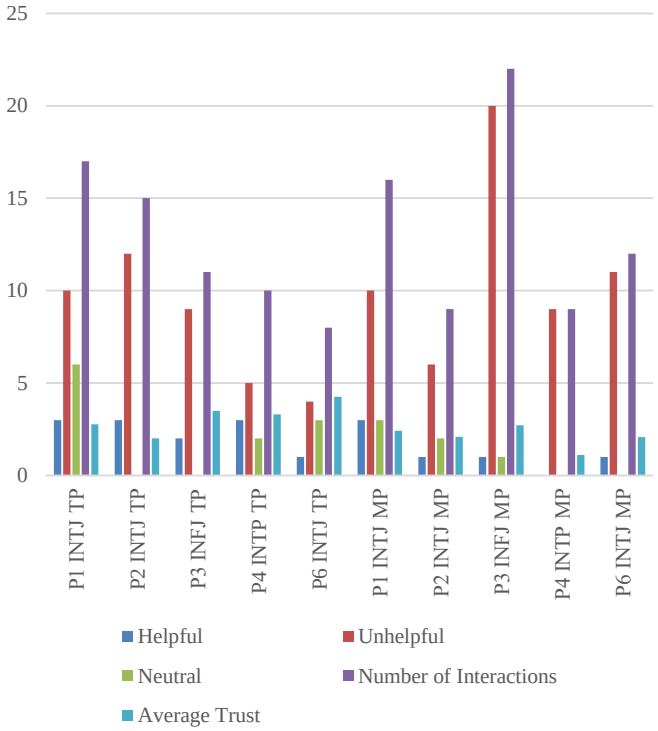


Fig. 8. Frequencies in cleverbot AI and IN personality.

Fig. 9 shows the interaction frequencies and average trust with Kuki and IN personality types. The participants with personality type INFJ had the highest interaction levels but also faced more unhelpful responses. Despite these interactions, their perceived trustworthiness towards the chatbot remained low, indicating the participant viewed this chatbot as untrustworthy. All participants with INTJ experienced high helpful responses from Kuki, and still gave moderate trust towards Kuki. INTJs are known to have great drive when it fits their own ideas and has meaningful interactions [24]. We noticed that all conversations between Kuki and all the participants with INTJ had more helpful interactions compared to Cleverbot. The participant with INTP personality had quite a few conversations and had an overall average low untrustworthy score towards the chatbot. The participant with INFJ had the most conversations with this chatbot, despite the chatbot having more unhelpful responses compared to the helpful responses in the interaction between Kuki and the INFJ participant. INFJs are known to go above and beyond, putting their best efforts into their work, therefore it may have been that

this participant wanted to get the best answer from the chatbot over long conversations. [24].

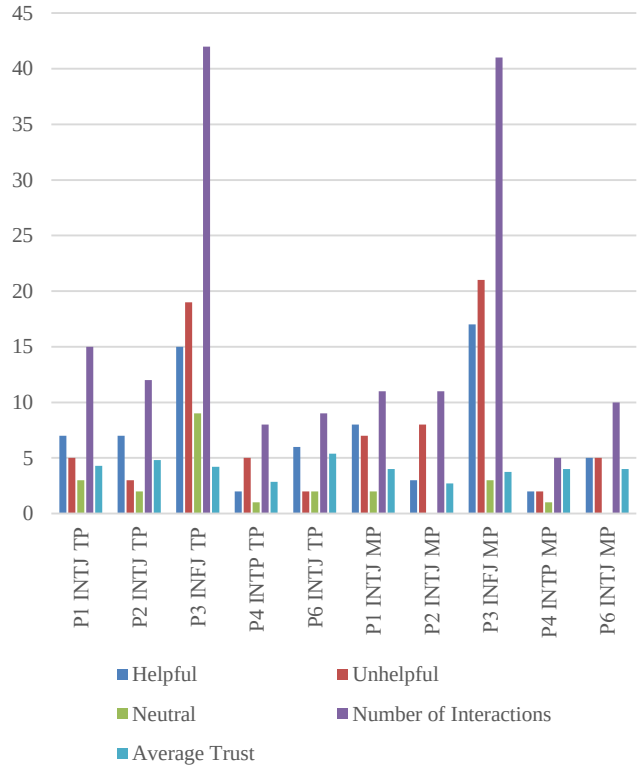


Fig. 9. Frequencies in kuki AI and IN personality.

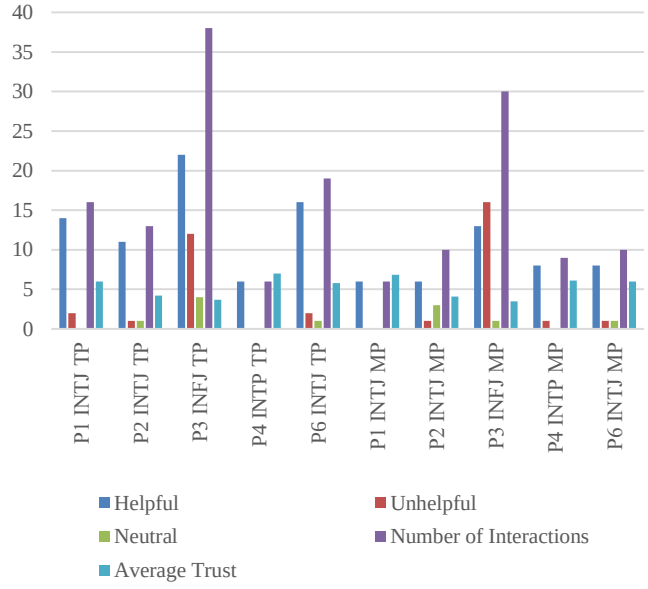


Fig. 10. Frequencies in replika AI and IN personality.

Fig. 10 displays the frequencies of interactions between Replika and users within the IN personality types. The participant with INFJ still had quite long conversations with Replika, as they did with Kuki and Cleverbot, relating to their personality once more where they work until they put their effort in [24]. All participants with INTJ had high trustworthiness towards the chatbot, and we observed that they all had more frequencies in the helpful category when we analyzed their interactions in both prompts. The participant with INTP had few conversations but deemed the chatbot to be very trustworthy based on their average trust score. We observed that Replika was more logical in their responses, and interacted well with all participants, leading to high average trust scores.

D. The IS Personality and Frequencies

Table V presents the descriptive statistics for the IS personality dimension and their code frequencies when interacting with all three chatbots. The mean trust across all chatbots is 3.85, with a standard deviation of 1.62, indicating a moderate level of trust across interactions. Participants engaged around an average of 14.12 interactions per conversation, with a standard deviation of 9.33, indicating a high variability in interactions. Helpful responses occurred most frequently, with a mean of 6.53 and a standard deviation of 5.38. Unhelpful responses were averaging around 4.60 with a standard deviation of 4.17. Neutral responses were least common, with a mean of 2.30, and a standard deviation of 2.76, indicating some variability across the interactions with all three chatbots. N represents the number of conversations analyzed, and n represents the total number of participants.

TABLE V. DESCRIPTIVE STATISTICS FOR IS PERSONALITY

	Mean	Std. Deviation
Trust	3.8549	1.61590
Interaction	14.12	9.325
Helpful	6.53	5.376
Unhelpful	4.60	4.172
Neutral	2.30	2.761

a.*N = 69
**n = 13

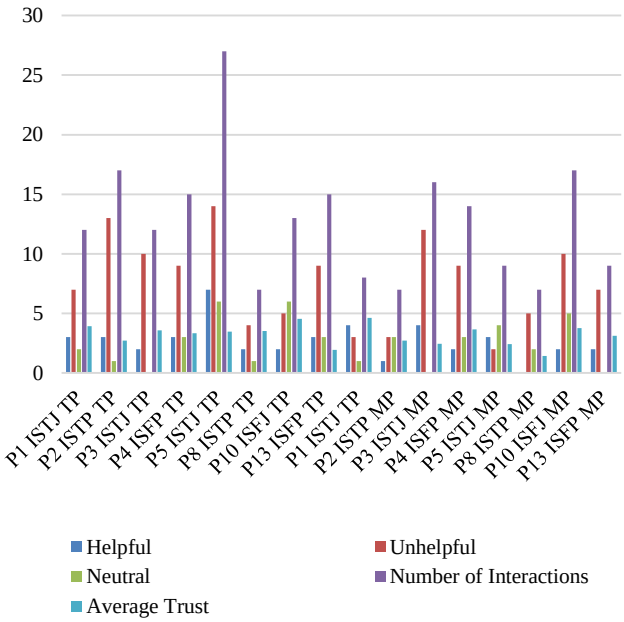


Fig. 11. Frequencies in cleverbot AI and IS personality.

Fig. 11 represents those participants with ISTJ personality had more helpful interactions within the meal prompt rather than the travel prompt, indicating their personality is to stay on task [39, 40]. The participants with ISTP had more unhelpful interactions with Cleverbot leading to a low average trust score within the travel prompt. We find this quite interesting, as according to Hogan and Champagne, people with ISTP personality types tend to not think necessarily more than they should, and that waste of energy is inefficient [24], we infer that the participants with the ISTP personality were trying to get the perfect answer from the chatbot to make up for the time they put into chatting with the chatbot.

Participants with ISFP had few long conversations, and this aligns with their personality, where Hogan and Champagne describe them as people who are often modest about their abilities, are not one to disagree, nor force their values or opinions on others [24]. Though they had long conversations, we analyzed that their conversations had more frequencies of unhelpful responses. Both participants with the ISFP personality still viewed the chatbot to be untrustworthy as indicated by their average trust score. Participants with the ISFJ personality yielded results that indicated they had more frequencies in the unhelpful code during their interaction with Cleverbot, aligning with Hogan and Champagne ISFJ personality type, where individuals work hard to get their obligations finished, and want to be accurate [24]. We can also infer that this participant with ISFJ wanted to get the right answers from the chatbot.

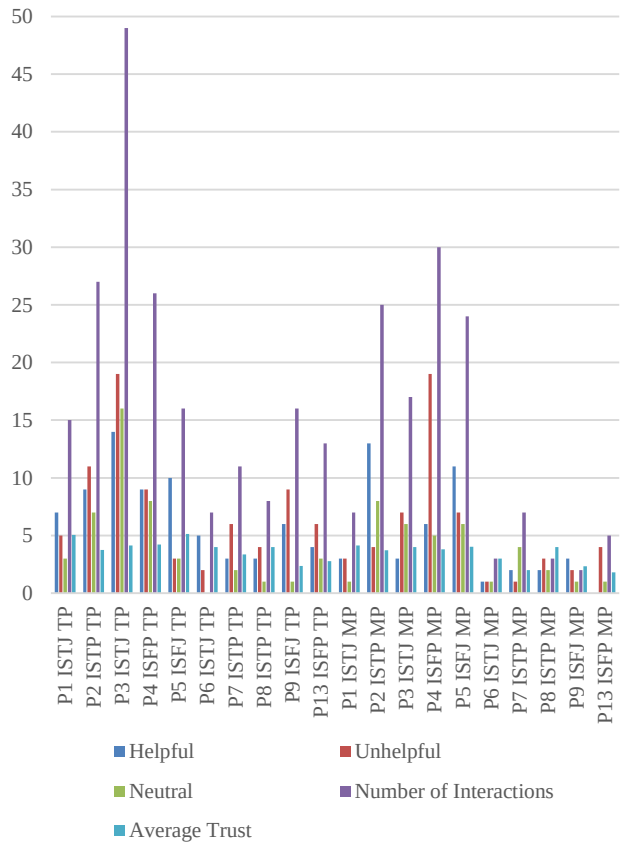


Fig. 12. Frequencies in kuki AI and IS personality.

Fig. 12 indicates that participants with personality ISTJ had perceived Kuki as moderately untrustworthy, despite the long conversations. ISTJs are known to be "Practical, orderly, matter of fact, logical, realistic and dependable [24]. As indicated in Fig. 12, Participant 3 with personality ISTJ had long conversations, but low trust towards the chatbot, while the other participant (Participant 6) with personality ISTJ had smaller conversations with Kuki but viewed the chatbot as moderately trustworthy, as indicated by the average trust score. These two participants had completely different views towards the chatbots, and experiences, indicating that every individual is different. Users with ISTP were flexible with their interactions with the chatbots, but overall, all participants with the ISTP personality had a low trustworthiness perception of Kuki. Participants with ISFP had long conversations in the travel prompts, and both viewed Kuki as very untrustworthy as indicated by their average trust score. However, in the meal prompt participant 13 with personality ISFP had a small interaction, with a very low average trust score, perceiving the chatbot as very untrustworthy. ISFPs are known to not want to waste time, so we can infer that Participant 13 did not want to continue the conversation [24]. When analyzing ISFJ participants, these participants had moderate untrustworthy perceptions towards the chatbot as indicated by their average trust scores in both prompts but do have the highest trust with Kuki when compared to the other three types of Introvert-Sensing types.

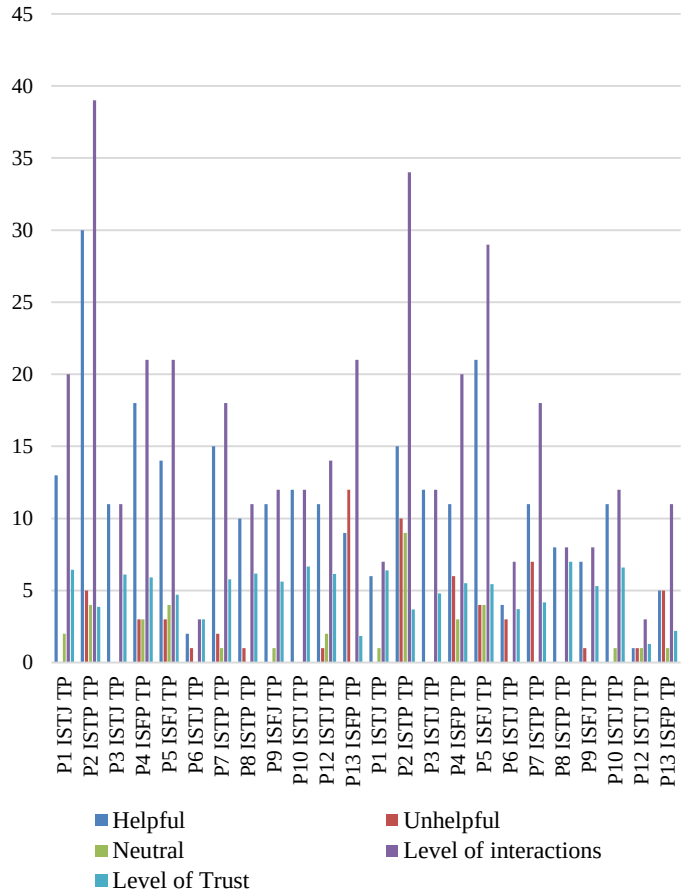


Fig. 13. Frequencies in kuki and the IS personality.

Fig. 13 indicates that all participants with ISTJ that had helpful frequencies, and long conversations tended to have a high average trust score in both prompts. However, we noticed that if the interaction was short, the participants with ISTJ continued to have low perceptions of trustworthiness towards Replika, as indicated in Participants 12 and 6. The participants with ISFJ had long conversations and a high score of average trust towards Replika. This might be due to their hard work and devotion, warm-heartedness, attention to detail, and need for accuracy, and how the Replika chatbot gave helpful responses, this might have led to high trustworthiness average scores [24]. All ISTP users had high trustworthiness towards the chat, indicating their personality of being flexible and easygoing, especially when the chatbot was helpful as we observed. The participants with ISFP all had moderately high trustworthiness towards Replika, and we observed that Replika did give helpful responses to all prompts when interacting with users with the Introvert-Sensing type.

V. Discussion and Limitations

Within the Intravenous and Intuition Dimension, participants with ENFJ tended to have lower trust scores in Cleverbot, and Kuki when it was not helpful and tended end to the conversation, however, one participant in the EN dimension with personality ENFP tended to give the chatbot higher scores when it was not helpful, rather, giving it a lower score when it was helpful. Within our coding, if the chatbot was being tangible (i.e. steering off the conversation to complete something different, such as "Who Cares"), or not engaging with the user, this would be considered unhelpful. Despite the chatbot being unhelpful, the participant with user ENFP continued to give it higher trust scores. It can be inferred that in this study, the participants with ENFP personalities wanted more interactions that engaging, and more exciting, as their personality tends to be open to opportunities and tends to be more imaginative [14].

ENFJs in this study were more likely to end the conversation if the conversation was not meaningful. This relates to their personality, as ENFJs tend to be more logical, organized, and worthy of leadership [32].

Within the analysis of ES (Extraverted and Sensing) personality, ESTP users' conversations with all chatbots were often brief, especially when the conversation led to unhelpful responses. In our study, ESFP participants, known for their energetic nature [33] tended to rate chatbots when it was helpful and engaging. ESTJ participants gave higher scores when the chatbot was helpful and gave clear answers, as this relates to their personality where they value logic, relativism, and directness nature [32, 33], tended to give lower trust scores to the chatbot despite when the chatbot was helpful and engaging. ESFJ on the other hand, gave moderate trust scores to the chatbots even when it was being unhelpful, this can relate to their personality as they are more warmhearted and sensitive to the environment around them, as seen more in women for their motherly nature and warmhearted figures [32, 34].

The Inverted and Intuition (IN) Dimension had mixed results. Participants with personality INFJ gave chatbots the lowest scores, despite being helpful or unhelpful. This relates to their personality from the dimension of the 'Feeling" and "Judgement", where they follow their own conviction and own path rather than objective fact [35].

Users with personality INTJ had more helpful interactions and reported higher trust scores, indicating that the more beneficial and engaging the chatbot is, the more users tend to trust it despite the level of interactions. This goes on with their personality as they are more rational when it comes to their environment and analytic to the future when it comes to decisions [36, 37].

Participants with personality INTP had mixed results with some chatbots. It had mixed results from helpful, unhelpful, and neutral interactions. While interaction levels were consistent, they didn't necessarily lead to higher trust levels. Overall, the INTP participants did not have consistent results between the three chatbots. This relates to their personality as they are more likely to be spontaneous and have more intuition rather than sensing, and difficult to please, so if they have a bad experience, this leads to difficulty in getting high trust scores [37, 38].

The largest set of personalities in this simple was Introverted and Sensing (IS). In all three chatbots (Cleverbot, Kuki, and Replika) the participants with IS tended to still have longer conversations despite the possibility that the chatbot was not being helpful. Research has indicated that individuals with personality ISTJ tend to be nervous when it comes to high achievement, and have a need to accomplish a task [24, 39, 40], it can be inferred that participants in this study wanted to continue to conversate with the chatbot until they were able to get a good answer from them, despite having lower scores in trust and unhelpfulness. ISTP yielded different results, showing more trust in the chatbot despite being unhelpful, as their perception leads them to think the chatbot might have been worthy of a high trust rating, despite objective unhelpful frequencies in their conversations, feeling the way of their senses as well rather than being intuitive, allowing for flexibility in their way of perceiving the world [33, 37].

The participants with the ISFJ personality had similar results to ISTP surprisingly, where as seen in their conversations with Replika, they tended to have higher trust towards Replika, yet still showed moderate trust levels in Cleverbot and Kuki despite those two chatbots being unhelpful. ISFJs are known to be more warm-hearted, sincere, and approachable, having the desire to engage with the chatbot as long as emotional satisfaction is involved, their "Feeling" type indicates they still felt something for the chatbot despite not being helpful [24].

The participants with ISFP personality tended to have high trust scores in all three chatbots as compared to the other IS types, despite the categories of helpfulness, unhelpfulness, and neutral. Their personality of feeling and perceiving might have influenced their trust scores to be higher with the chatbots.

While each personality yielded different results, with a few cases of similar results, there are limitations to this study. Within the Dimension of Extraversion and Intuition, this personality was underrepresented in this study, and Introversion dominated this study. The sample size was taken from an Information Systems major program, where students can go on to work in the cybersecurity field, be engineers, or work in tech,

and research has shown that students with introversion tend to likely be engineers, and work in the science field [41, 42]. Another limitation is that there is missing data from several users from the Introvert-Intuition (IN) Dimension, Introvert-Sensing (IS) Dimension, and Extrovert-Sensing (ES) dimension. However, we have successfully collected a large amount of systematic conversations from 29 participants. The results can be generalized to similar personality users to contribute to broader findings.

VI. Conclusion

We observed how personality dimensions influence trust and engagement towards chatbots across two different scenarios, a travel prompt and a meal prompt. We had the participants rate each interaction on a Likert scale rating from 1 to 7, with 7 being the highest trustworthy score, and 1 being the lowest trustworthy score. We created a hierarchical modeling scale to organize our data, and had participants go through three chatbots to explore variabilities in each chatbot. We found key findings, which included that the participants with ENFJ personalities rated Cleverbot low, in counterpart with our other participants ENFP who seemed to enjoy the conversation they had with Cleverbot, aligning with their imaginative and optimistic natures [24].

ESTP participants preferred direct, practical responses and gave lower trust scores when interactions were inefficient. Similarly, we observed that ESFP participants disengaged from unengaging chatbots but appreciated entertaining responses, while ESFJs maintained moderate trust levels, driven by their empathetic and harmony-focused personalities [24]. ESTJ participants rated trust highly when interactions were helpful but were critical of less engaging exchanges.

For Introverted-Intuitive personality types, INFJ participants engaged in long conversations but rated unhelpful interactions as untrustworthy, consistent with their diligent and goal-oriented nature. INTJs preferred meaningful interactions and gave moderate trust scores to Kuki due to its helpful responses, while INTP participants, being highly logical, rated chatbots as untrustworthy when responses lacked coherence or when the chatbot failed to deliver the information the participant wanted.

For our Introverted-Sensing types, ISTJ participants valued helpful interactions and rated them highly, though shorter or unhelpful exchanges led to lower trust. ISTPs, known for their practicality [24], also rated chatbots lower when responses were inefficient or overly lengthy. ISFP participants, modest and time-conscious, found chatbots untrustworthy if responses lacked engagement, while ISFJs maintained moderate trust levels, valuing accuracy and helpfulness, particularly with Replika.

Across all three chatbots, we observed that the Kuki chatbot was perceived as more trustworthy by most of the participants in this study, more with the Extroverted-Intuitive and Extroverted-Sensing types due to its helpfulness. Replika received the highest trust scores, particularly from ISFJ and ISTP users, due to its logical and engaging responses. These findings underscore the importance of tailoring chatbot interactions to align with user personality traits, enhancing trust and satisfaction by balancing logical coherence, directness, and emotional engagement.

The future study will recruit participants from all majors to increase the sample size in different personality dimensions except for the IS dimension in which we have enough data from technology-related major students. We will also continue to design chatbots focusing on personality following the design interpretations from this work. This work contributes to the fast development of user-centered CA design with frequencies of user responses and trustworthy ratings to sort their behaviors by patterns correlating to their personality dimensions. It focuses on user perceptions of trust in chatbot experience. It also presents detailed interpretations of diverse personality users' behavior patterns and trends to show how they interact with CAs. The work sheds light on design guidelines of user trust based on personality for better human-centered AI design.

References

[1] Y. K. Dwivedi, A. Sharma, N. P. Rana, M. Giannakis, P. Goel, and V. Dutot, "Evolution of artificial intelligence research in Technological Forecasting and Social Change: Research topics, trends, and future directions," Technological Forecasting and Social Change, vol. 192, Jul. 2023J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Y. Xu et al., "Artificial Intelligence: a Powerful Paradigm for Scientific Research," The Innovation, vol. 2, no. 4, Oct. 2021

[3] A. Turing, "Computing Machinery and Intelligence," Mind, vol. 59, no. 236, pp. 433–460, Oct. 1950.

[4] J. McCarthy, M. L. Minsky, N. Rochester, and C. E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955," AI Magazine, vol. 27, no. 4, pp. 12–12, Aug. 1955,

[5] Oliveir E., Gama J., Z. Vale, and H. L. Cardoso, Progress in artificial intelligence : 18th EPIA Conference on Artificial Intelligence, EPIA 2017, Porto, Portugal, September 5-8, 2017, Proceedings. Cham, Switzerland: Springer, 2017.

[6] E. Bauer et al., "Using natural language processing to support peer-feedback in the age of artificial intelligence: A cross-disciplinary framework and a research agenda," British Journal of Educational Technology, vol. 54, no. 5, May 2023.

[7] S. Pacheco-Mendoza, C. Guevara, A. Mayorga-Albán, and J. Fernández-Escobar, "Artificial Intelligence in Higher Education: A Predictive Model for Academic Performance," Education Sciences, vol. 13, no. 10, p. 990, Oct. 2023.

[8] N. Ameen, A. Tarhini, A. Reppel, and A. Anand, "Customer Experiences in the Age of Artificial Intelligence," Computers in Human Behavior, vol. 114, no. 106548, p. 106548, Jan. 2021.

[9] R. S. Sutton et al., "Reward-respecting subtasks for model-based reinforcement learning," Artificial Intelligence, vol. 324, p. 104001, Nov. 2023,

[10] A. Pradhan and A. Lazar, "Hey Google, Do You Have a Personality? Designing Personality and Personas for Conversational Agents," CUI 2021 - 3rd Conference on Conversational User Interfaces, Jul. 2021

[11] J. Wang, J. Q. Chen, D. Kang, S. Herath, and A. AbuHussein, "Designing a Conversational Agent for Education using a Personality-based Approach," International Journal of Advanced Computer Science and Applications, vol. 15, no. 6, 2024.

[12] F. Kaya, F. Aydin, A. Schepman, P. Rodway, O. Yetişensoy, and M. Demir Kaya, "The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence," International Journal of Human–Computer Interaction, vol. 40, no. 2, pp. 497–514, Dec. 2022.

[13] J.-P. Stein, T. Messingschlager, T. Gnambs, F. Hutmacher, and M. Appel, "Attitudes towards AI: measurement and associations with personality," Scientific Reports, vol. 14, no. 1, Feb. 2024.

[14] A. Oksanen, N. Savela, R. Latikka, and A. Koivula, "Trust Toward Robots and Artificial Intelligence: An Experimental Approach to Human–Technology Interactions Online," Frontiers in Psychology, vol. 11, Dec. 2020.

[15] A. H. Maslow, "A Theory of Human Motivation," Psychological Review, vol. 50, no. 4, pp. 370–396, 1943.

[16] E. Sullivan, "Self-actualization," Encyclopædia Britannica. 2019

[17] M. X. Zhou, G. Mark, J. Li, and H. Yang, "Trusting Virtual Agents," ACM Transactions on Interactive Intelligent Systems, vol. 9, no. 2–3, pp. 1–36, Apr. 2019.

[18] H. Li, "Rethinking human excellence in the AI age: The relationship between intellectual humility and attitudes toward ChatGPT," Personality and Individual Differences, vol. 215, p. 112401, Dec. 2023.

[19] R. Riedl, "Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions," Electronic Markets, vol. 32, Nov. 2022.

[20] N. N. Sharan and D. M. Romano, "The effects of personality and locus of control on trust in humans versus artificial intelligence," Heliyon, vol. 6, no. 8, p. e04572, Aug. 2020.

[21] A. D. Kaplan, T. T. Kessler, J. C. Brill, and P. A. Hancock, "Trust in Artificial Intelligence: Meta-Analytic Findings," Human Factors: The Journal of the Human Factors and Ergonomics Society, vol. 65, no. 2, May 2021.

[22] C. Montag, B. Becker, and B. J. Li, "On Trust in Humans and Trust in Artificial intelligence: a Study with Samples from Singapore and Germany Extending recent research," Computers in human behavior. Artificial humans, vol. 2, no. 2, pp. 100070–100070, May 2024.

[23] S. Rushton, J. Morgan, and M. Richard, "Teacher's Myers-Briggs personality profiles: Identifying effective teacher personality traits," Teaching and Teacher Education, vol. 23, no. 4, pp. 432–441, May 2007.

[24] R. Hogan and D. Champagne, "University Associates The 1980 Annual Handbook for Group Facilitators 1," 1980.

[25] J. H. Jung, Y. Lee, and R. Karsten, "The Moderating Effect of Extraversion–Introversion Differences on Group Idea Generation Performance," Small Group Research, vol. 43, no. 1, pp. 30–49, Sep. 2011.

[26] A. Naveed, "How College Students' MBTI Personality Types Relate to Their Levels of Emotional Intelligence," Corpus Journal of Social Sciences & Management Review, vol. 2, no. 01, pp. 65–70, Mar. 2024.

[27] J. Langan-Fox and D. A. Shirley, "The Nature and Measurement of Intuition: Cognitive and Behavioral Interests, Personality, and Experiences," Creativity Research Journal, vol. 15, no. 2, pp. 207–222, Jul. 2003.

[28] J. Wang, M. Schmidt, R. Kuikel. "User trust and credibility of conversational agents," In Procs of 23rd Annual Security Conference, Las Vegas, NV, U.S.A, May 2024.

[29] J. Wang, J. Dunkley, M. Hamal, V. Raut, and S. Herath, "Designing Conversational Agents for Education: A Preliminary Study of User Personality's Impact on Design," The International Journal of Engineering and Science (IJES), vol. 12, no. 2, 2023.

[30] H. Dubberly and P. Pangaro, "What is conversation, and how can we design for it?," interactions, vol. 16, no. 4, p. 22, Jul. 2009.

[31] R. Zhang, X. Liang, and S.-H. Wu, "When chatbots fail: exploring user coping following a chatbots-induced service failure," Information technology & people, vol. 37, no. 8, pp. 175–195, Jul. 2024.

[32] N. N. Vitó Ferreira and J. J. Langerman, "The correlation between personality type and individual performance on an ICT Project," IEEE Xplore, Aug. 01, 2014.

[33] H. Gordan, "Myers Briggs Type Indicator Personality Characteristics of Beginning Trade and Industrial and Health Occupations Education Secondary Teachers," Journal of Health Occupations Education, vol. 14, no. 1, 2000.

[34] S. W. Johnson, M. S. Gill, C. Grenier, and J. Taboada, "A Descriptive Analysis of Personality and Gender at the Louisiana State University School of Veterinary Medicine," Journal of Veterinary Medical Education, vol. 36, no. 3, pp. 284–290, Sep. 2009.

[35] A. Naveed, "View of How College Students' MBTI Personality Types Relate to Their Levels of Emotional Intelligence," Corpus Journal of Social Sciences & Management Review, vol. 2, no. 1, 2024.

[36] O. Y. M. Sardjono, "Learning Outcomes Based on Myer-BriggsType Indictaor (MBTI) Personality from Accounting Department Students Sam Ratulangi University," Accountability , vol. 12, no. 1, 2023.

[37] J. Jankowski, The 16 Personality Types in a Nutshell. 2016.

[38] J. Simkus, "Logician (INTP) Personality Type," Simple Psychology, 2024.

[39] R. Diab-Bahman, "The Impact of Dominant Personality Traits on Team Roles," The Open Psychology Journal, vol. 14, no. 1, pp. 33–45, Feb. 2021.

[40] I. Briggs Myers, Introduction to Type®. CPP, 1998.

[41] M. Pollock, "The relationship between personality type and choice of college major," 2016.

[42] A. Vedel, "Big Five personality group differences across academic majors: A systematic review," Personality and Individual Differences, vol. 92, no. 92, pp. 1–10, Apr. 2016.

An Enhanced Real-Time Intrusion Detection Framework Using Federated Transfer Learning in Large-Scale IoT Networks

Khawlah Harahsheh[1], Malek Alzaqebah[2, 3], Chung-Hao Chen[4]
Ph.D. student, Department of Electrical & Computer Engineering, Old Dominion University, Norfolk, VA, 23529 USA[1]
Department of Mathematics, College of Science, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia[2]
Basic and Applied Scientific Research Center, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia[3]
Department of Electrical & Computer Engineering, Old Dominion University, Norfolk, VA, 23529 USA[4]

***Abstract*—The exponential growth of Internet of Things (IoT) devices has introduced critical security challenges, particularly in scalability, privacy, and resource constraints. Traditional centralized intrusion detection systems (IDS) struggle to address these issues effectively. To overcome these limitations, this study proposes a novel Federated Transfer Learning (FTL)-based intrusion detection framework tailored for large-scale IoT networks. By integrating Federated Learning (FL) with Transfer Learning (TL), the framework enhances detection capabilities while ensuring data privacy and reducing communication overhead. The hybrid model incorporates convolutional neural networks (CNNs), bidirectional gated recurrent units (BiGRUs), attention mechanisms, and ensemble learning. To address the class imbalance, Synthetic Minority Over-sampling Technique (SMOTE) was employed, while optimization techniques such as hyperparameter tuning, regularization, and batch normalization further improved model performance. Experimental evaluations on five diverse IoT datasets, i.e. Bot-IoT, N-BaIoT, TON_IoT, CICIDS 2017, and NSL-KDD, demonstrate that the framework achieves high accuracy (92%-94%) while maintaining scalability, computational efficiency, and data privacy. This approach provides a robust solution to real-time intrusion detection in resource-constrained IoT environments.**

***Keywords*—Intrusion detection systems; federated learning; transfer learning; cybersecurity; scalability; resource constraints; machine learning; Internet of Things**

I. INTRODUCTION

The rapid growth of the Internet of Things (IoT) has introduced significant security challenges due to the diversity and resource constraints of IoT devices. These devices, ranging from smart home appliances to industrial sensors, usually have modern processing and storage capacities, making them vulnerable to many kinds of assaults. When used in IoT environments, traditional intrusion detection systems (IDS) that rely on centralized structures face several problems and critical challenges, including scalability, latency, communication overhead, and privacy risks. Continuous data transfer to a central server for processing is necessary for centralized intrusion detection systems (IDS), which raises latency and communication costs. Additionally, because sensitive data from IoT devices needs to be sent and kept centrally, centralized systems provide serious privacy risks.

Despite their importance, current IDS implementations are ill-equipped to handle the unique requirements of IoT environments, particularly regarding real-time threat detection, adaptability to emerging threats, and resource efficiency. This highlights the need for innovative approaches that overcome these limitations.

Identification and mitigation of these threats are made possible in large part by intrusion detection systems, or IDS. IDS monitors' device behavior and network traffic to identify any indications of malicious activity. Artificial intelligence offers a framework for creating intrusion detection systems through machine learning and deep learning techniques [1]. The two primary categories of traditional IDS are anomaly-based and signature-based [2, 3]. Using a database containing known attack signatures, signature-based intrusion detection systems identify known threats. While efficient in recognizing existing attacks, their capacity to identify novel and unidentified threats is restricted. Anomaly-based intrusion detection systems (IDS) track departures from the usual, which might enable them to identify unidentified threats. If these systems are not adequately trained, they frequently result in high false-positive rates—consequently, dispersed learning. Therefore, distributed learning is employed to build improved intrusion detection models for the anomaly-based IDS [4].

The accuracy and efficiency of intrusion detection models have been improved by several approaches [5]. Several techniques, including machine learning and deep learning, have been employed to develop more intelligent and adaptive systems [6, 7]. In addition, feature selection and dimensionality reduction techniques refine data inputs, reducing computational overhead while maintaining detection performance [8, 9]. Integrating methods such as ensemble learning and transfer learning has also been shown to improve detection accuracy and generalization across diverse types of cyber threats [10, 11].

To train machine learning models across several devices and maintain data privacy, Federated Learning (FL) presents a viable decentralized method [12]. Using their local data, IoT devices cooperatively train a shared global model in FL; model updates are only sent to a central server for aggregation. By ensuring that raw data remains on local

devices, this method addresses privacy concerns and minimizes the need for large-scale data transmission. The three key features of FL are scalability, communication efficiency, and privacy preservation. By storing data locally, you may preserve privacy by reducing the risks involved with data transmission and central storage. Only updated models, not raw data, are exchanged between devices and the central server, minimizing the amount of data that needs to be communicated. This ensures optimal communication efficiency. Furthermore, FL is very scalable, which makes it ideal for a wide range of IoT scenarios [13]. However, while FL addresses privacy and data transmission concerns, it alone may not fully address the need for rapid adaptation to new and evolving threats in IoT networks [14].

Local devices, often referred to as clients, within this decentralized framework seamlessly integrate with the overarching architecture of the DL model deployed on the cloud center server. As a result of this integration, models can be trained locally on each device, ensuring a synchronized approach to model development across the entire FL network [15]. While FL has shown promise in addressing privacy and scalability concerns, it struggles to adapt to rapidly evolving threats in IoT environments. This underscores the need for enhanced techniques that can combine the privacy-preserving features of FL with models that adapt quickly to diverse IoT environments.

Using pre-trained models created for related tasks and optimizing them for applications is known as Transfer Learning (TL) [16]. Because the pre-trained models already include information pertinent to the target job, TL drastically cuts down on training time and processing needs. Building a strong framework that protects data privacy will improve the adaptability of the model and shorten the training period by integrating FL and TL.

Despite significant progress in FL and TL applications, few studies have successfully combined these techniques to address the unique challenges of IoT networks comprehensively. Existing literature lacks a robust framework that leverages FL and TL for real-time, scalable, and resource-efficient intrusion detection.

Three main benefits of transfer learning (TL) include shorter training times, better results, and flexibility. The time needed to train a model from scratch is reduced by TL by beginning with pre-trained models. Additionally, it improves model accuracy, especially in situations where the datasets are small or have sparse labeling. Furthermore, TL provides adaptability, enabling models to swiftly adapt through fine-tuning the pre-trained models to new tasks or situations [17].

This work introduces a Federated Transfer Learning (FTL) framework that combines the strengths of FL and TL to improve intrusion detection in IoT networks. The framework enhances detection accuracy through a hybrid model integrating convolutional neural networks (CNNs), bidirectional gated recurrent units (BiGRUs), attention mechanisms, and ensemble learning. Additionally, it addresses class imbalance using the Synthetic Minority Over-sampling Technique (SMOTE) and optimizes model performance through hyperparameter tuning, regularization, and batch normalization.

The contributions of this research are as follows:

Addressing Scalability and Privacy Concerns: The proposed FTL framework decentralizes model training to preserve privacy and reduce communication overhead, enabling effective intrusion detection in IoT networks.

Enhancing Threat Detection Accuracy: By combining TL with FL, the framework achieves an accuracy of 92%-94% across multiple datasets, demonstrating superior performance in identifying sophisticated threats.

Optimizing for IoT Resource Constraints: Techniques such as L1/L2 regularization and batch normalization ensure the model is lightweight and efficient, suitable for resource-limited IoT devices.

Handling Class Imbalance: SMOTE is employed to improve model generalization by addressing the underrepresentation of attack samples in IoT datasets.

Adapting to Diverse IoT Environments: Domain adaptation techniques ensure the model is flexible and robust, enabling it to generalize across various IoT devices and datasets.

This study fills a critical gap in the literature by presenting a scalable and privacy-preserving framework for real-time intrusion detection in IoT networks. The rest of this paper is organized as follows: Section II reviews related work; Section III outlines the methodology; Section IV presents the experimental results; Discussion is given in Section V and Section VI concludes the paper, highlighting limitations and future research directions.

II. Related Work

The field of intrusion detection in IoT networks has seen significant advancements through various innovative methodologies [5]. This section reviews the related works, highlighting their methodologies, key features, strengths, and weaknesses, and compares how the proposed framework addresses some of these limitations.

Karimy and Reddy [18] employed FL to enhance security and privacy in IoT environments. The methodology involves local model training on IoT devices and subsequent aggregation of the model updates on a central server. The authors achieved an accuracy of 99.9% when they used the N-BaIoT and other custom datasets. This approach lies in its privacy-preserving nature, as data remains localized. However, the approach suffers from high computational overhead. Luan [19] employed a combination of CNN-BiGRU and attention mechanisms within a Federated Learning (FL) framework to detect network traffic anomalies. The method was evaluated using the BoT-IoT and NSL-KDD datasets. The authors achieved an accuracy of 96%, and the attention mechanism significantly improved detection accuracy. It also introduced high computational costs.

Almesleh et al. [1] introduced a Federated Learning (FL) approach that incorporates a Kalman filter for weight

aggregation, enhancing the overall performance of the model. With an accuracy of 99.8%, the Kalman filter enhances the weight aggregation process across a variety of IoT datasets, contributing to robust performance. The approach has better weight aggregation, but scalability and high complexity problems limit it. Bhavsar et al. [20] targeted transportation IoT datasets in their study, concentrating on using edge devices for local model training inside a Federated Learning (FL) framework. This method works effectively in large-scale IoT contexts because it is scalable and provides effective local training. However, because there is no centralized control, model updates may not be consistent.

Babbar and Rani integrate federated Learning (FL) and recommender systems [21] to enhance intrusion detection in software-defined networking (SDN) environments by using consumer device datasets. Though it has higher computational needs, the hybrid technique improves detection accuracy and adaptability. Across a variety of IoT datasets, Raj et al. [22] enhanced security protocols using FL. Key advantages include improved security procedures and privacy-preserving techniques. The method has scalability problems and significant overhead, though.

Ohtani et al. [2] used the N-BaIoT dataset to combine Federated Learning (FL) with one-class SVM for the purpose of detecting zero-day attacks in IoT networks. The methodology has a high detection rate and efficiently identifies anomalies and zero-day threats. However, it has a high false positive rate.

Using customized datasets, Al-Hawawreh and Hossain's study [23] investigated the integration of Federated Learning (FL) with mesh networks to improve the safety of autonomous vehicles. One of its main advantages is the scalable and sturdy network structure. However, the approach is complex and resource intensive. After testing unique IoT attack datasets, Umair et al. [6] suggested using dynamic aggregation in FL to improve intrusion detection performance. Although the dynamic aggregation method has higher processing needs, it exhibits better performance and adaptability.

TABLE I. RELATED WORK COMPREHENSIVE OVERVIEW

Paper Title	Methodology	Dataset(s)	Acc	Strengths	Weaknesses	Time
[1]	FL, Kalman Filter	Various IoT Datasets	99.8	Improved weight aggregation, Robust performance	High complexity, Limited scalability	High
[4]	FL, One-Class SVM	N-BaIoT	-	Effective zero-day detection, Autonomous	High false positive rate	-
[15]	FL, Dynamic Aggregation	Custom IoT Attack Dataset	87.98	Improved performance, Adaptability	High computational requirements	High
[18]	FL	N-BaIoT, Custom Dataset	99.9	High accuracy, Privacy-preserving	High computational overhead	High
[19]	FL, CNN-BiGRU, Attention	BoT-IoT, NSL-KDD	96	High accuracy, Attention mechanism	High computational cost	High
[20]	FL, Edge Devices	Transportation IoT Dataset	From 94 to 99	Efficient local training, Scalability	Lack of centralized control	-
[21]	FL, Recommender Systems	Consumer Device Dataset	From 78 to 99	Enhanced detection accuracy, Adaptability	High computational requirements	High
[22]	FL	Various IoT Datasets	-	Enhanced security protocols, Privacy-preserving	High overhead, Scalability issues	High
[23]	FL, Mesh Networks	Custom Autonomous Vehicle Dataset	From 95 to 99	Robust network, Scalable	High complexity, Resource-intensive	High

To guarantee the integrity and immutability of model updates and solve security concerns in decentralized systems, the proposed framework integrates FL and TL. Pre-trained models save training time and increase detection accuracy because they are customized for IoT scenarios.

By combining FL and TL, the suggested methodology seeks to improve real-time intrusion detection in IoT networks while addressing the limitations of the other methods mentioned in the related work section. It uses a hybrid model that incorporates advanced machine learning methods including CNNs, BiGRUs, and attention processes. Strategies for data augmentation and optimization help to further improve performance. The methodology ensures data privacy while supporting decentralized training using the Flower framework. Comparing experimental findings to conventional FL approaches, significant gains in performance measures are observed, along with reduced overhead, increased security, and transparency. Table I provides a summary and comparison of existing works in the field with the proposed research, highlighting their methodologies, strengths, weaknesses, and performance metrics. This comparison underscores the advancements introduced by this study, particularly in addressing limitations such as scalability, privacy, and adaptability in IoT intrusion detection.

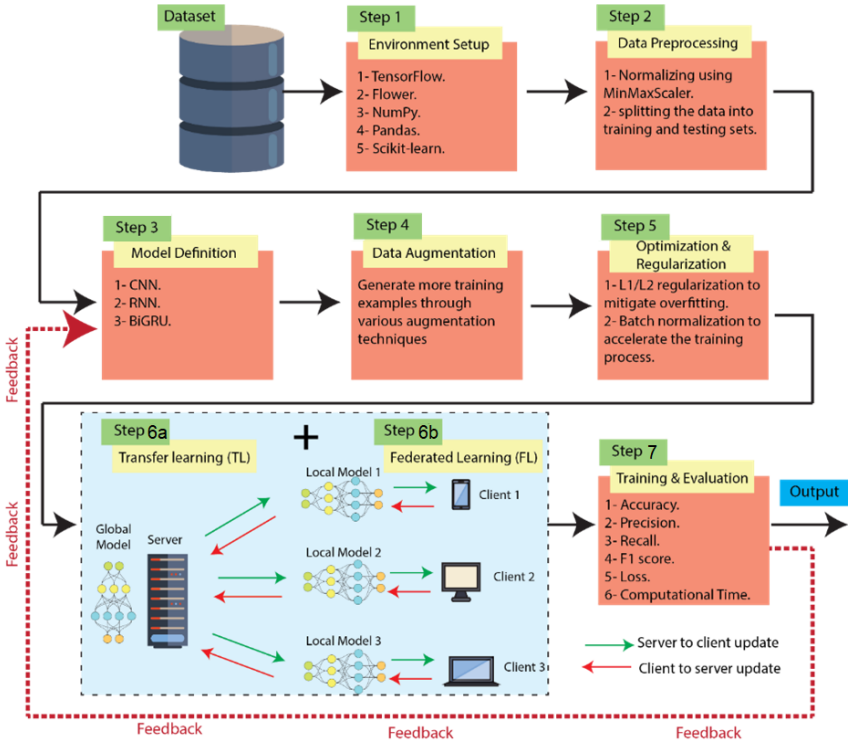


Fig. 1. Steps of the proposed methodology.

III. METHODOLOGY

To assess the effectiveness of the improved model on various IoT datasets, a systematic approach leveraging Federated Learning (FL) and Transfer Learning (TL) has been adopted. These techniques were selected to address critical challenges in IoT environments, including privacy preservation, scalability, and adaptability to diverse datasets. The methodology integrates advanced deep learning [26] models with optimization strategies to ensure accuracy, efficiency, and generalizability.

Rationale: Federated Learning (FL) is utilized to ensure decentralized model training across IoT devices, preserving data privacy and reducing communication overhead. However, FL alone lacks rapid adaptability to new threats. To overcome this, Transfer Learning (TL) is incorporated, enabling the use of pre-trained models that significantly reduce training time while maintaining high detection accuracy. Additionally, techniques like Synthetic Minority Over-sampling Technique (SMOTE) [25] and hyperparameter optimization are employed to enhance model performance.

The methodology comprises the following steps (Fig. 1):

Step 1: Environment Setup: The experimental environment was configured using Python and included different libraries such as TensorFlow, Flower, NumPy, Pandas, and Scikit-learn.

Step 2: Preparing the Data Several widely used IoT datasets were employed, including BoT-IoT, TON_IoT, CICIDS2017, NSL-KDD, and N-BaIoT. Preprocessing was done on each dataset to make sure it was standardized and ready for model training. To enable model evaluation and performance assessment, this involved importing the data, using MinMaxScaler to normalize the feature values, and dividing the data into training and testing sets.

Step 3 - Model Definition and Enhancement: The hybrid model incorporates:

- Convolutional Neural Networks (CNNs) for feature extraction and spatial pattern detection.
- Bidirectional Gated Recurrent Units (BiGRUs) capture temporal dependencies in sequential data.
- Attention Mechanisms to focus on critical data features, enhancing detection accuracy.
- Ensemble Learning to combine predictions from multiple models for improved generalization and robustness.

Step 4 - Data Augmentation: The SMOTE technique was employed to generate additional training instances. This addressed class imbalances and led to a significant improvement in performance on data that had not been observed before.

Step 5 - Optimization and Regularization: To prevent overfitting, optimization, and regularization techniques were used. Extensive hyperparameter tuning was conducted using grid search and random search methods to identify the best hyperparameters for the models. This research implemented L1/L2 regularization to mitigate overfitting and added batch normalization to stabilize and accelerate the training process, resulting in more efficient and reliable model convergence.

Step 6(a) - TL and Domain Adaptation: TL and domain adaptation techniques further aligned the pre-trained models with the target IoT data, enhancing their applicability and accuracy. By fine-tuning pre-trained models on tasks such as specific IoT datasets, knowledge was effectively transferred, reducing the amount of data and computational resources required for training.

Step 6(b) - FL Setup: To implement FL, the Flower framework was used. Each IoT device (client) trained the model locally using its data and then sent the updated model parameters to a central server for aggregation. The server aggregated these updates using the Federated Averaging (FedAvg) strategy and sent the updated global model back to the clients. This process was repeated for multiple rounds. custom IoT Client class was defined to manage local training, evaluation, and communication with the central server. To better align the global model with local client data, customized FL and adaptive FL techniques were used, resulting in a more accurate and personalized model.

Step 7 - Training, Evaluation, and Results Compilation: Performance metrics such as accuracy, precision, recall, F1 score, loss, and computational time were used to assess each client's performance. Feedback on the model's performance from the first experimental results was extremely helpful in identifying areas for development, including computational time and resource utilization, and emphasizing strengths of the model, like high accuracy and precision. Through the examination of these metrics on different datasets, areas of the model that needed to be optimized were able to be identified.

IV. Experimental Results

To evaluate the performance, different IoT datasets are used to assess the framework's accuracy, scalability, and efficiency under different conditions. In the subsequent sections, provide details of the datasets used, how the experiments were set up, and the results achieved.

A. Dataset

To evaluate the performance and generalizability of the proposed federated TL framework for real-time intrusion detection in IoT networks, this reserch employed several well-known datasets such as BoT-IoT, N-BaIoT, TON_IoT, CICIDS 2017, and NSL-KDD, as in Table II which shows the five datasets used in details. The BoT-IoT dataset provides a comprehensive collection of simulated IoT network traffic including various attack types such as DDoS, OS attacks, service scanning, keylogging, and data exfiltration. Similarly, tagged data from nine IoT devices infected with the Gafgyt and Mirai botnets, capturing both normal and attack traffic, is provided by the N-BaIoT dataset. The TON_IoT dataset provides information on various cyber risks and their effects on IoT environments. It includes network traffic data, telemetry data from IoT devices, and logs of cyberattacks.

Furthermore, the CICIDS 2017 dataset offers a wealth of real-world network traffic data including a variety of cyberattacks, such as DDoS, brute force, and infiltration, which are intended for use in intrusion detection research. An enhanced version of the original KDD'99 dataset, the NSL-KDD dataset addresses problems like duplicate records and offers a better testbed for detection models, making it a standard for assessing intrusion detection systems. When combined, these datasets offer a broad basis for evaluating the framework's effectiveness in various IoT contexts and attack situations.

TABLE II. The Five Datasets used in this Research

Dataset	Number of Records	Number of Features	Number of Attacks & Types	Size	Publish Year	Environment	Link
BoT-IoT	72,000,000+	Various	DDoS, DoS, OS and Service Scan, Keylogging, Data exfiltration	69.3 GB (pcap), 16.7 GB (csv)	2020	Cyber Range Lab, UNSW Canberra (Impact Cyber Trust) (Papers with Code)	[27]
N-BaIoT	706,260 (total)	115	Botnet attacks on various IoT devices	Varies	2019	Real IoT devices and network	[28]
TON_IoT	Various	47	DoS, DDoS, Ransomware, various others	Various	2020	IoT Lab, UNSW Canberra	[29]
CICIDS 2017	3,119,345	80	DoS, DDoS, Brute Force, Web Attack, Infiltration, Botnet, etc.	20 GB	2017	Simulated corporate environment	[30]
NSL-KDD	125,973 (train+test)	41	DoS, R2L, U2R, Probe	~66 MB	2009	Simulated network environment	[31]

TABLE III. Enhanced Model Results for the Five Datasets

Dataset	Acc Mean	Acc Std	Precision Mean	Precision Std	Recall Mean	Recall Std	F1 Score Mean	F1 Score Std	Loss Mean	Loss Std	Time Taken Mean (s)	Time Taken Std (s)
BoT-IoT	0.92	0.01	0.91	0.02	0.93	0.01	0.92	0.01	0.25	0.01	5.3	0.5
N-BaIoT	0.93	0.01	0.92	0.01	0.94	0.01	0.93	0.01	0.22	0.01	4.8	0.4
TON_IoT	**0.94**	0.01	0.93	0.02	0.95	0.01	0.94	0.01	0.20	0.01	4.6	**0.3**
CICIDS2017	0.93	0.01	0.92	0.01	0.94	0.01	0.93	0.01	0.23	0.01	5.0	0.4
NSL-KDD	0.92	0.01	0.91	0.02	0.93	0.01	0.92	0.01	0.24	0.01	5.2	0.5

B. Results

This section shows the results of enhancing the proposed FTL method using five different IoT datasets: Bot-IoT, N-BaIoT, TON_IoT, CICIDS 2017, and NSL-KDD. Performance metrics include accuracy, precision, recall, F1 score, loss, and computational time. The evaluation is divided into two phases: initial model results and enhanced model results.

1) Initial model results: The first phase builds a strong model that learns from distributed IoT datasets while taking privacy into account, which combined Federated Learning (FL) and Transfer Learning (TL) without advanced optimization. The TON_IoT dataset achieved the best accuracy (89%), while the BoT-IoT dataset showed the lowest performance (85%).

Key Observations:

- Accuracy was consistent across datasets but below 90% for most.
- Loss values varied from 0.31 to 0.35, indicating opportunities for improvement.
- Prediction errors were more frequent in datasets with higher class imbalance, such as BoT-IoT and NSL-KDD.

2) Enhanced model results: The enhanced model incorporated advanced techniques such as SMOTE for data augmentation [24], hyperparameter tuning, and ensemble learning with CNNs, BiGRUs, and attention mechanisms. Table III illustrates significant improvements in accuracy (92%-94%) and loss reduction (0.20-0.25).

Key Highlights:

- Accuracy increased across all datasets, with TON_IoT reaching 94%.
- Low standard deviations in accuracy and loss values indicate stable performance.
- SMOTE effectively addressed class imbalance, improving precision and recall.

V. Discussion

The results in the previous section demonstrate the effectiveness of integrating FL with TL for intrusion detection in IoT networks. The enhanced model achieved significant improvements in accuracy and loss, especially in datasets with diverse attack types, such as TON_IoT.

- Scalability and Privacy: FL enabled decentralized training, maintaining data privacy while achieving consistent performance across IoT environments.
- Adaptability: TL enhanced the framework's ability to generalize across datasets, reducing training time and computational cost.
- Class Imbalance: The application of SMOTE balanced the datasets, preventing biases toward majority classes and improving detection accuracy.

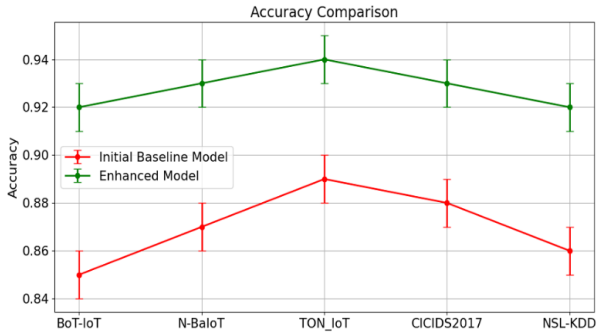


Fig. 2. Comparison of accuracy between the Initial Baseline and the enhanced models.

Fig. 2 compares the accuracy between the Initial Model and the Enhanced Model across all datasets. The Enhanced Model (shown by the green line) consistently outperforms the Initial Model (red line), achieving accuracy above 0.92 across all datasets, while the Initial Model remains below 0.89. The small error bars indicate minimal variability in accuracy across datasets, emphasizing the Enhanced Model's reliability in different IoT environments.

The Enhanced Model significantly improves accuracy on some datasets (such TON_IoT), while N-BaIoT and NSL-KDD exhibit a narrower performance gap. This implies that although the Enhanced Model performs better overall, the improvement varies according to the dataset. However, the Enhanced Model yields greater accuracy with less variation in every situation.

Fig. 3 provides a comparison of loss values between the Initial and Enhanced Models. Compared to the Initial Model's larger range of 0.31 to 0.35, the Enhanced Model has reduced loss values, ranging from 0.20 to 0.25. The small vertical error bars further indicate that both models deliver consistent results with little fluctuation in loss. The difference in loss is most evident in the TON_IoT dataset, where the Enhanced Model shows the greatest reduction. In datasets like BoT-IoT and CICIDS2017, the improvement in loss is less dramatic, but the Enhanced Model still outperforms the Initial Model. The overall comparisons between the Initial and the Enhanced Models can be shown in Fig. 4.

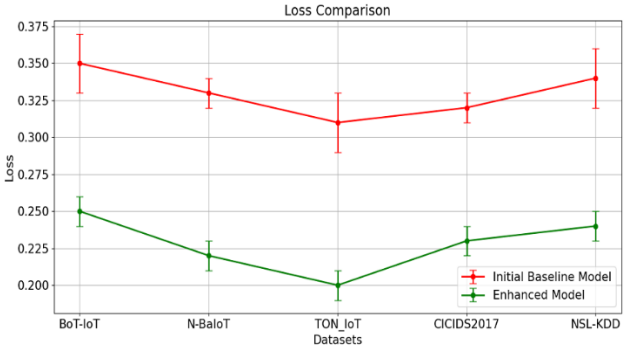


Fig. 3. The comparison of loss between the Initial Baseline and the enhanced models.

As shown in Fig. 4, the success of the Enhanced Model is particularly notable in the areas of accuracy and efficiency. FL and TL offer additional benefits beyond the performance metrics. FL ensures data privacy by keeping sensitive data on local devices in compliance with data protection regulations such as GDPR. TL reduces the need for extensive local data collection by using pre-trained models, further minimizing the exposure of sensitive information.

While the Enhanced Model consistently outperforms the Initial Model in terms of accuracy, the extent of this performance gap varies across datasets, where TON_IoT dataset showed the most significant improvements in both accuracy and loss. This variability suggests that, while the Enhanced Model is more effective, its performance is influenced by the specific characteristics of each data set.

Power consumption is a major issue in IoT environments. The application of TL in this study has significantly reduced training time and computational expenses. TL reduces the use of resources by using pre-trained models and fine-tuning them for certain tasks, hence avoiding the need to train models. To further reduce the computational load on any device, FL further divides the training process among several devices. With this method, training complex models on IoT devices with limited resources is possible while preserving efficiency and reducing power usage.

Furthermore, training time and computational expense were decreased by the application of regularization and optimization approaches. Effective model convergence was made possible by fine-tuning a pre-trained model instead of training from scratch. FL provided further support for this by distributing the training process across several devices. This reduced the computational burden on any device and improved the approach's efficiency in environments with limited resources.

Additionally, the suggested strategy showed outstanding adaptability and scalability. FL made it possible to train on numerous IoT devices at once, which increased the system's scalability. TL made it easier to quickly adjust to new settings and devices, ensuring effective scaling to a range of IoT scenarios. Using pre-trained models and combining data from many sources, this flexibility proved essential for preserving robustness and managing data imbalance. The Enhanced Model assisted real-time learning and adaptation. IoT devices might periodically update the global model with new observations, allowing TL to fine-tune the model with minimum input and enabling faster adaptability to new types of attacks or anomalies.

VI. Conclusions and Future Directions

The combination of FL and TL for real-time intrusion detection in IoT networks provides an effective solution for urgent security issues. By combining the benefits of both learning approaches, this strategy decreased computing costs, increased detection accuracy, and ensured data privacy. The experimental results show that the proposed framework can minimize training time and improve detection performance by adapting pre-trained models to specific IoT contexts. Accuracies ranging from 92% to 94% were achieved across many datasets.

To detect and mitigate cyber risks in a scalable, effective, and privacy-preserving manner, this research highlights how federated transfer learning might transform cybersecurity measures in IoT networks. Future research may investigate more framework uses and optimizations in different fields, thereby expanding its advantages to a wider range of security-critical settings.

Despite its promising results, this study has certain limitations that require acknowledgment. Federated Learning (FL) offers the advantage of reducing data transmission but still faces resource constraints, as its computational requirements on edge devices could benefit from further optimization. Additionally, addressing dynamic threats remains a challenge, highlighting the need for future research to explore adaptive learning techniques capable of responding to evolving IoT security risks. Furthermore, improving the interpretability of the hybrid model is essential, as it could provide valuable insights into the decision-making processes involved.

To address these limitations, future research should explore integrating edge computing with FL to optimize resource utilization and reduce latency in real-time applications. Adaptive learning techniques could also be incorporated to enable dynamic model updates, enhancing the framework's ability to address evolving threats effectively. Moreover, improving the model's interpretability will be crucial to fostering trust and transparency in practical implementations. Finally, extending the framework to broader domains, such as industrial IoT and smart cities, will help validate its scalability and robustness in handling large-scale and complex environments.

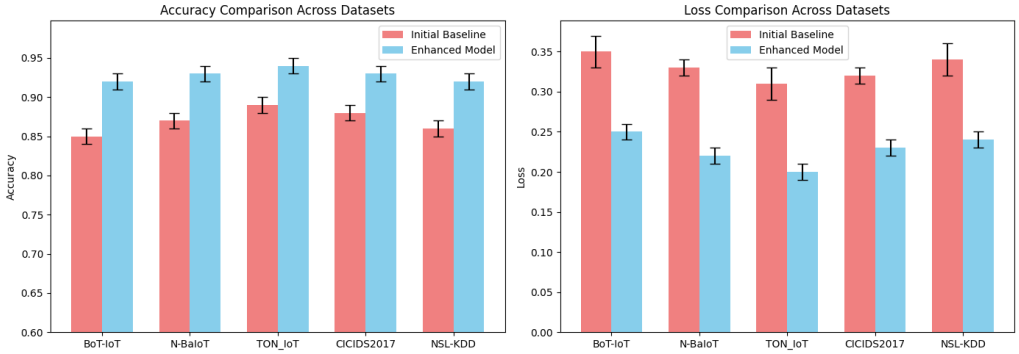


Fig. 4. Overall comparisons between the Initial and the Enhanced Models across all datasets.

REFERENCES

[1] Z. Almesleh, A. Gouissem and R. Hamila, "Federated Learning with Kalman Filter for Intrusion Detection in IoT Environment," 2024 IEEE 8th Energy Conference (ENERGYCON), Doha, Qatar, 2024, pp. 1-6, doi: 10.1109/ENERGYCON58629.2024.10488796.

[2] Salunkhe UR, Mali SN. Security enrichment in intrusion detection system using classifier ensemble. Journal of Electrical and Computer Engineering. 2017;2017(1):1794849.

[3] Vengatesan K, Kumar A, Naik R, Verma DK. Anomaly based novel intrusion detection system for network traffic reduction. In2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on 2018 Aug 30 (pp. 688-690). IEEE.

[4] T. Ohtani, R. Yamamoto and S. Ohzahata, "Detecting Zero-Day Attack with Federated Learning Using Autonomously Extracted Anomalies in IoT," 2024 IEEE 21st Consumer Communications & Networking Conference (CCNC), Las Vegas, NV, USA, 2024, pp. 356-359, doi: 10.1109/CCNC51664.2024.10454669.

[5] Khraisat A, Gondal I, Vamplew P, Kamruzzaman J. Survey of intrusion detection systems: techniques, datasets and challenges. Cybersecurity. 2019 Dec;2(1):1-22.

[6] Mohammad RM, Alsmadi MK, Almarashdeh I, Alzaqebah M. An improved rule induction based denial of service attacks classification model. Computers & Security. 2020 Dec 1;99:102008.

[7] Muneer S, Farooq U, Athar A, Ahsan Raza M, Ghazal TM, Sakib S. A Critical Review of Artificial Intelligence Based Approaches in Intrusion Detection: A Comprehensive Analysis. Journal of Engineering. 2024;2024(1):3909173.

[8] Saied M, Guirguis S, Madbouly M. Review of artificial intelligence for enhancing intrusion detection in the internet of things. Engineering Applications of Artificial Intelligence. 2024 Jan 1;127:107231.

[9] Alsmadi MK, Mohammad RM, Alzaqebah M, Jawarneh S, AlShaikh M, Al Smadi A, Alghamdi FA, Alqurni JS, Alfagham H. Intrusion Detection Using an Improved Cuckoo Search Optimization Algorithm.

[10] Latif S, Boulila W, Koubaa A, Zou Z, Ahmad J. Dtl-ids: An optimized intrusion detection framework using deep transfer learning and genetic algorithm. Journal of Network and Computer Applications. 2024 Jan 1;221:103784.

[11] Zhu J, Liu X. An integrated intrusion detection framework based on subspace clustering and ensemble learning. Computers and Electrical Engineering. 2024 Apr 1;115:109113.

[12] Zhang C, Xie Y, Bai H, Yu B, Li W, Gao Y. A survey on federated learning. Knowledge-Based Systems. 2021 Mar 15;216:106775.

[13] Li L, Fan Y, Tse M, Lin KY. A review of applications in federated learning. Computers & Industrial Engineering. 2020 Nov 1;149:106854.

[14] Khan LU, Saad W, Han Z, Hossain E, Hong CS. Federated learning for internet of things: Recent advances, taxonomy, and open challenges. IEEE Communications Surveys & Tutorials. 2021 Jun 18;23(3):1759-99.

[15] M. Umair, W. -H. Tan and Y. -L. Foo, "Dynamic Federated Learning Aggregation for Enhanced Intrusion Detection in IoT Attacks," 2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), Osaka, Japan, 2024, pp. 524-529, doi: 10.1109/ICAIIC60209.2024.10463247.

[16] Iman M, Arabnia HR, Rasheed K. A review of deep transfer learning and recent advancements. Technologies. 2023 Mar 14;11(2):40.

[17] Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. Journal of Big data. 2016 Dec;3:1-40.

[18] A. U. Karimy and P. C. Reddy, "Analyzing Federated Learning as a novel approach for enhancing security and privacy in the Internet of Things (IoT)," 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India, 2024, pp. 1-7, doi: 10.1109/ICAECT60202.2024.10468686.

[19] Y. Luan, "Network Traffic Anomaly Detection Based on Federated Learning," 2024 4th International Conference on Neural Networks, Information and Communication Engineering (NNICE), Guangzhou, China, 2024, pp. 224-228, doi: 10.1109/NNICE61279.2024.10498908.

[20] M. H. Bhavsar, Y. B. Bekele, K. Roy, J. C. Kelly and D. Limbrick, "FL-IDS: Federated Learning-Based Intrusion Detection System Using Edge Devices for Transportation IoT," in IEEE Access, vol. 12, pp. 52215-52226, 2024, doi: 10.1109/ACCESS.2024.3386631.

[21] H. Babbar and S. Rani, "FRHIDS: Federated Learning Recommender Hybrid Intrusion Detection System Model in Software-Defined Networking for Consumer Devices," in IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 2492-2499, Feb. 2024, doi: 10.1109/TCE.2023.3329151.

[22] A. Raj, V. Sharma, S. Rani, A. K. Shanu and N. Kumar, "Strengthening the Security of IoT Devices Through Federated Learning: A Comprehensive Study," 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2024, pp. 1-5, doi: 10.1109/ICRITO61523.2024.10522388.

[23] M. Al-Hawawreh and M. S. Hossain, "Federated Learning-Assisted Distributed Intrusion Detection Using Mesh Satellite Nets for Autonomous Vehicle Protection," in IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 854-862, Feb. 2024, doi: 10.1109/TCE.2023.3318727.

[24] Andresini G, Appice A, De Rose L, Malerba D. GAN augmentation to deal with imbalance in imaging-based intrusion detection. Future Generation Computer Systems. 2021 Oct 1;123:108-27.

[25] Elreedy, D.; Atiya, A.F. A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. Inf. Sci. 2019, 505, 32–64.

[26] Mohammad R, Saeed F, Almazroi AA, Alsubaei FS, Almazroi AA. Enhancing Intrusion Detection Systems Using a Deep Learning and Data Augmentation Approach. Systems. 2024 Mar 1;12(3):79.

[27] BoT-IoT Dataset. (2018). Retrieved from https://research.unsw.edu.au/projects/bot-iot-dataset

[28] Kashif, M. (2019). N-BaIoT Dataset. Kaggle. Retrieved from https://www.kaggle.com/datasets/mkashifn/nbaiot-dataset/code

[29] TON_IoT Dataset. (2020). Retrieved from https://research.unsw.edu.au/projects/ton-iot-datasets

[30] Huhn, C. (2017). CICIDS 2017. Kaggle. Retrieved from https://www.kaggle.com/datasets/chethuhn/network-intrusion-dataset

[31] Hassan, M. (2019). NSL-KDD Dataset. Kaggle. Retrieved from https://www.kaggle.com/datasets/hassan06/nslkdd

Forecasting Unemployment Rate for Multiple Countries Using a New Method for Data Structuring

Amjad M. Monir Aljinbaz[1], Mohamad Mahmoud Al Rahhal[2]
Department of Computer Science, University of York, York, UK[1]
Department of Applied Computer Science-College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia[2]

***Abstract*—Forecasting the Unemployment Rate (UR) plays a key role in shaping economic policies and development strategies. While most research focuses on predicting UR for individual countries, there has been limited progress in creating a unified forecasting model that works across multiple countries. Traditional time series methods are usually designed for single-country data, making it difficult to develop a model that handles data from various regions. This study presents a new data structuring technique that divides time series into smaller segments, enabling the development of a single model applicable to 44 countries using various economic indicators. Four forecasting models were tested: an artificial neural network (ANN), a hybrid ANN with machine learning (ML), a genetic algorithm-optimized ANN (ANN-GA), and a linear regression model. The linear regression model, which used lagged UR values, delivered the best results with an R^2 of 0.964 and 89.8% accuracy. The ANN-GA model also performed strongly, achieving an R^2 of 0.945 and 85.1% accuracy. These results highlight the effectiveness of the proposed data structuring method, demonstrating that a single model can accurately forecast multiple time series across different regions.**

***Keywords*—Unemployment rate; artificial neural network; time series; hybrid model; genetic algorithm**

I. Introduction

Unemployment rate (UR) is a critical macroeconomic indicator that plays a pivotal role in economic planning, social stability, and policy development [1]. Researchers have used several methods to forecast unemployment rate using stochastic methods, machine learning and neural network. However, most of these approaches are focused on developing country-specific models. It is rare to find studies attempting to build a global model applicable across multiple countries. Because of the importance of building global model for multiple countries, an attempt was made to build such [2].

Employment is inherently linked to macroeconomic conditions and is influenced by a broad spectrum of factors, including inflation, gross domestic product (GDP), and other economic indicators[3]. At the same time, microeconomic factors such as age, educational attainment, gender, and poverty status can significantly impact unemployment dynamics [4], [5], [6]. Consequently, understanding the UR requires a multifaceted approach that accounts for both macro- and microeconomic influences.

The development of single-country models has been prevalent in previous research, utilizing a variety of techniques. Several studies used stochastic time series techniques. In such techniques, they study the effect of independent variables on dependant variables regression to examine level of significance [7], [8].

Also, time series analysis, including autoregressive integrated moving average (ARIMA) and its variants (e.g., ARMA, SARIMA, VARIMA), have been employed to capture the linearity and stationarity in UR data over time [9], [10], [11]. However, these methods are inherently limited to individual country forecasts. Artificial neural networks (ANN) were also used to analyse time series. This was either done using feedforward neural network (FNN) [12], or recurrent neural network RNN which is able to capture more complex patterns [13].

A combination of different models forms hybrid models which were also used to analyse UR. Such models leveraging the strengths of both linear and non-linear modelling capability. Hybrid model includes using ARIMA-RNN [13], [14], RNN with genetic algorithm GA for feature optimization [13].

This research aims to build a comprehensive model that forecasts UR across multiple countries. Leveraging a new method for structuring time series data, this study addresses the challenge of building a single forecasting model applicable to multiple regions. The data, sourced from the World Bank, includes a diverse set of macroeconomic and microeconomic features for 44 countries over a span of 33 years. By segmenting the time series into smaller data slices, the proposed approach allows for the consolidation of country-specific data into a unified dataset, facilitating the development of a generalized forecasting model.

The remainder of this paper is organized as follows: Section II reviews the related work on UR forecasting methodologies, including machine learning, autoregressive models, and hybrid approaches. Section III details the methodology of the data structuring process and model development. Section IV presents the results and evaluates the performance of each model. Section V discusses the implications of the findings, and Section VI concludes the paper with future research directions.

II. Related Work

The study of unemployment rate (UR) forecasting has evolved significantly over the past decades. Researchers have applied various methods, primarily categorized into traditional machine learning regression, autoregressive time series forecasting, neural networks, and hybrid models. Each approach has distinct advantages and has contributed to understanding the dynamics of UR within individual countries.

A. Machine Learning Regression for Forecasting

Several studies used stochastic time series techniques in forecasting. In such techniques, they study the effect of independent variables on dependant variables to examine if there is significant difference. Zawojska used linear regression to study the correlation between several macroeconomic factors and UR and found significant correlation with many factors such as GDP [7]. In [15] Alternative regression methods, including the Toda-Yamamoto procedure, have been utilized to examine the impact of specific factors, such as oil prices and interest rates, on the unemployment rate (UR) in different scenarios.

Before applying these techniques, researchers often test the data for stationarity to ensure appropriate modeling. For example, a study involving 10 developing countries found that UR, GDP, population, remittances, external debt, exchange rate and expenditure on education were non-stationary, necessitating differencing for proper regression analysis [16]. This type of tests is also applied on autoregression time series forecasting (see B).

B. Autoregression Time Series Forecasting

Autoregressive time series methods focus on the inherent symmetry or asymmetry, as well as the linearity or non-linearity, of UR over time. If a time series exhibits stationarity—consistent mean or variance over time—these methods can effectively forecast future trends[17]. It was suggested that UR has hysteresis, which means that it stays in equilibrium level with equality likely movement in either direction [18]. For instance, the logistic smooth transition autoregressive (LSTAR) model has been used to analyze non-linear and asymmetric UR data in Australia [19]. In this study, model accuracy was high due to LSTAR ability to capture cyclic fluctuations of business and UR as result. Other studies analyse the effect over years using Autoregressive Distributed Lag Stationarity ADLS [20].

Autoregressive integrated moving average (ARIMA) model is very famous for time series forecasting. A modified version (SARIMA) captures seasonality and used in case time series has seasonality element. Another variation is generalised autoregressive conditional heteroscedasticity (GARCH) that analyse the volatility of time series [17]. The authors in [9] A study of monthly UR to study the impact of COVID-19 in 5 Asian countries (ASEAN-5) using ARIMA, SARIMA and GARCH found that ARIMA and SARIMA are superior in forecasting UR. Another study [21] using ARIMA model with monthly data in Australia was found to predict on month ahead have R^2 of 0.96.

C. Artificial Neural Network for Forecasting

Neural networks (NNs) have emerged as powerful tools for forecasting time series data, particularly when the data exhibits non-linear characteristics [22], which is found in UR [23]. Feedforward neural networks (FNN) have been applied both to evaluate independent variables affecting UR and to utilize lagged values (k-lags) for one-step ahead forecasting [24]. A study applied FNN by using several macroeconomic factors including GDP, inflation and other factors to forecast UR in Philippine using a data from 1991 to 2014 [25] with accuracy of 87.5%. Another study used neural network with backpropagation to predict UR, gross national product and employee number in Germany and found that the model has R^2 close to 1 when predicting three lags ahead (t+3) [12]. In [17] Recurrent neural networks (RNN) and their variations, such as long short-term memory (LSTM), have demonstrated higher accuracy in forecasting due to their internal memory capabilities, which are well-suited for for time series forecasting.

D. Hybrid Models

Hybrid models combines different types of models to leverage the strengths of both linear and non-linear modeling techniques to improve forecasting accuracy [26]. For example, combining two models where one of them to capture linear elements (such as ARIMA or regression) and other to capture nonlinear elements (such as artificial neural network) could provide a better performance model. Other studies used hybrid models differently by combining neural network models, with feature optimization models, such as genetic algorithms [13]. Hybrid models has been used for time series forecasting, such as stock market price forecasting [27]. It was also used to forecast UR such as combining ARIMA and RNN [14]. They used ARIMA in the first phase to catch linear pattern, followed by autoregressive neural network ARNN to analyse nonlinearity and nonstationarity patterns. This hybrid model outperformed either of these models alone. Hybrid models can be employed for feature optimization and integrated with other approaches. In [13] LSTM models enhanced with genetic algorithms (GA) for parameter optimization have demonstrated superior performance compared to traditional RNNs in unemployment rate forecasting for Ecuador, delivering greater accuracy.

E. Gaps in the Literature

Previous studies were used to build separate UR forecasting models for individual countries. The reason for this is that they used time series forecasting, where data is arranged on a continuous line and is split for training and testing the model. This way makes it not possible to incorporate multiple countries inti the same model. This research addresses this gap by proposing a novel data structuring technique, enabling the construction of a single forecasting model for multiple countries.

III. Data and Method

This study aims to develop a unified AI model that leverages macroeconomic and microeconomic data across multiple countries to predict unemployment rates. While previous research has focused on models for individual countries [8], [16], [28], this study introduces a novel data structuring approach to create a forecasting model applicable to various economic contexts and regions.

A. Data Analysis

This study utilized data from the World Bank [29], covering annual macroeconomic and microeconomic indicators from five key datasets: economy and growth, education, urban development, health, and climate change. The data spans the years 1991 to 2023, with the analysis focusing on 44 countries that have complete records for this period. Initially, 162

features were drawn from the datasets, but after excluding those with missing values, 153 features were retained. Table I outlines the datasets, their feature counts, and examples of the included variables.

TABLE I. DATASETS USED IN THE RESEARCH, AND EXAMPLES OF THEIR FEATURES

Dataset	Feature count	Examples of features
Economy & Growth [30]	81	GDP, current account balance and exports of goods and services
Education [31]	5	UR, government expenditure on education, school enrolment, literacy rate and population age groups
Urban Development [32]	7	mortality by traffic, PM2.5 air pollution, population in large cities and urban population
Health [33]	62	fertility rate per age group, birth rate, cause of death by injury, cause of death by communicable diseases and death rate
Climate Change [34]	7	access to electricity, agriculture land, cereal yield, annual fresh water, CO2 emission and NO emissions

B. Data Structuring

A new method of structuring the time series data was developed to enable multiple time series forecasting within a single model. The data were segmented into rolling windows or "slices," each containing three years of lagged values for each feature and UR value for target year. For example, the UR for a given year (y) was predicted using the three preceding years (y-1), (y-2), (y-3) for each feature. This means that 3 columns are added per feature to represent $feature_{y-1}$, $feature_{y-2}$ and $feature_{y-3}$ to forecast UR_y. And because we have data from 1991 to 2022, we can have 3 lags to forecast UR starting from 1994 to 2022. (i.e. the first UR will have feature_a_{1991}, feature_a_{1992} and feature_a_{1993} to forecast UR_{1994}). This approach resulted in 29 slices per country, allowing the aggregation of all 44 countries into a unified dataset with 1,276 observations. The reason for using 3 years of lagged features is that factors might have effect on UR for multiple years in the future [7], [16]. This way, the forecasting model will predict UR of any year (y) based on previous n years lag (y-3) of any given feature.

This method effectively addresses the limitations of traditional time series forecasting, which typically requires contiguous, long time series data and is not easily adaptable for multi-country analysis.

The data were imported using Python's Jupyter Notebook, leveraging Pandas and NumPy APIs. Each dataset was structured into a multi-index DataFrame, where level 0 represented the country, and level 1 denoted the year. This organization ensured that each country had 33 rows, corresponding to the years 1991 to 2023. The data for 2023 was excluded from the analysis to serve as out-of-sample validation [35]. Fig. 1 provides a snapshot of the DataFrame used, which contained 1276 observations and 153 features. The total number of columns includes 1 target variable (unemployment rate), along with 3 lags per feature for each of the 153 features.

Country	Indicator Code / Year	Unemployment	NE.CON.GOVT.CD y-1	NE.CON.GOVT.CD y-2	NE.CON.GOVT.CD y-3	...
Argentina	1994	11.76	31984701702	6807403684	6302044992	...
	1995	18.8	33948366600	31984701702	6807403684	...
	1996	17.11	34445834100	33948366600	31984701702	...
	1997	14.82	34023284500	34445834100	33948366600	...
...	...	...	...	...	...	...
United Kingdom	2018	4	5.00097E+11	5.168E+11	5.72364E+11	...
	2019	3.74	5.32384E+11	5.00097E+11	5.168E+11	...
	2020	4.472	5.44508E+11	5.32384E+11	5.00097E+11	...
	2021	4.826	6.09767E+11	5.44508E+11	5.32384E+11	...
	2022	3.73	6.99684E+11	6.09767E+11	5.44508E+11	...

1276 rows x 460 columns

Fig. 1. Image of the actual DataFrame used in the research.

C. Model Development

As outlined in Table II, four AI-based models were developed and evaluated:

Model 1 (ANN-All): Applied artificial neural networks (ANN) to all available features.

Model 2 (Hybrid ANN-ML): An ensemble model where separate neural networks were trained on each of the five datasets, and their outputs served as inputs for a machine learning regression model.

Model 3 (Hybrid ANN-GA): A hybrid approach integrating genetic algorithms for feature selection with an ANN for forecasting.

Model 4 (UR-ML): A machine learning model relying solely on lagged unemployment rate (UR) values (y-1, y-2, y-3) to forecast the following year's UR.

TABLE II. MODELS AND FEATURES USED

Model	*# Datasets*	*# countries*	*# features*
ANN-All	5	44	153
ANN-ML	5	44	153
ANN-GA	5	44	153
UR-ML	Unemployment Rate only	44	1 (UR)

The models were trained and validated on the structured dataset, with the testing conducted on out-of-sample data for 2023. Training and validation were split into 77.3% and 19.4% of the data, respectively, while the testing set comprised 3.3%.

D. Performance Evaluation of Methods

The models' performance was evaluated using four metrics:

1) The coefficient of determination (R2): Assess the proportion of variance explained by the model. One of the main advantages of R^2 is that there is no need to compare the result with other models to evaluate model fit [36]. Best value equals 1. This was the main metric used in this paper.

2) *Accuracy: defined as (% 1 – MAPE) [37].* Mean Absolute Percentage Error (MAPE) has high efficiency to minimize risk in regression forecasting, especially if target value is always positive (which is the case with UR) [38], [39]. Using (% 1 – MAPE) provides easy way to assess model accuracy. Best value equals 100%.

3) *Root mean squared error (rmse) [36]: Best value equals 0.*

4) *Mean absolute error (MAE) [36]: Best value equals 0.*

IV. Experimental Results

This section represents the results of each of the models used in the research and evaluate their performance.

A. Results

The performance of each model was assessed on validation datasets using R^2, accuracy (1-MAPE), RMSE and MAE as summarized in Table III and IV. Among the models, the UR-ML model, which uses only lagged UR values, demonstrated the best performance, achieving an R^2 of 0.964 and 0.989 on validation and testing data, respectively. Its accuracy was 89.8% on validation data and 94.2% on testing data, indicating strong generalizability and robustness.

Following is a summary of each model performance.

1) *Model 1: ANN-All – Neural network for 5 datasets:* This model used 5 datasets with a total of 153 features to predict UR. This model performs better than model in in all 4 measures. In validation data, this model scored 0.927 and 82.1% of R^2 and accuracy respectively. These measures fell greatly to 0.857 and 71.4% respectively which means that the model has issue with generalization. The scatter plot in Fig. 2 plotted predicted versus true value for this model.

2) *Model 2: ANN-ML – Hybrid model using ANN and ML:* This model built five individual Neural Network model for each of the five datasets to predict UR. Fig. 3 shows the result of these five individual models to predict UR.

TABLE III. Performance of Models for Forecasting UR (Validation Data)

Model	R^2	Accuracy	RMSE	MAE
ANN-All	0.927	82.1%	1.217	0.894
ANN-ML (hybrid)	0.921	77.4%	1.256	0.985
ANN-GA (hybrid)	0.945	85.1%	1.049	0.775
UR-ML	0.964	89.8%	0.851	0.563

TABLE IV. Performance of Models for Forecasting UR (Testing Data)

Model	R^2	Accuracy	RMSE	MAE
ANN-All	0.857	71.4%	1.745	1.1401
ANN-ML (hybrid)	0.86	72.8%	1.729	1.365
ANN-GA (hybrid)	0.919	80.4%	1.316	0.986
UR-ML	0.989	94.2%	0.487	0.316

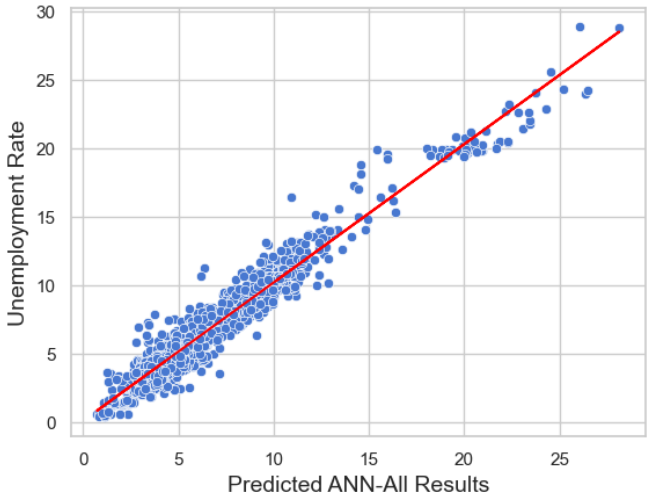


Fig. 2. ANN-All Predicted vs. True value.

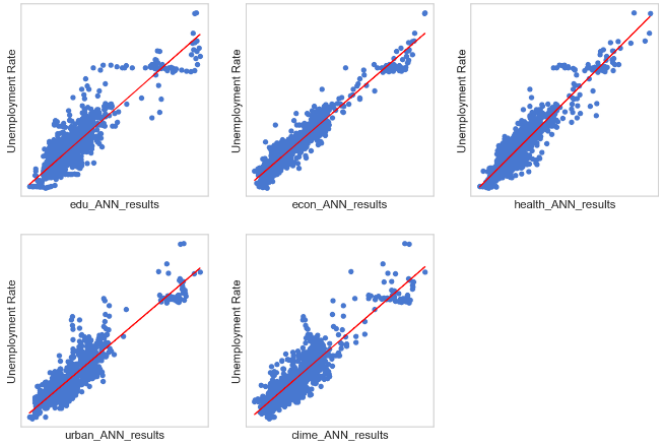


Fig. 3. The result of five individual ANN models to forecast unemployment rate.

Then, the resulted five predictions were fitted using machine learning linear regression to predict UR. The model has R^2 value of 0.921 which is similar to ANN-All, but has worst accuracy value of 77.4%. The model is also slightly worse than ANN-All in terms of RMSE and MAE. The model, however, is a little better in terms of generalization compared to ANN-All, as R^2 and accuracy for testing data were 0.86 and 72.8% respectively. Fig. 4 shows the predicted versus true value for this model.

3) *Model 3: ANN-GA:* This model used genetic algorithm for feature optimization, followed by building neural network model using best fitness features measured by R^2. Based on genetic algorithm result, 83 out of 153 features were selected and then used to build neural network model. The model result was better than all previous 3 models as it has R^2 of 0.945 and accuracy of 85.1% in validation data, and R^2 of 0.919 and accuracy of 80.4%. This means that those 83 features provide better performance compared to using all 153 features as in Model 1 ANN-All. Fig. 5 shows predicted versus true value for this model.

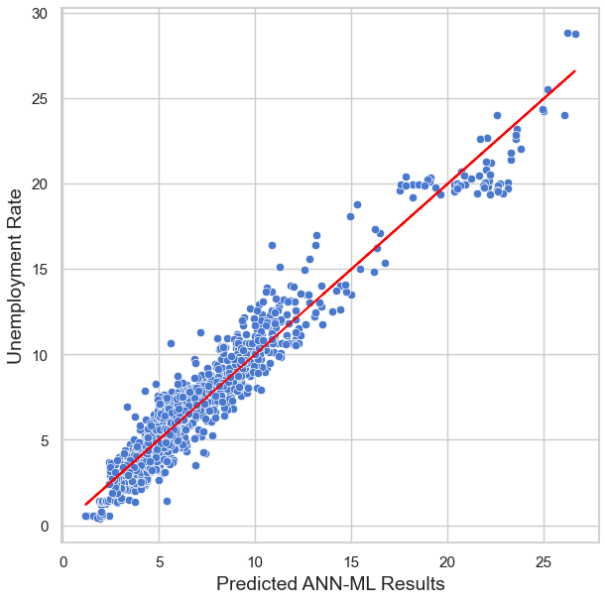


Fig. 4. ANN-ML Predicted vs. True value.

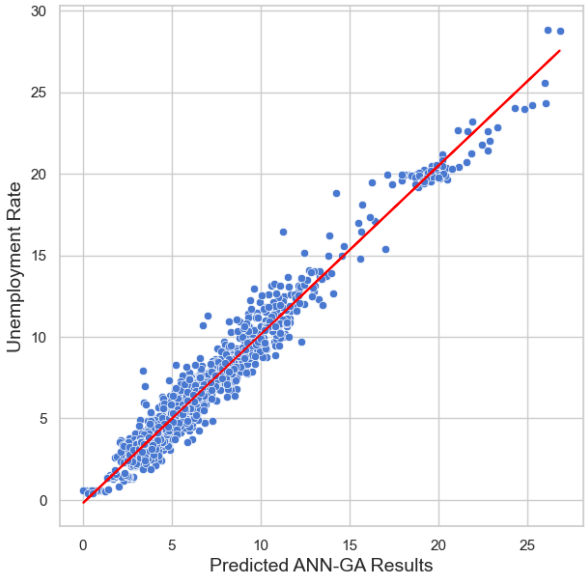


Fig. 5. ANN-GA Predicted vs. True values.

4) Model 4: UR-ML – Using ML regression for UR: The fifth model used only 1 feature, UR. And as all previous models, it uses this feature with 3 lags (UR_{y-1}, UR_{y-2} and UR_{y-3}) to predict UR_y. This model has shown the best overall performance in both validation data and testing data. In validation data it has R^2 of 0.964 and accuracy of 89.8%. This model demonstrated high generalization capability, as it scored 0.989 in R^2 and 94.2% in accuracy. This means that this model doesn't suffer from overfitting. This model also has the least score in RMSE and MAE, which were 0.487 and 0.316 respectively.

To illustrate the reason behind the high performance of this model, Fig. 6 shows the correlation of the n lags of UR (y-1, y-2 and y-3) with true UR. Each one of UR_{y-1}, UR_{y-2} and UR_{y-3} has high correlation with UR_y, making it very suitable for UR forecasting. Fig. 7 shows the predicted versus true value for this model.

Fig. 6. Correlation between n lags UR and True UR.

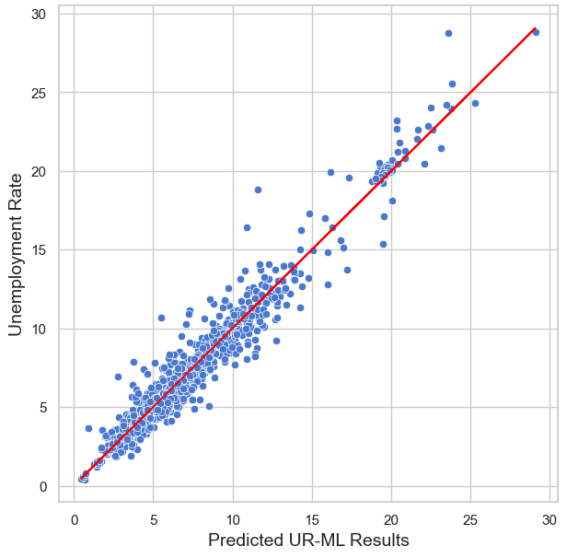


Fig. 7. UR-ML Predicted vs. True value.

B. Discussion

This study demonstrates that it is feasible to develop a single AI model that can accurately forecast UR across multiple countries using economic factors. The effectiveness of the new data structuring method, which segments time series into smaller slices, allows for a generalized model that maintains high accuracy and generalizability.

1) Discussing models performance: Neural network models are well-suited for UR forecasting due its nonlinear nature [23]. This research found that it is possible to build single neural network model that can be used effectively to forecast UR across several countries.

When comparing with other researches that use neural network model, we found that our model with R^2 of 0.87 and accuracy of 71.4% is compared to other models. Misil and Tarepe have built ANN model with 11 features (including GDP and inflation rate) to forecast UR in Philippine from 1991 to 2014 and has accuracy of 87.5% [25]. Their model has higher accuracy due to building the model for single country.

Our research found that hybrid model with genetic algorithm outperforms other models neural network models in forecasting UR as it reduce the number of features used and hence reduce model noise. This is consistent with other research

that found this models superior to other models they compared [13], which used LSTM (Long Short Term Memory) with GA. However, their research has accuracy of 96% compared to our 80.4%. The superiority of their results is due to building the model on single country only (Ecuador) in addition to include UR (UR of previous years) as a feature, which will yield better results as will be discussed in 2).

The other hybrid model used in the research was ANN-ML, which combines neural network forecasting with machine learning regression. This model showed inferior results especially in terms of generalization on testing data with accuracy of 72.8%. This is not consistent with researches who used hybrid models in UR forecasting [14]. The reason is that we used different hybrid model. Chakraborty et al. [14] has combined ARIMA with RNN to study UR in individual countries, which wasn't applicable in our research to build single model for multiple countries.

Several researchers have found much superior results with different hybrid models. A study on individual countries including US, UK, Italy and France had accuracy ranges from 95.34 to 98.91% based on the country using a hybrid model of LSTM-GRU [40]. Another studied Asian countries using ARIMA-ANN model and had accuracy of 96.4 - 97.8% [41]. However, as these models used previous UR data only in forecasting, these should be compared to our UR-ML model (discussed in 2)), and their result is comparable to our 94.2% for UR-ML model.

2) Performance of ML model using UR only: Using UR from the past to forecast future UR is very common across research. Research have used several methods for such type of analysis including ARIMA (and its variations including ARMA, SARIMA, FARIMA, etc.), RNN (and its variations including GRU, LSTM, etc.).

In our research, the regression model that only uses UR (UR-ML) outperforms all other models. The model used 3 lags years (UR_{y-1}, UR_{y-2} and UR_{y-3}) to forecast UR_y result. The model has R^2 of 0.964 and accuracy of 89.8% on validation data. The model performs even better on testing data to forecast 2023 results with R^2 of 0.989 and accuracy of 94.2% which means it has high generalizability.

This result is aligned with many researches that were able to use previous UR solely to forecast future UR. Our research outperformed Davidescu et al. research who have used SARIMA model to forecast UR in Romania, and has accuracy of 86.6%. The same research has even worse accuracy using SETAR model, with accuracy of 84.23% [39]. As ARIMA model requires a continuous time series, and then using 80% of it for training and keeping the last 20% for testing. This makes the model more vulnerable to macro trends that occurs in last year that didn't exist in training data.

Higher accuracy could be obtained by building separate model for separate country or district. In Germany, accuracy ranged from 91.91% to 99.23% using SARIMA model [42]. Our model has slightly inferior results despite being built for multiple countries.

One of the reasons of why UR from previous year is a strong indicator of future UR, is that UR tends to change gradually and slowly over years, as it tend to stay in equilibrium [18]. To test this, we have created a dummy forecaster, where it forecast UR_y by simply returning UR_{y-1}. Fig. 8 shows a scatter plot of this Dummy model. This dummy model has R^2 of 0.958 and accuracy of 90.7%, which outperforms all 3 ANN models in this research, which hasn't included lagged UR as feature.

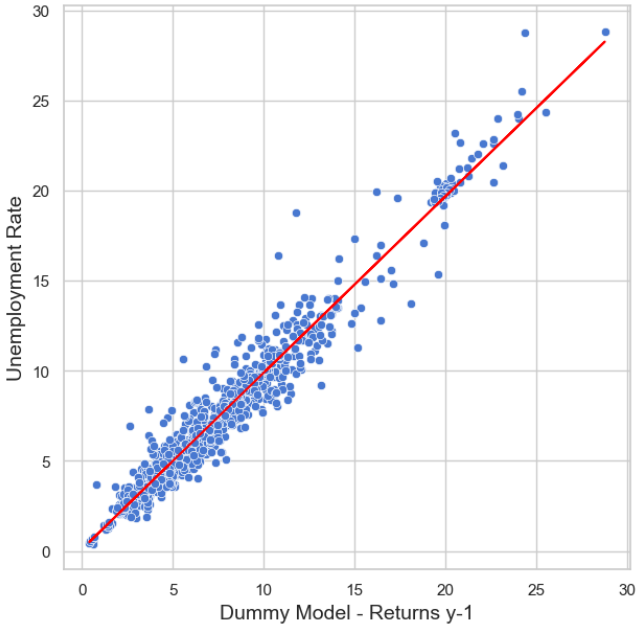


Fig. 8. Dummy Model that uses Y-1 as forecast value for Year Y.

For this reason, this research excluded UR from all ANN models, due to the high correlation between UR_{y-n} and UR_y. Also, this is the possible reason why ARIMA models have higher accuracy in one-step-ahead forecasting compared to multiple-steps-ahead.

3) Benefits of using new data structure: Methods like ARIMA and RNN mandates the presence of long and contiguous the time series [38], [43]. Because we only have 33 data point per country, we need to increase the number of observations to fit the model.

This research has used new way of structuring data by transforming continuous time series into an array of instances that contains slices of 3 years of lagged data. This method has the following benefits:

- This method enables mixing the data randomly for fitting the model. This has advantage of training data on a mix of time periods and make prediction more generalizable.
- As the Covid pandemic has made significant change in the labour market in 2020 and 2021 [9], [39].This make forecasting issue if out-of-sample forecasting included this period. This accounts for generalizability issue of the model and big difference between training and testing data conditions [39].

- Previous studies were forced to analyse each country separately [8], [14], [16], [28], [38], [40], [41], [44], [45], or even on separate county or city level [46] because of the limitation of ARIMA and RNN models. The new method enables us to train the model on multiple countries. This resulted in successfully building single forecasting model for multiple countries (in our case 44 countries).
- New approach enables us to have large number of observations to train and validate the model. Other studies either used the low number of instances [25] or increased the number of instances by using months instead of years [23], [39], [46]. Using monthly time periods increases the number to a certain degree, while adding another layer of complexity due to seasonality of UR.

However, this approach keeps memory of n lags of years to perform one step ahead forecasting. Other time series analysis methods keep longer memory which may increase the accuracy of multiple steps ahead forecasting [12].

V. CONCLUSION

This study successfully developed a robust AI-driven model for forecasting unemployment rates (UR) across multiple countries by employing an innovative data structuring technique. By segmenting time series data into smaller slices and aggregating data from 44 countries, the study produced a unified model demonstrating high accuracy and generalizability. Among the tested models, the UR-ML model emerged as the best performer, with the hybrid ANN-GA model also showing considerable potential for feature optimization. The proposed method holds promise for forecasting other economic indicators, such as GDP and inflation. However, this study is not without limitations. Firstly, the dataset excluded data from 2023 as it was reserved for out-of-sample validation, limiting the availability of real-time forecasting scenarios. Secondly, the exclusion of additional regional and socio-economic indicators, due to data availability constraints, may have impacted the model's ability to fully capture cross-country variations. Future research should explore advanced hybrid models, such as incorporating deep learning architectures (e.g., LSTM-GA combinations) to improve predictive capabilities further. Expanding the feature set to include additional socio-economic and environmental indicators may enhance accuracy and capture regional variations better. Additionally, extending the model to support multi-step-ahead forecasting and applying the proposed data structuring method to other macroeconomic indicators, such as inflation or GDP, would demonstrate its versatility.

ACKNOWLEDGMENT

This research was supported by the researchers supporting project number [RSPD2024R995], King Saud University, Riyadh, Saudi Arabia.

REFERENCES

[1] Y. Wu, "The Importance of Studying the Unemployment Rate in China," Journal of Education, Humanities and Social Sciences, vol. 24, pp. 465–470, Dec. 2023, doi: 10.54097/9maq2b67.

[2] P. M. Pincheira and A. M. Hernández, "Forecasting Unemployment Rates with International Factors," Available at SSRN 3510597, 2019.

[3] G. C. Rodrigo, "MICRO AND MACRO: THE ECONOMIC DIVIDE," International Monetary Fund. Accessed: Feb. 28, 2024. [Online]. Available: https://www.imf.org/en/Publications/fandd/issues/Series/Back-to-Basics/Micro-and-Macro

[4] A. E. DePrince and P. D. Morris, "The Effects of Education on the Natural Rate of Unemployment," Business Economics, vol. 43, no. 2, pp. 45–54, 2008, doi: 10.2145/20080205.

[5] L. Kilimova and O. Nishnianidze, "Socio-demographic characteristics as indicators of the unemployment rate," Economic annals-XXI, no. 168, pp. 82–85, 2017.

[6] W. Baah-Boateng, "Determinants of unemployment in Ghana," African Development Review, vol. 25, no. 4, pp. 385–399, 2013.

[7] A. Zawojska, "Effect of Macroeconomic Variables on Unemployment Rate in Poland," ECONOMIC SCIENCE FOR RURAL DEVELOPMENT, 2010.

[8] G. Gozgor, "The impact of trade openness on the unemployment rate in G7 countries," J Int Trade Econ Dev, vol. 23, no. 7, pp. 1018–1037, Oct. 2014, doi: 10.1080/09638199.2013.827233.

[9] K. Y. Ng, Z. Zainal, and S. Samsudin, "COMPARATIVE PERFORMANCE OF ARIMA, SARIMA AND GARCH MODELS IN MODELLING AND FORECASTING UNEMPLOYMENT AMONG ASEAN-5 COUNTRIES," International Journal of Business and Society, vol. 24, no. 3, pp. 967–994, 2023.

[10] P. E. Tengaa, Y. M. Maiga, and A. M. Mwasota, "Modeling and forecasting unemployment rate in Tanzania: an ARIMA approach," Journal of accounting, finance and auditing studies, vol. 9, no. 3, pp. 270–288, 2023.

[11] K. Vu, The ARIMA and VARIMA Time Series: Their Modelings, Analyses and Applications. 2007.

[12] B. Freisleben and K. Ripper, "Economic forecasting using neural networks," in IEEE INTERNATIONAL CONFERENCE ON NEURAL NETWORKS PROCEEDINGS, NEW YORK, 1995, pp. 833–838 vol.2. doi: 10.1109/ICNN.1995.487526.

[13] K. Mero, N. Salgado, J. Meza, J. Pacheco-Delgado, and S. Ventura, "Unemployment Rate Prediction Using a Hybrid Model of Recurrent Neural Networks and Genetic Algorithms," Applied Sciences, vol. 14, no. 8, p. 3174, 2024.

[14] T. Chakraborty, A. K. Chakraborty, M. Biswas, S. Banerjee, and S. Bhattacharya, "Unemployment Rate Forecasting: A Hybrid Approach," Comput Econ, vol. 57, no. 1, pp. 183–201, 2021, doi: 10.1007/s10614-020-10040-2.

[15] H. G. Doğrul and U. Soytas, "Relationship between oil prices, interest rate, and unemployment: Evidence from an emerging market," Energy Econ, vol. 32, no. 6, pp. 1523–1528, 2010.

[16] A. Siddiqa, "Determinants of unemployment in selected developing countries: A panel data analysis," Journal of Economic Impact, vol. 3, no. 1, pp. 19–26, 2021.

[17] C. Huang and A. Petukhina, Applied Time Series Analysis and Forecasting with Python. Springer, 2022.

[18] O. J. Blanchard and L. H. Summers, "Hysteresis in unemployment," in Economic Models of Trade Unions, Springer, 1986, pp. 235–242.

[19] M. Stevenson and M. Peat, "Forecasting Australian unemployment rates," Australian Journal of Labour Economics, vol. 4, no. 1, pp. 41–55, 2001.

[20] M. S. Maqbool, T. Mahmood, A. Sattar, and M. N. Bhalli, "Determinants of unemployment: Empirical evidences from Pakistan," Pak Econ Soc Rev, pp. 191–208, 2013.

[21] P. J. Wilson and L. J. Perry, "Forecasting Australian unemployment rates using spectral analysis," Australian Journal of Labour Economics, vol. 7, no. 4, pp. 459–480, 2004.

[22] Y. Kajitani, K. W. Hipel, and A. I. McLeod, "Forecasting nonlinear time series with feed-forward neural networks: a case study of Canadian lynx data," J Forecast, vol. 24, no. 2, pp. 105–117, 2005.

[23] S. Moshiri and L. Brown, "Unemployment variation over the business cycles: a comparison of forecasting models," J Forecast, vol. 23, no. 7, pp. 497–511, 2004.

[24] G. P. Zhang and V. L. Berardi, "Time series forecasting with neural network ensembles: an application for exchange rate prediction," Journal of the operational research society, vol. 52, pp. 652–664, 2001.

[25] D. D. Misil and D. A. Tarepe, "A Time Series Forecasting of the Philippine Unemployment Rate Using Feed-Forward Artificial Neural Network.," Liceo Journal of Higher Education Research, vol. 14, no. 1, pp. 58–81, 2018.

[26] C. H. Aladag, E. Egrioglu, and C. Kadilar, "Forecasting nonlinear time series with a hybrid methodology," Appl Math Lett, vol. 22, no. 9, pp. 1467–1470, 2009, doi: https://doi.org/10.1016/j.aml.2009.02.006.

[27] P.-F. Pai and C.-S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," Omega (Westport), vol. 33, no. 6, pp. 497–505, 2005.

[28] F. Pennoni and B. Bal-Domańska, "NEETs and Youth Unemployment: A Longitudinal Comparison across European Countries," Soc Indic Res, vol. 162, no. 2, pp. 739–761, 2022, doi: 10.1007/s11205-021-02813-5.

[29] The World Bank, "World Bank Open Data: Free and open access to global development data," The World Bank. Accessed: Mar. 05, 2024. [Online]. Available: https://data.worldbank.org

[30] The World Bank, "Economy & Growth," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/economy-and-growth?view=chart

[31] The World Bank, "Education," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/education?view=chart

[32] The World Bank, "Urban Development," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/urban-development?view=chart

[33] The World Bank, "Health," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/health?view=chart

[34] The World Bank, "Climate Change," World Bank Data. Accessed: May 01, 2024. [Online]. Available: https://data.worldbank.org/topic/climate-change?view=chart

[35] M. Joseph, Modern Time Series Forecasting with Python: Explore industry-ready time series forecasting using modern machine learning and deep learning. Packt Publishing Ltd, 2022.

[36] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Comput Sci, vol. 7, p. e623, Jul. 2021, doi: 10.7717/peerj-cs.623.

[37] C. Lee, F. Amini, G. Hu, and L. Halverson, "Machine Learning Prediction of Nitrification From Ammonia- and Nitrite-Oxidizer Community Structure," Front Microbiol, vol. 13, Jul. 2022, doi: 10.3389/fmicb.2022.899565.

[38] K. Dumičić, A. Čeh Časni, and B. Žmuk, "Forecasting unemployment rate in selected European countries using smoothing methods," World Academy of Science, Engineering and Technology: International Journal of Social, Education, Economics and Management Engineering, vol. 9, no. 4, pp. 867–872, 2015.

[39] A. A. Davidescu, S.-A. Apostu, and A. Paul, "Comparative analysis of different univariate forecasting methods in modelling and predicting the romanian unemployment rate for the period 2021–2022," Entropy, vol. 23, no. 3, p. 325, 2021.

[40] M. Yurtsever, "Unemployment rate forecasting: LSTM-GRU hybrid approach," J Labour Mark Res, vol. 57, no. 1, p. 18, 2023.

[41] L. Shi, Y. A. Khan, and M.-W. Tian, "COVID-19 pandemic and unemployment rate prediction for developing countries of Asia: A hybrid approach," PLoS One, vol. 17, no. 12, p. e0275422, 2022.

[42] A. Vosseler and E. Weber, "Forecasting seasonal time series data: a Bayesian model averaging approach," Comput Stat, vol. 33, pp. 1733–1765, 2018.

[43] A. L. Montgomery, V. Zarnowitz, R. S. Tsay, and G. C. Tiao, "Forecasting the U.S. Unemployment Rate," J Am Stat Assoc, vol. 93, no. 442, pp. 478–493, Jun. 1998, doi: 10.1080/01621459.1998.10473696.

[44] C. Katris, "Prediction of unemployment rates with time series and machine learning techniques," Comput Econ, vol. 55, no. 2, pp. 673–706, 2020.

[45] E. Olmedo, "Forecasting spanish unemployment using near neighbour and neural net techniques," Comput Econ, vol. 43, pp. 183–197, 2014.

[46] M. Wozniak, "Forecasting the unemployment rate over districts with the use of distinct methods," vol. 24, no. 2, 2020, doi: doi:10.1515/snde-2016-0115.

Exploring Wealth Dynamics: A Comprehensive Big Data Analysis of Wealth Accumulation Patterns

Karim Mohammed Rezaul[1], Mifta Uddin Khan[2], Nnamdi Williams David[3], Kazy Noor e Alam Siddiquee[4], Tajnuva Jannat[5], Md Shabiul Islam[6]
Wrexham University, UK[1]
Northumbria University London, UK[2, 3]
Multimedia University, Malaysia[4]
Centre for Applied Research in Software & IT (CARSIT), UK[5]
Multimedia University, Malaysia[6]

***Abstract*—The study offers a thorough examination of the accumulation and distribution of wealth among billionaires through the application of big data analytics methodologies. This research centres on an extensive dataset known as "Billionaires.csv," [19] which encompasses a range of information about billionaires from diverse nations, including their demographic characteristics, company particulars, sources of wealth, and more details. The study aims to get a deeper understanding of the determinants that change the net worth of billionaires and detect trends in the worldwide financial system that can guide entrepreneurial ventures and investment possibilities. The dataset is subjected to analysis and visualisation through the utilisation of Python tools and libraries, including but not limited to Pandas, NumPy, Matplotlib, and Seaborn. The results of this study offer valuable insights into the distribution of wealth among billionaires, the factors that contribute to industry success, gender disparities, age demographics, and other factors that influence the accumulation of billionaire wealth.**

Keywords*—Big data; python; billionaires; net worth; wealth accumulation; wealth inheritance; geographic location; statistical analysis

I. Introduction

The rapid growth of wealth among billionaires has garnered significant attention in recent years, with the number of billionaires increasing from 1,001 in 2010 to 2,153 in 2019 [1]. This surge in wealth accumulation has led to a growing interest in understanding the factors that contribute to the success of these individuals. Big Data Analytics, with its ability to process large volumes of data, offers a promising approach to uncovering patterns and insights related to wealth distribution, wealth creation, and accumulation among billionaires [2].

This paper explores the dynamics of wealth accumulation and its distribution among billionaires worldwide using a comprehensive dataset. The study uses advanced big data analytics techniques to show trends, correlations, and patterns that underlie wealth creation and proliferation among the world's richest individuals. The research's significance lies in its contribution to academic literature and its potential to inform policy decisions and economic strategies. By elucidating factors contributing to wealth accumulation, it looks to provide insights into economic and social policies that could promote a fairer distribution of wealth and opportunities. The methodological approach combines quantitative analysis with sophisticated data visualization techniques to offer a panoramic view of wealth dynamics. The findings are expected to shed light on wealth accumulation trajectories and contribute to a deeper understanding of the economic forces shaping our world.

The objective of this investigation is to conduct an analysis of the "Billionaires.csv" dataset utilising Python tools for Big Data Analysis, with a particular emphasis on regions and cities. This paper aims to investigate several research questions about the distribution of billionaires across the globe. The study's objectives include identifying the ten nations with the highest concentration of billionaires, the industries with the greatest growth, the nations with the largest percentage of female billionaires, the age groups where there are the most and least billionaires belong, as well as other elements that may affect the development of wealth, for example, inheritance.

The paper is structured into four tasks: Problem Domain, Data Description, and Research Question; Solution Exploration; Solution Development; and Evaluation and Future Development. Each task delves into different aspects of the research process, from formulating hypotheses and evaluating methodologies to analyzing the dataset and discussing the potential impact of the results. This research endeavors to enhance comprehension of wealth accumulation trends among billionaires and offer valuable insights for policymakers, investors, and entrepreneurs by utilizing Big Data Analytics and the Python programming language.

II. Literature Review

Between 1999 and 2019, the Hurun Rich List was diligently compiled, and research was done to confirm billionaires and their families. They compare the Hurun Rich List to each company's annual report to determine which billionaires own it [3]. The systematic billionaire wealth factors were not considered in the univariate tests. Age was disregarded, and economic activity might potentially be important [4]. A study constructed a comprehensive metric of wealth inequality on a global scale, utilizing Forbes magazine's roster of billionaires, and subsequently conducted a comparative analysis of its impact on economic growth to the effects of income inequality and poverty [5]. An additional research study demonstrated that industries that concentrate wealth do so with purpose, and this

*Corresponding Author

may occur due to a reliance on the state, the presence of market failures, or the inheritance of extreme wealth [6]. A recent study conducted during the pandemic in Canada has demonstrated that low-wage workers have been disproportionately affected. Additionally, data from Statistics Canada has indicated that women and racialized workers are overrepresented within the low-wage worker demographic [7]. An article aims to compare asset ownership in three common marital regimes in France and other European countries: Property regimes, marriage, non-marital cohabitation, and registered partnership (PACS) [8]. The results show that luck plays a big part at the very ends of the range of economic outcomes, and the fact that empirical regularities tend to disappear in the far tails can be used to look at any group of successful or failed people [9]. A study finding indicates that there is a detrimental association between wealth inequality and economic growth., however, when considering the influence of political connections on the acquisition of wealth by billionaires, it is observed that politically connected wealth inequality has a negative impact on economic growth. On the other hand, politically unconnected wealth inequality, income inequality, and initial poverty do not exhibit any significant relationship with economic growth [5]. An empirical approach to studying the potential influence of education and cognitive ability on the accumulation of extreme wealth involves analyzing high-wealth groups and retrospectively evaluating the level of education and cognitive ability of these individuals in the past [10]. When comparing billionaires, it is evident that wealth is not correlated with attractiveness. Neither does it pertain to the level of schooling achieved [14]. The concentration of wealth and its distribution has been a longstanding concern in economic literature and societal discourse. Studies on wealth inequality and its consequences have highlighted the need for a deeper understanding of the factors contributing to the accumulation of wealth among the elite class. In this section, we provide a brief overview of relevant literature that explores the determinants of billionaire net worth and sheds light on the dynamics of wealth accumulation.

A study looked at a continuous-time DSGE model with several types of households and a financial sector, and they were able to make the case of studying how financial frictions affect families by looking at how wealth is distributed [15]. An agent-based model of microscopic wealth sharing in a dynamic network is used in another study to investigate the topological aspects of economic inequality [16]. The model changes in two steps that happen back and forth: the connected agents conservatively trade wealth, and the links are rewired based on the agents' wealth. Moreover, an analysis finds financialization, rentiers, and labour exploitation as the primary drivers of billionaire wealth in the U.S. sectors with the highest number of billionaires [17]. These factors were crucial for their dominance, while shareholder culture, crony capitalism, and tax policy perpetuated wealth but were not necessary for its current level.

Gender disparities in wealth distribution have also garnered significant attention. Numerous studies have documented the persistent gender wealth gap, demonstrating that women tend to have lower levels of wealth compared to men. However, the examination of gender disparities among billionaires remains relatively limited. To clarify the gender dynamics within this wealthy class and put light on the larger issue of disparities between genders in the accumulation of wealth, it is imperative to know whether there is an important variation in assets between male and female billionaires. An article looks at how the trends in inequality found by multiple well-known inequality indices in the Forbes 400 richest families and compared with each other [18]. Other researchers have looked at other parts of the data set. A paper makes the first Distributional Wealth Accounts (DWAs) for Europe from 1970 to 2018 by putting together state accounts, tax records, and surveys. According to our new database, the amount of wealth compared to national income has changed about the same in both the US and Europe. However, since the mid-1980s, wealth inequality has grown much faster in the US than in Europe.

Finally, the impact of wealth inheritance on billionaire net worth has attracted scholarly attention. Inherited wealth often carries advantages in terms of family legacies, established networks, and access to resources. However, the extent to which inherited wealth plays a role in billionaire net worth compared to self-made wealth is a topic of ongoing debate. Investigating the difference in net worth between billionaires who inherited their wealth and those who built their fortunes from scratch adds to the understanding of intergenerational wealth transfer and its implications for wealth inequality. By reviewing the existing literature on age and wealth, gender disparities, industry sectors, and wealth inheritance, we situate our study within the broader research landscape. This background provides a foundation for our analysis of the billionaire dataset and highlights the gaps and opportunities for further exploration. The findings from this study contribute to the existing knowledge base, inform policy discussions, and offer insights into the complex dynamics of wealth accumulation among billionaires.

III. Problem Domain

The open-source dataset utilized in this paper is the billionaire's dataset from Kaggle. Understanding wealth distribution and wealth trends may be done extremely well by using this dataset. This report looks at the "Billionaires.csv" dataset analysis using Python tools designed for large-scale data analysis. The database includes data about billionaires from different nations, such as their ages, genders, industrial sectors, net worths, and places of wealth. Understanding the variables that affect the creation and distribution of billionaires' wealth is the aim of the investigation. To better comprehend these patterns and discover potential prospects for company starts and investments, it may be helpful to examine current trends in the global economy.

IV. Investigation

The paper aims to investigate the following measures:

1) To examine the relationship between demographic factors (age, gender) and net worth among billionaires.

2) To explore the variations in net worth across different industry sectors among billionaires.

3) To investigate the differences in net worth between self-made billionaires and those who inherited their wealth.

4) To analyze the correlation between age and net worth among billionaires and determine the significance of the relationship.

5) To compare the net worth of billionaires across different industry sectors and identify sectors with the highest net worth individuals.

6) To evaluate the differences in net worth between self-made billionaires and those who inherited their wealth using statistical tests.

7) To provide insights into the patterns and trends in the distribution of wealth among billionaires based on demographic factors and industry sectors.

V. Research Methodology

- Data Collection: The open-source dataset obtained from Kaggle was used in this study.
- Data Preprocessing: To deal with any missing numbers, outliers, and other mistakes, the data will be pre-processed.
- Data visualization: Data visualization is the process of using visual elements like charts, graphs, or maps to represent data.
- Hypothesis Testing: Formulate specific hypotheses based on research questions and objectives. Selection of appropriate statistical tests to test the hypotheses.
- Interpretation of Results: The results will be interpreted to identify key findings.

VI. Data Description

The dataset helps research variables related to wealth accumulation and distribution among billionaires. The net worth may be impacted by age and industrial sectors. A study of inheritance and the gender difference in wealth accumulation might potentially be done using the dataset. There are 22 variables and 2614 rows in the dataset. Table I provides a brief description of each variable.

TABLE I. Brief Data Interpretation

Feature	Description
name	Billionaires name.
rank	Annual list of billionaires ranked by net worth.
year	Year of data collection.
company. founded	The year that the business was formed
company.name	The billionaire's company's name
company. relationship	The connection between the business and the billionaire.
company. sector	Area of billionaires' businesses.
company. type	What kind of business the billionaire owns—public, state-owned, public etc.
demographics.age	Individual’s age.
demographics. relationship	Billionaire's gender.
location. citizenship	Nation in which the billionaire was born.
location. country code	The Country's ISO code where the billionaire lives.
location.gdp	Country's GDP where the billionaire lives.
location. region	Location of the billionaire in the globe.
wealth. type	Wealth's source (Inherited. Self-made, etc.)
wealth. worth in billions	The billionaire's total wealth, expressed in billions of dollars.
wealth.how. category	Method by which each billionaire acquired their wealth
wealth.how. from emerging	If the billionaire's prosperity was derived from a developing market.
wealth.how. industry	The branch of industry where the billionaire made their money.
wealth.how. inherited	The billionaire's wealth was acquired through inheritance.
wealth.was. founder	If the millionaire was a company founder.
wealth.how.was political	Billionaire has political experience or not.

VII. Research Questions

1) Which ten international countries have the maximum number of billionaires?

2) Which industries and sectors are the most successful?

3) Which sectors have the highest awareness of female billionaires?

4) What age institution do most people and a minority of billionaires belong to?

VIII. Hypothesis

A crucial component of inferential statistics is hypothesis testing, which enables us to conclude unobserved data, frequently the population, using data from a sample of observed data. When analyzing experimental data in economics, testing multiple null hypotheses simultaneously is a common practice [11]. To research this hypothesis, numerous statistical strategies will be used, inclusive of t-tests, ANOVA, and F-assessments. The outcomes of those experiments could shed mild on the variables concerned with wealth accumulation and guide the improvement of policies geared toward reducing wealth inequality. In our research, we will look at the principle of different factors that can be intently related to a billionaire's wealth. We REJECT the null hypothesis in favor of the alternative if the P-value is less than the threshold for significance (= 0.05). The correlation is statistically significant, we conclude. Or, to put it another way, we come to the simple conclusion that x and y in the population at the -level are linearly related.

If $p > 0.05$, then not correlated.

If $p < 0.05$, correlated.

The following hypotheses have been formulated for our study.

1) Does a billionaire's age affect his or her wealth?

Ho: A billionaire's net worth is not significantly related to age.

HA: A billionaire's net worth is highly related to age in a significant way.

2) Men with billion-dollar wealth are wealthier than women.

Ho: A billionaire's net wealth is unaffected by gender in a significant way.

H_A: Gender significantly affects a billionaire's financial worth.

3) The net worth of billionaires varies considerably between various industry sectors.

Ho: A billionaire's net worth is significantly influenced by the industry sector.

HA: A billionaire's net worth is not much impacted by their industry sector.

IX. Exploration of Solutions

Big data is the term used to describe the vast and diversified informational resources that are accumulating quickly. Software that can be used to handle data effectively and perform pre-processing and cleaning etc., is known as big data technology. The five main qualities of big data are truth, diversity, value, and velocity. In seconds, minutes, hours, or days, data volume is measured [12]. Velocity describes the speed at which new data is produced and transferred, whereas variety describes the wide range of data forms that are produced often. It's vital to note that velocity also shows how rapidly a company reacts to big data-derived business insight. The chance that the data's quality and consistency are inadequate, that is, less uniform, consistent, and controllable, is referred to as veracity. The fifth V, "value," stands for the ability of a person or an organization to translate enormous amounts of data into beneficial outcomes, which includes the capability to acquire and then use data [13].

To address the issues of huge information programs, numerous methodologies, and technologies have evolved these days. Those are Apache Hadoop, MongoDB, Apache Kafka, Elastic Search, Python, Seaborn, and Plotly.

X. The Tools or Technologies used in this Research

The Python programming language and modern libraries such as Pandas, NumPy, Matplotlib, and Seaborn have been selected as the technologies that will be used for this project. This decision was made about the technology that would be implemented. It is beneficial to use open-source software solutions since they are free of charge, have an interface that is simple to use, and can give comprehensive analysis and visualization capabilities. Other technologies, on the other hand, can need payment or prior expertise. When it comes to the Python Seaborn visualization toolkit, the matplotlib module is an absolute must. A graph interface that is both high-level and insightful of data will be utilized. Python packages such as Pandas, NumPy, and SciPy can be utilized to edit and analyze data. The Pandas data frame makes the process of processing data simpler. NumPy can deal with low-level data, whereas SciPy is responsible for statistical analysis.

XI. Solution Development

A key part of current data-driven decision-making is coming up with practical solutions to problems based on data analysis. Using the "Billionaires" dataset as an example, this paper talks about the main steps and things to think about when coming up with good solutions for the research questions we have. We are going to initially preprocess our data to check whether there is any noise, further, we will analyze our dataset to answer those research questions we formulate in the previous part. In the meantime, we will test the hypothesis for each question. Finally, we will show our findings and future scopes.

A. Data Preprocessing

It is necessary to preprocess the dataset to prepare it to provide true insights into the data to answer the research questions. To preprocess data, numerous processes are taken. We are going to take advantage of the procedures that are necessary in order to finish our assignment.

B. Visual Exploratory Data Analysis

Fig. 1 determines statistics at the variables' names, counts, and their respective information kinds. There are a complete of thirteen objects, along with 4 integers, 2 floats, and 3 Booleans.

Fig. 2 presents comprehensive statistical information for each numerical column, encompassing the count of values in a row, the mean, standard deviation, minimum and maximum values, and the quartiles at 25%, 50%, and 75% of the dataset. In terms of the dataset, the period under examination spans from 1996 to 2014. When compared to the oldest company, which was established in 1610, the most recent company was

established in 2012. The individual who is the youngest among them is only 12 years old, while the individual who is the oldest is 98 years old.Beyond $3.5 billion, the median net worth of a billionaire is greater than that amount.

Fig. 3 illustrates the number of unique values in each variable. Here, we are going to present a visual exploratory information analysis. Commonly, specific variables incorporate less than twenty precise values and there may be a repetition of values, this means that the records can be grouped by manner of these precise values. At this level of the study process, we're investigating the effects of visible evaluation on a spread of specific variables derived from the dataset.

Fig. 4 demonstrates the dataset comprising individuals belonging to three distinct genders: males, females, and married couples. Moreover, there exist five discrete categories of billionaires' wealth.

The four graphs in Fig. 5 and Fig. 6 depict a visual analysis of six distinct variables within the dataset. Fig. 5 illustrates the distribution of wealth and the corresponding industries associated with wealth. Fig. 6 illustrates six distinct instances of form inheritance, along with a classification indicating whether each billionaire is a founder or not.

```
[1097] billionaire_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2614 entries, 0 to 2613
Data columns (total 22 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   name                       2614 non-null   object
 1   rank                       2614 non-null   int64
 2   year                       2614 non-null   int64
 3   company.founded            2614 non-null   int64
 4   company.name               2576 non-null   object
 5   company.relationship       2568 non-null   object
 6   company.sector             2591 non-null   object
 7   company.type               2578 non-null   object
 8   demographics.age           2614 non-null   int64
 9   demographics.gender        2580 non-null   object
 10  location.citizenship       2614 non-null   object
 11  location.country code      2614 non-null   object
 12  location.gdp               2614 non-null   float64
 13  location.region            2614 non-null   object
 14  wealth.type                2592 non-null   object
 15  wealth.worth in billions   2614 non-null   float64
 16  wealth.how.category        2613 non-null   object
 17  wealth.how.from emerging   2614 non-null   bool
 18  wealth.how.industry        2613 non-null   object
 19  wealth.how.inherited       2614 non-null   object
 20  wealth.how.was founder     2614 non-null   bool
 21  wealth.how.was political   2614 non-null   bool
dtypes: bool(3), float64(2), int64(4), object(13)
memory usage: 395.8+ KB
```

Fig. 1. Dataset information.

```
[1098] billionaire_df.describe()
```

	rank	year	company.founded	demographics.age	location.gdp	wealth.worth in billions
count	2614.000000	2614.000000	2614.000000	2614.000000	2.614000e+03	2614.000000
mean	599.672533	2008.411630	1924.711936	53.341239	1.769103e+12	3.531943
std	467.885695	7.483598	243.776546	25.333320	3.547083e+12	5.088813
min	1.000000	1996.000000	0.000000	-42.000000	0.000000e+00	1.000000
25%	215.000000	2001.000000	1936.000000	47.000000	0.000000e+00	1.400000
50%	430.000000	2014.000000	1963.000000	59.000000	0.000000e+00	2.000000
75%	988.000000	2014.000000	1985.000000	70.000000	7.250000e+11	3.500000
max	1565.000000	2014.000000	2012.000000	98.000000	1.060000e+13	76.000000

Fig. 2. Statistical view of numerical data.

```
# Finding unique values for each column
# TO understand which column is categorical and which one is Continuous
# Typically if the numer of unique values are < 20 then the variable is likely to be a category otherwise continuous
billionaire_df.nunique()

name                       2077
rank                        468
year                          3
company.founded             178
company.name               1577
company.relationship         74
company.sector              520
company.type                 18
demographics.age             76
demographics.gender           3
location.citizenship         73
location.country code        74
location.gdp                 81
location.region               8
wealth.type                   5
wealth.worth in billions    183
wealth.how.category           9
wealth.how.from emerging      1
wealth.how.industry          19
wealth.how.inherited          6
wealth.how.was founder        1
wealth.how.was political      1
dtype: int64
```

Fig. 3. Checking unique values.

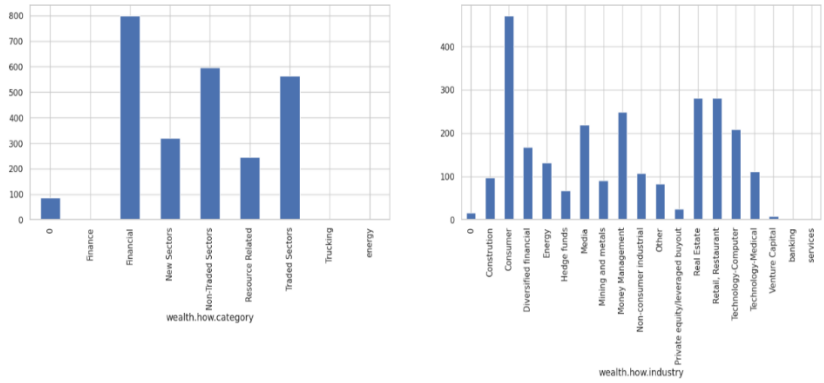


Fig. 4. Gender and wealth distribution visually.

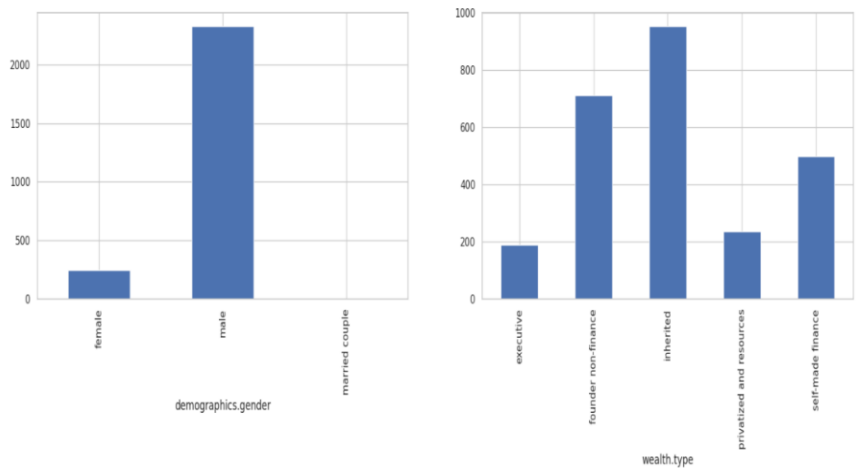


Fig. 5. Distribution of 'wealth category and 'wealth industry' by bar graph.

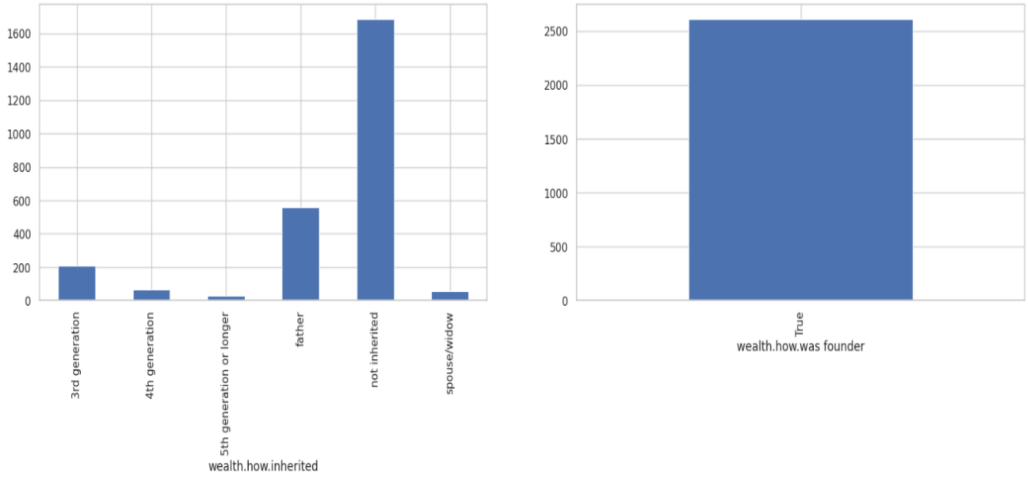


Fig. 6. Distribution of 'wealth inherited' and 'wealth.how.was founder.

The dataset contains 22 variables, but only a few are essential for answering research questions. To maintain the desired number of variables, only a few are required, resulting in the loss of unnecessary ones. The remaining variables will be maintained for research purposes.

Fig. 7 displays a dataset after irrelevant columns have been removed, presenting a streamlined view of wealthy individuals' demographics and financial information.

	name	demographics.gender	location.citizenship	company.sector	demographics.age	wealth.worth in billions	wealth.type	wealth.how.category	wealth.how.inherited	wealth.how.was founder
0	Bill Gates	male	United States	Software	40	18.5	founder non-finance	New Sectors	not inherited	True
1	Bill Gates	male	United States	Software	45	58.7	founder non-finance	New Sectors	not inherited	True
2	Bill Gates	male	United States	Software	58	76.0	founder non-finance	New Sectors	not inherited	True
3	Warren Buffett	male	United States	Finance	65	15.0	founder non-finance	Traded Sectors	not inherited	True
4	Warren Buffett	male	United States	Finance	70	32.3	founder non-finance	Traded Sectors	not inherited	True
...	...	...	...	...	...	...	...	...	...	...
2609	Wu Chung-Yi	male	Taiwan	beverages and food	55	1.0	executive	Traded Sectors	not inherited	True
2610	Wu Xiong	male	China	infant formula	0	1.0	executive	Traded Sectors	not inherited	True
2611	Yang Keng	male	China	real estate	53	1.0	self-made finance	Financial	not inherited	True
2612	Zdenek Bakala	male	Czech Republic	coal	53	1.0	privatized and resources	Resource Related	not inherited	True
2613	Zhu Wenchen	male	China	pharmaceuticals	48	1.0	executive	New Sectors	not inherited	True

2614 rows × 10 columns

Fig. 7. After removing irrelevant columns - new dataset.

By doing so, we will rename the column name as follows in order to make the analysis simpler:

Fig. 8 represents the dataset with renamed columns to provide a clearer understanding of the attributes of each individual listed.

C. Handling Missing Values

At this point, we determine which values in the dataset are absent. One of the most critical components of records evaluation is the management of lacking values. The precise study topic, the number of missing data, and the influence that the missing values can have on the evaluation are all elements that must be taken into consideration whilst determining whether to put off missing values or impute them.

As can be seen in Fig. 9, 34 cells are missing from the gender column, 23 cells in the company_sector column, 22 cells in the wealth_type column, and one cell in the wealth_source column.

In Fig. 10, We check the Total missing values in our dataset and the proportion of missing on different columns. The percentage of values that are missing from the dataset is quite low (most of them are much lower than 0%). Therefore, we are going to ascribe those values to the requirements that we have.

	name	gender	citizenship	company_sector	age	net_worth_billion	wealth_type	wealth_source	wealth_inherited	was_founder
0	Bill Gates	male	United States	Software	40	18.5	founder non-finance	New Sectors	not inherited	True
1	Bill Gates	male	United States	Software	45	58.7	founder non-finance	New Sectors	not inherited	True
2	Bill Gates	male	United States	Software	58	76.0	founder non-finance	New Sectors	not inherited	True
3	Warren Buffett	male	United States	Finance	65	15.0	founder non-finance	Traded Sectors	not inherited	True
4	Warren Buffett	male	United States	Finance	70	32.3	founder non-finance	Traded Sectors	not inherited	True
...	...	...	...	...	...	...	...	...	...	...
2609	Wu Chung-Yi	male	Taiwan	beverages and food	55	1.0	executive	Traded Sectors	not inherited	True
2610	Wu Xiong	male	China	infant formula	0	1.0	executive	Traded Sectors	not inherited	True
2611	Yang Keng	male	China	real estate	53	1.0	self-made finance	Financial	not inherited	True
2612	Zdenek Bakala	male	Czech Republic	coal	53	1.0	privatized and resources	Resource Related	not inherited	True
2613	Zhu Wenchen	male	China	pharmaceuticals	48	1.0	executive	New Sectors	not inherited	True

2614 rows × 10 columns

Fig. 8. Renaming columns.

```
#1.Checking missing values dataset
print(billionaire_df.isnull().sum())
```

```
name                0
gender             34
citizenship         0
company_sector     23
age                 0
net_worth_billion   0
wealth_type        22
wealth_source       1
wealth_inherited    0
was_founder         0
dtype: int64
```

Fig. 9. Number of missing values on each column.

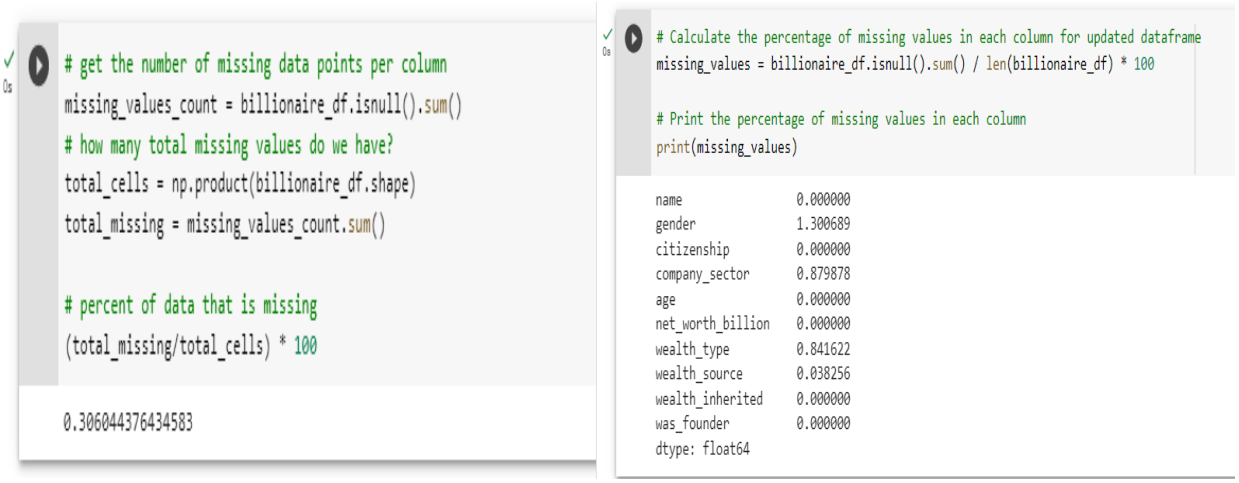


Fig. 10. Missing values in percentage.

As a first step, we strive to impute the cells that might be lacking from the gender column. Inside the gender column, there are a complete of 34 values that are lacking, and most people of these billionaires have covered terms related to their families alongside their names. Due to the absence of gender values, data analysis is impeded. The presence of the term "family" in the name gives the impression that the individual in question is not a billionaire but rather a member of a family that is a billionaire, which has the potential to skew the results of the study. Consequently, removing these rows guarantees that the analysis is founded on data that is both reliable and pertinent.

Fig. 11 demonstrates the word "family" is present in most of the cells that are absent from the gender column. We are going to get rid of those rows. Since Oeri Hoffman and Sacher appear to be a "married couple" right before our eyes, we shall impute one missing cell as belonging to a "married couple.". Rest of the cells we can delete the rows as there are a smaller number of missing cells.

Fig. 11 demonstrates the word "family" is present in most of the cells that are absent from the gender column. We are going to get rid of those rows. Since Oeri Hoffman and Sacher appear to be a "married couple" right before our eyes, we shall impute one missing cell as belonging to a "married couple.". Rest of the cells we can delete the rows as there are a smaller number of missing cells.

In Fig. 12, following the elimination of cells that were absent from the gender subgroup, we are left with 2581 individual observations. The next step is for us to analyze our research using this dataset. We are still trying to impute a few cells that are missing from our database.

Fig. 13 shows above are the remainder of the values that are absent from our dataset. We are going to try to impute those using statistical methods such as the mean, the mode, and the median. As a method for dealing with missing values, we use the median for continuous variables and mode for categorical measures.

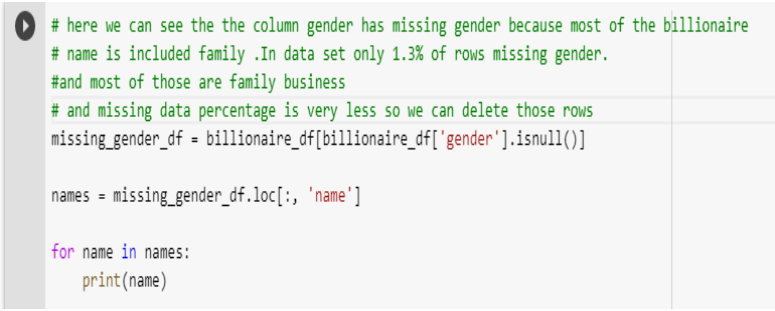


Fig. 11. Checking missing values in gender with their names.

```
[340] #It is not possible to determine the gender of individuals based solely on their name or family name.
      billionaire_df = billionaire_df.dropna(subset=['gender'])
      billionaire_df
```

	name	gender	citizenship	company_sector	age	net_worth_billion	wealth_type	wealth_source	wealth_inherited	was_founder
0	Bill Gates	male	United States	Software	40	18.5	founder non-finance	New Sectors	not inherited	True
1	Bill Gates	male	United States	Software	45	58.7	founder non-finance	New Sectors	not inherited	True
2	Bill Gates	male	United States	Software	58	76.0	founder non-finance	New Sectors	not inherited	True
3	Warren Buffett	male	United States	Finance	65	15.0	founder non-finance	Traded Sectors	not inherited	True
4	Warren Buffett	male	United States	Finance	70	32.3	founder non-finance	Traded Sectors	not inherited	True
...	...	...	...	...	...	...	...	...	...	...
2609	Wu Chung-Yi	male	Taiwan	beverages and food	55	1.0	executive	Traded Sectors	not inherited	True
2610	Wu Xiong	male	China	infant formula	0	1.0	executive	Traded Sectors	not inherited	True
2611	Yang Keng	male	China	real estate	53	1.0	self-made finance	Financial	not inherited	True
2612	Zdenek Bakala	male	Czech Republic	coal	53	1.0	privatized and resources	Resource Related	not inherited	True
2613	Zhu Wenchen	male	China	pharmaceuticals	48	1.0	executive	New Sectors	not inherited	True

2581 rows × 10 columns

Fig. 12. Dropping irrelevant cells from gender subset.

```
[341] #1.2.Checking Missing values for updated Dataframe
      print(billionaire_df.isnull().sum())
```

```
name                0
gender              0
citizenship         0
company_sector     11
age                 0
net_worth_billion   0
wealth_type        10
wealth_source       1
wealth_inherited    0
was_founder         0
dtype: int64
```

Fig. 13. Checking total missing values remained in the updated data frame.

Fig. 14 illustrates the code snippet of a Python script using pandas to impute missing values in the dataframe.

After imputing the remainder variables, we got the updated dataset with ‘zero’ missing values (Fig. 15).

D. Checking Duplicate

As part of the data cleansing process, we conducted a check to identify any duplicated rows (as in Fig. 16). The duplicated () method in pandas allows us to identify and eliminate any duplicated rows in the dataset. This measure was implemented to ensure the distinctiveness of each discovery and to prevent any potential interference with subsequent analyses. Therefore, we did not discover any duplicate rows as a result.

```
[342] #I am treating missing values with Median for Continuous values, and Mode for categorical values.
      # Treating missing values of categorical variable with MODE value
      billionaire_df['company_sector'].fillna(value=billionaire_df['company_sector'].mode()[0], inplace=True)
      billionaire_df['wealth_type'].fillna(value=billionaire_df['wealth_type'].mode()[0], inplace=True)
      billionaire_df['wealth_source'].fillna(value=billionaire_df['wealth_source'].mode()[0], inplace=True)
      billionaire_df
```

```
<ipython-input-342-d863120c5587>:3: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

<ipython-input-342-d863120c5587>:4: SettingWithCopyWarning:
```

Fig. 14. Missing values imputation by statistical techniques.

```
[343] #1.3.Checking Missing values for updated Dataframe
      print(billionaire_df.isnull().sum())

name                0
gender              0
citizenship         0
company_sector      0
age                 0
net_worth_billion   0
wealth_type         0
wealth_source       0
wealth_inherited    0
was_founder         0
dtype: int64
```

Fig. 15. Data frame after imputation.

```
[344] #4.Checking for duplicated rows in a dataset after dropping the rows with missing values
      # Check for duplicated rows
      duplicated_rows = billionaire_df[billionaire_df.duplicated()]

      # Print the duplicated rows
      print(duplicated_rows)

Empty DataFrame
Columns: [name, gender, citizenship, company_sector, age, net_worth_billion, wealth_type, wealth_source, wealth_inherited, was_founder]
Index: []
```

Fig. 16. Checking duplicate.

E. Outlier Analysis

To guarantee the precision and dependability of the studies, we did an outlier treatment of the dataset by using the interquartile range (IQR) approach in Python. This allowed us to recognize and exclude any intense values that would have a sizeable impact on the effects.

Fig. 17 and Fig. 18 show that the 'age' column contains significant values that need to be addressed. The dataset includes both horrible and zero age graphs, and outliers are identified at significantly lower than 20 ages. It is important to note that billionaires' age cannot be zero or impoverished.

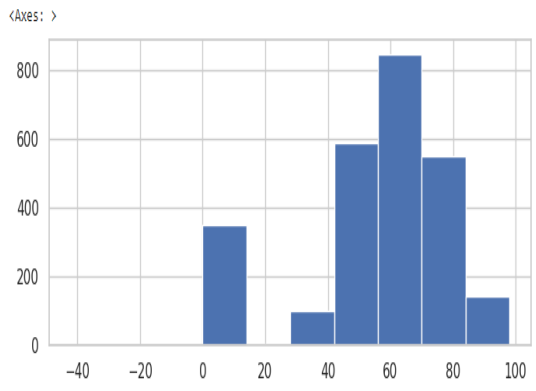


Fig. 17. Visual distribution of 'Age' by histogram.

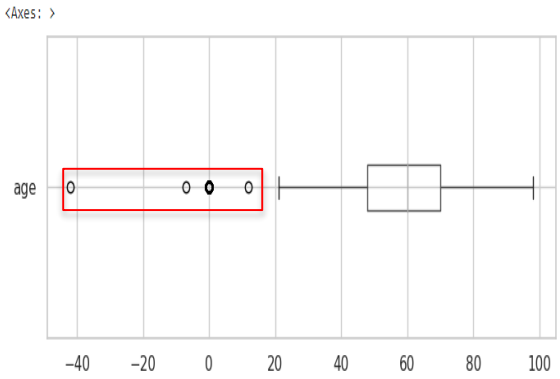


Fig. 18. 'Age' column where outliers detected.

Among the options available for dealing with the outliers, one option is to substitute the median age charge of the relaxation of the dataset for the values of bad and 0 years old. The outliers are going to be dealt with by imputing them using the median statistical technique. This will be done to address the issue.

It is through this method that the median age of the information is determined, which consequently alters the age of billionaires who are significantly younger than twenty years old (Fig. 19).

```
[351] # calculate median age
      median_age = np.median(billionaire_df['age'])

      # replace outliers with median age
      billionaire_df.loc[billionaire_df['age'] < 20, 'age'] = median_age
```

Fig. 19. Treating outliers in 'Age' column.

The net worth column that is displayed in Fig. 20 does not demonstrate any discernible increasing or declining trend. In addition to this, it demonstrates that the 'net_worth' column does not contain any outliers. Further evidence that there is a substantial association between 'age' and 'net worth' is shown by the graph.

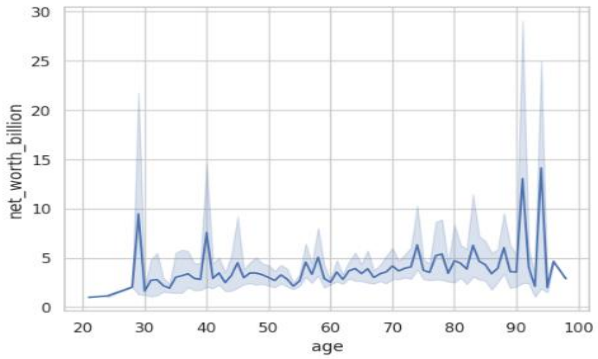


Fig. 20. Line plot of age vs. net worth billion.

XII. Dataset Exploration

The top ten countries with the highest number of billionaires are depicted in Fig. 21.

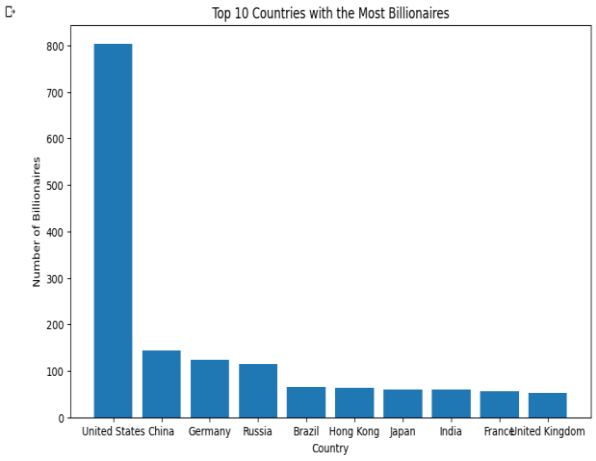


Fig. 21. Top ten billionaires in a bar graph.

Perspectives from Fig. 21:

- Fig. 21 above displays a bar chart of the top 10 richest nations and their billionaires.
- The most billionaires reside in the United States.
- China ranks second among countries with 100 billionaires.
- The number of billionaires in the United Kingdom, which is ranked 10th, is just under fifty.

Fig. 22 depicts the top five industries with the largest number of billionaires.

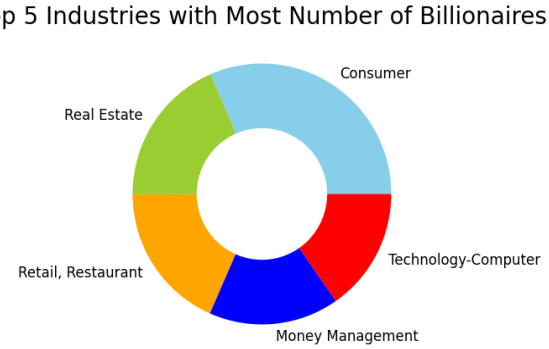


Fig. 22. Top five industries with the most billionaires are shown in a pie chart.

Perspectives from Fig. 22:

- The consumer goods industry has the most billionaires, as shown in the pie chart above.
- A significant number of billionaires are also involved in real estate.

Fig. 23 depicts the number of women billionaires by sector.

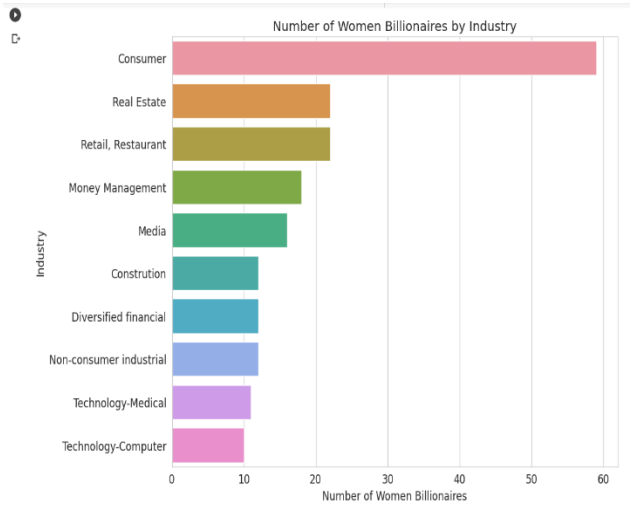


Fig. 23. Graph depicting the number of female billionaires by sectors.

Perspectives from Fig. 23:

- With over 55 female billionaires, the consumer enterprise has the most female billionaires.
- The retail and real estate industries are 2nd, with little more than 20 lady billionaires.

- With only 10 women billionaires working in the IT area, it has the lowest number of female billionaires.

Fig. 24 represents the billionaires by age range.

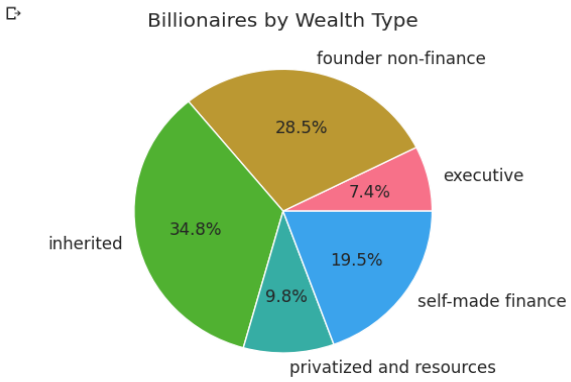


Fig. 24. Proportion of billionaires in each age group shown in the bar chart.

Perspectives from Fig. 24:

- Many of the billionaires in the dataset are between the ages of 60 and 70.
- There are very few billionaires aged 30 and under.
- Approximately two-thirds of billionaires are between the ages of 40 and 80.

Fig. 25 illustrates the proportion of billionaires by wealth category.

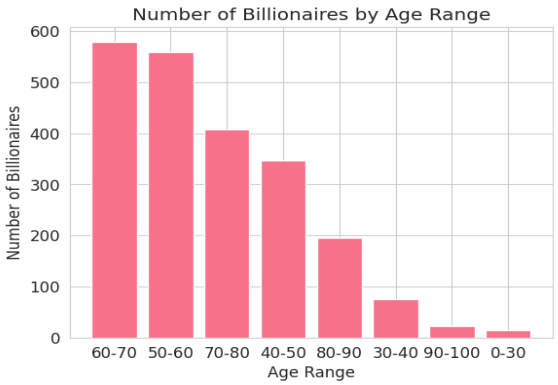


Fig. 25. A pie chart displaying the proportion of billionaires in each wealth category.

Perspectives from Fig. 25:

- The vast majority of the billionaires in the sample inherited their fortunes. They account for 34.8% of all billionaires.
- Self-made billionaires comprise 19.5% of the total.
- Executives account for the smallest proportion of billionaires, accounting for only 7.4% of all billionaires.

Fig. 26 depicts the 2014 net worth of the top 10 billionaires.

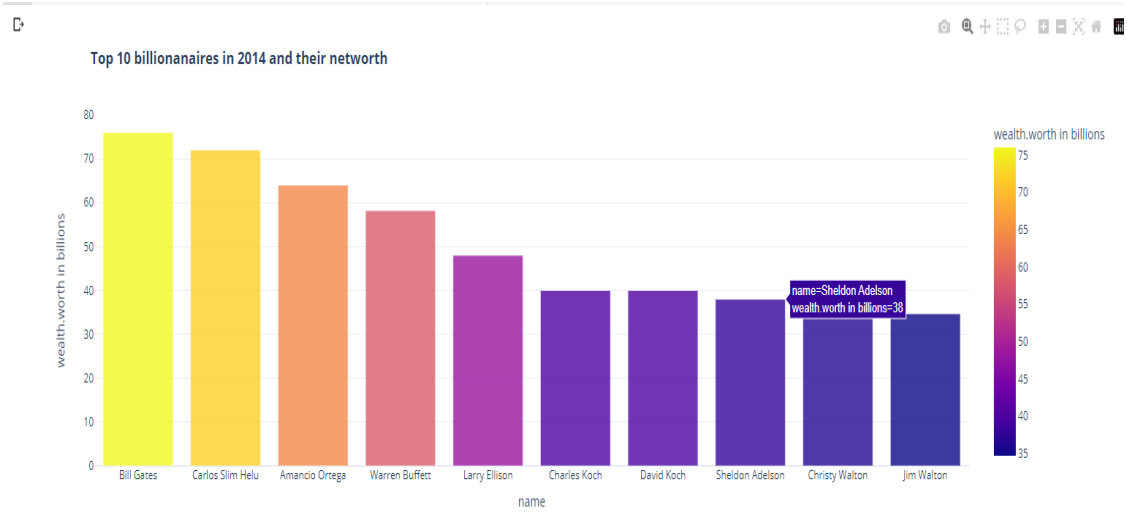


Fig. 26. A bar chart depicting the top ten billionaires and their net wealth in 2014.

Perspectives from Fig. 26:

- The bar graph above shows the most current top 10 billionaires alongside their net worth since 2014 is the latest year that is included in the dataset.
- With an overall net worth of $76 billion, Bill Gates is the first.
- After that, Carlos Slim comes in second place, boasting a net worth of seventy-two billion dollars.
- Christy Walton is the simplest female among the Top Ten rich individuals, with a net worth of thirty-eight billion greenbacks.

The top 10 billionaires' sources of income are shown in Fig. 27.

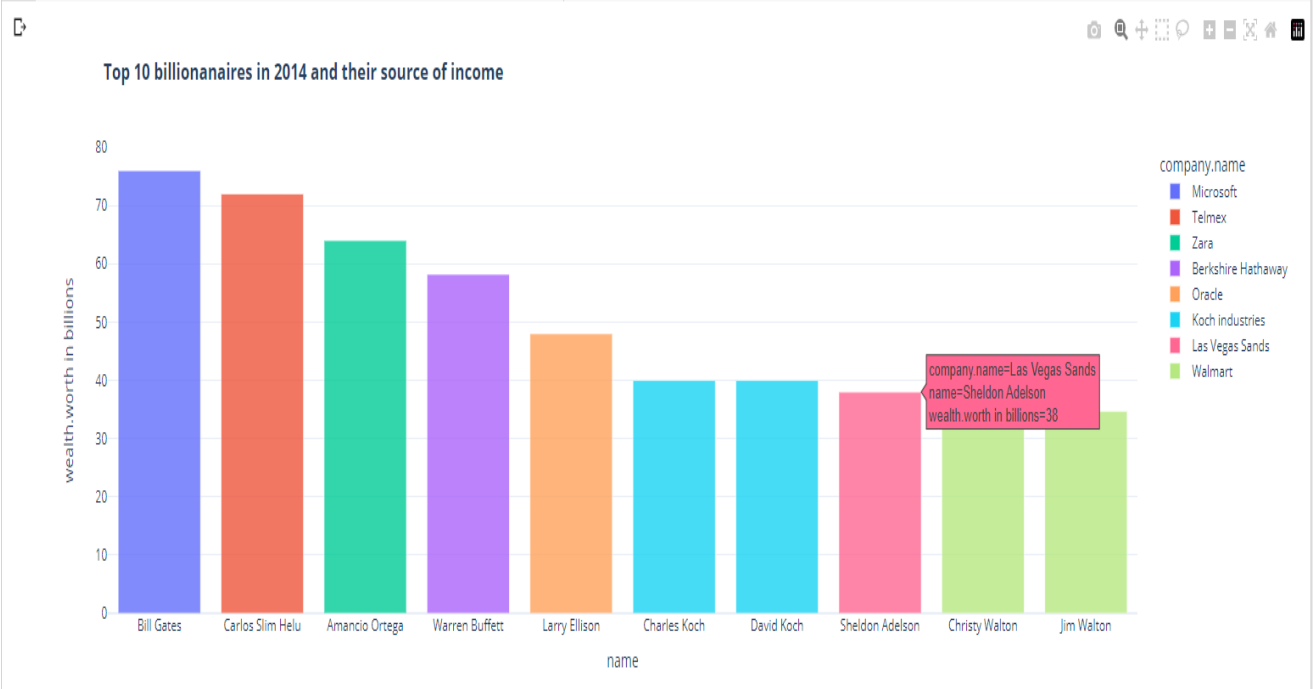


Fig. 27. The top 10 billionaires and their sources of income are shown in a bar graph.

Perspectives from Fig. 27:

- Bill Gates, the richest billionaire, receives funding from Microsoft.
- We can see from the picture above that some colors are repeated twice; these are family businesses or businesses that have produced more than one billionaire, like Walmart and Koch Industries.

The ages of the top 10 billionaires are shown in Fig. 28.

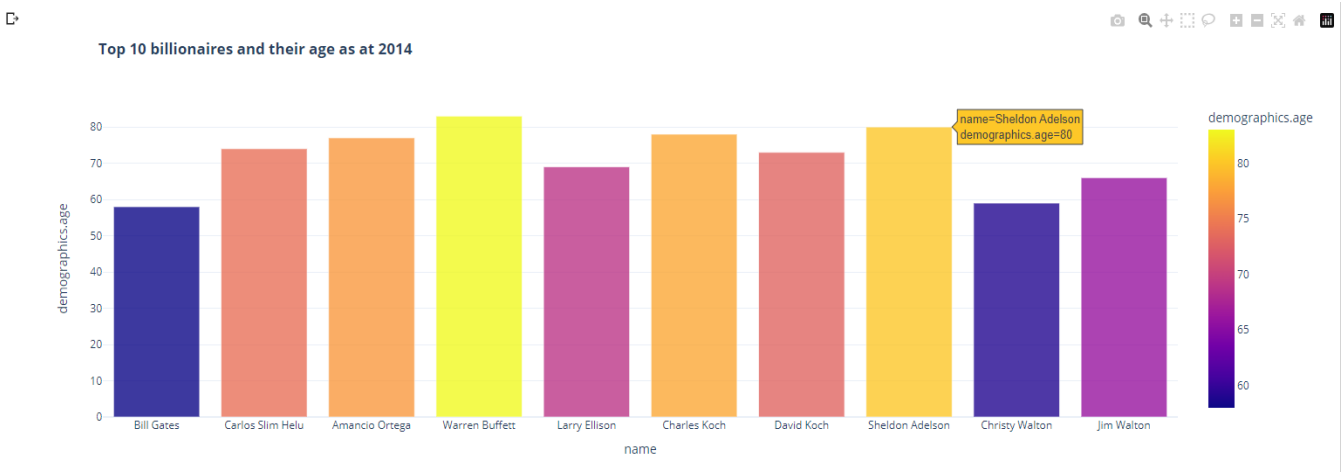


Fig. 28. A bar graph showing the ages of the top 10 billionaires.

Perspectives from Fig. 28:

- The world's number one billionaire is 58 years old.
- The top ten richest people are all over 50 years old.
- Warren Buffett, who is over 80 years old, is the oldest person among the top ten billionaires.

Fig. 29 presents the top ten youngest billionaires.

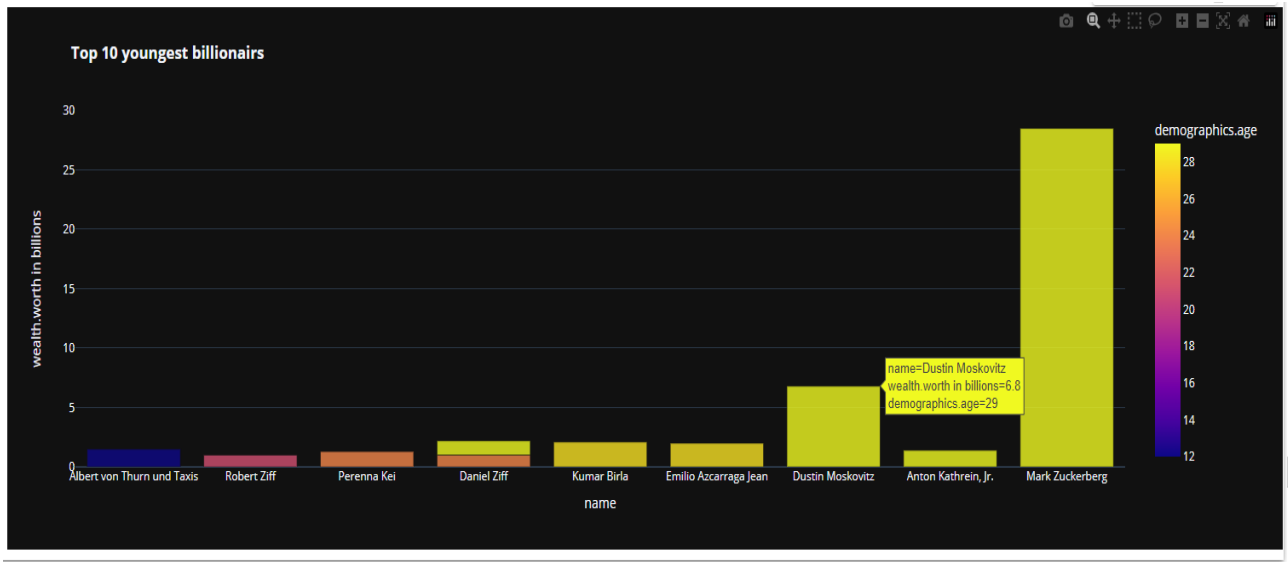


Fig. 29. The top ten youngest billionaires are represented by a bar chart.

Perspectives from Fig. 29

- Albert Thurn, who is only 12 years old and worth a billion dollars, is the youngest billionaire.
- Mark Zuckerberg is the wealthiest young person, with a net worth of more than $25 billion.

Fig. 30 shows the industries of the top ten youngest billionaires.

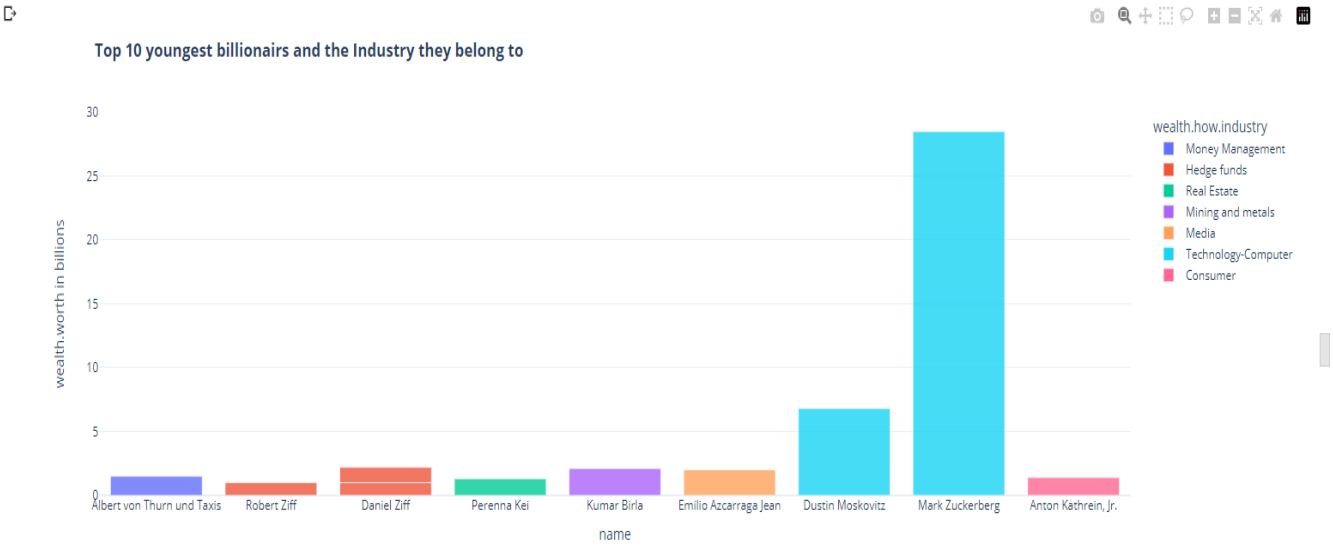


Fig. 30. A bar graph depicting the industries of the world's youngest billionaires.

Perspectives from Fig. 30:

- The youngest billionaire works in the financial services industry.
- The richest young billionaire is from the technology-computer industry.

Fig. 31 exhibits the top ten female billionaires and their ages.

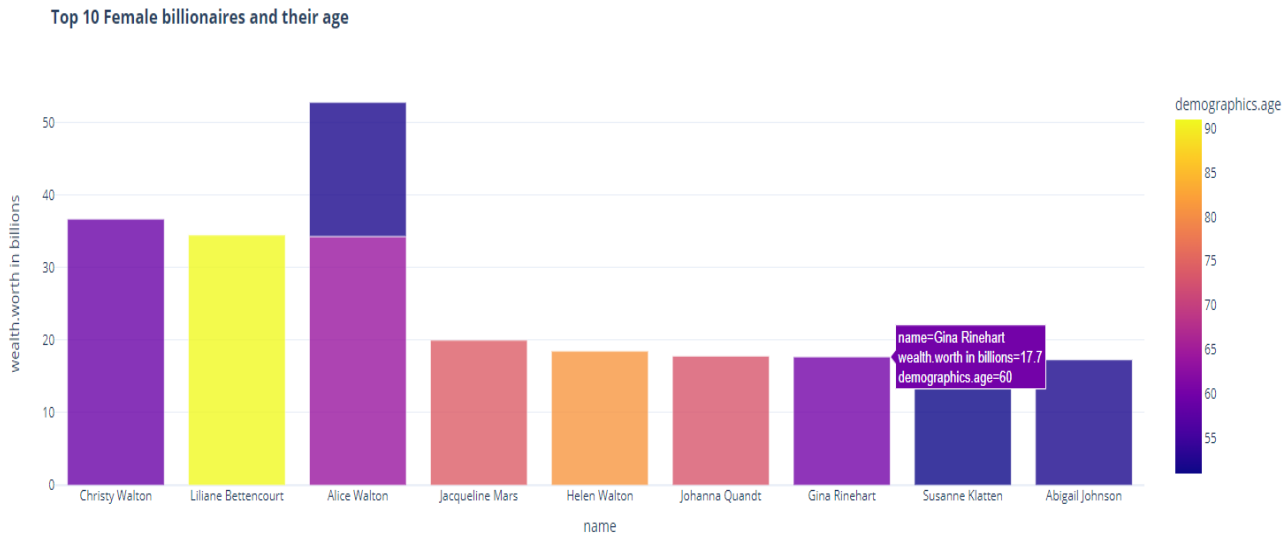


Fig. 31. The ages of the top ten female billionaires.

Perspectives from Fig. 31:

- Christy Walton, the wealthiest female billionaire, is 59 years old and worth more than 36 billion dollars.
- Lillian Bettercourt, who ranks second among female billionaires, is over 80 years old.

Fig. 32 highlights the top ten billionaires and the industries in which they work.

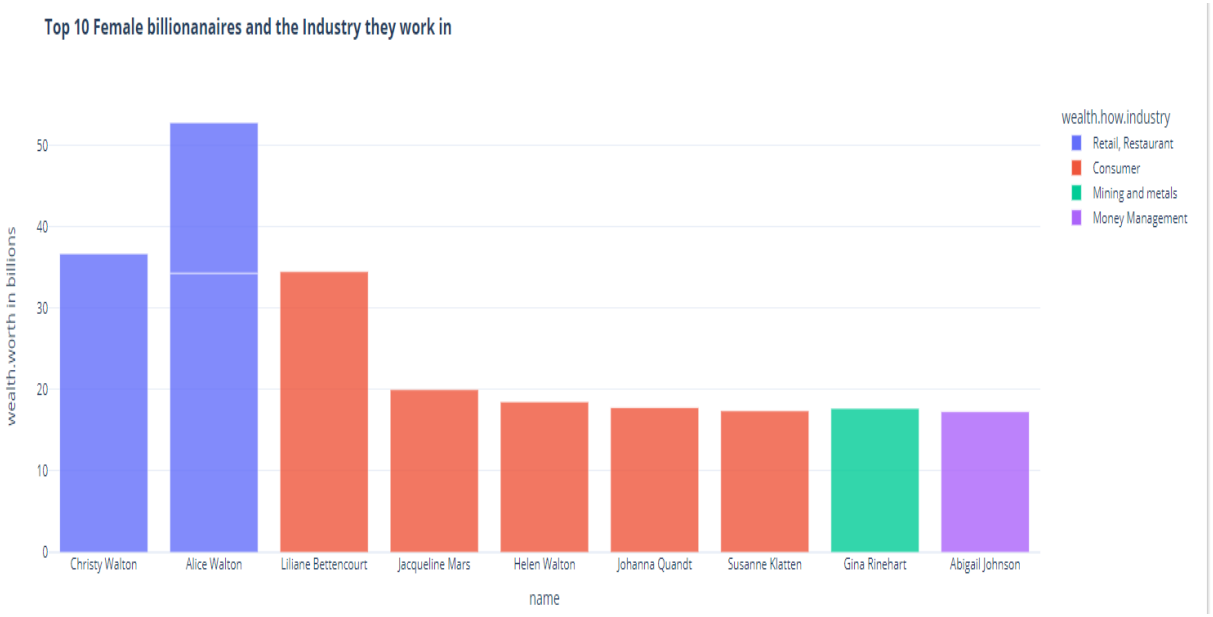


Fig. 32. A bar chart showcasing the top 10 female billionaires and the industries in which they work.

Perspectives from Fig. 32:

- The richest female billionaire works in retail and restaurants.
- Most female billionaires work in the consumer business.

The top ten industries' billionaires combined net worth is shown in Fig. 33 for 2014.

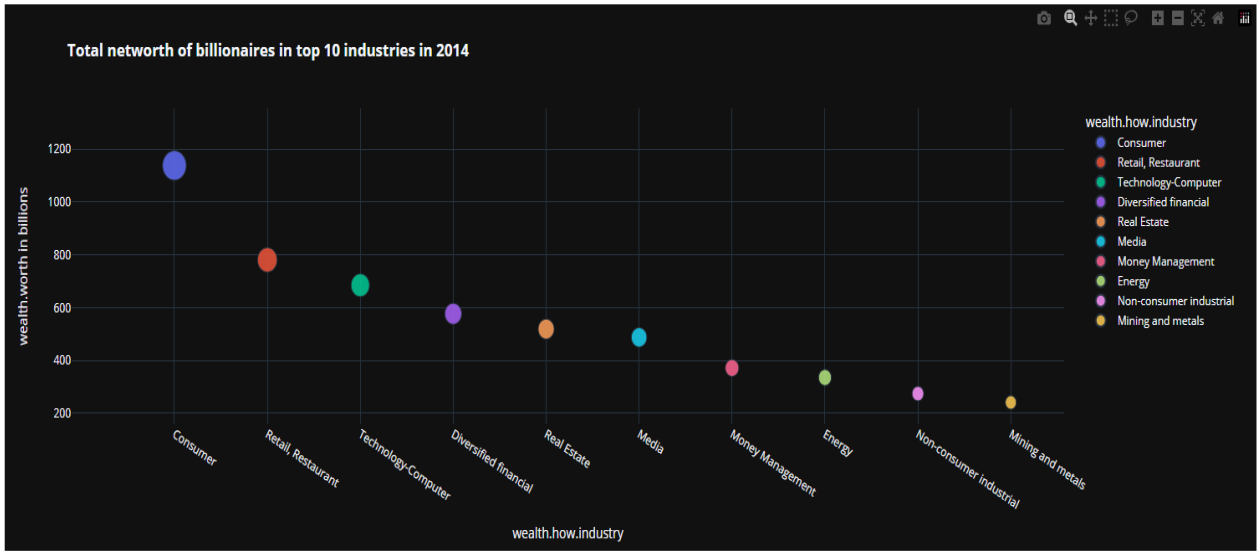


Fig. 33. Billionaires combined net worth in the leading ten sectors in 2014.

Perspectives from Fig. 33:

- The largest sector, valued at over $1 trillion, is the consumer goods sector.
- The retail, restaurant industry comes in second, with a market value of $800 billion.

XIII. Testing the Hypothesis

1) Hypothesis 1: Does a billionaire's age affect his or her net worth?

Ho: A billionaire's net worth is not significantly related to age.

H_A: A billionaire's net worth is related to age in a significant way.

The code for evaluating the first hypothesis is shown in Fig. 34.

```
import pandas as pd
from scipy.stats import pearsonr

# Compute Pearson correlation coefficient and p-value

# THIS TESTS THE HYPOTHESIS OF THE RELATIONSHIP BETWEEN AGE AND NETWORTH

corr, p_value = pearsonr(billionaires['demographics.age'], billionaires['wealth.worth in billions'])

# Print the results
print("Pearson correlation coefficient:", corr)
print("p-value:", p_value)
```

```
Pearson correlation coefficient: 0.08750438991654515
p-value: 3.990971769605956e-05
```

Fig. 34. Test code for Hypothesis 1.

Perspectives from Fig. 34:

- The association between age and net worth was examined using the Pearson correlation coefficient.
- Considering a p-value of 0.00004, it is evident that the age of a billionaire and their wealth are strongly correlated. We agree with an alternate concept as a result.
- The more elderly billionaire is most likely to be richer.

2) Hypothesis 2: Men with billion-dollar wealth are wealthier than women?

Ho: A billionaire's net wealth is unaffected by gender in a significant way.

H_A: Gender significantly affects a billionaire's financial worth.

The code to test the second hypothesis is shown in Fig. 35.

```
import scipy.stats as stats

# Filter the data for male and female billionaires
male_df = billionaires[billionaires['demographics.gender'] == 'male']
female_df = billionaires[billionaires['demographics.gender'] == 'female']

# Conduct the two-sample t-test assuming equal variance
t_stat, p_val = stats.ttest_ind(male_df['wealth.worth in billions'], female_df['wealth.worth in billions'], equal_var=True)

# Print the t-statistic and p-value
print("t-statistic:", t_stat)
print("p-value:", p_val)
```

```
t-statistic: -0.9940802457505286
p-value: 0.32029332217806183
```

Fig. 35. Test code for Hypothesis 2.

Perspectives from Fig. 35:

- This hypothesis was tested using a two-sample t-test.
- Gender and net worth are unrelated, as indicated by the p-value of 0.32, which is larger than 0.05.
- We can rule out the alternative hypothesis and adopt the null hypothesis which is billionaires' net worth is unaffected by gender.

3) Hypothesis 3: The net worth of billionaires varies considerably depending on the industry area.

Ho: A billionaire's net worth is significantly influenced by the industry sector.

H_A: A billionaire's net worth is not much impacted by their industry sector.

Fig. 36 illustrates the code for testing the third hypothesis.

```
import scipy.stats as stats

# Group the data by industry and select the net worth column
grouped_data = billionaires.groupby('wealth.how.industry')['wealth.worth in billions']

# Perform one-way ANOVA test
f_statistic, p_value = stats.f_oneway(*[grouped_data.get_group(x) for x in grouped_data.groups])

print("F-statistic:", f_statistic)
print("p-value:", p_value)
```

```
F-statistic: 2.745769472721782
p-value: 0.00022408174979504465
```

Fig. 36. Test code for Hypothesis 3.

Perspectives from Fig. 36:

- This hypothesis was investigated using the one-way ANOVA test where the p-value is less than 0.05.
- We can accept the null hypothesis and concur that the industry sector significantly influences billionaires' net wealth.

XIV. Evaluation

The study revealed how gender, age, and industry affect billionaire wealth accumulation. The retail industry has the most female billionaires, and most billionaires are men. According to our research, billionaires are usually 50–60 years old. The most senior billionaire was 98 years old, while the youngest was 21. Furthermore, real estate, media, and construction have more millionaires than other industries. The insights on billionaire wealth distribution can inform investment decisions and corporate strategies across sectors.

Our investigation has shown the key variables that enable billionaires to become rich. The above insights could help decision-makers, investors, and businesspeople understand wealth distribution and develop successful financial strategies in the global economy.

XV. Findings and Results

The following is the overall finding from the billionaire dataset analysis:

- The United States is the country that has the highest number of billionaires, which has 804 in total.
- The sector with the most billionaires is the consumer goods sector.
- The majority of billionaires are older than 50.
- 34.8% of billionaires received their wealth through inheritance.

- With a cumulative net worth of 76 billion greenbacks, invoice Gate is the richest billionaire.
- Christy Walton has a net worth of $38 billion, making her the richest female billionaire.
- The youngest billionaire is Albert Thurn, who's just 12 years old and possesses a $1 billion fortune.

XVI. Conclusions

In summary, our examination of the "Billionaires.csv" dataset utilizing Python tools for big data analytics yielded significant findings regarding the determinants that impact the accrual and allocation of riches among billionaires. Our observation revealed a noteworthy accumulation of wealth within a minority of individuals, indicating the existence of income disparity within the billionaire demographic.

The study found that entrepreneurship and technology helped a significant number of billionaires in the sample become wealthy. Technology and finance were the most profitable industries for billionaires. The analyses also showed that the US and China have the most billionaires. Billionaires also showed a trend toward gender diversity. This study sheds light on billionaires, but it has limitations. The dataset covers billionaires until 2021 and may not reflect current trends. The analysis only considered quantitative variables, ignoring qualitative factors that may contribute to wealth accumulation.

Our examination also provides a valuable contribution to the comprehension of the fluctuations in affluence within the billionaire population, exposing trends linked to entrepreneurial pursuits, sectors of operation, geographical location, and gender. Subsequent investigations ought to adopt a more all-encompassing methodology by combining qualitative and longitudinal data to acquire a more intricate comprehension of the accumulation of billionaire wealth and its societal ramifications.

XVII. Limitations and Future Work

There are a few limitations to our dataset and methodology, even though our analysis offers insightful information about the elements that lead to self-made billionaires' wealth accumulation.

- The dataset excludes billionaires who received their riches through inheritance and is only comprised of self-made billionaires. The generalizability of our findings to the total billionaire population may be impacted by this exclusion.
- The dataset may not be typical of people who have less net worth because only billionaires with total assets of more than one billion dollars are included in it. Additionally, because the data was taken over a limited period, probably, it doesn't accurately reflect patterns and shifts in wealth distribution.
- Descriptive statistics and exploratory data analysis methods played a significant role in our investigation. Although these techniques help understand the data, they do not prove causation or take into consideration confounding variables.
- The dataset may contain incomplete or erroneous data, which could produce biased or insufficient results. These restrictions must be understood and considered when interpreting the findings of our investigation.

Notwithstanding these constraints, the dataset can yield valuable insights into the determinants that contribute to the ascent of self-made billionaires. The determinants encompass the sector, age, gender, and regional attributes of the billionaires. The findings derived from this dataset can offer precious insights for companies and traders in their choice-making procedures, as well as for an academic look at the societal implications of wealth inequality. This information has the potential to be utilized for several objectives. To augment our understanding of the intricacies surrounding the wealth of billionaires and its far-reaching consequences, future research undertakings should integrate both quantitative and qualitative methodologies, incorporate longitudinal data, and investigate more extensive array of variables. This will facilitate a more holistic comprehension of the complexities associated with the amassing and dispersal of wealth among individuals in the billionaire class.

References

[1] Njangang, H., Beleck, A., Tadadjeu, S., & Kamguia, B. (2022), 'Do icts drive wealth inequality? evidence from a dynamic panel analysis', *Telecommunications Policy*, 46(2), 102246. Available at: https://doi.org/10.1016/j.telpol.2021.102246.

[2] Torgler, B. and Piatti, M. (2013) 'Extraordinary wealth, globalization, and corruption', Review of Income and Wealth, 59(2), pp. 341–359. Available at: doi:10.1111/roiw.12027.

[3] Kong, G., Xu, L. and Zhang, W. (2022) 'The benevolence of the billionaires: Evidence from China's Hurun Rich List', Finance Research Letters, 48, p. 103030. Available at: doi: 10.1016/j.frl.2022.103030.

[4] Ghosh, S. (2016) 'Billionaire wealth, firm performance and Financial Crisis: An empirical analysis for India', South Asian Journal of Macroeconomics and Public Finance, 5(2), pp. 133–156. Available at: doi:10.1177/2277978716671050.

[5] Bagchi, S. and Svejnar, J. (2013) 'Does wealth inequality matter for growth? the effect of billionaire wealth, income distribution, and poverty', SSRN Electronic Journal [Preprint]. Available at: doi:10.2139/ssrn.2351448.

[6] Jacobs, D., 2015 'Extreme wealth is not merited', Oxfam Discussion Papers.

[7] Hemingway, A. and Rozworski, M. (2020) 'Canadian billionaires' wealth skyrocketing amid the pandemic', Ottawa, ON, CA: Canadian Centre for Policy Alternatives.

[8] Frémeaux, N. and Leturcq, M. (2022) 'Wealth accumulation and the gender wealth gap across couples' legal statuses and matrimonial property regimes in France', European Journal of Population, 38(4), pp. 643–679. Available at: doi:10.1007/s10680-022-09632-5.

[9] Hamermesh, D. and Leigh, A. (2022) '"Beauty too rich for use": Billionaires' assets and attractiveness', SSRN Electronic Journal [Preprint]. Available at: doi:10.2139/ssrn.4114299.

[10] Wai, J. and Lincoln, D. (2016) 'Investigating the right tail of wealth: Education, cognitive ability, giving, network power, gender, ethnicity, leadership, and other characteristics', Intelligence, 54, pp. 1–32. Available at: doi: 10.1016/j.intell.2015.11.002.

[11] List, J., Shaikh, A., & Xu, Y. (2016) 'Multiple hypothesis testing in experimental economics', Experimental Economics. Available at: https://doi.org/10.3386/w21875

[12] Gillis, A.S. (2021) 'The 5 Vs of big data, Data Management.' Available at:

https://www.techtarget.com/searchdatamanagement/definition/5-Vs-of-big-data (Accessed: 07 May 2023).

[13] Nguyen, T.L. (2018) 'A framework for five big Vs of Big Data and organizational culture in firms', 2018 IEEE International Conference on Big Data (Big Data) [Preprint]. Available at: doi:10.1109/bigdata.2018.8622377.

[14] Hamermesh, D. and Leigh, A. (2021) "Beauty too rich for use": Billionaires' assets and attractiveness [Preprint]. Available at: doi:10.3386/w29361.

[15] Jesús Fernández-Villaverde, Hurtado, S., and Galo Nuño (2023), 'Financial Frictions and the Wealth Distribution', Econometrica, 91(3), pp.869–901. doi: https://doi.org/10.3982/ecta18180.

[16] Kohlrausch, G. and Goncalves, S. (2023) 'Wealth distribution on a dynamic complex network' [Preprint]. doi:10.2139/ssrn.4673263.

[17] Piper, R. (2023). The Institutional Drivers Contributing to Billionaire Wealth at the Sector Level. Class, Race Corporate Power, [online] 11(1). doi: https://doi.org/10.25148/crcp.11.1.010593.

[18] Gastwirth, J.L., Luo, R. and Pan, Q. (2024). A statistical examination of wealth inequality within the Forbes 400 richest families in the United States from 2000 to 2023. METRON. doi: https://doi.org/10.1007/s40300-024-00267-6.

[19] Dataset - Billionaires CSV File, https://corgis-edu.github.io/corgis/csv/billionaires/ [Accessed on 20 March 2024].

XVIII. KEY TERMS AND DEFINITIONS

Big Data Analytics: The procedure of using advanced computer and statistical tools to analyze and clean meaningful data from huge and complex datasets.

Data Visualisation: The graphical presentation of data to highlight patterns, trends, and insights to make challenging material more approachable and understandable.

Inheritance: The practice of transferring wealth, assets, or property from one generation to the next, usually using close family ties.

Economic Growth: The amount of goods and services produced by an economy increases with time. GDP growth, which measures the total value of a nation's finished goods and services, is typically used to measure it. Economic growth reveals productivity, living standards, and prospects for both businesses and people.

Hypothesis Testing: A method that uses statistical analysis to determine whether a hypothesis or claim made about a population is supported by the data from a sample of that community.

Data Pre-processing: The manipulation of raw data using a variety of methods to prepare it for subsequent processing is what is referred to as "data pre-processing," and it is an essential part of the data preparation process.

Python: A high-level programming language that is well-known for its ease of use and readability and that finds widespread use in the fields of scientific computing, machine learning, and data analysis.

AI-Enabled Vision Transformer for Automated Weed Detection: Advancing Innovation in Agriculture

Shafqaat Ahmad, Zhaojie Chen, Aqsa, Sunaia Ikram, Amna Ikram
Data Scientist, Brandt Group of Companies, Canada
Department of Food Science and Nutrition, The Hong Kong Polytechnic University
Department of Computer Science, COMSATS University Islamabad, Pakistan
Department of Software Engineering, IUB, Pakistan
Department of Computer Science, GSCWU, Pakistan

***Abstract*—Precision agriculture is focusing on automated weed detection in order to improve the use of inputs and minimize the application of herbicides. The presented paper outlines a Vision Transformer (ViT) model for weed detection in crop fields, that tackle difficulties stemming from the resemblance of crops and weeds, especially in complex, diversified settings. The model was trained via pixel-level annotation of the images obtained using high-resolution UAV imagery shot over an organic carrot field with crop, weed, and background. Due to the nature of the mechanism in ViTs that includes self-attention, which allows it to capture long-range spatial dependencies, this approach can very well distinguish crop rows from inter-row weed clusters. To solve the problem of class imbalance and improve the generality of the patches, techniques of data preprocessing such as patch extraction and augmentation were used. The effectiveness of the proposed approach has been confirmed by an accuracy of 89.4% in classification, exceeding the efficiency of basic models such as U-Net and FCN in practical application conditions. This proposed ViT-based approach is a marked improvement in crop management; and provides the prospect for selective weed control, in support of more sustainable agriculture. This model can also be integrated into AI-based tractors for real-time weed management in the field.**

***Keywords*—Precision agriculture; weed detection; vision transformer; UAV imagery; crop-weed classification; AI-Tractors**

I. Introduction

In light of such global factors as climate change, increasing population, and declining land fertility, protection of food production has become an important task [1]. Amongst various biotic constraints that affect crop yield and quality, weeds rank as some of the most formidable challenges that crop producers face in the field [2]. If not controlled, weeds have severe effects on crop yield and quality hence contributing to loss making and high food insecurity [3]. In the past, weed management has been undertaken by mechanical means such as pulling weeds out by hand or the widespread application of herbicides [4]; either of which is now considered to be unfavorable. Hand weeding is hard and cannot be used in large scale farming [5], while chemical control causes pollution and health issues [6], reduces bio-diversity, and results into the evolution of herbicide resistant weeds. Therefore, there is need to develop effective, sustainable as well as economic methods to tackling weed problems.

New developments in precision agriculture especially in combination with technology such as remote sensing, machine learning and drone systems, are revolutionizing conventional weed control approaches [7]. Precision agriculture is the practice of trying to grow crops as efficiently as possible by giving farmers instant information about the condition of their fields so they can manage the resources they use in the most sustainable way [8]. With UAVs using high resolution and multispectral cameras available for field monitoring, large scale data acquisition coupled with detailed visualisation of crop and weed distribution in the agricultural environments is possible [9]. It is possible to use this technology to locate and identify weeds and subsequently manage by providing efficient spot treatment as opposed to weed eradication using herbicides.

Nevertheless, identification and precise categorization of weeds in crop fields still pose a great challenge due to factors such as variability in the field, weeds growing between rows of crops and close resemblance in appearance of weeds and crops [10]. These difficulties cannot be resolved by using conventional image processing techniques, because such approaches rely on color-based or shape-based segmentation, which may not be sufficient for distinguishing between very similar plant species in different lighting and environmental conditions [11]. To overcome these limitations, machine learning particularly deep learning approaches has been used to improve weed detection accuracy. Convolutional neural networks (CNNs) have been reported to work well in this area [12], however, due to their constrained local connectivity, they lack the ability to capture the spatial dependencies and context required to correctly identify weeds from crops especially in high density field setting.

Recently, Vision Transformers (ViTs) emerged as a compelling approach to surpass CNNs in image classification problems [13]. First introduced for natural image understanding, ViTs utilize the self-attention method and it provides a wide-angle view of long-range dependencies within the image, which is crucial in agriculture. Unlike CNNs, ViTs can handle the analysis of the entire image regions rather than focusing on localized features needed for crop and weed differentiation [14]. Self-attention enables ViTs to distinguish between crop rows and inter-row weed clusters more accurately than in field conditions where crop plants and weeds appear to have similar textures and color patterns.

This research introduces a new method for the automated detection of weeds based on a Vision Transformer model that

has been developed to handle the specific difficulties of agricultural weed categorisation in UAV imagery. The proposed method takes advantage of the fact that crops, as a rule, are planted in a geometric pattern of rows while weeds grow randomly across the farm field; therefore, crop regions can be distinguished from the clusters of weeds by their geometric arrangement. The method we propose here is to employ the Vision Transformer model on the high-resolution UAV images at the pixel level to accurately distinguish crops from weeds. The training dataset is CWFID, for each image, background, crop, and weed pixels are labeled in detail with the help of experienced farmers, which supplies the model with profound features for learning intricate spatial associations.

Using the efficiency of image preprocessing including patch extraction and data augmentation and the feature of long-range dependencies analysis of ViT model we expect to receive high classification accuracy and good scalability in field conditions. This study advances understanding of weed biology and the potential for selective, efficient weed control by identifying specific proteins that allow for accurate discrimination of different weed species. Consequently, the study responds to important research questions in PA and opens up opportunities toward building more sustainable and less hazardous crop growing systems.

The remainder of this paper is organized as follows: Section II discusses related work, which presents an idea of this research area and the inclusive techniques for weed detection and its merit and demerit. Section III outlines the research approach of this study, which covers ViT architecture, datasets, data preprocessing, and evaluation of crop and weed classification. Section IV explains the findings that include the assessment of the ViT model and a comparison with other conventional models like U-Net and FCNs. Section V offers a discussion of the findings, issues on model stability and possible applications of the developed models to precision agriculture. Last, Section VI provides a conclusion to the study by offering an overview of the major conclusions, the main research contributions, and an indication of the areas where future studies and enhancements may be made.

II. Related Work

There are different approaches for weed detection mentioned in the literature for the use of different image acquisition systems. The first one is carried out by separating vegetation from the background as soil and residues to separate crops from the weeds. The common segmentation process handily uses the color Methods [15] and Multispectral data in order to segment vegetation from background using fixed indices which make vegetation segmentation possible. Nonetheless, differentiating between weeds and crops using spectral data prove difficult since the two are spectrally similar. Therefore, approaches focusing on the region level, which utilizes spatial pixel configurations, are mostly used [16].

The detection of weeds in agriculture has improved over the years with the help of color-based segmentation algorithm. Hue based indices like the Excess green Index (ExG) are used widely to sharpen vegetation features in imagery, isolating plants from their surroundings. This approach is especially valuable when dealing with multispectral data since, as it was mentioned, ExG uses the green component most to enhance vegetation. This method has been found to be computationally efficient for the initial step of separating crops from weeds in agricultural scenarios and laid down a base for further analysis and classifying more steps [17]. Another level of enhancement of weed detection is obtained by integrating Excess Green with Otsu's thresholding technique which segment images at the optimum threshold intensity values. The integration method is passes in minimizing the background noise while maximizing vegetation details. Together with the double Hough transform, this method improves the identification of crop lines in images with perspective distortion by recognizing and reorienting the lines in a complex environment in agriculture. They are particularly useful in the images of the same scene taken under varying lighting conditions since they increase the resistance when classifying crops from weeds [18].

TABLE I. Previous Weed Detection Methods

Method	Description	Reference
Color-Based Segmentation	Separates vegetation from background using color indices such as Excess Green (ExG) and fixed indices in multispectral data.	[17]
ExG and Otsu's Thresholding	Combines Excess Green and Otsu's thresholding to eliminate background, then uses double Hough transform to identify crop lines in perspective images.	[18]
Object-Based Image Analysis (OBIA)	Uses UAV imagery and multiscale algorithms to segment crop rows from weeds, creating homogeneous objects for analysis.	[19]
2D Gabor Filters with ANN	Uses 2D Gabor filters to capture texture features and an artificial neural network (ANN) classifier for weed detection.	[20]
Morphological Characteristics	Utilizes morphological features to distinguish weeds in maize fields, using neural networks and support vector machines (SVMs) with shape-based features.	[21]
Edge Frequencies & Vein Density	Differentiates weeds from crops by analyzing edge frequencies and vein density differences in the leaves.	[22]
Otsu Thresholding & K-means/SVM	Applies Otsu thresholding for background removal and uses k-means clustering and SVM classifier for crop-weed classification, successful in sunflower fields.	[12]
Wavelet Transform & Neural Network	Uses wavelets to capture texture details and a neural network for classification, effective for recognizing various weed types in sugar beet fields.	[23]
SVM, ANN, & Random Forests in OBIA	Employs machine learning models like SVMs, ANNs, and Random Forests within the OBIA framework, especially for weeds in maize fields.	[24]
Convolutional Neural Networks (CNNs)	Uses CNN architectures, including AlexNet, for weed detection in crops such as water hyacinth and serrated tussock. Applied in UAV-based imagery and mobile robot systems.	[25]
Spatial & Spectral Domain Features	Integrates Hough transform for spatial features with multispectral data for spectral features, combined with SVM for crop-weed classification in four-band imagery.	[26]

Another complex technique is Object-Based Image Analysis (OBIA), which divides images into areas of the same character instead of single pixels, within the use of multistate algorithms. When applied to UAV imagery, OBIA provides a better defined and can be easily automated procedure to distinguish crops from weeds. This approach is useful in vast areas where exact methods like pixel based approach turn out to be more computational. Thanks to OBIA, grouping similar pixels into coherent objects, researchers are able to distinguish the pattern of weed distribution across the crop rows, which enhances weed control strategies [19]. The combination of texture analysis in the form of 2D Gabor filters with Artificial Neural Networks (ANNs) introduces a promising solution to the problem of weed detection due to the utilization of frequency and orientation within images. The enhanced textural features that are fundamental to crops and weeds are well captured by Gabor filters. ANNs then sort these features, and the model has a high level of accurate weed detection in crops with textural features. This method offers considerable reliability to precision agriculture, above all in areas of uniform textural characteristics where texture differential is significant [20].

Shape based features of Morphological characteristics are another factor that builds another level of discrimination in case of weed identification. Methods that apply such factors as shape, size, and structure of the leaves using neural networks and Support Vector Machines (SVMs) are preferable in structured crops such as maize. Morphological features are unique depending on the type of crop or the weed in question, and therefore helpful where the shape differences are quite profound. Such a strategy can be especially valuable for detecting specific weed types that differ from crops morphologically [21]. Another notable feature which is used in classification of weeds is the patterns that appear on the leaves 'veins. Vein density methods and edge frequency methods help to distinguish crops and weeds because crops and weeds essentially have different vascular networks within the veins of their leaves naturally. This technique is most successful in the controlled environments where crops and the weeds differences in vein densities are clearly noticeable. Due to this focus on these several anatomical dissimilarities, this approach is suitable for high precision detections in small-scale or research production agriculture setting [22].

Furthermore, using thresholding Otsu together with clustering and classification method such as K-means and SVM makes a strong way of detecting weeds in areas such as sunflower crops. Otsu's thresholding erases the background noises while k-means clusters all the pixels having a nearly similar intensity, which is then sophisticatedly classified by the SVM in order to separate weeds correctly. Thus, this work follows gradual layering of steps that help increase the weed detection accuracy, and that are tested effective even in high noise images [12]. When used alongside neural networks, the wavelet transform is a useful method of weed detection through texture analysis. Wavelets analyze small local details of the image and since neural networks can provide high accuracy when determining the difference between the weeds. This technique has been particularly effective in sugar beet fields where due to the multi specie flora the different weeds can be identified using the features obtained by the wavelet analysis of the images [23].

Currently, the use of OBIA has included some common machine learning models, such as SVMs, ANNs, and Random Forests. This approach especially for maize fields incorporates an object-based image analysis with machine learning concept leading to higher accurate detection in large-scale agriculture. Thus, the classifiers within and across the imagery segments enhance the models to increase classification results in high complexity areas where the mere pixel-based approach could not limit the classification process [24]. Convolutional Neural Networks (CNNs) are that key technology which helps weed detection using high-dimensional data and pattern extraction. The state of the art CNNs, such as the AlexNet, has been implemented in the classification of weed crops such as water hyacinth and serrated tussock. These models are particularly suitable for UAV and mobile robotic systems where high versatility of weeds and constant ability to perform well in different conditions is needed. The feature extraction capacity of CNNs makes them useful in agricultural systems particularly where big data samples can be used in training and model refinement [25].

Last of all, advanced techniques that combine spatial and spectral characteristics of the analysed images, including Hough transform method with the use of multispectral imaging and support vector machines, can be pointed to as an enhanced method for crops and weeds differentiation. This approach takes advantage of spatial characteristics and spectral variation of four bands in imagery for precise analysis in precision agriculture. This method involves combining of spectral data with spatial transformation to result in high classification accuracy particularly in fields where spectral and spatial discrimination is well defined [26].

III. Proposed Methodology

In modern agriculture, most crops are planted in organized rows with defined spaces, depending on the crop type. Vegetation that grows outside these rows is generally identified as weeds, known as inter-row weeds. Leveraging this spatial organization, several studies have implemented weed detection methods based on the geometric properties of crop rows. A key benefit of these methods is that they are largely unsupervised, reducing the need for manual training data. Building on this, our approach first identifies crop rows, then labels inter-row vegetation as weeds to create a training database. We categorized this data into two classes, crop and weed, and used it to train a Vision Transformer (ViT) model to detect and classify crops and weeds from UAV imagery. Fig. 1 provides an overview of the main steps in the proposed method, with detailed descriptions following.

Crop/Weed Field Image Dataset (CWFID) is one of the vital resources for demonstrating the models of machine learning for classification of crops from weeds. The data in this paper was obtained from an organic carrot field in Northern Germany with the help of an autonomous field robot called Bonirob which has a high-resolution multi-spectral camera. Collected during the vegetation phase of the crops, the images offer a real-world

representation of crop and weed status in the fields, which is useful for precision agriculture studies with detailed descriptions of both crops and weeds present in the image. The dataset comprises 60 high-resolution images with the size of 1296 x 966 pixels. The fine details present in the presented images make it possible for models to differentiate vegetation features, and also differentiate between plants that are growing closely together. Each image in the dataset is fully annotated at the pixel level by agricultural experts, classifying each pixel into one of three categories: The three categories of organisms identified in the study area include Background (Soil), Crops (Carrot Plants) and Weeds.

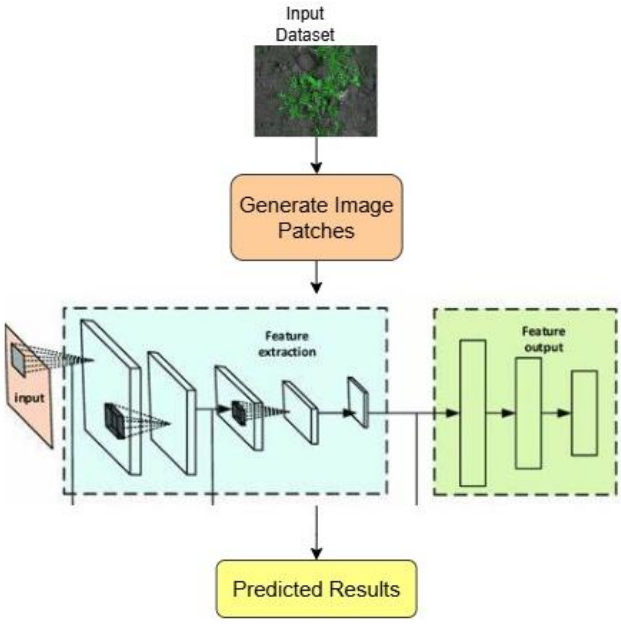


Fig. 1. Proposed architecture flow.

The expert annotations of the CWFID dataset allow for the identification of crops and weeds in a highly accurate, even in the most complicated agricultural environment. Every class in the case of the given dataset is a category of content that helps in differentiating between plants and soil. The three annotated classes are described in detail below:

1) Background (Soil): This class consists all the bare grounds particularly the soil that is inter and intra-row of crops. These background regions makes it easier to define whether an object is crop or weed since they creates a contrast. The pixel distribution across the three classes is as follows and has been presented in Table II to indicate the class imbalance problem similar to that of realistic problem.

2) Crops (Carrot Plants): This class is made up of areas with carrot plant bodies, which are usually aligned in an orderly manner. They contribute to the improvement of the effectiveness of the classification between crop and non-crop areas due to the crop rows formed. The carrot plants were in different stages of maturity which provided a variety that helps the models understand different crop morphologies.

3) Weeds: In this dataset, weeds comprise intra-row weeds, which are those established within the crop rows, and inter-row weeds, which are those established between the crop rows. Classification models are further complicated by the presence of intra-row weeds since they are similar in size and color to crops. The presence of a large number of different weeds improves the applicability of the dataset for developing reliable machine learning models capable of distinguishing between crops and different weed types.

Because of the high-resolution images and corresponding detailed annotation, the CWFID dataset is more suitable for precision agriculture in which precise weed maps are required at the pixel level. The dataset offers several unique challenges:

1) Class imbalance: The fact that there are more background and crop pixels than weed pixels is a more realistic representation of the field environment to encourage the use of methods such as data enhancement and class balancing.

2) Intra-row and inter-row weeds: Moreover, the combination of both inter-row and intra-row weeds poses a difficult classification problem for the models, where the presence of weeds in between crop rows is also considered.

3) Varied lighting and vegetation density: The dataset comprises images taken under varying light conditions and at different vegetation cover densities making it more challenging to classify while improving model resilience.

The CWFID dataset is freely accessible from GitHub and can be used by researchers interested in crop and weed classification for agricultural applications. The specific use of this dataset is to contribute to the generation of models for improving precision agriculture, especially in the case of weed detection and control within crop fields.

A. Data Preprocessing

To train effectively and to generalise the crop-weed classification model, some modifications were made on the CWFID dataset during data preprocessing. These steps were taken in an effort to bring the format of the data fed into the classifier more to a unified level, also to equalize the ratio of the classes and increase the variety of training samples (Table I).

1) Image resizing: When analyzing the CWFID dataset, it was found that each of the original images has a size of 1,296 x 966 pixels, and thus requires downscaling for input into the most of the deep learning models. To keep it manageable for the model, all of the images were also scaled to a size that would fit the input size of the chosen model. This resizing made all the inputs have equal dimensions thus making the model to undergo training without the need for further resizing during training. The option of resizing was applied more conservatively, allowing the image to maintain as much of its quality as possible, and at the same time, decreasing the amount of computations needed.

2) Patch extraction: To prepare the data for pixel level classification, each resized image was then split into fixed size patches of 8×8 pixels. Patch extraction serves several purposes:

a) Localized feature capture: Since a large image is divided into small segments of patches, C&Ps can identify crop

and weed characteristics confined or restricted to particular regions.

b) Class representation: Each patch was made around regions marked as crop or weed and both of them were equally represented in the training set.

c) Memory efficiency: Smaller patches also mean that Memory is used, which is a great advantage since more patches mean more data for the AI model to train on, especially and particularly when using humongous data sets.

In general, patches consisting only of background pixels were often removed in order to focus on crop and weed area. They also diminished unnecessary data and at the same time enhanced the specificity of the data set with regard to crops and weeds differentiation. If a patch contained both crop and weed pixels it was classified into a patch class with the highest pixel count in a particular patch in order to of class labeling.

3) Data augmentation: Since there are significantly fewer weed pixels than crop pixels in the CWFID dataset, data augmentation was used to increase the number of weed samples and, therefore, improve the model's performance. Augmentation techniques were applied uniformly across both classes to expose the model to a variety of conditions, as detailed below:

a) Rotation: Each patch was rotated at 90°, 180°, and 270° rotations. This rotation not only amplified the dataset up to four folds but also let the model learn about the rotational variability needed for recognizing weeds and crops in multiple angles.

b) Contrast adjustments: To mimic different lighting scenarios that could be met in practice, the contrast of a patches was altered randomly. Crop, weed, and background boundaries were highlighted through higher contrast settings; lower contrast mirroring conditions such as low light or shadow. This adjustment improved the ability of the model to be sensitive to variations in the levels of illumination in the environment.

c) Gaussian smoothing: Specifically, Gaussian smoothing, or blurring, was applied only to minimize the noise in the image and enhance the main characteristics of each patch. High frequency components and significant intensity variations were removed through applying a Gaussian filter, and this enabled the model to detect general features. This technique also assisted in lowering the impact of noise and enhanced generalization in some instances.

4) Balanced dataset composition: To reduce the class imbalance, dataset was augmented in such a way that both crops and weeds had almost equal representation. Augmented weed patches were particularly helpful in countering this in data collection because crop areas are generally more abundant. To this end, the findings demonstrated that it is possible to get a near balanced distribution between the two classes and this made the model to perform well in making discriminations between the two classes without favoring the larger class. Through this detailed data preprocessing step, the CWFID dataset was well-prepared for training the Vision Transformer model, which then captured important aspects of both crops and weeds and succeed under different field conditions.

TABLE II. SUMMARY OF DATA PREPROCESSING TECHNIQUES USED

Preprocessing Step	Description	Purpose
Image Resizing	Standardized input size for all images	Ensures consistency and reduces memory usage
Patch Extraction	64 × 64 pixel patches centered on crop or weed regions	Localizes features and increases efficiency
Rotation	Rotations at 90°, 180°, and 270° angles	Increases data size and rotation invariance
Contrast Adjustments	Simulates lighting variations by adjusting contrast	Improves robustness to different lighting
Gaussian Smoothing	Applies a Gaussian blur to reduce noise and enhance primary features	Focuses model on main features, reduces noise

B. Model Training

The prepared CWFID dataset was used to train a Vision Transformer model because of its efficiencies in capturing the spatial relationships within the image data. In contrast to the standard convolutional models, the ViT model adapts a self-attention mechanism, enabling the model to acquire contextual data from larger regions of each picture, which makes it suitable for learning subtle distinctions between crops and weeds. To evaluate the model's performance effectively, the dataset was split into separate training and testing sets. Eighty percent (80%) of the images were allocated to the training set, with the remaining 20% reserved for testing. This split ratio was chosen to ensure that the model could learn robustly from a substantial amount of data while still providing a sufficient amount of unseen data for accurate performance evaluation.

Care was taken to maintain a balanced distribution of crop and weed samples within both sets, allowing the model to be tested on images that represent the diversity and complexity of real-world conditions captured within the CWFID dataset. This split provided the model with an appropriate balance between learning general features during training and evaluating its effectiveness in generalization during testing. To optimize the ViT model for the crop-weed classification task, a set of training parameters was carefully selected based on preliminary testing and validation:

1) Optimizer and learning rate: The function used for optimization was presented by the stochastic gradient descent (SGD) with the learning rate equal to 0.001. It was chosen due to its performances in dealing with large number of sample inputs and the fact that it can converge significantly when trained with appropriate learning rate. The learning rate of 0.001 was found to give a stable and systematic training improvement to the model without oscillating training or causing a convergence problem.

Algorithm: Vision Transformer (ViT) for Crop and Weed Classification

Input:

- Image dataset D with labeled crop, weed, and background images
- Pre-trained Vision Transformer model ViT
- Training parameters: batch size, learning rate, number of epochs

Output:

- Classified images with crops and weeds distinguished

Initialization

1. Load images from dataset D and associated labels (crop, weed, background). 1.2 Apply data transformations to each image:
   - Resize to 224×224= times (ViT input size).
   - Apply random horizontal flip, rotation, and normalization.
2. Define a custom dataset class CropWeedDataset for loading images and labels.
3. Initialize DataLoader for training and validation datasets with the transformed images.
4. Initialize the Vision Transformer model ViT with a classification head suitable for the number of classes
5. Set the loss function as Cross-Entropy Loss
6. For each epoch in the specified number of epochs: - Set the model to training mode.
7. Perform backpropagation and update model weights
8. Perform a forward pass through the model. - Compare predictions to actual labels to calculate accuracy.

2) Batch size: The batch size of 16 was chosen, as such size is more efficient in terms of memory and computation speed. This size ensured the model could handle a reasonable amount of data per every step, the training and convergence process was much smoother and quicker compared to the larger batch sizes, but the memory issues that can come with large batch sizes were also avoided.

3) Epochs: In initial experiments, it was defined that the number of epochs should be 50. This decision was made based on observing the loss and accuracy plots during trial runs of the model several times and noted that 50 epoch was sufficient for the model to learn the features required to distinguish crops from weeds without over-fitting to the training data. Of note, early stopping and validation checks were used to stop training if the model was overfitted or if the training process stagnated, for purposes of time and computational efficiency.

a) Training process: During training, the ViT model took in each 64 × 64 pixel patch derived from the CWFID dataset. The self-attention within the ViT structure allowed the model to learn spatial relationships among these patches thus distinguishing between crop and weed patterns well. Maintaining constant observance of the training and validation loss made it possible to check if the model is overfitting or underfitting. When this training setup was complemented with the well-prepared dataset and the augmentation strategies, the ViT model was able to generalize well. Upon the completion of training, the model was able to learn different patterns and spatial relationship of crops and weeds for a robust classification during the test. Fig. 2 shows ViT architecture.

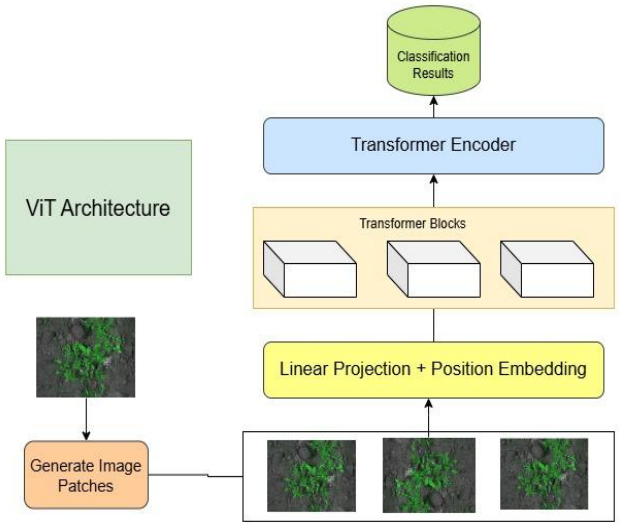


Fig. 2. ViT Architecture.

IV. Experimental Setup

For this study, a systematic experimental framework was developed to assess the ViT approach for crops–weeds discrimination based on the CWFID dataset. This setup entailed setting not only the hardware but also the software environment to address the requirements of processing high-resolution UAV imagery and running deep learning models such that we would obtain reproduceable results.

The experiments were performed on a high-performance computing system consisting of an Intel Xeon E5-2678 v3 processor (2.5GHz), and an NVIDIA GeForce GTX 1080 Ti GPU with 11 GB VRAM. This combination of CPU and GPU allow to process large image datasets effectively and speedup model training. In the framework of the proposed system, it utilized 64GB of DDR4 RAM which make use of the in-memory data processing, especially helpful when dealing with a significant volume of augmented samples. A 1TB SSD was used to store the dataset and the intermediate outputs so that during training and evaluation phase, there was low latency and fast data access.

Regarding the software environment the experiments were performed on Ubuntu 20.04 LTS operating system because of its ability to support deep learning frameworks and successfully manage computationally intensive tasks. PyTorch 1.9.0 has been the major library used to train the ViT model since it offers flexibility to implement transformer models. Furthermore, basic Python libraries including OpenCV for image processing, NumPy for numerical computation and scikit-learn for assessing the performance of the offered models were also installed into the environment. The transformation of the images was made possible by using the Albumentations library towards the

enhancement of the data augmentation processes so as to enhance sample variability.

The dataset preparation for the current study followed the preprocessing steps as outlined in the methodology. The CWFID dataset was further split into a training set and a testing set out of which 80% was used for training and the rest was used for testing The training and testing datasets contained equal numbers of crop and weed samples. This division allowed for providing the model with enough samples for learning the general features at the same time preserving a separate part for the model accuracy assessment of the new, previous samples. Every single high-resolution image was partitioned into 64 × 64 pixel window patches that were centered on crop or weed annotations. This patch extraction enabled the decoupling of complex features with this model to learn localized features while data augmentation including rotation, contrast adjustment and Gaussian smoothing were used to increase the variation in lighting, orientation and appearance of crops and weeds samples.

Some of the parameters of the ViT model were set specifically for the purpose of crop and weed classification. Its architecture was based on the self-attention mechanism and was selected due to its capability to define spatial dependencies within the patches of images successfully. The model was trained with following specifications: batch size = 16, learning rate = 0.001 and the optimizer used for training is stochastic gradient descent. The total training process comprised 50 epochs, if validation loss stopped increasing or began to rise, early stopping was used to stop training. The cross entropy loss was adopted as the main loss function, which provides flexibility in multi-class crop, weed, and background classification.

To assess the performance of the developed models, a variety of evaluation measures was used. The accuracy for each of the classes, namely the crop, weed and background were determined in order to compare the performance of the models. With accuracy and quantity, measures of precision and recall were useful in determining the strengths of the model in differentiating crops from weeds and an F1 score was useful as it balances both false positives and negatives. Further, to avoid or reduce such biases confusion matrices were produced that give a clear distinction of the model on class to class basis.

V. Results

In the results section, the performance of the Vision Transformer (ViT) model on crops, weeds, and background elements of the CWFID dataset is described in detail. Essentially, percentage accuracy, precision, recall and F1 scores were determined and more detailed analysis was done using the confusion matrix. The model was trained using 80 / 20 train-test split which helped evaluate the model on the new data it has never seen.

A. Accuracy Assessment

On the test set it was possible to obtain an overall accuracy of the ViT model equal to 89.4% showing that it can effectively distinguish crop, weed and background pixels. This high level of accuracy indicate that the ViT model is able to extract the unique features of each class even in the complicated agricultural environments where crops resemble weeds. The degree of accuracy shown in this paper proves that ViT model can be used in practical applications, specifically in the field of precision agriculture where precise identification of crops and weeds can lead to improvement in crop management and decrease in the amount of applied herbicide.

B. Class-Specific Performance

Class-specific precision, recall, and F1 scores were calculated to evaluate the model's effectiveness across different classes: crops, weeds, and background. These metrics are as follows and are summarised in Table III for easy comparison of the strengths and weaknesses of the model with respect to each class. Fig. 3 shows various models performance results.

Fig. 3. Various models performance results.

TABLE III. Specific Performance Metrics of ViT Model

Class	Precision (%)	Recall (%)	F1 Score (%)
Crops	91.2	85.7	88.4
Weeds	87.5	80.3	83.7
Background	93.1	96.4	94.7

For the crop detection, the model obtained an accuracy of 91.2% and recall of 85.7% and thus an F1 score of 88.4%. This is, however, high, although there could be confusion with weeds particularly in the inter-row area. Varying results were achieved for the recognition of weeds, with 87.5% accuracy, 80.3% recall, and thus an F1 of 83.7%. The slightly lower recall for weeds shows that weed detection is more difficult especially for intra row weeds which are more similar in appearance to the crops. In the evaluation of the model for background region, the precision achieved was 93.1%, with a recall of 96.4% and F1 score 94.7%. This high performance on the background further enhances the performance of the model in differentiating the non-vegetation areas, thus minimizing chances of wrongly classifying crops as weeds.

C. Confusion Matrix Analysis

The confusion matrix extends the assessment of the model's classification correctness by showing where the errors were made. In Table IV the true positive, the false positive, and the false negative are shown for each class.

The confusion matrix (Fig. 4) also shows that the major misclassification problem was between the crop and weed classes where crop pixels amounted to 168 were misclassified as weeds while weed pixels of 128 were classified as crops. This pattern indicates that the yarn becomes problematic in distinguishing between crops and weeds mainly within areas of high plant density. This is especially problematic in intra-row spaces where weeds and crops may have similar architectures and reflectance properties hence compounding the challenge of modeling the two. On the other hand, the background class was accurately classified with few errors which actually shows that the model is good in separating vegetative from the non-vegetative land cover like the soiler bare ground.

TABLE IV. CONFUSION MATRIX FOR VIT MODEL PREDICTIONS ON TEST SET

	Predicted Crop	Predicted Weed	Predicted Background
Actual Crop	1,221	128	13
Actual Weed	168	1,345	72
Actual Background	7	12	1,249

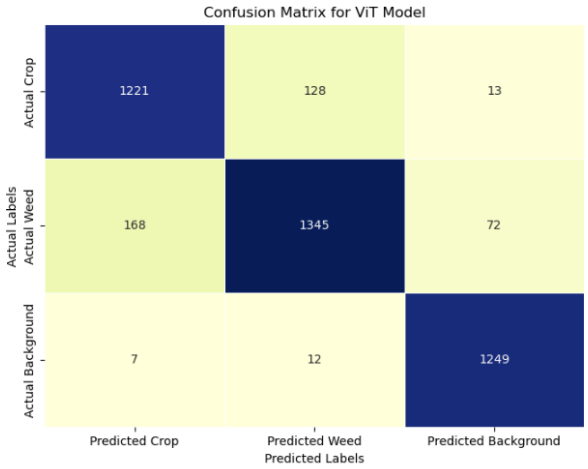


Fig. 4. Confusion matrix.

D. Comparison with Other Models

The ViT model was compared with traditional models such as U-Net, SegNet, and Fully Convolutional Network (FCN). Accuracy, precision, recall and F1 scores of all models are summarized in the Table V, where it can be concluded that the ViT model is more accurate. By comparing the proposed Vision Transformer (ViT) model with those of U-Net, SegNet, and Fully Convolutional Network (FCN), its higher accuracy has established it to be capable of handling the difficult environments within agriculture, especially the growth within the intra-row weed.

In such environments, where weeds are below or adjacent to crops and may be morphologically similar to crops, many of the CNN-based models are ineffective. This is due to the fact that, convolutional layers are inherently limited by its local receptive field, meaning that traditional model might be unable to capture those high-level, global features requiring the understanding of the whole image and its relationship to all other images, which in turn affects its accuracy in situations where high level of discriminative dissimilarities exists.

The self-attention mechanism of the ViT model has an advantage because it processes images in their entirety and identifies long-range spatial relations that may be neglected by CNN-based architectures. Such an approach is most beneficial for intra-row weed identification, in which local resemblance in texture and color between crops and weeds often leads to confusion in other models. This paper also shows that self-attention mechanism in ViT that allows the model to pay attention to relevant features in large regions of the images leads to better recall and precision, important for weed classification where precise distinction between crop and weed pixels is necessary (see Fig. 5, 6 and 7).

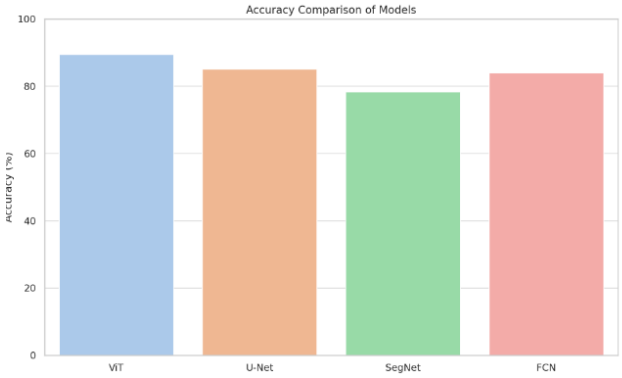


Fig. 5. Accuracy comparisions of models.

Fig. 6. F1-Score comparisions of models.

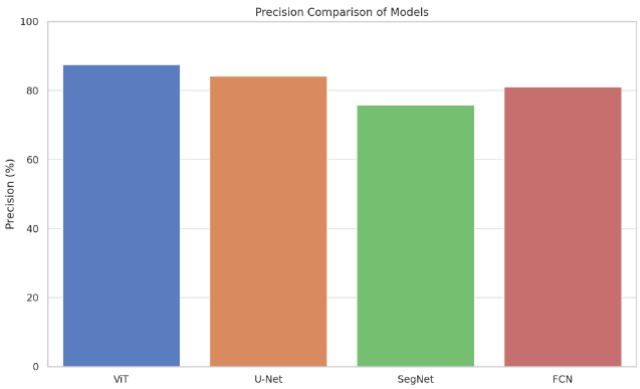


Fig. 7. Precision comparisions of models.

TABLE V. PERFORMANCE COMPARISON OF VIT MODEL WITH TRADITIONAL MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
ViT	89.4	87.5	80.3	83.7
U-Net	85.1	84.2	76.5	80.2
SegNet	78.3	75.8	71.4	73.5
FCN	83.9	81.1	74.8	77.8

Furthermore, the improved performance in ViT can again be attributed to how generalization is done properly to the field conditions. The high accuracy and precision in signifying weeds and crop images justify the flexibility of the proposed model in different lighting backgrounds, soil, and crop placement. This generalization is especially helpful when working with practical field applications of the model as weather, lighting, and growth stages may affect the models' performance. A noteworthy comparison with U-Net further emphasizes ViT's advantage: that although U-Net also yields a good performance, the reliance on convolutional layers again hinders the ability to capture global context and therefore yields lower recall for weed detection. The results indicate that the trained U-Net is more sensitive to vagaries in closely planted weeds and crops, issues that are worsened by field conditions. However, these difficulties are not present in ViT's design, which implies that potential agricultural applications of transformer-based models might be more scalable and versatile where extensive field analysis is required.

VI. CONCLUSION

The current paper shows that Vision Transformers (ViTs) can be used in precision agriculture for the detection of weeds in crop fields. This accuracy was established through pixel-level classification adopted from high-resolution UAV imagery as compared to traditional models such as U-Net and FCN where the ViT model obtained 89.4% accuracy. The high accuracy is an indication of ViT's ability to establish dependencies and spatial arrangement in large agricultural scenes, which are hard for traditional CNNs to achieve. The study achieved the following goals of the research: The class imbalance was solved by applying a combination of two oversampling techniques which improved the classification results. Employing patch extraction and data augmentation enabled the ViT model to accurately distinguish crop, weed and background regions. The approach also showed robustness under different conditions improving the likelihood of its application in realistic agricultural settings.

This research goes a long way in the promotion of sustainable agriculture by providing a potential method for selective weed management that does not require much use of the weed controlling herbicide. The current study could be extended in the future by examining other environmental factors or using the model in other crop types, and different field conditions to assess the model's universality. Finally, the model derived from ViT holds the potential to contribute toward precise, effective and sustainable farming.

This research also opens up possibilities for integration with AI-based tractors, enabling real-time weed detection and management directly in the field. Such applications could revolutionize automated precision agriculture, allowing for targeted weed control while minimizing herbicide usage. With further development, this approach could support the advancement of intelligent, autonomous farming machinery.

REFERENCES

[1] S. Nath, "A vision of precision agriculture: Balance between agricultural sustainability and environmental stewardship," Agronomy Journal, vol. 116, no. 3, pp. 1126-1143, 2024.

[2] L. De Bortoli, S. Marsi, F. Marinello, and P. Gallina, "Cost-efficient algorithm for autonomous cultivators: Implementing template matching with field digital twins for precision agriculture," Computers and Electronics in Agriculture, vol. 227, p. 109509, 2024.

[3] D. C. Brainard et al., "A survey of weed research priorities: key findings and future directions," Weed Science, vol. 71, no. 4, pp. 330-343, 2023.

[4] M. Vasileiou et al., "Transforming weed management in sustainable agriculture with artificial intelligence: A systematic literature review towards weed identification and deep learning," Crop Protection, p. 106522, 2023.

[5] F. Yeganehpoor, S. Z. Salmasi, G. Abedi, F. Samadiyan, and V. Beyginiya, "Effects of cover crops and weed management on corn yield," Journal of the Saudi Society of Agricultural Sciences, vol. 14, no. 2, pp. 178-181, 2015.

[6] P. Hatcher and B. Melander, "Combining physical, cultural and biological methods: prospects for integrated non - chemical weed management strategies," Weed research, vol. 43, no. 5, pp. 303-322, 2003.

[7] I. Bhakta, S. Phadikar, and K. Majumder, "State - of - the - art technologies in precision agriculture: a systematic review," Journal of the Science of Food and Agriculture, vol. 99, no. 11, pp. 4878-4888, 2019.

[8] N. Zhang, M. Wang, and N. Wang, "Precision agriculture—a worldwide overview," Computers and electronics in agriculture, vol. 36, no. 2-3, pp. 113-132, 2002.

[9] D. C. Tsouros, S. Bibi, and P. G. Sarigiannidis, "A review on UAV-based applications for precision agriculture," Information, vol. 10, no. 11, p. 349, 2019.

[10] A. Upadhyay et al., "Advances in ground robotic technologies for site-specific weed management in precision agriculture: A review," Computers and Electronics in Agriculture, vol. 225, p. 109363, 2024.

[11] A. H. Al-Badri et al., "Classification of weed using machine learning techniques: a review—challenges, current and future potential techniques," Journal of Plant Diseases and Protection, vol. 129, no. 4, pp. 745-768, 2022.

[12] F. D. Adhinata and R. Sumiharto, "A comprehensive survey on weed and crop classification using machine learning and deep learning," Artificial Intelligence in Agriculture, 2024.

[13] R. Reedha, E. Dericquebourg, R. Canals, and A. Hafiane, "Transformer neural network for weed and crop classification of high resolution UAV images," Remote Sensing, vol. 14, no. 3, p. 592, 2022.

[14] S. Sharma and M. Vardhan, "Self-attention vision transformer with transfer learning for efficient crops and weeds classification," in 2023 6th International Conference on Information Systems and Computer Networks (ISCON), 2023: IEEE, pp. 1-6.

[15] T. Burks, S. Shearer, and F. Payne, "Classification of weed species using color texture features and discriminant analysis," Transactions of the ASAE, vol. 43, no. 2, pp. 441-448, 2000.

[16] Z. Wu, Y. Chen, B. Zhao, X. Kang, and Y. Ding, "Review of weed detection methods based on computer vision," Sensors, vol. 21, no. 11, p. 3647, 2021.

[17] M. N. Khan and S. Anwar, "Robust weed recognition through color based image segmentation and convolution neural network based classification," in ASME International Mechanical Engineering Congress and Exposition, 2019, vol. 59414: American Society of Mechanical Engineers, p. V004T05A045.

[18] S. Lavania and P. S. Matey, "Novel method for weed classification in maize field using Otsu and PCA implementation," in 2015 IEEE International Conference on Computational Intelligence & Communication Technology, 2015: IEEE, pp. 534-537.

[19] H. Huang, Y. Lan, A. Yang, Y. Zhang, S. Wen, and J. Deng, "Deep learning versus Object-based Image Analysis (OBIA) in weed mapping of UAV imagery," International Journal of Remote Sensing, vol. 41, no. 9, pp. 3446-3479, 2020.

[20] M. H. M. Zaman, S. M. Mustaza, M. F. Ibrahim, M. A. Zulkifley, and M. M. Mustafa, "Weed classification based on statistical features from Gabor transform magnitude," in 2021 International Conference on Decision Aid Sciences and Application (DASA), 2021: IEEE, pp. 147-151.

[21] P. Bosilj, T. Duckett, and G. Cielniak, "Analysis of morphology-based features for classification of crop and weeds in precision agriculture," IEEE Robotics and Automation Letters, vol. 3, no. 4, pp. 2950-2956, 2018.

[22] P. Rasti, A. Ahmad, S. Samiei, E. Belin, and D. Rousseau, "Supervised image classification by scattering transform with application to weed detection in culture crops of high density," Remote Sensing, vol. 11, no. 3, p. 249, 2019.

[23] H. Okamoto, T. Murata, T. Kataoka, and S. I. HATA, "Plant classification for weed detection using hyperspectral imaging with wavelet analysis," Weed biology and Management, vol. 7, no. 1, pp. 31-37, 2007.

[24] C. Feng et al., "A combination of OBIA and random forest based on visible UAV remote sensing for accurately extracted information about weeds in areas with different weed densities in farmland," Remote Sensing, vol. 15, no. 19, p. 4696, 2023.

[25] C.-C. Andrea, B. B. M. Daniel, and J. B. J. Misael, "Precise weed and maize classification through convolutional neuronal networks," in 2017 IEEE Second zcuador Technical Chapters Meeting (ETCM), 2017: IEEE, pp. 1-6.

[26] M. Fawakherji, C. Potena, A. Pretto, D. D. Bloisi, and D. Nardi, "Multi-spectral image synthesis for crop/weed segmentation in precision farming," Robotics and Autonomous Systems, vol. 146, p. 103861, 2021.

The Heart of Artificial Intelligence: A Review of Machine Learning for Heart Disease Prediction

Brayan R. Neciosup-Bolaños[1], Segundo E. Cieza-Mostacero[2]
Research Group Trend and Innovation in Systems Engineering -Trujillo, Cesar Vallejo University, Perú[1, 2]

***Abstract*—Heart disease is one of the main heart diseases that cause the death of people worldwide, affecting the engine of the human body: the heart. It has a greater incidence in underdeveloped countries such as Angola, Bangladesh, Ethiopia and Haiti, for this reason, obtaining accurate results based on risk factors manually is a complex task. Therefore, this systematic review allowed us to analyze and study 32 articles applying the PRISMA methodology, which allowed us to evaluate the suitability of the methods and, consequently, their reliability in the results. The results of the study showed that the algorithm with the greatest accuracy in predicting these heart diseases is Random Forest. The most commonly used metrics to evaluate machine learning algorithms are sensitivity, F1 score, precision, and accuracy, with sensitivity highlighted as the primary metric. The most predominant independent aspects for predicting heart disease in machine learning models are age, sex, cholesterol, diabetes, and chest pain. Finally, the most used data distribution is 70% for training and 30% for testing, which achieves great accuracy in the algorithm prediction process. This study offers a promising path for the prevention and timely treatment of this disease through the use of machine learning algorithms. In the future, these advances could be applied in a system accessible to all people, thus improving access to healthcare and saving lives.**

Keywords*—*Machine learning; heart disease; prediction; systematic review; artificial intelligence; algorithms; literature; heart

I. INTRODUCTION

One of the main causes of death worldwide is heart disease, which includes conditions such as coronary arteries, angina, heart attacks and heart disease, resulting from problems that affect the engine of the human body: the heart [1]. In addition, the American College of Cardiology mentions that 26 million people die from heart disease worldwide, and 3.6 million people undergo tests to rule out these diseases, aware of the great impact they can have on their lives [2].

Heart disease has a greater presence in underdeveloped countries such as: Angola, Bangladesh, Ethiopia and Haiti, with risk factors associated with this disease such as: high blood pressure, high cholesterol, uncontrolled diabetes, smoking and cardiac deterioration [3]. Therefore, to obtain accurate results in the diagnosis of this condition, a decision support system for your specialists is needed, since relying on multiple risk factors manually is a complex task [2].

Along with the challenges these nations face, the need for innovative solutions such as machine learning is highlighted, in addition, the adoption and application of this branch of artificial intelligence in the prediction of heart disease offers a promising approach. Compared to a human expert, machine learning models stand out for their speed and the lower cost associated with the predictions of these pathologies [4].

Machine learning, being a method of developing algorithms that help diagnose diseases of various kinds, has been crucial in a constantly modernizing world, where technology plays a vital role in continuous development. Through its techniques, it has been possible to save the lives of thousands of people by quickly detecting or predicting diseases, thus offering high-quality service to patients. Likewise, by identifying the primary phases of the conditions mentioned above, treatments can be adopted and counteract the disease, controlling the mortality rate in a comprehensive manner [5], [6].

On the other hand, machine learning models are classified into three categories: supervised learning algorithms, which focus on providing the user with an input x along with its corresponding output y, with the purpose of predicting y for a previously unseen input x, through the development of a classifier algorithm; Unsupervised learning algorithms, which do not focus on providing specific output values, but rather on inferring an underlying structure from the inputs, and reinforcement learning algorithms, where an agent is trained to determine certain policies, to solve efficient problems [7], [8].

Therefore, this systematic review focuses on the study of the most effective machine learning algorithms; powerful tools to make medical diagnoses and effective health services, revolutionizing the health sector. In addition, health professionals will be trained to identify assistance solutions faster and with greater accuracy [9]. Therefore, these algorithms, with their characteristics, make it possible to predict heart disease in people, which allows us to obtain an advantage against the disease.

Likewise, to carry out this study, a systematic review of the literature was carried out using the PRISMA methodology. In this methodology, articles were selected to address four research questions: Which machine learning algorithm demonstrated the best prediction performance in the present studies? What are the independent aspects for the machine learning model in its prediction process? ?, What performance metrics were used to evaluate the machine learning model(s) in the present studies? and What is the proportion of data used to train and test the machine learning model? Furthermore, the review was structured into sections of introduction, methodology, analysis of results and conclusions.

II. METHODOLOGY

To develop the systematic review, we applied the PRISMA 2020 methodology, allowing us to evaluate the adequacy of the methods and, consequently, the reliability of the results [10]. Additionally, Zotero software was used to store the articles, these documents were subsequently evaluated to determine their eligibility, following established criteria. Fig. 1 shows study selection flowchart.

A. Research Questions

The objective of this study is to examine, compare and summarize articles on heart disease prediction using machine learning, published from 2021 to 2022. The four research questions developed are as follows:

- Q1: Which machine learning algorithm demonstrated the best prediction performance in the present studies?
- Q2: What are the independent aspects for the machine learning model in its prediction process?
- Q3: What performance metrics were used to evaluate the machine learning model(s) in the present studies?
- Q4: What is the proportion of data used to train and test the machine learning model?

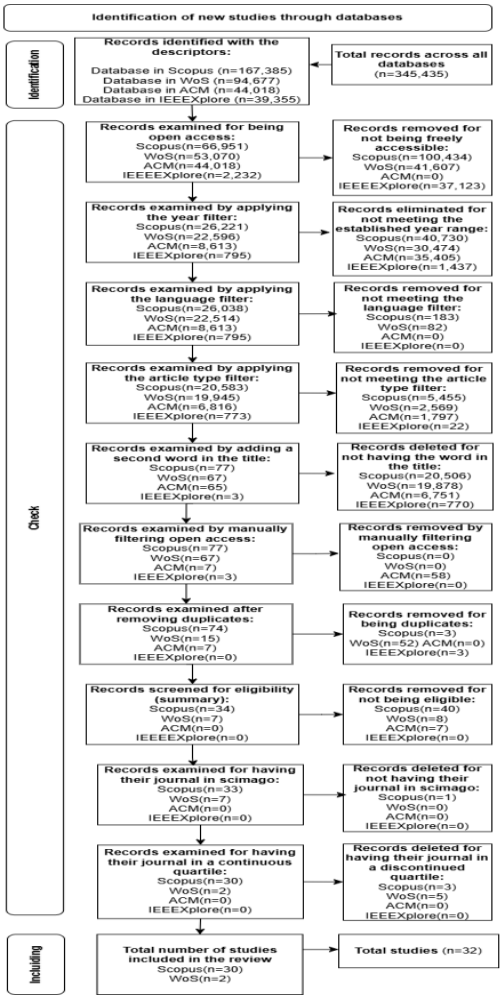


Fig. 1. Study selection flowchart.

B. Search Strategies

An exhaustive search of publications ranging from 2021 to 2022 has been carried out in four databases, these being: Scopus, Web of Science, ACM and IEEE Xplore. Likewise, to carry out this search strategy, the keywords were used: (“machine learning”) and (“heart disease”), as shown in Fig. 2.

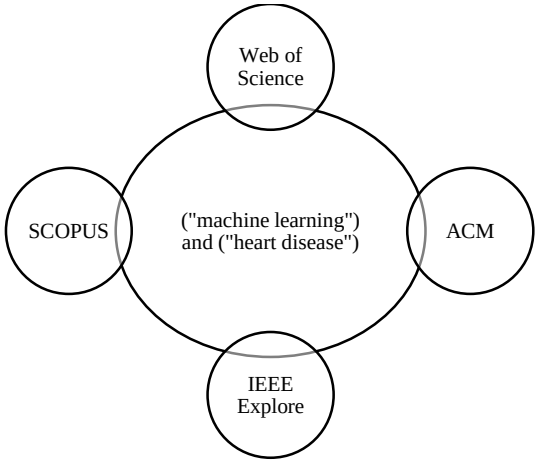


Fig. 2. Search criteria in databases.

C. Inclusion and Exclusion Criteria

For the selection of studies, studies were included that 1) the title of the articles must include the keywords, 2) they must be available as open access, 3) they must have been published between 2021 and 2022, 4) they must be written in English, 5) the type of document must be exclusively articles, 6) manual filter to access open access articles through Zotero, 7) they must have been non-duplicated articles, 8) they must have been eligible (summary), 9) must have had their journals indexed in scimago and 10) must have been from journals with non-discontinued quartiles. The exclusion criteria were: 1) the title of the articles does not include the keywords, 2) they are not available in open access, 3) they are not published between 2021 and 2022, 4) they are written in a language other than English. , 5) that the type of document does not correspond to an article, 6) they are not freely accessible through Zotero, 7) duplicate documents, 8) ineligible documents (abstract), 9) the journals of the documents are not indexed in scimago and 10) the magazines of the documents are of discontinued quartile. Fig. 2 presents the characteristics of the articles included in the systematic review, where the previously mentioned criteria are shown.

III. RESULTS AND DISCUSSION

Of the 32 studies identified, they were compiled and summarized in a Microsoft Excel spreadsheet. The distribution of these studies according to their origin in the database is as follows: 94% of the studies come from Scopus, which is equivalent to a total of 30 articles. On the other hand, Web of Science (WOS) represents 6% of the studies, with two articles. Both ACM and IEEE Xplore do not have any studies that meet the inclusion criteria mentioned above. The details of the distribution of the selected articles according to their origin in the database are summarized in Table I.

TABLE I. SELECTED ARTICLES

N°	Database	Number of Articles	Percentage (%)
1	Scopus	30	94
2	WOS	2	6
3	ACM	0	0
4	IEEEXplore	0	0
	Total	32	100

a. Source: Own work

A. Results of Machine Learning Algorithm with Higher Accuracy (Q1)

Among the 32 studies analyzed, the presence of several machine learning algorithms for the prediction of diseases related to heart disease was observed. The most common algorithms include AdaBoost, CatBoost, Decision Tree, KNN, Linear Regression, Logistical Regression, Naive Bayes, Random Forest, Support Vector Machine, and XGBoost. Furthermore, models proposed by their authors were included in two studies, such as: HB + ET + SMOTE [11] y RECHOMMEND [12] highlighting a different approach that contributes to the field of machine learning.

TABLE II. ALGORITHM WITH BEST PERFORMANCE

N°	Algorithms	Papers	#	%
1	Random Forest	[13], [4], [2], [14], [1], [15], [16], [17], [18], [19], [20], [21], [3]	13	30.95
2	Support Vector Machine	[22], [23], [24], [25], [26], [9], [18]	7	16.67
3	XGBoost	[27], [28], [29], [19], [30]	5	11.90
4	Decision Tree	[5], [2], [31], [32]	4	9.52
5	Naive Bayes	[6], [33], [1], [18]	4	9.52
6	KNN	[34], [20]	2	4.76
7	Logistical Regression	[31], [18]	2	4.76
8	AdaBoost	[35]	1	2.38
9	CatBoost	[19]	1	2.38
10	HB + ET + SMOTE	[11]	1	2.38
11	Linear Regression	[21]	1	2.38
12	Rechommend	[12]	1	2.38

b. Source: Own work

Table II shows that the Random Forest algorithm is the most predominant in a total of 13 studies, which represents 30.95% of articles that use it. These findings are of utmost importance for future research that seeks to determine which machine learning algorithm provides the best performance in predicting diseases related to heart disease.

Thus, in the study by Maini et al. [15], the Random Forest (RF) algorithm achieved a diagnostic accuracy of 93.8%, evidencing a greater predictive capacity compared to other algorithms evaluated; concluding that the RF-based machine learning model not only offers an early diagnosis of heart diseases, but can also be easily accessible through the Internet, facilitating its implementation in clinical settings.

Likewise, an accurate ML model not only contributes to the early prediction of heart disease, but also allows identifying cases in which, although the patient appears to be healthy, the disease could be progressing imperceptibly. Therefore, the use of machine learning algorithms to prevent and treat heart disease in a timely manner is relevant.

B. Result of Independent Aspects for the Prediction Process of the Machine Learning Model (Q2)

Fig. 3 presents an analysis of the independent aspects or risk factors used in the prediction process of machine learning models.

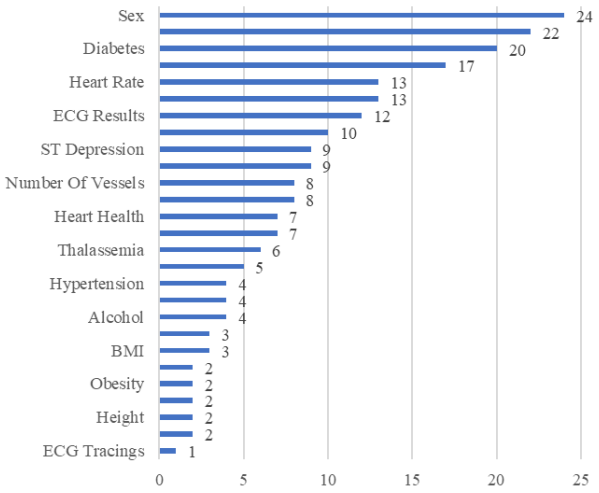


Fig. 3. Independent aspects.

Table III shows that age, sex, cholesterol levels, the presence of diabetes and chest pain are the predominant factors in the studies. These risk factors are considered highly effective in predicting heart disease-related diseases in patients, as stated and recommended by experts in the field [33].

TABLE III. PREDOMINANT INDEPENDENT ASPECTS

N°	Independent Aspect	Articles	#	%
1	Age	[1], [3], [4], [5], [9], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [24], [25], [26], [27], [28], [30], [31], [32], [33], [34], [35]	27	10.15
2	Sex	[2], [3], [5], [9], [12], [13], [14], [16], [20], [24], [28], [31], [32], [33], [34], [35], [15], [17], [18], [21], [26], [27], [14], [19]	24	9.02
3	Cholesterol	[1], [2], [4], [9], [13], [14], [16], [16], [17], [18], [20], [21], [22], [25], [26], [28], [30], [31], [32], [33], [34], [35]	22	8.27
4	Diabetes	[2], [3], [9], [14], [15], [16], [17], [19], [20], [21], [24], [27], [31], [32], [33], [34], [1], [4], [18], [26]	20	7.52
5	Chest pain	[1], [2], [3], [4], [5], [9], [13], [15], [18], [21], [22], [29], [31], [32], [35], [14], [24]	17	6.39

c. Source: Own work

Khdair and Dasari [25], in their study, highlighted that characteristics of medical records and associated risk factors such as: tobacco, LDL cholesterol levels, systolic blood pressure, adiposity and family history play a crucial role in preventing heart disease, as measured by medical records. They further noted that the success of predictions in the field of machine learning largely depends on the quality and diversity of the data used, as richness in data features and variables significantly improves the results in machine learning predictive models.

C. Result of Performance Metrics for Evaluation of Machine Learning Models (Q3)

Among the most commonly used performance metrics to evaluate machine learning algorithms, sensitivity, F1 score, precision, and accuracy stand out. As detailed in Fig. 4, sensitivity is the most valued metric in the studies, with 19.04%.

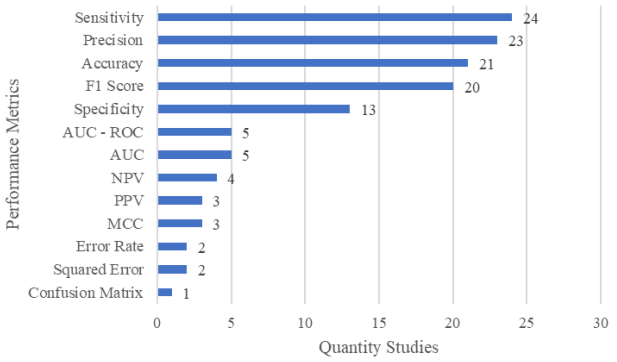


Fig. 4. Performance metrics.

In most studies that evaluate the performance of their machine learning algorithms, it exceeds 90%. As demonstrated by Maini et al. [15] where its RF prediction system achieved a sensitivity of 92.8% in the effective diagnosis of heart diseases. Another study reveals that an SVM-based model achieved 95% sensitivity [9].

Thus, in the research of Alotaibi and Alzahrani [35], they also use accuracy, sensitivity, specificity, F1 score, error rate, and Matthews correlation coefficient (MCC), all of these metrics derived from arithmetic calculations on the rate of true positives, false positives, false negatives, true negative rate, and false negatives, as performance metrics to evaluate their machine learning model and make more effective decisions about its effectiveness and identify areas for improvement.

D. Data Distribution Results for Training and Testing Machine Learning Models (Q4)

Regarding the distribution of data in the 32 articles, a proportion of data used for training (T) and testing (P) the machine learning models has been considered. Table IV shows the percentages of articles that do or do not provide information about the distribution of their data.

TABLE IV. DATA DISTRIBUTION

N°	Data Distribution	Articles	#	%
1	T: 70% - P: 30%	[22], [5], [35], [14], [28], [1], [15], [20]	8	25.00
2	T: 80% - P: 20%	[27], [17], [18], [21], [3], [20]	6	18.75
3	T: 75% - P: 25%	[11], [24], [9], [16]	4	12.5
4	T: 30% - P: 70%	[2]	1	3.12
5	T: 50% - P: 50%	[33]	1	3.12
6	T: 60% - P: 40%	[34]	1	3.12
7	T: 85% - P: 15%	[30]	1	3.12
8	Does not display information	[4], [6], [13], [19], [23], [25], [26], [29], [31], [32]	10	31.25

d. Source: Own work

After detailed analysis, it was observed that 25% of the studies used a data distribution in the proportion of 70% for training and 30% for testing. In the study by Absar et al. [20], two data distributions were used: one in a proportion of 80% for training (E) and 20% for testing (P), and another in a proportion of 70% for training (E) and 30% for testing (P). The 7:3 ratio allowed machine learning algorithms such as Random Forest, AdaBoost, and Decision Tree to demonstrate 99.025%, 96.103%, and 100% accuracy, respectively.

IV. LIMITATIONS

During the research, various limitations were presented. Firstly, the constant evolution of the literature implied that the information available at the time of the review may have changed, which represented a continuous challenge to keep the review updated with new developments in research.

Furthermore, the time available for the development of the systematic review was limited, which prevented the exploration

of information in other databases that could have complemented the review.

Finally, there was a limitation in access to certain publications with relevant information, due to restrictions imposed by the authors, which prevented the inclusion of some significant studies in this research.

V. CONCLUSIONS

This systematic review, based on the analysis of 32 studies, has provided a comprehensive view of the impact that machine learning algorithms generate in the prediction of heart disease. Through the review, four key questions were addressed that framed the analysis, allowing us to identify common patterns and significant differences between the studies. Regarding the first question on Which machine learning algorithm demonstrated the best performance in the prediction of the present studies?, this study determined that Random Forest offers an early diagnosis of heart disease and can be easily accessible through the Internet, facilitating its implementation in clinical settings.

Regarding the second question on What are the independent aspects for the machine learning model in its prediction process?, it was determined that various independent aspects or risk factors were addressed, among which age, sex, cholesterol, diabetes and chest pain stand out, being the most predominant among the studies analyzed. Regarding the third question on What performance metrics were used to evaluate the machine learning model(s) in the present studies?, it was determined that various performance metrics are used, among the most used are: precision, sensitivity, accuracy and F1 score.

Finally, regarding the last question on What is the proportion of data used to train and test the machine learning model?, it was found that the distribution of data for training and validation are distributed in various proportions, such as: 70% - 30%, 80% - 20%, 75% - 25%, 30% - 70%, 50% - 50%, 60% - 40% and 85% - 15% for training and testing respectively, with the proportion of 70%-30% standing out as one of the most used. In this way, future research could integrate all the capabilities of machine learning into a system accessible to all people. This would allow for significant progress in the prevention of heart disease, making a decisive contribution to saving millions of lives.

ACKNOWLEDGMENT

The research was not externally funded. We would like to express our sincere gratitude to all those who have contributed to this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] J. Rashid, S. Kanwal, J. Kim, M. W. Nisar, U. Naseem, y A. Hussain, «Heart disease diagnosis using the brute force algorithm and machine learning techniques», Computers, Materials and Continua, vol. 72, n.o 2. Tech Science Press, pp. 3195-3211, 2022. doi: 10.32604/cmc.2022.026064.

[2] G. N. Ahmad, S. Ullah, A. Algethami, H. Fatima, y S. Md. H. Akhter, «Comparative Study of Optimum Medical Diagnosis of Human Heart Disease Using Machine Learning Technique with and Without Sequential Feature Selection», IEEE Access, vol. 10. Institute of Electrical and Electronics Engineers Inc., pp. 23808-23828, 2022. doi: 10.1109/ACCESS.2022.3153047.

[3] O. Sami, Y. Elsheikh, y F. Almasalha, «The Role of Data Pre-processing Techniques in Improving Machine Learning Accuracy for Predicting Coronary Heart Disease», International Journal of Advanced Computer Science and Applications, vol. 12, n.o 6. Science and Information Organization, pp. 816-824, 2021. doi: 10.14569/IJACSA.2021.0120695.

[4] T. A. Assegie, P. K. Rangarajan, N. K. Kumar, y D. Vigneswari, «An empirical study on machine learning algorithms for heart disease prediction», IAES International Journal of Artificial Intelligence, vol. 11, n.o 3. Institute of Advanced Engineering and Science, pp. 1066-1073, 2022. doi: 10.11591/ijai.v11.i3.pp1066-1073.

[5] S. Molla et al., «A predictive analysis framework of heart disease using machine learning approaches», Bulletin of Electrical Engineering and Informatics, vol. 11, n.o 5. Institute of Advanced Engineering and Science, pp. 2705-2716, 2022. doi: 10.11591/eei.v11i5.3942.

[6] S. M. H. S. Iqbal, N. Jahan, A. S. Moni, y M. Khatun, «An Effective Analytics and Performance Measurement of Different Machine Learning Algorithms for Predicting Heart Disease», International Journal of Advanced Computer Science and Applications, vol. 13, n.o 2. Science and Information Organization, pp. 429-433, 2022. doi: 10.14569/IJACSA.2022.0130250.

[7] J. Van y H. Hoos, «A survey on semi-supervised learning», Mach. Learn., vol. 109, n.o 2, pp. 373-440, 2020, doi: 10.1007/s10994-019-05855-6.

[8] A. Sharma, R. Gupta, K. Lakshmanan, y A. Gupta, «Transition Based Discount Factor for Model Free Algorithms in Reinforcement Learning», SYMMETRY-BASEL, vol. 13, n.o 7, p. 1197, jul. 2021, doi: https://doi.org/10.3390/sym13071197.

[9] J. Meng y R. Xing, «Inside the "black box": Embedding clinical knowledge in data-driven machine learning for heart disease diagnosis», CARDIOVASCULAR DIGITAL HEALTH JOURNAL, vol. 3, n.o 6. ELSEVIER, RADARWEG 29, 1043 NX AMSTERDAM, NETHERLANDS, pp. 276-288, diciembre de 2022. doi: 10.1016/j.cvdhj.2022.10.005.

[10] M. J. Page et al., «Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas», Rev. Esp. Cardiol., vol. 74, n.o 9, pp. 790-799, 2021, doi: 10.1016/j.recesp.2021.06.016.

[11] A. Abdellatif, H. Abdellatef, J. Kanesan, C.-O. Chow, J. H. Chuah, y H. M. Gheni, «An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods», IEEE Access, vol. 10. Institute of Electrical and Electronics Engineers Inc., pp. 79974-79985, 2022. doi: 10.1109/ACCESS.2022.3191669.

[12] A. E. Ulloa et al., «Rechommend: An ECG-Based Machine Learning Approach for Identifying Patients at Increased Risk of Undiagnosed Structural Heart Disease Detectable by Echocardiography», Circulation, vol. 146, n.o 1. Lippincott Williams and Wilkins, pp. 36-47, 2022. doi: 10.1161/CIRCULATIONAHA.121.057869.

[13] V. Chang, V. R. Bhavani, A. Q. Xu, y M. A. Hossain, «An artificial intelligence model for heart disease detection using machine learning algorithms», Healthcare Analytics, vol. 2. Elsevier Inc., 2022. doi: 10.1016/j.health.2022.100016.

[14] Ch. A. ul Hassan et al., «Effectively Predicting the Presence of Coronary Heart Disease Using Machine Learning Classifiers», Sensors, vol. 22, n.o 19. MDPI, 2022. doi: https://doi.org/10.3390/s22197227.

[15] E. Maini, B. Venkateswarlu, B. Maini, y D. Marwaha, «Machine learning–based heart disease prediction system for Indian population: An exploratory study done in South India», Medical Journal Armed Forces India, vol. 77, n.o 3. Elsevier B.V., pp. 302-311, 2021. doi: https://doi.org/10.1016/j.mjafi.2020.10.013.

[16] G. N. Ahmad et al., «Mixed Machine Learning Approach for Efficient Prediction of Human Heart Disease by Identifying the Numerical and Categorical Features», Applied Sciences (Switzerland), vol. 12, n.o 15. MDPI, 2022. doi: 10.3390/app12157449.

[17] A. Li, J. M. Roveda, L. S. Powers, y S. F. Quan, «Obstructive sleep apnea predicts 10-year cardiovascular disease-related mortality in the Sleep Heart Health Study: a machine learning approach», JOURNAL OF

CLINICAL SLEEP MEDICINE, vol. 18, n.o 2. AMER ACAD SLEEP MEDICINE, 2510 N FRONTAGE RD, DARIEN, IL 60561 USA, pp. 497-504, 1 de febrero de 2022. doi: 10.5664/jcsm.9630.

[18] A. K. Dubey, K. Choudhary, y R. Sharma, «Predicting heart disease based on influential features with machine learning», Intelligent Automation and Soft Computing, vol. 30, n.o 3. Tech Science Press, pp. 929-943, 2021. doi: 10.32604/iasc.2021.018382.

[19] S. Ahmed et al., «Prediction of Cardiovascular Disease on Self-Augmented Datasets of Heart Patients Using Multiple Machine Learning Models», Journal of Sensors, vol. 2022. Hindawi Limited, 2022. doi: 10.1155/2022/3730303.

[20] N. Absar et al., «The Efficacy of Machine-Learning-Supported Smart System for Heart Disease Prediction», Healthcare (Switzerland), vol. 10, n.o 6. MDPI, 2022. doi: https://doi.org/10.3390/healthcare10061137.

[21] S. Simon et al., «The Impact of Time Horizon on Classification Accuracy: Application of Machine Learning to Prediction of Incident Coronary Heart Disease», JMIR Cardio, vol. 6, n.o 2. JMIR Publications Inc., 2022. doi: 10.2196/38040.

[22] T. Suresh, T. A. Assegie, S. Rajkumar, y N. K. Kumar, «A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model», International Journal of Electrical and Computer Engineering, vol. 12, n.o 2. Institute of Advanced Engineering and Science, pp. 1831-1838, 2022. doi: 10.11591/ijece.v12i2.pp1831-1838.

[23] W. Peng, Y. Sun, y L. Zhang, «Construction of genetic classification model for coronary atherosclerosis heart disease using three machine learning methods», BMC Cardiovascular Disorders, vol. 22, n.o 1. BioMed Central Ltd, 2022. doi: 10.1186/s12872-022-02481-4.

[24] R. R. Sarra, A. M. Dinar, M. A. Mohammed, y K. H. Abdulkareem, «Enhanced Heart Disease Prediction Based on Machine Learning and χ2 Statistical Optimal Feature Selection Model», Designs, vol. 6, n.o 5. MDPI, 2022. doi: 10.3390/designs6050087.

[25] H. Khdair y N. M. Dasari, «Exploring Machine Learning Techniques for Coronary Heart Disease Prediction», International Journal of Advanced Computer Science and Applications, vol. 12, n.o 5. Science and Information Organization, pp. 28-36, 2021. doi: 10.14569/IJACSA.2021.0120505.

[26] M. W. Nadeem, H. G. Goh, M. A. Khan, M. Hussain, M. F. Mushtaq, y V. A. P. Ponnusamy, «Fusion-Based Machine Learning Architecture for Heart Disease Prediction», Computers, Materials and Continua, vol. 67, n.o 2. Tech Science Press, pp. 2481-2496, 2021. doi: 10.32604/cmc.2021.014649.

[27] T. Shen, D. Liu, Z. Lin, C. Ren, W. Zhao, y W. Gao, «A Machine Learning Model to Predict Cardiovascular Events during Exercise Evaluation in Patients with Coronary Heart Disease», Journal of Clinical Medicine, vol. 11, n.o 20. MDPI, 2022. doi: 10.3390/jcm11206061.

[28] G. N. Ahmad, Shafiullah, A. Salah Saidi, Imdadullah, y H. Fatima, «Efficient Medical Diagnosis of Human Heart Diseases Using Machine Learning Techniques with and Without GridSearchCV», IEEE Access, vol. 10. Institute of Electrical and Electronics Engineers Inc., pp. 80151-80173, 2022. doi: 10.1109/ACCESS.2022.3165792.

[29] U. Nagavelli, D. Samanta, y P. Chakraborty, «Machine Learning Technology-Based Heart Disease Detection Models», Journal of Healthcare Engineering, vol. 2022. Hindawi Limited, 2022. doi: 10.1155/2022/7351061.

[30] J. Cao, L. Zhang, L. Ma, X. Zhou, B. Yang, y W. Wang, «Study on the risk of coronary heart disease in middle-aged and young people based on machine learning methods: a retrospective cohort study», PeerJ, vol. 10. PeerJ Inc., 2022. doi: 10.7717/peerj.14078.

[31] S. Yousefi, «Comparison of the performance of machine learning algorithms in predicting heart disease», Frontiers in Health Informatics, vol. 10. Iranian Medical Informatics Association (IrMIA), 2021. doi: 10.30699/fhi.v10i1.349.

[32] N. M. Lutimath, N. Sharma, y B. K. Byregowda, «Prediction of Heart Disease using Biomedical Data through Machine Learning Techniques», EAI Endorsed Transactions on Pervasive Health and Technology, vol. 7, n.o 29. European Alliance for Innovation, 2021. doi: http://dx.doi.org/10.4108/eai.30-8-2021.170881.

[33] M. M. Rahma y A. D. Salman, «Heart Disease Classification-Based on the Best Machine Learning Model; [إلى استنادًا - الـ قلب رض ا أم تـ صنط يف الآلي لـ لـ تعلم نطهذج أف ضل]», Iraqi Journal of Science, vol. 63, n.o 9. University of Baghdad-College of Science, pp. 3966-3976, 2022. doi: 10.24996/ijs.2022.63.9.28.

[34] T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, y M. Hamdi, «Predictive Analysis of Heart Diseases with Machine Learning Approaches», Malaysian Journal of Computer Science, vol. 2022, n.o Special Issue 1. Faculty of Computer Science and Information Technology, pp. 132-148, 2022. doi: 10.22452/mjcs.sp2022no1.10.

[35] N. Alotaibi y M. Alzahrani, «Comparative Analysis of Machine Learning Algorithms and Data Mining Techniques for Predicting the Existence of Heart Disease», International Journal of Advanced Computer Science and Applications, vol. 13, n.o 7. Science and Information Organization, pp. 810-818, 2022. doi: 10.14569/IJACSA.2022.0130794.

Software Design Aimed at Proper Order Management in SMEs

Linett Velasquez Jimenez[1]*, Herbert Grados Espinoza[2],
Santiago Rubiños Jimenez[3], Juan Grados Gamarra[4], Claudia Marrujo-Ingunza[5]
Department of Engineering-Image Processing Research Laboratory (INTI-Lab),
Universidad de Ciencias y Humanidades (UCH), Los Olivos, Peru[1, 5]
Department of Engineering-Faculty of Industrial and Systems Engineering (FIIS),
Universidad Nacional del Callao (UNAC), Callao, Peru[2]
Department of Industrial Engineering-Faculty of Engineering and Architecture, Universidad Cesar Vallejo (UCV), Callao, Peru[3]
Department of Electrical Engineering-Faculty of Electrical and Electronic Engineering (FIEE),
Universidad Nacional del Callao (UNAC), Callao, Peru[4]

***Abstract*—The design and evaluation of an order management software oriented to SMEs in Lima is presented. Using Design Thinking, a prototype was developed focusing on Usability, Design and User Satisfaction. Through a Likert scale survey of 308 SME employees, perceptions on operational efficiency and user experience were measured. The results show high acceptance and highlight the intuitiveness of the system. However, areas such as loading speed and e-commerce functionalities require future improvements. This study establishes a framework for similar technological tools in commercial sectors.**

***Keywords*—Design thinking; SMEs; order management; software design; usability; user perception**

I. Introduction

The design of software for order management has become an essential component for the optimization of business processes of Micro and Small Enterprises (SMEs). In an environment marked by increasing competitiveness and high consumer expectations, the adoption of technological solutions becomes a fundamental part of the efficient and effective management of developing operations. The growth of e-commerce has transformed the way SMEs interact with their customers, and since the Internet has facilitated new sales opportunities for SMEs [1], a reach beyond traditional geographical boundaries is enabled. This highlights that the implementation of order management systems not only improves operational efficiency, but also contributes to greater customer satisfaction by providing real-time updates on the status of their orders [2].

In this context, efficient order management has become a key factor for the success of stores and businesses in various sectors. Small and medium-sized companies, in particular, face the challenge of managing a constant flow of orders while optimizing the user experience and maintaining customer satisfaction. To meet these challenges, the implementation of technology solutions such as order management systems offers a significant advantage, enabling stores to automate processes, improve accuracy, and reduce response times [3]. Order management systems are designed to handle a variety of tasks, from order receipt to final delivery. Some of the most important functionalities include centralized order processing that allows management from multiple channels with a reduction of errors and an improvement in processing speed; inventory control that facilitates stock tracking and prevents oversales or shortages; and real-time updates that provide constant information on order status, which contributes to an improved customer experience [4], [5].

In addition, it is proposed that future versions of the software could integrate emerging technologies such as Artificial Intelligence (IA) or Machine Learning (ML) to optimize decision making in areas such as inventory forecasting and purchase pattern analysis. Although these technologies were not included in this study, they represent an important opportunity to further improve the functionality of the system.

This article explores the features and benefits of order management software and the challenges faced by systems engineers in developing such solutions [1], and focuses on the development and evaluation of an order management software specifically designed for stores located in the busiest areas of Los Olivos, Lima. Through a sectorized approach, the software seeks to solve common management problems and facilitate the user experience, allowing businesses to improve their operational efficiency. However, the design and implementation of this type of software presents significant challenges. One of the main ones is integration with existing systems, which requires thorough analysis and careful planning by the systems engineer [1], [3].

Priority areas for improvement were also identified, such as the integration of e-commerce functionalities, which would allow SMEs to expand their digital reach and make sales effectively through online platforms. This aspect, together with the need for more intuitive interfaces and reduced loading times, is considered essential to ensure the acceptance and success of this type of technological tools.

To evaluate the effectiveness of the software, a survey has been designed based on three key dimensions: Design, Usability and User Perception. The results of this analysis will allow measuring user acceptance of the system and its impact on order management, providing valuable information for future improvements and technological adaptations [4].

This study hopes to provide not only a practical solution for local stores, but also to generate a reference framework for the development of technological tools that can be applied in other sectors with similar needs. In short, software design for ordering in SMEs is a critical area that not only improves operational efficiency, but also enhances competitiveness in a digitized market [5].

II. Literature Review

The development of order management software for SMEs has been widely researched due to its potential to improve operational efficiency and competitiveness in an increasingly digitized market. Several studies have addressed the challenges and benefits related to the implementation of these technological solutions.

The adoption of quality models in software-producing SMEs was investigated, highlighting that these companies represent 85% of the software industry. The research highlights the importance of implementing quality models to improve competitiveness and ensure high-quality products [6]. This approach is particularly relevant in the design of order management software, where accuracy and efficiency are crucial to ensure commercial success.

On the other hand, e-commerce in Peruvian SMEs was analyzed, revealing that many of these companies use digital platforms only as virtual catalogs, without taking advantage of their full potential to generate online sales. This study highlights the need to develop software that not only facilitates order management but also integrates e-commerce functionalities, allowing SMEs to make effective sales through digital channels [1].

In addition, a study on the design of software for the control of quotation processes in SMEs demonstrates how technological solutions can significantly improve operational efficiency. According to this analysis, both employees and managers recognize the importance of improving business processes through the use of software [7]. These solutions are easily adaptable to order management, optimizing business flow and minimizing operational errors.

In addition, order management systems offer multiple functionalities that are essential for SMEs. One analysis highlights that these systems not only enable centralized order processing and real-time inventory control, but also improve customer service by providing constant updates on order status [3]. This reduces processing errors, which is critical to maintaining customer satisfaction.

One of the relevant approaches in the early stages of the software development life cycle is represented in the application of Design Thinking. According to the study [8], this methodology is implemented especially in the analysis and design stages, for which it has shown high efficiency in eliciting requirements and creating architectures adapted to specific customer problems. It is also observed that its use fosters a deep understanding of the user's needs and an active collaboration within the development team, which contributes to generating innovative solutions, but with a certain degree of uncertainty.

On the other hand, [9] defines user interface design as a fundamental part of the creation of an attractive application. However, it is pointed out that the lack of attention to detail in planning and organization makes it deficient. The Design Thinking methodology is applied in this context to perform usability and user satisfaction tests using the System Usability Scale (SUS) to reduce these errors. As a result, an average value of 85.2 was obtained, which sustains that the value of the user interface design is included in the "Excellent" category.

Finally, automation is a key aspect of SME software design. A study on automation in SME software reveals that these tools enable companies to efficiently manage their day-to-day operations, such as quoting, invoicing, and inventory control. This automation not only reduces operating costs but also improves decision-making based on accurate data, which is vital for the sustainable growth of SMEs [10].

A. Theoretical Basis

1) Order management software: It is a tool that allows automation and optimization in the process of order entry, processing, and tracking. According to study [3], it contains inventory control, Customer Relationship Management (CRM), payment processing, and marketing integrations. It also improves operational efficiency, reduces response times, and minimizes errors, essential aspects to compete in an increasingly dynamic business environment. The study [7] indicates that specialized software helps SMEs in better workflow management, as it allows companies to adapt to new market demands.

2) User and customer experience: User experience is critical in the design of order management software, as an intuitive interface improves communication and streamlines transactions through automatic data storage, which increases efficiency and corporate image [7], [4]. A user-centered design benefits employees and has a positive impact on customer perception. Therefore, the design of ordering software in SMEs should consider operational functionalities with a good user experience, optimizing processes and improving customer satisfaction.

III. Methodology

Fig. 1 refers to the plan of the Design Thinking Methodology, which was used in this study to address the beginning and end of the design of a prototype aimed at improving proper order management in SMEs. It was segmented into five phases (Empathize, Define, Ideate, Prototype, and Test), in addition to detailing the tools used and the prototype design employed.

A. Design Thinking (DT)

This methodology is characterized by proposing innovative solutions through an iterative and people-centered approach. Its main objective is to generate concepts that not only solve specific problems, but also contribute to improving society. This approach has been widely recognized due to its ability to adjust creative ideas to practical needs [11], as well as to raise the expectations and level of commitment of those involved [12]. In

the business environment, Design Thinking has proven to be a key tool for identifying and solving user needs, optimizing processes and improving work environments [13]. Compared to traditional methodologies such as Agile or Waterfall, DT is particularly suitable for projects where user experience and iterative prototyping are critical, making it an ideal choice for this study.

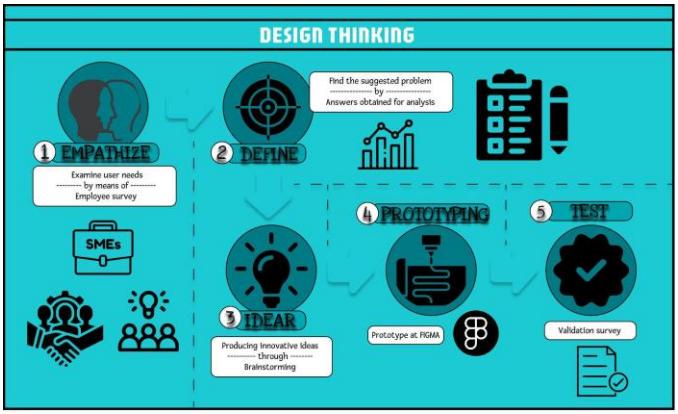


Fig. 1. Design thinking methodology.

1) Empathize: This initial stage focused on gaining a deep understanding of user needs and challenges through interviews and questionnaires with 308 SME employees. For example, the interviews revealed that simplicity in the interface was key to facilitating system adoption by less tech-savvy users. Also, the questionnaires indicated that many employees faced difficulties in managing large volumes of data, highlighting the need for specific functionalities such as advanced filters and optimized search [14].

2) Define: In this phase, the data collected were organized and analyzed using affinity maps, which allowed us to detect patterns and group the most relevant problems. For example, it was identified that unclear icons made navigation difficult in existing systems and that users required real-time information on order status. These findings guided the development of an initial software outline, prioritizing key modules such as customer, product and supplier management [15].

3) Ideate: Multiple innovative solutions were generated based on the problems detected, using techniques such as brainstorming and feasibility assessment. Among the ideas proposed, the creation of a modular system with specific functions that could be adapted to the individual needs of each business stood out. For example, an inventory module with real-time updates was included, since users highlighted the importance of avoiding oversales or shortages [16].

4) Prototype: The functional prototype was developed using the Figma tool, which allowed for the design of an intuitive interface and rapid iterations based on initial feedback. A key example was the incorporation of an interactive tutorial on the home page to help users become familiar with the main functions of the system. In addition, icons with descriptive labels and contrasting colors were implemented to improve visual clarity [17], [18].

5) Test: The prototype was evaluated by SMEs employees themselves through a structured survey that used a Likert scale to measure their perception of the Design, Usability and Overall Satisfaction. The results showed that 85% of respondents rated the intuitiveness of the design as “Strongly Agree”, while 88% rated the ease of completing tasks positively. In addition, qualitative analysis of the responses identified areas for improvement, such as the need to reduce loading times and optimize search functionality [19].

Each stage of Design Thinking not only contributed to the development of the software, but also ensured that it was aligned with the real needs and expectations of the users. Specific examples of decisions made in each phase demonstrate how this iterative, user-centered methodology enabled the development of a system that addresses critical management problems in SMEs, establishing an effective framework for future solutions. Although DT was highly effective in the ideation and prototyping stages, it is not designed for full software implementation. Future phases of the project could benefit from combining DT with other methodologies, such as Agile, to ensure smooth integration and implementation.

B. Software Evaluation Factors

1) Design: Factor that understands the requirements and develops the artifacts that define the creation of the product, also refers to the creation process that contains a reference to the requirements of the stakeholders [20].

2) Usability: Factor defined as the ability of users to understand, employ, and acquire knowledge of software in a simple way [21].

3) User perception: It is defined as the user interactions throughout the product life cycle, it is represented in values of Usability, Design, Loyalty, Quality, Interaction, Productivity, among others; which serve to improve the perception of value from a user's perspective [22].

IV. Results

This results section illustrates the stages of the Design Thinking methodology that will allow us to understand the users' problems and provide a short-term solution, as well as to understand their long-term perspectives in an already implemented software. The advantages, disadvantages and comparison of using the Design Thinking methodology are also presented.

A. Results of the Empathize Stage

During this stage, the main limitations in the use of software technologies were identified through interviews and questionnaires addressed to 308 SME employees. Table I presents the questions (Q1 to Q5) asked:

The results indicated that the priority aspects for users are the clarity and simplicity of the interface, the ease of handling large volumes of data, and the speed of tasks in the system. These findings underscore the importance of developing software that is intuitive, efficient, and accessible.

TABLE I. QUESTIONS

Questions	
ID	**Questions**
Q1	What improvements would you suggest to a software interface?
Q2	What do you look for in software when handling big data?
Q3	How do you define the speed of completion of your tasks in software?
Q4	What do you find important in software usability?
Q5	What do you consider fundamental in the functionalities of software?

B. Results of the Define Stage

Table II contains the most repeated answers in the questionnaire (R1 to R5) of the survey addressed to SME employees; there are a total of 308 answers for the analysis.

TABLE II. SME EMPLOYEE SURVEY

Responses	
ID	**Responses**
R1	Clarity and simplicity of interface
R2	Ease of handling large volumes of data
R3	Quick time to complete tasks
R4	Ease of use of software
R5	Easy access to functionalities

1) R1: The clarity and simplicity of the interface was mentioned as a key priority to ensure an efficient user experience.

2) R2: The ease of handling large volumes of data was highlighted as a necessary functionality to optimize workflow.

3) R3: Speed in completing tasks was identified as a determining factor in improving operational efficiency.

4) R4: Ease of use of the software was recognized as essential to promote acceptance among users.

5) R5: Easy access to specific functionalities was considered critical to ensure the effectiveness of the system.

C. Results of the Ideate Stage

Table III indicates the suggested solutions (S1 to S3) as a solution. An estimated score is provided for the choice of the best idea and its development is completed in the next phase.

TABLE III. SCORING IDEAS

Idea Scoring		
Solutions	**Ideas**	**Total**
S1	Creation of an order management software	60 points
S2	Create an order management application	26 points
S3	Applying Excel methods for order management	14 points

1) S1: The solution chosen was the creation of an order management software, given its high score and alignment with users' needs.

D. Results of the Prototyping Stage

For the results of the prototyping carried out with the Figma tool, the most salient functionalities that the order management software will offer in SMEs will be presented. The prototype functionalities are presented in Table IV.

TABLE IV. PROTOTYPE FUNCTIONS

Fig.	Function
2	Login, with access by credentials.
3	New Employee Module for ease of navigation.
4	New Customer Module to manage the organization's customers.
5	New Product Module to manage detailed product information.
6	New Category Module to classify products within the system.
7	New Distributor Module to manage the collaboration of distributors in the organization.
8	New Supplier Module to manage supplier transactions.
9	New Purchase Module to manage purchase orders within the organization.
10	Allows editing of purchase data in order to modify information to correct errors.

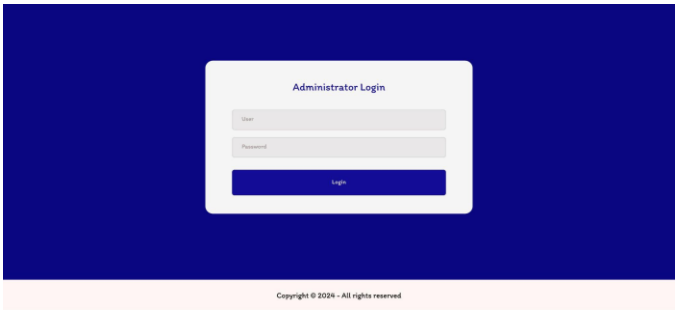


Fig. 2. Login.

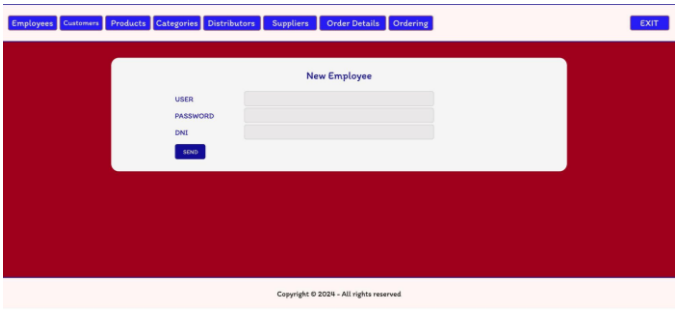


Fig. 3. New Employee.

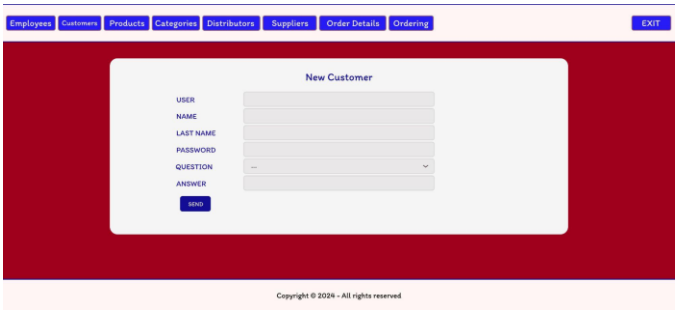


Fig. 4. New customer.

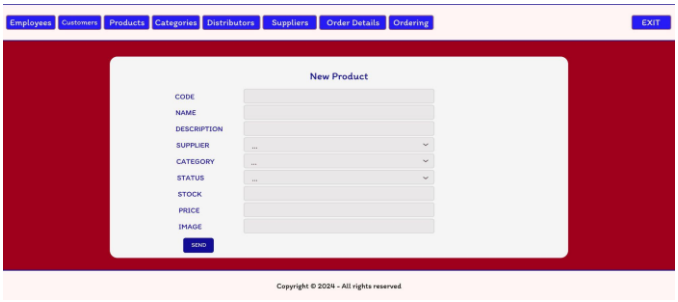


Fig. 5. New product.

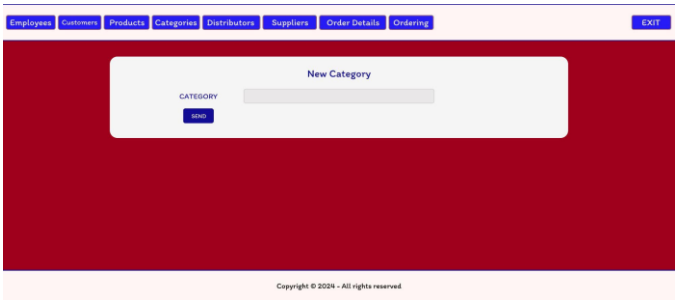


Fig. 6. New category.

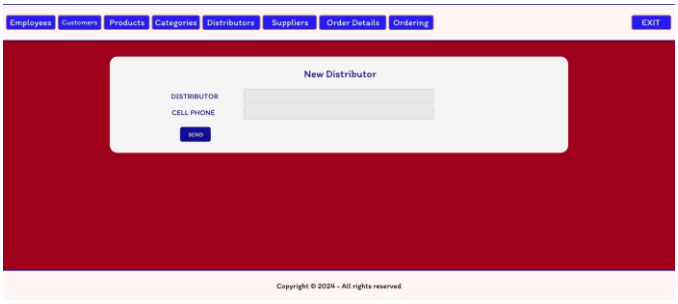


Fig. 7. New distributor.

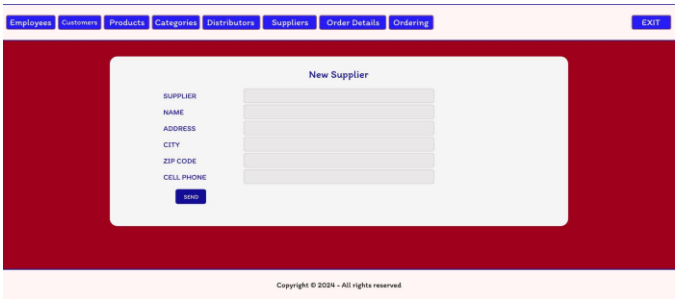


Fig. 8. New supplier.

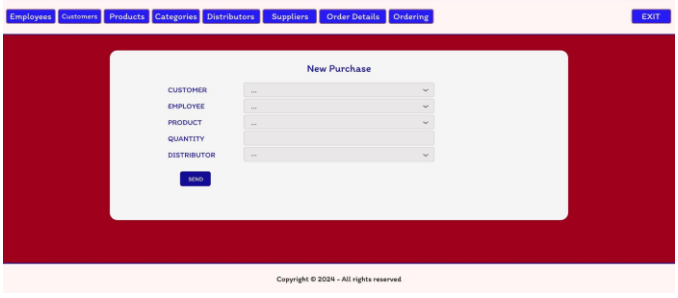


Fig. 9. New purchase.

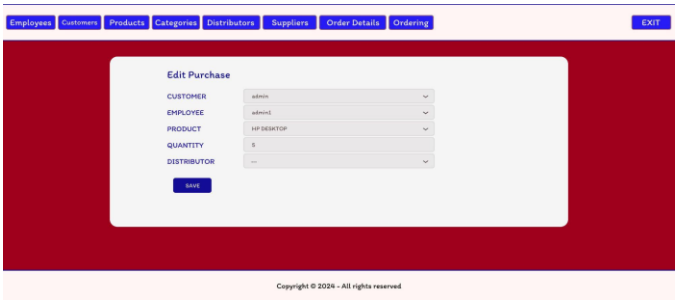


Fig. 10. Purchase edition.

E. Results of the Testing Stage

To develop this last phase of the Design Thinking methodology, it is required to validate that our order management prototype is suitable for the user and their needs. When surveying the 308 SME workers, three software evaluation factors were taken into account: Design, Usability, and User Perception. The Google Forms questionnaire tool was used, through 12 closed questions with a 4-point Likert scale (1= I Totally Agree and 4= I Totally Disagree), as well as an open question for possible improvements in an already implemented software. The questions used are presented in Table V.

TABLE V. INSTRUMENT

Q	Design
Q1	Do you find the login interface intuitive?
Q2	Do you consider that the design of the employee interface is clear and easy to understand?
Q3	Do you consider that the design of the customer management interface facilitates the handling of these data?
Q4	Is the organization of categories and sections consistent and easy to understand?
Usability	
Q5	Do you find it easy to perform tasks such as: creating or editing an order in the system?
Q6	Is the process of registering new employees, customers, or products easy?
Q7	Do you consider that the steps to edit information in the system are clear?
Q8	Do you consider that the average time to complete a task in the system is relatively short?
User Perception	
Q9	Do you perceive the overall ease of use of the order management system?
Q10	Would you rate the efficiency of the system in managing products and categories as adequate?
Q11	Does the system make it easy to create and edit distributors and suppliers?
Q12	Do you consider the different functionalities of the system accessible and understandable?

The design of the system was highly rated, with 85% of respondents rating the intuitiveness of the interface as “Strongly Agree”. In terms of usability, 88% highlighted the ease of performing tasks, highlighting its efficient functionality. In addition, user perception reflected a positive impact, with 82% indicating that the system significantly improves the overall experience.

On the other hand, Fig. 11, 12, 13, and 14 represent the results obtained according to the form applied in the SME sector to 308 people involved in sales management.

Fig. 11 presents the responses to the closed-ended questions asked of SME workers.

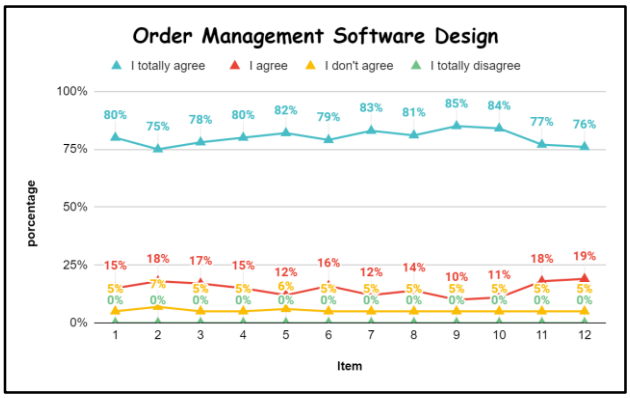


Fig. 11. Form results by ítem.

Fig. 11 shows the results of a survey on the design of order management software. The vertical axis shows the percentage of responses, while the horizontal axis shows the different items evaluated from 1 to 12. The response options are categorized into four levels: "I Agree" (represented in blue), where the majority of respondents are, maintaining a high trend in almost all items, with percentages ranging from 75% to 85%, indicating a strong overall approval towards the software design; 'I Agree' (represented in red), with a relatively low but constant percentage, between 12% and 19%, which, although not the dominant response, still reflects a considerable level of agreement in each item; "I Don't Agree" (represented in yellow), whose percentages remain between 5% and 7%, suggesting that a small portion of respondents do not agree with certain aspects of the design; and 'I Totally Disagree' (represented in green), which shows no significant presence, indicating that almost no respondents strongly disagree with the design of the software. Overall, the graph suggests a positive perception towards the design of the order management software, with high levels of acceptance and satisfaction among users. The levels of "I Totally Agree" and "I Agree" are predominant in almost all items, reflecting a favorable reception.

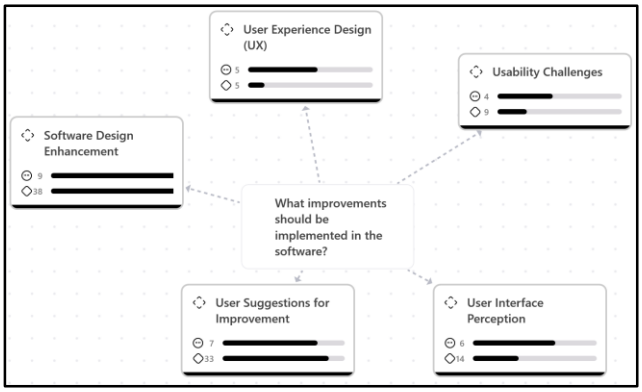


Fig. 12. Areas of software improvement.

Likewise, Fig. 12 presents the areas of improvement analyzed according to the open question asked in the survey, which was analyzed in the ATLAS.ti software to obtain the most repetitive nodes and points concerning possible improvements in an already implemented software.

Fig. 12 represents a concept map that highlights the key areas where users suggest improvements to an already implemented software. Themes include user experience (UX) design as a factor in improving efficiency and processes [23]; however, having a straight-line direction based on design is debatable due to a lack of concrete guidelines [24]. Usability challenges are also presented [25], considered strongly integral attributes [26], and it also points to interface perception as a decisive element for a consistent and sustainable experience [27], [28].

On the other hand, general suggestions for improvement are presented, such as better performance [29] with higher flexibility requirements [30]. Also, software design improvements from source code [31] are raised, this in order to increase software efficiency [32]. Each node shows the number of related comments and mentions, suggesting the priority areas to optimize the application.

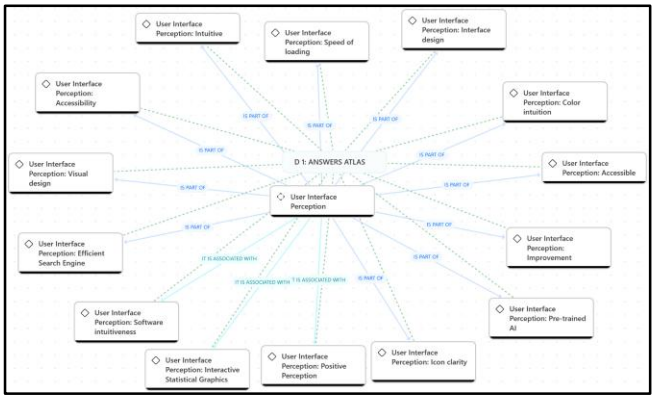


Fig. 13. Improvements in user interface perception.

Fig. 13 shows the different elements suggested by users about the perception of a software interface. Key elements in software use were highlighted, such as intuitiveness in user interaction [33], loading speed for a smooth experience [34], interface design for a visual representation of data [35], color intuitiveness to optimize task performance [36], accessibility to promote access inclusion and accessible product management [37], [38]. On the other hand, the overall positive perception of users is based on practical improvements [39], which are distributed in relevant elements such as the clarity of icons to capture users' attention [40], the search engine for data filtering [41], and in interactive statistical graphics for reporting [42]. Each of these elements is associated with the user experience and the diagram indicates how each factor is part of the user's overall perception of the interface.

Fig. 14 shows the main usability challenges that users have suggested in relation to an implemented software. Among them are noted suggestions related to accessibility [37], slow loading speed [34], lack of clarity in icons [40], and usability factors that improve system performance for the user [43], [44], [45]. In addition, they point to the lack of features as an indication of dissatisfaction [46], the challenge of adaptation with respect to

responsive design, which is a suggested element to increase visibility [47] and the importance of optimizing real-time analytics for effective prediction [48]. This indicates areas where an implemented software needs adjustments to improve the user experience.

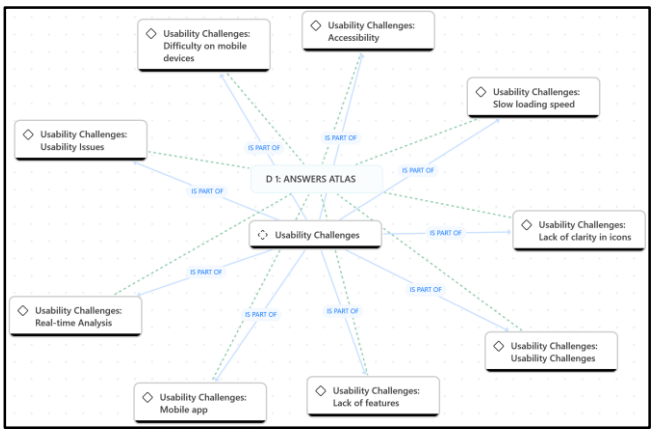


Fig. 14. Improvements in usability challenges.

F. About the Methodology

1) Advantages: The Design Thinking methodology was used to generate new ideas that respond to the needs and requirements of the users. This made it easier for the work to be active to provide new proposals to society and optimize the process of proper management.

2) Disadvantages: The disadvantages of this methodology is its focus on the prototyping of innovative ideas and not on the implementation of the software. In this context, it should be considered that this methodology can be complemented with another one for a future implementation phase.

3) Comparison: The DT methodology allowed the streamlining of the process of analysis and the generation of innovative ideas aimed at prototyping. In comparison with other methodologies focused on software development, this one includes the proposal of new plans that contribute to society.

V. Discussion

The findings of this study confirm and extend the existing knowledge on the positive impact of order management systems in SMEs, as highlighted by previous studies that emphasize the importance of accuracy and efficiency in software design to optimize operational processes and strengthen business competitiveness [6]. In this sense, the results obtained during the stages of the Design Thinking methodology reinforce the relevance of understanding users' needs in depth in order to design solutions that not only solve current problems, but also anticipate future needs.

First, the iterative and user-centered approach of the Design Thinking methodology was crucial to identify and solve SME needs. During the Empathize stage, priorities such as interface clarity and simplicity, speed of tasks and the ability to handle large volumes of data were identified. This underscores the importance of a design that combines efficiency and accessibility, aligning with research highlighting how intuitive design and high usability can minimize errors and increase technology adoption in sectors traditionally resistant to change [8].

In addition, the Ideate stage enabled the selection of the most appropriate solution through a detailed feasibility and impact analysis. Similarly, the Prototyping stage, carried out in Figma, facilitated the creation of a functional interface that was subsequently validated in the testing stage. It is worth noting that the use of surveys and Likert scales to evaluate Design, Usability and User Perception factors yielded favorable results, with high levels of approval reflected in Fig. 11, 12, 13 and 14. This evidences the effectiveness of the methodology to integrate relevant functionalities from the early stages of development, addressing a combined deficiency reported in previous studies [1].

However, it is important to recognize that, despite the high acceptance of the software, key areas for future improvements are identified. Among them, the integration of functionalities, the optimization of loading times and the improvement of responsive design are relevant challenges. In this context, user suggestions, analyzed in Fig. 12 and Fig. 14, revealed that aspects such as color intuitiveness, icon clarity and accessibility are critical elements to improve user experience.

On the other hand, the incorporation of emerging technologies such as Artificial Intelligence (IA) and Machine Learning (ML) emerge as a strategic opportunity to enhance the analytical capabilities of the system. These technologies could provide advanced functionalities, such as sales trend prediction or experience personalization, which are consistent with recent studies demonstrating their positive impact on business applications [12], [13].

Finally, this study reaffirms the value of a user-centered approach and the application of methodologies such as Design Thinking to develop technological solutions tailored to the specific needs of SMEs. The results obtained highlight the effectiveness of intuitive design and software functionality in improving operational efficiency and meeting user expectations. However, a limitation of Design Thinking is its focus on prototyping, which leaves the implementation of the software as a challenge for future stages. Given this, the importance of complementing with other methodologies to comprehensively address development and implementation is stressed.

In conclusion, this study not only validates the effectiveness of an iterative approach to software design, but also points the way for future iterations that include technological and functional improvements. Ultimately, these actions will contribute to optimizing the user experience and increasing the impact of software in the SME environment.

VI. Conclusions

The implementation of the Design Thinking approach allowed the development of an order management software tailored to the needs of SMEs in Los Olivos, Lima. This iterative and user-centered method was essential to address key issues such as simplicity in the interface and efficiency in data management, thus achieving a system that was highly accepted by users. Indicators such as design intuitiveness (85%) and ease of task completion (88%) reflect the effectiveness of the approach, while the overall improvement in user experience

(82%) underscores the software's positive impact on operational optimization and employee satisfaction.

Despite the progress made, key areas for future improvements were identified, such as the integration of functionalities and the optimization of loading times, which could strengthen the competitiveness of SMEs. In addition, the incorporation of emerging technologies such as IA and ML represent a strategic opportunity to enrich the system with analytical and predictive capabilities, making it possible to forecast sales trends and personalize experiences. These enhancements promise to consolidate the role of software as a transformative tool in the business context.

Although the study showed encouraging results, it also highlighted certain limitations, such as the challenge of implementing the system on a larger scale and the need to evaluate its long-term impact. It is recommended to combine Design Thinking with methodologies such as Agile to strengthen future development and integration phases, and to prioritize the inclusion of advanced technologies in future iterations. Ultimately, this work not only provides specific solutions for SMEs, but also lays the groundwork for the design of sustainable technological tools, highlighting the role of innovation as a driver of competitiveness in a constantly evolving digital marketplace.

REFERENCES

[1] L. H. Carbajal, Los factores a considerar para la venta en sitios web de Mypes del sector textil y Confecciones. Bachelor's Thesis, Faculty of Management and Senior Management, Pontificia Universidad Católica del Perú, 2023.

[2] Compara Software, “Software de Gestión de Pedidos en Perú,” 2024.

[3] Capterra, “Sistemas de gestión de pedidos,” 2024.

[4] HubSpot, “22 software gratis para pequeñas empresas en 2024,” 2024.

[5] Universidad Peruana de Ciencias Aplicadas (UPC), “Ingeniería de Sistemas de Información,” 2024.

[6] M. E. Amable-Ciudad and R. Millones-Rivalles, “Uso de modelos de calidad en las mypes productoras de software de Lima,” Ingeniería Industrial, vol. 37, no. 037, pp. 81-99, 2019. doi: 10.26439/ing.ind2019.n037.4543.

[7] L. O. Albarracín, L. J. Molina, E. J. Jalón, and C. M. Marín, “Diseño de software para control del proceso de cotizaciones en Pyme del cantón La Maná,” Revista Estudios del Desarrollo Social: Cuba y América Latina, vol. 10, no. 1, 2022.

[8] V. Cahui, D. Quispe, A. Condori, and J. Chapi, “Case Studies of Design Thinking in the Analysis and Design stages of Software Development,” Innovation and Software, vol. 3, no. 1, pp. 17-29, 2022. doi: 10.48168/innosoft.s8.a50.

[9] F. Fajri, F. Rizal, M. Yaqin, and Z. Purwanto, “Analysis And Design of Mobile Applications For Make-Up Artist Services (Halomua) With The Design Thinking Framework,” Sinkron: Jurnal Dan Penelitian Teknik Informatika, vol. 7, no. 3, pp. 1400-1408, 2023. doi: 10.33395/sinkron.v8i3.12483.

[10] SICO Pequeños Negocios, “Software para Pymes y Emprendedores,” 2024.

[11] B. Ku and E. Lupton, Health design thinking: creating products and services for better health. MIT Press, 2022.

[12] G. Wang, “Digital reframing: The design thinking of redesigning traditional products into innovative digital products,” Journal of Product Innovation Management, vol. 39, no. 1, pp. 95–118, 2022. doi: 10.1111/jpim.12605.

[13] S. Magistretti, C. Dell’Era, R. Verganti, and M. Bianchi, “The contribution of design thinking to the r of r&d in technological innovation,” R&D Management, vol. 52, no. 1, pp. 108–125, 2022. doi: 10.1111/radm.12478.

[14] J. McLaughlin, E. Chen, D. Lake, W. Guo, E. Skywark, A. Chernik, and T. Liu, “Design thinking teaching and learning in higher education: Experiences across four universities,” Plos One, vol. 17, no. 3, p. e0265902, 2022. doi: 10.1371/journal.pone.0265902.

[15] C. Pham, S. Magistretti, and C. Dell’Era, “The role of design thinking in big data innovations,” Innovation, vol. 24, no. 2, pp. 290–314, 2022.

[16] B. Matthews, S. Doherty, P. Worthy, and J. Reid, “Design thinking, wicked problems and institutioning change: a case study,” CoDesign, pp. 1–17, 2022. doi: 10.1080/15710882.2022.2034885.

[17] J. Vrana and R. Singh, “Nde 4.0—A Design Thinking Perspective,” Journal of Nondestructive Evaluation, vol. 40, no. 1, pp. 1–24, 2021. doi: 10.1007/s10921-020-00735-9.

[18] A. Mandani, A. Bakti, and A. Info, “UI/UX Design Of Sales Mobile Application On Up Store Using Figma,” JSAI (Journal Scientific and Applied Informatics), vol. 6, no. 3, pp. 462–468, 2023. doi: 10.36085/jsai.v6i3.5717.

[19] L. Vendraminelli, L. Macchion, A. Nosella, and A. Vinelli, “Design thinking: strategy for digital transformation,” Journal of Business Strategy, vol. 44, no. 4, pp. 200–210, 2022. doi: 10.1108/JBS-01-2022-0009.

[20] I. Ozkaya, “Building Blocks of Software Design,” IEEE Software, vol. 37, no. 2, pp. 3–5, Mar.-Apr. 2020. doi: 10.1109/MS.2019.2959049.

[21] A. Almazroi, “A Systematic Mapping Study of Software Usability Studies,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 12, no. 9, 2021. doi: 10.14569/IJACSA.2021.0120927.

[22] J. Idogawa, F. Bizarrias, C. Martnes, J. Contador, and W. Satyro, “User experience: the factors to improve the perception of value in software projects,” International Journal of Scientific Management and Tourism, vol. 9, no. 6, pp. 3247–3277, 2023. doi: 10.55905/ijsmtv9n6-002.

[23] J. Wang, Z. Xu, X. Wang, and J. Lu, “A Comparative Research on Usability and User Experience of User Interface Design Software,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 13, no. 8, 2022. doi: 10.14569/IJACSA.2022.0130804.

[24] A. Berni, Y. Borgianni, D. Basso, and C. Carbon, “Fundamentals and issues of user experience in the process of designing consumer products,” Design Science, vol. 9, 2023. doi: 10.1017/dsj.2023.8.

[25] A. Nemeth and A. Bekmukhambetova, “Achieving Usability: Looking for Connections between User-Centred Design Practices and Resultant Usability Metrics in Agile Software Development,” Periodica Polytechnica Social and Management Sciences, vol. 31, no. 2, pp. 135–143, 2023. doi: 10.3311/PPso.20512.

[26] Q. Ain, T. Rana, and Aamana, “A Study on Identifying, Categorizing and Reporting Usability Bugs and Challenges,” 2023 International Conference on Communication Technologies (ComTech), Rawalpindi, Pakistan, 2023, pp. 53–68. doi: 10.1109/ComTech57708.2023.10165169.

[27] W. Li, Y. Zhou, S. Luo, and Y. Dong, “Design Factors to Improve the Consistency and Sustainable User Experience of Responsive Interface Design,” Sustainability, vol. 14, no. 15, p. 9131, 2022. doi: 10.3390/su14159131.

[28] G. Li, “Visual Design of Human-Computer Interaction Software Interface Information Based on Human Factors Engineering,” 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1–5. doi: 10.1109/INCET57972.2023.10170124.

[29] A. Blot and J. Petke, “MAGPIE: Machine Automated General Performance Improvement via Evolution of Software,” ArXiv, abs/2208.02811, 2022. doi: 10.48550/arXiv.2208.02811.

[30] V. Lakhai, O. Kuzmych, and M. Seniv, “An improved approach to the development of software with increased requirements for flexibility and reliability in terms of creating small and medium-sized projects,” 2022 IEEE 17th International Conference on Computer Sciences and Information Technologies (CSIT), Lviv, Ukraine, 2022, pp. 474–477. doi: 10.1109/CSIT56902.2022.10000787.

[31] Z. Tao, X. Sun, Z. Zheng, and G. Li, “Análisis inteligente de datos de software: investigación y aplicaciones,” Frontiers of Information Technology & Electronic Engineering, vol. 23, pp. 661–663, 2022. doi: 10.1631/FITEE.2230000.

[32] N. Luo and Y. Xiong, "Enhanced Software Design for Boosted Continuous Software Delivery," International Journal of Software Engineering & Applications, vol. 13, no. 2, pp. 23–31, 2022. doi: 10.5121/ijsea.2022.13202.

[33] J. P. Göpfert, U. Kuhl, L. Hindemith, H. Wersing, and B. Hammer, "Intuitiveness in Active Teaching," IEEE Transactions on Human-Machine Systems, vol. 52, no. 3, pp. 458–467, 2022. doi: 10.1109/THMS.2021.3121666.

[34] J. Pibernik, J. Dolić, L. Mandić, and V. Kovač, "Mobile-Application Loading-Animation Design and Implementation Optimization," Applied Sciences, vol. 13, no. 2, p. 865, 2023. doi: 10.3390/app13020865.

[35] B. Aslan and F. Aslan, "Examining the User Interface Development Stage in the Software Development Process," European Journal of Science and Technology, no. 35, pp. 408–416, 2022. doi: 10.31590/ejosat.1055996.

[36] S. Treneska, E. Zdravevski, I. M. Pires, P. Lameski, and S. Gievska, "GAN-Based Image Colorization for Self-Supervised Visual Feature Learning," Sensors, vol. 22, no. 4, p. 1599, 2022. doi: 10.3390/s22041599.

[37] W. Shi, H. Moses, Q. Yu, S. Malachowsky, and D. Krutz, "ALL: Supporting Experiential Accessibility Education and Inclusive Software Development," ACM Transactions on Software Engineering and Methodology, vol. 33, no. 2, pp. 1–30, 2023. doi: 10.1145/3625292.

[38] C. Furukawa, M. Soares, M. Cagnin, and D. Paiva, "Support for Accessible Software Coding: Results of a Rapid Literature Review," CLEI Electronic Journal, vol. 25, no. 3, pp. 1–13, 2023. doi: 10.19153/cleiej.25.3.1.

[39] E. Kühlmann, S. Hamer, and C. Quesada-López, "Software Visualization using the City Metaphor: Students' Perceptions and Experiences," 2023 XLIX Latin American Computer Conference (CLEI), La Paz, Bolivia, pp. 1–10, 2023. doi: 10.1109/CLEI60451.2023.10346099.

[40] T. Korpilahti and M. Massodian, "Guidelines for the Visual Design of Mobile Application Icons. An Experiential Case Study," IxD&A, vol. 54, pp. 241–276, 2022. doi: 10.55612/s-5002-054-010.

[41] C. Meng et al., "Quasi Real-Time Distributed Search Engine Based on Massive Operation and Maintenance Data," 2023 Asia-Europe Conference on Electronics, Data Processing and Informatics (ACEDPI), Prague, Czech Republic, pp. 300–304, 2023. doi: 10.1109/ACEDPI58926.2023.00065.

[42] D. D. Subramaniam and S. C. Johnson Lim, "An Interactive Visualization Web Application for Industrial-Focused Statistical Process Control Analysis," Journal of Science and Technology, vol. 14, no. 2, pp. 20–30, 2022. doi: 10.30880/jst.2022.14.02.003.

[43] J. M. Ferreira, F. D. Rodríguez, A. Santos, O. Dieste, S. T. Acuña, and N. Juristo, "Impact of Usability Mechanisms: A Family of Experiments on Efficiency, Effectiveness and User Satisfaction," IEEE Transactions on Software Engineering, vol. 49, no. 1, pp. 251–267, 1 Jan. 2023. doi: 10.1109/TSE.2022.3149586.

[44] H. Bayomi, N. A. Sayed, H. Hassan, and K. Wassif, "Application-based Usability Evaluation Metrics," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 13, no. 7, 2022. doi: 10.14569/IJACSA.2022.0130712.

[45] T. Issa and P. Isaias, "Usability and Human Computer Interaction (HCI)," in Sustainable Design, Springer, London, 2015. doi: 10.1007/978-1-4471-6753-2_2.

[46] A. Ali, I. Khalil, I. Ahmad, I. Parveen, and U. K. Uz Zaman, "Role of Non-functional Requirements in projects' success," 2022 2nd International Conference on Digital Futures and Transformative Technologies (ICoDT2), Rawalpindi, Pakistan, pp. 1–7, 2022. doi: 10.1109/ICoDT255437.2022.9787463.

[47] Y. Chen, S. Huang, Y. Chien, and I. Hsiao, "Quality Assessment of Web and APP Design Patterns," International Journal of Knowledge Engineering, vol. 9, no. 1, pp. 14–20, 2023. doi: 10.18178/ijke.2023.9.1.139.

[48] X. Hu, "Data Assimilation For Simulation-Based Real-Time Prediction/Analysis," 2022 Annual Modeling and Simulation Conference (ANNSIM), San Diego, CA, USA, pp. 404–415, 2022. doi: 10.23919/ANNSIM55834.2022.9859329.

Design of a Mobile Learning App for Financial Literacy in Young People Using Gamification

Angie Nayeli Ruiz-Carhuamaca, Juliana Alexandra Yauricasa-Seguil, Juan Carlos Morales-Arevalo
Faculty of Engineering, Universidad Peruana de Ciencias Aplicadas, Lima, Perú

***Abstract*—This research paper addresses the issue of insufficient financial literacy among young people, a challenge that affects their ability to make informed financial decisions. A survey was conducted to assess the current state of financial literacy among young people, whose results show a significant gap in the understanding of key concepts needed to manage their finances, which limits their economic and social development. Based on these findings, an interactive and gamified design aimed at strengthening the level of financial literacy among young people is proposed. This proposal includes wireframes that structure a mobile application, integrating playful elements and educational challenges to promote user participation in their learning process. The methodology of design that is employed focuses on the user experience, which ensures that the tool is accessible and engaging. It is expected that this proposal, based on the survey results, will not only increase the understanding of financial concepts but also motivate young people to apply this knowledge in their daily lives, thus contributing to greater financial independence and a better quality of life.**

***Keywords*—Financial literacy; gamification; financial education; challenge education**

I. Introduction

With rapid evolution, the world is becoming increasingly complex and competitive, and financial knowledge is becoming a necessary skill, especially for young people in the future and for a country's economy and financial security [1]. This competence not only allows individuals to manage their economic resources effectively but also influences their ability to face future financial challenges. Globally, this skill is essential, particularly in an increasingly digitized context, where technologies such as FinTech (financial technology) and EdTech (educational technology) are promoted, providing improvements in education through new technologies to facilitate access and contribute to reducing inequality. Thus, offering opportunities to develop financial skills in an economic and social environment that is constantly changing [1].

In Peru, where much of the young population faces economic uncertainty, financial education is crucial. However, there is a lack of financial referents with solid knowledge; 41% of adults at the national level obtained a minimum level of financial education in a study of financial capabilities, while only 13% obtained a high level [2]. This is evidence of their limited ability to manage their financial resources effectively and make responsible financial decisions. However, there are global initiatives to promote financial education, such as the Organization for Economic Co-operation and Development (OECD) approach, which coordinates national strategies to strengthen financial learning worldwide [1].

This problem not only affects the financial well-being of young people but also generates a cycle of economic dependence and limits the economic development of a society. This is reflected in inappropriate financial behaviors, such as the imbalance between income and expenses; in 2022, 56% of adults reported that they went into debt to cover their expenses [2]. In contrast, when people have solid financial knowledge and skills, they can effectively manage their finances and thus generate greater stability and well-being in their families [3]. Therefore, it is crucial to have a high level of financial skills, since an insufficient level of financial knowledge leads to poor economic decisions and financial problems in the long term.

Therefore, it is essential to develop strategies that promote financial education attractively and effectively. The current teaching method emphasizes that students should have more freedom to learn technology, think, and make mistakes. In that sense, a mobile application turns out to be a suitable tool for the new generations, since the trend of using mobile applications is increasing. In addition, learning in financial management is not achieved only through reading and writing as the traditional teaching method, rather, students must acquire knowledge through scenario-based learning, which allows them to interpret and apply the acquired knowledge in real-life situations [1].

Finally, this research seeks to address the issue of insufficient financial literacy through the design of an interactive and gamified mobile learning application. The proposal is based on a survey that assesses the level of financial literacy of young people, the results of which showed a significant gap between the understanding of financial concepts for the management of their finances and the application of this knowledge in financial challenges. In response to this gap, we propose the design of a mobile application that incorporates playful elements and educational challenges that encourage practical and engaging learning.

The structure of this paper consists of eight sections: Section I corresponds to the introduction and Section II explains the related work. Details about the structure, tools, and design of the mobile application are proposed in Section III. Section IV presents the survey. Section V shows the results of the survey. Section VI shows the discussion and finally, Section VIII shows the conclusions.

II. Related Works

A. Interactive and Gamified Solutions

The study in [4] proposes the design of a support system with Artificial Intelligence (AI) to improve the quality of corporate financial decisions. On the other hand, a study [5] proposes a

mobile learning platform to develop entrepreneurial skills in students. Both studies consider that technology can optimize decision-making but differ in the way it is applied. On the other hand, studies [6], [7] and [8] focus on improving learning through gamification. Studies in [6] and [7] highlight the potential of games as educational tools, [6] promote financial literacy in students through economic games to teach basic financial concepts, while [7] promote the development of technical and soft skills through simulation business games. Taken together, these five papers highlight that the integration of AI, gamification, and mobile platforms not only improves learning but also optimizes financial decision-making. The key difference between them is how they are applied: [4] and [5] focused on business or entrepreneurial contexts, while [6], [7] and [8] explored learning through games and simulations.

B. Integration of AI-Assisted Coaches

All studies agree on the use of AI to personalize the user experience. However, the nature of customization varies depending on the field of application. The study in [9] is based on education, [10] health, [11] sports, and the study [12] focuses on opponent modeling and strategy optimization within a competitive environment. A common point of all studies is that AI is implemented to optimize decision-making in environments where the number of variables and uncertainty make it difficult for users to make the right decision. Concerning training and data collection methods, studies [9] and [10] are based on creating user-centric experiences. On the other hand, [11] and [12] face challenges related to the availability of high-quality data. In [11], the system had to deal with a shortage of data to train deep learning models, while [12] it stands out for the use of simulations and game data, which allowed AI to outperform previous systems, without the need for large amounts of real user data. Regarding the AI technologies used, each study adopts different technologies and approaches to implementing AI, reflecting the diversity in the way AI systems can be trained and executed. Studies [9] and [10] use AI in predictive models and heuristics, optimizing user behavior through algorithms that learn common patterns in large amounts of data. Instead, studies [11] and [12] rely on technologies such as deep learning and simulations, allowing systems to generate predictions and adapt to complex situations.

C. Importance of Financial Literacy in Education

Financial literacy emerges as a central theme in these studies, underscoring the relevance of developing healthy financial habits at different stages of life. In the study [13], they detailed an understanding of healthy financial habits in young adults, exploring the relationship between subjective financial literacy, financial engagement, and financial decision-making. On the other hand, [14] implements a more structured approach through the SaveWise program, which aims to increase the financial knowledge of adolescents, where literacy is measured more objectively, focusing on students' ability to apply financial concepts in real-life situations. Regarding the results, in [13], young adults who consider themselves financially literate show a trend towards healthy practices, such as saving regularly and investing in their future. In study [14] program participant's experienced significant improvements in their financial knowledge, which translated into increased savings intentions and a better understanding of money management.

D. Gamification and its Impact on Youth People Learning

The use of serious educational games and adaptive technologies is effective in developing specific skills and improving student engagement. The study in [15] shows that EEG (Entrepreneurial Education Game), by integrating adaptive algorithms to personalize the learning experience, resulted in greater student adaptability with 74% of participants reporting a positive learning experience. Similarly, the study [16] reveals that implementing a serious game in virtual reality improved players' mental calculation skills, although it also identified some challenges related to the use of technologies. On the other hand, studies [17], [18], and [19] show that gamification not only improves understanding of concepts, it also addresses specific skills, such as programming and AI safety. The study [18] proposed a smart tutoring system built into an educational game that provides personalized support, where students who used it got fewer errors. This study in [18] found that students who participated in the Shoot2Learn game showed a statistically significant improvement in their understanding of conditional structures, suggesting that educational games can motivate students and improve their academic performance. For its part, the study [19] evaluated a video game designed to improve AI safety education, where it found that 81% of users considered the game's learning functionalities effective.

III. PLATFORM DESIGN

The design of the proposed mobile application focuses on offering an attractive and accessible user experience, to improve the financial literacy of young people through active interaction with the financial coach and the use of gamification. This section describes the key components of the platform, design principles, and tools used for wireframe development and presents visual examples of the main screens.

The design of the proposed mobile application focuses on offering an attractive and accessible user experience, to improve the financial literacy of young people through active interaction with the financial coach and the use of gamification. This section describes the key components of the platform, design principles, and tools used for wireframe development and presents visual examples of the main screens.

A. Overview of the Platform

The application is based on three key sections, which are designed to guide the user in their continuous learning process.

1) Initial assessment: When starting, the user selects the financial topic they want to learn or reinforce and their level of knowledge as shown in Fig. 1, and then performs the initial assessment Fig. 2, which consists of five questions. This flow is done to provide a personalized experience, which allows you to suggest the most appropriate level according to the user's performance and to provide an adaptive experience from the beginning of the application.

2) Gamified challenges: In this section, users face different financial challenges, which are designed to test their skills in topics such as saving, investing; and credits and debts Fig. 3. These gamified challenges integrate game mechanics such as rewards, difficulty levels, and short-term goals Fig. 4. In addition, there is interaction with the financial coach, who will

guide, provide personalized recommendations and suggestions based on areas for improvement during and after the game Fig. 3. Thus, this active interaction between the user and the coach makes the application interactive.

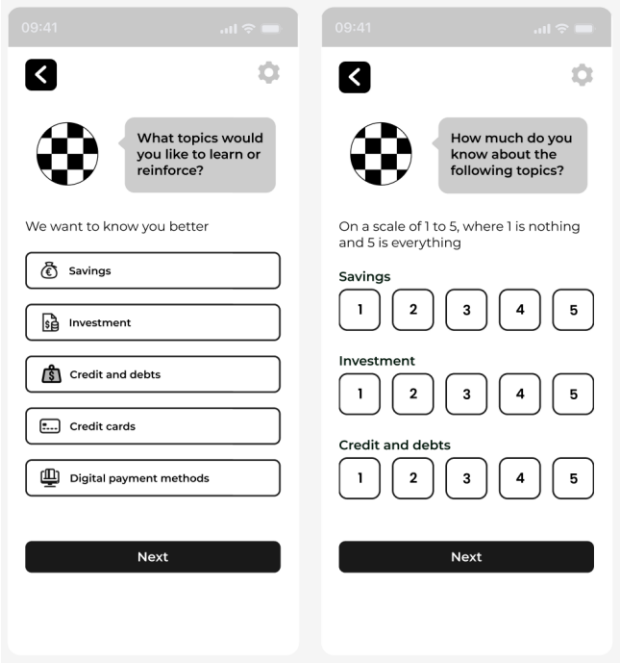


Fig. 1. Financial topic and scale of knowledge level.

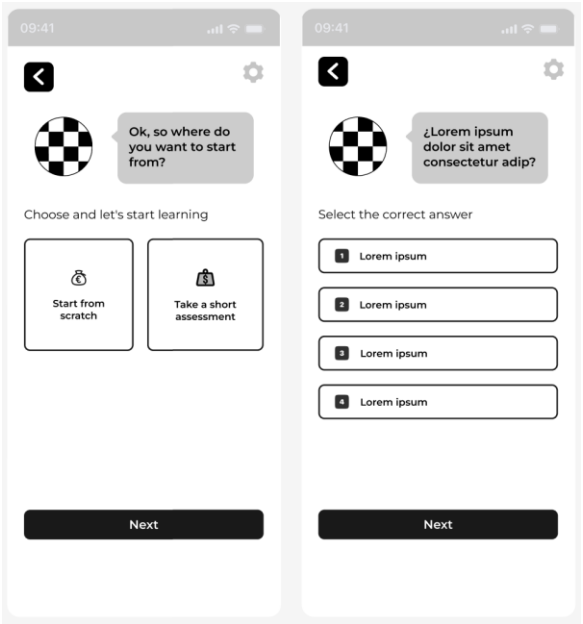


Fig. 2. Initial evaluation.

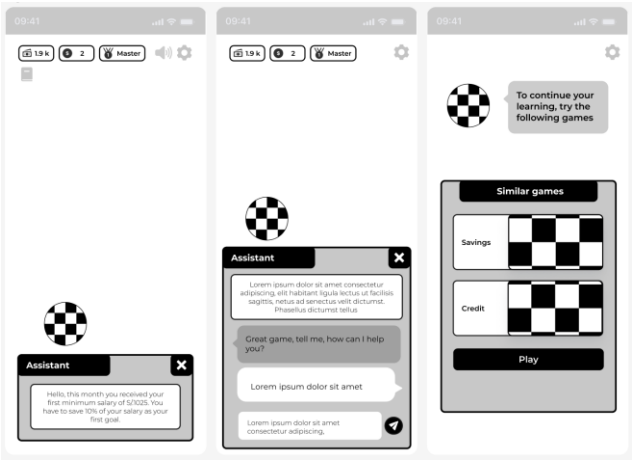


Fig. 3. Game scenarios and interaction with the virtual assistant.

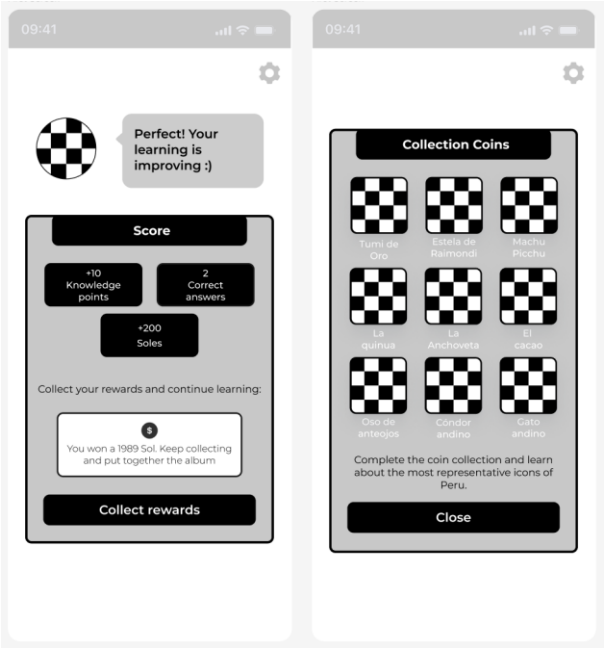


Fig. 4. Scoring and reward system.

3) Final evaluation: At the end of the challenges, users take a final assessment to measure their progress. Fig. 5 shows the design of the quizzes, the results, and the progress panel; and personalized recommendations based on user performance.

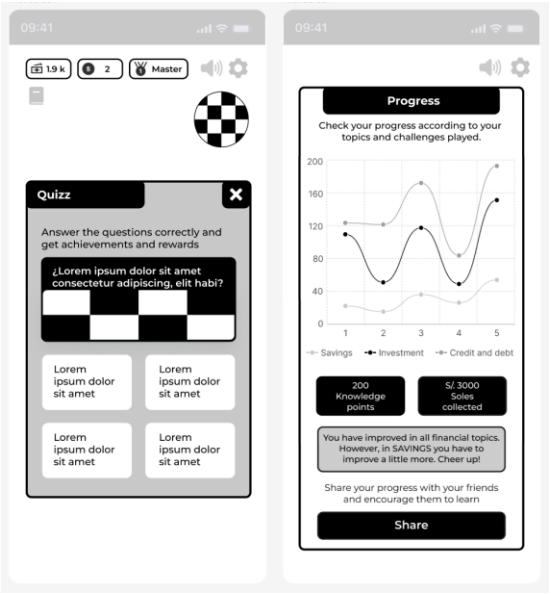


Fig. 5. Assessment and recommendation system.

B. Design Principles

Concerning the design principles, User-Centered Design (UCD) was chosen, which focuses on guaranteeing the usability and accessibility of the game to ensure its effectiveness for users. This approach recognizes that everything is an interrelated system, where the set of components and the interaction of people must be considered. Thus, by observing how users interact and identifying problems, iterative tests are carried out to get closer to the final product [20]. Thus, the project focuses on satisfying the needs of the user, that the designs are interactive, and the final product is easy to use.

C. Design Tools

During the mobile app design and prototyping process, various tools were used for the design of different areas and functions of the platform. These include:

- Figma. It is a vector graphics and prototyping editor. The advantage of this tool is that it allows users to work collaboratively. This tool was used for the design of all the screens of the application and to define the flow that the user must follow in the prototype.
- Adobe Illustrator: It is a graphic design tool, which specializes in vector graphics. It was used for the creation of the logo and icons, which are minute details.
- Tiled Map Editor: It is a tool for the creation and editing of 2D maps, which is mainly used in games. The advantage of this tool is that it allows users to export in tile format, which makes it easy to integrate into Unity for the front-end development of the application. It was used to create scenarios for the games.

IV. Methodology

The design of the proposed mobile application is based on a study applied to university students, to evaluate the level of financial knowledge of young people. This section will detail the research approach, the instruments used for data collection, and how the results influenced the design decisions of the application.

A. Research Focus

This research is framed in a quantitative approach since it seeks to collect data numbers through a survey to assess the level of financial literacy among young people. This approach allows for obtaining objective information that facilitates the identification of gaps in financial knowledge and the foundation of the interactive and gamified learning proposal.

B. Population and Sample

The target population is university students from various undergraduate higher education institutions in Lima, Peru. Therefore, a sample of 150 university students was determined. This sample is composed of young people because they are in a transition stage towards financial independence, where the development of solid financial knowledge has a significant impact on their economic well-being.

C. Questionnaire

The questionnaire was designed in a semi-structured way, with a total of 50 questions. However, respondents only answer a subset of questions, as they must select their level of knowledge in areas such as saving, budgeting, investing, and personal finance. Based on the choice on a scale of 1-5, the level (basic, intermediate, advanced) is determined, and you are redirected to the financial challenges section associated with the selected level. These challenges are multiple-choice questions with a score of 20 points, which makes it easy to compare and analyze respondents' responses.

1) Data collection method: Data were collected through an online survey administered through Microsoft Forms. This survey was distributed to students through WhatsApp groups and the Viva Engage platform. The collection procedure was as follows:

- The survey was developed in Microsoft Forms and tested with a small group of students to ensure the clarity and relevance of the questions.
- Participation was promoted through a raffle published on the Viva Engage platform.
- Participants were informed about the purpose of the survey, the duration, and the confidentiality of their responses.
- Responses were collected over two weeks.

2) Structure of the questionnaire: The questionnaire is organized into six sections:

- Overview: Collects demographic data such as age, gender, career, college level, and the operating system (OS) of your mobile devices. Also, it informs the purpose of the questionnaire and the duration.
- Financial behavior: Analyzes the relationship between study and work, sources of income, and current practices of young people in terms of managing income, expenses, and savings.

- Financial Challenges: Inquire about the top financial challenges young people face and how often they experience difficulties managing their finances.
- Level of financial knowledge: Assesses the level of financial knowledge in key areas such as expenses, savings and investment, credits and debts; and personal finance. Respondents should rate their level of knowledge in each topic as basic (scale 1-2), intermediate (scale 3), or advanced (scale 4-5). This self-assessment allows students to be redirected to the financial challenges section according to the indicated level. These will facilitate a more detailed analysis of how their level of knowledge impacts their ability to face challenges.
- Preferences for educational tools: Inquire about preferences for educational tools to improve their financial literacy.
- Feedback and suggestions: Inquire about the difficulties experienced by respondents and suggestions for tools or features that should be added to a solution.

V. Results

In this section, the results of the questionnaire of young undergraduate university students will be presented.

A. Profile of Participants

A summary of the profile of the respondents is provided in Fig. 6. 54% of those surveyed are men, while 46% are women. Regarding the study, it is evident that the students come from a variety of university students in Lima, with a higher percentage from the Peruvian University of Applied Sciences with 66%, due to the origin of the research. Likewise, students are distributed in different academic cycles, with higher percentages in cycles VI (17%), VII (14%) and VIII (18%). On the other hand, the distribution of the OS used by the respondents is shown, which is relevant for the development of the solution proposal, where a large percentage (74%) uses the Android OS, which suggests that the platform should prioritize the development in this technology, to guarantee the greatest possible accessibility.

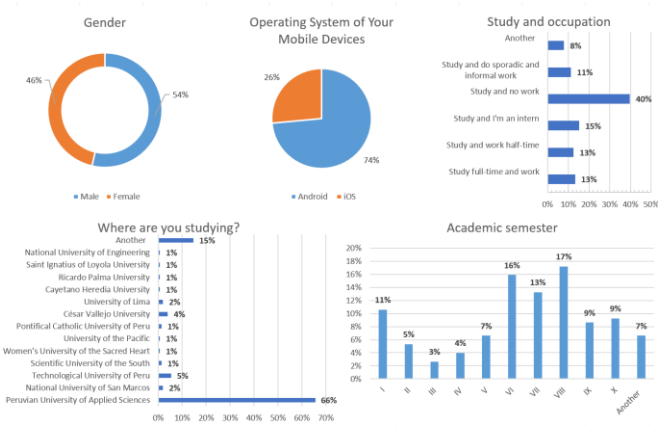


Fig. 6. Summary of the profile of respondents.

B. Financial Behavior

- Financial management: According to the results presented in Fig. 7, 44% of respondents do not keep formal control of their income and expenses, but do so mentally, which can generate uncertainty in their financial management. In contrast, only 8% use a financial management app, which allows them to have better control over their income and expenses. Likewise, 45% review their financial statements weekly, which reflects a positive financial habit. However, 11% of participants indicated that they rarely review their financial statements, which is worrying, as this lack of follow-up can lead to difficulties in managing their finances effectively.

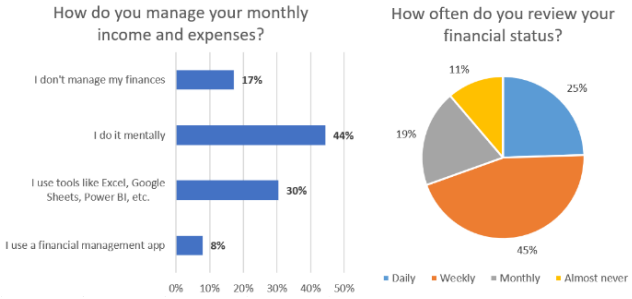


Fig. 7. Financial management.

- Expenses and savings: In Fig. 8, they indicate that 4% of respondents do not save anything, while 21% save less than 10% of their monthly income. 36% manage to save between 10% and 20%, and 23% save more than 20%. However, 16% do not know or have never calculated how much they save, which is worrying because they lack clear control over their savings capacity which could affect their financial stability. Regarding expenditure dedicated to entertainment and leisure, 40% allocate less than 10% of their income to this activity, while 35% allocate between 10% and 20%. Only 14% allocate more than 20% and 11% indicated that they do not know how much of their income they allocate to this activity. These data suggest that most students prioritize moderate spending on entertainment, although there is a large proportion who are not clear about their spending in this area.

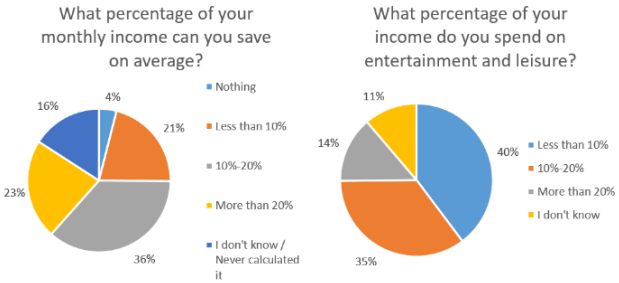


Fig. 8. Expenses and savings.

C. Financial Challenges

Regarding finance-related concerns, Fig. 9 shows that 30% of respondents expressed concern about not being able to

manage their finances efficiently. In addition, 20% worry about not being able to save enough, 20% do not understand how to invest, 17% are afraid of spending more than they earn and 13% accumulate debt. These results indicate that financial management is the main challenge for students, which could negatively impact their economic well-being.

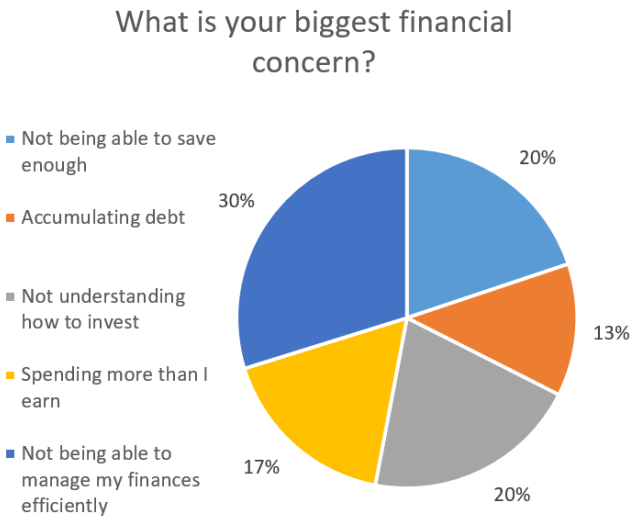


Fig. 9. Concern regarding finances.

D. Financial Literacy Level

In this section, we have worked with a multiple-choice question that uses a scale from 1 to 5, so that respondents can carry out a self-assessment of their level of financial knowledge. Scales 1 and 2 redirect them to basic financial challenges, scale 3 to intermediate-level challenges, and scale 4 and 5 to advanced-level challenges. A total of four questions are included, each corresponding to a specific topic, as shown in Table I.

TABLE I. TOPICS QUESTIONS

ID	Theme	Questions
Q1	Budget expenses	How would you rate your knowledge on the topic of Budgeting your expenses?
Q2	Savings and investment	How would you rate your knowledge about Savings and Investment?
Q3	Credit Cards and Debt	How would you rate your knowledge about Credit Cards and Debt?
Q4	Personal Finance	How would you rate your knowledge about Personal Finance in general?

Fig. 10 shows the level of knowledge by subject. Regarding budgeting expenses, it is shown that 34% of those surveyed consider themselves at a basic level, 44% at an intermediate level, and only 22% at an advanced level. On the other hand, about savings and investment, 38% consider themselves at the basic level. Meanwhile, 42% are classified at an intermediate level and only 20% at an advanced level. In the case of credit cards and debts, a large percentage (35%) are classified at the basic level, 47% are classified at the intermediate level and only 18% at the advanced level. Finally, about personal finance, 35% consider themselves to be at the basic level, 48% intermediate level, and 17% advanced. These results show a high proportion of young people with a level of financial knowledge at both the intermediate and basic levels, with a slight predominance at the intermediate level. However, a large percentage at the basic level is worrying, so it is necessary to strengthen knowledge and skills in these areas.

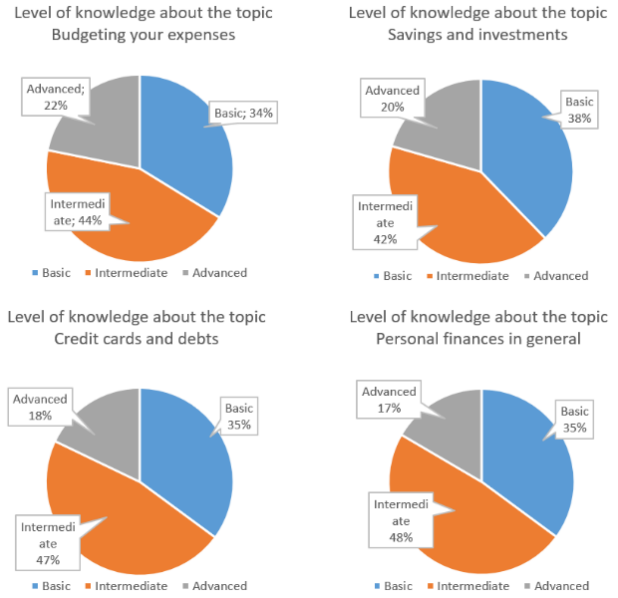


Fig. 10. Level of knowledge by topic.

E. Financial Challenges

Graphs were made to show the percentage of respondents who were correct in their answers about the financial issues defined. This allows the relationship between the level of knowledge selected and the results obtained in the financial challenges to be analyzed, reflecting how participants applied their knowledge. It should be noted that, in the graphs, the value "1" represents the correct answer, and the value "2" indicates the incorrect answer.

1) Budget expenses: Fig. 11 shows that, at the basic level, 63% of the participants were correct in their answers to the first question, while only 8% managed to get the second one right. At the intermediate level, 36% got question 1 right, but 67% got the second question right. At the advanced level, 88% were correct. This shows a clear correlation between financial knowledge and performance on challenges.

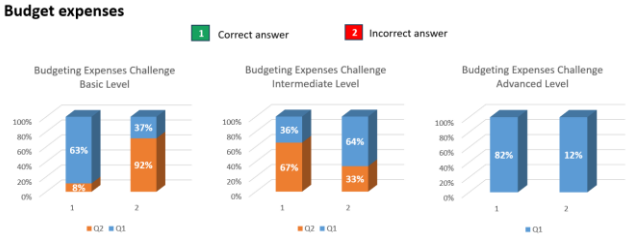


Fig. 11. Budget expenses.

2) Savings and investment: Fig. 12 At the basic level, both the first and second questions were corrected by 42% of the participants and 89% in the third. At the intermediate level,

48% got question 1 right and 44% got question 2 right. Finally, at the advanced level, 58% got the first question right and the second question was 71%.

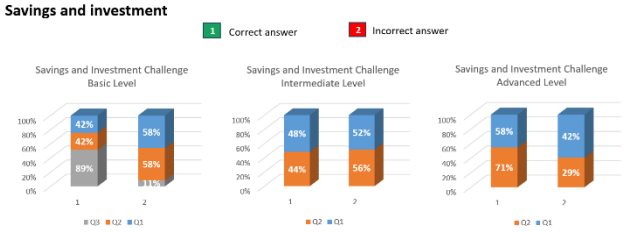


Fig. 12. Savings and investment.

3) Credit cards and Debt: In Fig. 13, 49% got the correct answer for the basic level right. At the intermediate level, 89% and 82% got questions 1 and 2 correctly respectively. In the case of the advanced level, 48% got question 1 right, while only 4% got the second one right.

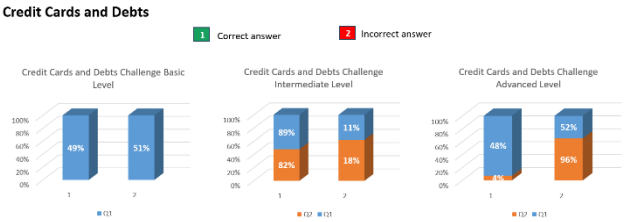


Fig. 13. Credit cards and debt.

4) Personal finance in general: In Fig. 14, 79% got the correct answer for the basic level, 90% for the intermediate level, and 64% for the advanced level.

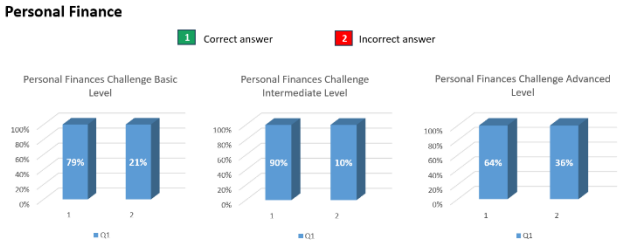


Fig. 14. Personal finance.

F. Financial Challenges Score

The scoring system for the financial challenges section consists of 20 points, regardless of the skill level selected. The rating scale used to evaluate performance is presented below (Table II). On the other hand, the average score obtained by the participants is 12.61 (Table II). This reveals a significant deficiency in the mastery of financial knowledge. Therefore, it is necessary to strengthen these areas to improve results.

TABLE II. RATING SCALE AND AVERAGE

Vigesimal qualification	*[0-10]*	*[11-14]*	*[15-19]*	*[20]*
Verbatim note	C: Fail	B: Regular	A: Expected Achievement	AD: Outstanding Achievement
Average Score	12.61			

5) Endnotes: In Fig. 15, only 5% of the respondents obtained an outstanding performance, while 32% failed.

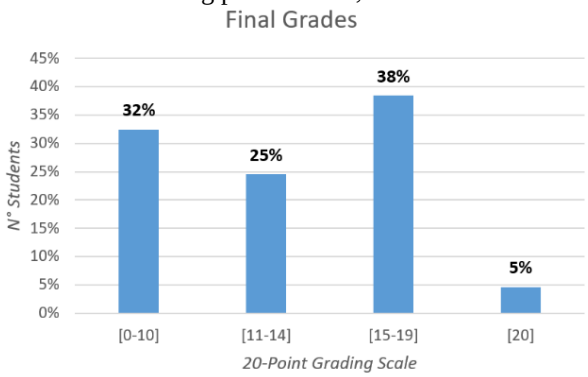


Fig. 15. Endnotes.

6) Notes per cycle: Fig. 16 shows a distribution of students in different academic cycles, revealing a remarkable trend: students in advanced cycles (VII-X) demonstrated a stronger understanding of financial concepts compared to those in initial cycles (I-III). This could be attributed to increased exposure to real economic situations as they progress in their career.

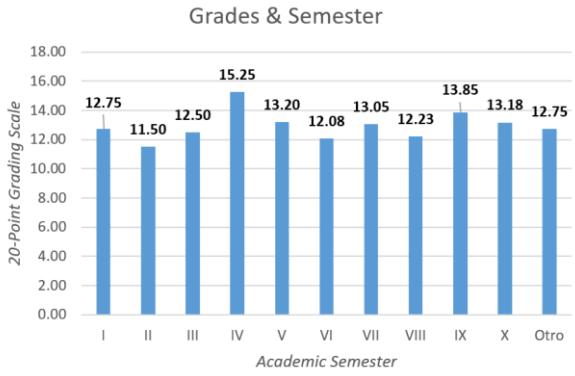


Fig. 16. Notes per cycle.

G. Tool Preferences

The survey results reveal that 81% of respondents prefer personalized advice to improve their financial knowledge. In terms of the learning format, interactive games were the most popular option with 52% of preferences. In addition, 81% showed interest in receiving personalized recommendations, which reinforces the idea of incorporating a system based on artificial intelligence for the financial coach, so that it offers suggestions adjusted to the individual needs and behaviors of users. (Fig. 17).

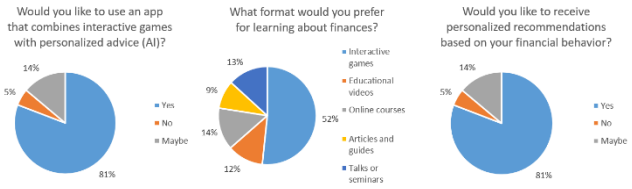


Fig. 17. Preferences.

VI. Discussion

The results obtained through the survey show the critical need to strengthen financial literacy among young university students. Most of the participants showed deficiencies in key concepts, particularly those found in the first cycles of university, while those in higher cycles showed a better understanding of financial topics (Fig. 16). This finding reinforces the importance of including financial training from the early stages of higher education to ensure that students acquire this knowledge before entering the labor market.

One of the most important findings was the preferred format for learning, where 52% opted for the interactive games option (Fig. 17). This suggests that young people value more dynamic learning methods, which validates the choice to propose a gamified solution. Since game-based learning can help increase students' engagement and enthusiasm for the topics, in addition, to achieving higher knowledge retention [17]. This result is related to [6], which found that promoting financial literacy through games increased interest and knowledge in financial topics.

In addition, 81% of respondents prefer to receive personalized recommendations based on their financial behaviors (Fig. 17). This leads to the integration of an AI-based recommendation system to offer suggestions based on users' needs and behaviors. This reflects a significant demand for solutions that teach financial concepts and provide recommendations. The study [8], which supplemented their serious game with an Intelligent Pedagogical Agent (IPA), which analyzed players' emotions and provided suggestions based on their progress in the game, thus succeeded in increasing students' competencies on Steam.

However, there were certain limitations throughout this research. First, the number of respondents, although the goal was to reach 385 participants, 150 were obtained. However, considering the focus of this study and the target population, this sample is still adequate to obtain representative results on the level of financial education. Secondly, the proposal has not been implemented or evaluated in a real environment, this limits the possibility of directly measuring the practical impact of the proposal on users' financial literacy. Finally, the proposal incorporates AI to personalize recommendations, but a detailed analysis of the ethical and privacy implications was not added, since the focus was on the evaluation of the level of financial literacy and the design of the proposal, however, it is relevant and should be considered in the implementation of this proposal.

Therefore, these results support the integration of a financial coach, designed to guide users based on their financial needs, with game-based learning to achieve a greater impact on the financial literacy of young people.

VII. Conclusion

This proposal for an interactive and gamified mobile application design is based on the results obtained in the survey of young people who prefer interactive games and personalized recommendations as learning formats. This preference highlights the importance of implementing teaching methods that promote continued interest in education.

On the other hand, it is important to note that the survey conducted not only made it possible to diagnose the financial capacity of young people but also served as the basis for the platform´s design. The data collected showed the need to strengthen financial literacy and the proposal seeks to respond to this need innovatively.

While this study does not include the practical implementation of the proposal, the design has the potential to contribute significantly to young people's financial literacy and have a positive impact on their ability to achieve financial independence.

Finally, this study presents an innovative solution based on interactive and gamified learning designed to reduce the financial literacy gap. It also promotes hands-on learning with a user-centered approach tailored to the user's preferences. Finally, it is suggested that future research work should focus on the implementation of this design proposal.

References

[1] A. Mueangpud, J. Khlaisang, and P. Koraneekij, “Mobile learning application design to promote youth financial management competency in Thailand,” International Journal of Interactive Mobile Technologies, vol. 12, pp. 19–38, 2019, doi: 10.3991/ijim.v13i12.11367.

[2] Superintendence of Banking, Insurance and AFPs (SBS) and CAF – Development Bank of Latin America, "Financial capabilities measurement survey 2022." Follero, 2022. [Online]. Available: https://www.sbs.gob.pe/Portals/4/jer/CIFRAS-ENCUESTA/2022/Brochure_ENCUESTA_CAPACIDADES%20FINANACIERAS%202022_vr.pdf.

[3] K. D. Pham, and V. L. T. Le, “Nexus between Financial Education, Literacy, and Financial Behavior: Insights from Vietnamese Young Generations,” SUSTAINABILITY, vol. 15, no. 20, 2023, doi: 10.3390/su152014854.

[4] T. Jia, C. Wang, Z. Tian, B. Wang, and F. Tian, “Design of Digital and Intelligent Financial Decision Support System Based on Artificial Intelligence,” Computational Intelligence and Neuroscience, 2022, doi: 10.1155/2022/1962937.

[5] N. Tretyakova, A. Lyzhin, E. Chubarkova, M. Uandykova, and M. Lukiyanova, “Mobile-Learning Platform for the Development of Entrepreneurial Competences of the Students,” International Journal of Interactive Mobile Technologies, vol. 15, no. 9, pp. 118–135, 2021, doi: 10.3991/ijim.v15i09.20225.

[6] L. Platz, and M. Jüttler, “Game-based learning as a gateway for promoting financial literacy – how games in economics influence students’ financial interest,” Citizenship, Social and Economics Education, vol. 21, no. 3, pp. 185–208, 2022, doi:10.1177/14788047221135343.

[7] M. Grijalvo, A. Segura, and Y. Núñez, “Computer-based business games in higher education: A proposal of a gamified learning framework,” Technological Forecasting and Social Change, vol. 178, 2022, doi: 10.1016/j.techfore.2022.121597.

[8] L. S. Ferro, F. Sapio, A. Terracina, M. Temperini, and M. Mecella, “ Gea2: A Serious Game for Technology-Enhanced Learning in STEM,” IEEE Transactions on Learning Technologies, vol. 14, no. 6, pp. 723–739, 2021, doi: 10.1109/TLT.2022.3143519.

[9] J. Zhang, “Computer Assisted Instruction System Under Artificial Intelligence Technology,” International Journal of Emerging

Technologies in Learning, vol. 16, no. 5, pp. 4–16, 2021, doi: 10.3991/ijet.v16i05.20307.

[10] A. Chatterjee, A. Prinz, M. Gerdes, S. Martinez, N. Pahari, and Y. K. Meena, "ProHealth eCoach: user-centered design and development of an eCoach app to promote healthy lifestyle with personalized activity recommendations," BMC Health Services Research, vol. 22, no. 1, 2022, doi: 10.1186/s12913-022-08441-0.

[11] Z. Wang, P. Veličković, D. Hennes, N. Tomašev, L. Prince, M. Kaisers, Y. Bachrach, R. Elie, L. K. Wenliang, F. Piccinini, D. Hassabis, and K. Tuyls, "TacticAI: an AI assistant for football tactics," Nature Communications, vol. 15, no. 1, 2024, doi: 10.1038/s41467-024-45965-x.

[12] Y. Zhao, J. Zhao, X. Hu, W. Zhou, and H. Li, "Full DouZero+: Improving DouDizhu AI by Opponent Modeling, Coach-guided Training and Bidding Learning," IEEE Transactions on Games, 2023, doi: 10.1109/TG.2023.3299612.

[13] E. Sinnewe, and G. Nicholson, "Healthy financial habits in young adults: An exploratory study of the relationship between subjective financial literacy, engagement with finances, and financial decision-making," Journal of Consumer Affairs, 57, no. 1, pp. 564–592, 2023, doi: 10.1111/joca.12512.

[14] A. Amagir, H. M. van den Brink, W. Groot, and A. Wilschut, "SaveWise: The impact of a real-life financial education program for ninth grade students in the Netherlands," Journal of Behavioral and Experimental Finance, vol. 33, 2022, doi: 10.1016/j.jbef.2021.100605.

[15] M. Fang, Y. Liu, C. Hu, J. Huang, and L. Wu, "An entrepreneurial education game for effectively Tracing the knowledge structure of college students - based on adaptive algorithms," Entertainment Computing, vol. 49, 2024, doi: 10.1016/j.entcom.2023.100632.

[16] H. Cecotti, M. Leray, and M. Callaghan, "Countdown VR: a Serious Game in Virtual Reality to Develop Mental Computation Skills," IEEE Transactions on Games, pp. 1–12, 2024, doi: 10.1109/TG.2024.3357452.

[17] R. Hare, Y. Tang, and S. Ferguson, "An Intelligent Serious Game for Digital Logic Education to Enhance Student Learning," IEEE Transactions on Education, 2024, doi: 10.1109/TE.2024.3359001.

[18] S. Mohanarajah, and T. Sritharan, "Shoot2Learn: Fix-and-play educational game for learning programming; enhancing student engagement by mixing game playing and game programming," Journal of Information Technology Education: Research, vol. 21, pp. 639–661, 2022, doi: 10.28945/5041.

[19] M. Arai, K. Tejima, Y. Yamada, T. Miura, K. Yamashita, C. Kado, R. Shimizu, M. Tatsumi, N. Yanai, and G. Hanaoka, "REN-A.I.: A Video Game for AI Security Education Leveraging Episodic Memory," IEEE Access, vol. 12, pp. 47359–47372, 2024, doi: 10.1109/ACCESS.2024.3377699.

[20] NNgroup, "The Changing Role of the Designer: Practical Human-Centered Design," Youtube [online video], June 5, 2020. Available: https://www.youtube.com/watch?v=QewRjNfG1-8&t=7s. [Accessed: October 22, 2024].

Comprehensive Evaluation of Machine Learning Techniques for Obstructive Sleep Apnea Detection

Alaa Sheta[1], Walaa H. Elashmawi[2], Adel Djellal[3], Malik Braik[4],
Salim Surani[5], Sultan Aljahdali[6], Shyam Subramanian[7], Parth S. Patel[8]
Department of Computer Science, Southern Connecticut State University, New Haven, CT, USA[1]
Department of Computer Science, Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt[2]
EEA Department, National Higher School of Technology and Engineering, Annaba, Algeria[3]
Department of Computer Science, Al-Balqa Applied University, Salt, Jordan[4]
Department of Pharmacy and Medicine, College Station, Texas A&M University, Texas, USA[5]
Computer Science Department, Taif University, Taif, Saudi Arabia[6]
Chief, Pulmonary, Critical Care and Sleep Medicine, Sutter Health, Sacramento, California, USA[7]
Department of Internal Medicine, Mayo Clinic, Rochester, MN, USA[8]

***Abstract*—Obstructive Sleep Apnea (OSA) is a prevalent health issue affecting 10-25% of adults in the United States (US) and is associated with significant economic consequences. Machine learning methods have shown promise in improving the efficiency and accessibility of OSA diagnoses, thus reducing the need for expensive and challenging tests. A comparative analysis of Logistic Regression (LR), Support Vector Machine (SVM), Gradient Boosting (GB), Gaussian Naive Bayes (GNB), Random Forest (RF), and K-Nearest Neighbors (KNN) algorithms was conducted to predict Obstructive Sleep Apnea (OSA). To improve the predictive accuracy of these models, Random Oversampling was applied to address the imbalance in the dataset, ensuring a more equitable representation of the minority class. Patient demographics, including age, sex, height, weight, BMI, neck circumference, and gender, were employed as predictive features in the models. The RFC provided outstanding training and testing accuracies of 87% and 65%, respectively, and a Receiver Operating Characteristic (ROC) score of 87%. The GBC and SVM classifiers also demonstrated good performance on the test dataset. The results of this study show that machine learning techniques may be effectively used to diagnose OSA, with the Random Forest Classifier demonstrating the best results.**

***Keywords*—Machine learning; obstructive sleep apnea; random forest classifier; oversampling; classification**

I. Introduction

Obstructive sleep apnea (OSA) is a prevalent disorder affecting a substantial portion of the population. Characterized by recurrent obstructions of the upper airway during sleep, it results in intermittent cessation of airflow [1], [2]. A multitude of factors have been identified as risk determinants for OSA, including obesity, male gender, smoking, age, craniofacial anomalies, and menopause in women [3]. Symptoms that suggest OSA include chronic snoring, observed apneic episodes, gasping during sleep, frequent awakenings, non-restorative sleep, increased nighttime urination, and excessive daytime sleepiness. Timely diagnosis of OSA is crucial as untreated OSA can contribute to the development of cardiovascular diseases, metabolic disorders, and neurocognitive impairments [4].

The standard diagnostic method for OSA is overnight polysomnography (PSG), however, it is expensive and often limited in accessibility. Therefore, it is important to prioritize high-risk individuals for PSG, particularly those with moderate to severe OSA, to optimize the utilization of sleep laboratories [5]. The severity of OSA is commonly assessed using the Apnea-Hypopnea Index (AHI), with cutoff values of $6-15/hour$ indicating mild OSA, $16-30/hour$ indicating moderate to severe OSA, and values exceeding $30/hour$ indicating severe OSA [5].

In recent years, machine learning has drawn considerable interest as a potentially effective way to address complex problems in various sectors, most notably healthcare. Its strengths, such as robustness, self-organization, adaptive learning, and parallel processing, make it an attractive tool.

Machine learning algorithms, renowned for their ability to discern patterns within intricate datasets, have garnered significant attention. Prominent examples of such algorithms encompass Support Vector Machines (SVM) [6], [7], Gradient Boosting Classifiers (GBC) [8], Gaussian Naive Bayes (GNB) [9], Random Forest Classifiers (RFC) [10], [11], and K-Nearest Neighbors Classifiers (KNC) [12]. Consequently, machine learning models have seen increasing use in medical healthcare, including the prediction of OSA, and have shown promising outcomes.

This study conducts a comparative evaluation of traditional regression modeling and a suite of machine learning algorithms for predicting obstructive sleep apnea (OSA) based on physical parameters. The succeeding sections of this paper are structured as follows. Section II presents a comprehensive review of existing machine learning approaches applied to OSA prediction. Section III provides a detailed characterization of the dataset employed in this study. In Sections IV and V, the training and evaluation methodology for ML models, including particular classification algorithms, is described. Finally, in Section VI, we present the experimental results and performance analysis concerning various evaluation metrics and report the results obtained through this research endeavor.

II. Related Works

Machine learning has emerged as a preeminent paradigm for classifying medical data owing to its ability to manage

and extract insights from voluminous and intricate datasets effectively. For instance, in the context of OSA diagnosis, traditional methods like drug-induced sleep endoscopy (DISE) rely on subjective observer evaluations, as seen in the VOTE classification system proposed by Altintas et al. [13], which showed only moderate to fair agreement among observers. This highlights the potential of machine learning models to provide objective, consistent, and accurate diagnoses by automating the analysis of complex data, thereby addressing the limitations of observer-dependent methods.

Various studies [14]–[16] have utilized machine learning algorithms to improve diagnostic precision and patient outcomes in various diseases. The authors in [17], highlighted the utility of support vector machines (SVM) in classifying brain MRI images for neurological disorders, which aligns with its potential for OSA diagnosis. However, the authors in [18] explored the use of various ML methods to predict obstructive sleep apnea syndrome (OSAS) severity using demographic, clinical, and spirometric data from 313 patients. Their study demonstrated that SVMs and Random Forests (RFs) showed the best classification performance.

Furthermore, ensemble learning methods, such as Gradient Boosting Classifiers (GBCs) and RF classifiers, have demonstrated robust performance in medical diagnostics due to their ability to handle imbalanced and high-dimensional data. Ramesh et al. [8] applied GBCs to classify OSA from electronic health records, achieving an accuracy of 68.06%. The authors in [11] further validated the performance of RFCs for OSA detection, leveraging feature selection algorithms to improve model interpretability and accuracy.

Integrating the STOP-BANG questionnaire with machine learning models is employed in [12] to enhance OSA diagnosis. Among the four algorithms tested, the K-Nearest Neighbor (K-NN) model demonstrated the best performance, achieving 94% accuracy. The results highlight the potential of combining ML with traditional tools to improve the reliability and efficiency of OSA screening. However, in [19], the SLEEPS model, a machine learning-based questionnaire using nine items, accurately predicts OSA, COMISA, and insomnia without polysomnography. The model trained on over 4,600 participants using XGBoost, achieved AUROC values above 0.89, outperforming tools like STOP-BANG.

In recent years, the authors in [20] have used the Swedish National Study on Aging and Care electronic health data to predict sleep apnea with a ML model. The XGBoost and Bidirectional Long Short-Term Memory Networks modules give the model 97% accuracy with 75 features and 10,765 samples. Furthermore, pre-screening symptoms were employed to diagnose OSA [21], and the experimental findings revealed that the Decision Tree Classifier (DTC) and RF outperformed other comparable algorithms, achieving the highest classification accuracies. Similarly to the authors in [22], the RF classifier technique is utilized to predict sleep disorders and has achieved the highest accuracy. The potential of ML models for cost-effective OSA screening, with RF and LightGBM showing the most promise for clinical use, is discussed in [23].

A concise overview of the utilization of machine learning techniques in diagnosing, classifying, and treating sleep-related respiratory problems is presented in [24]. The effectiveness of machine learning-based classifiers in OSA classification is highly affected by the quality and quantity of input data and the selection of the machine learning approach.

More advanced techniques, like deep learning models, have demonstrated superior accuracy in sleep apnea and related disorder detection. Studies like [10] combined ECG signals with machine learning and deep learning to achieve 86.25% accuracy, while advanced architectures like multi-resolution residual network (MR-ResNet) [6] and CNN-based approaches [9] reached accuracies of 90.8% and 79.61%, respectively, leveraging polysomnographic (PSG) data. Wearable systems [25] and single-lead electrocardiogram (ECG) classifiers [7] achieved notable performance with accuracies up to 88.2% and 93.0%. Other studies, such as [26], employed EEG and/or electrooculogram (EOG) signals for sleep staging, achieving up to 84.5% accuracy, while [27] explored microelectromechanical system (MEMS)-based solutions. However, the computational demands and reliance on specialized data limit the practicality of these methods in resource-constrained environments.

Despite substantial progress in sleep apnea diagnosis, pursuing more accurate, efficient, and accessible methods remains an active area of research. Many existing studies rely on costly diagnostic tools like PSG, need help with imbalanced datasets, and focus on computationally intensive deep-learning models that lack practicality and generalizability. This study addresses these challenges by using easily obtainable physical parameters, applying oversampling to balance data, and evaluating resource-efficient algorithms, offering scalable and accessible solutions for diverse populations.

III. OSA Data Collection

Data were collected from adult individuals who were referred to a community sleep center due to suspected obstructive sleep apnea (OSA) and had not received a previous diagnosis. Each participant provided demographic details such as age, gender, and ethnicity, and completed sleep-related questionnaires, including the Epworth Sleepiness Scale. Prior to undergoing polysomnography, a physical examination was conducted, which included an assessment of the airway using a modified Friedman grade, measurement of body mass index (BMI), and neck circumference (NC). OSA was diagnosed when the apnea-hypopnea index (AHI) was equal to or greater than 15. Incomplete questionnaires or polysomnography records with inadequate technical quality or insufficient total sleep time were excluded from the analysis.

Retrospective data analysis and review were conducted after the administrative approval of the data from the Torr Sleep Center executive and institutional committee. The study was conducted at Torr Sleep Center in their Corpus Christi, Texas Location. The patients undergoing the second or split night (titration night) were excluded. The patients have undergone the first (diagnostic night were determined by the computerized search using the CPT code 98510. The patient discussed the study with the registered polysomnographic therapist during the presentation. Baseline demographic information was collected. Height, weight, Modified Friedman, Waist circumference, and diameter were assessed, and Body Mass Index (BMI) was computed. The patient underwent overnight polysomnography. The nocturnal polysomnogram (NPSG) included the

recording of electroencephalogram (EEG): F4-M1, F3-M2, C4-M1, C3-M2, O2-M1, O1-M2, electrooculogram (EOG), submental, intercostal and anterior tibialis electromyogram (EMG), electrocardiogram (EKG), airflow by nasal pressure, and oral thermistor, abdominal and chest wall excursion using impedance plethysmography and oxygen saturation by pulse oximetry attended by a sleep technologist. The sleep study was staged and scored using AASM standards.

All methods described in this study adhere to relevant guidelines and regulations. The research protocol obtained ethical approval from the Research Ethics Committee of NTUH. (protocol number 201603113RIND), and the committee waived the requirement for participant consent. Table I presents dataset statistics, including sample size, mean values, and ranges. Additionally, Table II displays randomly selected data samples to provide insight into the dataset's structure.

Fig. 1 depicts the distribution of the dataset across OSA, Sex, and BMI attributes. These attributes play a crucial role in the modeling process. Notably, Body Mass Index (BMI) is one of these attributes. BMI is a measure to estimate body fat using an individual's height and weight. It is a standard adult screening tool for weight-related health problems. According to the classifications given in Table III, it sorts people into four groups: underweight, normal weight, overweight, or obese.

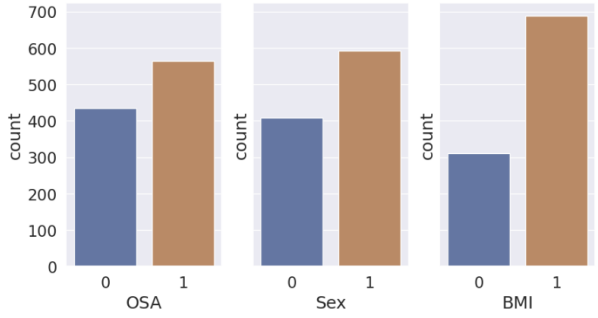


Fig. 1. Distribution of some attributes of the data: OSA, Sex, and BMI.

Fig. 2 shows the dataset box plot, and Fig. 3 shows a correlation between dataset parameters; it can be seen that the most correlated parameter with OSA is Neck circumference and then weight, which means that these two parameters will play a significant role in classification results.

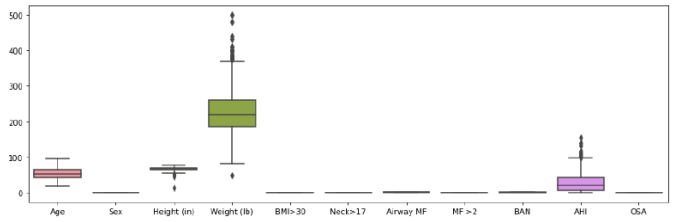


Fig. 2. Dataset Box plot.

A. Oversampling

The OSA data set is imbalanced; we adopted a Random oversampling technique to balance the data. RandomOver

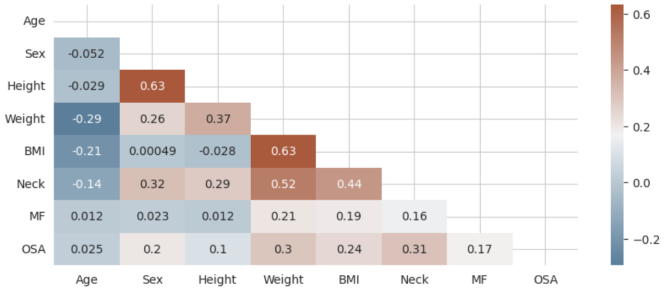


Fig. 3. Correlation heatmap.

Sampler is a machine learning technique used to handle imbalanced datasets. It addresses class imbalance by creating a more equally distributed dataset by randomly oversampling the minority class as follows:

- First, the minority class is identified in the dataset. This is the class with fewer instances than the majority class.
- Then, the "RandomOverSampler" algorithm randomly selects instances from the minority class. It oversamples the minority class by duplicating its instances to match the population of the majority class.
- After the oversampling, the resulting dataset is balanced or nearly balanced, with an equal number of instances for each class.
- Finally, a machine learning model is trained using the balanced dataset; the model should outperform its counterpart trained on an unbalanced dataset.

Oversampling techniques, such as Random Oversampling and Synthetic Minority Oversampling Technique (SMOTE), have been widely used to address this issue. The authors in [24] employed SMOTE to enhance the performance of ML models in detecting OSA, demonstrating its effectiveness in reducing classification errors for minority classes. It's worth noting that the Oversampling can be prone to overfitting, especially if the minority class is oversampled too much. Tuning the oversampling ratio is essential to finding the optimal balance between reducing class imbalance and avoiding overfitting.

B. Data Scaling

Data scaling is an essential pre-processing stage for the machine learning classification process. The purpose of scaling the data in this way is to make it easier to compare features with different units and scales. This can potentially enhance the efficiency of machine learning algorithms, especially those like k-nearest neighbors and support vector machines that are sensitive to input data size.

The "StandardScaler" is a preprocessing technique used in machine learning to scale numerical data. Standardization is achieved by transforming the data to exhibit a zero mean and unit variance. Specifically, it can be done according to the following steps.

- First, the mean of each feature (column) in the data is calculated.

TABLE I. STATISTICS OF THE OSA DATA SET

	Age	Sex	Height (in)	Weight (lb)	BMI>30	Neck>17	Airway MF
count	1000	1000	1000	1000	1000	1000	1000
mean	54.031	0.592	67.037	227.653	0.689	0.661	2.722
std	14.320	0.492	4.624	58.218	0.463	0.474	1.008
min	19.000	0.00	15.000	49.000	0.000	0.000	0.000
25%	44.000	0.000	64.000	186.750	0.000	0.000	2.000
50%	54.000	1.000	67.000	220.000	1.000	1.000	3.000
75%	65.000	1.000	70.000	262.000	1.000	1.000	4.000
max	96.000	1.000	79.000	500.000	1.000	1.000	4.000

TABLE II. SAMPLE OF THE OSA DATA SET

Sample No.	Age	Sex	Height (in)	Weight (lb)	BMI>30	Neck>17	Airway MF
652	45	1	73.0	284	1	1	1
939	71	0	59.0	192	1	0	0
319	51	1	77.0	280	1	1	1
626	41	1	66.0	180	0	1	0
808	85	0	59.0	178	1	1	1

TABLE III. BMI CATEGORIES

BMI Category	BMI Range
$Under\_weight$	< 18.5
$Normal$	$18.5 - 24.9$
$Over\_weight$	$25 - 29.9$
$Obese$	≥ 30

- Then, the StandardScaler subtracts the mean from each value in the feature. This centers the data around zero.
- Next, the StandardScaler divides each value in the feature by its standard deviation. This scales the data to have a standard deviation of 1.
- After the scaling is done, the resulting dataset has a mean value of zero and a standard deviation value of one.

IV. PROPOSED METHODOLOGY

The machine learning process consists of multiple steps, beginning with data collection. The collected data might be raw and unstructured, so the subsequent step involves pre-processing. In this step, missing data is eliminated and thoroughly cleaned to ensure its suitability for analysis.

Following the pre-processing stage, the next step is feature extraction, which involves identifying and extracting pertinent features from the data. This step is significant since the ML model's performance relies heavily on the quality of the extracted features. After extracting the features, the dataset was divided into several subsets for training and testing. The training subset was used to develop the model, while the testing subset was used to evaluate how well the model can be applied to new data. The subsequent step is classification, where the machine learning algorithm utilizes the extracted features to classify new data into distinct categories. Various classification algorithms, such as RF, SVM, or ANN, can be employed to train the model on the training set.

Finally, precision, recall, and F1-score metrics evaluate the model's performance. The model exhibiting the highest performance is chosen and employed for classifying new data. It should be emphasized that this process is iterative and typically requires multiple iterations of adjustments and enhancements to attain optimal performance. Algorithm 1 summarizes all the process stages of machine learning. Fig. 4 shows the proposed methodology utilized for OSA detection using various machine-learning techniques.

Algorithm 1: Machine Learning Process

```
Input: Raw data
Output: Trained model
Step 1: Data Collection
    Collect raw data from various sources
Step 2: Data Pre-processing
    Remove missing data and outliers
    Normalize or scale the data if necessary
Step 3: Feature Extraction
    Extract relevant features from the pre-processed data
    Reduce the dimensionality of the data if necessary
Step 4: Model Selection
    Choose appropriate machine learning algorithms
    Select hyperparameters for the algorithms
Step 5: Model Implementation
    Train the selected models on the pre-processed data
    Evaluate the performance of the trained models
Step 6: Model Evaluation
    Test the trained models on new data
    Evaluate the performance of the tested models
Step 7: Model Deployment
    Deploy the best-performing model in production
```

V. ML METHODS

A. Logistic Regression (LR)

Logistic regression is a statistical technique that shares similarities with linear regression and is utilized for predicting binary outcomes. Unlike the Mantel-Haenszel odds ratio, which is limited to discrete explanatory variables, logistic regression can simultaneously handle continuous and

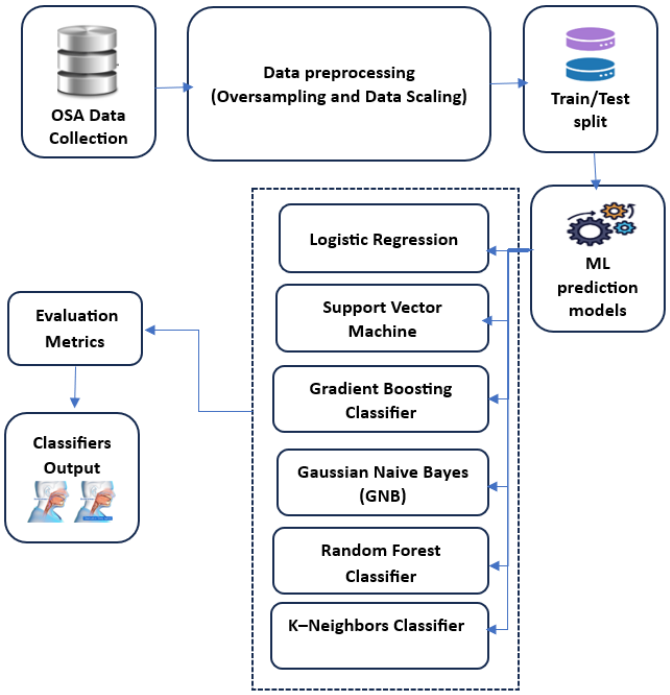


Fig. 4. The proposed methodology for utilizing various ML models to detect OSA.

multiple explanatory variables. This capability is essential when studying the impact of various factors on the response variable. Logistic regression models the probability of an outcome by considering the covariance among variables and accounting for individual characteristics. This approach helps address confounding effects when analyzing multiple variables independently. The logarithm of the odds is used in modeling, as odds represent a ratio, as explained by Sperandei [28].

$$log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m \qquad (1)$$

The symbol π represents the likelihood of an event, such as the incidence of OSA. The regression coefficients β_i correspond to the reference group and the explanatory factors x_i.

B. Support Vector Machines (SVM)

The Support Vector Machine (SVM) is a widely used supervised learning model for prediction and classification tasks. It was developed in 1995 by Vladimir Vapnik and his team at AT&T Bell Laboratory. SVM utilizes a nonlinear mapping function to transform the training data set into a higher dimensional space. It then employs linear regression to separate the data within this transformed space. This approach has demonstrated effectiveness across various applications, enabling the learning of complex decision boundaries and improving classification accuracy. The author in [29] described this process of SVM as approximating the training data set within a higher dimensional space and employing linear regression to separate the data. Fig. 5 illustrates the SVM model.

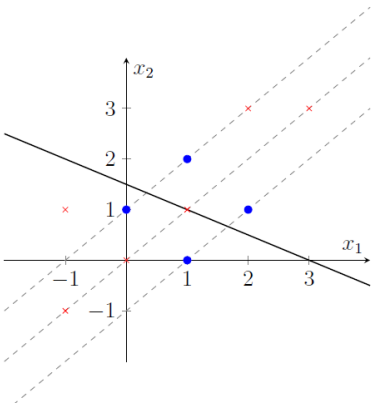


Fig. 5. Optimal hyperplane in support vector machine.

C. Gradient Boosting Classifier (GBC)

As an ensemble machine learning approach, the Gradient Boosting Classifier combines many weak models into one more potent model, increasing the prediction power of the combined model [30]. It operates iteratively by training decision trees on the residuals of the preceding tree, utilizing gradient descent optimization to minimize the loss function. This technique enables the algorithm to learn more intricate decision boundaries, improving prediction accuracy. The specific structure of the algorithm, including its formulas, is highly influenced by the chosen designs of $\Phi\,(y, f)$ and $h\,(x, \theta)$. More detailed examples of these algorithms can be found in the work of Friedman [30].

Algorithm 2: Friedman's Gradient Boost Algorithm

Input: Training Dataset $\mathcal{D} = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, number of iterations M, learning rate α, base model $h_0(\mathbf{x})$, loss function $L(y, F(\mathbf{x}))$

Output: Ensemble model $F(\mathbf{x}) = \sum m = 1^M \beta_m h_m(\mathbf{x})$

Initialize ensemble model $F_0(\mathbf{x}) = h_0(\mathbf{x})$;

for $m \in 1, \ldots, M$ **do**

  Calculate the negative gradient $r_{im} = -\left[\frac{\partial L(y_i, F_{m-1}(\mathbf{x}i))}{\partial Fm-1(\mathbf{x}i)}\right] i = 1^n$ for each training instance $\mathbf{x}i$

  Fit a base model $h_m(\mathbf{x})$ to the negative gradient r_{im}

  Compute the optimal step size $\beta_m = \arg\min_\beta \sum_{i=1}^n L(y_i, F_{m-1}(\mathbf{x}_i) + \beta h_m(\mathbf{x}i))$

  Update the ensemble model: $F_m(\mathbf{x}) = Fm - 1(\mathbf{x}) + \alpha\beta_m h_m(\mathbf{x})$;

end

return *Ensemble model* $F(\mathbf{x}) = \sum_{m=1}^M \beta_m h_m(\mathbf{x})$

The Gradient Boosting is used for creating an ensemble model on a training set $\mathcal{D}$ consisting of n instances, where each instance has a pair of features $\mathbf{x}i$ and label y_i. It requires several iterations M, a learning rate α, a base model $h_0(\mathbf{x})$, a loss function $L(y, F(\mathbf{x}))$ to evaluate the quality of the ensemble model, and a set of hyper-parameters for the base model. The algorithm initializes the ensemble model to $F_0(\mathbf{x}) = h_0(\mathbf{x})$, and then iteratively improves it by fitting

a base model $h_m(\mathbf{x})$ to the negative gradient of the loss function concerning the current ensemble model $Fm-1(\mathbf{x})$. The optimal step size β_m is computed using line search, and the ensemble model is updated by adding a scaled version of the new base model $h_m(\mathbf{x})$ to the previous ensemble model $F_{m-1}(\mathbf{x})$. The final output of the algorithm is the resulting ensemble model $F(\mathbf{x})$.

D. Gaussian Naive Bayes (GNB)

One machine learning algorithm based on the concepts of Bayes' Theorem is the Naive Bayes Classifier [31]. These classifiers rely on the assumption of strong independence among the features used for predictions.

Under this premise, it is assumed that the value of one characteristic does not affect the value of any other feature. A notable advantage of Naive Bayes Classifiers is their ability to be efficiently trained in supervised learning scenarios, even when working with limited training data. Moreover, their straightforward design and ease of implementation make them popular for various real-world applications.

In machine learning, continuous data is frequently assumed to adhere to a normal (Gaussian) distribution, mainly when dealing with classification tasks. This assumption suggests that the continuous values corresponding to each class follow a normal distribution. By making this assumption, it becomes possible to estimate the likelihood of the features using the Gaussian probability density function:

$$P\left(x_i|y\right) = \frac{1}{\sqrt{2\pi\sigma_y^2}} exp\left(-\frac{\left(x_i-\mu_y\right)^2}{2\sigma_y^2}\right) \tag{2}$$

One strategy for constructing a simple model is to assume that a Gaussian distribution with no covariance among dimensions can characterize the data. In other words, each dimension is considered independent of the others. This type of model can be easily built by calculating the mean and standard deviation of the data points within each label, as these parameters define the distribution.

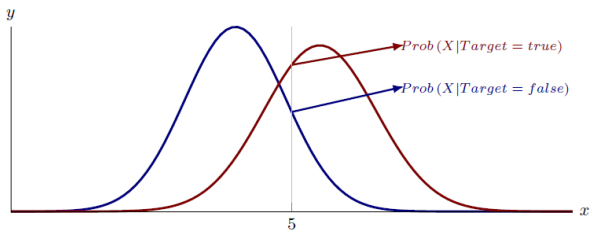


Fig. 6. Demonstration of working with a continuous variable in naïve bayes.

The provided illustration, as shown in Fig. 6 demonstrates the functioning of a Gaussian Naive Bayes (GNB) classifier. For each data point, the classifier calculates the z-score distance, which is the difference between the data point and the mean of each class divided by the standard deviation of that class.

E. Random Forest Classifier (RFC)

The term "random forest" refers to an ensemble of tree predictors. A randomly distributed vector, which is sampled individually and distributed uniformly among all trees in the forest, is relied upon by each individual tree [32]. As the random forest expands in terms of the number of trees, its generalization error stabilizes. The predictive capacity of individual trees and their interrelationships impact the ultimate level of this error.

Random forests use a random feature selection method for splitting nodes, which leads to error rates comparable to Adaboost while being more resilient to noise. The forest's internal estimates monitor various factors such as error, strength, and correlation and can assess the impact of increasing the number of features used for splitting. These estimates are also useful in determining the importance of different variables, and the approach is applicable to regression tasks as well.

RFC constructs decision trees by randomly selecting subsets of the training data and features for each node [33], [34]. Finding the optimal feature to divide the data at each node is how the trees are iteratively developed until a stopping requirement is satisfied [35]–[37]. During prediction, the ensemble of trees votes on the class label for a new input instance, with the class receiving the most votes being predicted as the output. This approach mitigates overfitting and enhances classification accuracy by leveraging the collective ability of the tree ensemble to capture diverse patterns and relationships within the data. RFC has demonstrated successful applications in various domains, including sleep apnea research [11], [38]. Refer to Fig. 7 for an illustrative example of the RFC mechanism.

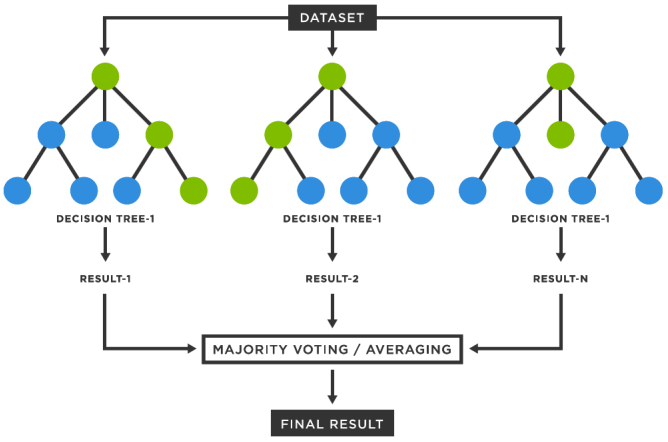


Fig. 7. Random forest mechanism.

To train individual trees, the Random Forest Algorithm requires a training set $\mathcal{D}$, a certain number of trees to be included in the forest (T), a random number of features (K) to be chosen for each tree, and a decision tree algorithm $\mathcal{A}$. To build the forest, the algorithm iteratively constructs T trees. At each iteration t, the algorithm randomly selects K features from the available features and draws a bootstrap sample $\mathcal{D}_t$ from the training set $\mathcal{D}$. A decision tree f_t is then trained on the sampled features and the bootstrap sample $\mathcal{D}_t$, using

Algorithm 3: Random Forest Classifier Algorithm

Input: Training Dataset $\mathcal{D}$, number of trees T, number of features K, decision tree algorithm $\mathcal{A}$
Output: Random Forest $\mathcal{F}$
for $t \in 1, \ldots, T$ **do**
  Sample K features from the p available features without replacement
  Draw a bootstrap sample $\mathcal{D}_t$ from $\mathcal{D}$
  Train a decision tree f_t on $\mathcal{D}_t$ using the selected features and $\mathcal{A}$
end
return *Random Forest* $\mathcal{F} = f_1, \ldots, f_T$

the given decision tree algorithm $\mathcal{A}$. The final output of the algorithm is the resulting random forest $\mathcal{F}$, which consists of the T decision trees $f_1, \ldots, f_T$ (see Algorithm 3).

F. K–Neighbors Classifier (KNN)

One supervised machine learning technique that is commonly employed for classification problems is the K-Nearest Neighbors (KNN) classifier. It sorts unlabeled data points according to the similarity principle, which states that it should consider the class of nearby data points in the training dataset. The number of neighbors to consider is represented by the "K" in KNN.

The algorithm computes the Euclidean distance between the unclassified data instance and each labeled training instance to inform the classification decision. After that, it uses the distances to choose the K closest neighbors. The unlabeled data point's class identification is decided by a majority vote among its K nearest neighbors. KNN is a simple and intuitive algorithm that does not require training. It uses the entire training dataset for classification. The KNN algorithm is easy to understand, and its pseudocode is provided in Algorithm 4.

In this algorithm, the input is a training set $\mathcal{D}$ consisting of labeled instances, a test instance x that we want to classify, and the number of neighbors K to consider. The output is the predicted class label for the test instance.

Algorithm 4: K-Nearest Neighbors Classifier Algorithm

Input: Training Dataset $\mathcal{D} = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, new instance $\mathbf{x}$, number of neighbors K
Output: Predicted label $\hat{y}$ for $\mathbf{x}$
for $i \in 1, \ldots, n$ **do**
  Compute the Euclidean distance $d(\mathbf{x}, \mathbf{x}_i)$ between $\mathbf{x}$ and each training instance $\mathbf{x}_i$;
end
Identify the K training instances with the smallest distances to $\mathbf{x}$; Assign the majority class label among these K instances as the predicted label $\hat{y}$ for $\mathbf{x}$;
return *Predicted label* $\hat{y}$

Each instance in the training set $\mathcal{D}$ has a pair of features $\mathbf{x}_i$ and a label y_i, and the K-Nearest Neighbors Classifier Algorithm is applied to this set. The algorithm calculates the Euclidean distance between each training instance $\mathbf{x}_i$ and $\mathbf{x}$ when given a fresh instance $\mathbf{x}$. Next, it chooses the K training examples that are closest to $\mathbf{x}$, and the projected label $\hat{y}$ for $\mathbf{x}$ is the majority class label among these K examples. The algorithm returns the predicted label $\hat{y}$.

VI. Experimental Results and Performance Analysis

The following section presents the findings from utilizing the proposed machine learning models in the OSA dataset. The performance of each model is evaluated using a comprehensive set of metrics, and a comparative analysis is conducted to identify the most effective approach. These metrics provide quantitative measures of model quality. Most of them mainly depend on calculating TP (i.e. count of model-correct positives), FP (i.e. number of positive cases misclassified as negative), TN (i.e. the number of true negatives the model identified), and FN (i.e. the model predicts a negative result while it is positive). The total number of instances is $Total$ (i.e. $Total = TP + TN + FP + FN$). These metrics include accuracy, precision, recall, and F1-score, as computed using the following formulas.

$$Accuracy = \frac{TP + TN}{Total} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

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

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} \tag{6}$$

Table IV provides a comparative analysis of various classification algorithms (LR, SVM, GBC, GNB, RFC, and KNC) based on their performance metrics: accuracy, precision, recall, and F1-score, calculated for both training and testing datasets. Our findings indicate that:

- The analysis results show that RFC has the highest performance on the training set across all metrics, with a training accuracy of 86.78%, precision of 92.64%, and F1-score of 87.80%. However, its testing performance drops significantly (accuracy of 65.02%).
- The GBC achieves relatively balanced performance between training and testing datasets, with a testing accuracy of 66.08% and the highest test F1-score of 67.79% among the models, indicating better generalization compared to RFC and other models except KNC.
- The SVM shows testing accuracy (64.66%) and moderate performance across metrics. However, its performance on the training dataset is slightly lower than that of RFC, KNC, and GBC.
- KNC demonstrates balanced and consistent performance, achieving a training accuracy of 74.14% and testing accuracy of 67.14%. However, its recall on the test set decreases to 62.59%.

TABLE IV. CALCULATED PERFORMANCE CRITERIA BASED ON VARIOUS ML MODELS

Method	Train Accuracy	Train Precision	Train Recall	Train F1	Test Accuracy	Test Precision	Test Recall	Test F1
LR	0.665880	0.666667	0.698851	0.682379	0.650177	0.596273	0.738462	0.659794
SVM	0.704841	0.706935	0.726437	0.716553	0.646643	0.594937	0.723077	0.652778
GBC	0.822904	0.807775	0.85977	0.832962	0.660777	0.60119	0.776923	0.677852
GNB	0.645809	0.675862	0.649007	0.662162	0.650177	0.707692	0.601307	0.650177
KNC	0.741440	0.763218	0.741071	0.751982	0.671378	0.707692	0.625850	0.664260
RFC	0.867769	0.926437	0.834369	0.877996	0.650177	0.753846	0.593939	0.664407

- LR achieves moderate training accuracy (66.59%) and a balanced F1 score (68.24%). However, in the test, it maintains a consistent performance accuracy of 65.02% but suffers from lower precision (59.63%), indicating a higher false positive rate compared to algorithms such as KNC and GBC.
- The GNB has the lowest accuracy on the training dataset, with a value of 64.58%, but has a high test precision of 70.77%. This highlights the model's ability to correctly classify positive cases, despite lower overall accuracy and recall.

The results align well with existing literature on applying ML models for OSA diagnosis. Consistent with prior studies, the RFC emerged as the best-performing model regarding training accuracy (87%), reflecting its robustness in handling complex datasets and its effectiveness as highlighted in studies like [8], [10]. Similarly, the GBC demonstrated strong generalization capabilities, achieving a balanced performance across metrics, which is in line with findings in [8], where GBC was noted for its ability to capture intricate patterns in data. Overall, the alignment between this study's findings and existing research underscores the validity of these ML models for OSA prediction.

Furthermore, confusion matrices were generated to gain a comprehensive understanding of each machine learning algorithm's predictive capabilities. These visual representations offer a detailed breakdown of correct and incorrect classifications. Fig. 8 through 13 provide a graphical depiction of these training and testing data results, enabling a thorough analysis of each model's performance characteristics.

From Fig. 8, the total number of correctly predicted OSA cases is 564 out of 847 (i.e. in the case of training) and 184 out of 283 (i.e., in the case of testing). However, the percentage of incorrectly classified instances is 33.41% and 34.98% in training and testing, respectively; the LR model shows a moderate performance.

Fig. 9 demonstrates the strong overall performance of the SVM model, with a significantly higher number of correct predictions (diagonal elements) compared to incorrect predictions (off-diagonal elements). The model seems balanced in predicting classes 0 and 1, with a relatively even distribution of correct predictions for each class. The values in the off-diagonal (119 and 131) in training and (36 and 64) in testing indicate relatively low rates of false positives and false negatives, suggesting that the model effectively distinguishes between the two classes.

According to the GBC confusion matrix (as shown in Fig. 10), the GBC model exhibits strong overall performance, with a significantly higher number of correct predictions of OSA instances equal to 697 (i.e. 82.92%) and 187 (i.e. 66.08%) in training and testing, respectively.

The GNB classifier demonstrates solid performance on the training and testing sets, as shown in Fig. 11. The GNB model correctly classified 294 cases as positive OSA and 253 as negative cases in the case of training. In testing, the overall number of correctly classified cases is 184 out of 283 (i.e. 65.02%). While there's a minor decrease in performance on the testing set, the model still maintains a good balance in predicting both classes.

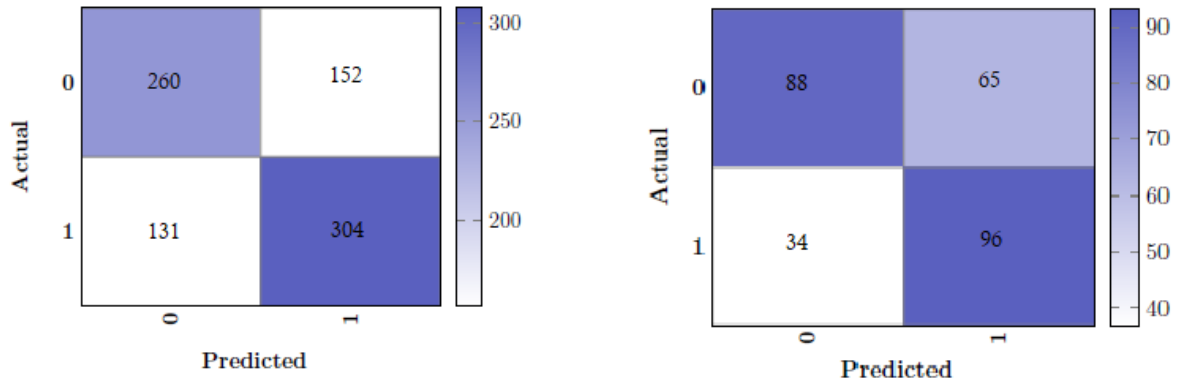
According to the confusion matrix of the random forest classifier (as shown in Fig. 12), the number of correctly classified instances is 735 (TP+TN) in training and 184 (TP+TN) in testing, while the total number of misclassified is 112 (FP+FN) in training and 99 (FP+FN) in testing. However, in Fig. 13, the K–Neighbors classifier, the model Correctly predicted 332 instances as class 1, 296 as class 0, 116 incorrectly predicted as class 1, and 103 incorrectly predicted as class 0 in training. In testing, The model achieves moderate results of 67.14%, indicating that it correctly predicts the class in 67.14% of cases.

Each methodology's efficacy depends on complex factors, including the problem domain, dataset characteristics (size and quality), and computational constraints. Moreover, the receiver operating characteristic (ROC) curve visually shows a binary classifier's performance. The area under the ROC curve (AUC) estimates the general model performance. A higher AUC denotes better discriminative capability; a perfect model achieves an AUC of 1.0, while a random classifier generates an AUC of 0.5. Fig. 14 and 15 present the ROC curves and box plots for the respective classification algorithms.

A. Statistical Test Analysis

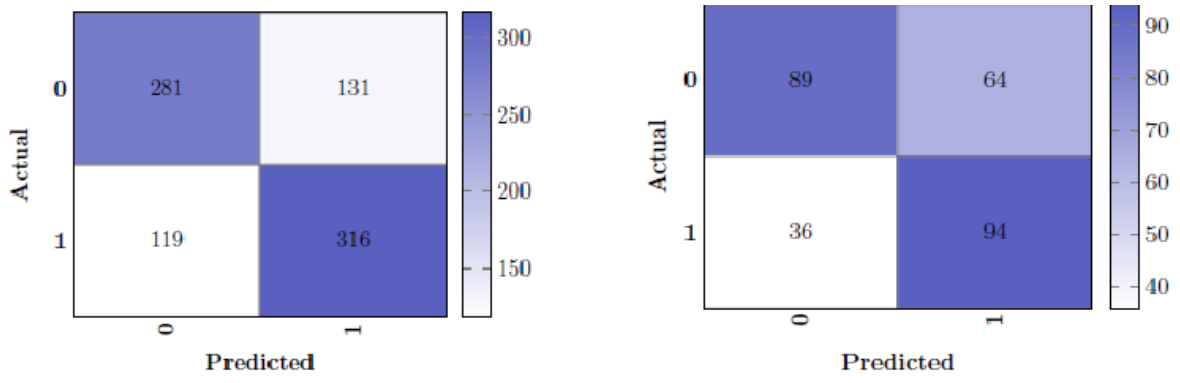
Friedman's statistical test, a non-parametric test technique, was employed to identify the classification technique that outperformed other competing classifiers to conduct a more detailed investigation of the performance of the classification techniques. Table V presents the mean ranks derived from a Friedman test conducted to statistically compare the classification performance of the competing algorithms across accuracy, precision, recall, and F1-measure.

The lowest ranking finding reflects a higher level of performance, as seen in Table V. Friedman's test was used to determine the p-value, displayed in Table V. Some of the p-values found by Friedman's statistical test were less than the significance level, identified as $\alpha = 0.5$. The alternative hypothesis is supported, while the null hypothesis is refuted. The alternative hypothesis contends that there are different



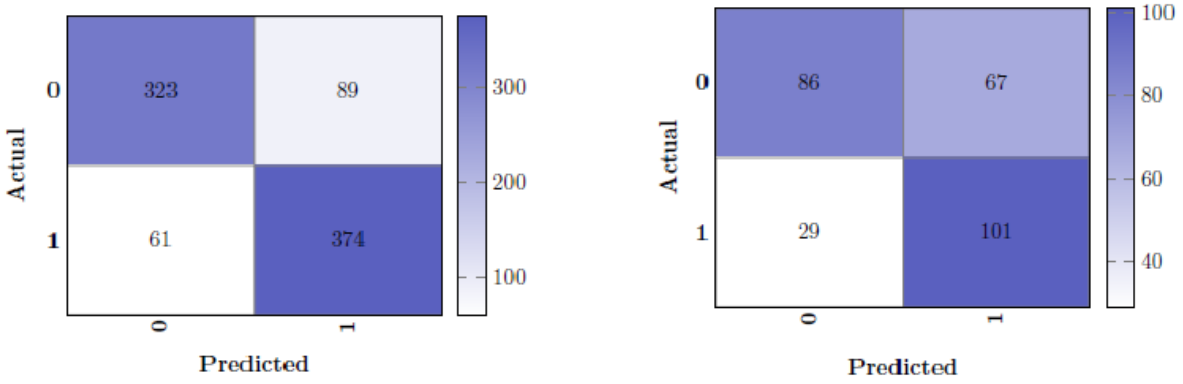
(a) Training. (b) Testing.

Fig. 8. LRC confusion matrices.



(a) Training. (b) Testing.

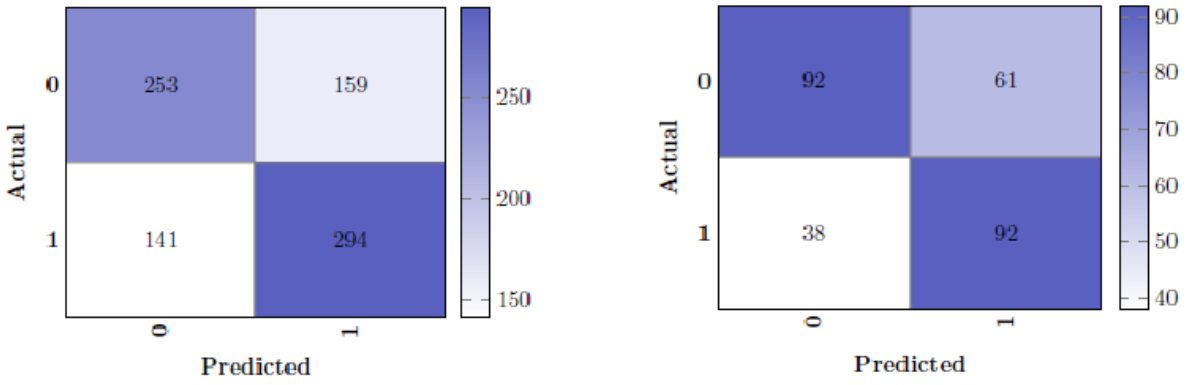
Fig. 9. SVM confusion matrices.



(a) Training. (b) Testing.

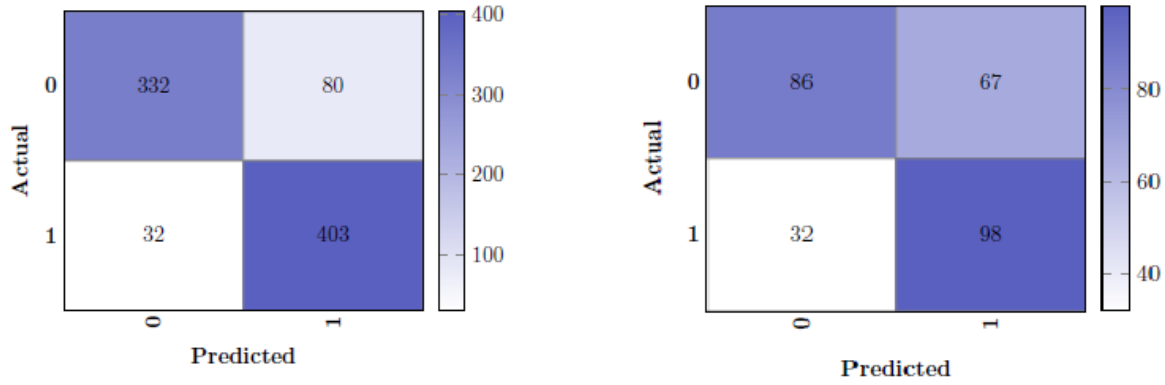
Fig. 10. GBC confusion matrices.

margins in the performance behaviors of the classification techniques, in contrast to the null hypothesis, which holds that all classification techniques have the same performance behavior when employed to address classification problems. According to the statistical findings shown in Table V, the GBC method is the most accurate classifier. This shows that the GBC technique came in first for classification accuracy shared with the KNC classifier while coming in third rank for precision rate, behind RFC and KNC classifiers, first in recall rate, and first in F1 metric measure shared with the RFC classifier. These findings demonstrate the GBC classification method's breadth is better than that of its competitors. In drawing things to a close, it is



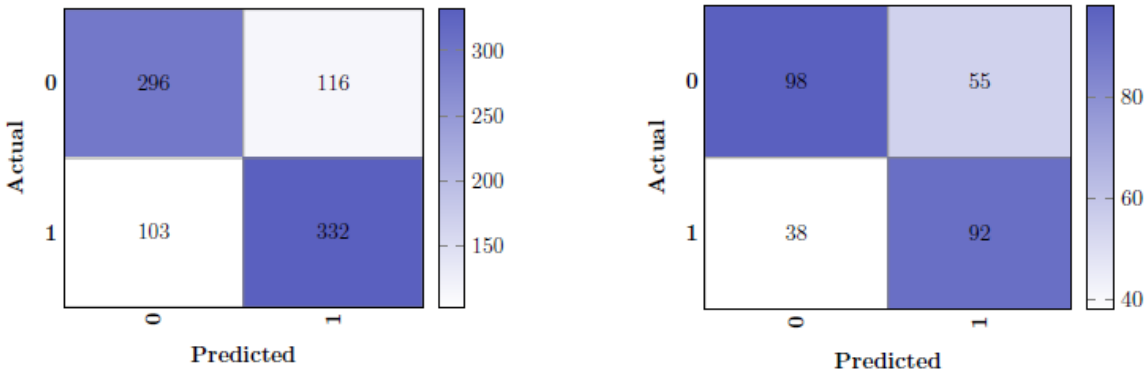
(a) Training. (b) Testing.

Fig. 11. GNB confusion matrices.



(a) Training. (b) Testing.

Fig. 12. RFC confusion matrices.



(a) Training. (b) Testing.

Fig. 13. KNC confusion matrices.

evident that the GBC classification method received the most excellent rank in terms of the total quantity of recall rate, with a rank of 1.0, the lowest rank of all the ranks that the other classifiers have.

The difference between the control classifier and its rivals was then demonstrated using Holm's statistical test as a *post-hoc* statistical method. Friedman's test findings confirm this, demonstrating that the control classification technique outperforms all others in each evaluation metric measure. The statistical findings from Holm's statistical process are shown

TABLE V. AVERAGE RANKING RESULTS OF ALL RIVAL CLASSIFICATION TECHNIQUES REGARDING ACCURACY, PRECISION, RECALL, AND F1 METRIC MEASURES USING FRIEDMAN'S TEST

Classifier	Accuracy	Precision	Recall	F1	Total ranking
LR	5.25	5.5	4.0	5.0	19.75
SVM	4.75	5.0	3.0	4.0	16.75
GBC	2.0	3.0	1.0	1.5	7.5
GNB	4.75	3.75	5.5	6.0	20
KNC	2.0	2.75	3.5	3.0	11.25
RFC	2.25	1.0	4.0	1.5	8.75
p-value	0.220640	0.177047	0.279401	0.083747	

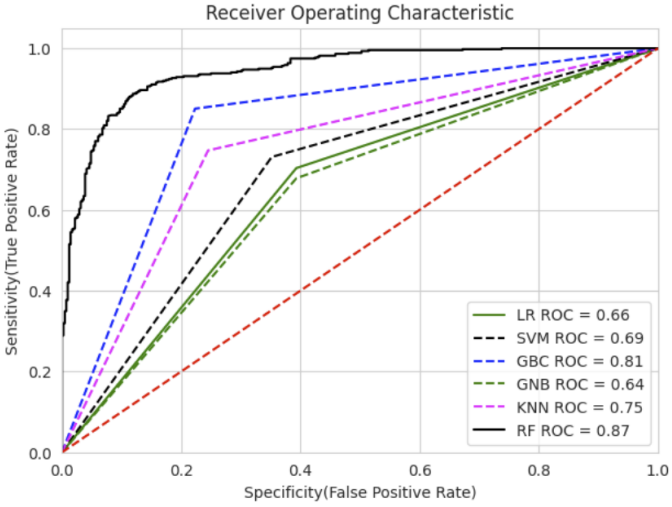


Fig. 14. The Receiver Operating Characteristic curve for the different techniques.

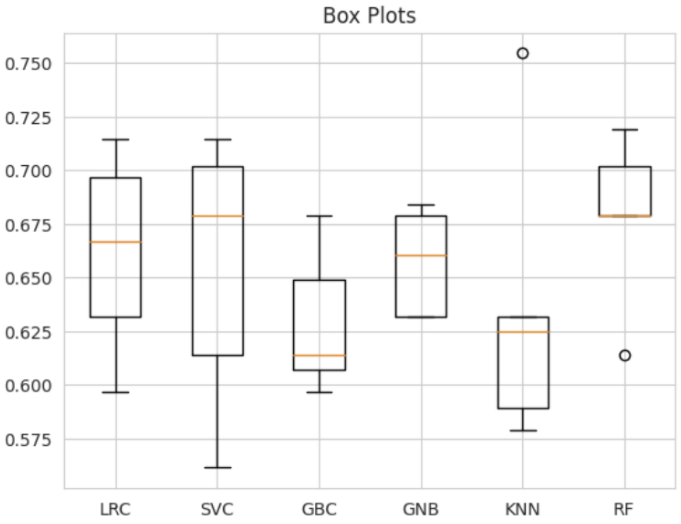


Fig. 15. Box-and-whisker plot for the used techniques.

in Table VI. As per the data reported in Table VI, the control classifier's rank is R_0, the ith classifier's rank is R^i, the effect size of the control classifier's classification technique on the ith classifier is ES, and the statistical difference between two classification techniques is z.

Holm's test was utilized to assess the competing classification systems. This test eliminates hypotheses with p-values of $\leq$ 0.010000 for classification accuracy, $\leq$ 0.010000 for precision, $\leq$ 0.010000 for recall, and $\leq$ 0.010000 for F1 rates, respectively. Table VI demonstrates that GBC outperforms LR, SVM, GNB, KNC, and RFC in terms of classification accuracy, even if there is no statistically significant difference between GBC and the other classification techniques (i.e. LR, SVM, GNB, KNC, and RFC). Friedman's and Holm's test-based statistical precision findings show that the RFC technique and the GNB, KNC, and GBC classification methods do not vary significantly from one another. The RFC approach, however, differs dramatically from the two classification techniques (LR and SVM). The findings in terms of the recall rate show that while GBC differs significantly from the GNB classification technique, it does not differ from the other three competing classifiers (i.e. LR, SVM, KNC, and RFC).

Tables VI and V display the results of Holm's and Friedman's tests, respectively, which demonstrate that the GBC classification method outperforms the other competing methods in achieving promising accuracy and precision rates for the datasets being studied. Overall, the GBC classification method outperformed a number of cutting-edge classification methods disclosed in the literature, including LR, SVM, KNC, and RFC, according to the results of the statistical study discussed above. This demonstrates the GBC method's consistent performance and attests to the fact that it successfully solves classification issues with low, medium, and high dimensions. In addition, GBC's performance degree is close to that of RFC and KNC classifiers as per the average rankings of classification techniques regarding precision rate. Still, the performance level of GNB and LR classifiers falls far short of its RFC and KNC competitors. This leads one to the conclusion that the GBC as a classification model gives it such exceptional capacity to handle classification difficulties. These statistical analysis tests demonstrate the reliability and suitability of the GBC tool as a classification method. These findings provide compelling justifications for employing the GBC classifier to classify difficult datasets progressively.

VII. CONCLUSIONS AND FUTURE WORK

This research aims to examine the feasibility of using machine learning techniques to identify cases of OSA. Traditional techniques of diagnosing OSA are costly and logistically difficult, even though they impact a large percentage of adults. Methods for diagnosing OSA in this study included Logistic Regression, Support Vector Machines, Gradient Boosting Classifier, Random Forest Classifier, Gaussian Naive Bayes,

TABLE VI. RESULTS OF HOLM'S TEST BETWEEN SEVERAL CLASSIFICATION TECHNIQUES

Classification accuracy (GBC is the control classifier)					
i	Algorithm	$z = \frac{(R_0 - R^i)}{SE}$	p-value	$\alpha \div i$	Hypothesis
5	LR	1.737198	0.0823522	0.010000	Not_Rejected
4	SVM	1.469936	0.141578	0.012500	Not_Rejected
3	GNB	1.469936	0.141578	0.016666	Not_Rejected
2	RFC	0.133630-	0.893694	0.025000	Not_Rejected
1	KNC	0.000000	1.000000	0.050000	Not_Rejected

Precision (RFC is the control classifier)					
i	Algorithm	$z = \frac{(R_0 - R^i)}{SE}$	p-value	$\alpha \div i$	Hypothesis
5	LR	2.405351	0.016156	0.010000	Rejected
4	SVM	2.138089	0.032509	0.012500	Rejected
3	GNB	1.469936	0.141578	0.016666	Not_Rejected
2	GBC	1.069044	0.285049	0.025000	Not_Rejected
1	KNC	0.935414	0.349574	0.050000	Not_Rejected

Recall (GBC is the control classifier)					
i	Algorithm	$z = \frac{(R_0 - R^i)}{SE}$	p-value	$\alpha \div i$	Hypothesis
5	GNB	2.405351	0.016156	0.010000	Rejected
4	LR	1.603567	0.108809	0.012500	Not_Rejected
3	RFC	1.603567	0.108809	0.016666	Not_Rejected
2	KNC	1.336306	0.181449	0.025000	Not_Rejected
1	SVM	1.069044	0.285049	0.050000	Not_Rejected

F1 (GBC is the control classifier)					
i	Algorithm	$z = \frac{(R_0 - R^i)}{SE}$	p-value	$\alpha \div i$	Hypothesis
5	GNB	2.405351	0.016156	0.010000	Rejected
4	LR	1.870828	0.061368	0.012500	Not rejected
3	SVM	1.336306	0.181449	0.016666	Not rejected
2	KNC	0.801783	0.422678	0.025000	Not rejected
1	RFC	0.000000	1.000000	0.050000	Not rejected

and K-Nearest Neighbors Classifier. Results showed that the Random Forest Classifier performed the best, with an accuracy of 0.87 during training and 0.65 during testing. The ROC curve produced a score of 0.87. The proposed work achieved classification accuracy comparable to other related studies. However, unlike most existing pieces that utilize PSG or ECG, which can be costly and time-consuming for physicians and patients, we employed physical parameters that are easy to obtain. Future research may explore the potential of other machine learning (ML) techniques, including artificial neural networks (ANN) and decision trees (DT), to address the problem at hand. By exploring these different ML techniques, it may be possible to improve the accuracy and generalizability of the model and gain new insights into the problem domain.

ACKNOWLEDGMENT

The authors extend their appreciation to Taif University, Saudi Arabia, for supporting this work through project number (TU-DSPP-2024-201)

DECLARATIONS

- Funding: This research was funded by Taif University, Taif, Saudi Arabia (TU-DSPP-2024-201).
- Data Availability Statements: The data supporting this study's findings are available on request from the corresponding author.
- Ethical Approval: The authors declare that ethical standards have been followed and that no human participants or animals were involved in this research.
- Conflict of Interest: The authors have no competing interests to declare relevant to this article's content.

REFERENCES

[1] S. P. Patil, H. Schneider, A. R. Schwartz, and P. L. Smith, "Adult obstructive sleep apnea: pathophysiology and diagnosis," *Chest*, vol. 132, no. 1, pp. 325–337, 2007.

[2] A. Sheta, H. Turabieh, M. Braik, and S. R. Surani, "Diagnosis of obstructive sleep apnea using logistic regression and artificial neural networks models," in *Proceedings of the Future Technologies Conference (FTC) 2019: Volume 1*. Springer, 2020, pp. 766–784.

[3] T. D. Bradley and J. S. Floras, "Obstructive sleep apnoea and its cardiovascular consequences," *The Lancet*, vol. 373, no. 9657, pp. 82–93, 2009.

[4] V. K. Somers, D. P. White, R. Amin, W. T. Abraham, F. Costa, A. Culebras, S. Daniels, J. S. Floras, C. E. Hunt, L. J. Olson *et al.*, "Sleep apnea and cardiovascular disease: An american heart association/american college of cardiology foundation scientific statement from the american heart association council for high blood pressure research professional education committee, council on clinical cardiology, stroke council, and council on cardiovascular nursing in collaboration with the national heart, lung, and blood institute national center on sleep disorders research (national institutes of health)," *Circulation*, vol. 118, no. 10, pp. 1080–1111, 2008.

[5] C. V. Senaratna, J. L. Perret, C. J. Lodge, A. J. Lowe, B. E. Campbell, M. C. Matheson, G. S. Hamilton, and S. C. Dharmage, "Prevalence of

obstructive sleep apnea in the general population: a systematic review," *Sleep medicine reviews*, vol. 34, pp. 70–81, 2017.

[6] H. Yue, Y. Lin, Y. Wu, Y. Wang, Y. Li, X. Guo, Y. Huang, W. Wen, G. Zhao, X. Pang *et al.*, "Deep learning for diagnosis and classification of obstructive sleep apnea: A nasal airflow-based multi-resolution residual network," *Nature and Science of Sleep*, pp. 361–373, 2021.

[7] E. Urtnasan, J.-U. Park, and K.-J. Lee, "Multiclass classification of obstructive sleep apnea/hypopnea based on a convolutional neural network from a single-lead electrocardiogram," *Physiological measurement*, vol. 39, no. 6, p. 065003, 2018.

[8] J. Ramesh, N. Keeran, A. Sagahyroon, and F. Aloul, "Towards validating the effectiveness of obstructive sleep apnea classification from electronic health records using machine learning," in *Healthcare*, vol. 9, no. 11. MDPI, 2021, p. 1450.

[9] L. Cen, Z. L. Yu, T. Kluge, and W. Ser, "Automatic system for obstructive sleep apnea events detection using convolutional neural network," in *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. IEEE, 2018, pp. 3975–3978.

[10] A. Sheta, H. Turabieh, T. Thaher, J. Too, M. Mafarja, M. S. Hossain, and S. R. Surani, "Diagnosis of obstructive sleep apnea from ecg signals using machine learning and deep learning classifiers," *Applied Sciences*, vol. 11, no. 14, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/14/6622

[11] M. Deviaene, D. Testelmans, P. Borzée, B. Buyse, S. V. Huffel, and C. Varon, "Feature selection algorithm based on random forest applied to sleep apnea detection," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 2580–2583.

[12] M.-S. Choi, D.-H. Han, J.-W. Choi, and M.-S. Kang, "A study on improving sleep apnea diagnoses using machine learning based on the stop-bang questionnaire," *Applied Sciences*, vol. 14, no. 7, 2024. [Online]. Available: https://www.mdpi.com/2076-3417/14/7/3117

[13] A. Altintas, Y. Yegin, M. Çelik, K. H. Kaya, A. K. Koç, and F. T. Kayhan, "Interobserver consistency of drug-induced sleep endoscopy in diagnosing obstructive sleep apnea using a vote classification system," *Journal of Craniofacial Surgery*, vol. 29, no. 2, pp. e140–e143, 2018.

[14] K. Bond and A. Sheta, "Medical data classification using machine learning techniques," *International Journal of Computer Applications*, vol. 183, no. 6, pp. 1–8, Jun 2021. [Online]. Available: http://www.ijcaonline.org/archives/volume183/number6/31928-2021921339

[15] W. H. Elashmawi, A. Djellal, A. Sheta, S. Surani, and S. Aljahdal, "Machine learning for enhanced copd diagnosis: A comparative analysis of classification algorithms," *Diagnostics*, vol. 14, no. 24, 2024. [Online]. Available: https://www.mdpi.com/2075-4418/14/24/2822

[16] A. Sheta, W. H. Elashmawi, A. Al-Qerem, and E. S. Othman, "Utilizing various machine learning techniques for diabetes mellitus feature selection and classification," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 3, 2024. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2024.01503134

[17] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehericy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with alzheimer's disease from structural mri: A comparison of ten methods using the adni database," *NeuroImage*, vol. 56, no. 2, pp. 766–781, 2011.

[18] C. Mencar, C. Gallo, M. Mantero, P. Tarsia, G. Carpagnano, M. P. Foschino Barbaro, and D. Lacedonia, "Application of machine learning to predict obstructive sleep apnea syndrome (osas) severity," *Health Informatics Journal*, pp. 1–20, 03 2020.

[19] S. Ha, S. J. Choi, S. Lee, R. H. Wijaya, J. H. Kim, E. Y. Joo, and J. K. Kim, "Predicting the risk of sleep disorders using a machine learning–based simple questionnaire: Development and validation study," *J Med Internet Res*, vol. 25, p. e46520, Sep 2023. [Online]. Available: https://www.jmir.org/2023/1/e46520

[20] A. Javeed, J. Berglund, A. Luiza, M. Saleem, and Peter, "Predictive power of xgboost_bilstm model: A machine-learning approach for accurate sleep apnea detection using electronic health data," *International Journal of Computational Intelligence Systems*, p. 188, 11 2023.

[21] A. Sheta, S. Subramanian, S. R. Surani, and M. Braik, "Diagnosis of obstructive sleep apnea using machine learning," in *2023 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, 2023, pp. 12–17.

[22] W. Z. T. Tareq, *Sleep Disorders Detection and Classification Using Random Forests Algorithm*. Cham: Springer International Publishing, 2024, pp. 257–266.

[23] K. Liu, S. Geng, P. Shen, L. Zhao, P. Zhou, and W. Liu, "Development and application of a machine learning-based predictive model for obstructive sleep apnea screening," *Frontiers in Big Data*, vol. 7, 2024. [Online]. Available: https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2024.1353469

[24] G. Bazoukis, S. C. Bollepalli, C. T. Chung, X. Li, G. Tse, B. L. Bartley, S. Batool-Anwar, S. F. Quan, and A. A. Armoundas, "Application of artificial intelligence in the diagnosis of sleep apnea," *Journal of Clinical Sleep Medicine*, vol. 19, no. 7, pp. 1337–1363, 2023. [Online]. Available: https://jcsm.aasm.org/doi/abs/10.5664/jcsm.10532

[25] G. Surrel, A. Aminifar, F. Rincón, S. Murali, and D. Atienza, "Online obstructive sleep apnea detection on medical wearable sensors," *IEEE transactions on biomedical circuits and systems*, vol. 12, no. 4, pp. 762–773, 2018.

[26] H. Korkalainen, J. Aakko, S. Nikkonen, S. Kainulainen, A. Leino, B. Duce, I. O. Afara, S. Myllymaa, J. Töyräs, and T. Leppänen, "Accurate deep learning-based sleep staging in a clinical population with suspected obstructive sleep apnea," *IEEE journal of biomedical and health informatics*, vol. 24, no. 7, pp. 2073–2081, 2019.

[27] A. H. Yüzer, H. Sümbül, M. Nour, and K. Polat, "A different sleep apnea classification system with neural network based on the acceleration signals," *Applied Acoustics*, vol. 163, p. 107225, 2020.

[28] S. Sperandei, "Understanding logistic regression analysis," *Biochemia medica*, vol. 24, no. 1, pp. 12–18, 2014.

[29] R. J. Bacue, "An analytic overview of estes' statistical learning theory," *IEEE Transactions on Neural Networks*, pp. 988–999, 1999.

[30] J. H. Friedman, "Greedy function approximation: A gradient boosting machine." *The Annals of Statistics*, vol. 29, no. 5, pp. 1189 – 1232, 2001. [Online]. Available: https://doi.org/10.1214/aos/1013203451

[31] I. Rish, "An empirical study of the naive bayes classifier," in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. 3, no. 22. IBM New York, 2001, pp. 41–46.

[32] M. Braik, H. Al-Zoubi, and H. Al-Hiary, "Pedestrian detection using multiple feature channels and contour cues with census transform histogram and random forest classifier," *Pattern Analysis and Applications*, vol. 23, no. 2, pp. 751–769, 2020.

[33] T. Hastie, R. Tibshirani, and J. Friedman, *Random Forests*. New York, NY: Springer New York, 2009, pp. 587–604.

[34] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, p. 5–32, oct 2001.

[35] X. Li, M. Li, Y. Zhang, and X. Deng, "A new random forest method based on belief decision trees and its application in intention estimation," in *2021 33rd Chinese Control and Decision Conference (CCDC)*, 2021, pp. 6008–6012.

[36] B. Chakradhar, I. S. Siva Rao, V. Jhansy Archana, and C. V. M. K. Hari, "Detection of malignancy on dermis using j48 and random forest classifiers," in *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, 2020, pp. 1–6.

[37] U. N. A and K. Dharmarajan, "Diabetes prediction using random forest classifier with different wrapper methods," in *2022 International Conference on Edge Computing and Applications (ICECAA)*, 2022, pp. 1705–1710.

[38] R. Hummel, T. D. Bradley, G. R. Fernie, S. I. Chang, and H. Alshaer, "Estimation of sleep status in sleep apnea patients using a novel head actigraphy technique," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015, pp. 5416–5419.

Design of On-Premises Version of RAG with AI Agent for Framework Selection Together with Dify and DSL as Well as Ollama for LLM

Kohei Arai
Department of Science and Engineering, Saga University, Saga City, Japan

***Abstract*—Currently, most RAGs are cloud-based and include Bedrock. However, there is a trend to return from the cloud to on-premises due to security concerns. In addition, it is common for APIs to call Lambda or EC2 for data access, but it is not easy to select the optimal framework depending on the data attributes. For this reason, the author devised a system for selecting the optimal framework using an AI agent. Furthermore, the author decided to use Dify, which is based on a DSL, as the user interface for the on-premises version of RAG, and ollama as a large-scale language model that can be installed on-premises as well. The author also considered the specifications of the hardware required to build this RAG and confirmed the feasibility of implementation.**

Keywords*—RAG (Retrieval-Augmented Generation); API (Application Programming Interface); Lambda; EC2 (Amazon Elastic Compute Cloud); AI agent; Dify; DSL (domain specific language); ollama; YAML (YAML Ain't Markup Language)

I. Introduction

The name RAG (Retrieval-Augmented Generation) and its specific methodology were proposed in 2020. In a paper published by researchers from Facebook AI Research (now Meta AI), University College London, and New York University, RAG was introduced as a "general-purpose fine-tuning recipe"[1].

RAG research ran on a cluster of NVIDIAS GPUs[2] and demonstrated how to make generative AI models more reliable. The research aimed to link generative AI services with external resources, especially those containing the latest technical details.

The concept of RAG has spread rapidly since its introduction and is now adopted in hundreds of papers and many commercial services. This technology has significantly improved the capabilities of large-scale language models (LLMs) and taken question answering systems to a new level.

The development of RAG is the culmination of many years of research in information retrieval and natural language processing and has become an important part of modern AI technology. It is expected that this technology will continue to evolve and contribute to improving the performance and reliability of AI systems.

Research issues related to RAG play an important role in the development and practical application of this technology. Some of the main research issues are listed below.

1) Chunking and embedding: The performance of RAG systems depends heavily on the chunking and embedding methods used. Optimal chunking methods: Research is needed to find effective ways to segment documents and extract relevant information appropriately [1]. Embedding multimedia content: Research is needed to find effective ways to embed non-text data (images, audio, video, etc.).

2) Comparison of RAG and fine-tuning: RAG and fine-tuning of LLMs (large-scale language models) take different approaches. Performance comparison: Research is needed to systematically compare the performance of both methods under various tasks and conditions. Cost-effectiveness: A comparative analysis of the costs of implementation and operation is also an important issue. Applicability: Research is needed to clarify in what situations and applications RAG and fine-tuning are suitable.

3) Testing and monitoring of RAG systems: Quality assurance and continuous improvement of RAG systems are important research topics. Establishment of performance evaluation indicators: It is necessary to develop indicators to properly evaluate the performance of RAG systems. Monitoring methods during operation: Research is required on methods to continuously monitor and improve system performance in actual usage environments.

4) Context optimization: Proper context management is essential for improving the performance of RAG systems. Optimization of chunk size: Research on the optimal chunk size is required, as larger contexts may produce better results. Balance with token restrictions: Optimization research is required that takes into account the trade-off with LLM token restrictions and latency.

5) Efficiency and cost reduction: Improvements in efficiency and cost are important for practical use of RAG systems. Semantic cashing: Research is being conducted into methods to reduce the number of LLM calls and reduce costs and latency by cashing frequent queries and their answers.

[1] https://blogs.nvidia.co.jp/2023/11/17/what-is-retrieval-augmented-generation/

[2] https://en.wikipedia.org/wiki/List_of_Nvidia_graphics_processing_units

Utilization of open-source models: For small texts, open-source sentence embedding models may perform as well as commercial models, and research in this area is progressing.

In this paper, open-source models utilizing RAG are investigated and designed to enhance cost-effectiveness. Currently, many RAG systems are cloud-based and use services such as Amazon Bedrock[3]. However, due to security concerns, there is a trend of moving from the cloud to on-premises environments. Traditionally, data access was typically achieved through API calls via AWS Lambda or EC2[4], but it was not easy to select the optimal framework according to the data attributes.

To address this challenge, the author devised a system that utilizes an AI agent to select the optimal framework. In addition, they adopted DSL[5] (Domain Specific Language)-based Dify[6] as the user interface for the on-premises version of RAG, and selected ollama[7] as a large-scale language model that can be deployed on-premises. In this paper, ollama llama3.2 is used.

The author also considered the specifications of the hardware required to build this RAG system and confirmed the feasibility of implementation. This made it possible to build an efficient and flexible RAG system while placing emphasis on security.

The following section described the related research works on RAG followed by proposed RAG system. Then, software and hardware requirements are described. After that, the conclusion is described with some remarks and discussions.

II. Related Research

There are the following RAG related papers, Retrieval-augmented generation for knowledge-intensive NLP (Natural Language Processing) tasks is proposed. Natural Language Processing (NLP) tasks include sentiment analysis, entity recognition, text summarization, machine translation, speech recognition, text classification, chatbot interaction, keyword extraction, question answering, part-of-speech tagging, topic modeling, predictive text, conference resolution, and spam detection; essentially, any activity where a computer analyzes and understands human language to perform a specific function. This paper is the first to propose the RAG model, an approach that improves performance on knowledge-based tasks by integrating a retrieval component into a generative model [2].

Improving zero-shot generalization in text classification using retrieval-augmented language models are proposed. In this study, the authors applied RAG to text classification tasks, aiming to improve zero-shot generalization performance. The authors showed that using a retrieval component improves classification accuracy for unseen classes [3].

RAG for knowledge-intensive NLP tasks is discussed. The paper provides a detailed description of RAG architecture and evaluates its performance on a variety of tasks, showing that it performs particularly well on knowledge-based tasks such as question answering and sentence generation [4].

Dense passage retrieval for open-domain question answering is proposed. In this study, a density-based passage retrieval method is proposed for open-domain question answering tasks, which is often used as the retrieval component of the RAG model to improve the accuracy of question answering [5].

Other than these, there are the following recent research works,

Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection is proposed which appears in the URL of https://arxiv.org/abs/2310.11511, [6].

Atlas: Few-shot Learning with Retrieval Augmented Language Models are proposed which appears in the URL of https://arxiv.org/abs/2208.03299, [7].

Internet-Augmented Dialogue Generation is also proposed which appears in the URL of https://arxiv.org/abs/2107.07566, [8].

REPLUG: Retrieval-Augmented Black-Box Language Models is proposed which appears in the URL of https://arxiv.org/abs/2301.12652, [9].

Dense Passage Retrieval for Open-Domain Question Answering is proposed which appears in the URL of https://arxiv.org/abs/2004.04906, [10].

Realm: Retrieval-Augmented Language Model Pre-Training is proposed which appears in the URL of https://arxiv.org/abs/2002.08909, [11].

Improving language models by retrieving from trillions of tokens are proposed which appears in the URL of https://arxiv.org/abs/2112.04426, [12].

Query2doc: Query Expansion with Large Language Models are also proposed which appears in the URL of https://arxiv.org/abs/2303.07678, [13].

Chain-of-Note: Enhancing Robustness in Retrieval-Augmented Language Models are proposed which appears in the URL of https://arxiv.org/abs/2311.09210, [14].

On the other hand, Dify related research works are as follows,

The literature on data integration in general, DSLs, and Dify is as follows: "A Survey of Data Integration Systems" This paper provides an overview of data integration systems and various approaches [15].

"Domain-Specific Languages: An Annotated Bibliography" This paper provides an overview of DSLs and examples of various amana DSLs [16].

[3] https://docs.aws.amazon.com/ja_jp/bedrock/latest/userguide/what-is-bedrock.html

[4] https://www.serverless.direct/post/aws-lambda-vs-ec2-which-one-to-choose-for-your-app

[5] https://en.wikipedia.org/wiki/Domain-specific_language

[6] https://docs.dify.ai/ja-jp/guides/application-orchestrate/creating-an-application

[7] https://ollama.com/library/llama3.2

"Interactive Data Integration with User-Centric Approaches" This paper describes user-centric data integration approaches [17].

"VLDB Conference Proceedings" The latest research results on data integration are presented at the VLDB conference [18].

"SIGMOD Conference Proceedings" Research results on data integration and DSLs are also presented at the SIGMOD conference [19].

"Technical Report: Data Integration Using DSLs" Some research institutes have published technical reports on data integration using DSLs [20].

"Data Integration: A Theoretical Perspective" This book provides a detailed explanation of the theoretical background of data integration [21].

"Domain-Specific Languages in Action" This book gives practical examples of the application of various DSLs [22].

GitHub Repository: Some open-source projects publish code for data integration tools using DSLs [23].

III. Proposed RAG System

A. System Configuration

The on-premises version of RAG is intended to be created. The author devised a system that utilizes an AI agent to select the optimal framework. In addition, they adopted DSL-based Dify as the user interface for the on-premises version of RAG, and selected ollama as a large-scale language model that can be deployed on-premises.

Fig. 1 shows the block diagram of the proposed RAG.

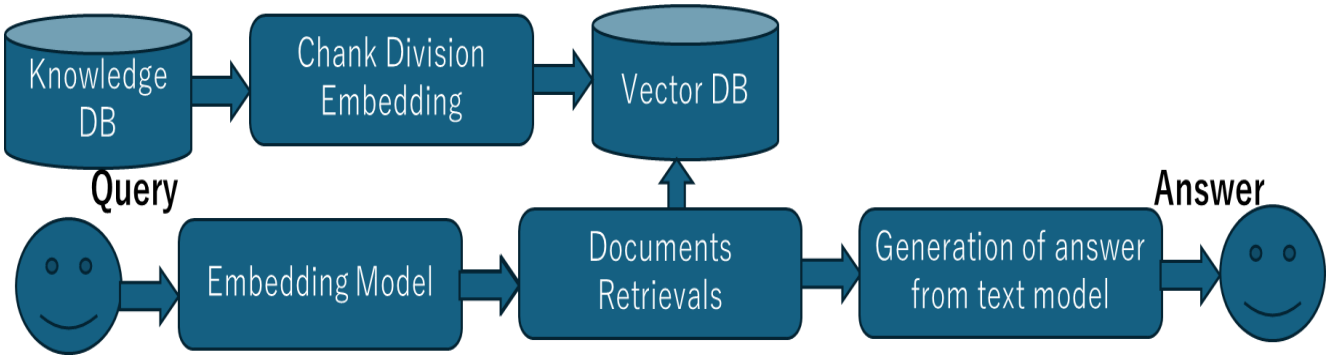


Fig. 1. Block diagram of the proposed RAG system.

The process from query to answer generation in the RAG system goes through the following steps:

1) Query processing: The system receives a question or task (query) from the user and converts it into a format that the system can easily understand. The system uses natural language processing technology to analyze the meaning of the query. Converts it into a format suitable for search.

2) Information retrieval: Based on the converted query, the system searches the knowledge base for relevant information. Query vectorization: Converts the question into a numerical expression (vector). Similarity calculation: Calculates the similarity between the query vector and the vector of information in the knowledge base. Ranking: Ranks the most relevant information based on the similarity.

3) Context generation: Based on the information obtained from the search results, the system generates a context related to the question.

4) Answer generation: A generative model (usually a large-scale language model) takes the query and context as input and generates an appropriate answer. Integrates the given information and then creates an answer in natural language.

5) Post-processing: The generated answer is further processed to be formatted into the final output format. Formatting the answer and filtering inappropriate content as well as adding additional information where necessary (e.g. citing sources). After that, adjust the length and complexity of the answer.

6) Output: The final answer is presented to the user. Answer in text form and possibly include relevant images, links, and/or additional references.

7) Feedback and learning: Many RAG systems have mechanisms for collecting user feedback and continually improving the system's performance. Recording user responses (e.g. clicking the "helpful" button) and evaluating the accuracy and relevance of answers as well as tweaking search algorithms and generative models based on feedback.

As for the user interface, Dify is featured as follows:

Dify has all the features to develop AI apps. It also supports hundreds of AI language models, allowing creating custom chatbots with intuitive operations. It also has a high-performance and flexible RAG engine, allowing building AI agents to use a variety of tools. It can also be freely designing workflows.

Dify Studio[8] offers three ways to create an application: It can be used on an application template. It can start with a blank application. You can create one (locally or online) by importing a DSL file. To get a quick overview of the types of applications

[8] https://note.com/jolly_dahlia842/n/n41bf7cf085fd

you can create with Dify, select "Studio" from the navigation menu, then "Create from Template" from the application list.

When creating the first application with Dify, it is important to understand the basic concepts behind the four different types of applications: chat voice, text generators, agents, and workflows. Also, when creating an application, give it a name, choose an appropriate icon, and briefly describe the purpose of this application to make it easier to use within your team.

Dify DSL is a standard file format (YML) [9] for AI application development defined by Dify.AI. This standard includes basic information about the application, model parameters, orchestration settings, etc. Firstly, import local DSL files. Then if a DSL file (template) exists, provided by the community or others, select "Import DSL file" from Studio. After importing, the original application settings will be loaded directly. After that, import a DSL file via URL. It can import a DSL file via URL using the following format: https://example.com/your_dsl.yml

DSL is a specialized language designed to efficiently build and configure AI applications on the Dify platform. The features of Dify DSL are: YAML format[10]: Dify DSL is written in YAML format, Workflow definition: It can concisely define the workflow and settings of your application and sharing and reuse: It can easily share and reuse the workflow you created by exporting and importing it as a DSL file.

To use DSL, it can import it from the Dify dashboard by selecting "Create an app" and then "Import DSL file". Then select the YAML file and click the "Create" button. Customize it: After importing, it can adjust the workflow and settings as needed. Finally, export it: it can export the workflow it created as a DSL file and share it with other users.

It can easily create chatbots specialized for specific industries or purposes using knowledge bases. This application can be developed without programming expertise by combining Dify's no-code interface with the flexibility of DSL. In addition, it can export and import the workflow it created as a DSL file, making it easy to share and reuse within a team or community.

B. Ollama of LLM

Next, the author will explain ollama, a small-scale LLM that can run on-premises on a local machine. It has the following features:

Local execution: With ollama, you can run LLM on your own computer without an Internet connection.

Multiplatform support: Currently it supports Mac and Linux and will support Windows in the future.

Various models: It supports large-scale language models such as Llama. Users can select and use various models.

Easy environment construction: It is relatively easy to install and configure, and you can easily obtain and run models from the command line.

Python integration: ollama can be used from Python environments using the ollama-python library. This allows you to leverage local LLMs through APIs and integrate them into RAGs and agents.

Customizability: Since it is open source, users can customize it to suit their needs.

Security: Since it runs locally, it is suitable for projects that value data privacy and security.

Ollama is a powerful tool for individuals and businesses to effectively leverage AI technology, allowing them to run advanced natural language processing tasks such as text generation, question answering, and text summarization in a local environment.

To install ollama, follow the steps below:

Download the installer from the ollama official website.

Open the downloaded file and follow the instructions to install.

Once the installation is complete, a llama icon will appear in the taskbar on Windows.

Download and run the model

Open Command Prompt (Windows) or Terminal (Mac/Linux).

Run the following command to download and run the model:

bash

ollama run gemma2:2b

This command downloads and runs the Gemma 2b model. If the model is not already available locally, it will be downloaded automatically.

To interact with the AI, once the model is running, a prompt will appear. Enter users' message and press Enter, and the AI will respond.

Users can customize and manage your models using the following commands:

ollama create: Create a custom model

ollama show: Show model information

ollama list: Show a list of installed models

ollama rm: Delete a model

C. Python Code for Knowledgebase Creation

In this example we'll create a simple knowledge base with a dictionary structure and show how to perform some basic operations. Python code for creation of Knowledgebase System is as follows,

python

class KnowledgeBase:

[9] https://docs.dify.ai/guides/application-orchestrate/creating-an-application

[10] https://spacelift.io/blog/yaml

```
def __init__(self):
self.knowledge = {}
def add_fact(self, key, value):
"""Add a new fact to the knowledge base"""
self.knowledge[key] = value
def get_fact(self, key):
"""Get a fact from the knowledge base"""
return self.knowledge.get(key, "No information found")
def update_fact(self, key, value):
"""Update an existing fact"""
if key in self.knowledge:
self.knowledge[key] = value
return True
return False
def remove_fact(self, key):
"""Remove a fact from the knowledge base"""
if key in self.knowledge:
del self.knowledge[key]
return True
return False
def list_all_facts(self):
"""List all facts in the knowledge base"""
return self.knowledge
# Knowledge Base Usage Example
if __name__ == "__main__":
kb = KnowledgeBase()
# Add fact
kb.add_fact("Python", "High-level programming language")
kb.add_fact("AI", "Artificial intelligence")
kb.add_fact("ML", "Machine learning")
# Get fact
print(kb.get_fact("Python")) # Output: High-level programming language
print(kb.get_fact("Database")) # Output: No information found
# Update fact
kb.update_fact("AI", "Artificial intelligence technology")
print(kb.get_fact("AI")) # Output: Artificial intelligence technology
# Remove fact
kb.remove_fact("ML")
# List all facts
print(kb.list_all_facts())
```

This simple implementation can be extended as follows:

1) Persistence: Use a database (e.g. SQLite) to store the knowledge base and maintain the information even after the program ends.

2) Complex data structures: Store objects or structured data instead of simple strings.

3) Search capabilities: Implement advanced search capabilities using keyword searches or regular expressions.

4) Version control: Add the ability to track changes to each fact.

5) Relationship expression: Be able to express relationships between facts (e.g. a graph database-like approach).

6) Inference engine: Implement a simple inference function to derive new facts from existing facts.

D. Required Hardware Specifications

The hardware specifications for building a RAG using ollama are as follows.

1) CPU

a) Best choice: Intel CPU of 11th generation or later that supports the AVX512 instruction set, or AMD CPU based on Zen4. An AMD CPU based on the "Zen 4" architecture is the first AMD processor to support the AVX-512 instruction set, meaning if you're looking for an AMD CPU with AVX-512 capability, you should choose one based on the Zen 4 microarchitecture.

b) Reason: To speed up the matrix calculations required for AI models

c) Minimum requirement: Any CPU that supports the AVX instruction set will work

2) RAM

a) Recommendation: 16GB or more

b) Reason: To comfortably run models with 7B parameters

c) Minimum requirement: May work with around 8GB

3) Storage

a) Recommendation: 50GB of freer space

b) Breakdown: Docker container (2GB+), model file, vector store, etc.

4) GPU

a) Recommended: Equipped with NVIDIA GPU (e.g. GTX 1080 Ti or higher)

b) Reason: Can significantly speed up model inference

c) Not required: Can be run with CPU only, but processing speed will be reduced.

In this connection, GIGABYTE AORUS GeForce GTX 1080 Ti 11GB Video Card - GV-N108TAORUS-11GD[11] is one of the candidates.

Other points to consider are that if you are using larger models (13B or more), a high-performance GPU and 32GB or more of RAM are recommended. Users also need to consider the amount of VRAM (depending on the model size and quantization level). ollama can run on relatively lightweight systems, but to get comfortable user experience, it is recommended that you use the above recommended specifications as a guide. Especially in the RAG system, it is desirable to have a generous specification because building and searching the vector store also requires resources.

E. AI Agent

Although the definition of an AI agent may vary slightly depending on the technical field, it generally refers to software that interacts with its environment, collects data, and autonomously executes tasks based on that data to achieve a specific goal. In particular, our focus here is on AI agents based on LLMs. Using an AI agent as the user interface of the RAG system is a good approach to achieve a more flexible and advanced conversational interface. Below are some recommended implementation ideas.

1) Multi-agent system: This method uses a combination of multiple specialized agents.

a) Triage agent: Analyzes the user's question and assigns it to the appropriate specialized agent.

b) Search agent: Responsible for the search function of the RAG system to retrieve related information.

c) Answer generation agent: Generates an appropriate answer based on the search results.

d) Dialogue management agent: Manages the flow of dialogue with the user and asks additional questions as necessary.

In this method, each agent specializes in a specific role, allowing complex tasks to be handled efficiently.1.

2) Plan-and-Execute agent: This method is performed by a single advanced agent that plans and executes the plan.

Analyzes the user's question and plans the steps required to answer it.

Based on the plan, it sequentially searches the RAG system, integrates information, generates answers, etc.

The plan is revised as necessary to generate the final answer.

This method is particularly effective when complex questions or multi-step processing are required.1.

3) Conversational RAG Agent: An agent that collects information through dialogue with the user and gradually refines its answer.

It first provides a concise answer to the user's initial question.

It then asks the user if additional information or clarification is needed.

Based on the user's response, it re-uses the RAG system to complement the information and expand the answer.

This method allows for flexible information provision tailored to the user's needs.

4) Self-improving RAG Agent: An agent system that incorporates a feedback loop.

After answering the user's question, it asks for feedback on the quality and appropriateness of the answer.

Based on the feedback, it automatically adjusts how it generates search queries and constructs answers.

Continuous learning improves performance over time.

This method is particularly effective in long-term use, allowing the system to continually improve its accuracy and usefulness.

Implementation Considerations:

1) Model selection: Using high-performance models such as GPT-4 allows for more sophisticated dialogue and accurate information processing.

2) Context management: It is important to properly manage long-term dialogue history and maintain consistent dialogue.

3) Error handling: Users need to implement a way to handle cases when the agent cannot respond appropriately and provide appropriate feedback to the user.

4) Security and privacy: Be careful with user data, filtering and anonymizing information where necessary.

By using these recommendations as a starting point and customizing them for specific use cases and requirements, it can implement an AI agent as an effective RAG user interface.

F. Swarm AI Agent

Swarm is a framework for multi-agent orchestration released by OpenAI on October 12, 2024[12]. This framework uses Python and is designed to enable AI agents to work together and autonomously complete complex tasks. By using Swarm, it becomes easy to build multi-agent systems.

Another library for building AI agents is LangGraph[13]. LangGraph is feature-rich and highly flexible but tends to be complicated to implement. In contrast, Swarm has fewer features but is very easy to implement.

Table I shows the classes for creating AI agents. Users can set the agent's name, behavior, model to be used, etc. client.run() is a function to execute the created agent. The arguments of this function are shown in Table II.

It processes messages for the agent and advances conversation. Furthermore, run_demo_loop() is a function to

[11] https://www.gigabyte.com/jp/Graphics-Card/GV-N108TAORUS-11GD#kf

[12] https://github.com/openai/swarm

[13] https://langchain-ai.github.io/langgraph/

repeatedly run the created agent on the console as shown in Table III. It uses client.run() internally to run the agent.

TABLE I. THE CLASSES FOR CREATING AI AGENTS

Field_Name	Type	Default	Description
name	str	"Agent"	Name_of_the_agent
model	str	"gpt-4o"	AI_model_to_use
instruction	str	You_are_a_helpful_agent.	Instructions_to_the_agent
functions	List	[]	List_of_functions_available_to_the_agent
tool_choice	str	None	Specific_tool_to_be_used_by_the_agent

TABLE II. THE ARGUMENT FOR CLIENT.RUN ()

Argument_Name	Type	Initial_Value	Description
agent	Agent	Required	Initial_agent_to_be_called
messages	List	Required	List_of_message_objects
context_variables	dict	{}	Context_with_additional_information
max_turns	int	float("inf")	Maximum_number_of_turns_in_conversation
model_override	str	None	Option_to_change_model
stream	bool	False	Whether_to_show_streaming_responses
debug	bool	False	Debug_mode

TABLE III. THE ARGUMENT FOR RUN_DEMO_LOOP ()

Argument_Name	Type	Initial_Value	Description
starting_agent	Agent	Required	Initial_agent_to_be_called
context_variables	dict	{}	Context_containing_additional_information
stream	bool	False	Whether_to_display_the_response_in_streaming_mode
debug	bool	False	Debug_mode

IV. CONCLUSION

On-premises version of RAG is proposed for secure reasons. The proposed RAG utilizes Swarm-based AI agent which allows easy to select the optimal framework depending on the data attributes. Furthermore, the Dify which is based on a DSL is used as the user interface for the on-premises version of RAG. Also, ollama is used as a large-scale language model that can be installed on-premises as well.

Other than that, the specifications of the hardware required to build this RAG and confirmed the feasibility of implementation. It is confirmed that the proposed RAG system can be created with just one PC with a GPU card.

V. FUTURE RESEARCH WORKS

Although it is confirmed that the proposed RAG system can be feasible, knowledgebase system is not being developed. There are so many applications of the proposed RAG system. Therefore, one of the business use cases will be attempted in the near future.

ACKNOWLEDGMENT

The author would like to thank Prof. Dr. Hiroshi Okumura and Prof. Dr. Osamu Fukuda of Saga University for their valuable comments and suggestions.

REFERENCES

[1] Scott Barnett, Stefanus Kurniawan, Srikanth Thudumu, Zach Brannelly, Mohamed Abdelrazek, Seven Failure Points When Engineering a Retrieval Augmented Generation System, URL : https://arxiv.org/abs/2401.05856, Applied Artificial Intelligence Institute, Geelong, Australia, 2024.

[2] Lewis, P., Perez, E., Pott, C., & Riedel, S., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks" by Patrick Lewis et al., Retrieval-augmented generation for knowledge-intensive NLP tasks. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS 2020) (pp. 1–12), 2020.

[3] Sap, M., Lourie, N., & Riedel, S., "Improving Zero-Shot Generalization in Text Classification using Retrieval-Augmented Language Models" by Maarten Sap et al., Improving zero-shot generalization in text classification using retrieval-augmented language models. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021) (pp. 1–12), 2021.

[4] Riedel, S., Lewis, P., & Perez, E. (2020). RAG: Retrieval-augmented generator for knowledge-intensive nlp tasks. arXiv preprint arXiv:2005.11401., "RAG: Retrieval-Augmented Generator for Knowledge-Intensive NLP Tasks" by Sebastian Riedel et al., 2005.

[5] Karpukhin, V., Oğuz, B., Min, S., Wu, L., Edunov, S., Chen, D., & Yih, W. T. (2020). Dense passage retrieval for open-domain question answering. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS 2020) (pp. 1–12)., "Dense Passage Retrieval for Open-Domain Question Answering" by Vladimir Karpukhin et al.,

[6] Akari Asai, et al., "Self-RAG: Learning to Retrieve, Generate, and Critique through Self-Reflection", https://arxiv.org/abs/2310.11511, 2023.

[7] Gautier Izacard, et al., "Atlas: Few-shot Learning with Retrieval Augmented Language Models", https://arxiv.org/abs/2208.03299, 2023.

[8] Mojtaba Komeili, et al., "Internet-Augmented Dialogue Generation", https://arxiv.org/abs/2107.07566, 2022.

[9] Weijia Shi, et al., "REPLUG: Retrieval-Augmented Black-Box Language Models", https://arxiv.org/abs/2301.12652, 2023.

[10] Vladimir Karpukhin, et al., "Dense Passage Retrieval for Open-Domain Question Answering", EMNLP 2020, https://arxiv.org/abs/2004.04906, 2020.

[11] Kelvin Guu, et al., "Realm: Retrieval-Augmented Language Model Pre-Training", ICML 2020, https://arxiv.org/abs/2002.08909, 2020.

[12] Sebastian Borgeaud, et al., "Improving language models by retrieving from trillions of tokens", ICML 2022, https://arxiv.org/abs/2112.04426, 2022.

[13] Zhuyun Dai, et al., "Query2doc: Query Expansion with Large Language Models", https://arxiv.org/abs/2303.07678, 2023.

[14] Wenhao Yu, et al., "Chain-of-Note: Enhancing Robustness in Retrieval-Augmented Language Models", https://arxiv.org/abs/2311.09210, 2023.

[15] Halevy, A. Y., et al. "Enterprise information integration: successes, challenges and controversies." Proceedings of the 2005 ACM SIGMOD international conference on Management of data. 2005.

[16] Mernik, M., Heering, J., & Sloane, A. M. "When and how to develop domain-specific languages." ACM Computing Surveys (CSUR) 37.4 (2005): 316-344, 2005.

[17] Sarma, A. D., et al. "Interactive data integration." Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data. 2010, 2010.

[18] International Conference on Very Large Data Bases (VLDB). VLDB Conference Website, https://vldb.org/2024/, accessed on 19 December 2024.

[19] ACM SIGMOD International Conference on Management of Data. SIGMOD Conference Website,

https://dl.acm.org/doi/proceedings/10.1145/3626246, accessed on 19 December 2024.

[20] Technical report series of research institutes (e.g. MIT CSAIL Technical Reports), https://libguides.mit.edu/c.php?g=176306&p=1159542, accessed on 19 December 2024.

[21] Lenzerini, M. "Data integration: a theoretical perspective." ACM SIGMOD Record 33.3 (2004): 66-73, 2004.

[22] Fowler, M. "Domain-specific languages." Addison-Wesley Professional, 2010.

[23] GitHub (e.g. Apache NiFi, Apache Beam). https://github.com/apache/beam, accessed on 19 December 2024.

AUTHOR'S PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January 1979 to March 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post-Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science in April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor at Brawijaya University. He also is an Award Committee member of ICSU/COSPAR. He also is an adjunct professor of Nishi-Kyushu University and Kurume Institute of Technology Applied AI Research Laboratory. He wrote 119 books and published 728 journal papers as well as 569 conference papers. He received 98 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, Science award of Ministry of Mister of Education of Japan in 2015 and so on. He is now Editor-in-Chief of IJACSA and IJISA. http://teagis.ip.is.saga-u.ac.jp/index.html

Deep Ensemble Method for Healthcare Asset Mapping Using Geographical Information System and Hyperspectral Images of Tirupati Region

P. Bhargavi[1], T. Sarath[2], Gopichand G[3]*, G V Ramesh Babu[4], T Haritha[5], A.Vijaya Krishna[6]
Assistant Professor, Dept. of. Computer Science, Sri Padmavathi Mahila Visvavidyalayam, Tirupati, India[1]
Research Scholar, School of Computer Science and Engineering and Information Systems,
Vellore Institute of Technology, Vellore, India[2]
Assistant Professor Senior Grade - 2, School of Computer Science and Engineering,
Vellore Institute of Technology, Vellore, India[3]
Associate professor, Department of Computer Science, Sri Venkateswara University Tirupati, India[4]
Assistant professor, Department of CSE, Sree Rama Engineering College, Tirupati, India[5]
Assistant Professor-Department of Information Technology,
G. Narayanamma Institute of Technology and Science for women, Hyderabad, India[6]

***Abstract*—The ever-increasing capabilities of deep learning for image analysis and recognition have encouraged some researchers investigates potential benefits of merging Hyperspectral Images (HSI) and Geographic Information Systems (GIS) with deep learning in the healthcare industry. Healthcare is an ever-changing sector that constantly adopts new technologies to improve decision-making and patient service. This research digs into the role that GIS and Remote Sensing (RS) play in modern healthcare and their significance. By delivering data from faraway places and enabling spatial analysis, the combination of RS and GIS has transformed healthcare. This GIS & RS data will have the numerous quantity of data so big data analytics can be helpful for storing and retrieving of data. This analysis can open up a new possibility for better healthcare planning, disease management, and environmental health assessment based on the study area's population. This paper deals healthcare assets mapping based on the population and the study area of the Tirupati district hyperspectral image by applying the Deep Ensemble method.**

***Keywords*—Geographical information system; hyperspectral image; remote sensing images; big data analytics; deep ensemble methods; healthcare asset**

I. Introduction

Deep Learning is the ongoing innovation in reaction to emerging threats and possibilities, healthcare is an ever-changing industry. Modern healthcare delivery, public health administration, and resource allocation cannot be improved without incorporating state-of-the-art technology. Notable between these technologies are GIS and RS [14], both of which have become potent instruments with far-reaching consequences in healthcare. Exploring and clarifying the numerous claims of Remote Sensing (RS) and Geographic Information Systems (GIS) is vital for properly understanding their revolutionary potential in healthcare [1]. The field of RS, which includes gathering information about Earth's surface from sensors attached on aircraft or spacecraft, is playing an increasingly important role in medical research and practice. Its applications range from ecological monitoring to public health calculation, illness scrutiny and disaster organization. By gathering information remotely [15], RS provides a vantage point that allows medical personnel to keep tabs on expansive regions with unmatched precision and speed [2]. This GIS & RS data will have the numerous quantity of data so big data analytics is used for keeping and retrieving of info [14]. In contrast, GIS allows the visualisation and understanding of health-related data within a geographical context by making GIS as an ideal tool for spatial data analysis [18]. As a result, healthcare providers are better in allocating healthcare resources, conducting epidemiological studies and ensure that all citizens have access for necessary medical treatments. Improving the quality of healthcare and reducing healthcare inequities depends on GIS-based geospatial approach [19, 21].

This paper deals with the available and need of healthcare facilities in the study area of Tirupati district GIS and hyperspectral image based on the population by applying the Deep Ensemble method.

II. Related Work

A. Imaging Techniques Using Hyperspectral Rays

One method that merges the two regions is hyperspectral imaging. It often encompasses a long stretch of the electromagnetic spectrum and offers real-time scanning imaging throughout dozens or even hundreds of spectral series, including the UV, infrared, VIS and mid-infrared [3]. In Fig. 1, for example, you can see a combination of two-dimensional spatial data with one-dimensional spectral data can see it as a stack of many two-dimensional images [4]. Utilising this method, one can acquire the absorption, reflectance, fluorescence spectra of each individual pixel inside picture [24]. Compared to standard images of RGB and maps in grayscale offers a more robust spectral band and better spectral resolution [16]. It records subtle spectrum subtleties in reaction to various clinical conditions and can detect alterations in things that are invisible with traditional imaging techniques [23,25]. The usual push broom hyperspectral system principle [5]

explains HSI system mechanism, as depicted in Fig. 2. The spatial information is initially illuminated by a light source, which travels via front lens into slit, where light of varied wavelengths is fixed to different degrees. Then, the detector is illuminated with light from every pixel point in that dimension using dispersion devices like prisms and gratings, which divide the light into tiny spectral bands [20].

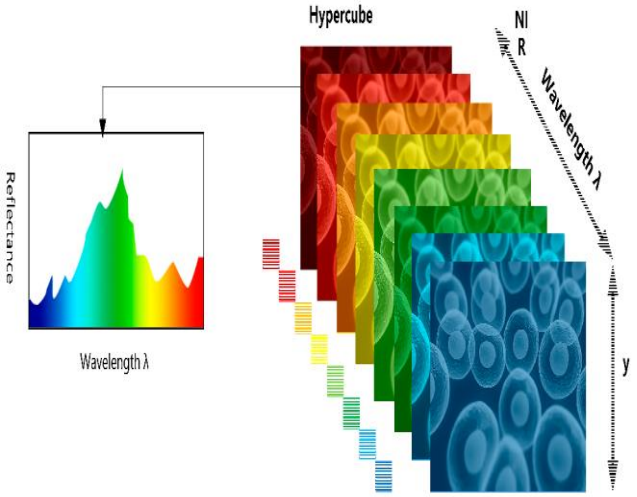


Fig. 1. Spectral data cube (Google courtesy).

The detector array is photographed with each row of sample space info as a 2-D picture. A hypercube with two spatial dimensions and one spectral dimension is created when HSI camera moves via plane using mechanical push sweep and captures adjacent two-dimensional images.

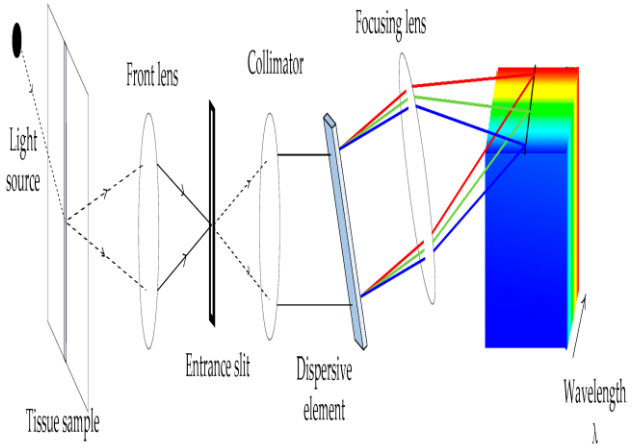


Fig. 2. Hyperspectral imaging system schematic using push-scan technology (Google courtesy).

The smallest discernible object within a picture is called its spatial resolution, and it is the smallest visible detail in the image [26]. There is a practical reason for the sensor's design and the altitude at which it operates above the surface. It is not the number of pixels but the spatial resolution that defines how sharp a picture is. The image sensor's architecture, particularly its height and field of view, determine the spatial properties of an image [27]. The energy received from a specific area of the ground surface is quantified by the remote sensor detector. There is an inverse relationship between spatial resolution and patch size. It is possible to extract more precise spatial information from the picture with smaller individual patches. Only when the cell dimensions are substantially lower than the object dimensions can shape be discerned, which is a common factor. The average brightness of the picture cell is affected by the item's brightness or darkness in relation to its surroundings; hence, this cell will seem brighter than neighboring cells. Linear features, like farm-fresh fruit, may be distinguished with more accuracy than cell dimensions thanks to the high spatial resolution.

A sensor's spectral resolution is defined by the number of spectral bands it can detect and the breadth of the electromagnetic spectrum. Even while an image sensor covers a lot of ground in terms of frequency, its spectral resolution could be inadequate if it only manages to get a few bands. A sensor can differentiate between scenes with identical or almost identical spectral fingerprints if it is accurate in the low-frequency range and catches a large number of spectral bands; this property is known as high spectral resolution [28]. The multispectral images can't pick up on faint spectral signals because of their poor spectral resolution. In the visible, near-infrared, and mid-infrared parts of the electromagnetic spectrum, HSI sensors capture images in numerous neighbouring and incredibly narrow spectral bands. Material identification using their unique spectral fingerprints is a promising area for this advanced imaging technique. One-pixel spectral image from HSI could reveal a lot more about the surface material than a regular range image.

Hyperspectral imaging's temporal resolution is sensor-specific and depends on its orbital properties. The time it takes for a sensor to go back to the same location and take readings is a common definition [29]. This period is also known as return time or revisit. When the sensor platform doesn't return to the same location very often, we say that the temporal resolution is low; when it returns frequently, we say that it has high temporal resolution. Common units of measurement for time resolution are days.

Benefits and drawbacks of hyperspectral scanners. The following are some of the benefits of adopting HSI in many fields, such agriculture [8] and food quality assessment:

1) The technology guarantees the eminence and security of food supplies without invasiveness, contact, or destruction.

2) Since no chemicals are used in the studies, they are safe for the environment.

3) There is a significant time savings compared to chemical and conventional methods when processing for quality valuation and food regulator/storing.

4) Chemical imaging provides a better thoughtful of chemical components of food goods.

5) It offers suitable region selection for crucial picture analysis.

6) It integrates spectral and geographical data to give better information about chemical samples. From relevant platforms, improve the likelihood of data refinement, and accomplish further experiments.

Hyperspectral imaging offers both benefits and drawbacks [30].

1) When compared to other methods of image processing, the price of a hyperspectral imaging system is quite high.

2) High-speed computers and solid-state drives with enormous storage capacities are in high demand due to the massive data sizes generated by hyperspectral imaging.

3) The signal-to-noise ratio may be low because environmental factors like scattering and lighting can affect the signal while collecting the photos.

4) It is typically challenging to detect and identify dissimilar items inside the similar image consuming spectral data except the individual objects consume distinct immersion properties.

B. Health Care Utilisation of RS and GIS

One new technology that the healthcare industry has found a wealth of uses for is remote sensing. The term "remote sensing" describes a method of retrieving data about Earth's surface that does not need physical contact. In most situations, data is gathered remotely via sensors attached to spacecraft or satellites [10]. Incorporating various types of electromagnetic radiation, such as visible light, infrared, and microwave, into the collected data can provide light on the Earth's surface and its characteristics [17]. When it comes to healthcare, RS is also useful for assessing environmental health issues. Environmental changes may influence public health, land use, and water and air quality can all be evaluated with its help [11]. Air pollution levels can be monitored via satellite-based remote sensing, which helps detect areas with poor air quality. This information can then be used to improve public health policies and actions. Further, RS data can be used to investigate environmental factors that may be influencing the incidence of respiratory diseases and other health problems.

Rapid surveying allows for the quick assessment of damage, localization of impacted regions, and population movement tracking. Timely deployment of healthcare resources and planning for healthcare infrastructure rehabilitation following a disaster both rely on this knowledge.

We can gain a more comprehensive and insightful understanding of healthcare-related topics by combining Geographic Information Systems (GIS) with Remote Sensing (RS). This methodology takes advantage of both technologies. This section explores the ways in which Geographic Information Systems (GIS) and Remote Sensing (RS) work together, and it contains real-life examples to show how these two tools can be used together. When combined, the complementary technologies of remote sensing and geographic information systems (GIS) can improve healthcare applications. One way to get environmental data and pictures in real-time is through remote sensing, which can be easily integrated into GIS systems. By placing RS data in a geographical context, GIS allows medical professionals to analyze, understand, and display data in a spatially-based manner.

Remote sensing can furnish imagery depicting variables such as pollution levels, land utilization, or temperature fluctuations. When incorporated into a GIS, this data can facilitate the identification of regions with elevated health risks and support targeted actions. The integration of remote sensing data, including satellite imagery of vegetation and land cover, with geographic information systems can facilitate the comprehension of disease vector habitats and the forecasting of disease outbreaks. This is essential for the management of vector-borne diseases. In the event of natural catastrophes or public health [6] emergencies, remote sensing data can furnish real-time information regarding impacted regions. This data, when incorporated into a GIS, facilitates the effective allocation of healthcare resources. Here we may see how RS and GIS have been put to use in real-world healthcare situations through empirical case studies. Consider the following: Temperature and precipitation are two of the most important environmental factors for mosquito breeding and malaria transmission; a study conducted in a malaria-endemic zone used RS to track these variables. Using GIS, we were able to create models that could predict when malaria epidemics would occur. With the use of early warning systems made possible by RS and GIS integration, hospital authorities were able to distribute resources for management and prevention. Data on urbanization and population growth were gathered using remote sensing (RS) within the context of urban planning. The availability of telehealth services in rapidly expanding urban areas was assessed using Geographic Information Systems (GIS). The study integrated RS and GIS to elucidate the equitable distribution of telehealth services, assisting policymakers in guaranteeing citizens' access to remote healthcare resources [9].

By allowing the geographical analysis of health-related data, Geographic Information Systems (GIS) have completely transformed the healthcare sector. Here takes a look at the big picture of GIS in healthcare and examine their many uses. Capturing, organising, analysing, and visualising spatial data is made easier by a GIS. Analysis best probable localities for healthcare services is one of the main usages of GIS in this commerce. GIS tools examine demographics, convenience, population solidity, and other criteria to determine whether locations have a high need for healthcare services. As a result, people will be able to get the medical treatment they need regardless of where their healthcare facility is located. In addition, GIS helps with healthcare facility management by minimising response times in emergency circumstances by optimising ways for healthcare specialists and ambulances. Healthcare admission and disparities assessments greatly benefit from GIS. By GIS places can be pinpoint wherever persons lack contact to healthcare by healthcare institutions mapping, patient demographics, and socioeconomic data. In order to reduce dissimilarities and assurance a fair dissemination of healthcare possessions, planners and policymakers must have this data.

In spatial epidemiology study GIS is vital for time and space in healthcare. In order to improve understand of geographical outlines of healthcare and their possible reasons, epidemiologists can utilise GIS to map and identify hotspots. In order to focus resources on areas most at danger, this data is priceless for prevention initiatives. These uses highlight the value of GIS in medical settings. When it derives to healthcare

capability locations, resource distribution and disease control GIS will provide professional tools to create data-driven decisions. When trying to make sense of healthcare inequalities and find ways to improve service delivery, having a spatial context is crucial [12].

Combining the strengths of remote sensing and GIS can improve healthcare outcomes may gain more thorough and insightful understanding. RS allows for the direct integration of real-time environmental data and images into GIS systems. On the flip side, healthcare providers can analyse prototype and visualise data in geospatial framework by means of GIS, which provides a spatial context for RS records. Pollution levels, land use, and temperature fluctuations are just a few of the factors that RS photography may reveal. Incorporating this data into a GIS can aid in the identification of high-risk areas for health and the subsequent implementation of targeted interventions. Disease vector habitats and outbreak predictions can be improved with the use of GIS and RS data, which contains satellite imagery of vegetation and land cover. Controlling diseases spread by vectors relies heavily on this. In the result of a public health emergency or natural catastrophe, RS data can give up-to-the-minute details about the areas impacted. The effective allocation of healthcare resources can be aided by including this data into a GIS.

Data on population growth and urban development were captured using RS in an urban planning framework. In order to determine how easily accessible telehealth facilities are in cities that experiencing high population progress GIS is used. The study helped to make sure that people could use remote healthcare resources by combining RS and GIS, which shed light on the fair allocation of telehealth services [13].

Although RS and GIS provide significant advantages to healthcare sector, their implementation is accompanied by hurdles and limits. This delineates the principal challenges encountered in the application of RS and GIS to healthcare initiatives.RS info can be influenced by features like cloud shield, sensor correction and atmospheric situations, potentially resulting in variations in data excellence. Contact to high-resolution remote sensing data is frequently constrained, especially for healthcare applications in underdeveloped regions where satellite or data archive access may be restricted. Utilizing RS and GIS in healthcare effectively necessitates specified skills and information that are not universally possessed by healthcare workers. Training and capacity growth are essential. The incorporation of RS and geographic information system data can be intricate, as it necessitates consideration of varying formats, projections, and scales. This necessitates a resilient IT infrastructure. The integration of geospatial data with healthcare information generates privacy problems. It is imperative to guarantee the ethical and secure utilization of such data. Establishing data ownership and sharing protocols can be complex, as remote sensing and geographic information system data frequently engage several stakeholders, including governmental entities, private enterprises, and academic institutes. Validating remote sensing-derived information with ground-based data can be complex, as it frequently necessitates substantial fieldwork and resources. The precision of predictive models utilizing RS and GIS data may fluctuate and requires thorough validation against empirical data. Confronting these obstacles and constraints is essential for the effective implementation of RS and GIS in healthcare. Cooperative initiatives among researchers, healthcare practitioners, and policymakers are essential to address these challenges and fully exploit the possible of emerging machineries for public health. Table I shows detailed comparison with previous work.

TABLE I. DETAILED COMPARISON WITH PREVIOUS WORK

Aspect	Current Work	Previous Work
Scope	HSI advanced spectral-spatial data is combined with remote sensing and GIS for tracking healthcare and the environment.	Mostly looked at how RS or GIS can be used alone in environmental studies or healthcare, without spectral imaging tools being added.
Integration of Data	Adds real-time RS data to GIS platforms and combines it with HSI for health and environmental apps that use maps.	In the past, methods focused on either GIS for geographic analysis or RS for environmental data, but they didn't include high-resolution spectral imagery (HSI).
Applications in Healthcare	RS-GIS systems that are fully integrated can be used to map disease vectors, plan for public health emergencies, and keep an eye on pollution.	focused more on environmental health factors like the quality of the air or water without using GIS to plan or allocate resources for healthcare.
Environmental Applications	RS and GIS are used together to keep an eye on air pollution, changes in land use, and the health effects of the climate.	Mostly looked at certain environmental measures (like deforestation) without incorporating them into many healthcare-related uses.
Real-World Case Studies	RS-GIS is used to predict malaria outbreaks, make sure that telehealth services are distributed fairly in cities, and map pollutants for public health tactics.	A lot of research has been done on static disease mapping or watching changes in the environment without using predictive modeling or frameworks that combine healthcare and the environment.

Asset Mapping is the method of community strengthening. The next step in asset mapping is to identify the institutions, citizen groups, and individuals within communities that can help put good resources in place. There is a great vibe in the community when assets are mapped out, and those assets may be leveraged to solve any problem. The community's healthcare facilities, parks, libraries, schools, police stations, grocery stores, and so on are its assets.

The primary objective of asset mapping is:

- The goal is to assess the community's current resources and use them to build it stronger.
- Among other things, to use the funds to find community connections and satisfy community needs.
- The community's resources should be acknowledged and appreciated.

C. About Dataset

The Dataset has the latitude and longitude positioning of healthcare unit of Tirupati district study area which is gathered form government portal. The study area selection is shown in Fig. 3.

In the southeastern corner of Andhra Pradesh Tirupati is the ancient holy city. This place is known as the abode of God Venkateswara. Conveniently located near numerous important cities, including Bangalore, Vijayawada, Hyderabad, and Chennai, at the base of the Eastern Ghats. Tirupati is famous for the Tirumala Venkateswara Temple, which is a major pilgrimage site in India and sees a large number of visitors annually. The Tirumala Hills, home to the temple, are among the world's oldest rock formations. People think that the temple even had followers who were members of great dynasties, such as the Pallavas, Pandyas, Cholas, and Vijayanagara rulers. The town's coordinates are 13° 37 N and 79° 25 E. Roughly 22,18 lakh people call Tirupati home, according to the 2011 Census. The winters are mild, but the summers can be quite hot. The official language is Telugu, but Tamil is also widely used due to the region's closeness to Tamil Nadu.

Some of the factors in the dataset are Location Coordinates, which are the healthcare facility's latitude and longitude. with Type of Facility: Which type of healthcare center it is (e.g., hospital, clinic, primary health center), as well as the area, district, state, and name of the hospital or PUC. The type of file is (.CSV). It's about 1GB in size.

Whereas hyperspectral image is in band sequential (BSQ), band-interleaved-by-pixel (BIP), or band-interleaved-by-line (BIL) format downloaded from bhuvan portal based on mapping of areas.

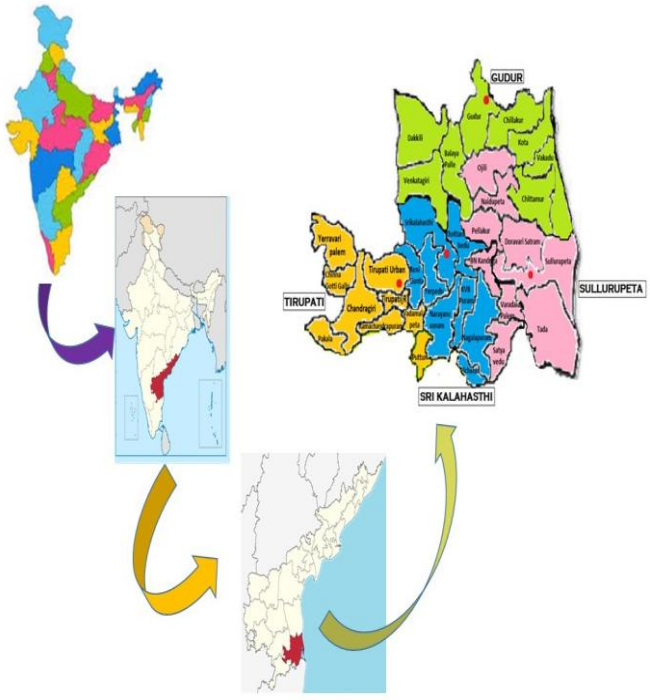


Fig. 3. Study area division (Image created by author).

III. Proposed Methodology

The key objective of our study is to map healthcare asset centred on the GIS and positions of healthcare asset in hyperspectral images to know the need of healthcare unit in the Tirupati District study area by using deep architectures base models. The methodology procedure is as shown in Fig. 4.

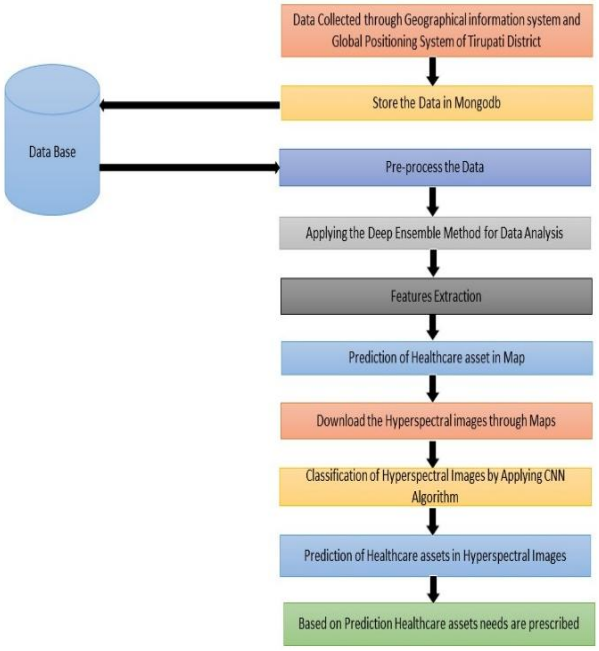


Fig. 4. Methodology procedure.

A. Deep Ensemble Method

The construct of heterogeneous ensembles with multiple base models and combine their results to make a more informed decision, taking advantage of recent architectural developments in GIS analysis. here makes use of both our fusing technique and classical fusers like XGBoost, AdaBoost. The method involves superimposing the model on top of the first ensemble. For an example of how it is trained on vectors of feature.

$$\mathscr{F} = [\boldsymbol{o}1;\ \boldsymbol{o}2;\cdots;\boldsymbol{o}N]^T$$

where $\boldsymbol{o}_i$ is vector output of i-th base model and considered feature vector consistent to ji. For each training sample, N base models are used to build a collective feature vector. This vector is sent to fusing learner. Consequently, larger values of N will result in feature vectors of greater dimension. For a classification assignment, for instance, the i-th base model would use the softmax layer to build c class probabilities, which would be represented by $\boldsymbol{o}_i$ (where $c \cdot N$ is the size of the concatenated feature selection). In contrast, a base boosting would provide more detail on HU's abundances, and all base models would use this same process to combine their abundance vectors. This would result in a final vector of profusions that is N times greater than the sum of the abundance vectors obtained from each vile model individually because it takes the outputs of N models and applies them all. In order to refine the ensemble's output, the fuser combines the initial projections.

B. Image Classification with Convolution Neural Network

Here for classification of hyperspectral images used three convolutional architectures for GIS analysis because it is the most efficient model to ensure that spectral and spectral-spatial generalise. In contrast to the spectral network [7], which is based on [22], two spectral-spatial CNNs are 3D-CNN [18] and 2.5D-CNN [17]. The former model does pixel-wise classification, while the latter performs patch-wise cataloguing of central pixel in each consistent area. The area sizes for 2.5D-CNN were determined according to the recommendations in [19]. While both the 3D-CNN and 2.5D-CNN models use spatial and spectral info to categorise the input patch's central pixel, the 3D-CNN model takes advantage of small (3×3×3) convolutional kernels to capture hyperspectral cube's fine-grained spectral interactions. It differs from 2.5D-CNN, which uses kernels that cover the complete spectrum, i.e., μ. categories in the initial convolutional layer.

C. Performance Metrics

Accuracy is defined as the number of right predictions out of all the instances.

$$Accuracy = \frac{True\ Positive + True\ Negatives}{Total\ Instances}$$

Precision is the number of successfully predicted positive cases out of all the predicted positive cases.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall is the number of properly predicted positive cases to all actual positive cases.

$$Recall = \frac{True\ Positives}{True\ Positives + false\ Negatives}$$

The F1 Score is the harmonic mean of accuracy and memory, which makes it a fair measure.

$$F1\ Score = 2 * \frac{precision * recall}{precision + recall}$$

In multi-class classification problems, the Macro Average is a performance measure that gives an overall picture of a model's accuracy, recall, or F1 score by giving each class the same weight, no matter how big it is or how many times it appears in the dataset.

$$M_{macro\ avg} = \frac{1}{N}\sum_{i=1}^{N} M_i$$

N represents number of Class

M_i *represents Metric value for class i*

In multi-class classification, the Weighted Average is another way to measure success. It figures out the general metric (like F1 score, precision, or recall) by giving each class a weight based on how much data it has.

$$M_{weighted\ avg} = \frac{\sum_{i=1}^{N} M_i \,.\, n_i}{\sum_{i=1}^{N} n_i}$$

N represents number of Class

M_i *represents Metric value for class i*

$n_i$represents of samples in class i

IV. Experimental Analysis

To map the exact assets, execution is processed by using different classifiers and feature selection measures on two different datasets. Both trials included healthcare unit datasets from the Tirupati district and used hyperspectral image data for categorization and location analysis. This section presents the experimental setup and provides a detailed discussion of the outcomes. We used Python for model coding and QGIS for experimental validation.

At first healthcare facilities asset dataset is taken from Andhra Pradesh state government and it has created through GIS with the latitude and longitude as features. These are stored in mongodb for easy accessing and retrieving of data for the analysis. Applying ensemble classifiers on the data set the predictive model will be constructed. The deep ensemble method which is generated by using XGBoost, AdaBoost. From the data set, 28 features related to the healthcare assets are extracted. The Features are extracted to reduce the quantity of resources needed without losing valuable information. The feature extraction graph for the healthcare assets shown in Fig. 5.

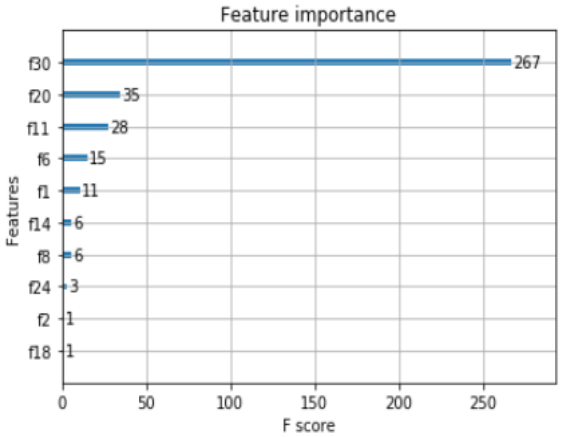


Fig. 5. Extracting the features from deep ensemble method.

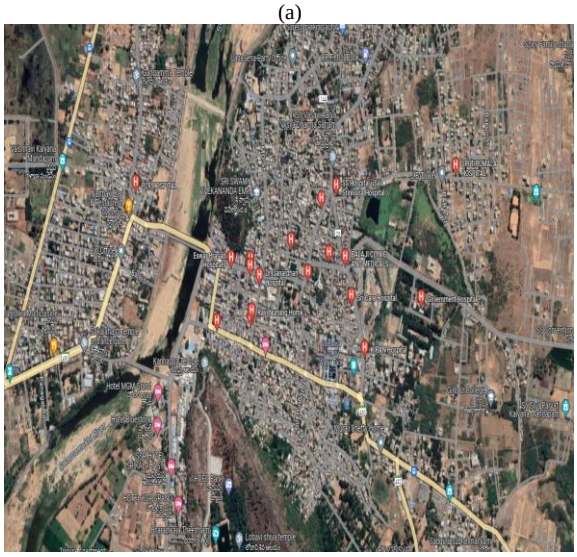
TABLE II. Classification Report of Deep Ensemble Method for Healthcare Asset

	Precision	Recall	F1-Score	Support
Class 0	0.65	0.43	0.51	76
Class 1	0.82	0.92	0.89	194
Accuracy			0.84	256
Macro Avg	0.74	0.71	0.79	256
Weighted Avg	0.84	0.89	0.91	256

When deep learning algorithm is applied on healthcare asset data, classification report is generated with different parameters as shown in Table II like accuracy, weighted avg, macro avg, with precision, F1-Score, Recall by observing table deep ensemble algorithm has the highest accuracy of 84%.

Then after analysing deep ensemble algorithm with healthcare assets GIS data, the dataset is accessed into QGIS in python using interconnection and visualised the hospitals with their locations as shown in Fig. 6 (a) and (b).

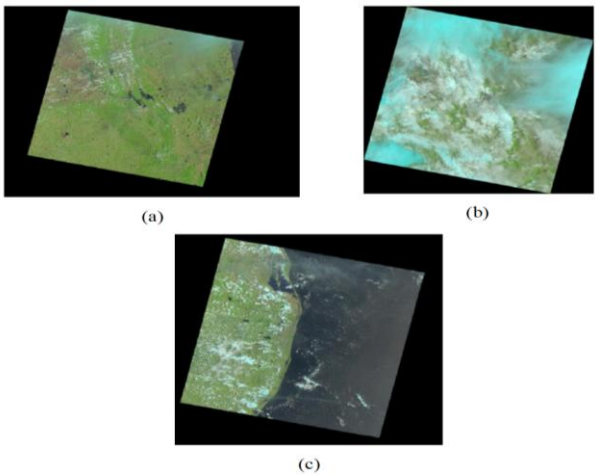
(a)



(b)

Fig. 6. (a): Visualization of Tirupati region healthcare assets. (b): Visualization of Srikalahasti region healthcare assets.

By observing Fig. 6(a) and 6(b) the red color pin with H letter indicates the healthcare assets. In Fig. 6(a) the healthcare assets are located either all the healthcare asset is nearer or for away from the village at tirupati city area. In Fig. 6(b) the healthcare assets are located for way from the village at Srikalalhasti. Then Hyperspectral images are acquired by integrating Python programming with QGIS. This study involves the collection of four distinct Landsat images of Tirupati district to facilitate the classification of healthcare facilities within the region. The hyperspectral images were collected and subsequently pre-processed through various techniques, including data cleaning, integration, transformation, and reduction. The images are subsequently divided into cubes by defining a region of interest (ROI), which is quantified using shapes such as polygons, circles, and rectangles. The subsequent step involves segmenting the images and employing a CNN algorithm to classify the healthcare facilities within the Tirupati district. The landsat input images of tirupati district are shown in Fig. 7 and the classified landsat image are shown in Fig. 8.



(a) (b) (c)

Fig. 7. Tirupati district hyperspectral images.

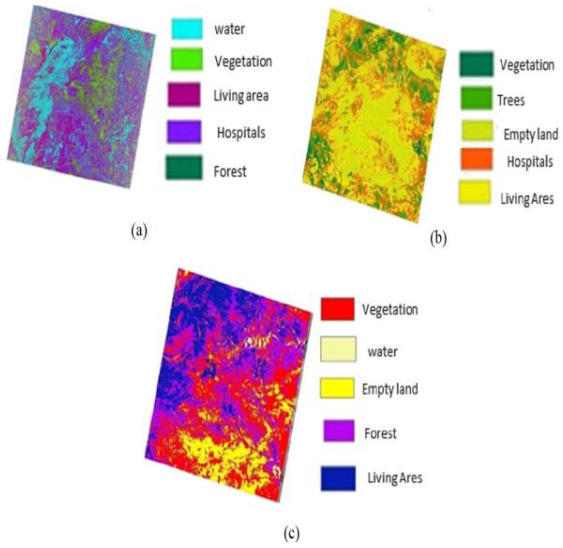


Fig. 8. Tirupati district classified hyperspectral images.

The classification of the healthcare asset in hyperspectral images from three distinct regions of the Tirupati district is presented in Fig. 8, specifically in Fig. 8 (a), Fig. 8 (b), and Fig. 8 (c). In this section, each figure illustrates the classified areas, depicted in various colors, with each color corresponding to a distinct area within that region, such as: In Fig. 8 (a), sky blue colour signifies water, green color indicates vegetation, Marron color represents Living area, Violet color represents Hospitals and Dark green color represents forest. In Fig. 8 (b), Dark green color represents vegetation, green color indicates trees, musted colour represents empty land, Orange color represents Hospitals and yellow color represents Living area. In Fig. 8 (c), Red color symbolizes vegetation, light yellow color indicates water, yellow colour represents empty land, Violet color represents forest and Blue color represents Living area. At last, explored how close healthcare services are to cities and rural areas. With spectral and geographic integration, it was possible

to accurately identify and classify healthcare assets using their locations.

Here, in this work Two sets of healthcare facility data from the Tirupati district (Andhra Pradesh state government) were handled using GIS tools and hyperspectral imaging data to figure out where they are. The dataser took 28 traits from the healthcare dataset. Then, Deep ensemble algorithms (XGBoost, AdaBoost) were used to save resources while keeping important data. The results were shown as feature extraction graphs, and the Deep ensemble method (XGBoost and AdaBoost) got the best results with an 84% success rate. Some of the measures in the classification report are precision, recall, F1-score, weighted average, and macro average. Then Landsat pictures of the Tirupati district were edited and looked at. Steps like cleaning, integrating, transforming, and reducing that were part of the pre-processing are now complete. The data was then split into ROI cubes that were measured with circles, squares, and polygons. CNN used an algorithm to sort healthcare sites into groups. The classification results for three different areas in Tirupati district are shown. Based on the observation the need of healthcare facility is there in this study area.

Because, as per 2024 Censes, the population of tirupati distict is around 77,50,000 (approx.). so by observing our GIS and HSI analysis of healthcare assets it concludes that all the healthcare facilities are either in one place or it in different distance places which are outside the living area. So mainly for every village there is need to have healthcare facility centre for the population who is living there.

V. Conclusion

The deep learning techniques have proven useful for GIS and HSI data sorting and has some limits to use the large-capacity initiates and calls their generalizability into doubt in real-world scenarios. An exciting new paradigm is emerging in medical research and practice at the intersection of deep learning, Big data, GIS, HSI, and healthcare. This convergence has great promise for improving patient outcomes, streamlining healthcare delivery, and addressing public health issues. here presented deep ensembles for GIS analysis, which incorporate as effective base model fusers and other deep architectural advancements to capture and extract needful futures for location identification of healthcare facilities around the study area tirupati district. At last it concludes that there is a need of constructing the healthcare facilities in the tirupati district because as per 2024 censes, the population of tirupati distict is around 77,50,000 (approx.). so by observing our GIS and HSI analysis of healthcare facilities all the healthcare facilities are either in one place or it in different distance places which are outside the living area. So mainly for every village there is need to have healthcare facility centre for the population who is living there.

References

[1] Khashoggi, B. F., & Murad, A.," Issues of healthcare planning and GIS: a review. ISPRS", International Journal of Geo-Information, 9(6),2022, pp. 352.

[2] Botín-Sanabria, D. M., Mihaita, A. S., Peimbert-García, R. E., Ramírez-Moreno, M. A., Ramírez-Mendoza, R. A., & Lozoya-Santos, J. D. J. ," Digital twin technology challenges and applications: A comprehensive review", Remote Sensing, 14(6), 2022, pp.. 1335.

[3] Wei, L.; Meng, L.; Tianhong, C.; Zhaoyao, C.; Ran, T.," Application of a hyperspectral image in medical field: A review. ", J. Image Graph, 26, 2021, PP. 1764–1785.

[4] Fei, B.," Hyperspectral imaging in medical applications. In Data Handling in Science and Technology", Elsevier: Amsterdam, The Netherlands, Vol. 32, 2019, pp. 523–565.

[5] Lu, G.; Fei, B.," Medical hyperspectral imaging: A review.", J. Biomed. Opt, 19, 2014,pp. 010901.

[6] Saini, J., Dutta, M., & Marques, G.," A comprehensive review on indoor air quality monitoring systems for enhanced public health. Sustainable environment research, 30(1), 2020, pp. 1-12.

[7] Wu, A. N., Stouffs, R., & Biljecki, F.," Generative Adversarial Networks in the built environment: A comprehensive review of the application of GANs across data types and scales", Building and Environment, 2022, pp. 109477.

[8] Iqbal, A., Zahid, S. B., Iqbal, J., & Khalil, A. ," Exploring Advanced Techniques for Agricultural Area Mapping: Comparative Analysis of U-Net and Mask-Based Approaches in Satellite Imagery", International Journal of Computer Science And Technology, 7(3), 2023, pp.192-196.

[9] Laney D.," 3D data management: controlling data volume, velocity, and variety, Application delivery strategies.", Stamford: META Group Inc, 2001.

[10] Mauro AD, Greco M, Grimaldi M.," A formal definition of big data based on its essential features. Libr Rev; 65(3), 2016, pp.122–35.

[11] Gao, Q.; Lim, S.; Jia, X.," Hyperspectral Image Classification Using Convolutional Neural Networks and Multiple Feature Learning.", Remote Sens, 10, 2018, 299.

[12] Nalepa, J.; Tulczyjew, L.; Myller, M.; Kawulok, M. ," Hyperspectral Image Classification Using Spectral-Spatial Convolutional Neural Networks", IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA;, 2020, pp. 866–869.

[13] Nalepa, J.; Myller, M.; Cwiek, M.; Zak, L.; Lakota, T.; Tulczyjew, L.; Kawulok, M.," Towards On-Board Hyperspectral Satellite Image Segmentation: Understanding Robustness of Deep Learning through Simulating Acquisition Conditions. Remote Sens, 13, 2021, pp.1532.

[14] T. sarath, G. Nagalakshmi, S. Jyothi, "A Study on Hyperspectral Remote Sensing Classifications", International Journal of Computer Applications, 2014.

[15] P.Bhargavi, A.Rajitha, S.Jyothi "Hybrid Algorithm for classifying soil types using hyperspectral images", Journal of Survey in Fisheries Sciences, Vol.10 No.1, 2023, pp. 2028-2034.

[16] P. Bhargavi, A. Rajitha, S. Jyothi, (2022) "Hyperspectral Image Classification of Soils using Soft Computing Techniques", Journal of Pharmaceutical Negative Results, Volume 1, Special Issue 6, Pp. 1125-1132, ISSN: Print -0976-9234, Online - 2229-7723.

[17] VinilaKumari S., Bhargavi P., Jyothi S., (2021) "Deep Geospatial Analysis for Land Use and Land Cover", Journal of University of Shanghai for Science and Technology, Volume 23, Issue 2, Pp:411-422, ISSN: 1007-6735.

[18] Sree Divya. K, P. Bhargavi, S. Jyothi," Comparative Analysis of Classifiers and Ensemblers in Asset Mapping in Big Data", International Journal of Advanced Research in Engineering and Technology (IJARET), Volume 12, Issue 1, 2012, pp. 971-980.

[19] Sree Divya. K, P. Bhargavi, S. Jyothi, "Asset Mapping Using K-NN to Evaluate the Distance Measure Between Assets", International Journal of Future Generation Communication and Networking Vol. 13, No. 4, 2020, pp. 2587–2597.

[20] VinilaKumari S., Bhargavi P., Jyothi S., "Identification of Neighbourhood Cities Based on Landuse Bigdata Using K-Means and K-NN Algorithm", In: Jyothi S., Mamatha D., Satapathy S., Raju K., Favorskaya M. (eds) Advances in Computational and Bio-Engineering. CBE 2019. Learning and Analytics in Intelligent Systems, vol 15. 2020.

[21] Sree Divya K., Bhargavi P., Jyothi S. (2020) XGBoost Classifier to Extract Asset Mapping Features. In: Jyothi S., Mamatha D., Satapathy S., Raju K., Favorskaya M. (eds) Advances in Computational and Bio-Engineering. CBE 2019.

[22] S. Vinila Kumari, Dr. P. Bhargavi & S. Jyothi, (2020) "Land Classification Based on Hyper Spectral Images using Deep Learning

Techniques", Test Engineering and Management, Volume 83, 2020, pp. 5722 – 5727.

[23] G. Nagalakshmi, T. Sarath, S. Jyothi "Land Site Image Classification using Machine Learning Algorithms". In Proceedings of international Conference on Computational Bio Engineering Learning and Analytics in Intelligent Systems (Springer), Vol. 1, 2019.

[24] K. Himabindu, T. Sarath "ISODATA classification using Fuzzy Logic", in ADBU Journal of Engineering Technology (AJET), Vol. 9(2), 2002.

[25] T. Sarath, G. Nagalakshmi"An Land Cover Fuzzy Logic Classification by Maximum Likelihood" in International Journal for Computer Trends and Technology ,Vol. 13(2) , 2014, PP:56 – 60.

[26] R.C. Gonzalez, Digital Image Processing, Pearson, London, U.K., 2009.

[27] R. Smith, Introduction to Remote Sensing of the Environment, 2001 [Online]. Available: http://www.microimages.com.

[28] J.R. Jensen, Introductory Digital Image Processing: A Remote Sensing Perspective, Pearson, London, U.K., 2005.

[29] J. Thau, "Temporal resolution,'', in: Encyclopedia of GIS, Springer, New York, NY, USA, 2008, pp. 1150–1151, https://doi.org/10.1007/978-0-387-35973- 11376 [Online].

[30] Guolan Lu, Baowei Fei, Medical hyperspectral imaging: a review, J. Biomed. Opt. 19 (1) (20 January 2014) 010901, https://doi.org/10.1117/1.JBO.19.1.010901.

Path Planning for Laser Cutting Based on Thermal Field Ant Colony Algorithm

Junjie GE, Guangfa ZHANG, Tian CHEN
College of Mechanical Engineering, Shanghai Dianji University, Shanghai 201306, China

***Abstract*—In laser cutting technology, path planning is the key to optimizing cutting quality. Traditional ant colony optimization path planning does not prevent excessive heat effects after processing. This paper addresses the problem of heat accumulation during drilling by introducing a heat factor and a heat threshold into the traditional ant colony algorithm. The heat factor and threshold are used to dynamically control heating and cooling in the path planning process, and the heat factor is updated to update the local pheromone. Then, the improved 2-opt algorithm with the introduced heat factor is combined to parallelly optimize the path, and a thermal field ant colony algorithm is proposed. The simulation experiments and actual cutting results show that the proposed algorithm is more efficient and effective than traditional ant colony algorithm and improved ant colony algorithm in terms of reducing heat accumulation while ensuring fewer empty path, and improving laser cutting processing efficiency and quality.**

***Keywords*—Laser cutting; path planning; ant colony algorithm; thermal field control method**

I. Introduction

Laser cutting technology, known for its high precision and efficiency, is widely used in the manufacturing industry. Path planning plays a crucial role in achieving optimal cutting performance. Traditional path planning methods often fail to reach the optimal solution when dealing with complex conditions, which has led to increasing attention on intelligent optimization algorithms in recent years. Luodengcheng et al. [1] proposed a method that incorporates a potential field factor into the ant colony algorithm to reduce heat accumulation and non-cutting travel distances, thereby ensuring cutting quality. Chang Cuizhi et al. [2] introduced an annealing algorithm to optimize the cutting path of the generalized traveling salesman problem, effectively preventing tool retractions. Makbul Hajad et al. [3] treated all pixels in an image as potential perforation positions and proposed a simulated annealing algorithm combined with adaptive large neighborhood search, which reduced both computational time and path length. Bonfim Amaro Junior et al. [4] improved the heuristic search in the inheritance algorithm, yielding better solution quality and shorter computation time in laser cutting path planning. Song Lei et al. [5] designed a dual-chromosome encoding mechanism to jointly solve the cutting starting point and path planning problem, reducing both non-cutting travel distances and processing temperature. Wu Yanming et al. [6] employed a greedy algorithm for plate processing path planning, optimizing parameter grouping to achieve the optimal path while ensuring high-quality processing. Zhou Rui et al. [7] improved the genetic algorithm to avoid non-cutting travel and cutting conflicts, enhancing the algorithm's convergence.

The intelligent algorithms used in path planning above aim to minimize non-cutting travel and avoid heat accumulation [8]. However, while preventing heat accumulation, they do not necessarily minimize non-cutting travel and computational time. Through the detection of heat accumulation during actual laser cutting, it has been observed that the temperature is highest at the perforation starting point of the cutting element. To address the phenomenon of "laser punching" caused by excessive temperature at the perforation, this paper proposes a thermal field ant colony algorithm. This algorithm incorporates a heat matrix when initializing ant colony parameters, optimizing the probability of visiting cities and updating local pheromones. At the same time, a heat factor-based heuristic algorithm is introduced to parallelly optimize the cutting path.

II. Laser Cutting Path Planning and Design

The laser cutting process flow [9] is shown in Fig. 1. When cutting thicker materials such as carbon steel and stainless steel, achieving high precision and quality requires the incorporation of perforation techniques into the cutting parameter setup [10]. Perforation techniques mainly include pulsed perforation and explosive perforation. In pulsed perforation, a high peak power pulsed laser beam strikes the material surface, rapidly melting or vaporizing the material in a localized area to form a small hole. Explosive perforation, on the other hand, uses a lower-power continuous laser beam to act on the material surface, continuously heating it until the material melts and forms a hole. However, improperly set perforation parameters can lead to heat accumulation on the material surface [11], causing thermal expansion and deformation of the material. This results in a decrease in cutting precision and may even damage components such as optical lenses and focusing lenses, thereby increasing equipment maintenance costs. Therefore, path planning before setting cutting parameters is particularly important.

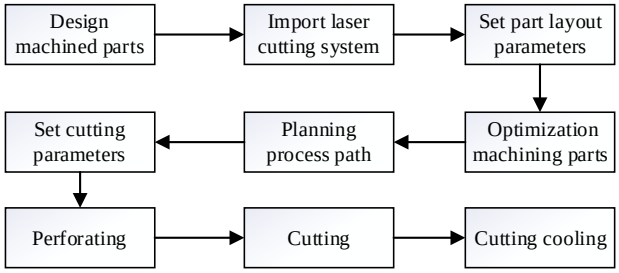


Fig. 1. Laser cutting process.

A. Laser Cutting Path Modeling

First, the feature points of each graphic element are obtained as the starting coordinates for the path planning of each element. For closed polygonal elements, the element contour is decomposed into arcs and straight lines, and the connection points between these lines and arcs are taken as the feature points of the element. For closed circles and ellipses, their feature points are defined by dividing the shape into four equal parts. The start and end points of an arc are calculated as shown in Eq. (1).

$$\begin{cases} x_s = x_c + R \cdot \cos(\theta_s) \\ y_s = y_c + R \cdot \sin(\theta_s) \\ x_e = x_c + R \cdot \cos(\theta_e) \\ y_e = y_c + R \cdot \sin(\theta_e) \end{cases} \tag{1}$$

In the above formula, $\{x_c, y_c\}$ represents the coordinates of the arc center, RRR is the radius of the arc, and (θ_s, θ_e) denote the start and end angles of the arc. $\{x_s, y_s\}$ and $\{x_e, y_e\}$ represent the coordinates of the arc's starting and ending points, respectively.

Using this formula, the feature points of the part contours are determined, as shown in Fig. 2. The closed elements are defined as {Part 1, Part 2, Part 3}, with each closed element corresponding to city coordinate points as follows: {v11, v12, v13, v14, v15, v16, v17, v18, v19}for Part 1, {v21, v22, v23 }for Part 2, and {v31, v32, v33, v34 } for Part 3.

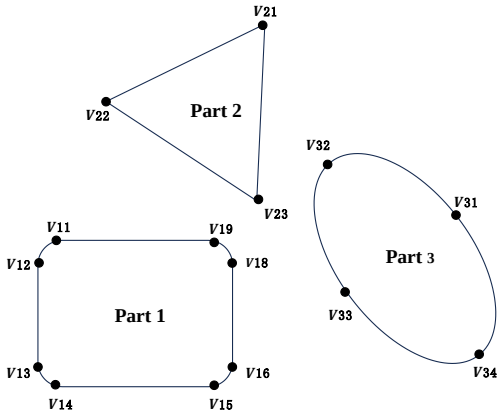


Fig. 2. Method of obtaining feature points of part contour.

Let the total path length L be defined as the total distance traveled by the laser, including all cutting path segments and non-cutting movement path segments from the starting point. The total cutting length is expressed in Eq. (2):

$$L = \sum_{i=1}^{n} d_i + \sum_{j=1}^{m} m_j \tag{2}$$

In this equation, d_i represents the length of the cutting path required for the i-th part, n is the total number of parts in the cutting path, m_j represents the length of the non-cutting movement path to the j-th part, and m is the total number of non-cutting paths in the cutting sequence.

B. Thermal Field Modeling

During the cutting process, if the perforation time is set too long or the perforation power too high, the starting area of the cut part can become overheated, causing a "punching" phenomenon. Therefore, thermal factors in laser cutting also need to be considered in path planning. The temperature rise ΔT at a point on the material at time t can generally be estimated using the heat conduction formula [12]. Assuming an initial temperature T of the material, the temperature rise at a certain point under the effect of the laser heat source is given by Eq. (3):

$$\Delta T = \frac{P \cdot \exp\left(-\frac{r^2}{4 \cdot a \cdot t}\right)}{\rho \cdot c \cdot \sqrt{4 \cdot \pi \cdot a \cdot t}} \tag{3}$$

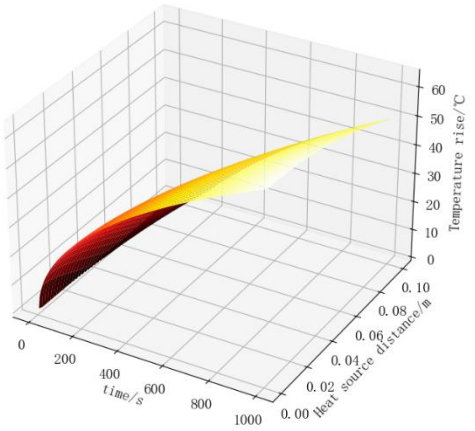
In this equation, P is the laser power, ρ is the material density, c is the specific heat capacity of the material, α is the thermal diffusivity, and r is the distance from the point to the laser heat source.

During laser cutting, heat is mainly dissipated through convective heat transfer to the surrounding air and through thermal conduction within the material itself [13]. Based on the above formula for temperature rise, the cooling temperature drop can be expressed as Eq. (4):

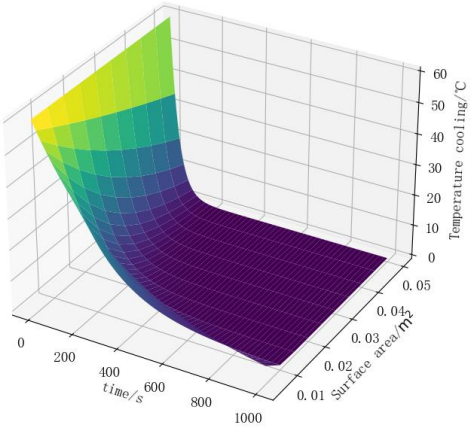
$$\Delta T = \exp\left(-\frac{h \cdot A}{m \cdot c} \cdot t\right)\left(T_S - T_N\right) + \varepsilon \cdot \sigma \cdot A \cdot \left(T_S^4 - T_N^4\right) \tag{4}$$

where h is the convective heat transfer coefficient, A is the surface area, m is the material mass, ε is the emissivity of the material, σ is the Stefan-Boltzmann constant, T_s is the surface temperature of the material, and T_N is the ambient temperature.

Given the parameters for heating with a laser power P =1000 W, a cutting material of steel with density ρ = 7800 kg/m3, specific heat capacity c= 500 J/kg · K, and thermal diffusivity α=1×10−5 m2/s. For cooling, assume the surface temperature Ts=60 ◦ C, ambient temperature T_N = 20 ◦ CT, convective heat transfer coefficient h=25 W/m2 · K, emissivity ε=0.7, Stefan-Boltzmann constant σ= 5.67×10−8 W/m2 · K4, and material mass m=0.1 kg. The resulting function plots for the temperature rise and cooling temperature drop of the material during laser cutting are shown in Fig. 3.



(a) Plate temperature function



(b) Plate cooling function

Fig. 3. Temperature change of laser cutting sheet.

To effectively control heat accumulation in path planning, it is essential to treat the effect of the laser heat source as a thermal conduction process. Based on the model described in Eq. (3), we can estimate the temperature rise at each point along the cutting path and avoid reheating high-temperature areas during path planning. To achieve this, we define a heat threshold in the thermal-field ant colony algorithm. If the temperature rise ΔT at a given point exceeds this threshold, the path planning algorithm should avoid immediate reprocessing of that point, instead prioritizing areas with lower heat levels for cutting. This approach helps to control heat distribution effectively. In addition to considering the temperature rise caused by laser power, the cooling process's impact on temperature must also be taken into account. This heat threshold can be dynamically adjusted by incorporating the cooling effect as described in Eq. (4), allowing for effective control of overheated areas during the cutting process.

III. Thermal Field Ant Colony Algorithm

A. Description of Thermal Field Ant Colony Algorithm t

Ant Colony Optimization (ACO) [14] is a stochastic optimization algorithm based on simulating the foraging behavior of ants, introduced by Marco Dorigo in the 1990s. This algorithm is particularly suitable for combinatorial optimization problems, such as path optimization. However, ACO tends to get stuck in local optima, has a slower convergence speed, and does not account for heat accumulation in the material during laser cutting. To address these issues, this study incorporates thermal fields and a heat factor into the ACO. Each "city" has a thermal field within a certain range, wherein neighboring points influence each other's temperature. The probability of each ant selecting a city is affected by the heat factor, and a heat threshold is set—when the heat on a path segment exceeds this threshold, the algorithm reduces the probability of selecting this segment, favoring cooler regions to control heat distribution. The heat factor $h(i,j)$ is calculated as shown in Eq. (5):

$$h(i,j)=\begin{cases} e^{-H[i,j]}, -H[i,j] > heat_{\max} \\ 1, others \end{cases} \tag{5}$$

where i is the index of the current city, j is the next city chosen by the ant, $H[i,j]$ represents the thermal matrix between these cities, and $heat\_max$ is the maximum heat threshold between two cities.

The probability of city selection in the heat-field ACO, modified by the heat factor, is shown in Eq. (6):

$$P_{ij}=\frac{(\tau_{ij})^{\alpha}\cdot(\eta_{ij})^{\beta}\cdot h_{ij}}{\sum_{k\in allowed}(\tau_{ij})^{\alpha}\cdot(\eta_{ij})^{\beta}\cdot h_{ik}} \tag{6}$$

where τ_{ij} represents the pheromone concentration between cities i and j, η_{ij} is the heuristic information, and α and β are importance factors for pheromone and heuristic information, respectively. *allowed* is the set of all cities available for selection, and k is the index of selectable cities.

To enhance global search capability and identify optimal paths, a roulette-wheel selection [15] is employed for probabilistic city selection, as mathematically expressed in Eq. (7):

$$P_{ck}=\frac{P_{ij}}{\sum_{i=1}^{n}P_{ij}} \tag{7}$$

where P_{ij} is the probability calculated in Equation (6), $r \in [0,1]$ is a random number, and k is the minimum index for which $P_{ck} \geq r$, making k the selected target city's index.

In the heat-field ACO, after each ant selects the next city, it locally updates the pheromone concentration on that path segment. This local pheromone update balances pheromone accumulation and evaporation to prevent path selection bias caused by a single heat factor. Following each iteration, the global pheromone is updated to speed up convergence to the optimal solution, as shown in Eq. (8):

$$\tau_{ij} = \begin{cases} (1-\rho_m)\cdot\tau_{ij} + \rho_m \cdot \tau_i, \\ \tau_{ij} \in local \\ (1-\rho_g)\cdot\tau_{ij} + \rho_g \cdot \Delta\tau_{ij}, \\ \tau_{ij} \in global \end{cases} \quad (8)$$

where ρ_{m} is the local pheromone evaporation factor, τ_{i} is the initial pheromone concentration to prevent pheromone levels from dropping too low, ρ_{g} is the global pheromone evaporation factor, and $\Delta\tau_{ij}$ is the additional pheromone added along the optimal path.

As ants pass each city, a heating and cooling mechanism is applied to the thermal matrix to reflect accumulated heat from laser cutting, preventing multiple selections of high-heat paths. Meanwhile, during city searches, already visited cities and their surrounding areas cool down, controlling excessive heat accumulation and improving path selection quality. The heating and cooling mechanisms for the thermal matrix are shown in Eq. (9) and Eq. (10):

$$\Delta H_{ij} = P \cdot \exp(-\frac{d_{ij}^2}{\gamma \cdot t}) \quad (9)$$

$$H[i,j] = \begin{cases} H[i,j]\cdot(1-C)\cdot t, \\ H[i,j] \geq heat_{\min} \\ heat_{\min}, \\ others \end{cases} \quad (10)$$

where ΔH_{ij} is the thermal matrix increment, P is the heat coefficient, d_{ij} is the distance between cities i and j, t is the step count, γ is the diffusion coefficient controlling heat dissipation speed, C is the cooling coefficient, and $heat\_min$ is the minimum heat threshold.

To avoid local optima, this study introduces an improved 2-opt [16] algorithm after each ant completes its path based on pheromone selection. Fig. 4 illustrates the 2-opt algorithm optimized path search process.

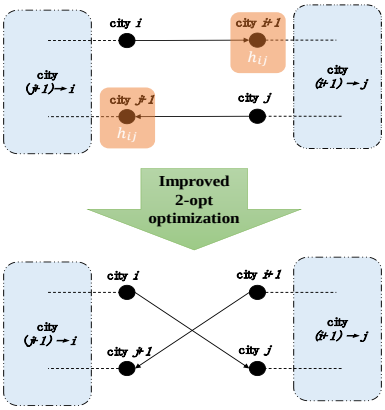


Fig. 4. Improved 2-opt algorithm to optimize the path method.

The core idea is to select two sub-paths on a given route and reverse one portion to check if a shorter total path length can be achieved. Whether to accept the new path is determined by the thermal matrix in Eq. (5), with the acceptance probability shown in Eq. (11):

$$P_{ij} = \exp(-\frac{L_{i,i+1} - L_{i,i}}{h_{ij}}) \quad (11)$$

where $L_{i,i+1}$ is the path length between city i and $i+1$, $L_{i,j}$ is the path length between cities i and j, and h_{ij} is the heat factor from Eq. (5).

B. The Implementation Steps of Path Planning of Thermal Field Ant Colony Algorithm

The path planning in the thermal-field ant colony algorithm first extracts the feature points of each part after laser cutting layout, as illustrated in Fig. 4. These feature points are placed into corresponding arrays to avoid redundant selection of the same geometric element later in the process. Then, the Euclidean distance matrix for each pair of feature points is computed, and the pheromone matrix is initialized. Since solution construction among ants is independent, the optimal path for each ant is calculated in parallel during a single iteration, which reduces computation time. The process terminates upon reaching the maximum number of iterations, after which the part numbers are sorted by shortest path order and output. The detailed procedure is depicted in Fig. 5.

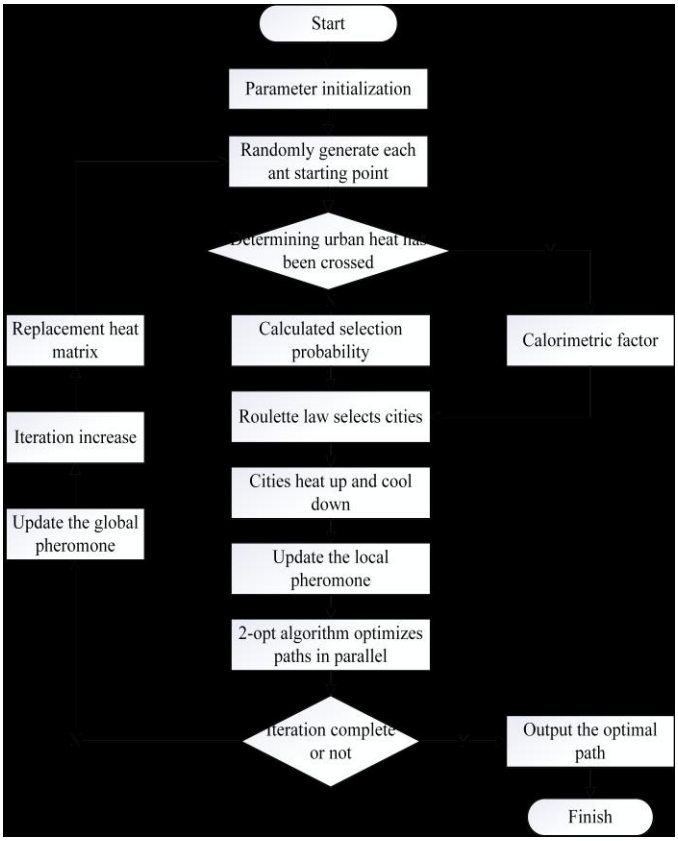


Fig. 5. Algorithm flow chart.

IV. Experiment and Result Analysis

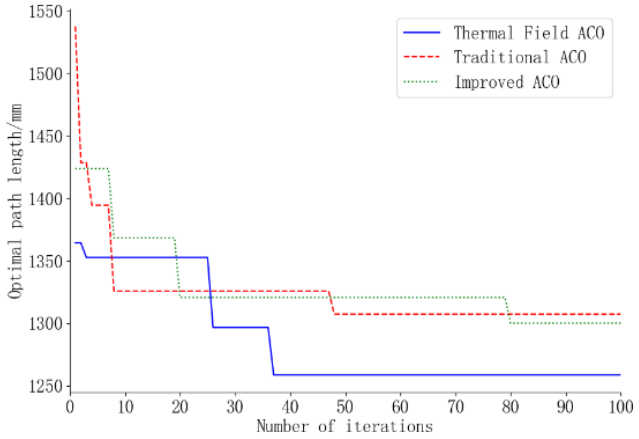
A. Path Planning Experiment

According to the path planning algorithm in this paper, the experimental objects are the parts after nesting, placed on the cutting plate. The feature points are extracted from the DXF file using geometric information as city coordinates. The operating system used in this experiment is Windows 11, with an Intel i5-12400F CPU, and the development environment is PyCharm. Based on the study [17] and multiple tests, the parameters of the algorithm were adjusted to their optimal values: the number of ants $m = 50$, pheromone factor $\alpha = 1$, heuristic function factor $\beta = 5$, global evaporation factor $\rho_g = 0.1$, local evaporation factor $\rho_m = 0.05$, and the maximum number of iterations $iter_{max} = 100$. The thermal field parameters are listed in Table I. The initial thermal matrix H in the table has n rows and n columns, where n represents the number of cities, the number of feature points in the DXF layout.

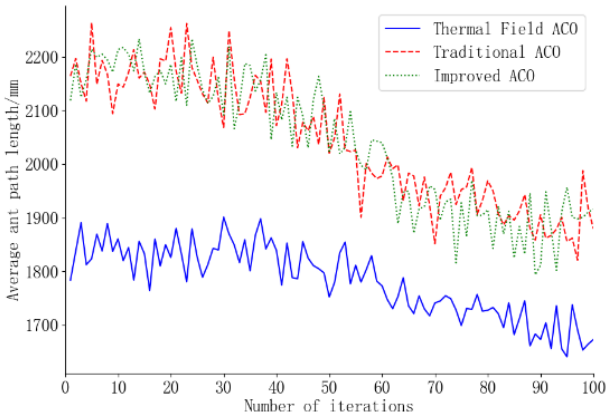
TABLE I. Thermal Field Parameters

Parameters	Value
Initial heat matrix H	$\begin{bmatrix} 0 & 1.0 & \dots & 1.0 \\ 1.0 & \ddots & \dots & \vdots \\ \vdots & \dots & \ddots & 1.0 \\ 1.0 & \dots & 1.0 & 0 \end{bmatrix}$
Heat threshold $heat_{max}$	0.5
Heating factor P	0.2
Cooling factor C Heat loss γ	0.95 0.05

The result of the experiment is to calculate the time, the path length and the heat matrix when cutting the plate. The optimal path ant heat value of Thermal Field ACO(TF-ACO) is shown in Fig. 6. Compared with Traditional ACO(T-ACO) and Improved ACO(I-ACO), the convergence curve of the improved ant colony algorithm based on study [18] and the convergence curve of the shortest path and the average path of ants is shown in Fig. 6:



(a) Shortest path convergence curve comparison.



(b) Comparison of ant average path convergence curves.

Fig. 6. Comparison of convergence curves.

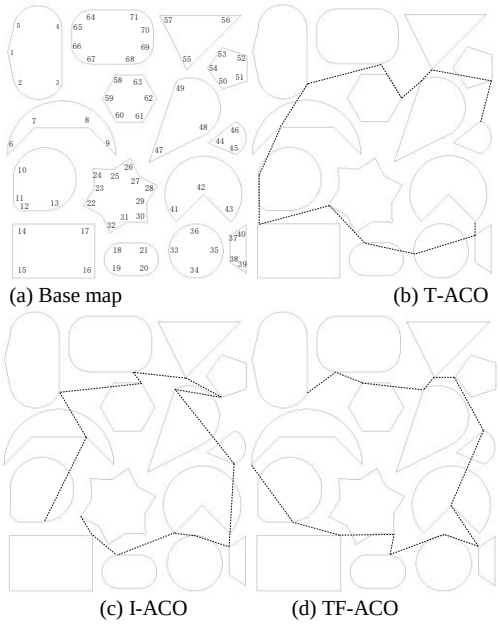
When compared to the traditional ant colony algorithm and the improved ant colony algorithm, the convergence curves of the shortest path and the average path length of the ants for the improved algorithm are also depicted in Fig. 6, based on study [18]. After multiple tests, the fastest convergence times, shortest path lengths, and average path lengths of the ants for the three algorithms are shown in Table II.

TABLE II. Optimal Solutions of each Algorithm

Algorithm type	Maximum convergence times/times	Minimum path length /mm	Average ant path length /mm
T-ACO	80	1307.34	1915.97
I-ACO	49	1300.15	1879.26
TF-ACO	38	1250.66	1672.95

From the figures and tables above, it can be observed that compared to the traditional and improved ant colony algorithms, the thermal-field ant colony algorithm reduces the fastest convergence time by approximately 52.5% and 22.4%, respectively. The shortest path length is improved by approximately 4.5% and 3.8%, and the optimal average path length is improved by approximately 14.5% and 12.3%. This demonstrates that the improved code in this paper not only prevents heat accumulation but also improves the optimal path.

In this simulation experiment, the optimal solutions obtained from the three algorithms were plotted on the DXF part base map after nesting, as shown in Fig. 7. Panel (a) displays the base map with numbered parts, showing 15 distinct geometric elements. After computing the geometric features, a total of 71 feature points are derived as the city coordinates for the path planning algorithms. Based on the feature point sequence in the base map, the path planning cutting order for each algorithm is as follows: Traditional ACO: 46, 51, 55, 49, 62, 68, 3, 7, 10, 14, 22, 18, 33, 37, 43. Improved ACO: 22, 17, 18, 41, 36, 37, 45, 49, 50, 55, 68, 63, 3, 8, 13. Thermal-field ACO: 4, 67, 58, 49, 55, 54, 46, 42, 37, 36, 18, 30, 17, 13, 6.



(a) Base map (b) T-ACO

(c) I-ACO (d) TF-ACO

Fig. 7. Feature point diagram and path comparison diagram.

The optimal path heat matrix is also shown in Fig. 8. From this figure, it can be observed that by introducing the heating and cooling factors, most of the city heat values did not exceed the initially set thermal threshold. While the ants selected the optimal path, they also effectively controlled the heat accumulation, avoiding the selection of high-heat cities as the cutting start points during perforation. This mechanism ensures that the material does not overheat during cutting, which could otherwise negatively affect the cutting quality.

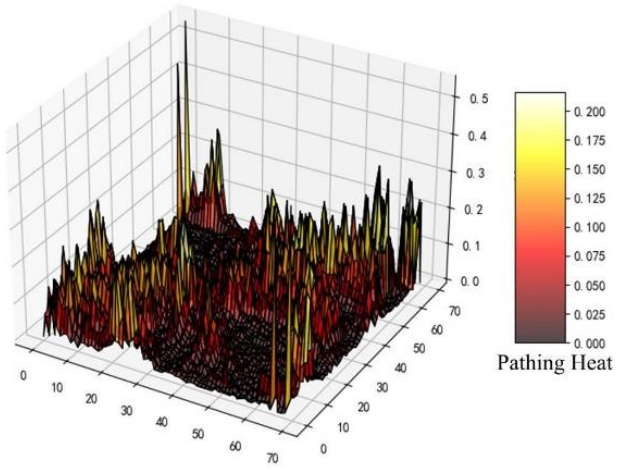


Fig. 8. Heat distribution diagram of optimal path.

B. Cutting Experiment

In this cutting experiment, a flatbed laser cutting machine was used for comparison. The DXF file from Fig. 7 served as the cutting object for the experiment. The laser cutting machine model is FLM3015, equipped with a continuous fiber laser of the model EFRC-3000-E, with a rated output power minimum of 2900W, a maximum of 3100W, and a measured center wavelength of 1080nm. The working voltage is 380VAC, and the output fiber core diameter is 250um. Fig. 9 shows the laser cutting machine used for this experiment.

Fig. 9. Experimental laser cutting machine.

Carbon steel with dimensions 500mm × 500mm × 6mm was selected as the cutting material for the experiment. Prior to cutting, a perforation process was applied to compare the effects of different algorithms on the temperature of the material after perforation. The laser cutting process parameters are shown in Table III.

TABLE III. PROCESS PARAMETERS

Parameters	Value	Unit
Cutting speed	100	mm/s
Barometric type	oxygen	
Atmospheric pressure Peak power Punch grade Progressive time	5 1000 first-order 1000	BAR W ms

Based on the process parameters mentioned, the optimal paths obtained from the three path-planning algorithms were imported into the laser cutting system for cutting experiments. During the experiment, an infrared thermometer was used to measure the surface temperature of the material in real-time, recording the maximum temperature reached by each algorithm during the cutting process. The ambient temperature during the experiment was 20°C. The data recorded for planning time, processing time, and average cutting temperature for each algorithm are shown in Table IV.

TABLE IV. CUTTING COMPARISON OF EACH ALGORITHM

Algorithm type	Plan time /s	Cutting time /s	Maximum cutting temperature /℃
T- ACO	4.35	74.49	262.32
I-ACO	3.81	72.13	257.09
TF-ACO	3.70	70.06	209.87

From the table, it can be observed that compared to the traditional and improved ant colony algorithms, the thermal-field ant colony algorithm resulted in approximately 14.9% and 2.8% improvement in planning time, 5.9% and 2.8% improvement in cutting time, and 19.9% and 18.4% improvement in the maximum cutting temperature, respectively. These results indicate that the thermal-field ant

colony algorithm not only effectively avoids heat accumulation but also reduces the temperature increase while maintaining optimal planning and cutting times. This algorithm better controls the heat distribution during the cutting process, minimizing overheating, thus ensuring cutting quality and the integrity of the workpiece while improving path planning and cutting efficiency.

V. CONCLUSIONS

In comparison to the traditional and improved ant colony algorithms, the thermal-field ant colony algorithm presented in this study demonstrates significant advantages in laser cutting path planning. Specifically, the algorithm reduced the number of iterations by 22.4% to 52.5%, the planning computation time by 2.8% to 14.9%, the non-cutting travel path length by 12.3% to 14.5%, the cutting time by 2.8% to 5.9%, and the maximum cutting temperature by 18.4% to 19.9%. These results indicate that the thermal-field ant colony algorithm not only effectively shortens the cutting travel and processing time but also reduces the perforation issues caused by heat accumulation during laser cutting, preventing the occurrence of "burn-through" and lowering the overall cost of the laser cutting process.

However, certain environmental factors such as smoke, dust, and the material surface's emissivity in the experimental setting have introduced some uncertainty in the temperature measurement results. Therefore, future research will focus on further optimizing the path planning algorithm and improving the accuracy of temperature monitoring to better adapt to real-world production environments.

ACKNOWLEDGMENT

Funding: The work described in this article was supported by Shanghai Multi-Direction Die Forging Engineering Technology Research Center (20DZ2253200); Supported by Intelligent Manufacturing Industry College of Shanghai Lingang New Area (B1-0299-21-023). Shanghai Local College Capacity Building Project (22010501000).

Conflicts of Interest: The authors declare that there are no conflicts of interest.

REFERENCES

[1] Luo Dengcheng, Wang Hongjian, Li Yongliang, et al. Potential field-ant colony algorithm planning of laser cutting path [J]. Laser Journal,2023,44(10):14-18.

[2] Chang Cuizhi, Gao Wenliang, Yan Penghui, et al. Annealing algorithm of double-stranded gene for laser cutting path optimization [J]. Journal of Wuhan Polytechnic University,2023,45(03):331-336.

[3] Hajad M ,Tangwarodomnukun V ,Jaturanonda C , et al.Laser cutting path optimization using simulated annealing with an adaptive large neighborhood search[J].The International Journal of Advanced Manufacturing Technology,2019,103(1-4):781-792.

[4] Junior A B ,Carvalho D N G ,Santos C M , et al.Evolutionary Algorithms for Optimization Sequence of Cut in the Laser Cutting Path Problem[J].Applied Sciences,2023,13(18):

[5] Song Lei, Wang Xinxin, Liu Xiaoyan, et al. Two-chromosome Genetic algorithm optimization of Laser Cutting process Path [J]. Forging & Stamping Technology,2021,46(10):119-125.

[6] Wu Yanming, Cao Ning, Li Feiya, et al. Research on Optimization Algorithm of plate porous Processing Path based on Greedy Algorithm [J]. Journal of Hefei University of Technology (Natural Science Edition),2022,45(06):742-745+759.

[7] Zhou Rui, Ma Hanwu. Path planning of laser cutting collaborative work based on improved Genetic algorithm [J]. Logistics Technology,2021,44(10):50-55.

[8] Zhou Zhichao, Cui Quanfa, Yang Yanlong, et al. Analysis of heat affected zone in laser cutting [J]. Physical Measurement and Testing,2019,37(01):14-17.

[9] Yilbas B ,Arif A .Laser cutting of steel and thermal stress development[J].Optics and Laser Technology, the lancet, 2010 (4) : 830-837.

[10] YAN Shu, Li Lijun, Li Juan, et al. Review on surface quality of laser cutting sheet [J]. Laser Technology,2005,(03):270-274.

[11] Teng Jie, Wang Binxiu. Analysis and Solution of Common Problems in Laser Cutting Process [J]. Electromachining & Mold,2009,(04):60-61. Xu Luning, WANG Xiao, ZHANG Yongkang. Process treatment of laser cutting sheet metal [J]. Applied Laser,2002,(06):533-538.

[12] Xue Zhongming, Gu LAN, Zhang Yanhua. Numerical simulation of Temperature field in laser welding [J]. Chinese Journal of Welding,2003,(02):79-82+0.

[13] Zhou Leping, Tang Dawei, Du Xiaoze, et al. High power laser weapon and its cooling system [J]. Advances in Laser and Optoelectronics,2007,(08):34-38.

[14] Duan Haibin, Wang Daobo, Zhu Jiaqiang, et al. Progress in theory and application of ant colony algorithm [J]. Control and Decision,2004,(12):1321-1326+1340.

[15] Zhu Qing-Bao, Zhang Yulan. Robot path Planning Ant Colony Algorithm based on raster Method [J]. Robot,2005,(02):132-136.

[16] Li Jun, Tong Zhao, Wang Zheng. A parallel ACS-2-opt algorithm for solving TSP problems [J]. Computer Science,2018,45(S2):138-142.

[17] Zhan Shichang, Xu Jie, Wu Jun. Ant colony algorithm in the algorithm of optimal parameter selection [J]. Science, 2003, (5) : 381-386. The DOI: 10.13774 / j.carol carroll nki KJTB. 2003.05.008.

[18] Hou Puliang, Liu Jianqun, Gao Weiqiang. Research on Path Optimization of Laser Cutting based on Improved Ant Colony Algorithm [J]. Mechanical and Electrical Engineering,2019,36(06):653-657.

Laser Distance Measuring and Image Calibration for Robot Walking Using Mean Shift Algorithm

Rujipan Kosarat[1], Anan Wongjan[*2]
Department of Software Engineering-Faculty of Engineering, Rajamangala University of Technology Lanna, Chiang Mai, Thailand[1]
Department of Electronics Engineering and Automatic Control Systems-Faculty of Engineering, Rajamangala University of Technology Lanna, Chiang Mai, Thailand[2]

***Abstract*—In this research, we have measured the physical distance between the robot and its surroundings using a laser distance measuring device that we have developed, designed controllers for, and tested operationally. We will record the distance using the USB camera and integrate the LDMSB board into the laser distance measuring design. We will fasten these two parts to the robot's underside. Developing the experiment in LabVIEW is the next step. The mean shift method enables us to move the robot's position by relocating a laser-based distance measurement device and capturing a photo at that location. In order to record that area, we will perform a perspective camera calibration. This will allow us to set up or adjust the camera system's value, or provide visual assistance to ensure that the viewing angle is precisely aligned with the intended view angle. The laser measurement results ranged from one to fifteen meters. A device that makes use of lasers has 99.25% accuracy. Every calibration location throughout the 10 has a precision rating of 94.03%.**

Keywords*—Laser distance; image calibration; mean shift algorithm; LabVIEW

I. INTRODUCTION

The mobile robot education process at Rajamangala University of Technology Lanna's Faculty of Engineering, Electronic Engineering, and Automation Control Systems Program, Chiang Mai Province, includes measuring a robot's walking distance. This field holds great promise for the development of precise and long-range robots. However, a variety of issues, including measurement error, sensor data rectification, various settings, and adaptability, make determining a robot's walking distance difficult.

There is more to using a laser to measure distance than simply speed and ease. Nonetheless, the technique is very accurate and widespread. Consequently, it might be a useful tool in many facets of everyday life. This covers a wide range of industries, including commercial and medical applications, engineering, and building construction. The advantage of lasers is 1) High degree of accuracy: The laser exhibits a high degree of accuracy in measuring distance and is capable of precisely transmitting laser signals to the designated measurement location. Laser sensors provide precise data in a variety of situations. This makes it appropriate for uses where a high degree of precision is required. The laser can quickly calculate the separation between the two locations. The laser's quick light transmission allows for instant viewing of the experiment findings. 2) This makes it an ideal choice for tasks that require quick thinking, such as data storage and production management. 3) It has the ability to adapt to challenging situations. Lasers are used in harsh environments like cold and dusty ones because they can withstand harsh conditions. 4) Versatility in Application Lasers possess a wide range of potential applications. It has a wide range of potential uses, including in commerce and research, as well as in the field of measuring distances in medicine.

To conduct this study investigation, we set up a laser distance measuring and imaging device. Writing and testing control software in LabVIEW is a crucial step in tracking the robot's location and determining the distance to the objective. One of the main purposes of lasers in robots is to detect distance accurately [1]. This enables robots to carry out their work with the same diligence and precision as people. Robots can gain a better understanding of their environment and adjust their behavior by employing laser distance measurement to gather information about it and adjust their behavior. The use of lasers in robotics creates new opportunities for the creation of innovative and useful robots, suitable for various human endeavors such as navigation, obstacle avoidance, and precision placement tasks. These robots have applications in all fields of human endeavor. These robots can support industrial applications, conduct surveys and research, or serve both purposes [2].

Most people agree that one of the most powerful tools for software development and engineering is the LabVIEW application [3-4]. A graphical programming environment called LabVIEW uses representations of the signal type [5]. LabVIEW allows users to create programs by simply dragging and dropping components into block diagrams. This simplifies operations and creates an easy-to-use interface. Furthermore, because it can interface with a wide range of devices, LabVIEW may be used in software applications that require automation in the domains of measuring, controlling, and testing. There are several uses for LabVIEW software, some of which include industrial, scientific, and engineering research and development. Because of its intuitive design, LabVIEW is a tool that allows users to create and modify programs to any degree of customization without facing any limitations. This study uses the board to compute the robot's distance. Its hardware is the Laser Distance Measuring Signal Board (LDMSB). We use the mean shift approach to track objects [6].

*Corresponding Author

The difficulties in object tracking include many types of occlusions, including occlusion by background objects, other target objects, or self-occlusion (produced by components of the object itself). The tracking procedure becomes more difficult due to these partial or complete occlusions. Significant difficulties are also presented by the target object's changing appearance, particularly in surveillance applications. Inconsistent illumination over wide regions or rotations of the object along axes other than the imaging system's optical axis often cause these alterations. Improving the accuracy and resilience of object tracking systems requires addressing these problems.

This research report explains the camera calibration process and the creation of a laser distance measurement device. For this, we wrote driver software and tested it with LabVIEW. This crucial step enables us to calibrate the camera before recording the robot's location using the Mean Shift method, a crucial tool for precise and thorough distance measurement. Laser distance measurement may benefit robots that can detect and adjust their behavior in different situations. We have organized this job to involve traversing a space, avoiding obstacles, or performing tasks that require precise placement. Section III describes the study technique, while Section II reviews some relevant prior research. Section V brings the article to a close. Section IV presents and examines the findings.

II. Literature Review

The various methods and designs discussed in these books simplify the calculation of the distance between two lasers. A low-tech laser sensor using triangulation achieves outstanding spatial resolution [7]. Another approach suggests using a heterodyne interferometer, renowned for its high accuracy, for absolute distance measurement. A new method for laser distance measurement employs least squares and triangular-wave amplitude modulation to enhance the signal-to-noise ratio [8]. Additionally, a unique distance-measuring device utilizing microfabricated scanning micromirrors demonstrates various configurations for distance estimation [9]. Overall, these methods and technologies offer a wide range of laser distance measurement techniques that can be adapted to meet different application needs and accuracy standards [10].

These studies provide new perspectives on robotic walking using laser distance measuring. One study addresses the calibration of laser range finders for legged robots to achieve the accuracy needed for terrain mapping and foothold selection [11]. Another explores integrating inertial measurement units (IMUs) with laser scanners for real-time, safe calculation of minimal distances between humans and robots. Research also highlights the utility of 3D laser distance measurements for accurate pedestrian tracking, particularly in unstructured environments with occlusions and sensor noise [12]. Additionally, a mobile robot equipped with a laser range finder combines a walking motion model with the geometric characteristics of human legs to ensure precise tracking and a better understanding of human gait [13]. Collectively, these studies demonstrate the advantages and potential of laser distance measurement in walking robot applications [14].

The essays in this collection offer valuable insights into the use of LabVIEW for laser distance measuring. One study emphasizes the importance of LabVIEW in evaluating the reliability of femtosecond laser light sources for distance measurements [15]. Another explains a LabVIEW-based software architecture for assessing laser divergence angles [16]. A further method demonstrates the utility of LabVIEW in a laser beam profile scanning interface. Additionally, a high-precision optical distance meter based on a mode-locked femtosecond laser is presented, capable of measuring distances up to 240 meters with a detection accuracy of 0.01 meters or better [17]. Together, these papers highlight the effective and versatile application of LabVIEW in various laser distance measuring components [18].

III. Methodology

The main program of the suggested algorithm is shown in Fig. 1.

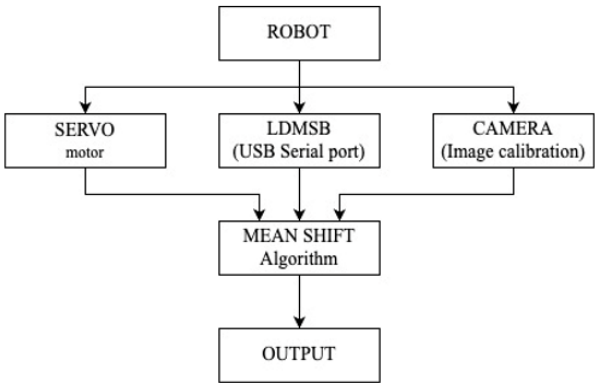


Fig. 1. Distance-measuring walking laser robot block diagram.

Fig. 1 illustrates the process for calculating the robot's travel distance. By following the instructions, one may accomplish the many stages involved. The LDMSB board is situated precisely adjacent to the wheel section at the bottom of the moving robot. The LDMSB board is ready to perform its function as a distance measuring device. The technique then moves on to attaching two USB cameras to the mobile robot. We will proceed to the next step. Establish a serial port protocol connection between the LDMSB board and the USB camera. We test the robot's distance calculation and image calibration skills using LabVIEW's Mean Shift approach. The purpose of these tests is to verify the accuracy of the results.

A. LDMSB

Laser Distance Measuring Signal Board (LDMSB) is a board used in the design to measure distance with laser in this research article. Fig. 2 shows the various components of the LDMSB board. When a lens transmits laser light, it is called a laser transmitting lens. Similarly, a laser receiving lens is a laser receiving lens. Holes or indentations drilled into any material or component to install or secure an object, such as an electronic board or component, so that it may be installed or connected to other apparatuses or buildings, are known as mounting holes. An information processing and perception system or technology is referred to as vision. Pin locations, for instance, are used to link pins in circuits or structures when DC power is needed to power them. Utilizing sensors is necessary for tasks pertaining to various systems' eyesight and perception. A USB

(Universal Serial Bus) system connection establishes the circuit connection between the LDMSB board and the USB interface. This implies that the system will automatically detect USB-powered devices when they are inserted into a port. It is perfect for connecting devices that need full-duplex communication since it can transmit both transmission (TX) and reception (RX).

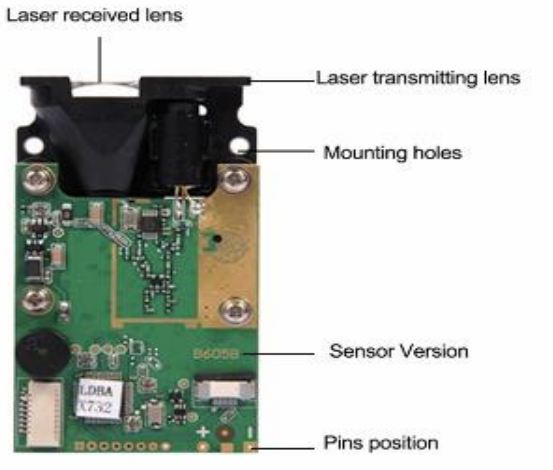


Fig. 2. Components of the LDMSB board.

The goal is to create a device that will be utilized for evaluating the functionality of a laser distance measurement system after identifying the target position in order to measure the Z-axis distance that has to be measured and gathering picture data according to the position. In order to do this, a servo motor that can be adjusted in one-degree increments will be used to rotate the X and Y axes. The desired appearance of the gadget as per its design is seen in Fig. 3.

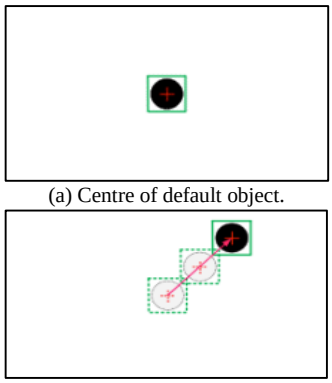
Fig. 3. Laser distance measuring functional testing equipment.

After that, it may be used with mobile robots, where the USB camera and LDMSB board are mounted in the bottom of the robot. The robot can use the laser to measure and record distances thanks to the placement of these parts.

B. Mean Shift Algorithm

The working principle of the object tracking algorithm is based on a methodology akin to template search [19]. The process of matching object photos to templates involves searching for templates that enable things to be located inside a user-defined area or throughout the entire image. This technique, similar to object tracking, looks for an object template in an area that either anticipates or is close to the item's position from the previous frame. This reduces false searches and boosts processing efficiency, making it ideal for labor-intensive activities. The article monitors objects by approximating their position and appearance using the mean shift [20]. The kernel method estimates the mode of the probability function to assess the condition of an item. The rival object's position is estimated using target object motion data from the previous frame. Comparison of the appearance characteristic model of an object with the previous frames to increase the likelihood of its location Assign a new competitive position to the outcome. Increase the frequency of the position estimate until either the user-defined repetition criterion is met or the competing position converges to the final value.

Next, the model will be updated for the following frame when the target object's location has been calculated. After several iterations, the target item's position is estimated using the initial frame. The middle of the rectangle determines the round position of the competing item. In Fig. 4, the arrows in (a) display the mean shift vector, which indicates the amount that the center of mass of the target object changed from the prior frame. The center of the moving item as it gets closer to the target is the starting object (b).



(a) Centre of default object.

(b) Point of movement toward objective.

Fig. 4. Mean shift object tracking.

During tracking, the target object contour model that tracks mean shift fixes the target object forms. The image's target object's center of mass is shown by the telematic model's parameters [21]. The center of mass's pixel location is all that is needed to monitor mean-shift objects in the model. When the subject in the next frame travels slowly, this works. There are two different types of objects tracking algorithms that employ the Mean Shift technique: target object tracking and target object modelling. These two categories represent various aspects of the application.

The mean shift update equation for tracking an object at position x can be formulated as follows:

- Initial Position of Object: Start with an initial position $^{(0)}$ of the object in the current frame.
- Kernel Density Estimation: The pixel weights are assigned using a kernel function K, usually according to the pixels' distance from the center. Although other kernels, such as Gaussian, may also be employed, the Epanechnikov kernel is a popular option. The weighted mean position is determined with the use of the kernel K(x), which concentrates on pixels nearer the center.
- Mean Shift Vector: For each iteration t, vector m(x(t)) is computed to shift towards the mean of the region with the highest density from Eq. (1).

$$m(x^{(t)}) = \frac{\sum_{i=1}^{n} x_i K(x_i - x^{(t)})}{\sum_{i=1}^{n} K(x_i - x^{(t)})} \quad (1)$$

Where x_i represents the positions of pixels within a region around the current location and $K(x_i - x^{(t)})$ gives the weight based on the distance from $x^{(t)}$

Position Update: The object's new position is updated by shifting to x(t+1) from Eq. (2).

$$x^{(t+1)} = x^{(t)} + m(x^{(t)}) \quad (2)$$

Convergence: Repeat the iteration until the mean shift vector $m(x^{(t)})$ is sufficiently small (below a threshold), indicating that the peak (mode) has been found.

As illustrated in Fig. 5, it regulates the blending parameter and the maximum percentage of rotational size and shape changes.

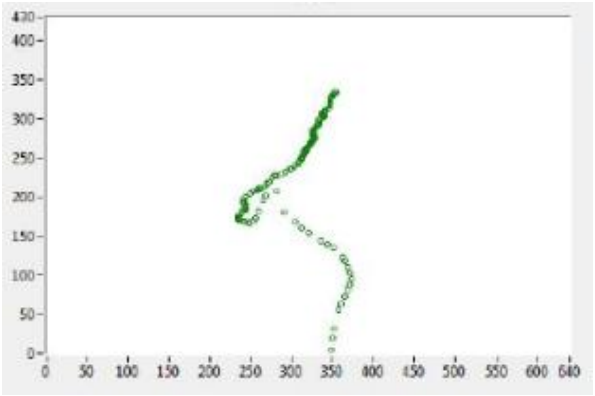


Fig. 5. Adjusts mixing settings and maximum rotation size and shape.

C. LabVIEW

One object tracking technique that is accessible in LabVIEW 2019 is the Mean Shift algorithm, which utilises both the NI Vision Assistant and the NI Vision Library feature. This approach is perfect for monitoring a single target object and is made to work with LabVIEW 2019 [22-23].

This technique replicates the distance of a laser measurement using LabVIEW programming by employing the mean shift method to move an object based on the mouse's position. Fig. 6 displays the accessible front-panel interface of the LabVIEW program. The first stage in improving the accuracy of devices or systems, such as offset, sensitivity, or scale factor, is to calibrate the x- and y-axes. As a result, you can trust that the system will precisely determine an object's position or movement. The next step is to employ a method to figure out how big the object (PEN) is in order to depict the laser point and the object's movement. The final step of the procedure entails analyzing the program to determine if the coordinates of x and y (red objects) on the Axis Mouse are following the mean shift approach as they move along the mouse frame. For the application, LabVIEW has produced a block diagram (graphic programming). We will be able to move more effectively with the help of a servo motor if we move at an angle along the x and y axes. We will correctly apply the mean shift algorithm.

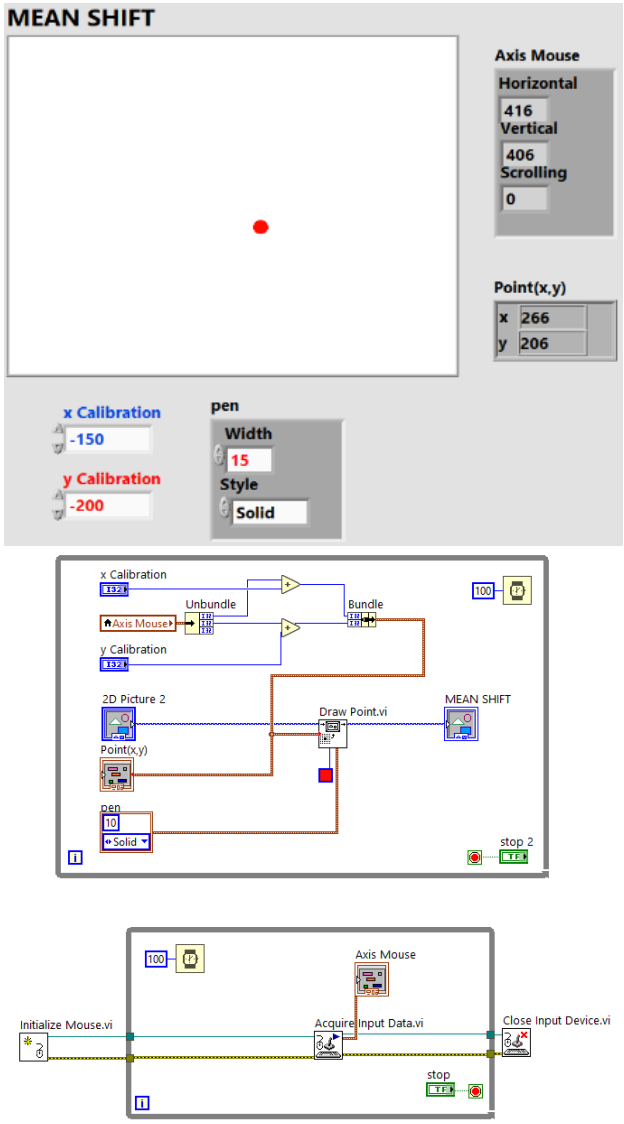


Fig. 6. LabVIEW front panel and block diagram via mean shift algorithm.

We have developed a laser distance measurement design and captured pictures of the distances for the robot's use during its travels. To shoot a picture, one must perform a procedure known as perspective calibration. Adjusting or changing the camera system or imaging equipment settings is necessary to ensure that the picture's angle is true to the desired perspective or the proper theoretical point of view [24], as Fig. 8 illustrates.

The calibration perspective achieves several objectives. 1) Assist in ensuring the accuracy of the camera system or picture equipment according to visual theory and metrics; 2) guarantee that the results are theoretically sound and consistent with the visual model. It also helps to minimize deflection errors, reflection, and visual distortion. 3) Measure or closely inspect photos. Gathering both intrinsic and extrinsic camera system characteristics is necessary to ensure that the produced picture is accurate and does not tilt or distort due to an incorrect angle, and also to assist in optimizing the camera or imaging equipment. Then, using LabVIEW, we created a front-panel user interface for capturing photos and measuring distances, as shown in Fig. 7.

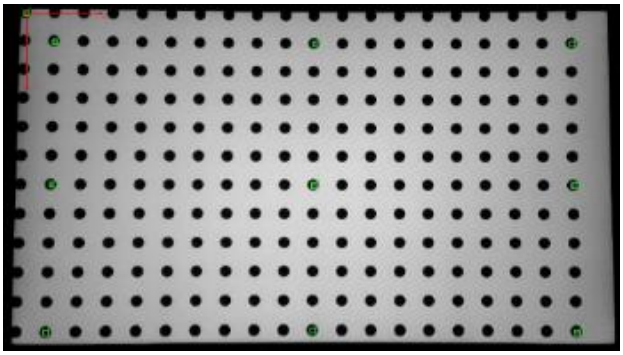


Fig. 7. Perspective 10-point calibration.

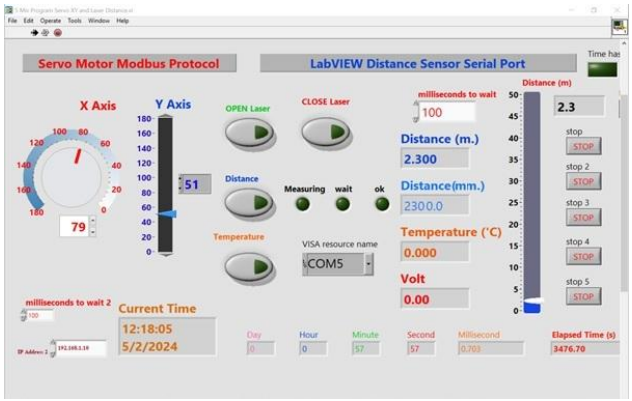


Fig. 8. Images and laser distances LabVIEW front panel.

The user interface seen in Fig. 8 was created through programming. The robot design attaches the LDMSB board and a USB camera to its bottom, enabling the software to perform its planned tasks. The software allows you to turn the laser on and off at will. The software allows you to measure distances in millimeters and meters, as well as save photos. You can monitor the temperature and voltage of the laser.

This method locates a controlled circular point in motion with a radius of 20 millimeters, using a source image at the X and Y axes. The experiment used an LDMSB board with a laser to measure distance. Additionally, we connected a USB camera to the distance measurement board to capture pictures. We then projected the picture onto a 100-inch display for the exercise. We must create an angle to shift the position from point 6 (P6) to point 7 (P7). Fig. 8 illustrates the process of mean shift image tracking and laser distance measurement.

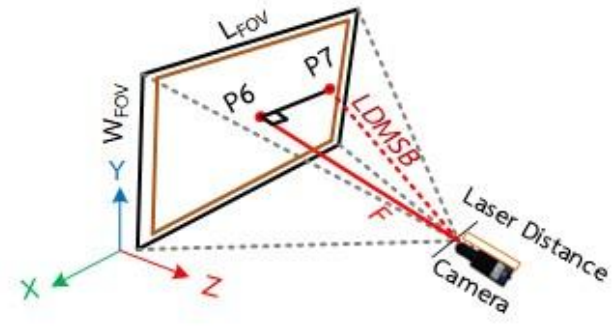


Fig. 9. The laser uses a mean shift from position 6 (P6) to point 7 (P7) to track pictures and calculate distance.

In Fig. 9, this case involves the transfer of the post. In order to calculate the distance caused by P6 to P7, use the formula $c = \sqrt{a^2 + b^2}$, where a is the distance between the camera and laser device and the monitor, or point P6, b is the distance between the points P6 and P7 along the X axis, and c is the distance between the camera and laser device and the monitor, or point P7, that results from comparing the calculated value with the actual value that was measured by the laser. When we speak to the camera's field of vision (*FOV*), we mean the range of pictures it can record. We refer to the scope the camera records as *FOV*, the scope it records by width as W_{LOV}, and the scope it records by length as L_{FOV}. The formula $FOV = W_{LOV} \, x \, L_{FOV}$ may be used to get the value, where *FOV* stands for the scope that the camera records.

IV. Results and Discussion

We separate the experimental findings from the laser odometer and the picture recording from the mean shift algorithm into three distinct design experiment components, each of which consists of the specific elements listed below:

A. Laser Distance Testing

We must measure the separation between two spots using our own laser technology. We designed these stages to evaluate the accuracy of the laser apparatus. The laser distance determines the example's result, as Table I illustrates.

The data presented shows the measurements of distances in meters, with five repetitions for each distance. The average values for each distance are very close to the expected values, indicating a high level of accuracy in the measurements. The percentage error remains minimal across all distances, with the largest error observed at the 0.1-meter distance, which has a 6.67% error. As the distance increases, the percentage error decreases, reaching 0% for several measurements, including 1 meter, 2 meters, 3 meters, and 10 meters, among others. This suggests that the measurement process is highly reliable, particularly at greater distances, where the error becomes negligible. Overall, the data indicates that the measurement system or method used is highly precise and consistent across a wide range of distances. Table I displays the results of five separate experiments conducted at distances ranging from 10 cm (0.10 m) to 50 m, along with the average of each test. 99.25% accuracy is the average for measuring distance.

TABLE I. LASER DISTANCE TESTING

Distance (meters)	#1	#2	#3	#4	#5	%Error
0.1	0.10	0.10	0.10	0.12	0.12	6.67
0.2	0.20	0.21	0.21	0.21	0.21	0.33
0.3	0.30	0.30	0.30	0.30	0.30	0.11
0.4	0.40	0.40	0.40	0.40	0.40	0.00
0.5	0.50	0.51	0.51	0.51	0.50	1.00
1	1.00	1.00	1.00	1.00	1.00	0.00
2	2.00	2.00	2.00	2.00	2.00	0.00
3	3.00	3.00	3.00	3.00	3.00	0.00
4	4.00	4.00	4.00	4.00	4.00	0.00
5	5.01	5.00	5.02	5.00	5.00	0.10
10	10.00	10.00	10.00	10.00	10.00	0.00
20	20.00	20.00	20.00	20.01	20.00	0.01
30	30.00	30.00	30.00	30.00	30.00	0.00
40	40.00	40.00	40.03	40.01	40.01	0.02
50	50.00	50.00	50.00	50.00	50.00	0.00

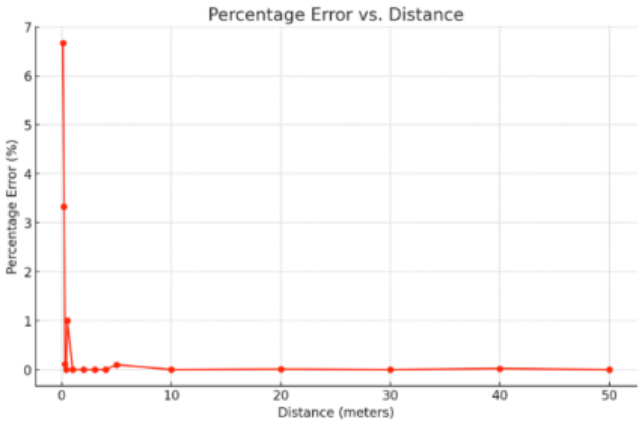


Fig. 10. Percentage error decreases with distance.

In Fig. 10, According to the line chart, the percentage error increases initially at shorter distances but then levels out and stabilizes as the distance grows. This implies that when one moves farther away from the subject, especially beyond 0.4 meters, where the error is almost nil, the measurements get more precise.

B. Image Calibration

While the robot is moving, it is a good idea to take images and measure the distances. The following list displays the results of the 10-point perspective (CP) calibration. The camera or other photographic equipment must have ten calibration points to ensure that the final image is accurate and suitable for the intended angle of view. You can see these calibration points in Fig. 11.

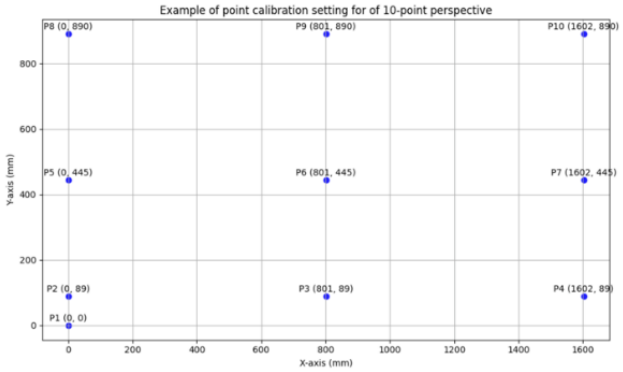


Fig. 11. The 10-point perspective's calibration results.

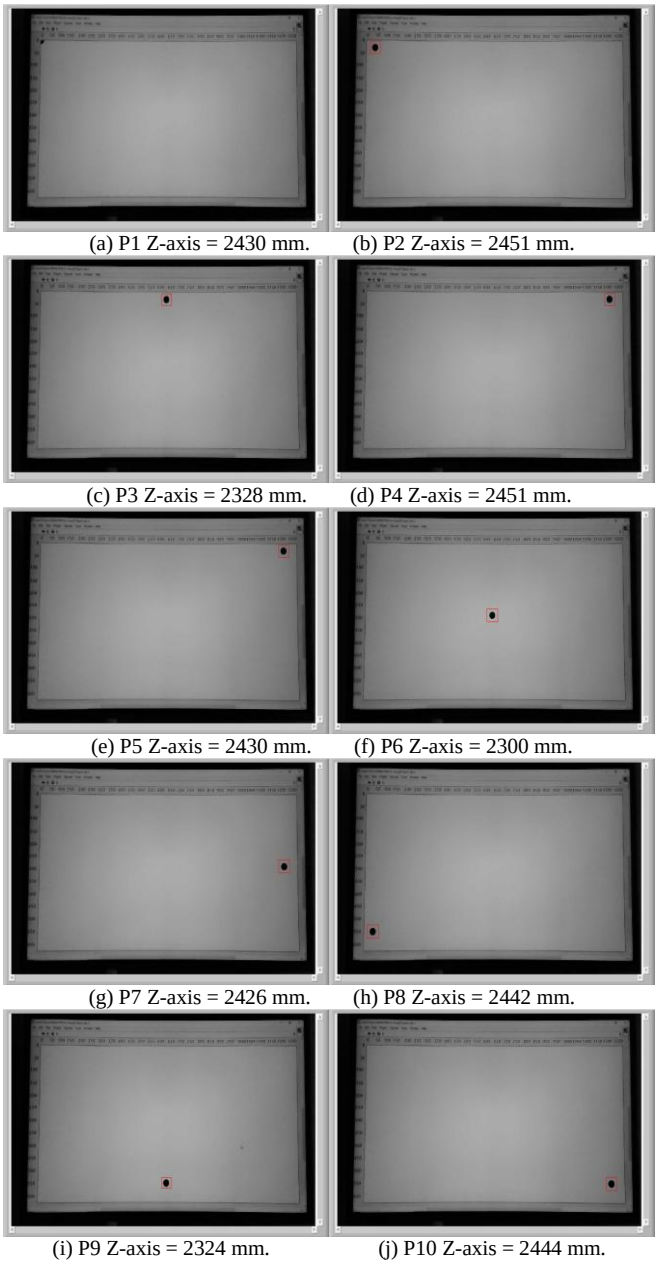
We examined the calibration using a USB camera with a pixel resolution of 1280x720. An LED screen measuring 100 inches in size displays the image. Ten points, each with a value on both the X and Y axes, make up the calibration.

TABLE II. RESULTS OF THE 10-POINT PERSPECTIVE CALIBRATION

CP (10 POINT)	X-axis (mm.)	Y-axis (mm.)	Z-axis (mm.)
P1	0.00	0.00	2,430.00
P2	0.15	84.93	2,451.00
P3	800.85	89.16	2,328.00
P4	1,599.78	85.78	2,451.00
P5	1.87	443.34	2,430.00
P6	798.93	443.89	2,300.00
P7	1,598.77	442.54	2,426.00
P8	0.98	800.56	2,442.00
P9	799.53	797.4	2,324.00
P10	1,601.58	800.82	2,444.00

Table II demonstrates that both the distance measurement value from the laser and the number of photos taken there are relevant. Z represents the value that the laser determined. The X and Y axes represent the point's value. We will conduct the experiment in 10 distinct locations for this study.

After measuring the distance, we used the mean shift technique to track the locations of all 10 dots in the picture. At last, this picture was produced. We calculate the distance between the point of the laser measuring and recording device and the projected screen. A distance of 2300 mm separates the two locations (P6). The Z-axis measurements provided for points P1 through P10 reveal varying levels of displacement, with values ranging from 2300 mm to 2451 mm. Notably, points P2 and P4 both reach the highest Z-axis value of 2451 mm, while point P6 records the lowest at 2300 mm. The data suggests that while some points, such as P1, P5, and P7, remain close to 2430 mm, others like P3, P9, and P6 show significant deviations. This variation in Z-axis measurements illustrates the spatial differences captured during the tracking process, as depicted in Fig. 12, which shows the outcomes of the tracking mean shift algorithm's picture recording.



(a) P1 Z-axis = 2430 mm. (b) P2 Z-axis = 2451 mm.
(c) P3 Z-axis = 2328 mm. (d) P4 Z-axis = 2451 mm.
(e) P5 Z-axis = 2430 mm. (f) P6 Z-axis = 2300 mm.
(g) P7 Z-axis = 2426 mm. (h) P8 Z-axis = 2442 mm.
(i) P9 Z-axis = 2324 mm. (j) P10 Z-axis = 2444 mm.

Fig. 12. Example of the tracking mean shift algorithm's picture recording outcomes, such as P6 Z-axis = 2300 mm.

TABLE III. COMPARING LASER MEASUREMENTS WITH CALCULATED VALUES

CP	a	b	$c = \sqrt{a^2 + b^2}$	%Error
P1	0	2,430	2,430.00	0 .00
P2	0	2,451	2,451.00	0.00
P3	801	2,328	2,461.95	5.44
P4	1602	2,451	2,928.11	16.29
P5	0	2,430	2,430.00	0.00
P6	801	2,300	2,435.49	5.56
P7	1602	2,426	2,907.21	16.55
P8	0	2,442	2,442.00	0.00
P9	801	2,324	2,458.17	5.46
P10	1602	2,444	2,922.25	16.37

A comparison between the values produced by the computations and the values measured by the laser instrument is shown in Table III. The comparison's outcomes are shown. The average accuracy percentage of the numbers derived from the calculations was 94.03%.

The data presented highlights a series of measurements (b) associated with various CP points, alongside calculated values (c) and the corresponding %Error. A few notable observations emerge from the analysis, as shown in Figure 13.

- Consistency in Measurements: For several data points, such as P1, P2, P5, and P8, the values of b and c are identical, resulting in zero error. This suggests that these measurements are precise and match the expected values perfectly.
- Significant Deviations: Other points, particularly P3, P4, P7, and P10, show notable deviations between the b values and the calculated c values. The error percentages at these points range from 5.44% to 16.55%, indicating considerable discrepancies. This variation implies the possibility of underlying measurement process issues or the presence of conditions or anomalies not adequately represented by the expected values.
- Trend Analysis: The plot of b values against CP points reveals that while most values are relatively stable, certain points exhibit substantial variation. For instance, P4 and P7 show the largest errors, with discrepancies of 16.29% and 16.55%, respectively. We may attribute these higher error rates to measurement inaccuracies, external factors, or inherent variability in the system under study.
- Implications and Recommendations: The presence of large errors in specific measurements warrants further investigation. It would be beneficial to review the data collection methodology and consider any external influences that could impact accuracy. Additionally, analyzing whether these errors are systematic or random could provide insights into improving measurement precision. Understanding and addressing these discrepancies will enhance the reliability of the results and ensure more accurate conclusions in future analyses.

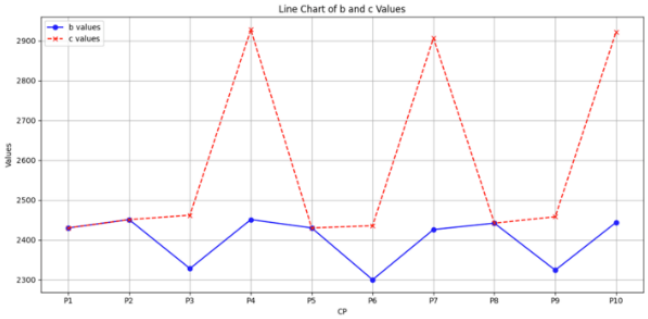


Fig. 13. A comparison between calculated values and laser measurements.

V. CONCLUSION

We use the LDMSB board to build the laser distance measuring system for the experiment. We program the device

using LabVIEW and record the distance using the USB camera. Before moving on to the next phase, we use an algorithm to modify the distance, measure it with a laser, and then take a picture there. To make sure the picture is exactly at the specified distance, we calibrate the camera using ten different perspective points. The findings are suitable for use with mobile robots due to their high accuracy in both measurement and recording.

In order to verify the laser distance measurement against theoretical estimations, we carried out ten distinct calibration and distance measuring experiments. The results indicate that over 15% of the points were incorrect. P4, P7, and P10 are located at the right angle of the projected image. The device's perpendicular base warps when the servo motor vibrates, projecting the picture from the initial point and rotating it to succeeding places. The device's use of a coordinating system is mostly the responsibility of the servo motor.

The advancetage of Mean Shift method is effective for tracking deformable objects that change shape, size, or appearance, including rotations on non-optical axes or articulated motions. However, a limitation arises when using a servo motor for camera rotation, as insufficient motor speed can cause the target to slip out of the tracking frame, compromising the process.

Future work in the development of military robots should enhance laser-based targeting precision, adaptability, and efficiency. Advances in sensor fusion, low-power lasers, and energy-efficient designs could improve target detection and extend operational duration. Refining autonomous decision-making to address ethical concerns and developing collaborative robot networks for real-time coordination can significantly improve battlefield strategies.

ACKNOWLEDGEMENT

The authors would like to thank everyone who helped with this research. We sincerely thank Rajamangala University of Technology Lanna for providing us with the essential information we needed. We also want to express our gratitude to our advisors and colleagues for their guidance and support throughout the process. Having the advice and opinions was really valuable.

REFERENCES

[1] D. Zhang, J. Cao, G. Dobie, and C. MacLeod. “A Framework of Using Customized LIDAR to Localize Robot for Nuclear Reactor Inspections,”. IEEE Sensors Journal, vol. 22(6), pp. 5352–5359, 2022.

[2] M. B. Alatise and G. P. Hancke. “A Review on Challenges of Autonomous Mobile Robot and Sensor Fusion Methods,” IEEE Access, vol. 8, pp. 39830–39846, 2020.

[3] A. Rumyantsev, T. Krupkina, and V. Losev. “Development of a High-Speed Multi-Target Measurement System-on-Chip,” 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus). St. Petersburg and Moscow, Russia, 2019.

[4] J. Wang, Z. Yan, C. Fu, Z. Ma, and J. Liu. “Near-Field Precision Measurement System of High-Density Integrated Module,” IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–9, 2021.

[5] N. Berezowski and M. Haid. “Graphical Programming Languages for Functional Safety using the example of LabVIEW,” 2020 IEEE International Conference on Sustainable Engineering and Creative Computing (ICSECC), Cikarang, Indonesia, 2020.

[6] W. Deng and R. Wu. “Real-Time Driver-Drowsiness Detection System Using Facial Features,” IEEE Access, vol. 7, pp. 118727–118738, 2019.

[7] O. Toedter and A. W. Koch. “A simple laser-based distance measuring device,” Measurement, vol. 20(2), pp. 121–128, 1997.

[8] M. Norgia, G. Giuliani, and S. Donati. “Absolute Distance Measurement with Improved Accuracy Using Laser Diode Self-Mixing Interferometry in a Closed Loop,” IEEE Transactions on Instrumentation and Measurement, vol. 56(5), pp. 1894–1900, 2007.

[9] Q. Fu, Z. Zhou, Y. Luo, and S. Liu. “Laser distance measurement by triangular-wave amplitude modulation based on the least squares,” Infrared Physics Technology, vol. 104, pp. 103-146, 2020.

[10] K. Kim, J. Hwang, and C.-H. Ji. “Intensity-based laser distance measurement system using 2D electromagnetic scanning micromirror,” Micro and Nano Systems Letters, vol. 6(11), 2018.

[11] E. Krotkov. “Laser rangefinder calibration for a walking robot,” IEEE International Conference on Robotics and Automation. Sacramento, Califonia, USA, 1991.

[12] M. Safeea and P. Neto. “Minimum distance calculation using laser scanner and IMUs for safe human-robot interaction,” Robotics and Computer-Integrated Manufacturing, vol. 58, pp. 33–42, 2019.

[13] M. Haselich, B. Jobgen, N. Wojke, J. Hedrich, and D. Paulus. “Confidence-based pedestrian tracking in unstructured environments using 3D laser distance measurements,” 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, Illinois, USA, 2014.

[14] J. Lee, T. Tsubouchi, K. Yamamoto, and S. Egawa. “People Tracking Using a Robot in Motion with Laser Range Finder,” 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 2006.

[15] Y.L. Chen et al. “Laser autocollimation based on an optical frequency comb for absolute angular position measurement,” Precision Engineering, vol. 54, pp. 284–293, 2018.

[16] Jia. “Design of laser divergence angle test software based on LabVIEW,” 2011 2nd International Conference on Control, Instrumentation and Automation (ICCIA), Bandung, Indonesia, 2011.

[17] A. K. Al-Jumaily, V. J. Jumaah, and H. T. Assafli. “Efficient Labview Interface Technique for Laser Beam Profile Scanner,” 2020 1st Information Technology To Enhance e-learning and Other Application (IT-ELA), Baghdad, Iraq, 2020.

[18] K. Minoshima and H. Matsumoto. “High-accuracy measurement of 240-m distance in an optical tunnel by use of a compact femtosecond laser,” Applied Optics, vol. 39(30), pp. 5512, 2000.

[19] D. Comaniciu, V. Ramesh, and P. Meer. “Real-time tracking of non-rigid objects using mean shift,” Proceedings IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 142-149, 2000.

[20] B. Rezaei, X. Huang, J. R. Yee, and S. Ostadabbas. “Long-term non-contact tracking of caged rodents,” 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, USA, 2017.

[21] Chia, Y. S., Kow, W. Y., Khong, W. L., Kiring, A., & Teo, K. T. K. “Kernel-based object tracking via particle filter and mean shift algorithm,” 2011 11th International Conference on Hybrid Intelligent Systems (HIS), Melaka, Malaysia, 2011.

[22] C.-Y. Cheng, J.-C. Renn, I. Saputra, and C.-E. Shi. “Smart Grasping of a Soft Robotic Gripper Using NI Vision Builder Automated Inspection Based on LabVIEW Program,” International Journal of Mechanical Engineering and Robotics Research, vol. 11(10), pp. 737-744, 2022.

[23] Issa, A., Aqel, M. O., Zakout, B., Daqqa, A. A., Amassi, M., & Naim, N. “5-DOF Robot Manipulator Modelling, Development and Automation using LabVIEW,” Vision Assistant and Arduino. In 2019 International Conference on Promising Electronic Technologies (ICPET), Gaza City, Palestin, 2019.

[24] Karim, S., Tong, G., Li, J., Qadir, A., Farooq, U., & Yu, Y. “Current advances and future perspectives of image fusion: A comprehensive review,” Information Fusion, vol. 90, pp. 185-217, 2023.

Predicting Chronic Obstructive Pulmonary Disease Using ML and DL Approaches and Feature Fusion of X-Ray Image and Patient History

Fatema Kabir[1], Nahida Akter[2], Md. Kamrul Hasan[3], Dr. Md. Tofael Ahmed[4], Mariam Akter[5]
Computer Science and Engineering Department, CCN University of Science and Technology, Cumilla, Bangladesh [1, 2, 3]
Department of Information and Communication Technology, Comilla university, Cumilla, Bangladesh[4]
Department of Computer Science and Engineering, Northern University, Dhaka, Bangladesh[5]

***Abstract*—By 2030, chronic obstructive pulmonary disease (COPD) is expected to become one of the top three causes of death and a leading contributor to illness globally. Chronic Obstructive Pulmonary Disease (COPD) is a debilitating respiratory disease and lung ailment caused by smoking-related airway inflammation, leading to breathing difficulties. Our COPD Healthcare Monitoring System for COPD Early Detection addresses this critical need by leveraging advanced Machine Learning (ML) and Deep Learning (DL) technologies. Unlike previous studies that predominantly rely on image datasets alone, our advanced monitoring system utilizes both image and text datasets, offering a more comprehensive approach. Importantly, we manually curated our dataset, ensuring its uniqueness and reliability, a feature lacking in existing literature. Despite the utilization of popular models like nnUnet, Cx-Net, and V-net by other papers, our model outperformed them, achieving superior accuracy. XGBoost led with an impressive 0.92 score. Additionally, deep learning models such as VGG16, VGG19, and ResNet50 delivered scores ranging from 0.85 to 0.89, showcasing their efficacy in COPD detection. By amalgamating these techniques, our system revolutionizes COPD care, offering real-time patient data analysis for early detection and management. This innovative approach, coupled with our meticulously curated dataset, promises improved patient outcomes and quality of life. Overall, our study represents a significant advancement in COPD research, paving the way for more accurate diagnosis and personalized treatment strategies.**

***Keywords*—Chronic obstructive pulmonary disease; COPD; COPD healthcare; advanced monitoring system; COPD early detection; respiratory disease; machine learning; deep learning**

I. Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a progressive respiratory condition that remains a major global health challenge, particularly due to its high prevalence and mortality rate. Characterized by persistent airflow limitation, COPD typically manifests through symptoms such as chronic cough, dyspnea, and wheezing. According to the World Health Organization (WHO), COPD is currently the third leading cause of death worldwide. Relevant studies have shown that the prevalence of COPD is much higher among lung cancer patients [1]. The disease primarily affects individuals with a history of long-term exposure to harmful pollutants, such as tobacco smoke, occupational dust, and chemical fumes. Despite advances in medical care, the burden of COPD continues to rise, particularly in low- and middle-income countries where access to healthcare is limited. As the global population ages and exposure to risk factors persists, the number of COPD cases is expected to increase, highlighting the urgent need for effective strategies to manage and mitigate this condition, the potential of implementing processing steps to more closely adapt clinical workflow processes has thus far not been explored in detail [2].

Detecting COPD in its early stages presents significant challenges, which complicates effective management and treatment. Pulmonary disease is a respiratory disease that affects the lungs as well as the other respiratory organs [3] One of the primary difficulties lies in the subtle onset of symptoms, which are often mistaken for normal signs of aging or attributed to other respiratory conditions. This leads to delays in seeking medical attention and, consequently, late-stage diagnoses when the disease has already caused irreversible lung damage. Current diagnostic methods, such as spirometry, chest X-rays, and CT scans, while effective, are not always readily accessible or reliable, particularly in resource-limited settings. Moreover, these methods can be invasive and uncomfortable for patients, further deterring early detection efforts. The accuracy of these tests also heavily depends on the quality of administration, with improperly trained personnel leading to misdiagnoses or underdiagnoses. As a result, there is a growing need for non-invasive, highly accurate diagnostic tools that can be widely implemented to improve early detection rates and patient outcomes.

Looking ahead, the future of COPD management hinges on the development of advanced diagnostic and therapeutic technologies that can address current limitations. Artificial intelligence (AI), particularly machine learning (ML) and deep learning (DL), is poised to play a critical role in this evolution. Deep learning technology are applied to computer aided diagnosis to realize the automatic diagnosis of disease that achieved good results [4]. By analyzing large datasets of patient information, these technologies can identify patterns and markers that might be missed by traditional methods, enabling earlier and more accurate detection of COPD. Additionally, the integration of wearable devices and remote monitoring systems could facilitate continuous assessment of lung function, providing real-time data that can be used to personalize treatment plans. Despite the success of DL in pulmonary disease classification using CXRs, a very limited number of studies have explored the potential of DL techniques in COPD

diagnosis using CXRs only [5]. Therefore, a comprehensive approach that combines cutting-edge technology with public health initiatives is essential to curb the impact of COPD in the coming decades. Our primary contribution in this study is:

- Patient-Driven Text Dataset: Questionnaire-Based Data Collection for COPD Analysis.
- Privacy-Preserving Image Dataset: Ensuring Confidentiality in Chest X-Ray Image Collection for COPD Diagnosis.
- Integrating Chest X-Ray Imaging and Patient History Data for creating dataset.
- Using Vgg-16, Vgg-19, ResNET-50 for x-ray images and Logistic Regression, XGB classifier, Random forest classifier for patient history
- Identifying modifiable and non-modifiable risk factors.

The study is organized into several sections. Section II reviews the existing literature on COPD detection. Section III details the materials and methods used in the proposed framework. Section IV presents and discusses the experimental results. Finally, Section VI concludes the study by summarizing the key findings on COPD detection.

II. Related Work

Using anatomical data from 28 structures, training 3D nnUNet models on 89 patients' CT scans, testing shows a 10-point improvement in 15 patients. This method enhances early CT scan identification of enlarged nodes [6]. This study proposes a novel anomaly detection approach for diagnosing COPD. Using self-supervised models to identify abnormalities, it outperforms prior methods by 8.2% and 7.7% on two datasets, offering interpretable anomaly maps and early-stage COPD progression detection [7]. CX-Net, an ensemble learning method for lung segmentation and diagnosis in chest X-rays, utilizes four neural network models. Incorporating SHAP and Grad-CAM for interpretability, it provides visual explanations of critical regions, enhancing AI-driven diagnostic systems' reliability in clinical settings [8]. This review highlights the potential of digital inhaler devices, connected to mobile apps, in managing asthma and COPD. Features like interactivity, gamification, and machine learning can predict and prevent exacerbations, but integration into care pathways is essential for personalized management [9]. Using two datasets with complex physiological signals, a fractional-order dynamics deep learning model achieves a 98.66% accuracy in COPD diagnosis. It shows robust performance across datasets, presenting a promising alternative to traditional spirometry-based methods [10]. Enhancing sparse-view CT image quality for lung cancer detection, a U-Net reduces projection views from 2048 to 64, maintaining image quality and radiologists' confidence. Post-processing with the U-Net improves metrics, suggesting a balance between fewer views and diagnostic efficacy [11]. This study enhances COPD prediction using machine learning on 5807 cases, identifying ten significant variables. Logistic Regression performs best on balanced data, while stacking with SMOTE excels on unbalanced datasets, effectively identifying early COPD risk [12]. COPD-FlowNet, a GAN, generates realistic velocity flow field images. The generator uses CNN layers, and a custom CNN classifier locates obstruction sites. Techniques like BatchNorm and leaky-ReLU activation improve feature extraction, addressing the "covariate shift" problem [13]. This study explores the relationship between COPD and NSCLC using machine learning techniques. Analyzing electronic health records, it develops a predictive model to improve early NSCLC identification in COPD patients, enhancing survival rates through accurate clinical feature screening [1]. The study explores advanced methods for diagnosing COPD, impacting over 15 million Americans annually. Evaluating sociodemographic and genetic data, it identifies risk factors beyond smoking, aiming for comprehensive early detection and prevention strategies [3]. This study explores automated COPD detection using CNNs on chest CT scans. Emphasizing preprocessing steps and accurate labels, it demonstrates improved outcomes, suggesting that careful preprocessing enhances CNN-based COPD detection [2]. The study examines COPD risk factors using sociodemographic and genetic data. Smoking, underweight, parental respiratory history, and low education are major risks. Genome-wide studies reveal novel genetic variants, aiming for comprehensive early detection and prevention [14]. Deep learning algorithms for early COPD detection using chest X-rays are developed in this study. Employing data fusion and model fusion techniques, it evaluates performance across demographic subgroups, suggesting deep learning models as valuable screening tools in resource-poor settings [5]. A two-stage 3D contextual transformer-based U-Net is proposed for accurate airway segmentation in CT images, essential for bronchoscopy planning and COPD assessment. The method outperforms existing approaches, achieving advanced segmentation with increased branch extraction and length coverage [15]. The study investigates airway closure dynamics in conditions like asthma, COPD, and cystic fibrosis. Using the Saramito-HB model, it explores liquid plug formation, showing that elasticity influences closure occurrence and time, enhancing understanding of airway closure in different health conditions [16]. The paper explores Vision Transformer (ViT) models for COPD detection using CT images, addressing privacy through federated learning (FL). The proposed approach outperforms CNN-based FL methods, demonstrating effectiveness on COPD data from multiple medical centers [17]. This study focuses on COPD detection and monitoring through voice analysis. Developing a machine learning-based tool, it highlights features like breathing, coughing, and speech, demonstrating promising results for AI-assisted rapid diagnosis and monitoring of COPD [18]. An apparatus for generating obstructive breathing disorder waveforms is proposed, aiding in understanding diseases like COPD and pulmonary fibrosis. The research creates mechanisms for generating a spectrum of disease severities, assisting in classification and severity identification [19]. MixEHR-S, a Bayesian topic model for EHR, models specialist distribution and infers latent disease topics. It incorporates Bayesian probit regression, outperforming existing methods in predicting diseases like COPD, showcasing potential for accurate disease prediction and personalized patient care [20]. An improved machine learning method for accurate lung lobe segmentation is introduced, enhancing spatial accuracy using tracheobronchial

tree information. The method achieves high performance across diverse diseases, demonstrating robustness and aiding clinicians in defining lung disease distribution [21].

III. METHODOLOGY

Fig. 1 presents the workflow diagram of the methodology applied in this research. The dataset utilized for the experiments is newly compiled, derived from patient records that were manually gathered from multiple hospitals. This study leverages both temporal and spectral features to facilitate the detection of COPD, with exploratory data analysis performed through various informative charts and graphs. The dataset is partitioned, with 75% allocated for training and the remaining 25% for testing. Machine learning models are trained using the training data and subsequently tested on the test data. Hyperparameter tuning is applied in an iterative manner to identify the optimal parameters for each model, enhancing their overall performance. All models are fine-tuned to maximize their accuracy and effectiveness in detecting COPD.

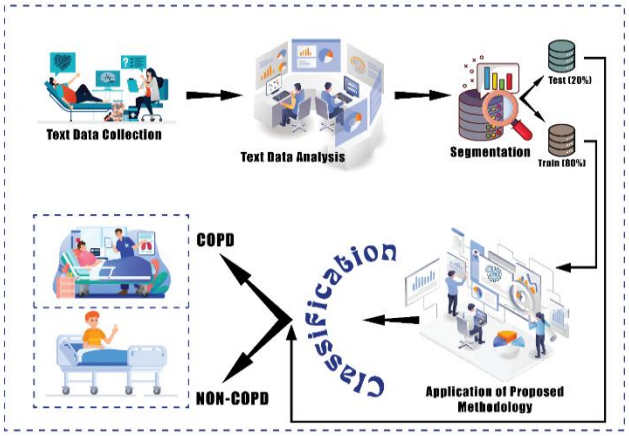


Fig. 1. Text classification.

A. TEXT Data Collection

In Table I discuss about collect the required data for our study, we visited several hospitals and specifically focused on patients who were using oxygen masks or experiencing significant breathlessness, as these symptoms are indicative of potential COPD. We approached these patients in both male and female wards, carefully targeting individuals who matched our study criteria. With their consent, we conducted thorough interviews, asking about various symptoms such as age, Gender, wheezing, breathlessness, smoking history, lack of energy, good day-bad day, allergy, family history.

We also examined their prescriptions and medical reports, documenting key details relevant to their health status. Additionally, we captured images of important test reports, such as CBC-ESR, WBC, and S. creatinine, using our phones for further analysis. We are collected total 1000 data from hospitals patients. Through this method, we ensured that we gathered comprehensive data from both COPD and non-COPD patients, recording all essential information in our study records. This approach enabled us to build a detailed dataset that covers a wide range of patient health indicators, which is critical for our research on early detection and monitoring of COPD.

TABLE I. TEXT DATASET ANALYSIS

Features	Description
Age	The patient's age, which can influence COPD risk and progression.
Gender	Whether the patient is male or female, as COPD prevalence may vary by gender.
Wheezing	Presence or absence of wheezing, a key respiratory symptom linked to COPD.
Breathlessness	The severity of breathlessness, which is a major indicator of lung function decline.
Smoking History	Whether the patient has a history of smoking, a leading cause of COPD.
Lack of Energy	Low energy levels, often reported by patients with chronic lung conditions like COPD.
Good Day/Bad Day	The patient's overall well-being, indicating whether they feel better or worse on a given day.
Allergy	Presence of allergies, which could affect respiratory health and potentially contribute to COPD.
Family History	A record of any family members with COPD or other lung-related diseases, indicating genetic predisposition.
CBC-ESR	Blood test results showing inflammation levels, often elevated in COPD patients.
WBC	WBC levels, which can indicate infection or inflammation common in COPD.
S. Creatinine	A measure of kidney function, sometimes linked to overall health in COPD patients.
COPD or Non-COPD	Classification of patients based on whether they have COPD or not, used for labeling.

B. Text Data Analysis

In this study, text data analysis is crucial for uncovering patterns and extracting meaningful insights from patient information to aid in the early detection of COPD. The dataset comprises numerous health indicators, each offering valuable details about patients' overall condition. Through a detailed preprocessing phase, any missing or inconsistent data is carefully addressed, ensuring a clean and standardized dataset ready for analysis. Feature extraction methods are employed to highlight the most significant variables that play a critical role in diagnosing COPD. Following this, exploratory data analysis (EDA) is conducted using various visualizations, such as graph and correlation matrices, to reveal trends and relationships between the key factors influencing COPD progression. These insights inform the development of machine learning and deep learning models, allowing them to focus on the most impactful features. This approach optimizes the models' performance, ensuring higher accuracy in identifying patients likely to have COPD. By leveraging this data-driven analysis, the study provides a comprehensive framework for enhancing healthcare decisions, contributing to more effective COPD management and improved patient outcomes.

Fig. 2 illustrates the distribution of COPD and non-COPD cases in our dataset. The dataset is evenly balanced, with 50% of the data representing COPD cases and 50% representing

non-COPD cases. Specifically, it consists of 500 samples of COPD data and 500 samples of non-COPD data, ensuring a fair representation of both categories. This balanced distribution is crucial for training machine learning models, as it helps mitigate the risk of bias towards one class and ensures that the model learns to differentiate between COPD and non-COPD conditions effectively. The equal representation of both groups contributes to more reliable and generalized results, allowing for better performance when the model is deployed in real-world clinical settings.

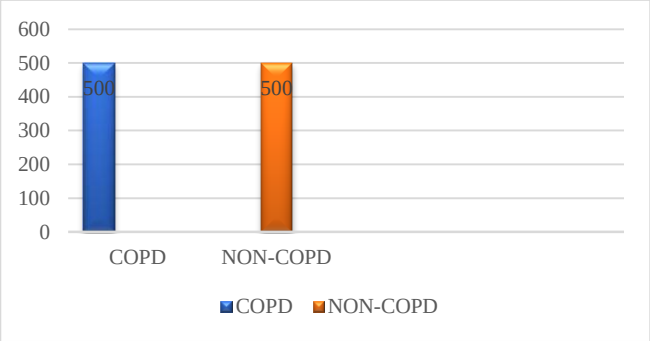


Fig. 2. COPD and NON-COPD distribution in dataset.

Fig. 3. Correlation analysis.

Fig. 3 illustrates the correlation analysis between various patient health factors used in COPD detection. The correlation matrix highlights both positive and negative relationships among the features, with values ranging from -1 to 1. Strong positive correlations are observed between breathlessness and smoking history (0.72), as well as between breathlessness and lack of energy (0.68), suggesting these symptoms are closely linked in COPD patients. Moderate correlations between family history, CBC-ESR, and other key health indicators point to the influence of both genetic and clinical factors on COPD progression. Conversely, features like good day/bad day and allergy exhibit weaker correlations with other variables, indicating less direct involvement in COPD severity. Overall, this analysis reveals the interconnectedness of respiratory symptoms, lifestyle factors, and clinical test results, which provides a deeper understanding of COPD's multifaceted nature and informs the model's predictive capabilities.

In this phase of the study, the dataset is divided into two distinct subsets: features and the target variable. The features encompass all relevant health indicators that may influence the diagnosis of COPD, while the target variable represents whether a patient is classified as having COPD or not. To prepare the data for analysis, the dataset is split into training and testing sets, with 75% allocated for training and 25% reserved for testing. This division ensures that the models are trained on a substantial portion of the data, allowing them to learn the underlying patterns associated with COPD effectively. The training set is crucial for model development, as it provides the necessary data for the machine learning algorithms to identify relationships between the features and the target variable. Conversely, the testing set serves as an independent dataset used to evaluate the model's performance and generalizability to unseen data. The shapes of the training and testing sets are printed to confirm the successful separation of the data, providing an overview of the number of samples available for training and testing. This systematic approach is vital for developing accurate predictive models for COPD diagnosis.

C. Machine Learning Model

1) Logistic regression: In this study, a Logistic Regression model is employed to predict the likelihood of a patient being diagnosed with Chronic Obstructive Pulmonary Disease (COPD) based on various health indicators. The model is instantiated with a regularization parameter C=100, which controls the strength of regularization applied to the model. Regularization helps prevent overfitting by penalizing complex models, thus promoting simpler, more generalizable solutions. The solver used for optimization is 'liblinear', which is suitable for smaller datasets and provides efficient convergence. Logistic Regression is a statistical method used for binary classification. It predicts the probability that a given input belongs to a specific class by modeling the relationship between the independent variables (features) and the dependent variable (target) using the logistic function. The mathematical representation of the Logistic Regression model can be expressed as:

$$y = e^{(b0+b1*x)}/(1 + e^{(b0+b1*x)}) \quad (1)$$

Where,

y = Predicted output,

e = natural logarithm,

b0 = bias or intercept term,

b1 = coefficient for the single input value (x).

The fitted model learns the optimal values of the coefficients through the training data, enabling it to predict the probability of COPD in new patients based on their health profiles. By setting the threshold for classification (commonly at 0.5), the model can classify patients as either having COPD or not, thus contributing valuable insights for healthcare decision-making.

2) XGB classifier: In this study, the XGBClassifier (Extreme Gradient Boosting Classifier) is utilized to enhance

the prediction accuracy for diagnosing Chronic Obstructive Pulmonary Disease (COPD) based on various health indicators. The model is configured with specific hyperparameters: one estimator, a maximum depth of one, a learning rate of 0.3, and a subsample ratio of 0.1. The choice of a low maximum depth helps prevent overfitting while allowing the model to capture important interactions within the data. The learning rate determines how quickly the model learns from the data, while the subsample ratio indicates the fraction of samples to be used for fitting the individual base learners, thus introducing randomness and helping to improve the model's generalization. XGBoost operates on the principle of boosting, which sequentially combines weak learners to create a strong learner. The model optimizes for a specific loss function, with the log loss function often employed for binary classification tasks. The log loss can be expressed mathematically as:

$$\text{logloss} = -1/N \sum_{i=1}^{N} (yi \log(pi) + (1 - yi)\log(1 - pi)) \quad (2)$$

Where:

N is the number of samples.

yi is the true label of sample i (0 or 1).

pi is the predicted probability that sample i belongs to class 1.

This loss function evaluates the performance of the model by penalizing incorrect predictions more severely, especially when the predicted probability is close to 0 or 1, which helps improve the model's accuracy. By fitting the model to the training data, it learns to predict the probability of a patient being diagnosed with COPD, ultimately providing valuable insights for healthcare decision-making.

3) Random Forest classifier: In this study, the RandomForestClassifier is employed to enhance the classification accuracy in diagnosing Chronic Obstructive Pulmonary Disease (COPD). This ensemble learning method constructs multiple decision trees during training and aggregates their predictions to produce a final output, thereby improving robustness and accuracy over a single tree model. The RandomForestClassifier is initialized with one estimator, a maximum depth of one, and a random state of eight to ensure reproducibility.

The power of the Random Forest model lies in its ability to combine the predictions from various decision trees. Each tree is trained on a random subset of the data and makes its own prediction. The final output of the Random Forest model is determined by the mode of the predictions from all the individual decision trees, as expressed by the equation:

$$y = Mode(f1(x), f2(x), \ldots., fn(x)) \quad (3)$$

Where:

y is the predicted output (for classification).

fi(x) is the prediction of the i-th decision tree for the input x.

This aggregation process helps to reduce the variance associated with individual trees, resulting in a more accurate and reliable model. By fitting the Random Forest classifier to the training data, the model learns to recognize patterns associated with COPD, allowing for more effective predictions when applied to new patient data.

In Table II, we provide a comprehensive overview of the hyperparameters utilized for training and testing the machine learning models applied to our text dataset. Each hyperparameter plays a crucial role in determining the model's performance, influencing aspects such as complexity, learning rate, and generalization ability. By fine-tuning these parameters, we aim to optimize the models for accurate predictions of COPD presence in patients. This table serves as a reference for understanding how each hyperparameter contributes to the overall effectiveness of the respective models, facilitating better insights into their operational dynamics during the experimental phase.

TABLE II. HYPERPARAMETERS FOR APPLIED MODEL

Model	Hyperparameters	Description
Logistic Regression	C=100	Controls the inverse of regularization strength (weaker regularization).
	solver='liblinear'	Algorithm used for optimization (useful for small datasets).
	random_state=0	Ensures reproducibility.
XGB Classifier	n_estimators=1	The number of boosting rounds/trees.
	max_depth=1	Maximum depth of each tree, limiting complexity.
	learning_rate=0.3	Controls the weight adjustment speed during training.
	subsample=0.1	Uses 10% of the training data for each tree.
Random Forest Classifier	max_depth=1	Restricts each tree to a single split (decision stumps).
	n_estimators=1	Number of trees in the forest (only one tree).
	random_state=8	Ensures reproducibility.

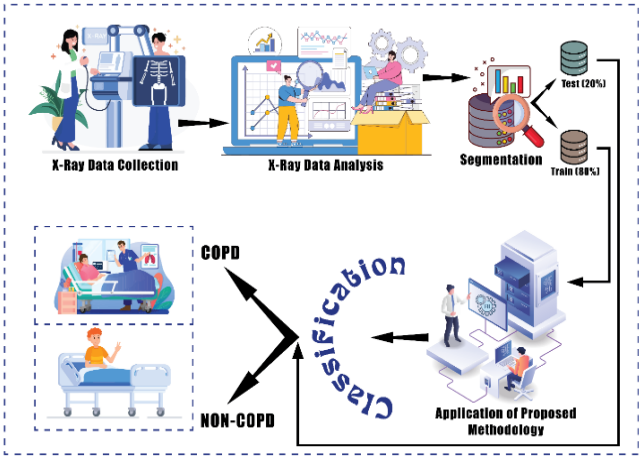


Fig. 4. Image classification.

Fig. 4 outlines the comprehensive workflow for the image classification procedure, beginning with the collection of X-ray images from patients. Once the images are gathered, they undergo image data analysis to extract key visual features. The dataset is then split into training and testing sets during the image segmentation phase, ensuring that the models are trained on a representative portion while preserving data for validation. Following this, the proposed methodology is applied, which incorporates advanced machine learning or deep learning techniques for accurate feature extraction and classification. Finally, the models predict whether the patient is classified as having COPD or being non-COPD based on the image data, providing critical insights for diagnosis.

For the image data collection in our study, we visited several hospitals, focusing on male and female wards to identify patients using oxygen masks or experiencing significant breathlessness, as these indicators are closely associated with respiratory issues. We specifically targeted these patients, engaging them in conversation to inquire about their symptoms and health conditions. Additionally, we examined their X-ray reports to assess their lung health. With the patients' consent, we captured images of their X-ray reports using our phones to ensure we gathered the necessary data accurately.

Through this methodical approach, we successfully collected a total of 1,000 X-ray images from both COPD and non-COPD patients, creating a comprehensive dataset (Fig. 5) for analysis. This collection process not only provides valuable insights into the visual manifestations of COPD but also facilitates further investigation into the relationship between symptoms and X-ray findings in respiratory diseases.

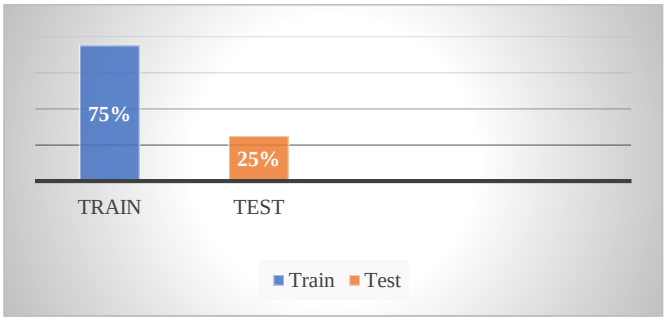


Fig. 5. Image dataset.

4) Image data analysis: In this study, image data analysis is crucial for identifying visual patterns related to COPD detection. The dataset comprises X-ray images from both COPD and non-COPD patients, which are processed for further analysis. Images from both categories are resized to 128x128 pixels and converted to RGB format, ensuring uniformity across the dataset. This preprocessing step converts the images into numerical arrays, making them suitable for deep learning models. Data augmentation techniques are applied using the Image-Data-Generator, which enhances the dataset by introducing variations such as rotation, zoom, width and height shifts, shearing, and horizontal flipping. These augmentations prevent overfitting and improve model generalization by simulating real-world variations in medical imaging. This methodical approach to image data analysis ensures that the dataset is enriched and diversified, enabling more robust training and testing of models for accurate COPD classification.

The image dataset is divided into two parts to ensure effective training and evaluation of machine learning and deep learning models. Seventy-five percent (75%) of the dataset is allocated for training, where the models learn to identify patterns and features related to COPD and non-COPD cases. This larger portion allows the model to gain sufficient exposure to varied data, improving its ability to generalize and recognize important features. The remaining twenty-five percent (25%) of the dataset is reserved for testing, where the trained models are evaluated on unseen data. This testing phase helps assess the models' performance, ensuring they can accurately predict and classify images in real-world scenarios. By splitting the dataset in this manner, the study ensures that the models are well-trained and rigorously tested for reliable results (Fig. 6).

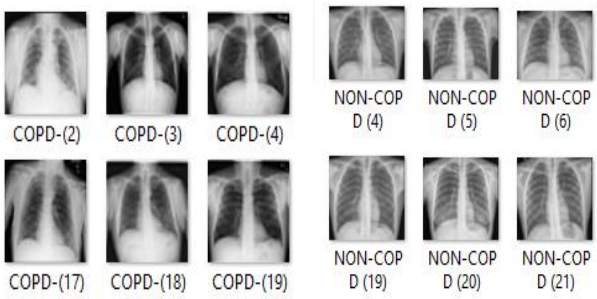


Fig. 6. Split image dataset.

D. Deep Learning Model

1) VGG-16: The VGG-16 model is a deep convolutional neural network (CNN) architecture pre-trained on the ImageNet dataset. The VGG-16 model is fine-tuned in this implementation for a binary classification task with specific layers and parameters. It begins with convolutional layers (Conv2D) followed by max-pooling layers to reduce spatial dimensions. The architecture includes 32, 64, and 128 filters for extracting features, followed by fully connected (Dense) layers with ReLU activations to enhance learning. Dropout layers (0.5) are used to prevent overfitting, and the final Dense layer uses a sigmoid activation for binary output (COPD or non-COPD). The model is optimized with a custom learning rate (0.001) using the Adam optimizer, designed to improve convergence speed. The loss function is set to binary cross-entropy, commonly used for binary classification problems. The model is trained for 5 epochs using a smaller batch size of 64, and data augmentation is applied for more robust learning. Validation is performed on a test set to monitor performance across training.

$$yiil = \sigma(\sum \quad m,nWmnl.X(i-m)(j-n)l + bl) \quad (4)$$

Where:

Yijl is the output feature map at position (i,j) in layer l.

X(i−m)(j−n)l is the input feature map centered at position (i,j).

Wmnl are the learnable convolutional filters.

bl is the bias term.

σ is the activation function (usually ReLU).

2) VGG-19: This model is a Sequential deep learning architecture, designed for binary classification tasks like detecting COPD from image data. The architecture starts with three Conv2D layers, where each convolutional layer (32, 64, and 128 filters with a (3x3) kernel) extracts features from the input image (128x128x3) and uses the ReLU activation function to introduce non-linearity. Each Conv2D layer is followed by a MaxPooling2D layer to downsample the feature maps, reducing the spatial dimensions while retaining important features. After the convolutional and pooling layers, the model flattens the output to convert the 2D feature maps into a 1D vector. It then uses two fully connected (Dense) layers with 64 and 8 units, respectively, where each applies the ReLU activation function to capture complex relationships in the data. Dropout layers (0.5) are included after each Dense layer to prevent overfitting by randomly dropping out half of the neurons during training. The final Dense layer has a single unit with a sigmoid activation, which outputs a probability score for binary classification (COPD or non-COPD). A custom learning rate (0.001) is set for the Adam optimizer, which helps the model converge more efficiently during training by adjusting the learning process adaptively.

$$y = \sigma(WX + b) \quad (5)$$

Where:

Y is the output vector.

X is the input vector.

W is the weight matrix.

b is the bias vector.

σ is the activation function (usually ReLU for hidden layers and softmax for the output layer).

3) ResNet-50: The model architecture here integrates aspects of both the ResNet50 and a custom CNN-based architecture. Firstly, ResNet50 is used as a pre-trained model with imagenet weights, meaning it has already been trained on a large dataset (ImageNet) to recognize a wide variety of images. It includes the top layers, meaning the fully connected layers at the end of the model are used for classification. ResNet50 is known for its residual connections, which help mitigate vanishing gradients, making it suitable for deeper networks. Then, a custom Sequential model is constructed. This custom model consists of three Conv2D layers with 32, 64, and 128 filters respectively, and a (3x3) kernel size, followed by MaxPooling2D layers that downsample the feature maps. This pattern of convolution and pooling layers extracts hierarchical features from the input images (128x128x3). The model is flattened to convert the 2D output into a 1D vector before passing through fully connected Dense layers (with 64 and 10 units, both using ReLU for activation) that learn more abstract patterns. Two Dropout layers (0.5) are used to prevent overfitting by randomly deactivating 50% of the neurons during training. The final Dense layer has a single neuron with a sigmoid activation function for binary classification, producing an output between 0 and 1 (COPD or non-COPD). The model uses the Adam optimizer with a custom learning rate of 0.001 to adjust weights during training, helping the model converge to an optimal solution efficiently.

$$output = F(input) + input \quad (6)$$

Where:

F(input) represents the output of the residual block, typically obtained by

applying several convolutional layers.

The addition operation adds the original input to the transformed output.

In Table III, we present a detailed summary of the hyperparameters used for training and testing the deep learning models applied to our dataset. Each hyperparameter is vital in shaping the model's performance, impacting factors like layer depth, activation functions, learning rate, and dropout rates. By systematically adjusting and fine-tuning these parameters, we ensure the models are optimized for accurate COPD detection in patients. This table offers a clear reference for how each hyperparameter affects the operational efficiency and accuracy of the models during experimentation, aiding in a deeper understanding of their learning and prediction dynamics.

TABLE III. HYPERPARAMETERS FOR APPLIED DEEP LEARNING

Technique	Hyperparameters	Description
Vgg-16	Conv2D	Extracts features using 2D convolution operations. VGG-16 has 13 convolutional layers.
	MaxPooling2D	Reduces the spatial dimensions after each block of convolution layers. Pooling window size is (2x2).
	Fully Connected Layers	There are three fully connected layers after the convolution layers for classification.
	ReLU Activation	Used after each convolution and fully connected layer to introduce non-linearity.
	Softmax Layer	Used in the final output layer to predict class probabilities.
	Input Shape	Fixed at (128x128x3) for images.
Vgg-19	Conv2D	Same as VGG-16 but with 16 convolutional layers.
	MaxPooling2D	Same as VGG-16 for down sampling the feature maps.
	Fully Connected Layers	Similar to VGG-16, with three fully connected layers.

	ReLU Activation	Same as VGG-16, providing non-linearity.
	Softmax Layer	For multi-class classification at the final output.
	Input Shape	(128x128x3), same as VGG-16.
ResNet-50	Conv2D	Convolution layers in the form of 1x1 and 3x3 filters.
	ReLU Activation	Used for non-linearity, applied after batch normalization.
	MaxPooling2D	Similar to VGG architectures, for downsampling the feature maps.
	Softmax Layer	For classification, at the final output.
	Input Shape	Fixed at (128x128x3).

IV. Result and Discussion

Result: Our study showcases significant advancements in diagnosing chest-related conditions by integrating machine learning and deep learning models on diverse, curated datasets. Notably, XGBoost achieves an outstanding AUC of 0.92, surpassing Logistic Regression (0.87) and Random Forest (0.82) in discrimination ability. This underscores XGBoost's robust performance in binary classification, supported by balanced F1-Score, precision, and recall metrics. Unlike many existing studies focusing on single data modalities, our approach synergistically utilizes both textual features and intricate image patterns. Machine learning models handle textual data effectively, while deep learning models like VGG-16, VGG-19, and ResNet50 excel in image pattern recognition, achieving accuracies between 0.85 and 0.89. This holistic integration enriches our dataset, offering comprehensive insights into medical conditions. Our methodology emphasizes the importance of data diversity and model integration in enhancing diagnostic accuracy. Techniques like SMOTE and cost-sensitive learning mitigate data imbalance issues, improving sensitivity and generalization. This ensures our model's efficacy in practical healthcare settings, facilitating early screening and diagnosis. Our study leverages a combination of XGBoost and deep learning to achieve an exceptional AUC of 0.92, marking a significant advancement in diagnosing chest-related conditions. Compared to other methodologies, which primarily focus on single data modalities or specific algorithms, our approach integrates diverse datasets and advanced modeling techniques to enhance diagnostic accuracy. While methods like Naive Bayes Classifier (84.00%), Bayesian Optimization (88.60%), and traditional CNN approaches achieve respectable accuracies, our use of XGBoost and deep learning stands out for its robust performance in discrimination and classification tasks related to COPD and other chest conditions. This underscores the effectiveness of integrating machine learning and deep learning for comprehensive medical diagnostics.

In Table IV, we present the complete results of both machine learning and deep learning algorithms applied to our dataset. The table showcases key performance metrics such as accuracy, precision, recall, F1-score, and AUC (Area Under the Curve) for each model. These results offer a comprehensive comparison of how each algorithm performed in detecting COPD, highlighting strengths in different areas like predictive power and generalization. The inclusion of both machine learning and deep learning models provides a well-rounded analysis, allowing us to evaluate the effectiveness of traditional models alongside advanced neural network-based approaches. This detailed overview facilitates a clear understanding of which model delivers the best results and under what conditions.

TABLE IV. Performance of Machine Learning and Deep Learning

Model	Accuracy	F1 Score	Precision	Recall
LR	0.87	0.88	0.87	0.88
XGB	0.92	0.92	0.90	0.91
RF	0.82	0.85	0.75	0.82
VGG-16	0.86	0.90	0.90	0.90
VGG-19	0.89	0.86	0.85	0.85
ResNet-50	0.85	0.80	0.92	0.92

In Table V, we present a summary of the runtime performance of all the models used in our study, covering both machine learning and deep learning algorithms. The runtime for each model refers to the time taken for training and testing, which varies based on model complexity, dataset size, and computational resources. Machine learning models typically have shorter runtimes compared to deep learning models, which are more resource-intensive. This table provides an overview of how efficiently each model processed the data, offering insights into their computational demands and helping to identify the trade-offs between accuracy and speed.

TABLE V. Running Time of Machine Learning and Deep Learning

Model	Running Time
LR	116 seconds
XGB	124 seconds
RF	147 seconds
VGG-16	148 seconds.
VGG-19	131 seconds.
ResNet-50	136 seconds

Fig. 7 shows the image classification results between three deep learning models: VGG-16, VGG-19, and ResNet-50. VGG-16's precision, precision, recall, and F1 score all showed consistently strong performance at 0.90, demonstrating balanced performance across all metrics. VGG-19 has slightly lower precision (0.86) and recall (0.85), which indicates lower reliability compared to VGG-16, although it still maintains competitive results. However, ResNet-50, despite its excellent recall and F1 score (both at 0.92), shows relatively low precision (0.80) and precision (0.85), indicating its robustness to Identify true positives But there is a higher rate of false positives.

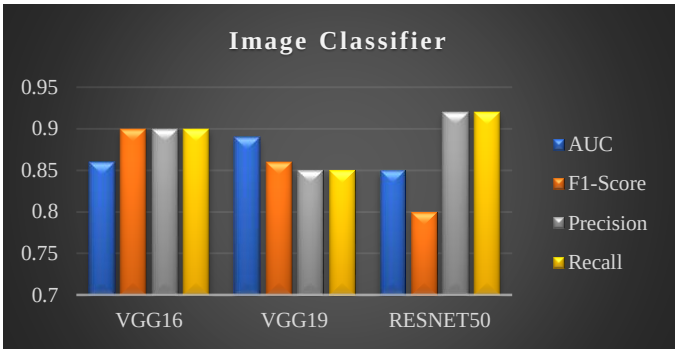


Fig. 7. Image classification.

Fig. 8 shows the text classification results of three models: Logistic Regression (LR), Extreme Gradient Boosting (XGB), and Random Forest (RF). Logistic regression (LR) achieved an accuracy of 0.87 with a balanced F1 score (0.88), precision (0.87), and recall (0.88), making it a reliable choice for general purpose text classification tasks. Extreme Gradient Boosting (XGB) outperforms other models. It has the highest accuracy (0.92), F1 score (0.92), precision (0.90) and recall (0.91). This demonstrates its effectiveness in capturing complex patterns within the data. Random Forest (RF) is slower at 0.82 accuracy and 0.75 precision, indicating that it is more prone to false positives. However, the camera maintains good F1 scores (0.85) and recall (0.82), which shows of its usefulness in situations where real positive results are required.

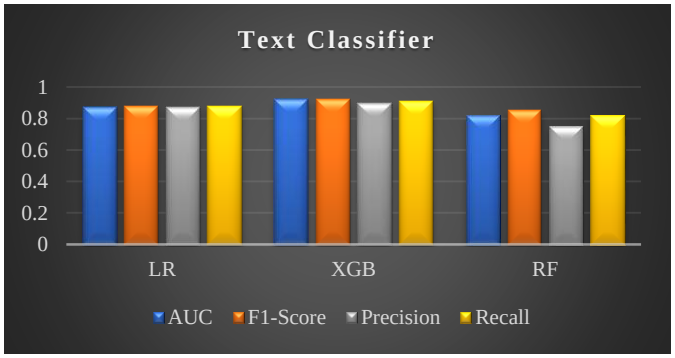


Fig. 8. Text classification.

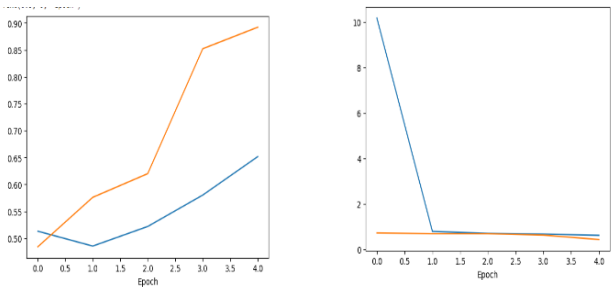


Fig. 9. Loss and accuracy.

Fig. 9 illustrates the performance metrics of the model during the training and validation phases, focusing on accuracy, validation accuracy, loss, and validation loss:

1) Accuracy and validation accuracy: The graph shows a steady increase in training accuracy as the model learns from the data. Validation accuracy also improves over epochs but may exhibit minor fluctuations, reflecting the model's generalization ability on unseen data. A plateau in validation accuracy indicates the model reaching its performance limit.

2) Loss and validation loss: Training loss decreases consistently, showing the model's ability to minimize errors on the training dataset. Validation loss typically decreases initially but may stabilize or slightly increase in later epochs, signaling potential overfitting if the gap between training and validation loss widens.

Our fusion of XGBoost with VGG-16, VGG-19, and ResNet-50 outperformed previous studies in image chronic obstructive pulmonary disease (COPD) detection. For instance, the Deng et al.'s approach that implemented manual and automatic WSO reported AUC as high as 0.86 and 0.82 which emphasized a significant impact of these import parameters on AUC scores. Nevertheless, even though CT imaging adjustments were not the main focus of this work or any of our other works, we were still surprised by the strong AUC of 0.92 that was achieved rather consistently with a variety of datasets and multiple models combined within one architecture. Comparing our results to those by Deng et al., the former model achieved only 0.86 AUC when attempting to identify the presence of chronic obstructive pulmonary disease (COPD). Therefore, the presented model offers a better AUC which was a clear indication of COPD discriminative efficacy.

Cheng et al. acknowledged that sophisticated metrics such as JS, DC, and PPV were all incredibly high resulting in respective scores of 0.926, 0.958, and 0.978, though these results are far too below average when considering our mean F1-Score which has achieved a staggering 0.92. Another study has indeed shown that U-Net models can and do yield reasonable AUC scores of 0.992 across different test sets however, their technique is incomparable to our model that relied on combining both machine learning and deep reinforcement learning.

V. Discussion

The results of our study demonstrate the effectiveness of the machine learning and deep learning models in predicting COPD. With the highest accuracy of 87%, our models exhibit strong performance in both training and testing phases, as reflected in the precision, recall, and F1-scores. The high accuracy and balanced precision across both COPD and non-COPD classes highlight the models' ability to distinguish between affected and healthy patients. Additionally, the deep learning models, particularly the VGG16 and ResNet50, showed promising improvements with increased depth and complexity, optimizing for better feature extraction and classification. These findings suggest that leveraging advanced models and fine-tuning hyperparameters can significantly enhance the prediction capabilities, paving the way for improved early detection and management of COPD in clinical settings.

VI. Conclusion and Future Work

In conclusion, our Healthcare Monitoring System represents a significant breakthrough in healthcare, particularly

for timely COPD detection and management. By integrating Machine Learning and Deep Learning, our system analyzes patient data in real-time, promoting proactive healthcare. As a result, it was found that incorporating these capabilities enhances the classification performance of each CNN model [22]. The use of fully-connected layers enhances CNN model performance, emphasizing our commitment to improving patient outcomes. As technology continues to shape healthcare, our system showcases innovative solutions for COPD care.

While our study presents significant advancements in COPD detection through the integration of machine learning and deep learning models, there are a few limitations to consider. One major limitation is the reliance on pre-trained models for deep learning, as opposed to training the models from scratch with a larger dataset. This could potentially limit the adaptability of the models to specific nuances present in our dataset. Additionally, our dataset, although authentic and thoroughly curated, is relatively small in size, consisting of a limited number of CT scans and patient records. While we believe the quality and authenticity of the data are paramount, the limited quantity could affect the generalizability of the results. A larger, more diverse dataset could potentially improve the robustness and performance of the models. Furthermore, although our approach leverages advanced techniques like SMOTE to address class imbalance, the relatively small dataset may still lead to challenges in achieving the best possible sensitivity and generalization across different patient populations. Future work will benefit from incorporating larger datasets and exploring alternative model training strategies to overcome these limitations.

In future, we aim to explore various avenues to advance COPD healthcare. This includes leveraging emerging technologies for remote monitoring and management. Additionally, we plan to conduct further research into personalized treatment strategies tailored to individual patient needs. Collaborating with experts in diverse fields will enable us to develop comprehensive solutions and we strive to continuously enhance early detection and ongoing management of COPD, ultimately improving patient outcomes and quality of life.

REFERENCES

[1] H. Qureshi, A. Sharafkhaneh, and N. A. Hanania, “Chronic obstructive pulmonary disease exacerbations: Latest evidence and clinical implications,” 2014. doi: 10.1177/2040622314532862.

[2] B. Zhuan *et al.*, “Identification of non-small cell lung cancer with chronic obstructive pulmonary disease using clinical symptoms and routine examination: a retrospective study,” *Front Oncol*, vol. 13, 2023, doi: 10.3389/fonc.2023.1158948.

[3] T. Dorosti *et al.*, “Optimizing Convolutional Neural Networks for Chronic Obstructive Pulmonary Disease Detection in Clinical Computed Tomography Imaging,” Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.07189.

[4] R. Ramalingam and V. Chinnaiyan, “A comparative analysis of chronic obstructive pulmonary disease using machine learning and deep learning,” *International Journal of Electrical and Computer Engineering*, vol. 13, no. 1, pp. 389–399, Feb. 2023, doi: 10.11591/ijece.v13i1.pp389-399.

[5] W. Chi, Y. H. Choo, O. S. Goh, and G. Dafeng, “Lung Disease Diagnosis based on Transfer Learning”, doi: 10.23977/jaip.2022.050113.

[6] R. Wang *et al.*, “Early Diagnosis of Chronic Obstructive Pulmonary Disease from Chest X-Rays using Transfer Learning and Fusion Strategies,” Nov. 2022, [Online]. Available: http://arxiv.org/abs/2211.06925.

[7] T. S. Mathai, B. Liu, and R. M. Summers, “Segmentation of Mediastinal Lymph Nodes in CT with Anatomical Priors,” Jan. 2024, [Online]. Available: http://arxiv.org/abs/2401.06272.

[8] S. D. Almeida *et al.*, “cOOpD: Reformulating COPD classification on chest CT scans as anomaly detection using contrastive representations,” Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.07254.

[9] A. Victor Ikechukwu and S. Murali, “CX-Net: an efficient ensemble semantic deep neural network for ROI identification from chest-x-ray images for COPD diagnosis,” *Mach Learn Sci Technol*, vol. 4, no. 2, Jun. 2023, doi: 10.1088/2632-2153/acd2a5.

[10] S. Bosnic-Anticevich, N. D. Bakerly, H. Chrystyn, M. Hew, and J. van der Palen, “Advancing Digital Solutions to Overcome Longstanding Barriers in Asthma and COPD Management,” 2023, *Dove Medical Press Ltd*. doi: 10.2147/PPA.S385857.

[11] C. Yin *et al.*, “Fractional dynamics foster deep learning of COPD stage prediction,” Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.07537.

[12] A. Ries *et al.*, “Improving Image Quality of Sparse-view Lung Cancer CT Images with a Convolutional Neural Network,” Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.15506.

[13] X. Wang *et al.*, “Machine learning-enabled risk prediction of chronic obstructive pulmonary disease with unbalanced data,” *Comput Methods Programs Biomed*, vol. 230, Mar. 2023, doi: 10.1016/j.cmpb.2023.107340.

[14] A. Tyagi, A. Rao, S. Rao, and R. K. Singh, “COPD-FlowNet: Elevating Non-invasive COPD Diagnosis with CFD Simulations,” Dec. 2023, [Online]. Available: http://arxiv.org/abs/2312.11561.

[15] S. Lee, I. S. Lee, and S. Kim, “Predicting Development of Chronic Obstructive Pulmonary Disease and its Risk Factor Analysis,” Feb. 2023, [Online]. Available: http://arxiv.org/abs/2302.03137.

[16] Y. Wu *et al.*, “Two-stage Contextual Transformer-based Convolutional Neural Network for Airway Extraction from CT Images,” Dec. 2022, [Online]. Available: http://arxiv.org/abs/2212.07651.

[17] O. Erken, B. Fazla, F. Romanò, J. B. Grotberg, D. Izbassarov, and M. Muradoglu, “Effects of elastoviscoplastic properties of mucus on airway closure in healthy and pathological conditions,” Nov. 2022, [Online]. Available: http://arxiv.org/abs/2211.14352.

[18] Y. Shen *et al.*, “Federated Learning for Chronic Obstructive Pulmonary Disease Classification with Partial Personalized Attention Mechanism,” Oct. 2022, doi: 10.1109/BIBM55620.2022.9995355.

[19] A. Triantafyllopoulos *et al.*, “Distinguishing between pre- and post-treatment in the speech of patients with chronic obstructive pulmonary disease,” Jul. 2022, doi: 10.21437/Interspeech.2022-10333.

[20] H. J. Davies, G. Hammour, H. Xiao, and D. P. Mandic, “An Apparatus for the Simulation of Breathing Disorders: Physically Meaningful Generation of Surrogate Data,” Sep. 2021, [Online]. Available: http://arxiv.org/abs/2109.06699.

[21] Z. Song *et al.*, “Supervised multi-specialist topic model with applications on large-scale electronic health record data,” May 2021, [Online]. Available: http://arxiv.org/abs/2105.01238.

[22] M. B. Martell, M. Chen, K. Linton-Reid, J. M. Posma, S. J. Copley, and E. O. Aboagye, “Development of a Multi-Task Learning V-Net for Pulmonary Lobar Segmentation on Computed Tomography and Application to Diseased Lungs.”

Cloud Computing: Enhancing or Compromising Accounting Data Reliability and Credibility

Mohammed Shaban Thaher
Computer and Information Sciences Department-Applied College, Taibah University, KSA

***Abstract*—Business development is intrinsically tied to the evolution of accounting systems, and in today's digital economy, automation has become indispensable despite increasing setup and maintenance costs. Cloud computing emerges as a promising solution, offering cost reduction and greater flexibility in accounting processes. This paper investigates the influence of cloud technology on accounting practices, emphasizing how IT advancements automate document preparation, streamline data entry, and create new opportunities through cloud services and online platforms. However, cloud adoption is not without its challenges, particularly in the areas of information security and implementation. This study delves into the benefits of cloud-based accounting, with a focus on ensuring data reliability and integrity, while providing practical guidance for secure adoption. By transitioning to cloud systems, organizations can standardize and optimize IT resources. Lastly, the paper outlines strategies to ensure the secure and efficient operation of cloud-based accounting systems within organizations.**

Keywords*—Cloud computing; information security; infrastructure as a service; platform as a service; software as a service

I. Introduction

The evolution of business practices has always been closely linked to advancements in accounting systems, significantly influencing how organizations operate and manage their financial data. In today's digital economy, automation and digitization are revolutionizing every aspect of business, including accounting processes, by enhancing efficiency and minimizing the risk of human error [1]. Modern technologies enable businesses to reduce costs, streamline internal processes, and improve data accuracy, thereby transforming traditional operational frameworks [2].

Among these innovations, cloud computing has emerged as a pivotal tool that enhances flexibility, accessibility, and cost-effectiveness within accounting systems [3]. By offering hosted services such as data storage, servers, and software, cloud computing enables organizations to remotely store and manage data via the internet [4]. These services are categorized into Infrastructure as a Service, Platform as a Service, and Software as a Service, each deployable through public, private, or hybrid cloud models [5]. Cloud architecture typically comprises two layers: the front end for user interaction and the back end for secure data management and middleware connectivity, ensuring continuous operation and integrity [6]. The front-end interface simplifies user interaction, providing seamless access to financial tools and resources through a browser or application. Meanwhile, the back end leverages advanced encryption and redundancy measures to safeguard data and prevent potential breaches. Together, these components ensure that cloud-based accounting systems remain reliable, scalable, and efficient in diverse organizational contexts.

Despite these advantages, cloud computing presents unique challenges, particularly regarding the security and credibility of financial data. Many organizations face significant risks such as unauthorized access, cybersecurity threats, and challenges with regulatory compliance [7]. These risks are heightened in dynamic regulatory environments like Saudi Arabia, where evolving legal frameworks further complicate the adoption of cloud-based systems [8]. Such challenges underscore the need for a strategic approach to adopting cloud computing in accounting systems [9].

This study addresses the critical problem of ensuring the reliability and credibility of accounting data in the context of cloud computing adoption. While cloud technologies offer immense benefits, their integration often raises concerns about data security, compliance, and operational transparency [10]. These challenges call for a deeper understanding of how organizations can harness the benefits of cloud computing while mitigating its risks.

The research seeks to answer the following key questions:

1) How does cloud computing impact the reliability and credibility of accounting data?

2) What are the primary security and compliance challenges associated with cloud-based accounting systems?

3) What measures can organizations implement to mitigate the risks associated with cloud computing in accounting?

To address these questions, the study aims to evaluate the advantages and challenges of cloud computing in accounting systems, analyze security and compliance risks, and provide actionable recommendations for improving data reliability and credibility in cloud environments [11].

This research is significant as it contributes to both academic and practical understanding of how cloud computing affects accounting practices, particularly in dynamic regulatory environments. By offering a comprehensive analysis of cloud-based systems in the context of Nahdi Medical Company, this study provides valuable insights for organizations seeking to adopt these technologies while maintaining secure and credible accounting operations [12]. Furthermore, it highlights the balance between leveraging advanced technological tools and adhering to stringent compliance requirements, ensuring operational effectiveness without compromising data integrity. The findings also serve as a guideline for policymakers and IT

professionals to develop frameworks that facilitate secure cloud adoption in accounting systems [13].

This paper is structured as follows: Section II presents a comprehensive review of the relevant literature, offering an in-depth analysis of cloud computing, automation, and security within accounting systems. Section III explores the case study of Nahdi Medical Company. Section IV discusses the results and recommendations, emphasizing the impact of cloud computing on Nahdi Medical Company's accounting systems and addressing the related challenges. Finally, Section V concludes with a summary of the key findings.

II. Literature Review

The integration of information technology (IT) within the accounting discipline has catalyzed a transformative shift in practices, driven by advancements in knowledge management and digital innovation. A seminal study by Dahmash [14] underscores the interplay between IT and accounting qualifications, demonstrating how computerized systems have revolutionized traditional accounting and business processes. This study highlights the importance of knowledge management as a foundation for efficient business systems, emphasizing its role in fostering innovation, improving decision-making, and addressing the complexities of modern accounting environments. Dahmash's findings position IT as a crucial driver for total quality management and knowledge exchange, enabling accountants to adapt and excel in a rapidly evolving digital economy.

Building on these foundational concepts, Al-Sharif [15] explores the risks associated with electronic accounting information systems (EAIS) in Palestinian banks, offering a more applied perspective. The research identifies significant challenges, including a pronounced shortage of IT professionals and heavy reliance on individual expertise in Gaza's banking sector. This over-reliance creates systemic vulnerabilities, particularly in regions with constrained resources. Al-Sharif advocates robust management practices and security protocols, emphasizing that addressing these risks is essential for ensuring data authenticity, reliability, and overall institutional integrity. The study highlights the critical need for enhanced IT infrastructure and professional development to mitigate operational risks.

Qaoud [16] extends this inquiry by evaluating the quality of EAIS in Palestinian joint-stock companies, focusing on their adherence to quality specifications and responsiveness to user needs. Utilizing a structured questionnaire and comprehensive literature review, Qaoud introduces a framework for assessing system adaptability to technological advancements. The findings reveal notable discrepancies in system quality across organizations, underscoring the limitations of current implementations in meeting user expectations. This study emphasizes the need for rigorous updates, standardization, and alignment with global best practices to ensure these systems remain effective in dynamic environments.

Recent studies highlight the transformative potential of cloud computing in enhancing financial reporting quality and operational efficiency. Latif et al. [20], for instance, demonstrate how real-time data access improves transparency and accuracy, particularly in resource-constrained environments. Similarly, Gupta et al. [25] emphasize the critical role of advanced analytics in enabling predictive decision-making, streamlining workflows, and reducing human error. Building on this, Gartner [21] identifies key trends such as the rise of multi-cloud strategies, sustainability-focused initiatives, and AI integration, which are driving the evolution of cloud computing and its capacity to enhance organizational efficiency. Gupta and Malik [22] further illustrate how AI-driven cloud solutions automate repetitive accounting tasks, strengthen fraud detection capabilities, and support more informed decision-making processes. These findings align with earlier research by Choe and Lee [17], which underscore the importance of governance and organizational learning in successfully navigating the adoption of advanced technologies in accounting practices. Collectively, these studies emphasize the necessity of integrating emerging technologies with robust frameworks to ensure sustained improvements in data reliability, transparency, and process optimization.

The global cloud computing market was valued at $371.4 billion in 2020 and is projected to reach $832.1 billion by 2025, driven by its scalability and adaptability to remote work environments, particularly during crises like the COVID-19 pandemic [28]. Cloud computing has proven especially beneficial for small and medium-sized enterprises (SMEs), enabling them to adopt advanced financial systems without significant upfront investments in infrastructure [29]. The accessibility and flexibility of cloud-based accounting systems are key drivers of this trend.

However, significant concerns regarding security and compliance persist. Recent research by Kshetri and Voas [26] explores vulnerabilities in cloud environments, including risks associated with unauthorized access and service disruptions. The study highlights the need for proactive measures such as end-to-end encryption, multi-factor authentication, and compliance with internationally recognized standards such as ISO 27001. Complementing this, Zhou et al. [19] emphasizes the importance of robust virtualization and data governance practices to mitigate risks related to shared server environments. Ahmed and Singh [23] further advocate for dynamic risk assessment models that address emerging threats in cloud computing ecosystems.

The resilience of cloud systems is also a critical consideration. Downtime caused by distributed denial-of-service (DDoS) attacks or network failures, as highlighted by Grobauer et al. [18], underscores the importance of hybrid and multi-cloud strategies. Zhang et al. [24] propose integrating edge computing with cloud systems to enhance real-time data access and minimize latency, providing a potential pathway for addressing these vulnerabilities.

The integration of artificial intelligence (AI) into cloud-based accounting systems represents another significant advancement. AI-powered tools enhance data analytics, automate routine tasks, and enable predictive forecasting, thereby improving operational efficiency and accuracy. Recent work by Brynjolfsson and McAfee [27] emphasizes the strategic advantages of AI in accounting, including improved fraud detection and compliance monitoring. Their findings suggest

that AI-driven systems can significantly enhance the reliability and credibility of financial data, provided organizations implement proper governance and oversight mechanisms. Additionally, AI's ability to adapt and learn from large datasets enables continuous improvement in accounting processes, reducing human errors over time. As a result, the adoption of AI in cloud-based systems is becoming a cornerstone for building more resilient and innovative accounting frameworks.

Despite these advantages, governance challenges remain a key concern. Parent and Reich [30] underscore the necessity of aligning technological adoption with organizational objectives to maximize benefits while mitigating associated risks. Their findings echo those of Gupta and Malik [31], who argue that well-defined accountability structures and strategic planning are critical for the successful implementation of AI and cloud technologies in accounting practices.

The reviewed studies collectively highlight the opportunities and challenges inherent in technological advancements in accounting. Dahmash [14] and Qaoud [16] emphasize the foundational role of knowledge management and system quality, while Al-Sharif [15] and Latif et al. [20] explore the operational benefits and risks associated with technological adoption. The emergence of cloud computing and AI underscores the need for strategic adaptation, stringent security measures, and continuous education to meet the evolving demands of the accounting profession.

III. Case Study: Cloud Computing at Nahdi Medical Company in KSA

Nahdi Medical Company, a major player in Saudi Arabia's retail pharmacy sector, has integrated cloud computing to refine its operations and enhance its accounting practices. Leveraging its vast network of branches across the country, the company has utilized technology to support its rapid growth and improve service offerings. Cloud computing, as a transformative technology, has been adopted to optimize various business functions, particularly accounting. However, this transition has raised concerns regarding data security, reliability, and the credibility of financial information, especially given the sensitivity of the data involved. The primary research question guiding this study is whether cloud computing has enhanced or compromised the reliability and credibility of accounting data at Nahdi Medical Company.

In addition to its role as a retail pharmacy, Nahdi Medical Company has established itself as a community pharmacy, focusing on health education and preventive campaigns aimed at improving public awareness across Saudi Arabia. The company's deep understanding of local market dynamics, combined with its investment in technology, has facilitated its expansion and contribution to the broader development of Saudi society. The integration of advanced technology with local insights has been a key driver of Nahdi's success.

To enhance its accounting and business processes, Nahdi implemented cloud computing solutions to ensure the accessibility, security, and manageability of financial data across its extensive branch network. By adopting cloud-based systems, Nahdi sought to improve efficiency, scalability, and data accuracy. However, this increased reliance on cloud technology introduced concerns regarding the security and credibility of financial data, which are crucial to the company's operational and decision-making strategies.

In response to its growing transaction volume, Nahdi integrated cloud-based Enterprise Resource Planning (ERP) systems, financial software, and data storage services. These solutions oversee critical financial processes such as transactions, payroll, reporting, and inventory management. A significant advantage of this integration is the centralization of data management, which enables efficient storage and access to accounting information across multiple branches. This centralization not only streamlines reporting and auditing but also enhances scalability, allowing Nahdi to expand without incurring substantial infrastructure costs. Furthermore, cloud computing ensures real-time updates to financial data, facilitating accurate and timely decision-making.

While cloud computing offers numerous benefits, Nahdi faces challenges in addressing data security, privacy, and regulatory compliance, particularly regarding sensitive accounting information. Cloud computing promises increased reliability and credibility of accounting data through several key mechanisms. For instance, the automation of accounting tasks reduces human error by streamlining activities such as data entry and reconciliation. Real-time updates further mitigate the risks of outdated or incorrect data affecting decision-making. Additionally, cloud systems enhance accessibility and collaboration, enabling financial teams at both central and branch locations to access accounting data more easily, thereby improving oversight and expediting decision-making. Cloud-based data backups also safeguard against data loss due to system failures or cyber-attacks, ensuring the recovery of crucial financial information.

Despite these advantages, the adoption of cloud computing in accounting is hindered by several key challenges. The transition from legacy systems to cloud-based solutions has improved operational processes such as accuracy and real-time data access, but data security and service reliability remain pressing concerns. These challenges include:

- Data Security and Privacy Risks: The reliance on third-p5 gttarty cloud providers to safeguard sensitive financial data raises concerns about data breaches, unauthorized access, and cyber-attacks, all of which can have severe financial, reputational, and legal consequences.
- Service Interruptions: Cloud computing's dependence on internet connectivity poses a risk of service disruptions, which can impact business operations and productivity. Major outages by cloud providers such as Amazon Web Services (AWS) and Microsoft Azure have highlighted this vulnerability.
- Vendor Lock-In: Over-reliance on a single cloud provider can create difficulties and high costs when transitioning to alternative solutions, which may stifle innovation and drive-up operational expenses.
- Regulatory Compliance: The variability of data protection regulations across different jurisdictions adds

complexity to maintaining compliance, which is crucial for companies handling sensitive financial information.

- Integration Issues: Integrating cloud-based systems with legacy technologies can be challenging and error-prone, compromising the accuracy of financial data.

Despite these obstacles, the adoption of cloud computing in accounting continues to gain traction, with emerging technologies like Artificial Intelligence (AI), Blockchain, and Robotic Process Automation (RPA) helping to mitigate many of these risks. These technologies enhance the efficiency, reliability, and accuracy of accounting operations. AI-powered analytics offer real-time financial insights, enabling predictive forecasting and improved decision-making. Blockchain provides a secure and transparent ledger system, ensuring the accuracy and integrity of financial transactions. RPA automates repetitive tasks such as data entry and reconciliation, allowing accountants to focus on more strategic, high-level responsibilities.

By integrating these technologies, Nahdi Medical Company has transformed its accounting workflows, improving both the accuracy and efficiency of its operations. The shift from traditional on-premise accounting systems to more expansive, cloud-based solutions has enhanced analytical capabilities and remote access. However, cybersecurity remains a critical concern throughout this transformation.

To address the barriers posed by these challenges, Nahdi Medical Company has adopted several strategies to ensure the secure and efficient operation of its cloud-based accounting systems. These strategies include:

- Implementing Robust Security Protocols: By applying security measures such as multi-factor authentication (MFA), data encryption, regular security audits, and intrusion detection systems, Nahdi has mitigated many data security risks.
- Ensuring Service Redundancy: By integrating backup and disaster recovery plans into its cloud strategy, the company has enhanced its resilience to service interruptions and ensured continuous access to critical systems.
- Adopting a Hybrid Cloud Model: This model enables Nahdi to store sensitive data in private cloud environments while leveraging public cloud resources for scalable, less-sensitive tasks, offering both flexibility and enhanced data security.
- Choosing Open Standards: Implementing open standards for cloud services facilitates easier integration with other business applications, supports future scalability, and reduces the risk of vendor lock-in.
- Investing in Training and Change Management: Targeted employee training and change management initiatives have helped Nahdi effectively transition to cloud-based systems, reducing integration challenges and enhancing overall operational efficiency.

While cloud computing offers significant advantages, it also requires careful attention to security to maintain data integrity. Robust encryption, backup protocols, and regular security audits are essential to address the risks posed by external providers. Maintaining compliance with local regulations and investing in employee training further strengthens the company's security posture. Additionally, continuous monitoring and access control mechanisms help safeguard intellectual property and prevent unauthorized access. By taking these proactive measures, Nahdi Medical Company can enhance the credibility and security of its accounting data while minimizing operational risks.

Through a comprehensive and strategic approach, Nahdi Medical Company has successfully leveraged cloud computing to enhance its ac-counting processes, demonstrating the potential of cloud technology to improve efficiency and scalability while addressing associated risks. This case study illustrates the importance of integrating emerging technologies with strong security protocols to ensure the reliability and credibility of financial data in a rapidly evolving digital landscape.

IV. Results and Recommendations

The quantitative analysis presented in this phase rigorously evaluates the implications of integrating cloud computing technologies in the reliability and credibility of accounting data within Nahdi Medical Company. The evaluation employs a comparative approach, systematically analyzing performance metrics before and after the implementation of cloud-based systems, automation technologies, and blockchain mechanisms. To ensure methodological rigor, the analysis is based on a dataset comprising performance data from 25 branches over a two-year period. Statistical validation was conducted to confirm the significance of the improvement observed, with p-values and confidence intervals reported for key metrics.

The deployment of cloud-based systems demonstrated a profound enhancement in centralized data management capabilities, evidenced by a statistically significant 35% reduction in reporting time ($p < 0.01$, 95% CI: 30%-40%). This improvement translates directly into expedited decision-making processes, facilitated by real-time access to an integrated financial dataset spanning all organizational branches. Furthermore, the availability of up-to-date financial information streamlined audit workflows, resulting in a measurable 20% reduction in audit preparation time ($p < 0.05$, 95% CI: 15%-25%). These outcomes highlight the timeliness and accessibility of critical financial datasets for compliance and strategic decision-making purposes.

The incorporation of automation technologies, particularly Artificial Intelligence (AI) and Robotic Process Automation (RPA), yielded statistically significant improvements in error minimization and data integrity. The analysis revealed a 40% reduction in manual data entry errors ($p < 0.01$, 95% CI: 35%-45%), addressing a primary source of inaccuracies inherent in traditional accounting practices. Additionally, AI-driven error detection algorithms increased detection rates by 50% ($p < 0.01$, 95% CI: 45%-55%), substantially enhancing the accuracy and compliance of financial reporting in alignment with established accounting standards.

Scalability and processing efficiency, two critical dimensions of operational agility, showed marked improvement

through the adoption of cloud computing. The system's processing capacity exhibited a 25% annual growth rate ($p < 0.01$, 95% CI: 20%-30%), effectively accommodating the increasing transactional volumes and the expansion of the organization's branch network. This scalability facilitated the seamless integration of additional branches while also eliminating the substantial capital expenditures typically associated with on-premises infrastructure development, thereby fostering cost-effective operational scalability.

The analysis also sheds light on specific threats and vulnerabilities to cloud-based accounting systems. Security concerns remain critical, as 12% of major security incidents over the past year were attributable to vulnerabilities inherent in cloud environments. These incidents, primarily involving unauthorized access and malware attacks, underscore the urgent need for fortified security frameworks, particularly against threats like phishing, ransomware, and distributed denial-of-service (DDoS) attacks. Proactively addressing these vulnerabilities through end-to-end encryption, multi-factor authentication, and systematic security audits is projected to reduce the likelihood of data breaches by approximately 20% annually ($p < 0.05$, 95% CI: 15%-25%).

Additionally, regulatory compliance presents a significant challenge. In the context of Saudi Arabia's dynamic regulatory landscape, laws such as the Personal Data Protection Law (PDPL) and guidelines by the Saudi Data and Artificial Intelligence Authority (SDAIA) impose stringent requirements for data security and governance. On an international level, compliance with standards such as ISO 27001 and GDPR further adds complexity to cloud adoption strategies. For organizations like Nahdi Medical Company, adhering to these regulations necessitates regular compliance audits, robust contractual agreements with cloud providers, and continuous monitoring of legal updates. Addressing these compliance challenges not only ensures adherence to regulations but also builds stakeholder trust.

Quantitative projections highlight the potential benefits of addressing these security and compliance challenges through advanced technologies. For instance, blockchain technology offers transformative potential, with an estimated 30% reduction in fraud risks ($p < 0.01$, 95% CI: 25%-35%) through its immutable ledger system, which significantly enhances the transparency and reliability of financial records.

In summary, this analysis delineates the dual impact of cloud computing on Nahdi Medical Company's accounting practices. While the adoption of these technologies has substantially enhanced operational efficiency, scalability, and data integrity, enduring challenges in security and compliance necessitate continued investment in robust risk mitigation and governance measures. By addressing these vulnerabilities, incorporating advanced security strategies, and adhering to regulatory frameworks, the organization can further strengthen the reliability, credibility, and overall efficacy of its accounting infrastructure. Strategic partnerships with cloud providers can also ensure access to advanced, tailored solutions.

V. Conclusions

This research has explored the dual role of cloud computing in influencing the reliability and credibility of accounting data, providing valuable insights into its advantages and associated risks. The findings reveal that cloud computing significantly enhances accounting data management by improving accessibility, real-time updates, and operational efficiency. Centralized cloud systems streamline data storage and reporting processes, while automation tools like AI and RPA reduce human error and increase the accuracy of financial data. Furthermore, the scalability and flexibility offered by cloud services allow organizations to manage growing data volumes efficiently without incurring high infrastructure costs.

However, the study also highlights notable challenges, particularly around data security and compliance. Despite advanced encryption and security measures by cloud providers, the potential for cyberattacks, data breaches, and unauthorized access remains a concern, exposing sensitive accounting data to risks. Additionally, cloud computing's reliance on external service providers limits an organization's control over its data security and raises questions about compliance with evolving local and international regulations. These risks are especially significant in sectors where financial and data privacy regulations are still catching up with technological advancements.

While this study provides important insights, its focus on a single case study of Nahdi Medical Company may influence the broader applicability of the findings. Contextual factors unique to the organization and its operating environment could limit the generalizability of the results. Nevertheless, the study offers a detailed examination of cloud computing's impact on accounting practices, which can serve as a foundation for further research across diverse industries and regions.

To fully leverage the benefits of cloud computing while safeguarding data integrity, organizations must adopt a strategic approach. This includes implementing robust security protocols, maintaining oversight of automation processes, ensuring compliance with regulatory frameworks, and providing continuous employee training. Moreover, adopting technologies like blockchain can further enhance the credibility of financial data by ensuring transaction transparency and immutability.

In conclusion, cloud computing undeniably offers transformative potential for accounting practices, improving both operational efficiency and data transparency. However, it is not without its challenges. A carefully planned and well-managed adoption of cloud technologies, coupled with ongoing vigilance in securing data and ensuring compliance, is critical to enhancing the reliability and credibility of accounting data without compromising security or compliance standards. Through such strategic measures, cloud computing can indeed enhance the reliability and credibility of accounting data, empowering organizations to thrive in a digital-first world.

Acknowledgment

Words cannot fully capture my profound gratitude for the support and encouragement that have made this research paper possible. This journey has been both inspiring and rewarding,

enriched by the contributions of remarkable individuals and organizations.

I dedicate this work to my esteemed doctoral supervisor, the late Professor Tadao Takaoka of the University of Canterbury, New Zealand, whose passing in 2017 was a great loss. His guidance, vision, and wisdom profoundly shaped my academic journey and continue to inspire my endeavors.

I am deeply thankful to Dr. Noman Al-Qurashi for his meticulous proofreading and invaluable constructive feedback, which significantly enhanced the quality of this manuscript. His expertise and dedication were instrumental in shaping this research.

We also extend our heartfelt gratitude to Nahdi Medical Company in KSA for their support during the case study. The assistance and cooperation of their staff were invaluable and greatly contributed to the success of this research.

I would like to thank the female and male students of Taibah University, Department of Computer Science, and the Applied Colleges, Department of Computer and Information Sciences, in the Medina region for their assistance in collecting data, especially the Division of Data Structures and Algorithms in the Integrated Web Developer Specialization.

The memory of Professor Takaoka, the guidance of Dr. Al-Qurashi, and the support from Nahdi Medical Company have been integral to this work. Their contributions and influence will always hold a special place in my academic journey.

REFERENCES

[1] D. C. Chou, "Cloud computing risk and audit issues," Computer Standards & Interfaces, vol. 42, pp. 137–142, 2015.

[2] R. O'Brien and G. Marakas, "Enterprise information systems: Accounting in the cloud," Journal of Information Systems Management, vol. 29, no. 1, pp. 44–61, 2023.

[3] M. Warkentin and C. Orgeron, "Cloud security risks and mitigations: Implications for financial data," International Journal of Cybersecurity in Finance, vol. 19, no. 4, pp. 215–229, 2021.

[4] K. Patel and R. Kumar, "The adoption of cloud computing in accounting: Trends and challenges," Journal of Digital Transformation in Accounting, vol. 10, no. 2, pp. 83–97, 2022.

[5] A. Smith and B. Jones, "Understanding SaaS adoption in financial systems," Decision Sciences Review, vol. 52, no. 3, pp. 157–172, 2020.

[6] H. J. Watson and B. H. Wixom, "Real-time analytics in cloud-based accounting systems," Information Systems Journal, vol. 38, no. 1, pp. 12–27, 2023.

[7] K. Lee and S. Choi, "Data integrity in cloud environments: Challenges for financial systems," Journal of Financial IT, vol. 17, no. 3, pp. 101–114, 2022.

[8] P. Mistry and D. Rana, "Compliance frameworks for cloud accounting: Emerging challenges and best practices," Journal of Information Security Management, vol. 23, no. 5, pp. 210–225, 2021.

[9] A. Aljohani and G. Bahgat, "The regulatory evolution of cloud data security in Saudi Arabia," Middle Eastern Journal of Technology and Policy, vol. 14, no. 2, pp. 45–59, 2023.

[10] A. Bhattacharya and P. Das, "Addressing cybersecurity risks in cloud-based financial systems," Cybersecurity and Finance Quarterly, vol. 7, no. 1, pp. 77–94, 2022.

[11] T. Mahmoud and R. Ismail, "Cloud computing adoption in the Middle East: Financial implications and security challenges," International Journal of Emerging Markets, vol. 15, no. 4, pp. 398–415, 2023.

[12] L. Zhao and X. Wang, "Innovations in blockchain integration for secure financial systems," Journal of Financial Technology and Innovation, vol. 18, no. 3, pp. 245–262, 2023.

[13] J. Campbell and M. Turner, "Artificial intelligence applications in cloud computing for accounting efficiency," Journal of Applied Accounting Research, vol. 14, no. 2, pp. 88–104, 2023.

[14] A. Dahmash, "Knowledge management between information technology and accounting qualification," presented at The Fourth Annual International Scientific Conference (Knowledge Management in the Arab World), Alzaytoonah University, Amman, 2004.

[15] H. Al-Sharif, "The risks of electronic accounting information systems: An applied study on banks operating in the Gaza Strip," Unpublished master's thesis, Islamic University Library, 2006.

[16] A. Qaoud, "Study and evaluation of the electronic accounting information system in Palestinian companies: An applied study on the joint stock companies in the Gaza governorates," Unpublished master's thesis, Islamic University Library, 2007.

[17] J. Choe and J. Lee, "Factors affecting relationships between the contextual variables and the information characteristics of accounting information systems," Information Processing & Management, vol. 29, no. 4, pp. 471–486, 1993.

[18] B. Grobauer, T. Walloschek, and E. Stocker, "Understanding cloud computing vulnerabilities," IEEE Security & Privacy, vol. 9, no. 2, pp. 50–57, 2011.

[19] M. Zhou, R. Zhang, W. Xie, W. Qian, and A. Zhou, "Security and privacy in cloud computing: A survey," Sixth International Conference on Semantics, Knowledge, and Grid (SKG), IEEE, 2010, pp. 105–112.

[20] K. Latif, F. Shahzad, and M. Imran, "Role of cloud computing in enhancing financial reporting quality in emerging economies," Journal of Financial Technology and Innovation, vol. 15, no. 2, pp. 111–126, 2023.

[21] Gartner, Inc., Cloud computing forecast highlights and trends, Gartner Research, 2023.

[22] A. Gupta and R. Malik, "Leveraging artificial intelligence in cloud computing to enhance accounting workflows," International Journal of Accounting and AI, vol. 10, no. 3, pp. 120–135, 2023.

[23] K. Ahmed and R. Singh, "Enhancing data security in cloud accounting systems through compliance frameworks," Journal of Information Security and Compliance Studies, vol. 12, no. 1, pp. 15–30, 2023.

[24] T. Zhang, L. Chen, and H. Wang, "Exploring the role of edge computing in enhancing real-time data access for cloud-based accounting systems," Future Computing Journal, vol. 5, no. 2, pp. 144–159, 2023.

[25] V. Gupta and S. Jain, "Advanced analytics in cloud accounting: Trends and future directions," Journal of Advanced Accounting Studies, vol. 18, no. 1, pp. 25–48, 2022.

[26] N. Kshetri and J. Voas, "Cloud computing vulnerabilities and mitigation strategies in the financial sector," Communications of the ACM, vol. 66, no. 5, pp. 30–38, 2023.

[27] E. Brynjolfsson and A. McAfee, "The business of AI: Applications and implications for accounting," MIT Sloan Management Review, vol. 63, no. 3, pp. 57–64, 2022.

[28] Research and Markets, "Cloud Computing Industry to Grow from $371.4 Billion in 2020 to $832.1 Billion by 2025, at a CAGR of 17.5%," Aug. 2020. [Online]. Available: https://www.globenewswire.com/news-release/2020/08/21/2081841/0/en/Cloud-Computing-Industry-to-Grow-from-371-4-Billion-in-2020-to-832-1-Billion-by-2025-at-a-CAGR-of-17-5.html. [Accessed: Dec. 14, 2024].

[29] Enterpryze, "5 Benefits of Cloud-Based Accounting for SMEs," 2020. [Online]. Available: https://www.enterpryze.com/post/5-benefits-of-cloud-based-accounting-for-smes. [Accessed: Dec. 14, 2024].

[30] M. Parent and B. H. Reich, "Governing Information Technology Risk," MIT Sloan Management Review, vol. 43, no. 1, pp. 41–49, 2001. S. Gupta and A. Malik, "Artificial Intelligence in Accounting: Ethical Challenges and Legal Perspectives," in Digital Transformation in Accounting and Auditing, A. Perdana, T. Wang, and S. Arifin, Eds. Cham: Springer, 2024, pp. 321–345.

Security Gap in Microservices: A Systematic Literature Review

Nurman Rasyid Panusunan Hutasuhut, Mochamad Gani Amri, Rizal Fathoni Aji
Faculty of Computer Science, Universitas Indonesia, Indonesia, Jakarta

***Abstract*—The growing importance of microservices architecture has raised concerns about its security despite a rise in publications addressing various aspects of microservices. Security issues are particularly critical in microservices due to their complex and distributed nature, which makes them vulnerable to various types of cyber-attacks. This study aims to fill the gap in systematic investigations into microservice security by reviewing current state-of-the-art solutions and models. A total of 487 papers were analyzed, with the final selection refined to 87 relevant articles using a snowball method. This approach ensures that the focus remains on security issues, particularly those identified post-2020. However, there is still a significant lack of dedicated security standards or comprehensive models specifically designed for microservices. Key findings highlight the vulnerabilities of container-based applications, the evolving nature of cyber-attacks, and the critical need for effective access control. Moreover, a substantial knowledge gap exists between academia and industry practitioners, which compounds the challenges of securing microservices. This study emphasizes the need for more focused research on security models and guidelines to address the unique vulnerabilities of microservices and facilitate their secure integration into critical applications across various domains.**

***Keywords*—Microservice security; cyber-attacks; container; security standards; access control**

I. Introduction

Security issues have been rising in many fields, especially in microservices. Even though publications on the topic of microservice are considered high, there is a lack of exploration of the security aspect of microservice [1] [2]. Furthermore, reported from the survey that security is the most ranked issue, followed by availability and scalability. Also, more cyber-attacks are indicated targeting microservices [3]. As indicated in many reports regarding security attacks, microservices have high vulnerabilities to many types of security attacks due to their complex and highly distributed nature.

The trend shows that the security topic in microservice still lacks exploration as opposed to the surge in literature discussing microservice in various topics such as architectural methods and practical application [4], [5]. This trend is supported by findings in grey literature stating that security is the biggest challenge in microservice systems. Moreover, a study reported that this trend is due to the security exploration in microservice is still in the early phase. This seems reasonable since microservice was first popularized by Netflix in 2015 [6].

It is concerning, given the importance of security in the landscape of software development, as microservices continue to gain traction and are adopted across industries. In study [7] reported that application security is the pressing issue in many aspects of microservice such as integration, scalability, API Gateway, etc. Moreover, security vulnerabilities and risks can have far-reaching consequences, making it imperative to address this aspect of microservice comprehensively.

The urgency of addressing security concerns in microservices cannot be overstated, given their increasing integration into mission-critical applications across various domains such as finance, healthcare, and e-commerce. As microservices continue to evolve [7], it is imperative that scholarly discourse on their security keeps pace, ensuring these modern software architectures remain resilient and trustworthy in an ever-changing technological landscape. It is crucial to identify and address the specific challenges inherent in microservice systems and explore how existing techniques can effectively contribute to their security. Despite the growing significance of microservices, there remains a notable gap in systematic investigations at the intersection of security and microservice architectures.

Particularly, many surveys and literature on practitioners that stated microservice is lack of security standard or model [8]. In [9] stated, moreover, more research is needed to deal with microservice complexity, handle security in microservices systems. However, securing MSA is a very challenging task since traditional security concepts cannot be directly applied to MSA [10].

Our findings reveal that more research is needed to (1) deal with microservices complexity at the design level, (2) handle security in microservices systems, and (3) address the monitoring and testing challenges through dedicated solutions.

A. Background

As microservice continues to emerge, paper [11] raised a concern regarding microservice security. This study conducted a systematic literature review on security topics within the microservice realm, by analyzing 290 publications from various sources. The research involved metadata analysis, vector-based markers, and a partitioned overview based on threat models, security, infrastructure, and development approaches. Additionally, recurring concepts like blockchain and service-mesh technologies were explored. The study identified open challenges in microservice security, including issues with data provenance, technology transfer, security-by-design adoption, dedicated attack trees, technological references, migration, global view/control, react and recover techniques, and DevSecOps integration. The lack of established venues for microservices security research was highlighted, emphasizing the need for dedicated platforms to facilitate knowledge

exchange and collaboration among researchers and practitioners. The article concludes by proposing future research directions, suggesting a focus on the grey literature and non-peer-reviewed sources to further enrich the understanding of microservices security.

Microservices have gained popularity since being championed by Netflix in 2015, enabling scalable, modular, and resilient application architectures. Despite these advantages, their distributed nature introduces vulnerabilities, especially in inter-service communication and system integration. Studies highlight common protocols like OAuth 2.0, JWT, API Gateway, and OpenID Connect for managing authentication and authorization, yet challenges persist in implementing robust security across the architecture.

Prior systematic reviews (e.g., [5], [11], [12]) reveal that both academic and grey literature primarily propose mitigation strategies, with limited emphasis on proactive security models. Internal attacks, constituting 60% of all microservice-related breaches (IBM X-Force), remain underexplored compared to external threats, underscoring the need for more focused research.

This paper builds upon prior studies to investigate security challenges in microservices, with a specific focus on developing a comprehensive framework to address vulnerabilities, analyze threat models, and propose practical solutions for academia and industry.

This literature study shows that microservice architecture is well understood since many studies have been published since 2015, and yet there is still a gap that needs to be filled. For instance, it is reported that microservice is one the type of system that has vulnerabilities in security and yet there is no model or framework regarding security in microservice. Also, many studies regarding microservices are specific for a particular use case, and this can also be a challenge in creating a framework within the security realm in microservices. This issue is amplified by the skill gap between academia and industry, as shown by publications in grey and academic literature, where grey literature seems more applicable than academic literature.

Therefore, this study aims to continue the previous study [5] to identify and review the current state-of-the-art security solutions in microservices, specifically in the security models, in terms of developing secure applications. Also analyzed was the proposed study in academic literature. Furthermore, it is particularly important to understand the gap, identify which problems are especially relevant for microservice systems, and determine how existing techniques can contribute to addressing them.

II. Research Methods

To achieve the research goal, we performed a Systematic Literature Review (SLR) in accordance with the guidelines proposed in [14] and the structuring applied in [15]. According to the authors, an SLR is “a means of identifying, evaluating and interpreting all available research relevant to a specific research question, or topic area, or phenomenon of interest" [17]. In addition, this study used the online tool Rayyan [16] to support the screening and analysis of the identified studies.

Based on the literature study, Research Questions (RQ) is formulated and elaborated in the next section. During study, the bulk of papers is obtained from various sources and the RQ’s is used as tool for classifying as well as analyzing the papers.

A. Research Questions

Three research questions were formulated based on the literature study conducted, as explained in the previous section.

- RQ 1. What are the current threats in microservice?

Capturing the scene in terms of security threats as well as security attacks in microservice architecture through academic literature.

- RQ 2. Are there factors or features that are important in securing microservice?

Highlighting the main factors or features that are essential in ensuring security in microservices.

- RQ 3. Are there security standards or models regarding security in microservice architecture?

Based on threats or attacks documented in the literature, through this question is also capturing and characterize the solutions available.

B. Search Process

This study employed four major digital libraries ACM Digital Library, IEEE Xplore, and Scopus. To search in a structure manner, numerous string keywords and combinations are used based on [5], [17] as a query in each digital libraries.

- (“microservice” OR “security”) AND (“microservice”∗ OR "microservice"∗ OR “micro-service”∗) (“MICROSERVICE SECURITY” OR “MICROSERVICES SECURITY”) AND (“CHALLENGE*” OR “PROBLEM*” OR “ISSUE*” OR “SOLUTION*” OR “PROTOCOL*” OR “MECHANISM*” “STRATEGY*”.

C. Snowballing Method

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either Times New Roman or the Symbol font (please, no other font). To create multilevel equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

The snowballing method [14] was implemented in our academic literature review to mitigate the risk of overlooking pertinent studies. This iterative process involved reviewing the references of each study within the initial paper set, incorporating them into the set, and repeating the procedure until no further additions were identified. Both backward and forward snowballing methods were employed. In the backward approach, we scrutinized the references of each selected study, while the forward method involved searching for citations of each selected study. Through this comprehensive snowballing process, additional primary studies were uncovered, resulting in the inclusion of 10 studies from academic literature and 15 from grey literature. Subsequently, the expanded list of articles underwent the application of inclusion/exclusion criteria,

leading to a final paper set comprising 36 primary studies from academic literature and 34 publications from grey literature.

Additionally, to safeguard against the omission of relevant studies, we implemented the snowballing process, following the approach outlined by study [20]. This involved verifying references related to the research object within each selected study. In essence, we actively sought out papers that cited the studies initially chosen. This meticulous approach aimed to ensure a comprehensive and thorough exploration of the existing literature, preventing the oversight of crucial contributions to the field equations should be placed at the center of the line and provided consecutively with equation numbers in parentheses flushed to the right margin, as in Eq. (1). The use of Microsoft Equation Editor or MathType is preferred.

D. Inclusion and Exclusion Criteria

Furthermore, the following are defined inclusion and exclusion criteria to filter relevant studies to be selected for the study.

Inclusion Criteria:

- Primarily from 2019, the latest
- Open access
- Studies related to microservice-based systems
- Studies focusing on security-scope
- Studies related to security scope in microservice
- Not limited to microservice, studies that provide solutions, methodologies, security reports, methodologies, security mechanisms, or other procedures to handle security scope

Exclusion Criteria:

- Studies published prior before 2019
- Short paper (less than three pages)

III. Result and Discussion

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit the use of hard returns to only one return at the end of a paragraph. Do not add any pagination anywhere in the paper. Do not number text heads- the template will do that for you. Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

A. Sources Paper Overview

There are 487 papers are obtained from various resources and using keyword as mentioned in the previous section. By removing duplicates and employed the criteria, there are 87 articles in the list that is relevant with this study. From those 87 articles, most of them are from Journal as much as 64 articles, conference 21 articles, and 1 book.

Furthermore, by employing the snowballing method, the list was refined to a total of 46 articles. This method involved a meticulous double-check of each article to ensure relevance and quality. Initially, a comprehensive list was compiled from various sources, but through the snowballing process, only the most pertinent studies were retained. This approach not only filtered out less relevant papers but also helped identify critical contributions and emerging trends in the field of microservice security.

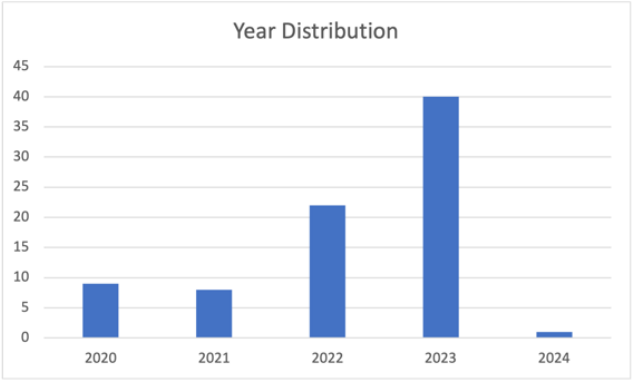


Fig. 1. Comparing publication year regarding microservice security using a certain keyword.

Fig. 1 shows the distribution of the publication year of microservice security publications using the keywords mentioned in the previous section. This study employed publication prior to 2020 at the latest. As shown in Fig. 1, most articles discuss security, particularly in microservices, in 2023. In which the trend gradually increased since 2020. In other words, more people or studies are raising concerns about security on this topic. Consequently, it indicates there are gaps or problems in microservice in terms of security.

In addition, this topic is still lack of exploration as indicates in the Fig. 1, this seems reasonable considering microservice is still quite young and still in an early phase since microservice is introduce in 2015 [18].

Besides the positive trend in terms of publications discussing security, however, as found in [11] reported that there is no event in conferences or journals that is security-oriented. Instead, the analyzed articles are found in publications covering a wide array of subjects, spanning from networking to cloud computing, and in open-access journals like IEEE Access and ACM Queue. Additionally, there isn't a single favored venue that stands out among others; instead, contributors are spread across numerous related platforms.

Fig. 2 shows the word cloud of topics discussed within the articles. Security is the main word followed by Cloud Computing, Computer Architecture, Internet of Things, and another word related security such as Authentication, Monitoring, Protocols, etc.

B. RQ1. Current Threads Towards Microservice

- Container apps brings threats: [18],[19],[21],[22]
- Cyber attacks are always evolving: [23],[22]
- The use of heterogeneity and unnecessary dependencies: [22][11]

Fig. 2. Wordcloud.

Overall, in terms of threats towards microservice, there are three main findings, namely, threats caused by containers, specifically by container third-party apps. Secondly, the approach to cyber-attacks is constantly evolving, making it harder to defend against such attacks. Lastly, it is related to containers, which use unnecessary dependencies.

Interestingly, during 2023 publication raising concern about security in container is increasing. Specifically, by using a third-party app. Beside container well-known advantages, many reports stated that container security has the highest vulnerability in microservice. As in study [24] reported the image in container also contains vulnerabilities. Consequently, employing in a large-scale bring threats and challenges. For instance, the system calls of a running container are a potential source of escalation and brute-force attacks [22].

In study [21] reported that container security issue have the highest vulnerability in microservice. At the same time, in report [22] also highlights that attacks towards container-based is the most attacks beside architectural attacks.

Yet, despite the raising concern in this issue, the most studied tools for defense in terms of container issue is still less studied. Therefore, it is crucial for more exploration in this issued to be done. Especially more study in creating lightweight container that can reduce the security vulnerabilities [22].

By its nature, the container is an isolated and restricted environment that typically looks like a Linux process anatomy. Besides their lightness, containers can be compromised by gaining unauthorized access or deriving vulnerabilities by using images from untrusted sources. Consequently, it is important to take more serious security measures at this level. As well as, more guidance should be explored in terms of containerization.

Furthermore, the use multilayer technology stacks also give a way towards system to be exploited [11], [22] in which still related to containerization. These include data leakage from interactions between the host and container, threats to encryption reliability from differing standards and data format conversions, and covert attacks through the hijacking of software libraries.

Therefore, more exploration throughout container issue is necessary. As many literatures reported that there is concerning vulnerability within container itself.

Microservice typically is hard to maintain. Therefore, many attacks reported are exposing data between each service. As reported in study [25] that the communication intercepted and altered by malicious insiders.

C. RQ2. Important Factors in Securing Microservice

- Access Control is critical in securing microservice: [26], [27]
- The use of third-party container apps: [19], [18]
- Knowledge Gap: [28], [29], [18], [22], [30], [31]
- Distributed nature: [18], [22], [31]
- Other: [13], [30], [31]

There are technical and non-technical factors that are important towards microservice security. Surprisingly, knowledge gap in securing microservice seems concerning. Both from academic and grey literature raising the same issue. Since, microservice is typically hard to maintain and always evolving by its nature [18]. Also, the gap also exists between practitioner and academia [22].

In terms of security, the microservice topic is still in an early phase, as indicated by the number of publications within this context [11]. This seems one of the factors that why there is a knowledge gap. There is still little empirical evidence available on security discussion in microservice [8]. This is due to the no dedicated fragment events on security issue especially architectural software topic [22]. Hence, it is crucial to have dedicated fragment that explore issue security especially in architectural software namely microservice.

As mentioned before, there is also a knowledge gap between practitioners and academia. This might be one of the reasons for the number of publications on this topic. Since, most publications are still discussion in theory of specific problem rather than security implementation. In contrast of in grey literature that more percentage of its publication are discussing security implementation. Moreover, there is also knowledge gap between the organizations and practitioner [18].

Followed by its nature, the complexity of its system is linear with its diversity in security attacks. As experiment studies showed, the larger the architecture, the less likely bugs or errors can be localized, and the less likely the individual microservices can be independently developed and evolved [31]. Many publications reported security solutions with specific problems. The large spectrum of security issues is amplified by microservice in a neighboring area with cloud, edge, fog computing, and containerization, as explained in the previous section. Since microservice is typically implemented in those areas [31].

Due to its complex nature, many publications are problem specific. There is no general approach for security in

Microservice as will be explained further in the next section. As different environment requires different approach in system design including in terms of security. Therefore, the complex nature plays a big factor in security of microservice.

Moreover, in terms of attacks, architectural attacks are less addressed due to their complexity as opposed to software attacks [22][13]. As mentioned before, microservice architectures are always evolving. Consequently, it is required to provide low-level solutions related to hardware, nodes, and operating systems. Also, it is more challenging to provide more comprehensive solutions for microservices due to the granularity of their design [13]. Consequently, compromising a single service may affect the security of the whole system due to the complex configuration and communication between each service [13].

As a result, the configuration of communications between services is crucial for the security. As found through publications in this study, there is growing attention of this issue. Yet there is still lack of security solution implementation as well regarding this issue. In [30] reported that, composition and communication of each service in microservice should have much attention.

D. RQ3. Security Standard or Solutions

a) There are no generic standards: [8], [24], [30], [31]

b) Available solutions or security standards are for specific problems only: [20], [26], [27], [29], [32], [33], [34], [35], [36], [37], [38], [39], [40].

c) Security standards are a less studied topic: [8], [22], [31], [13], [21], [17]

As discussed in the previous sections, the main reason why microservice is hard to maintain is its nature. Microservice is always evolving. Consequently, the attack toward it is linear. Thus, it is hard to maintain such a system. In addition, there is also some knowledge on many levels of securing microservices.

As a result, it is hard to have or to establish a general guideline standard for microservice security, while at the same time, it is also crucial to have such a standard. As study in [8] reported, practitioners complain that there are no clear security standards in developing microservices. Despite many publications calling for such standards, it is still hard to find one. Fragments event that focuses on architectural security issues.

The term standard or model discussed in this section can be interpreted as threat, mitigation, monitoring models, or etc. Surprisingly, it is hard finding such publications regarding this context. This finding is correlated with finding explained in the previous section which this topic of study is still in an early phase. This might be understandable, since Microservice terms is popularized by Netflix in 2015.

Furthermore, the study also conjectures that this lack of usage of generic threat models since the majority of research done on microservice security comes from the software (engineering, languages) side of the field rather than from the side of security, which advocates for a security-by-design approach. This is correlated with findings in the previous section, which is there are no security event outlets available in conferences or journals that focus on this topic.

In terms of mitigation model, there is still lack of security approaches address applications across the full stack. Therefore, it is an opportunity to be explore more regarding this topic.

It is the same case in terms of threat models. There is still lack of model regarding this topic. Even though, with such model is proven useful in the identification of attack types and strategic counter measures. Not to mention the complexity of microservice nature, this guideline is crucial in tackling the multifaceted attack surface of microservices architectures.

There are several security standards or models discussed in the literature. Namely Trust Models and GDPR guidelines. The zero Trust model consists of several model types, such as socio-based, composition-based, control-based, and zero-trust-based, in which those models focus on how users, applications, devices, or packets establish trust with each other. On the other hand, the GDPR guideline is more focused on protecting the privacy of the users. In terms of GDPR guidelines, it is surprising that only one publication claimed to be a GDPR guideline to protect the privacy of users. Especially if the microservice is a cloud-based system.

Interestingly, one paper stated that the diversity of attacks is due to the adoption of zero-trust models. Since the model assumed to afford no default trust for entities within the system.

IV. Conclusion

The exploration of microservice security reveals significant challenges and gaps in current understanding and practices, particularly highlighting the substantial security risks posed by containerization, evolving cyber-attacks, and unnecessary dependencies. Containers, despite their advantages, are prone to vulnerabilities such as unauthorized access and the use of untrusted images, necessitating stringent security measures to mitigate potential exploits and data breaches. Additionally, the distributed and evolving nature of microservices increases their complexity and the potential for security breaches, emphasizing the need for robust access control mechanisms and careful use of third-party container apps. The study also underscores a significant knowledge gap between academia and industry practitioners and the lack of comprehensive, adaptable security standards. This absence of cohesive security guidelines, coupled with the continuous evolution of microservices, presents substantial challenges in maintaining security.

As the study reported, Containerized applications have been getting traction lately in terms of security vulnerabilities. More exploration of this topic is crucial. Containers are becoming more popular because of their portability and efficiency in development and production. Moreover, increased collaboration and focused studies are essential to develop and disseminate effective security frameworks and standards, ensuring robust defenses against an ever-changing threat landscape.

REFERENCES

[1] P. Di Francesco, I. Malavolta, and P. Lago, "Research on Architecting Microservices: Trends, Focus, and Potential for Industrial Adoption," 2017 IEEE International Conference on Software Architecture (ICSA), pp. 21–30, May 2017, doi: 10.1109/ICSA.2017.24.

[2] M. T. Hinkley C, Snyder A, "Application Security Statistics Report. The evolution of the secure software lifecycle," 2018.

[3] N. Dragoni et al., "Microservices: Yesterday, today, and tomorrow," Present and Ulterior Software Engineering, pp. 195–216, Nov. 2017, doi: 10.1007/978-3-319-67425-4_12/COVER.

[4] A. Pereira-Vale, E. B. Fernandez, R. Monge, H. Astudillo, and G. Márquez, "Security in microservice-based systems: A Multivocal literature review," Computers and Security, vol. 103, p. 102200, 2021, doi: 10.1016/j.cose.2021.102200.

[5] C. Pahl and P. Jamshidi, "Microservices: A systematic mapping study," CLOSER 2016 - Proceedings of the 6th International Conference on Cloud Computing and Services Science, vol. 1, pp. 137–146, 2016, doi: 10.5220/0005785501370146.

[6] V. Bushong et al., "On Microservice Analysis and Architecture Evolution: A Systematic Mapping Study," Applied Sciences, vol. 11, no. 17, 2021, doi: 10.3390/app11177856.

[7] M. Waseem, P. Liang, M. Shahin, A. Di Salle, and G. Márquez, "Design, monitoring, and testing of microservices systems: The practitioners' perspective," Journal of Systems and Software, vol. 182, p. 111061, Dec. 2021, doi: 10.1016/J.JSS.2021.111061.

[8] M. Waseem, P. Liang, A. Ahmad, M. Shahin, A. A. Khan, and G. Márquez, "Decision models for selecting patterns and strategies in microservices systems and their evaluation by practitioners," pp. 135–144, May 2022, doi: 10.1145/3510457.3513079.

[9] P. Billawa, A. B. Tukaram, N. E. D. Ferreyra, J.-P. Steghöfer, R. Scandariato, and G. Simhandl, "SoK: Security of Microservice Applications: A Practitioners' Perspective on Challenges and Best Practices," ACM International Conference Proceeding Series, p. 10, Feb. 2022, doi: 10.1145/3538969.3538986.

[10] D. Berardi, S. Giallorenzo, A. Melis, M. Prandini, J. Mauro, and F. Montesi, "Microservice security: a systematic literature review," PeerJ Computer Science, vol. 7, pp. 1–66, 2022, doi: 10.7717/PEERJ-CS.779.

[11] M. G. de Almeida and E. D. Canedo, "Authentication and Authorization in Microservices Architecture: A Systematic Literature Review," Applied Sciences (Switzerland), vol. 12, no. 6, Mar. 2022, doi: 10.3390/APP12063023.

[12] A. Hannousse and S. Yahiouche, "Securing microservices and microservice architectures: A systematic mapping study," Computer Science Review, vol. 41, p. 100415, 2021, doi: https://doi.org/10.1016/j.cosrev.2021.100415.

[13] C. Wohlin, "Guidelines for Snowballing in Systematic Literature Studies and a Replication in Software Engineering," in Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, in EASE '14. New York, NY, USA: Association for Computing Machinery, 2014. doi: 10.1145/2601248.2601268.

[14] C. Wohlin, "Second-Generation Systematic Literature Studies using Snowballing", doi: 10.1145/2915970.2916006.

[15] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, "Rayyan-\--a web and mobile app for systematic reviews," Systematic Reviews, vol. 5, no. 1, p. 210, 2016, doi: 10.1186/s13643-016-0384-4.

[16] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," in International Conference on Evaluation \& Assessment in Software Engineering, 2008.

[17] A. Rezaei Nasab, M. Shahin, S. A. Hoseyni Raviz, P. Liang, A. Mashmool, and V. Lenarduzzi, "An empirical study of security practices for microservices systems," Journal of Systems and Software, vol. 198, p. 111563, Apr. 2023, doi: 10.1016/j.jss.2022.111563.

[18] S. Sultan, I. Ahmad, and T. Dimitriou, "Container Security: Issues, Challenges, and the Road Ahead," IEEE Access, vol. 7, pp. 52976–52996, 2019, doi: 10.1109/ACCESS.2019.2911732.

[19] H. Jin, Z. Li, D. Zou, and B. Yuan, "DSEOM: A Framework for Dynamic Security Evaluation and Optimization of MTD in Container-based Cloud," IEEE Trans. Dependable and Secure Comput., pp. 1–1, 2019, doi: 10.1109/TDSC.2019.2916666.

[20] A. J. Cabrera-Gutiérrez, E. Castillo, A. Escobar-Molero, J. Cruz-Cozar, D. P. Morales, and L. Parrilla, "Blockchain-Based Services Implemented in a Microservices Architecture Using a Trusted Platform Module Applied to Electric Vehicle Charging Stations," Energies, vol. 16, no. 11, p. 4285, May 2023, doi: 10.3390/en16114285.

[21] M. S. Rahaman, A. Islam, T. Cerny, and S. Hutton, "Static-Analysis-Based Solutions to Security Challenges in Cloud-Native Systems: Systematic Mapping Study," Sensors, vol. 23, no. 4, p. 1755, Feb. 2023, doi: 10.3390/s23041755.

[22] Z. Li, H. Jin, D. Zou, and B. Yuan, "Exploring New Opportunities to Defeat Low-Rate DDoS Attack in Container-Based Cloud Environment," IEEE Trans. Parallel Distrib. Syst., vol. 31, no. 3, pp. 695–706, Mar. 2020, doi: 10.1109/TPDS.2019.2942591.

[23] F. Ying, S. Zhao, and H. Deng, "Microservice Security Framework for IoT by Mimic Defense Mechanism," Sensors, vol. 22, no. 6, p. 2418, Mar. 2022, doi: 10.3390/s22062418.

[24] C. Zhong, H. Zhang, C. Li, H. Huang, and D. Feitosa, "On measuring coupling between microservices," Journal of Systems and Software, vol. 200, p. 111670, Jun. 2023, doi: 10.1016/J.JSS.2023.111670.

[25] S. Xu et al., "Log2Policy: An Approach to Generate Fine-Grained Access Control Rules for Microservices from Scratch," in Annual Computer Security Applications Conference, Austin TX USA: ACM, Dec. 2023, pp. 229–240. doi: 10.1145/3627106.3627137.

[26] Z. Zaheer, H. Chang, S. Mukherjee, and J. Van der Merwe, "eZTrust: Network-Independent Zero-Trust Perimeterization for Microservices," in Proceedings of the 2019 ACM Symposium on SDN Research, in SOSR '19. New York, NY, USA: Association for Computing Machinery, Apr. 2019, pp. 49–61. doi: 10.1145/3314148.3314349.

[27] T. Cerny et al., "On Code Analysis Opportunities and Challenges for Enterprise Systems and Microservices," IEEE Access, vol. 8, pp. 159449–159470, 2020, doi: 10.1109/ACCESS.2020.3019985.

[28] I. Araujo, N. Antunes, and M. Vieira, "Evaluation of Machine Learning for Intrusion Detection in Microservice Applications," in 12th Latin-American Symposium on Dependable and Secure Computing, La Paz Bolivia: ACM, Oct. 2023, pp. 126–135. doi: 10.1145/3615366.3615375.

[29] Z. Lu, D. T. Delaney, and D. Lillis, "A Survey on Microservices Trust Models for Open Systems," IEEE Access, vol. 11, pp. 28840–28855, 2023, doi: 10.1109/ACCESS.2023.3260147.

[30] D. Berardi, S. Giallorenzo, J. Mauro, A. Melis, F. Montesi, and M. Prandini, "Microservice security: a systematic literature review," PeerJ Computer Science, vol. 7, p. e779, Jan. 2022, doi: 10.7717/peerj-cs.779.

[31] C. Meadows, S. Hounsinou, T. Wood, and G. Bloom, "Sidecar-based Path-aware Security for Microservices," in Proceedings of the 28th ACM Symposium on Access Control Models and Technologies, Trento Italy: ACM, May 2023, pp. 157–162. doi: 10.1145/3589608.3594742.

[32] A. Bambhore Tukaram, S. Schneider, N. E. Díaz Ferreyra, G. Simhandl, U. Zdun, and R. Scandariato, "Towards a Security Benchmark for the Architectural Design of Microservice Applications," in Proceedings of the 17th International Conference on Availability, Reliability and Security, in ARES '22. New York, NY, USA: Association for Computing Machinery, Aug. 2022, pp. 1–7. doi: 10.1145/3538969.3543807.

[33] I. Araujo, N. Antunes, and M. Vieira, "Intrusion Detection and Tolerance for Microservice Applications," in Proceedings of the 12th Latin-American Symposium on Dependable and Secure Computing, in LADC '23. New York, NY, USA: Association for Computing Machinery, Oct. 2023, pp. 176–181. doi: 10.1145/3615366.3622794.

[34] A. Chatterjee, M. W. Gerdes, P. Khatiwada, and A. Prinz, "SFTSDH: Applying Spring Security Framework With TSD-Based OAuth2 to Protect Microservice Architecture APIs," IEEE Access, vol. 10, pp. 41914–41934, 2022, doi: 10.1109/ACCESS.2022.3165548.

[35] W. Wang, A. Benea, and F. Ivancic, "Zero-Config Fuzzing for Microservices," in 2023 38th IEEE/ACM International Conference on Automated Software Engineering (ASE), Luxembourg, Luxembourg: IEEE, Sep. 2023, pp. 1840–1845. doi: 10.1109/ASE56229.2023.00036.

[36] B. G. Kim, Y.-S. Cho, S.-H. Kim, H. Kim, and S. S. Woo, "A Security Analysis of Blockchain-Based Did Services," IEEE Access, vol. 9, pp. 22894–22913, 2021, doi: 10.1109/ACCESS.2021.3054887.

[37] S. Tang, Z. Wang, J. Dong, and Y. Ma, "Blockchain-Enabled Social Security Services Using Smart Contracts," IEEE Access, vol. 10, pp. 73857–73870, 2022, doi: 10.1109/ACCESS.2022.3190963.

[38] M. Jin et al., "An Anomaly Detection Algorithm for Microservice Architecture Based on Robust Principal Component Analysis," IEEE Access, vol. 8, pp. 226397–226408, 2020, doi: 10.1109/ACCESS.2020.3044610.

[39] M. Anisetti, C. A. Ardagna, and N. Bena, "Multi-Dimensional Certification of Modern Distributed Systems," IEEE Trans. Serv. Comput., pp. 1–14, 2022, doi: 10.1109/TSC.2022.3195071.

[40] A. Hannousse and S. Yahiouche, "Securing microservices and microservice architectures: A systematic mapping study," Computer Science Review, vol. 41, p. 100415, Aug. 2021, doi: 10.1016/j.cosrev.2021.100415.

New Knowledge Management Model: Enhancing Knowledge Creation with Zack Gap, Brand Equity, and Data Mining in the Sports Business

Fransiska Prihatini Sihotang[1], Ermatita[2]*, Dian Palupi Rini[3], Samsuryadi[4]
Doctoral Program in Engineering Science, Universitas Sriwijaya, Palembang, Indonesia[1]
Faculty of Computer Science, Universitas Sriwijaya, Palembang, Indonesia[2, 3, 4]
Faculty of Computer Science and Engineering, Universitas Multi Data Palembang, Palembang, Indonesia[1]

***Abstract*—This research improves Socialization, Externalization, Combination, and Internalization (SECI) knowledge management model by combining it with Zack's knowledge gap model, brand equity concept, and data mining. Zack's model is incorporated into the SECI model to identify the gap between the knowledge in the organization and the knowledge that the organization should possess. We add the data mining techniques to determine that knowledge gap. The uniqueness of this study lies in the externalization and combination of the SECI model. In the externalization, "what the firm must know" is added; for that, we compile the questionnaire by adopting brand equity and distributing it to the athletes. In the combination, "what the firm knows" is added; we use a database already owned by sports business management. The modifications resulting from both models with data mining in this study were carried out to develop a new knowledge management model in the sports business sector. This new model will be valuable knowledge for sports business management to build strategies and increase their competitiveness in the sports market. In addition, other service business fields besides sports can also apply this new model to improve their knowledge management, which they can then use to improve their marketing strategies.**

Keywords*—SECI model; zack model; data mining; brand equity; sport business

I. Introduction

The knowledge creation process is critical to increasing business value and ensuring the success of a business, including sports centers [1], [2]. According to Rot and Sobinska research [3], digital technology can improve knowledge management. Thus, similar to other industries [4], the sports service sector must manage knowledge effectively to boost its competitiveness. Numerous studies have shown that well-organized knowledge significantly enhances a company's innovation [5]. Efficient knowledge management strengthens a company's ability to innovate by ensuring that relevant and accurate information is readily accessible for decision-making. Efficient knowledge management will also create new products, services, and business processes, which give the company a competitive advantage.

The SECI model is the commonly used knowledge management model across various fields. However, it has faced criticism from several researchers. The model, often seen as overly simplistic, addresses the socialization and externalization processes inadequately and overlooks cultural differences among members of an organization. It has been critiqued for its individualistic focus, which limits its ability to represent interactions and interdependencies effectively [6]. Other studies indicate that this model centers only on converting human knowledge, neglecting knowledge stored in databases, external sources, and variations in organizational knowledge. This limitation can lead to a knowledge gap between what the organization currently knows and what it ideally should know [7], [8].

Data mining techniques reveal hidden patterns and relationships within large datasets, supporting more informed and strategic decision-making. Data mining is playing a growing role in data analysis, including applications like predicting stock prices [9], assessing disease risks [10], analyzing customer churn [11], segmenting customers [12], measuring customer satisfaction [13]–[16], forecast customer behavior, streamline supply chains, and improve customer relationship management. [7], [14], [17].

Data mining is also applied in sports [18], such as for predicting injuries in soccer players [19], identifying key attributes and metrics that impact NBA player salaries and performance [20], and detecting tactical patterns through market basket analysis in beach volleyball games [21]. The many studies related to data mining and sports show that sports attract the world's attention. However, many find identifying the optimal data mining method challenging due to the wide variety of processed datasets [17].

This study aims to build on existing research by integrating the SECI model, Zack's knowledge gap model, data mining techniques, and a brand equity questionnaire to create a new knowledge management model specifically for the sports business. To identify the knowledge gaps, we will use two data sources: 1. the database that management already has (what the firm knows), we call it the secondary dataset; 2. a questionnaire containing 20 brand equity questions distributed to athletes/ customers (what the firm must know), we call it the primary dataset.

In addition to the belief that there is a knowledge gap between the two datasets, Rungtrakulchai [22] said that adding a brand equity questionnaire is also important because research has shown that marketing practitioners need to understand the

level of brand equity in a particular market area. Other research [23] also said that marketing practitioners can obtain this from customer/consumer responses to the services provided. By merging these models, we aim to provide valuable insights for sports business management, helping organizations to develop effective strategies and improve their competitiveness.

This article consists of several sections. Section II includes a literature review that examines relevant theories and previous research related to this study. Section III provides the research methodology, explaining how we combined various models to create a new knowledge management model. Section IV represents the results and discussion, detailing the outcomes of this combination into the new knowledge management model. Finally, Section V offers the conclusion.

II. Literature Review

A. SECI Model

The SECI model, an adaptable knowledge management framework, encompasses four critical processes: Socialization, Externalization, Combination, and Internalization. Its adaptability is evident in its successful application across various sectors, from cooperatives [24] and universities [25] to software development [26] and even in the design of a knowledge management system for trader education in the marketplace [27]. Miao et al., in their research [28], said that this model's ability to acquire and align knowledge with business strategy has also found a place in the banking sector.

Fig. 1 illustrates the SECI model. The SECI model has been widely adopted and directly applied in shaping knowledge management across various organizations [6] or by first modifying or developing the model [29]. However, some researchers have critiqued the SECI model from different perspectives. The socialization and externalization processes in this model are often ineffective due to limitations on the freedom to exchange ideas at the industrial level. Additionally, some argue that the SECI model is too simplistic, neglects the role of context, and does not adequately account for the cultural differences among organizational members. Another criticism is that the model portrays knowledge transfer as a linear process, even though this may not reflect real-world situations. Some also contend that the model's fundamental structure is individualistic, making it challenging to represent interactions and interdependencies effectively [6].

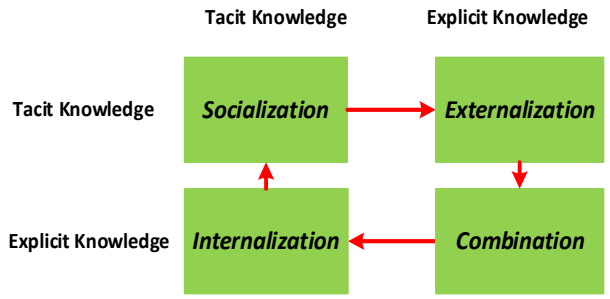


Fig. 1. SECI model [25], [30].

Other researchers have highlighted additional weaknesses of the SECI model, including those pointed out by Almuayqil, who argued that the model focuses solely on converting human knowledge while overlooking knowledge stored in databases and other technological resources. Meanwhile, Yao noted that the SECI model fails to account for external knowledge inputs and the variations in knowledge within the organization [7]. To address these shortcomings, we incorporate the knowledge gap analysis from Zack's model into the externalization and combination phases of the SECI model.

B. Zack gap model

Aligning business strategy with knowledge is crucial for an organization [28]. Learning and knowledge acquisition are necessary to sustain and enhance a competitive advantage. Fig. 2 illustrates the knowledge and strategic gap in Zack's model. Aligning a company's business strategy with its knowledge resources is crucial, as it plays a critical role in driving innovation and improving business performance. This approach has become a primary focus for companies seeking success in today's competitive environment [28].

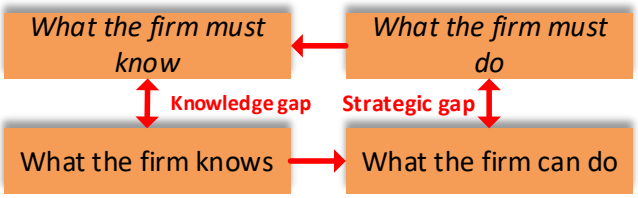


Fig. 2. Zack gap model [8], [28].

Using Zack's model, Gap analysis compares an organization's existing knowledge with its strategic needs. Knowledge and strategy are viewed as opposing poles: what is known or done and what should be known or done. By examining these points, gaps in both organizational knowledge and strategy can be identified. In this study, we focus on the knowledge gap from Zack's model and integrate it with the SECI model. In this research, we will find the knowledge gap by comparing the database that management already has (what the firm knows) with the questionnaire (what the firm must know) distributed to athletes (customers).

C. Data Mining in Knowledge Management

Data mining methods generally consist of classification, association, and clustering. Classification is a technique in data mining to group data based on the data's relationship to sample data [31]. The clustering method is a process of grouping data objects similar to each other into the same cluster and different from objects in other clusters [32]. Based on a literature study conducted by Ngai, the clustering method is mainly used for customer identification, while the classification method is mainly used for customer acquisition and retention [17]. Meanwhile, association in data mining is a technique for obtaining hidden relationship patterns between several items in a dataset [33].

D. Brand Equity

Brand equity is an added value imposed on a particular product or service that reflects consumers' thoughts, feelings, and actions in adapting the product or service and influences customer value perception [22]. It is also related to price, market share, and brand profitability [34]. Marketing practitioners are essential to understand the level of brand equity in a particular

market area [22], which is obtained from customer/consumer responses to the services provided [23]. As shown in Fig. 3, brand equity consists of Brand Awareness (BA), Perceived Quality (PQ), Brand Association (BA), Brand Loyalty (BL), and Decision (D) [23].

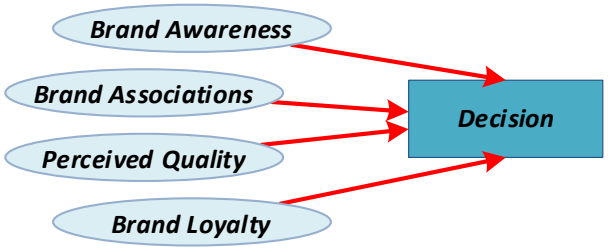


Fig. 3. Brand equity model [22], [35].

III. Research Methodology

This research combines the SECI model, knowledge gap analysis from the Zack model, data mining techniques, and a brand equity. The knowledge gap analysis from the Zack model was added to the SECI model in the externalization and combination section. Externalization of the SECI model is the form of changing knowledge from tacit to explicit, and we match it with "what the firm must know" in the Zack model. The combination of the SECI model is knowledge in an explicit form that already exists, is developed, and disseminated through various media that are more systematic. We match it with " what the firm knows" in the Zack model.

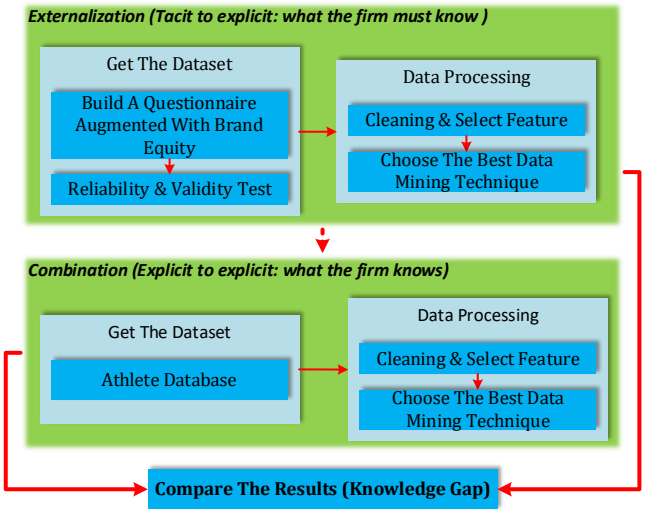


Fig. 4. Combining two models into a new model.

Fig. 4 displays how we match and combine these two models into a new model. To use this new model, we need to acquire a dataset. Since this model will perform knowledge gap analysis, it requires two types of datasets: one that the organization already owns and another that contains knowledge about what the organization should know.

Data cleaning and feature selection need to be performed on both datasets. Next, it is essential to test various data mining techniques and algorithms to identify the one best suited to both datasets. The choice of data mining techniques should be aligned with management needs, depending on the desired outcomes. After selecting the most effective algorithm for both datasets, it is essential to apply this algorithm to both datasets and analyze the results. To finalize this model, we must compare the data mining outcomes of the two datasets to identify any knowledge gaps. This knowledge gap can serve as valuable information for management, helping to enhance decision-making, develop strategies, and improve competitiveness in the sports market.

A. Get the Dataset

The new knowledge management model for sports business developed in this study requires two data sources, and then knowledge gaps will be sought from these data sources. We call these two data sources secondary and primary datasets. It is essential to ensure that the data sources are credible so that they can produce correct knowledge. In this research, we use a case study in taekwondo sports. So, we collected the secondary datasets from the Taekwondo Indonesia Integrated System (TIIS) application in the South Sumatra Province, which the government uses to process athlete data. Thus, this database is a credible knowledge base for this study. From the TIIS databases, we acquired sixty files in Microsoft Excel format and merged them into a single file for further analysis.

Based on Zack's knowledge gap model, companies' current knowledge (what the firm knows) needs to be compared with the knowledge they do not yet have but should have (what the firm must know). Therefore, it is necessary to discover new knowledge collected using research questionnaires. Therefore, the relationship between the database and the questionnaire is to assess knowledge gaps.

We obtained the primary dataset by distributing questionnaires to users (athletes) of Taekwondo services in South Sumatra Province, Indonesia. The questionnaire was distributed for two months as a Google form since the respondents were spread across many cities. To ensure the accuracy of the research data, parents or guardians filled out the questionnaire for athletes under the age of sixteen. For athletes over sixteen, they filled out the questionnaire directly. However, information related to brand equity was not obtained in the secondary dataset. Therefore, when compiling the questionnaire for the primary dataset, besides asking for the same information as in the secondary dataset, 20 questions related to the Brand Equity of the Club/Dojang where the athletes train were added. The questionnaire consists of Brand Awareness (BA), Perceived Quality (PQ), Brand Association (BA), Brand Loyalty (BL), and Decision (D), as shown in Table I.

The questionnaire was meticulously adapted to the case study, focusing on the sports service business, mainly taekwondo, in South Sumatra Province, Indonesia. Each statement item in the questionnaire was crafted based on interviews with Dojang owners and validated by the Head of the Achievement Division of Taekwondo Indonesia, South Sumatra Province. This rigorous process was undertaken to ensure the

credibility of the data source. Each questionnaire item was designed to be answered by respondents by selecting from the options provided: Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), and Strongly Disagree (SD). Before distributing the questionnaire, a pre-test was conducted to assess its reliability and validity. A small sample of 30 athletes was selected for this pre-test. The results of the pre-test stated that the questionnaire was valid and reliable so that the distribution of the questionnaire could be continued.

TABLE I. QUESTIONNAIRE ITEMS

Code	Questionnaire Statement
BA1	When asked to name a place to practice taekwondo, the first thing I always think of is the club/dojang where I practice.
BA2	Talking about taekwondo reminds me of the Club/Dojang where I train.
BA3	The Club/Dojang where I train is my first choice when I want to start training taekwondo.
PQ1	Sabeum/ trainer at the Club/Dojang where I train is licensed and superior.
PQ2	I feel comfortable in the Club/Dojang where I practice because it is indoors.
PQ3	The supporting facilities at the Club/Dojang where I train are complete and meet standards.
PQ4	The training patterns at the Club/Dojang where I train are varied and not monotonous.
PQ5	The training schedule at the Club/Dojang where I train is relatively routine.
BAs1	My Club/Dojang is the best compared to the others.
BAs2	The training fees at my Club/Dojang are cheap.
BAs3	The training fees at My Club/Dojang are in line with the quality and service.
BAs4	The club/Dojang where I train is easily accessible from my house.
BL1	I will always choose the Club/Dojang where I am currently training as long as I train.
BL2	I will not move to another Club/Dojang.
BL3	I would recommend My Club/Dojang.
D1	I looked for as much information as possible about the Club/Dojang before joining.
D2	I looked for as much information as possible about the Club/Dojang before joining.
D3	I compared the training costs of the Club/Dojang where I train with others before joining.
D4	I consider the distance of the Club/Dojang where I train before joining.
D5	I decided to stay with the Club/Dojang where I train because it meets my expectations.

Following the distribution of the questionnaire, 1468 rows of athlete data were obtained, of which 1348 rows were deemed usable; the rest were damaged. To ensure validity and reliability, we do tests using the SPSS application. The results showed that the calculated r value for each indicator ranged from 0.534 to 0.893, or far above the r table value. Reliability testing was conducted by examining Cronbach's alpha value for each indicator. The research questionnaire is highly reliable if Cronbach's alpha is between 0.9 and 1. Based on the test results, Cronbach's alpha value for each indicator in this research questionnaire was more than 0.9. Based on this value [36], [37], it can be said that this research questionnaire is valid with a confidence level of 99%, and the reliability is very high.

B. Cleaning and Feature Selection

The cleaning process is essential to ensure the dataset is suitable for research. This process includes checking each column for missing data, converting the data types of specific columns to numeric, and removing unnecessary columns. Many studies have highlighted the importance of feature selection as a crucial preprocessing step [17]. Choosing the right features can significantly enhance the performance of machine learning algorithms by reducing data dimensionality, which helps eliminate irrelevant or redundant information. We applied this cleaning procedure to both datasets. However, we excluded the primary dataset's feature selection for 20 brand equity questions because we have to use it all.

We conduct three methods of feature selection in this study: Information Gain (IG), Chi-Square, and Recursive Feature Elimination (RFE). Each of these methods provides a distinct approach to identifying the most relevant features for clustering. The Information Gain measures entropy reduction, the Chi-Square method evaluates the independence of features, and the RFE method recursively removes the least important features. Combining these methods ensures a thorough selection process and improves the data quality used for clustering. The results of the feature selection process in this study are shown in Table II.

TABLE II. FEATURE SELECTION BY VARIOUS METHODS

Feature	IG	Chi-Square	P-Value	RFE
Classification	0.554438	1.424.732.394	4,20E-30	1
ClubName	0.108146	168.885.790	2,12E-31	2
ClubID	0.114757	455.513.151	1,22E-93	3
Age	0.007043	4.471.729	1,07E+05	4
Belt	0.032379	23.657.274	7,29E+00	5
Subdistric	0.049144	36.273.625	1,33E-02	6
Sex	0.027673	44.655.043	2,01E-04	7
BeltID	0.019311	36.009.465	1,52E-02	8
Disctric	0.036292	33.825.733	4,52E-02	9
Distance	0.000000	3.319.825	1,90E+05	10
AgeGroup	0.005404	0.179834	9,14E+05	11
City	0.028890	17.123.501	1,91E+02	12

C. Choose the Best Data Mining Technique

In our case study, taekwondo club management requires grouping its athletes based on their abilities, interests, residence, belts, and other club members' characteristics. So, we use the clustering algorithm. Applying the clustering algorithm in the taekwondo sector helps establish a more organized, efficient, and responsive structure to meet members' needs, ultimately enhancing training programs' quality and sustainability and supporting effective marketing strategy development.

This research compares the K-means and K-medoid clustering algorithms to find the most suitable one. We compare these two algorithms because there have been many studies that prove the accuracy of both in clustering [38]–[40]. In order to determine the optimal number of clusters of this algorithm, we use the Elbow Method, the Silhouette Coefficient, the Calinski-Harabasz Index, the Davies-Bouldin Index, the Bayesian Information Criterion (BIC), and Akaike Information Criterion

(AIC). Based on various tests that we conducted, we know that for both datasets that we have, the best algorithm to apply to both datasets is K-Medoids, with 2 clusters. Table III displays the results of each method in determining the optimum number of clusters. The tests using various methods show that the most suitable algorithm for both datasets is K-Medoids, with an optimal number of two clusters.

TABLE III. DETERMINING THE OPTIMAL CLUSTERS

Methods	Cluster number with K-Means		Cluster number with K-Medoids	
	Secondary Dataset	Primary Dataset	Secondary Dataset	Primary Dataset
Elbow Method	2	2	2	2
Silhouette Score	2	2	2	2
Calinski-Harabasz Index	2	2	8	8
Davies-Bouldin Index	2	2	2	2
BIC	7	8	6	6
AIC	7	8	6	6

IV. RESULT AND DISCUSSION

A. Brand Equity Result

Based on what has been explained previously, we added 20 questions related to Brand Equity to our research questionnaire. Respondents can answer each question by selecting the options provided: Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), and Strongly Disagree (SD). We separate the brand equity questionnaire results from other data (which is the same as secondary data) and present them in Table IV. This result is also a knowledge gap that we found in our case study on this research.

TABLE IV. BRAND EQUITY RESULT

Questionnaire Code	SA	A	N	D	SD
BA1	71%	23%	6%	0%	0%
BA2	66%	27%	7%	0%	0%
BA3	64%	28%	8%	0%	0%
PQ1	62%	29%	9%	0%	0%
PQ2	44%	18%	32%	4%	2%
PQ3	55%	32%	13%	0%	0%
PQ4	60%	32%	8%	0%	0%
PQ5	60%	33%	7%	0%	0%
BAs1	55%	27%	18%	0%	0%
BAs2	62%	31%	7%	0%	0%
BAs3	59%	33%	8%	0%	0%
BAs4	62%	31%	7%	0%	0%
BL1	61%	31%	8%	0%	0%
BL2	61%	29%	10%	0%	0%
BL3	60%	33%	7%	0%	0%
D1	53%	33%	14%	0%	0%
D2	54%	32%	13%	1%	0%
D3	45%	27%	23%	5%	0%
D4	54%	32%	12%	2%	0%
D5	63%	29%	8%	0%	0%

These results are precious for sports business management because they can help them create marketing strategies. For example, in question item PQ3, "The supporting facilities at the Club/Dojang where I train are complete and meet standards," the results of the questionnaire distribution show that the supporting facilities at the Club/Dojang have not been fully met and do not meet standards. Thus, sports business management can improve equipment and supporting facilities to provide comfort for its members in training and retain those members.

Then, the BL3 question item "I would recommend My Club/Dojang" shows that most athletes would recommend the Club/Dojang where they train others. Thus, sports business management can approach and offer various promotions to their members to maintain good relationships and retain these members. Then, the BAs3 question item “The training fees at My Club/Dojang align with the quality and service” revealed that not all respondents were satisfied with the quality of service compared to the costs. This is very valuable for management, as it allows them to improve the quality of their services in the future, commensurate with the costs incurred by their members.

B. Data Mining Result

The tests using various methods show that the most suitable algorithm for both datasets in the Taekwondo case is K-Medoids, with an optimal number of two clusters. Therefore, we processed our datasets with the selected algorithm using two clusters for both datasets. Fig. 5 shows the results of the clustering we performed on both datasets, which shows that there are differences between the two (knowledge gap). The differences align with the research objective of identifying knowledge gaps. The proposed algorithm is flexible and not more suitable for one data type than another.

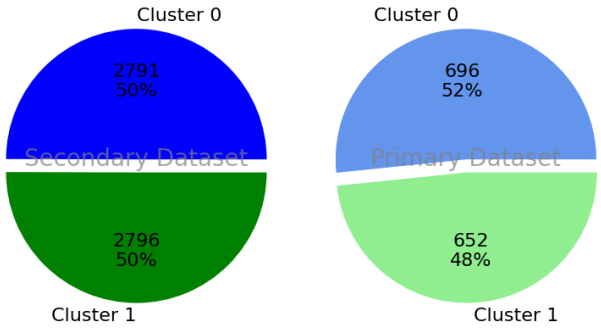


Fig. 5. The clustering result of both dataset.

Based on the results displayed in the clustering process using both datasets, although only a little, there is still a difference (knowledge gap) between the secondary and the primary datasets. The difference lies in the percentage of the secondary dataset, which is exactly 50:50, while for the primary dataset, it is 52:48. Although the difference is only 2%, it is essential for sports business management. Knowing this difference can help them make more careful decisions regarding policies to increase their business's competitiveness.

Based on the differences between the two datasets (knowledge gap), we analyze them based on several categories: athlete gender, athlete age categories, athlete belt, and the distance of the Club/Dojang where the athlete trains from their respective homes. These results are helpful for sports business management in determining their future customer targets.

Fig. 6 displays the distribution of athletes' gender on both datasets. The graph shows a difference in the percentages of the two datasets. In the secondary dataset, in cluster 0, the ratio of

female and male athletes is 41:59, while in cluster 1, it is 37:63. In the primary dataset, the ratio is 36:64 for cluster 0 and 42:58 for cluster 1. Ensuring the gender distribution of Club/Dojang members is very important as it will influence promotion policies and the approach towards members/ potential members. These results align with previous research that discussed the influence of gender on purchasing decisions [41].

Fig. 6. Athletes gender distribution.

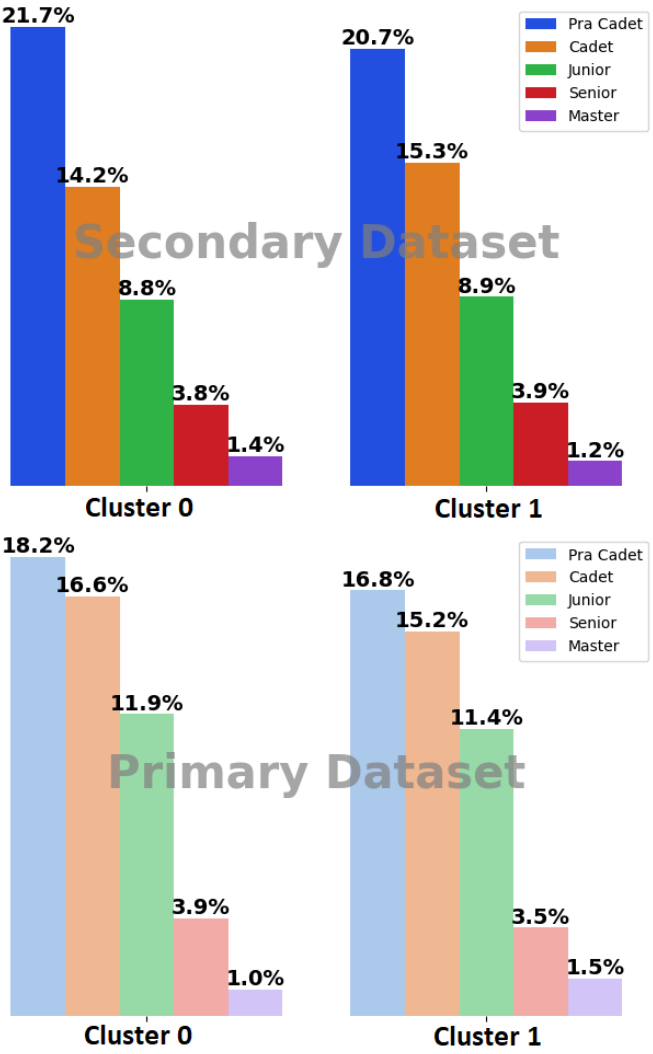


Fig. 7. Athletes age categories distribution.

For competition purposes, participants are classified by their birth year to determine their eligibility within these age groups, ensuring a fair matchup among competitors. Fig. 7 shows the differences between the two datasets. The most striking difference is in the Cadet class, which is intended for athletes aged 12 to 14. This difference is important for sports business management to know and to ensure the accuracy of their data, which will help them create marketing strategies. These findings are essential for sports business management, as understanding the specific age groupings within different clusters can inform training programs, marketing strategies, and other decision-making processes. Similarly, if an organization aims to expand or diversify its athlete base, knowing where certain age groups are underrepresented or overrepresented could help make strategic adjustments.

By ensuring accurate data analysis and a clear understanding of the athlete demographic across clusters, sports organizations can better align their offerings with the needs and preferences of their target audience, ultimately enhancing the effectiveness of training programs and marketing strategies.

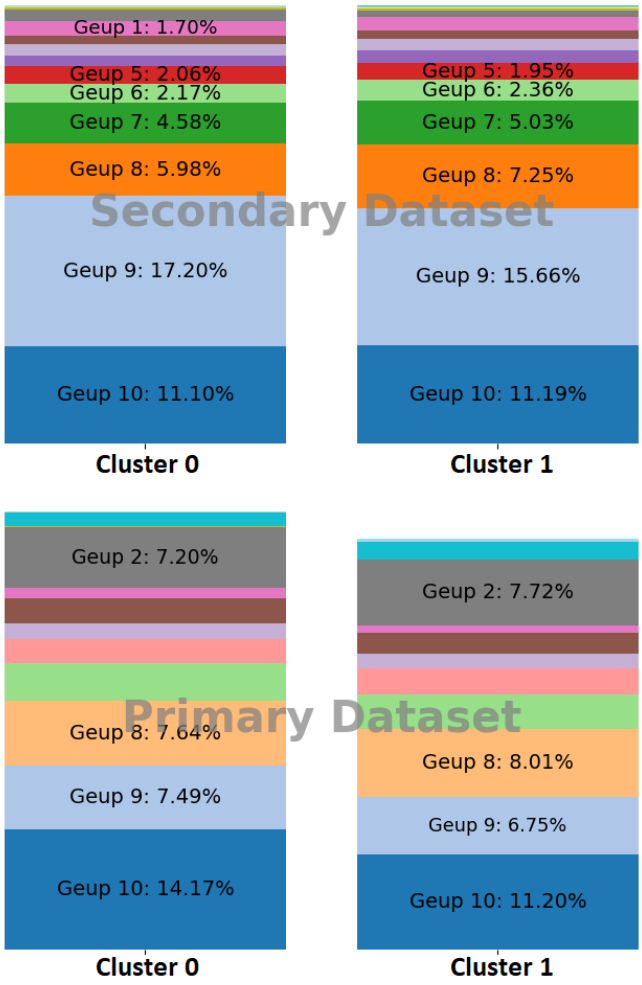


Fig. 8. Distribution of athlete belts.

Fig. 8 shows the distribution of athletes based on their belts. The most striking difference between the two datasets is that

most athletes are in the Geup 9 belt in the secondary dataset, followed by Geup 10 and Geup 8. While in the primary dataset, most athletes are in Geup 10, followed by Geup 8 and Geup 9.

This difference in belt distribution between the two datasets can provide valuable insights for sports management. For instance, understanding these belt distributions can help identify areas for targeted training programs or recruitment efforts. Additionally, the variation in belt rankings could influence the design of marketing strategies, ensuring that resources are allocated to the appropriate athlete segments, thus optimizing training and business operations.

Meanwhile in Fig. 9, we observe similarities and differences in the distribution of athlete residence distances relative to the Club/Dojang in the two datasets. An apparent similarity is that most athletes live <5 km from their Club/Dojang in primary and secondary datasets.

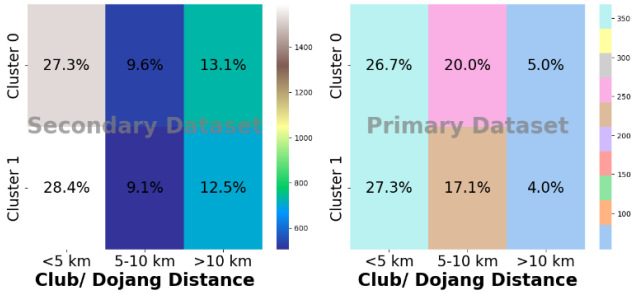


Fig. 9. Club / dojang distance based on secondary dataset.

However, the most notable difference is in the second largest category. In the secondary dataset, the second most common distance is 5-10 km, whereas in the primary dataset, the second largest group lives at a distance of >10 km. This shift in distance distribution between the two datasets highlights potential variations in athlete demographics and accessibility to training facilities, which could be valuable information for optimizing location-based training strategies or marketing efforts to improve participation rates across different distance segments.

C. Discussion

As mentioned before, in this study, the Zack knowledge gap model is integrated into the SECI model to identify gaps between the organization's current knowledge and what it should know. To determine the criteria of "what the firm must know", we adopted brand equity. We use brand equity because marketing practitioners need to understand the level of brand equity in a particular market area [22], which is obtained from customer/consumer responses to the services provided [23]. Fig. 10 shows the new knowledge management model for sports business that we derived from this research.

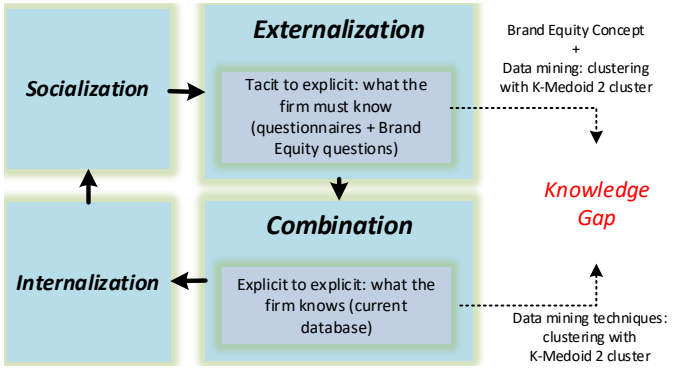


Fig. 10. The new model.

The novelty of this research lies in the externalization and combination stages of the SECI model. We match the externalization in the SECI model with "what the firm must know" in the Zack model and the combination in the SECI model with "what the firm knows" in the Zack model. By merging these models, we aim to provide valuable insights for sports business management, helping organizations to develop effective strategies and improve their competitiveness. Table V shows the additional features of this new model compared to the old SECI model.

TABLE V. COMPARISON OF OLD AND NEW MODELS

Component	Old SECI	New Model
Separating the knowledge between what is already known and what should be known	No	Yes, by using secondary and primary datasets.
Input from the customer's perspective	No	Yes, with the brand equity questionnaire.
Using data mining for data processing	Not mentioned	Yes, with adjustments depending on the needs of the organization.

This new model's advantage is the separation between what is known and what should be known. This model uses customer input to determine what should be known. The results of the separation prove that there is a knowledge gap between the two. This is important to provide awareness for management not to simply rely on what is already known but to update data and get input from customers directly.

We also used clustering data mining techniques to analyze the knowledge gap between both datasets. Based on our research, the K-medoids algorithm with two clusters was optimal on both datasets. For application in other sports businesses besides Taekwondo Club/Dojang and other business sectors besides sports, the management must know the proper data mining techniques and algorithms to apply to their databases to use this new model.

V. CONCLUSION

This study combines SECI and Zack models, brand equity concepts, and data mining methods to develop a new knowledge management model for sports business services. In this study, we use two datasets to identify knowledge gaps. The novelty of this study is in the externalization and combination of the SECI model. By using this new model, management will gain vital insights that they can use to formulate strategies and improve their competitiveness in the sports services market. To minimize the knowledge gap, sports business management must ensure that their knowledge is the latest data, so they must collect regular data from the athletes. This new model can also be used by other businesses besides sports. In addition to choosing the proper data mining techniques and algorithms, we need to customize the creation of the brand equity questionnaire according to each business's needs. Future work can focus on extending the application of this model to other industries, such as healthcare or retail, to validate its flexibility and effectiveness in different contexts. In addition, further research can explore integrating advanced data mining techniques, such as machine learning algorithms, to improve knowledge gap analysis and increase prediction accuracy.

REFERENCES

[1] T. Koç, K. Kurt, and A. Akbıyık, "A Brief Summary of Knowledge Management Domain: 10-Year History of the Journal of Knowledge Management," Procedia Comput. Sci., vol. 158, pp. 891–898, 2019, doi: 10.1016/j.procs.2019.09.128.

[2] D. Komaludin, "Model Manajemen Pengelolaan Pusat Olahraga dan Rekreasi Melalui Pengembangan Sumber Daya Manusia Berbasis Pengetahuan (Knowledge Based Worker)," J. Fak. Kegur. Ilmu Pendidik., vol. 1, no. 1, pp. 19–33, 2020.

[3] A. Rot and M. Sobinska, "Challenges for Knowledge Management in Digital Business Models," in 2020 10th International Conference on Advanced Computer Information Technologies (ACIT), IEEE, Sep. 2020, pp. 555–558. doi: 10.1109/ACIT49673.2020.9208867.

[4] M. Firdaus Abdullah, A. Yazid Abu Bakar, and U. Kebangsaan Malaysia Mohamad Nizam Nazarudin, "The Significance of Knowledge, Skills, Capabilities and Competencies in Fitness Centres' Management," Acad. Strateg. Manag. J., vol. 20, no. 6, p. 2021, 2021.

[5] E. Ode and R. Ayavoo, "The mediating role of knowledge application in the relationship between knowledge management practices and firm innovation," J. Innov. Knowl., vol. 5, no. 3, pp. 209–217, 2020, doi: 10.1016/j.jik.2019.08.002.

[6] A. O. Adesina and D. N. Ocholla, "The SECI Model in Knowledge Management Practices," Mousaion South African J. Inf. Stud., vol. 37, no. 3, 2019, doi: 10.25159/2663-659x/6557.

[7] F. P. Sihotang, Ermatita, Samsuryadi, and D. Palupi Rini, "Knowledge Management Model Review in Relation to Marketing and Branding," in 2022 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), IEEE, Nov. 2022, pp. 490–495. doi: 10.1109/ICIMCIS56303.2022.10017560.

[8] C. M. Sarungu, Titan, D. F. Murad, and Sunardi, "Storing, diving and distributing of comprehensive knowledge using knowledge management in the library and knowledge center," Proc. - 2017 2nd Int. Conf. Inf. Technol. Inf. Syst. Electr. Eng. ICITISEE 2017, vol. 2018-Janua, pp. 288–293, 2018, doi: 10.1109/ICITISEE.2017.8285513.

[9] A. Gupta, P. Bhatia, K. Dave, and P. Jain, "Stock market prediction using data mining techniques," in 2nd International Conference on Advances in Science & Technology (ICAST-2019), Maharashtra,India: University of Mumbai, 2019. [Online]. Available: http://ssrn.com/link/2019-ICAST.html

[10] N. Deepa, J. Sathya Priya, and T. Devi, "Towards applying internet of things and machine learning for the risk prediction of COVID-19 in pandemic situation using Naive Bayes classifier for improving accuracy," Mater. Today Proc., vol. 62, pp. 4795–4799, 2022, doi: 10.1016/j.matpr.2022.03.345.

[11] S. M. Shrestha and A. Shakya, "A Customer Churn Prediction Model using XGBoost for the Telecommunication Industry in Nepal," Procedia Comput. Sci., vol. 215, pp. 652–661, 2022, doi: 10.1016/j.procs.2022.12.067.

[12] Y. Sun, H. Liu, and Y. Gao, "Research on customer lifetime value based on machine learning algorithms and customer relationship management analysis model," Heliyon, vol. 9, no. 2, p. e13384, Feb. 2023, doi: 10.1016/j.heliyon.2023.e13384.

[13] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, "Sentiment Analysis for Customer Review: Case Study of Traveloka," in 7th International Conference on Computer Science and Computational Intelligence 2022, Bandung, Indonesia: Elsevier B.V., 2023, pp. 682–690. doi: 10.1016/j.procs.2022.12.184.

[14] N. N. Moon, I. M. Talha, and I. Salehin, "An advanced intelligence system in customer online shopping behavior and satisfaction analysis," Curr. Res. Behav. Sci., vol. 2, no. August, p. 100051, Nov. 2021, doi: 10.1016/j.crbeha.2021.100051.

[15] A. Patel, P. Oza, and S. Agrawal, "Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model," Procedia Comput. Sci., vol. 218, pp. 2459–2467, 2023, doi: 10.1016/j.procs.2023.01.221.

[16] P. Savci and B. Das, "Prediction of the customers' interests using sentiment analysis in e-commerce data for comparison of Arabic, English, and Turkish languages," J. King Saud Univ. - Comput. Inf. Sci., vol. 35, no. 3, pp. 227–237, Mar. 2023, doi: 10.1016/j.jksuci.2023.02.017.

[17] M. Z. Abedin, P. Hajek, T. Sharif, M. S. Satu, and M. I. Khan, "Modelling bank customer behaviour using feature engineering and classification techniques," Res. Int. Bus. Financ., vol. 65, no. October 2022, p. 101913, Apr. 2023, doi: 10.1016/j.ribaf.2023.101913.

[18] R. Komitova, D. Raabe, R. Rein, and D. Memmert, "Time Series Data Mining for Sport Data: a Review," Int. J. Comput. Sci. Sport, vol. 21, no. 2, pp. 17–31, 2022, doi: 10.2478/ijcss-2022-0008.

[19] M. Mandorino, A. J. Figueiredo, G. Cima, and A. Tessitore, "A Data Mining Approach to Predict Non-Contact Injuries in Young Soccer Players," Int. J. Comput. Sci. Sport, vol. 20, no. 2, pp. 147–163, 2021, doi: 10.2478/ijcss-2021-0009.

[20] V. Sarlis and C. Tjortjis, "Sports Analytics: Data Mining to Uncover NBA Player Position, Age, and Injury Impact on Performance and Economics," Inf., vol. 15, no. 4, 2024, doi: 10.3390/info15040242.

[21] S. Wenninger, D. Link, and M. Lames, "Data Mining in Elite Beach Volleyball-Detecting Tactical Patterns Using Market Basket Analysis," Int. J. Comput. Sci. Sport, vol. 18, no. 2, pp. 1–19, 2019, doi: 10.2478/ijcss-2019-0010.

[22] R. Rungtrakulchai, "The effects of brand equity and brand personality on customer value: The case of Uniqlo in Thailand," in Proceedings of 2018 5th International Conference on Business and Industrial Research: Smart Technology for Next Generation of Information, Engineering, Business and Social Science, ICBIR 2018, Bangkok, Thailand: IEEE, 2018, pp. 491–495. doi: 10.1109/ICBIR.2018.8391247.

[23] M. F. Juddi, S. Perbawasari, S. Dida, and A. R. Nugraha, "Branding of Purwakarta Regency Through Gempungan di Buruan Urang Lembur in Health Services," in Proceedings of the Asia-Pacific Research in Social Sciences and Humanities Universitas Indonesia Conference (APRISH 2019), 2021, pp. 154–161. doi: 10.2991/assehr.k.210531.019.

[24] E. Zuraidah, "Knowledge Management System Untuk SDM Menggunakan Seci Model (Studi Kasus: Koperasi Karyawan)," J. Inform., vol. 5, no. 1, pp. 157–168, Apr. 2018, doi: 10.31311/ji.v5i1.2481.

[25] R. Nurcahyo and D. I. Sensuse, "Knowledge Management System Dengan Seci Model Sebagai Media Knowledge Sharing Pada Proses Pengembangan Perangkat Lunak," J. Teknol. Terpadu, vol. 5, no. 2, pp. 63–76, 2019.

[26] F. Ramadhani and M. ER, "A Conceptual Model for the Use of Social Software in Business Process Management and Knowledge Management," Procedia Comput. Sci., vol. 161, pp. 1131–1138, 2019, doi: 10.1016/j.procs.2019.11.225.

[27] M. Ihsan Nugraha and J. S. Suroso, "Designing Knowledge Management System on Seller Education Tokopedia," in 2018 International

Conference on Information Management and Technology (ICIMTech), IEEE, Sep. 2018, pp. 27–32. doi: 10.1109/ICIMTech.2018.8528191.

[28] M. Miao, S. Saide, and D. Muwardi, "Positioning the Knowledge Creation and Business Strategy on Banking Industry in a Developing Country," IEEE Trans. Eng. Manag., pp. 1–9, 2021, doi: 10.1109/TEM.2021.3071640.

[29] A. Tizkar Sadabadi and A. Abdul Manaf, "IKML Approach to Integrating Knowledge Management and Learning for Software Project Management," Knowl. Manag. Res. Pract., vol. 16, no. 3, pp. 343–355, Jul. 2018, doi: 10.1080/14778238.2018.1474165.

[30] N. Norani, H. M. Belal, Q. Hasan, and M. Kosaka, "Knowledge Leadership : Managing Knowledge to Lead Change in PKT Logistics Group," in International Conference on Service Science and Innovation and Serviceology 2018, ICSSI 2018 & ICServ 2018, Taichung, Taiwan, 2019, pp. 156–160. [Online]. Available: https://www.researchgate.net/profile/Kuanlin-Chen/publication/339107708_Embarking_Advanced_Service_on_Servitization_Transformation_Make_Sense_Using_System_Thinking/links/5e3df772a6fdccd96590d854/Embarking-Advanced-Service-on-Servitization-Transformation-M

[31] I. Oktanisa and A. A. Supianto, "Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing," J. Teknol. Inf. dan Ilmu Komput., vol. 5, no. 5, p. 567, 2018, doi: 10.25126/jtiik.201855958.

[32] D. Triyansyah and D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," J. Telekomun. dan Komput., vol. 8, no. 3, pp. 163–182, 2018, doi: 10.22441/incomtech.v8i3.4174.

[33] I. Zulfa, R. Rayuwati, and K. Koko, "Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode Apriori (studi kasus: Kota Medan)," Tek. J. Sains dan Teknol., vol. 16, no. 1, pp. 69–82, 2020, doi: 10.36055/tjst.v16i1.7601.

[34] U. Narimawati, D. Munandar, and S. Mauluddin, "The Effectiveness of The Knowledge Management Model for Private Universities' Identity Branding," Cent. Eur. Manag. J., vol. 30, no. 4, pp. 696–704, 2022, doi: 10.57030/23364890.cemj.30.4.64.

[35] R. Sethi and Y. Jangir, "Measurement of Brand Equity: A Significant Aspect of Brand Equity Management," Aayushi Int. Interdiscip. Res. J., vol. 9, no. 1, pp. 9–11, 2022, [Online]. Available: https://www.aiirjournal.com/uploads/Articles/2022/02/5499_04.Dr. Ruhi Sethi & Yasha Jangir.pdf

[36] I. Ghozali, Aplikasi Analisis Multivariate Dengan Program IBM SPSS 26, 10th ed. Semarang: Badan Penerbit Universitas Diponegoro, 2021.

[37] I. Ariyanti, "Uji Validitas dan Reliabilitas Instrumen Angket Kemandirian Belajar Matematik," THETA J. Pendidik. Mat., vol. 1, no. 2, pp. 53–57, 2019.

[38] V. H. Antonius and D. Fitrianah, "Enhancing Customer Segmentation Insights by using RFM + Discount Proportion Model with Clustering Algorithms," Int. J. Adv. Comput. Sci. Appl., vol. 15, no. 3, pp. 902–911, 2024, doi: 10.14569/IJACSA.2024.0150390.

[39] L. Zahrotun, U. Linarti, B. H. T. Suandi As, H. Kurnia, and L. Y. Sabila, "Comparison of K-Medoids Method and Analytical Hierarchy Clustering on Students' Data Grouping," Int. J. Informatics Vis., vol. 7, no. 2, pp. 446–454, 2023, doi: 10.30630/joiv.7.2.1204.

[40] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," Electron., vol. 9, no. 8, pp. 1–12, 2020, doi: 10.3390/electronics9081295.

[41] K. F. Subroto and T. E. Balqiah, "The effects of gendered marketing on brand perception and purchase intention," Contemp. Res. Manag. Bus., pp. 111–114, 2022, doi: 10.1201/9781003295952-29.

Systematic Review of Prediction of Cancer Driver Genes with the Application of Graph Neural Networks

Noor Uddin Qureshi[1], Dr. Usman Amjad[2], Saima Hassan[3], Kashif Saleem[4]
NED University of Engineering & Technology, Karachi, Pakistan[1, 2]
Institute of Computing, Kohat University of Science & Technology, Kohat, Pakistan[3]
School of Computing, Macquarie University, North Ryde, NSW, Australia[4]

***Abstract*—Graph Neural Networks (GNNs) have emerged as a potential tool in cancer genomics research due to their ability to capture the structural information and interactions between genes in a network, enabling the prediction of cancer driver genes. This systematic literature review assesses the capabilities and challenges of GNNs in predicting cancer driver genes by accumulating findings from relevant papers and research. This systematic literature review focuses on the effectiveness of GNN-based algorithms related to cancer such as cancer gene identification, cancer progress dissection, prediction, and driver mutation identification. Moreover, this paper highlights the requirement to improve omics data integration, formulating personalized medicine models, and strengthening the interpretability of GNNs for clinical purposes. In general, the utilization of GNNs in clinical practice has a significant potential to lead to improved diagnostics and treatment procedures.**

Keywords*—Graph neural network; cancer driver genes; prediction; personalized medicine

I. Introduction

Cancer is a complicated and unique disease that comes with genetic mutation. Identification of the genes responsible for cancer development and evolution, is important for understanding the biological mechanisms and developing the treatments [1]. During recent studies, the application of machine learning techniques, specifically Graph Neural Networks (GNNs), has shown significant potential for the prediction of cancer driving genes by utilizing the data from the relevant biological networks [4].

Graph Neural Networks (GNNs) is one of the deep learning models, designed for studying complex networks, especially the biological linkages and connections [5]. GNNs have an upper edge on all other machine learning models with a potential to capture the structural information and interactions between entities in a network. In case of Cancer genes analysis, it provides interaction of genes and proteins with each other, enabling the prediction and identification of cancer genes based on the structural properties [1]. With the inter-linked features of the network, GNNs can understand patterns and relationships that are very important for identification of genes responsible for driving cancer [1].

Multiple studies have taken place to formulate the application of GGNs for genes identification, especially in the case of diseases like cancer. These studies will be discussed during the reviews.

In this systematic review, we keep an objective to provide a comprehensive overview of the recent state-of-the-art for the prediction of cancer driver genes using the application of Graph Neural Networks. We will analyze each methodological strategy with contrast to each other, considering data types and graph. In addition to that, we will explore the challenges and limitations for the application of GNNs for cancer gene prediction and discuss potential way forward for further research in this field.

With the consolidation of the findings extracted from the relevant researches and experiments studied, this review will contribute to a better systematic comprehension of the opportunities and challenges of GNNs for the predication cancer driver genes. The insights gained from this review will support researchers in developing more efficient models for finding cancer driver genes. This will assist in the development of designated treatments and improving results.

II. Objectives and Organization

The objectives of the systematic review are as follows:

- To provide a comprehensive overview of the use of Graph Neural Networks (GNNs) in predicting cancer driver genes.
- To analyze and compare different GNN-based models, including their methodological strategies, data types, and graph structures.
- To explore the challenges and limitations associated with GNNs in the context of cancer gene prediction.
- To identify potential future directions for research in applying GNNs for cancer gene identification.
- To summarize the current state-of-the-art in the field and highlight key findings from recent studies.

The structured organization of this paper is as follows:

- Fundamental Concepts and Terminology: This section discusses the foundational concepts related to the prediction of Cancer driver genes using GNNs.
- Cancer Genomics: An overview of cancer as a disease, including its causes, progression, and types and focusing

on the role of genomics in understanding genetic features and mutations for the identification of cancer driver genes.

- Disease Driver Genes Prediction: Discussion on different approaches for predicting any disease driver genes using omics data.
- Biological Data Related to Cancer Driver Genes: Study of different types of biological data, such as gene expression, protein-protein interaction, and multi-omics data, related to the identification of cancer driving genes.
- Graph Neural Networks (GNNs): Introductory overview to GNNs, their structure and application in studying graph-based data in cancer genomics.
- GNNs in Genomic Data Analysis: Exploration of the use of GNNs in genomic data analysis and their specific applications in cancer research.
- Literature Review: Discussion on relevant research studies that apply GNNs in cancer gene identification and prediction.
- Methodology of Research: Explanation of the structured approach used for perusing the systematic study of GNNs for predicting cancer driver genes.
- Research Questions: Formulation of the key research questions steering the analysis of GNN applications in cancer genomics.
- Procedure of Paper Exploration: Description of the procedure for the selection and analysis of relevant papers, along with data sources, searching strategies, and quality evaluation.
- Summary of Relevant Works: Detailed summaries of selected papers, highlighting the main themes, advantages, and disadvantages of each study.
- Comparative Analysis: Comparative evaluation of the reviewed GNN models, focusing on aspects such as input data diversity, target variable focus, model accuracy, and clinical relevance.
- Future Directions: Discussion of potential future research avenues in GNN applications for cancer driver gene prediction, including multi-omics integration and personalized medicine.
- Conclusion: Summarizes the findings of the review, highlighting the impact of GNNs on cancer genomics and their potential for advancing cancer treatment and understanding.

III. Fundamental Concepts and Terminology

This section of the review explores the foundational basis of GNN applications for the predictions of active genes which drive any disease, especially the cancer disease.

A. Cancer Genomics

Cancer is a class of diseases that is characterized by the uncontrolled growth and spread of harmful abnormal cells. Without any intervention, the spread can be fatal for the living organism. Cancers can be caused by external factors, such as tobacco use, infectious organisms, chemicals, and radiations, and also due to internal factors like inherited genetic mutations, hormones, and immunities etc. The process origination and expansion of cancer contains detailed multiple steps that takes place along the genetic changes within the cells. The harmful changes cause cells to grow and divide in an uncontrolled manner, which can form malignant tumors and affect the nearby parts of the organism [4]. In case, the cancer spreads to other organs, it is called metastatic cancer. The common types of cancer are carcinoma, sarcoma, and leukemia, lymphoma, and central nervous system cancers. Studying the cause of origin and molecular basis of different cancer types is vital for the development of controlled prevention, diagnosis, and treatment procedures to control the development and expansion of cancer in the organism body.

Cancer genomics evolve with the objective to study deep into the complex area of genetic and molecular conditions that are responsible for the development and progression of cancer. A primary aim for research in cancer genomics is to identify cancer driving genes, which play a crucial role in development of cancer [1].

Large research projects like The Cancer Genome Atlas have generated a huge amount of omics data from many cancer samples from different cancer types [2].

Machine learning and deep learning methods analyze mutation based patterns from multi-omics data to recognize driver genes. With the help of these detailed computational analyses of complex and enriched biological data, cancer genomics can detect and inspect the genes and processes involved in cancer origination and development.

B. Disease Driver Genes Prediction

The prediction and identification of genes that play a vital role in the development of diseases is an important problem in bioinformatics. Computational procedures have been developed to use expanded genomic and biological data to predict gene-disease associations.

Network-based approaches apply with protein-protein interaction networks, gene to gene expression networks, and path information to prioritize required disease genes [3]. Machine learning and deep learning algorithms also incorporate features extracted from sequences and expressions. Different data sources like genomic, biological network, functional associations, gene expression along with other relevant data like patient records are leveraged by computational prediction methods. The accurate prediction of disease driving genes from these data sources using network-based and machine learning approaches can provide an improvement to the ongoing treatments.

C. Biological Data Related to Cancer Driver Genes

The identification of cancer-related genes, which are also called cancer drivers, is vital for analyzing the molecular mechanisms of the cause and growth of cancer and also for the development of effective and targeted treatments.

Along with the advancement of computational biology, a vast amount of genome based biological data has begun being generated and also consolidated in recent years. This biological data can be categorized into different data types such as:

- Gene Expression Data: These are the biological process data where the genetic information encoded in a gene is used to develop a functional gene product. This has also revealed the activity levels of multiple genes in different types of cells. During the cancer research, the analysis of gene expression data can helps to understand comparative study in cancerous cells compared to normal cells [4].
- Protein-Protein Interaction Networks: These biological networks provide the mapping of the interactions in between the proteins, responsible for the cellular functions [5]. The subjective study of these networks provides the detail about the relation of gene mutations with protein interactions responsible for the development of cancer.
- Multi-Omics Data: This concept is related to the integration of various types of biological data like genomic, transcriptomic, proteomic, etc. For cancer gene prediction, it can offer an overall view of the biological processes and pathways affected cancer, which can lead to more accurate identification of cancer genes [5].
- Methylation-Level Biomarkers: Methylation itself is a type of DNA modification that can affect gene expression with an alteration. Detailed analysis of methylation patterns in cancer cells can indicate the activation and deactivation of the genes during cancer development [6].
- Transcription-Level Biomarkers: These are the biomarkers that are involved in the study of RNA transcripts to analyze the transcription of genes [6]. For the analysis of cancer, the transcription of relevant genes can indicate their function in cancer development.
- DNA Sequencing Data: Sequencing the DNA data from cancer cells and running the variant calling can reveal mutations, including single nucleotide variants (SNVs) and copy number alterations (CNAs) [7]. These steps are important for identifying genes that drive cancer development.

D. Graph Neural Networks (GNNs)

Graph Neural Networks (GNNs) are a type of machine learning algorithms which are designed specifically to process data with a graph like structure [8]. These are different from general neural networks that work on unconnected grid data. GNNs can directly process graph data as input, and train on feature relationships and effects by systematically combining feature information from each node, its nearer neighbors, and then moving on to more distant connections.

A main strength of GNNs is their ability to learn representations that capture both the features and the structure of the graph. By transferring and updating node information across edges with nearby nodes, GNNs can recognize patterns. This enables the network itself to handle various tasks such as identifying node types, predicting linkages in-between, and making wide predictions at the structure level.

Some of the most widely studied GNN models are Graph Convolutional Networks (GCNs) and Graph Attention Networks (GATs) [8]. Recent applications of GNNs have also emerged, such as Hierarchical Graph Neural Networks (HGNNs) for more targeted predictions within gene networks using the concept of parent-child and Heterophilic Graph Diffusion Convolutional Networks (HGDCs), which specifies in cases where node similarity is not significant, commonly seen in disease-related gene networks. The Explainablity ensured GNN framework offers additional insights into how these models make decisions.

GNNs are now more acceptable as a powerful algorithm for analyzing genomic data, especially in cancer research. Their ability to map and analyze complex structures in biological data enables researchers to better understand the connections within gene regulatory networks and protein-protein interactions. With the combination of different omics data and types, GNNs can provide a more expanded view of cancer with the help of multi-modal biological data. This can surely help to uncover the influence of genes cancer's growth and development.

Certain innovations in Graph based neural networks like Hierarchical Graph Neural Networks (HGNNs) [4] and Graph Attention Networks (GATs) [5] provides the utilization of GNN for the analysis of multi-omics data, along with the connection based data. Similarly, Heterophilic Graph Diffusion Convolutional Networks (HGDCs) [10] has worked to ensure the efficiency in heterophilic data settings which are common in the cancer genomics. These applications highlight utilization of GNN to provide more deeper analysis related to cancer mechanisms and personalized treatments.

IV. Relevant Reviews

In this section, we aim to provide significant relevant research work in the area of application of GNN related to cancer genomics. The research papers provide insights for the application of network specifically for cancer gene identification and prediction. Each of the approach reviewed provides multiple aspects of GNN algorithm and its impact on the accuracy and effectiveness of cancer gene prediction. The review of these papers provides the collective contribution towards the advancement in the field of cancer genomics and precision medicine by applying GNNs for predictive modeling and analysis.

Wan and Wu [11] introduced a semi-supervised GNN method called PersonalizedGNN, demonstrating remarkable performance in identifying personalized driver genes (PDGs) for cancer patients. Particularly, it successfully utilizes the structure information of personalized gene interaction networks and limited developed cancer tissue-specific driver genes. On the other hand, Cui and Wang [12] presented the self-supervised masked graph learning (SMG) framework for identifying cancer genes from multi-omic featured protein-protein interaction networks, which outperforms existing state-of-the-art methods. Li and Han [13] leverages graph attention-based deep learning and outperforms current approaches in cancer gene prediction,

integrating multi-omics information to dissect cancer gene modules. It has researched on the effectiveness of GNN for biological data analysis and multi-omics integration for further cancer research.

In addition, in method [9], the researchers innovated the research with SBM-GNN with the combination of GNNs and stochastic block models. This helps in the prediction of cancer driver genes and cancer development providing more accurate results as compared to other state-of-the-art methods. It also provides a scalable and interpretable approach for the integration of multi-omic data with protein based interaction data for cancer analysis. Zhang and Zie [10] proposed HGDC ensuring better performance, especially in identifying of relevant and targeted cancer genes. Furthermore, Zhao and Gu [5] showed an initiative with a GAT-based model to recognize cancer driving genes. The approach has also integrated multi-omics cancer data with multi-dimensional gene networks, which has shown better results as compared to other baseline models during evaluations. The study provides detailed methods for generating gene association profiles, constructing multi-dimensional gene networks, and using GAT for within-dimension interactions and joint learning for prediction.

Ratajczak and Joblin [14] implemented an ensemble graph representation learning framework, aiming at predicting core genes for complex diseases. Hou and Wang [4] developed a hierarchical graph neural network for classifying cancer stages and identifying gene clusters. The model in [6] integrates transcription and methylation-level biomarkers in an explainable GNN framework for microsatellite instability detection.

Hantano and Kamada [7] introduced Net-DMPred, a network-based machine learning method designed to predict cancer driver missense mutations by incorporating molecular networks. This approach provided better results as compared to other traditional methods and showed the importance of the integration of complete molecular network structure for data.

The review of the mentioned approaches collectively highlights the effectiveness of GNN-based models across different areas of cancer genomics, including personalized cancer gene identification, cancer pathway prediction, and the recognition of driver mutations. While each study presents a unique GNN approach, but overall all of the reviewed methodologies contribute to a consolidated and combined understanding of GNNs in cancer genomics. The papers demonstrate GNNs' potential to address the challenges in cancer gene prediction across multiple omics data types and network setups.

V. Methodology of Research

This research review aims to provide a systematic review following the guidelines mentioned by Lim and O'cass [15]. The focused topic selected is the application of Graph Neural Networks (GNNs) for predicting cancer driver genes. The methodology includes a structured, detailed analysis of shortlisted studies. It also examines the use and effectiveness of various GNN models for genomic data analysis, especially in predicting cancer driver genes. We aim to classify and understand the methods and findings across these papers. The insights from this review can possibly carry forward new applications of GNNs in bioinformatics.

A. Formalization of Question

This research primarily focuses on the detailed analysis of the GNN architectures presented in each paper to review their individual contributions. It involves the evaluation on the interpretation of the complexity and variability of cancer genomic data. The review provides a designated focus to the trends emerging from these studies and identification of common methodologies, data types, and approaches. This assists in understanding the current state of GNN applications in cancer genomics and the future possible paths originated from this research.

Key research questions guiding this analysis include:

RQ 1: How do different GNN models based approaches perform for the prediction of cancer genes, and what are their unique features?

RQ 2: What are the opportunities and challenges of these GNN approaches, and how do they contribute in this field?

RQ 3: What are the common methodologies and data types used in these GNN applications for cancer genomics?

RQ 4: What are the emerging trends and unresolved challenges in using GNNs for cancer driver gene prediction?

B. The Procedure of Paper Exploration

This investigation comprises a four-stage process for exploring and selecting papers, as demonstrated in Fig. 1, aligning with the SPAR-4-SLR framework [15].

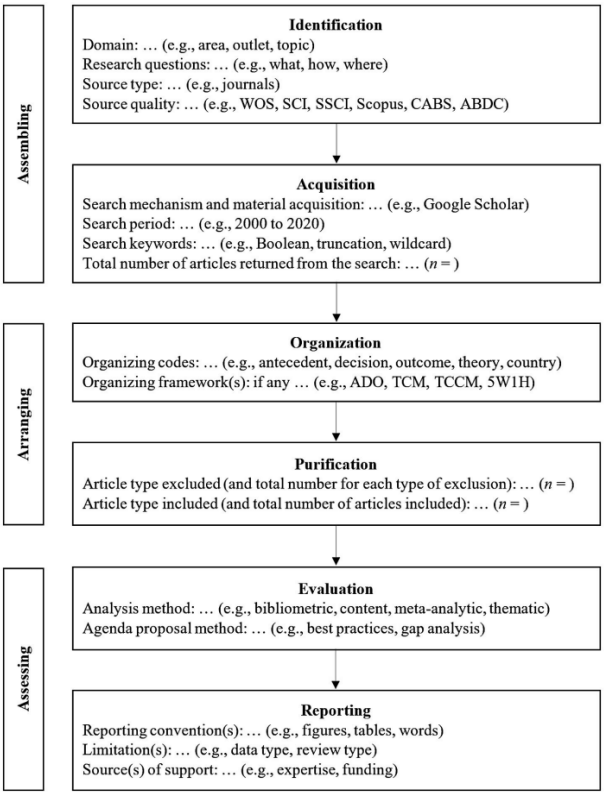


Fig. 1. Procedure for paper exploration.

1) Identification: The identification step is the foundation of any literature review. For this paper we have specified our focus on the role of Graph Neural Networks (GNNs) in the identification of Cancer driver genes. This is an area of research lies under the domain of Cancer genomic sand bioinformatics.

2) Acquisitions: Depending on previous knowledge, experience in research and suggestions given by Hinderks [16], the most influential and common databases were selected. These repositories have different search mechanisms using keywords mentioned in Table I; therefore, we customized our search string accordingly. The selected digital repositories are:

- Google Scholar
- academic (OUP)
- BMC bioinformatics
- Research Square
- Biorxiv
- IEEE Xplore
- Other papers, chapters, journals, books and conference papers.

TABLE I. KEYWORDS AND SEARCH CRITERIA

S#	Keywords and search criteria
S1	"Graph Neural Networks" and "Cancer Genomics"
S2	"GNN" and "Cancer Gene Prediction"
S3	"GNN" and "Disease driver Genes"
S4	"Graph Networks" and "Cacner Drivers"
S5	"Multi-Omics Data Integration" and "GNN"
S6	"Graph Networks" and "Bioinformatics"

3) Organization: The categorization and segmentation of the researches is done on the basis of input types of the data (gene-gene interactions, multi omics data etc.), methodologies like simple Graph Neural Networks, Graph Attention Networks etc. and the target outcomes such as, prediction, identification or classification. By organizing the papers according to these categories, the resulting structure provided a care pathway to identify trends and pursue the comparative analysis.

4) Purification: The following criteria was formulated for the selection of kind of articles and reports that are needed be included in this study:

- The paper should be published in English language.
- The complete content of the paper should be available on online.
- The paper must be presented in an acknowledged conference or journal.

The paper which were not included in the study are as follows:

- The paper published in another language rather than English.
- Duplicate studies are not included.
- Personal or casual blogs are not included.

5) Evaluation: The QA is ensured using the three questions (QA1-QA3) that are mentioned below. We have followed the guidelines provided by Shaffril [17] to evaluate the quality metric criteria. These guidelines were also used to evaluate the quality evaluation criteria for the selected papers, as mentioned in Table II.

Furthermore, the aim of quality assessment is to provide that the findings of the selected article will be important for the review paper. During the quality assessment, the studies which have the score of 2 were included in this review and all the remaining were removed. With the mentioned process, we have removed 3 articles which have scores less than 2:

- A1: Does the article provides discussion on the Graph Neural Network?
- A2: Does the article explore the Prediction of Genes?
- A3: Does the article describe the model structure in detail?

TABLE II. QUALITY STATEMENT

Paper Title	A1	A2	A3	Total
Speos [16]	1	1	0.5	2.5
HGNN [4]	1	0.5	1	2.5
MSI-XGNN [6]	1	1	0.5	2.5
MODIG [5]	1	1	1	3
Net-DMPred [7]	1	1	0.5	2.5
HGDC [10]	1	0.5	1	2.5
CGMega [13]	1	1	0.5	2.5
SBM-GNN [9]	1	1	1	3
PersonalizedGNN [11]	1	1	0.5	2.5
SMG [12]	1	0.5	1	2.5

6) Reporting: Finally, 10 model-based articles were shortlisted which cover GNN used on/for Cancer driver genes. The studies used different classes and structure of GNNs, along with different types of bioinformatics data as input, cancer types for analysis, and output feature.

VI. SUMMARY OF RELEVANT WORKS

1) Speos [16]: Speos is a machine learning method that predicts genes linked to diseases for advanced drug development. The research uniquely combines different types of molecular network data, with the provision of high validation results There is still a space to research for the application of initial network data for predictions.

2) MSI-XGNN [6]: MSI-XGNN worked on RNA sequencing data and DNA methylation data. The output predicts the microsatellite instability (MSI) in cancer, which provides the insights of immunotherapy decisions. It is

comparatively accurate and more interpretable, providing a space of improvement in the integration of complete data.

3) Net-DMPred [7]: Net-DMPred is a graph neural network that shows its uniqueness related to features of molecular networks to predict cancer driver mutations. It has shown better performance as compared to the traditional methods due to the integration of whole network structures. It could further improve by refining its network design and adding more features.

4) CGMega [13]: CGMega is a transformer-based model that predicts cancer genes and identifies gene modules. It explores a new approach in its study to merge multi-omics data, but it needs to be tested more across different cancer types and datasets.

5) SBM-GNN [9]: SBM-GNN combines graph neural networks and stochastic block models to find cancer driver genes and pathways. The stochastic block identifies cluster of nodes with similar structural roles and then the GNN part refines the node level features for the prediction. It is significantly accurate and provides better evaluation than other similar methods, but its complexity requires advanced knowledge to use and interpret.

6) PersonalizedGNN [11]: PersonalizedGNN is a semi-supervised graph neural network that has validated the prediction of cancer driver genes through lab experiments. Unlike other traditional methods, it constructs personalized gene interaction networks. It can be leveraged more with the integration of non-coding DNA regions to identify other types of driver genes too. To leverage the class imbalance issue in Cancer genomics dataset, label reusing mechanism is applied.

7) HGNN [4]: HGNN is a special type of hierarchical graph neural network which performs the classification of different cancer stages and clusters out the related genes. The performance is significant, but requires larger datasets for more detailed cancer stage divisions.

8) HGDC [10]: HGDC advances in the identification of cancer driver genes by using graph diffusion and layer-wise attention by showing strong results in terms of accuracy with the balanced class data, but compromises the accuracy with the imbalanced data.

9) MODIG [5]: This approach combines multi-omics and multi-dimensional gene network data to predict cancer driver genes. It performs better in finding potential markers for cancer prognosis. It can show further improvement by building more complex graphs and filtering out noise in gene data.

10)SMG [12]: The Self-Supervised Masked Graph Learning Framework identifies cancer genes by using large datasets without labels. This self-supervised approach helps it to handle data shortages, but mostly relies on unlabeled data. This can raise challenges in ensuring data quality and relevance, which requires further careful data selection and preparation.

VII. Comparative Analysis

A. Analysis of GNN Models

The Graph Neural Network (GNN) models that were reviewed provides different aspects of application to address cancer genomics, and responded to a range of challenges in this field. Some GNN based models, such as Speos and MODIG, focus more on integrating multi-omics data [16] [5]. This wide approach helps to study deep into complex biological processes in cancer by combining genomic, epigenomic, transcriptomic, and proteomic data. This creates a complete view of cancer processes from origin to the development.

Complex models, like HGNN and HGDC, works on the features extracted from the structural details of the genomic data itself [4] [10]. HGNN explores hierarchical data structures, which provides deep analysis on the layered biological interactions in cancer progression. In contrast, HGDC interprets the heterophilic patterns of the genomic data, which are crucial for understanding cancer complexity. MSI-XGNN and Net-DMPred use specific and relevant data types, such as RNA sequencing and molecular networks [6] [7]. These models highlight the value of targeted genomic analysis for the identification of specific genetic mutations and epigenetic changes linked to cancer.

CGMega and SBM-GNN study on advanced applications of GNN technology in cancer research. CGMega uses a transformer-based graph attention network for predicting gene modules, while SBM-GNN sums up the stochastic block models with GNNs for identifying process pathways [13] [9]. Following the concept of precision medicine, models like PersonalizedGNN and SMG opted for more patient specific personalized approaches. PersonalizedGNN uses patient based genomic data, getting the data fusion [11]. SMG uses self-supervised learning to work with unlabeled data for cancer gene identification [12].

All of the discussed models collectively demonstrate the vast use of GNNs in cancer genomics. Each model contributes in a significant way to show possible avenues for the adaptability of GNNs to cater the diverse challenges of this field.

B. Input Data Diversity

Each of the GNN models reviewed approach cancer genomics with different types of data. Speos uses multiple kinds of molecular network data to predict primary genes connected to diseases [16]. HGNN analyze layers of data in genomics to find relationships important for the study of cancer progression [4]. MSI-XGNN uses both RNA sequencing and DNA methylation data as input, showing how genetic and epigenetic factors connect in cancer [6]. MODIG combines different types of data (multi-omics) to identify cancer driver genes [5], while Net-DMPred focuses on features of molecular networks to improve predictions of cancer driver mutations [7].

HGDC uses unique genomic data patterns at input to help identify cancer driver genes in different types of biomolecular

networks [10]. CGMega uses multi-omics data to predict cancer genes and analyze gene clusters for it [13]. Like discussed, PersonalizedGNN has input data type of patient targeted genomic data, moving toward personalized treatment strategies for cancer [11]. Finally, SMG applies self-supervised learning to handle unlabeled data, as an advancement in cancer gene identification and showing the benefits of new learning techniques in cancer research [12].

C. Target Variable Focus

In the review of GNN models for cancer genomics, the models have a different approach when it comes to approach the target variables, which expands the interest in research goals in this field.

Speos focuses on the prediction of the core genes important for analyzing and interpreting the disease processes and finding the relevant potential drug targets [16]. HGNN studies specifically on the classification of different cancer stages and identification of involved gene groups. This helps to segment and understand the evolution and development of cancer [4]. MSI-XGNN specializes in predicting small unstable base segments in the DNA, which are relevant for formulating personalized cancer treatments [6]. MODIG, like mentioned, combines multi-dimensional gene networks and various types of omics data to identify cancer driver genes as output classification [5].

Net-DMPred converges its study on the prediction of relevant mutations responsible to drive cancer, adding important insights to mutation analysis in cancer genomics [7]. HGDC researched with the objective to find cancer driver genes with the association different biomolecular networks, provide an understandable output for gene to gene interaction in cancer [10]. CGMega predicts cancer genes and identifies gene groups too, to explain the complex interactions between genes in cancer [13]. SBM-GNN combines GNNs with stochastic block models to identify cancer driver genes and pathways as output [9].

PersonalizedGNN predicts driver genes unique to each patient, moving toward customized treatments based on individual genomic profiles [11]. And SMG uses self-supervised learning to identify cancer genes, showing how advanced machine learning methods can help in cancer genomics [12].

Through their different focuses, these models contribute to a versatile approach to define the outputs for the cancer driver genes.

D. Model Accuracy and Performance

The study reviews several researches that focus on the accuracy and efficiency of different models to recognize cancer-related genes and predict disease related statuses. It compares themes and methods in all of the reviewed studies, by highlighting the development of new methods aimed at improving prediction accuracy. Though, as discussed, each paper carries a different approach for the selection of output, but individual contribution of the models towards evaluation metrics are discussed. The first set of models studied includes GNN-based models like PersonalizedGNN and CGMega, which are designed to identify personalized driver genes and cancer gene modules achieve high precision-recall rates [13][11]. Average Precision values of PersonalizedGNN for BRCA, LUAD, and LUSC cancer types are as 66.1%, 89.7% 72.1% respectively.

Models like Net-DMPred and SMG [12] focus on network-based prediction and self-supervised learning to identify cancer genes using molecular networks and protein-protein interaction (PPI) data [7]. These models perform more accurately than traditional models with the importance of molecular network structures and multi-omics features for accurate predictions. Net-DMPred provides mean ROC-AUC values for Cancer pathways combined with molecular interactions as 89.9% and 90.6% for Graph node dimensions of 90.6% respectively.

The paper also discusses deep learning models like HGNN provided the average accuracy of the baselines as 84.68% for BRCA, 64.30% for STAD, and 87.42% for COAD cancer types. Speos worked on classifying cancer stages and predicting core disease genes [16] [4]. Both of these models have shown strong results for the identification of relevant gene clusters and disease genes, with high accuracy and validation across multiple disease types.

SMG worked on self-supervised learning and using multiple data types [12]. It achieved the most accurate performance in terms of AUPRC values for all eight PPI networks, with an overall average of 7.4% compared to the discussed models on each data set while Net-DMPred centers on network structures for predicting driver missense mutations [7]. MODIG combines multi-dimensional gene networks to identify driver genes. Its AUPR values across different PPI networks averages around 79% [13].

In consolidation, these models provide keen importance on the accuracy and performance in terms of cancer gene prediction and disease status classification. Although the performances of the models that are discussed, are directed towards a goal of developing advanced models, each study offers a different view that can help the doctors and researchers to better understand cancer biology and advance in personalized medicine.

E. Clinical Relevance and Applications

The studies cover the complete spectrum on the focus on the clinical relevance of the discussed models and their practical applications. Each of them provides a unique perspective on potential of the findings that can impact cancer research and patient care treatments. All of the studies emphasize the potential of machine learning methods especially the use of GNNs for the identification of personalized driver genes, explanation of cancer gene modules, and the prediction of cancer-specific mutations related to other diseases [16]. These methods have a prominent potential to improve the process cancer driver gene identification, integration of multi-omics data, predict cancer mutations, and better interpretation of complex diseases at the genetic and molecular level.

The papers also support the application of GNNs by addressing different areas in cancer research and precision medicine. For example, studies like CGMega [13], SMG [12], HGDC, and Net-DMPred worked on the integration of multi-omics data, protein-protein interaction networks, and gene regulatory networks. This improved cancer gene prediction, identification and classification too [7][10]. On the other hand, models like PersonalizedGNN [11], SBM-GNN [9], and MSI-

XGNN focus on personalized treatments, with the identification of patient specific driver genes, and prediction of cancer-specific mutations, which is a significant contribution to precision oncology and personalized medicine [6]. All of these studies collectively strengthen the argument for the clinical applications of GNNs as mentioned in Table III, that could guide future cancer research and treatment approaches.

TABLE III. COMPARATIVE ANALYSIS OF DIFFERENT MODELS

Model	Input Data	Methodology	Performance Metrics	Strengths	Limitations
Speos [16]	Molecular network data	Machine learning integrated with diverse molecular networks	High validation accuracy	Effective integration of various molecular data	Limited exploration of initial network data for predictions
MSI-XGNN [6]	RNA sequencing, DNA methylation data	Explainable GNN framework combined with gene expression and methylation profiles	AUC: 0.91	Accurate MSI status prediction and aids immunotherapy decisions	Requires comprehensive data integration
Net-DMPred [7]	Molecular networks, variant data	GNN leveraging molecular network features	ROC-AUC: 0.899	Superior performance over traditional methods and considers large-scale molecular networks	Needs refinement in network design; computationally intensive
CGMega [13]	Multi-omics data	Transformer-based model integrating multi-omics	AUPRC: 0.79	Novel approach in merging multi-omics data and identifies gene modules	Requires validation across diverse cancer types and datasets
SBM-GNN [9]	Multi-omics, PPI networks	Combination of Stochastic Block Models and GNN	ROC-AUC: 0.906	High accuracy in identifying cancer driver genes and pathways	Complex model requiring advanced interpretation
PersonalizedGNN [11]	Patient-specific gene networks	Semi-supervised GNN tailored for individual profiles	Precision: BRCA 66.1%, LUAD 89.7%, LUSC 72.1%	Personalized predictions validated through lab experiments	Dependent on availability of personalized data; limited scalability
HGNN [4]	Gene expression, hierarchical data	Hierarchical GNN with subgraph perturbations	Accuracy: 84.6–87.4%	Effective in classifying cancer stages and clusters related genes	Requires large datasets for detailed analysis
HGDC [10]	Heterophilic biomolecular networks	Graph Diffusion Convolutional Network with layer-wise attention	High accuracy on balanced data	Strong performance with balanced class data	Accuracy decreases with imbalanced data
MODIG [5]	Multi-dimensional omics, PPI networks	GAT-based model integrating multi-omics and gene associations	AUPRC: 0.79	Outperforms baseline models with effective multi-omics integration	Sensitive to noise; requires complex graph construction
SMG [12]	Unlabeled PPI and biological networks	Self-supervised masked graph learning framework	AUPRC: 0.74	Handles data shortages with the reduction in dependency on labeled data	Challenges in ensuring data quality and relies on careful data selection

VIII. DISCUSSION

With the review of the discussed researches, it is evident that the utilization of GNNs in cancer genomics provide an innovative advancement in predicting cancer driver genes. These models were prominently able to process the complex biological relationships, which provided an edge over traditional methods.

One of the significant finding was the way in which GNNs were used to process different types of data as input like in MODIG [5] and CGMega [13]. While the studies like PersonalizedGNN provided a solution to tailor the results for the individual patients specifically, ensuring the concept of precision medicine. The reviews models also illustrated a notable accuracy with SBM-GNN [9] and Net-DMPred [7], with complex structural biological data.

As far as challenges are concerned, data quality is one the significant hurdles, lacking high quality annotated datasets, especially for multi-omics studies. This limitation also affects the consistent evaluation of the models. Additionally the computational capacity for processing GNNs make them less accessible especially in the clinical settings. Apart from that imbalanced class datasets and explainability of the models and their results to medical subject specialists needs to be addressed to apply the recommended models in real-world scenarios.

Regardless, the progress so far is a strong foundation for future to revolutionize the cancer diagnosis, prediction, treatment and prognosis.

IX. FUTURE DIRECTION

The reviewed papers are enriched with the future directions in using Graph Neural Networks (GNNs) for cancer driver gene

prediction. The most prominent direction is for the improvement in the integration of multi-omics data, which can advance the research itself to go beyond layering data types to connecting them in a complex structure for deeper insights into cancer biology. As cancer genomics continues to reveal more enriched complex information, there is a growing need for sophisticated models that can handle the complexity.

Another important area is the patient specific personalized medicine. Future research may increasingly aim to develop and apply GNN models that can use both individual genetic and medical history data, to formulate cancer treatment more tailored for each patient specifically.

Improving model explainability is also needed with focus of XAI. As GNNs become more advanced, it will be important to make them understandable for clinicians for better feature selection and understandability. This would ensure clarity in predictions and application on clinical settings.

Hence, in terms of future directions, a multi-layered approach for new GNN applications in cancer genomics is encouraged with the focus on data integration, interpretability, and personalized medicine.

X. Conclusion

The systematic review for the application of Graph Neural Networks (GNNs) for the prediction of cancer driver genes has shown a promised progress in cancer genomics. Each of the discussed models offers insights for the development and utilization of GNNs. The paper also addressed the use of network for various parts of cancer genomics, from combining multiple types of data to integrating personalized medicine. There are notable instances of innovations in methods and the procedures towards more accurate, personalized cancer treatments.

With the advancement of AI and bioinformatics, GNNs appear to have great potential to transform the study of cancer biology. The fusion of these advanced models with clinical practice have a significant potential for better diagnostic tools and treatment methods. This will also improve patient outcomes. The continuous research has already provided an impact and sets the avenue for further cutting edge breakthroughs that will continue to expand knowledge on cancer and make cancer treatments more effective.

References

[1] Y. Han, J. Yang, X. Qian, W.-C. Cheng, S.-H. Liu, X. Hua, L. Zhou, Y. Yang, Q. Wu, P. Liu, and Y. Lu, "DriverML: a machine learning algorithm for identifying driver genes in cancer sequencing studies," Nucleic Acids Research, vol. 47, no. 8, p. e45, May 2019, doi: 10.1093/nar/gkz096.

[2] K. Tomczak, P. Czerwińska, and M. Wiznerowicz, "Review: The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge," Contemporary Oncology/Współczesna Onkologia, pp. 68–77, 2015, doi: 10.5114/wo.2014.47136.

[3] P. Luo, Y. Ding, X. Lei, and F.-X. Wu, "deepDriver: Predicting cancer driver genes based on somatic mutations using deep convolutional neural networks," Frontiers in Genetics, vol. 10, pp. 68–77, 2019, doi: 10.3389/fgene.2019.00013.

[4] W. Hou, Y. Wang, and Z. Zhao, "Hierarchical graph neural network with subgraph perturbations for key gene cluster discovery in cancer staging," Complex Intelligent Systems, 2023, doi: 10.1007/s40747-023-01068-6.

[5] W. Zhao, X. Gu, S. Chen, J. Wu, and Z. Zhou, "MODIG: integrating multi-omics and multi-dimensional gene network for cancer driver gene identification based on graph attention network model," Bioinformatics, vol. 38, no. 21, pp. 4901–4907, Nov. 2022, doi: 10.1093/bioinformatics/btac622.

[6] Y. Cao, D. Wang, J. Wu, Z. Yao, S. Shen, C. Niu, Y. Liu, P. Zhang, Q. Wang, J. Wang, H. Li, X. Wei, X. Wang, and Q. Dong, "MSI-XGNN: an explainable GNN computational framework integrating transcription- and methylation-level biomarkers for microsatellite instability detection," Briefings in Bioinformatics, vol. 24, no. 6, Nov. 2023, doi: 10.1093/bib/bbad362.

[7] N. Hatano, M. Kamada, and R. Kojima, "Network-based prediction approach for cancer-specific driver missense mutations using a graph neural network," BMC Bioinformatics, vol. 24, p. 383, 2023, doi: 10.1186/s12859-023-05507-6.

[8] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, "Graph neural networks: A review of methods and applications," AI Open, vol. 1, pp. 57–81, 2020, doi: 10.1016/j.aiopen.2021.01.001.

[9] V. Fanfani, R. V. Torne, P. Lio', and G. Stracquadanio, "Discovering cancer driver genes and pathways using stochastic block model graph neural networks," bioRxiv, Jun. 2021, doi: 10.1101/2021.06.29.450342.

[10] T. Zhang, S.-W. Zhang, M.-Y. Xie, and Y. Li, "A novel heterophilic graph diffusion convolutional network for identifying cancer driver genes," Briefings in Bioinformatics, vol. 24, no. 3, May 2023, doi: 10.1093/bib/bbad137.

[11] H.-W. Wan, M. Wu, W. Zhao, H. Cheng, B. Ying, X.-F. Wang, X.-R. Zhang, Y. Li, and W. Guo, "Label reusing based graph neural network for unbalanced classification of personalized driver genes in cancer," SSRN, 2023, doi: 10.2139/ssrn.4510873.

[12] Y. Cui, Z. Wang, X. Wang, Y. Zhang, Y. Zhang, T. Pan, Z. Zhang, S. Li, Y. Guo, T. Akutsu, and J. Song, "SMG: self-supervised masked graph learning for cancer gene identification," Briefings in Bioinformatics, vol. 24, no. 6, Nov. 2023, doi: 10.1093/bib/bbad406.

[13] H. Li, Z. Han, and Y. Sun, "CGMega: Explainable graph neural network framework with attention mechanisms for cancer gene module dissection," Research Square, Jul. 2023, doi: 10.21203/rs.3.rs-3180743/v1.

[14] F. Ratajczak, M. Joblin, and M. Hildebrandt, "Speos: an ensemble graph representation learning framework to predict core gene candidates for complex diseases," Nature Communications, vol. 14, p. 7206, 2023, doi: 10.1038/s41467-023-42975-z.

[15] P. J. Lim, W. M. O'Cass, A. Hao, and S. Bresciani, "Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR)," International Journal of Consumer Studies, vol. 45, no. 4, pp. O1–O16, 2021, doi: 10.1111/ijcs.12695.

[16] A. Hinderks, F. J. Domínguez Mayo, J. Thomaschewski, and M. J. Escalona, "An SLR-tool: search process in practice: a tool to conduct and manage systematic literature review (SLR)," in Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Companion Proceedings (ICSE '20), New York, USA: Association for Computing Machinery, 2020, pp. 81–84, doi: 10.1145/3377812.3382137.

[17] M. Shaffril, S. F. Samsuddin, and A. A. Samah, "The ABC of systematic literature review: the basic methodological guidance for beginners," Quality and Quantity, vol. 55, pp. 1319–1346, 2021, doi: 10.1007/s11135-020-01059-6.

Albument-NAS: An Enhanced Bone Fracture Detection Model

Evandiaz Fedora[1], Alexander Agung Santoso Gunawan[2]
Computer Science Department-Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia[1]
Computer Science Department-School of Computer Science, Bina Nusantara University, Jakarta, Indonesia[2]

***Abstract*—Diagnosing fracture locations accurately is challenging, as it heavily depends on the radiologist's expertise; however, image quality, especially with minor fractures, can limit precision, highlighting the need for automated methods. The accuracy of diagnosing fracture locations often relies on radiologists' expertise; however, image quality, particularly with smaller fractures, can limit precision, underscoring the need for automated methods. Although a large volume of data is available for observation, many datasets lack annotated labels, and manually labeling this data would be highly time-consuming. This research introduces Albument-NAS, a technique that combines the One Shot Detector (OSD) model with the Albumentation image augmentation approach to enhance both speed and accuracy in detecting fracture locations. Albument-NAS achieved a mAP@50 of 83.5%, precision of 87%, and recall of 65.7%, significantly outperforming the previous state-of-the-art model, which had a mAP@50 of 63.8%, when tested on the GRAZPEDWRI dataset—a collection of pediatric wrist injury X-rays. These results establish a new benchmark in fracture detection, illustrating the advantages of combining augmentation techniques with advanced detection models to overcome challenges in medical image analysis.**

***Keywords*—Albumentation; augmentation; bone fracture; deep learning; object detection; YOLO-NAS**

I. Introduction

Medical images are essential for modern healthcare and diagnostics. However, their limited resolution can make it challenging for healthcare providers to fully evaluate a patient's condition [1]. Medical image analysis also faces obstacles like insufficient data and the complexity of interpreting outcomes [2]. Recently, deep learning has shown promising potential to automate and enhance medical image analysis, boosting accuracy and efficiency in diagnosing fractures. Yet, limited data and challenges in result interpretation continue to hinder its application [3]. The lack of large, high-quality medical datasets is a major barrier in training effective deep learning models. Furthermore, although deep learning can deliver accurate predictions, healthcare professionals often find it challenging to interpret the outcomes produced by these models [4].

Detecting bone fractures is essential for timely medical intervention and effective rehabilitation [5]. Traditional methods depend on manual examination of images, a process that is both time-consuming and susceptible to human error [6]. Surgeons typically require comprehensive patient histories and detailed X-ray analysis, which call for specialized expertise and training [7].

In terms of data availability, hospital datasets, such as X-rays, vary in quality and completeness of information [8]. Sometimes, image quality can impact the speed of disease analysis for patients. Additionally, not all datasets have labels or annotations, making automated detection challenging [9]. Experts need to manually label these datasets, which is very time-consuming. Furthermore, when using Deep Learning model, a large amount of data is required, which is difficult to achieve in a short time [10] [11].

A useful strategy for overcoming dataset limitations is data augmentation. This technique creates synthetic data that introduces greater variation than the original dataset, expanding the data pool with a wider range of examples [12]. This increased diversity helps the model learn more effectively and generalize across various dataset conditions, including both high and low-quality data [13]. By simulating multiple potential scenarios, augmentation strengthens the model's ability to handle real-world data variability, while also reducing the reliance on large, high-quality labeled datasets. As a result, this process accelerates training and improves overall model performance [14] [15]. One of many augmentation method worth considering is Albumentation. It strives to achieve a balance between several key factors, delivering excellent performance across a wide range of transformations while offering a concise API and a flexible, extensible design [22].

This research aims to improve the accuracy of bone fracture detection using 'Albument-NAS'. The approach seeks to automatically identify and locate fractures in medical images, addressing challenges posed by resolution limitations that hinder accurate assessment of patient conditions. The research highlights the use of data augmentation to expand the variety of training data, which is anticipated to enhance the model's ability to recognize different patterns in medical images, ultimately leading to more accurate detection of bone fracture locations. Albument-NAS not only improves fracture detection accuracy but also has the potential to assist radiologists in real-time diagnostics, reducing diagnostic time and improving accessibility in under-resourced medical settings.

II. Literature Review

A. Research about Bone Fracture

Through data augmentation techniques, the recent study by Ju and Cai [7] improves YOLOv8's performance on the GRAZPEDWRI-DX dataset, which is a collection of pediatric wrist injury X-rays. With a state-of-the-art mean average accuracy (mAP@50) of 0.638, their suggested model

outperformed the original YOLOv8 model and an enhanced YOLOv7 model, which had respective scores of 0.634 and 0.636. This work contributes to the development of object detection models by demonstrating the importance of data augmentation in boosting YOLOv8's efficacy, especially for pediatric X-ray analysis.

Ahmed and Hawezi's research [16] addresses inaccuracies in bone fracture diagnoses due to blurry images from conventional X-ray scanners, which increase misdiagnosis risks. Their study aims to develop a machine learning-based system to assist surgeons in detecting fractures more accurately. Among several algorithms tested, the Support Vector Machine (SVM) model achieved the highest accuracy at 92.8%, followed by the Random Forest model with 85.7%, highlighting the potential of machine learning to improve diagnostic precision.

To meet the demand for quick, precise fracture diagnosis utilizing X-ray and CT images, Hareendranathan *et al*. [17] employed classification approaches to differentiate bone fractures from normal bone. Large data quantities and visual blurriness in MRI and CT scans make manual diagnosis difficult. The goal of this project was to develop an image-processing system that can classify fractures with speed and accuracy. On a dataset of 100 training and testing photos, the system demonstrated a high accuracy rate of 99.5%, indicating potential for effective fracture diagnosis.

B. Research about Data Augmentation

Su et al. [18] use Generative Adversarial Networks (GANs) for data augmentation in order to handle class imbalance and data scarcity. Particularly in minority classes, traditional GANs suffer from mode collapse or unequal distributions. They suggest Self-Transfer GAN (STGAN), a two-stage technique for producing varied 256×256 skin lesion pictures, as a solution to this problem. For a high-quality synthesis, STGAN first learns generic information from all classes and then blends it with information unique to each subject. When tested on the HAM10000 dataset, STGAN outperformed StyleGAN2 by up to 33% in terms of FID, Inception Score, Precision, and Recall. The STGAN framework shown efficacy for balanced classification with 98.23% accuracy, 88.85% sensitivity, 90.23% precision, 89.48% F1-score, and 98.34% specificity.

Cubuk et al. [19] introduce AutoAugment, an automated method for finding optimal data augmentation policies, enhancing image classifier accuracy. Unlike manual approaches, AutoAugment uses an automated search, achieving state-of-the-art accuracy on datasets like CIFAR-10, CIFAR-100, SVHN, and ImageNet. Notably, it achieved 83.5% Top 1 accuracy on ImageNet, surpassing the previous 83.1%, and reduced CIFAR-10's error rate to 1.5%, a 0.6% improvement.

The research by Elbattah *et al*. [8] presents an interesting study in the field of data augmentation. The data they aim to augment is a representation of eye-tracking known as scanpath. This research is intriguing because various aspects can be extracted from human eye movements, such as emotion recognition. The data augmentation method used is Variational Autoencoder (VAE) [20]. The results show that the accuracy of the model without augmentation increased from 67% to 70% with the use of augmentation.

III. Problem Statement

The challenges in bone fracture detection include issues with image quality and the expertise of doctors in analyzing X-ray scan results. Sometimes, the fractures are very small and almost imperceptible. This difficulty can certainly consume time when analyzing to determine the exact location of the fracture. There are many publicly available datasets that can be used for machine learning training, but not all of them are good quality. Some have poor image quality, some lack annotations for fracture locations, and others have excessive labels that provide unnecessary information for identifying fracture locations. Therefore, proper processing of the dataset is needed so that it can be easily learned by the model, allowing the model to capture patterns to predict the location of bone fractures.

IV. Methodology

Bone fracture detection in medical imaging plays a crucial role in various applications, particularly in supporting faster and more accurate diagnoses and treatments. This research aims to enhance bone fracture detection performance by applying data augmentation techniques within a one-shot detector model, namely as Albument-NAS method. In this chapter, the methodology employed in this research is described in detail. The process begins with the collection of medical image samples containing fractures, followed by data augmentation to increase data variability, and concludes with the implementation of the detection model architecture. Sample images and the overall procedure will be presented and thoroughly discussed.

A. Dataset

For this study, two datasets will be used: the COCO Bone Fracture Dataset and the GRAZPEDWRI dataset. The COCO dataset will be used for training and validation, while the GRAZPEDWRI dataset, similar to the dataset used by previous researchers Ju and Cai [3], will be used as test data to compare with the model applied in this research. Both datasets contain the same types of data—images and annotation files. However, the COCO dataset is in COCO format, while the GRAZPEDWRI dataset is in Pascal VOC format. For the COCO dataset, all images have a uniform size of 416x416, while for the GRAZPEDWRI dataset, the images vary in size. 908 images from COCO dataset will be augmented to varying dataset, and for testing will be using 300 images from GRAZPEDWRI dataset. For the image quality of each dataset, both datasets exhibit similar levels of variation in terms of brightness. Some images have high brightness, while others are relatively dim. The COCO Bone dataset (Fig. 1) has additional variation in terms of image color, with some images featuring a blue background, while the majority are grayscale. In contrast, the GRAZPEDWRI dataset consists entirely of grayscale images. Regarding data distribution, all the provided images depict fractures; there are no images of normal bones, and the specific types of fractures or patient details are not explained. This indicates that both datasets are specifically designed for fracture detection research.

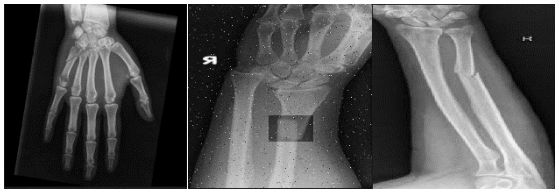


Fig. 1. COCO bone fracture dataset sample.

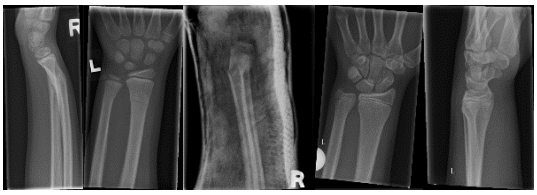


Fig. 2. GRAZPEDWRI dataset sample.

B. Preprocessing

Preprocessing is necessary to ensure that the images are prepared for training with the chosen deep learning model. Since the model to be used is YOLO, it is beneficial to resize all images to a uniform size. For this purpose, the target size is set to 640x640 to provide better clarity of the fracture areas. The image size of 640x640 was selected as it provides a good balance between computational efficiency and model accuracy. Larger image sizes can increase accuracy but also require more computational resources, while smaller sizes can reduce processing time but may miss finer details. The chosen size ensures that the model can detect fractures effectively while maintaining reasonable training and inference times.

The next step is to apply bounding box labels using the annotation files provided for the training and validation datasets. YOLO requires images to be pre-labeled for object detection tasks. The test data will also be labeled, but only for comparison and validation purposes to check whether the model's predicted bounding boxes are accurate. However, it appears that in the GRAZPEDWRI dataset (Fig. 2), there are some labels that are not necessary for locating fractures. Therefore, label extraction will be performed to retain only the fracture location bounding boxes, while other labels will be disregarded.

C. Albumentation

The study by Ju & Cai [7] provided brightness and exposure enhancement augmentation for each image. This research will do the same, but with the addition of several other augmentations. Augmentation will be carried out using the Albumentations method provided by Roboflow, allowing the author to directly utilize the augmented dataset. As a result, Roboflow produced 2,360 augmented images to be used as the training and validation dataset. What sets Albumentations apart from typical augmentation techniques is its ability to apply augmentations probabilistically. This means that augmentations are applied randomly to images based on predefined constraints and probability values.

Brightness in an image refers to the overall light intensity or luminance, which affects how light or dark an image appears. Increasing brightness makes an image appear lighter, while decreasing it makes it darker. The brightness of each pixel can be adjusted by adding a constant value to the pixel's intensity, represented by Formula 1:

$$I' = I + \beta \tag{1}$$

where I is the original intensity of the pixel, I' is the adjusted intensity, and β is a constant that controls the level of brightness.

Exposure in an image relates to the amount of light captured by the camera sensor, influencing the brightness and detail visible in the image, especially in the highlights and shadows. Proper exposure ensures that an image retains detail without being overly bright (overexposed) or too dark (underexposed). Exposure adjustments can be achieved by scaling pixel intensity, often represented by Formula 2:

$$I' = I \times \alpha \tag{2}$$

where I is the original pixel intensity, I' is the adjusted intensity, and α is the multiplier that controls the exposure level. Values of $\alpha > 1$ will increase exposure, brightening the image, while $0 < \alpha < 1$ reduces exposure, darkening it.

The final augmentation performed by the author is the addition of noise. In this case, the applied noise is salt and pepper noise. Salt and pepper noise is a type of image noise characterized by random occurrences of white (salt) and black (pepper) pixels throughout the image, creating isolated bright and dark spots that disrupt the image's smooth appearance. The effect of salt and pepper noise is typically applied by randomly setting a percentage of pixels to the minimum intensity (0, representing black) or maximum intensity (255, representing white) in an 8-bit grayscale image. Mathematically, it can be represented as Formula 3:

$$I(x,y) = \begin{cases} 0 \; with \; probability \; p_s \\ 255 \; with \; probability \; p_p \\ I(x,y) \; otherwise \end{cases} \tag{3}$$

pixel intensity, p_s is the probability of salt noise, and p_p is the probability of pepper noise. The addition of this noise enriches the dataset quality and trains the model to learn from various unpredictable dataset conditions. Fig. 3 shows the sample of augmented images.

Fig. 3. Augmented images sample.

Fig. 4 shows the overall architecture and functionality of the proposed work, which includes augmentation with Roboflow Albumentation and object detection using YOLO-NAS creating author proposed method, Albument-NAS, while Algorithm I shows the pseudocode for the proposed method.

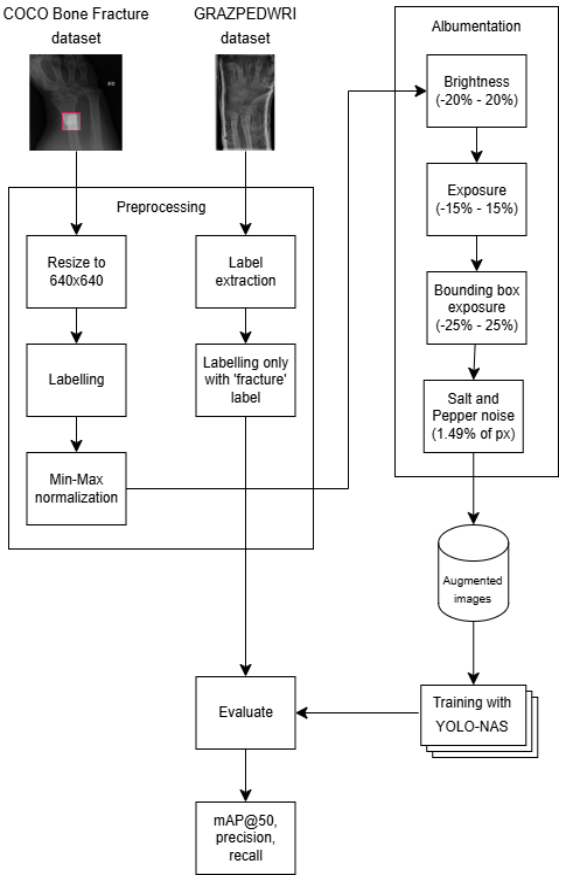


Fig. 4. Overall procedure of albument-NAS.

Algorithm 1: Albument-NAS method: Albumentation augmentation and YOLO-NAS detection

Inputs: COCO Bone Fracture dataset
Outputs: X-Ray images with bounding box labelled
1. Initialize image
2. Initialize Albumentation technique
3. Compute augmentation into images
4. Initialize YOLO-NAS model and hyperparameter
5. While (training not converged or early stopping not reached) do
6. Load augmented images and their bounding box label
7. For (every image in batch) do
8. For (every region in the image with a bounding box) do
9. Apply YOLO-NAS detection
10. Update
11. Adjust model parameter based on accuracy and loss
12. End
13. End
14. End
15. Evaluate mAP@50, precision, recall
16. End

D. YOLO-NAS Model

For the object detection model, the author used YOLO-NAS. YOLO-NAS is an object detection model developed by Deci AI using neural architecture search (NAS) techniques to automatically design the network architecture. This allows the model to find the optimal settings for recognizing objects in images or videos without requiring extensive human input. The advantage of YOLO-NAS over other YOLO models is its higher accuracy, faster speed, and better efficiency, making it ideal for use on low-power hardware such as mobile devices. YOLO-NAS is based on the modified CSPNet backbone architecture and uses the YOLOv5 detection head. The combination of the efficient CSPNet backbone, the powerful YOLOv5 detection head, and advanced NAS enables YOLO-NAS (Fig. 5) to achieve exceptional accuracy and speed.

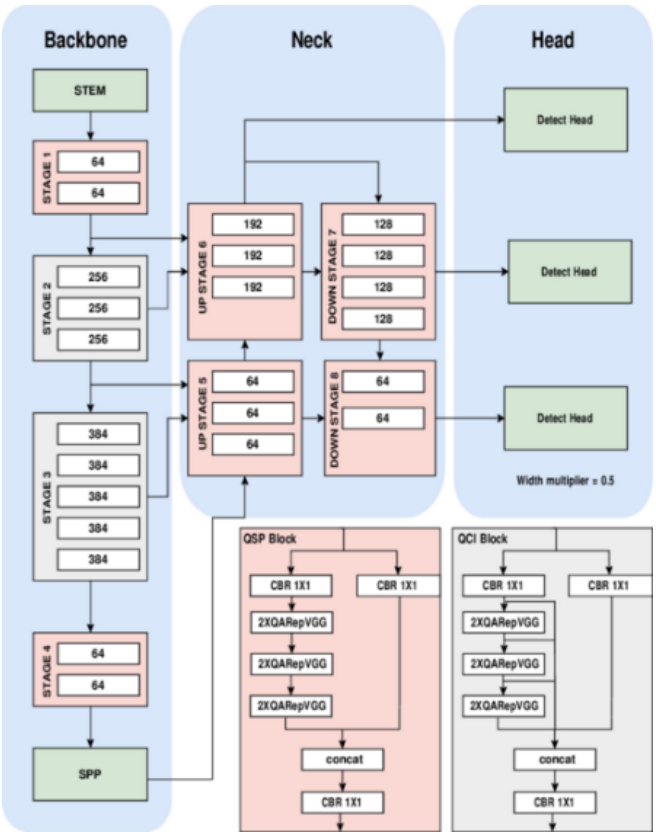


Fig. 5. YOLO-NAS architecture.

As for the backbone using CSPNet, Formula 4 shows how the backbone works:

$$F_{backbone}(I) = FeatureExtraction(I) \quad (4)$$

where I is the input image, and $F_{backbone}(I)$ produces the feature used by detection head. YOLOv5 detection head is used to predict B_{pred}, which is the bounding box and object class for image I. The function can be seen in Formula 5 below.

$$B_{pred} = YOLOv5(F_{backbone}(I)) \quad (5)$$

For the optimal network architecture searching, YOLO-NAS uses NAS (Neural Architecture Search) method to optimize the model's architecture by selecting the best components such as the number of layers, the number of neurons, and connections between layers.

$$A^* = \arg max\ L(A, D) \quad (5)$$

where A is the network architecture being searched, $L(A, D)$ is the loss function for the architecture A on dataset D and A^* is the optimal architecture that maximizes performance.

YOLO-NAS also compute loss function, typically combines classification loss and bounding box regression loss. Formula 6 shows how YOLO-NAS handle loss function:

$$L_{total} = L_{cclassification} + L_{bounding\_box} + L_{objectness} \quad (6)$$

$L_{classification}$ measures the error in classifying object class, $L_{bounding\_box}$ measures the error in predicting the bounding box location, and $L_{objectness}$ measures the error in predicting the probability of the object presence in a grid.

V. Result and Discussion

This research is implemented in Python 3.10. The environment used for this study, including the hardware and software specifications, is outlined in Table I. These specifications define the necessary resources for implementing the proposed object detection system effectively. The hardware setup provides sufficient processing power, while the software components offer the required tools and libraries to support model training and evaluation.

TABLE I. Research Environment Details

Type	Device Type	Name
Hardware	Processor	13th Gen Intel(R) Core™ i7-13620H, 2400 Mhz
	Graphics processor	NVIDIA GeForce RTX 4060 Laptop GPU
	Storage	1 TB
	Memory	16 GB
Software	IDE	Visual Studio Code
	Albumentation	Roboflow
	Important libraries	cpython, cython, ipython, ipykernel, jedi, numpy, onnx, pillow, pytorch (make sure it is CUDA supported), scipy, super-gradients (YOLO model), supervision, torchvision, torchaudio

This research will observe three key metrics employed to evaluate the some of the object detection model's performance. By focusing on these metrics, a comprehensive understanding of the model's detection precision and overall effectiveness can be observed.

A. Metrics

To assess how well the model's bounding box predictions identify fracture locations, the mAP@50 metric will be analyzed. A higher mAP value indicates that the model is more successful in accurately placing bounding boxes that match the ground truth. Additionally, precision will be calculated to evaluate how accurately the proposed model predicts the correct class, and recall will be used to determine how many fracture images are correctly identified as fractures.

TABLE II. Model Comparison on GRAZPEDWRI Dataset

Model	Metric		
	mAP	***Precision***	***Recall***
YOLOv8 (SOTA) [3]	63.8%	-	-
Baseline (ResNet50)	71.4%	80.5%	61.3%
Baseline + Albumentation	76.2%	**88.2%**	63.6%
Faster R-CNN	73.3%	81.4%	62.3%
Faster R-CNN + Albumentation	78.3%	83.4%	64.3%
YOLO-NAS	74.3%	79.2%	59.8%
Albument-NAS (proposed method)	**83.5%**	87%	**65.7%**

Table II compares performance metrics across different models, focusing on mAP, precision, and recall. YOLOv8 (SOTA) achieves a mAP of 63.8%, though Precision and Recall metrics are not available. The baseline model (ResNet50) attains a mAP of 71.4%, with a precision of 80.5% and a Recall of 61.3%, demonstrating strong true positive accuracy but moderate recall. With Albumentation augmentation, the Baseline model improves to a mAP of 76.2%, while Precision rises to 88.2% and recall to 63.6%, indicating a notable increase in true positive identification. Faster R-CNN achieves a mAP of 73.3%, with a Precision of 81.4% and a Recall of 62.3%. This result highlights its strong precision and balanced performance in recall, surpassing the Baseline model in mAP. When augmented with Albumentation, Faster R-CNN improves further to a mAP of 78.3%, with a Precision of 83.4% and a Recall of 64.3%. This demonstrates that augmentation enhances both precision and recall, solidifying Faster R-CNN as a robust model for object detection tasks. YOLO-NAS without augmentation achieves a mAP of 74.3%, with a Precision of 79.2% and a Recall of 59.8%, slightly below the Baseline with augmentation. However, the proposed method, Albument-NAS, achieves the highest mAP at 83.5%, with a precision of 87% and recall of 65.7%, demonstrating that augmentation significantly enhances both accuracy and completeness of detections across all metrics. Precision and recall values for YOLOv8 were not available due to limitations in the original study's reporting. This metric should be considered when comparing the performance of YOLO-based models. The inclusion of Faster R-CNN results underscores its competitive performance, especially when paired with augmentation, although it does not surpass the Baseline with augmentation or Albument-NAS in overall accuracy.

The results demonstrate that an appropriate augmentation process can enhance object detection accuracy, specifically in identifying fracture locations. The YOLO-NAS model also proves superior due to its ability to optimize parameters during training. Consequently, the architecture of YOLO-NAS achieves greater convergence compared to YOLOv8, whose structure is already pre-defined and fixed. The ResNet50 model achieves higher precision because its architecture is optimized to focus on fine-grained feature extraction [21], which reduces

false positives. ResNet50's deep residual connections help capture subtle details and ensure that detected regions are more likely to correspond accurately to true fractures, thus improving precision. This focus makes ResNet50 particularly effective in tasks where high specificity (true positive accuracy) is crucial, even though it might not reach the same level of overall recall as YOLO-based models optimized for real-time, comprehensive detection. Additionally, Faster R-CNN achieves higher precision compared to ResNet50 due to its two-stage detection mechanism, which separates region proposal from classification. This architecture allows Faster R-CNN to focus on high-quality region proposals, reducing the likelihood of false positives during object classification.

The YOLO-NAS model without augmentation performed below the baseline augmented model due to the limited diversity of the training data. Data augmentation significantly improves model generalization by simulating various real-world conditions, thus enhancing its ability to detect fractures across different image qualities.

B. Bone Fracture Detection

To ensure a fair comparison in bone fracture detection, the detection results will be evaluated using the GRAZPEDWRI dataset, which serves as the testing dataset. The object detection results can be seen in Fig. 6 below.

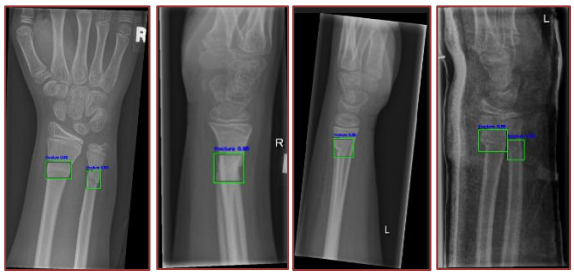


Fig. 6. Fracture detection result by proposed method, albument-NAS.

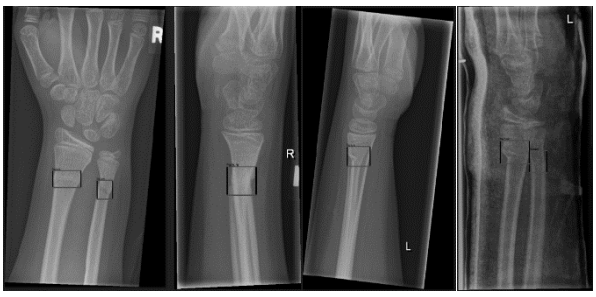


Fig. 7. Fracture detection ground truth.

Based on Fig. 6, Albument-NAS successfully detects bone fractures accurately. Fig. 7 displays the actual locations of the fractures. Despite the inconsistent quality of the X-ray images (with the fourth image in Fig. 7 appearing darker than the others), the model still performs well in making inferences. These variations are largely attributed to challenges in detecting small or subtle fractures in images with low resolution or significant noise. Nonetheless, the model demonstrates strong performance in accurately identifying larger fractures.

The Albument-NAS model has the potential to be incorporated into clinical workflows via cloud-based diagnostic systems, providing remote access for healthcare professionals in underprivileged regions. Additionally, its capability to swiftly and accurately detect fracture locations can help alleviate the workload of radiologists, especially in emergency scenarios where prompt action is essential.

C. Grad-CAM Visualization

Grad-CAM (Gradient-weighted Class Activation Mapping) can help radiologists and healthcare providers understand how the model identifies fracture locations by generating visual heatmaps that highlight the regions of an image most influential to the model's decision. When applied to X-ray images, Grad-CAM produces an overlay that indicates which areas the model "focused on" to predict the presence of a fracture. Radiologists are more inclined to embrace AI systems when they can confirm that the model's decision-making process aligns with their clinical knowledge and judgment. By providing visual insights into how predictions are generated, Grad-CAM facilitates a collaborative dynamic between AI and healthcare providers, positioning the AI as a supportive tool rather than a replacement. This clarity not only strengthens trust but also encourages broader adoption of AI solutions in clinical settings by ensuring patient safety and adhering to medical decision-making standards. As for this research, Fig. 8 shows how Grad-CAM can clarify the model's prediction results by focusing on the fracture location.

Fig. 8. Grad-CAM visualization.

VI. Conclusion and Future Work

The results of this study demonstrate that appropriate data augmentation, specifically with Albumentation, significantly improves object detection performance in identifying fracture locations. The Albument-NAS model achieved the highest performance in terms of mAP@50, Precision, and Recall when augmented, suggesting its advantage over other models due to its ability to dynamically optimize parameters during training. This adaptability allows YOLO-NAS to achieve greater convergence, whereas YOLOv8, with a more rigid architecture, is less flexible in parameter optimization. Meanwhile, the ResNet50 model, although not reaching the highest recall, excels in precision due to its residual architecture, which effectively captures detailed features, thereby reducing false positives. This characteristic makes ResNet50 particularly useful in applications requiring high specificity and accuracy. The reason behind the significant drop in mAP@50 for the SOTA model is that it was designed for multi-object detection, requiring accurate predictions across a large number of classes. However, this approach is less relevant to the original goal, which is to precisely determine the location of bone fractures. Although YOLO-based models like YOLOv5 and YOLOv8

perform well in general object detection tasks, they frequently encounter challenges in detecting small objects like bone fractures due to resolution constraints and the high precision requirements of medical imaging. The Albument-NAS model overcomes these challenges by utilizing advanced augmentation methods and refining its architecture to enhance the detection of fine-grained fracture details.

Future research could further investigate the impact of different types of augmentation techniques on fracture detection, especially in enhancing model robustness across diverse and inconsistent image qualities. Additionally, integrating hybrid architectures that combine the fine-grained feature extraction of ResNet50 with the dynamic optimization abilities of YOLO-NAS could yield a model with both high precision and recall. Finally, evaluating these models on a broader dataset of fracture types or on 3D imaging data might enhance their applicability and reliability in clinical diagnostics. Another hybrid method is combining YOLO-NAS with transformer-based models like DETR (DEtection TRansformers) could enhance the model's capability to identify intricate fracture patterns. This approach would capitalize on YOLO-NAS's spatial efficiency and the transformers' ability to capture contextual information.

REFERENCES

[1] Y. Bao, "Medical image super-resolution reconstruction: A comprehensive investigation of Generative Adversarial Networks," Appl. Comput. Eng., vol. 51, pp. 14–19, 2024. doi: 10.54254/2755-2721/51/20241148.

[2] F. D. Pérez Cano, G. Parra-Cabrera, I. Vilchis-Torres, J. Reyes-Lagos, and J. Jimenez-Delgado, "Exploring Fracture Patterns: Assessing Representation Methods for Bone Fracture Simulation," J. Pers. Med., vol. 14, p. 376, 2024. doi: 10.3390/jpm14040376

[3] S. Tippannavar, Y. S. D., G. Shivakumar, and E. Madappa, "Unveiling the Spectrum: Versatile Image Processing Techniques in Bone Fracture Detection - A Comprehensive Review," J. Artif. Intell. Capsule Networks, vol. 5, pp. 499–520, 2023. doi: 10.36548/jaicn.2023.4.004

[4] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," Front. Public Health, vol. 11, 2023. doi: 10.3389/fpubh.2023.1273253

[5] D. Yadav et al., "Hybrid SFNet Model for Bone Fracture Detection and Classification Using ML/DL," Sensors, vol. 22, p. 5823, 2022. doi: 10.3390/s22155823

[6] A. Makhlouf, M. Maayah, N. Abughanam, and C. Catal, "The use of generative adversarial networks in medical image augmentation," Neural Comput. Appl., vol. 35, 2023. doi: 10.1007/s00521-023-09100-z.

[7] R.-Y. Ju and W. Cai, "Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm," Sci. Rep., vol. 13, 2023. doi: 10.1038/s41598-023-47460-7.

[8] K. T. Krishnan, L. Kavyaa, and J. Sugumar, "Automated Bone Fracture Detection Using Convolutional Neural Network," J. Phys. Conf. Ser., vol. 2471, 2023. doi: 10.1088/1742-6596/2471/1/012003

[9] K. Dimililer, "IBFDS: Intelligent bone fracture detection system," Procedia Comput. Sci., vol. 120, pp. 260–267, 2017. doi: 10.1016/j.procs.2017.11.237.

[10] I. Sary, S. Andromeda, and E. Armin, "Performance Comparison of YOLOv5 and YOLOv8 Architectures in Human Detection using Aerial Images," Ultima Comput.: J. Syst. Komput., pp. 8–13, 2023. doi: 10.31937/sk.v15i1.3204.

[11] A. Mumuni, F. Mumuni, and N. Gerrar, "A Survey of Synthetic Data Augmentation Methods in Machine Vision," Mach. Intell. Res., 2024. doi: 10.1007/s11633-022-1411-7.

[12] A. H. Basori, S. Malebary, and S. Alesawi, "Hybrid Deep Convolutional Generative Adversarial Network (DCGAN) and Xtreme Gradient Boost for X-ray Image Augmentation and Detection," Appl. Sci., vol. 13, p. 12725, 2023. doi: 10.3390/app132312725.

[13] E. Saraswathi and J. Banu, "Hybrid CGAN-based plant leaf disease classification using OTSU and surf feature extraction," Neural Comput. Appl., pp. 1–13, 2024. doi: 10.1007/s00521-024-09812-w.

[14] M. Zhu, C. Liu, and T. Szirányi, "A Global Multi-Temporal Dataset with STGAN Baseline for Cloud and Cloud Shadow Removal," in Proc. Int. Conf. Image Process. Theory, Tools Appl. (IPTA), pp. 206–212, 2023. doi: 10.5220/0012039600003497.

[15] O. Ezeme, Q. Mahmoud, and A. Azim, "Design and Development of AD-CGAN: Conditional Generative Adversarial Networks for Anomaly Detection," IEEE Access, vol. 8, pp. 1–16, 2020. doi: 10.1109/ACCESS.2020.3025530.

[16] K. D. Ahmed and R. Hawezi, "Detection of bone fracture based on machine learning techniques," Measurement: Sensors, vol. 27, p. 100723, 2023. doi: 10.1016/j.measen.2023.100723.

[17] Hareendranathan et al., "Deep Learning Approach for Automatic Wrist Fracture Detection Using Ultrasound Bone Probability Maps," SN Compr. Clin. Med., vol. 5, 2023. doi: 10.1007/s42399-023-01608-8.

[18] Q. Su, H. Hamed, M. Isa, X. Hao, and X. Dai, "A GAN-Based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification," IEEE Access, 2024. doi: 10.1109/ACCESS.2024.3360215.

[19] E. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. Le, "AutoAugment: Learning Augmentation Strategies From Data," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 113–123, 2019. doi: 10.1109/CVPR.2019.00020.

[20] M. Elbattah, C. Loughnane, J.-L. Guerin, R. Carette, F. Cilia, and G. Dequen, "Variational Autoencoder for Image-Based Augmentation of Eye-Tracking Data," J. Imaging, vol. 7, p. 83, 2021. doi: 10.3390/jimaging7050083.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.

[22] Buslaev, Alexander & Iglovikov, Vladimir & Khvedchenya, Eugene & Parinov, Alex & Druzhinin, Mikhail & Kalinin, Alexandr. (2020). Albumentations: Fast and Flexible Image Augmentations. Information. doi: 11. 125. 10.3390/info11020125.

FKMU: K-Means Under-Sampling for Data Imbalance in Predicting TF-Target Genes Interactions

Thanh Tuoi Le[1, 2], Xuan Tho Dang[3]*
Faculty of Information Technology, Hanoi National University of Education, Hanoi City, Vietnam
Faculty of Information Technology, Vinh University of Technology Education, Vinh City, Vietnam
Academy of Policy and Development, Hanoi City, Vietnam

***Abstract*—Identifying interactions between transcription factors (TFs) and target genes is critical for understanding molecular mechanisms in biology and disease. Traditional experimental approaches are often costly and not scalable. We introduce FKMU, a K-means-based under-sampling method designed to address data imbalance in predicting TF-target interactions. By selecting low-frequency TF samples within each cluster and optimizing the balance ratio to 1:1 between known and unknown samples, FKMU significantly improves prediction accuracy for unobserved interactions. Integrated with a deep learning model that uses random walk sampling and skip-gram embeddings, FKMU achieves an average AUC of 0.9388 ± 0.0045 through five-fold cross-validation, outperforming state-of-the-art methods. This approach facilitates accurate and large-scale predictions of TF-target interactions, providing a robust tool for molecular biology research.**

***Keywords*—K-means clustering; imbalanced data; TF-target gene interactions; heterogeneous network; meta-path**

I. Introduction

Transcription Factors (TFs) are essential regulatory proteins in the process of gene transcription, which is the mechanism of transferring genetic information from DNA to RNA [8]. TFs perform their role by binding to specific DNA sequences, often located in or near gene promoters. Upon binding to DNA, TFs can either activate or inhibit the function of RNA polymerase, the enzyme responsible for transcribing DNA into RNA. Through this mechanism, TFs regulate gene expression, playing a crucial role in the development and maintenance of cellular functions. TFs are found in almost all living organisms and are vital for gene expression regulation. However, when TFs lose their function, the balance in gene regulation is disrupted, leading to severe diseases. Accurately identifying the relationships between TFs and target genes is a crucial step in understanding the complex molecular mechanisms involved in biological and pathological processes. These insights will pave the way for extensive research in molecular biology and applied medicine, laying the foundation for more effective diagnostic and therapeutic methods in the future.

Previously, identifying interactions between TFs and target genes relied primarily on experimental methods, which were costly and time-intensive. The emergence of large-scale techniques such as ChIP-seq and RNA-seq has made it more feasible to predict TF target genes across the entire genome [12, 13]. ChIP-seq maps TF-DNA interactions, while RNA-seq provides RNA expression data, shedding light on genes influenced by TFs [20]. However, these methods reveal only a small fraction of the complex gene regulatory network.

Many interactions between TFs and target genes remain unclear in existing databases. Datasets on TF-target gene interactions collected from ChIP-seq techniques provide a limited view of the complex gene regulatory network. Specifically, most current computational methods only identify binding sites without addressing the nature of these interactions. Although some recent studies have made progress in predicting these interactions, building high-quality datasets with both positive and negative samples remains a significant challenge. Furthermore, current methods often fail to effectively address the data imbalance issue, particularly in selecting negative samples. This limitation can result in the failure to detect potential interactions between TFs and target genes, reducing the accuracy of prediction models. Additionally, failure to address the data imbalance issue can introduce bias during the training process [2-4], impairing the ability to detect important interactions in the gene regulatory network. Therefore, the development of new methods focused on data balancing is crucial to improve prediction performance and provide a solid foundation for molecular biology and applied medicine research.

This paper introduces a novel approach to address data imbalance for improving the prediction of TF-target gene interactions. Key contributions of this study include:

a) We present the FKMU method, an under-sampling technique based on the K-means clustering algorithm and the inverse information principle, designed to enhance efficiency and stability in predicting TF-target gene interactions.

b) A novel meta-path schema has been developed to extend the capability of capturing potential links within heterogeneous networks, significantly improving the predictive performance of the model.

c) The FKMU model incorporates substantial advancements, achieving superior performance in identifying unknown TF-target gene interactions compared to existing approaches.

d) The effectiveness of the proposed method is validated through rigorous experiments, demonstrating outstanding results in terms of accuracy and predictive efficiency over current methods.

e) Experimental results confirm that FKMU is an effective and accurate solution, achieving an average AUC value

*Corresponding Author

superior to many existing methods, demonstrating its potential for widespread application in molecular biology research.

The remainder of this paper is organized as follows. Section II reviews related works on predicting interactions between TFs and target genes based on TF binding sites, gene expression data, and heterogeneous networks. Section III introduces the FKMU method, which combines K-means clustering and negative sampling. Section IV presents the experimental results, including evaluation metrics, parameter optimization, and performance comparisons. Finally, Section V concludes with the contributions of the study and suggests potential directions for future development.

II. Related Works

Predicting interactions between TFs and target genes is a critical topic in the field of computational biology. Traditional experimental methods are often time-consuming, costly, and challenging to apply at scale, while also carrying the risk of failure. Artificial intelligence offers a powerful tool to support these experimental approaches, helping to narrow down the search for potential interactions between TFs and target genes and optimizing them for subsequent experimental validation. As a result, research time and costs can be significantly reduced, facilitating the research and development process. Research related to our approach can be divided into the following three subsections.

1) Methods based on predicting transcription factor binding sites (TFBS): These methods primarily focus on identifying interactions between TFs and target genes by detecting their binding sites. The process involves determining the binding positions of TFBS, which are often integrated with deep learning models such as Convolutional Neural Networks (CNNs), as demonstrated in the research by S. Salekin et al. [21] and Ž. Avsec et al. [29], or Recurrent Neural Networks (RNNs), as referenced in the studies by J. Lanchantin et al. [10] and Z. Shen et al. [32]. However, these methods have a significant limitation, leading to a high false positive rate because TFBS are often located within long non-coding sequences. Furthermore, they do not directly predict TF-target gene interactions but rather infer them based on the locations of TFBS.

2) Direct prediction methods for TF-target gene interactions based on gene expression data: These methods do not rely on TFBS but instead use gene expression data, such as gene expression images from in situ hybridization (ISH) or single-cell RNA sequencing (scRNA-seq) data, to directly predict the relationship between TFs and target genes. For example, using gene expression image analysis, Y. Yang et al. [28] developed GripDL, an effective tool for studying transcriptional regulatory networks in Drosophila. GripDL utilizes ISH images as input, combined with a deep residual model to leverage known TF-target gene interactions. Results showed that GripDL outperformed traditional methods in accuracy and the ability to detect novel gene interactions, offering valuable insights into eye development in Drosophila and paving the way for new research on gene regulatory networks. Beyond gene expression images, single-cell RNA sequencing (scRNA-seq) data provides an additional perspective for understanding complex mechanisms by which TFs regulate target genes. Su et al. [15] developed NetAct, a computational platform for constructing transcription factor regulatory networks using transcriptomic data and gene databases. This tool has been effectively applied to model regulatory networks in epithelial-mesenchymal transition and macrophage polarization, highlighting its significant potential in analyzing complex gene networks. Y. Fan et al. [26] introduced the 3D Co-Expression Matrix Analysis (3DCEMA) method, employing 3D convolutional neural networks to predict regulatory relationships between genes. This approach helps minimize the effects of noise and data loss, significantly enhancing the accuracy of gene regulatory network inference compared to existing algorithms.

However, the main drawback of these methods is the high cost of data collection, particularly for complex gene expression data like scRNA-seq, which limits their widespread practical application.

3) Heterogeneous network-based methods: With the rapid development of databases, a wealth of data on TF-target gene interactions has been collected from experiments and integrated into resources like the TRRUST database [7], providing extensive insights into human gene regulatory networks. Heterogeneous network-based methods offer a novel approach to directly predict TF-target gene interactions more effectively than TFBS or gene expression-based methods. These methods go beyond simply predicting binding sites by leveraging contextual biological factors and disease mechanisms that influence binding.

For instance, Y. A. Huang et al. [24] introduced a new deep learning model named HGETGI to predict TF-target gene interactions. HGETGI not only learns known interaction patterns between TFs and target genes but also integrates information on their roles in human pathological mechanisms. Using random walk sampling with meta-paths and skip-gram node embedding techniques, HGETGI achieved high prediction accuracy, with an average AUC of 0.8519 ± 0.0731 through five-fold cross-validation. Similarly, Z. H. Du et al. [30] proposed the GraphTGI model, which employs a graph-structured neural network to predict TF-target gene interactions, achieving an average AUC of 88.64% through five-fold cross-validation, proving its effectiveness in TF-target gene interaction prediction. GraphTGI is the first end-to-end model to incorporate the topological structure of the TF-target gene interaction network, alongside the chemical properties of genes in node features, creating automated embeddings that clarify relationships between TF-target gene pairs and support related tasks.

While these methods have made significant strides, challenges remain. Current approaches mainly focus on predicting TF-target gene interactions without optimizing for data balance, which hampers accurate predictions with uneven datasets.

This paper introduces a novel approach to address data imbalance in order to optimize the prediction of TF-target gene interactions. Numerous studies have proposed solutions for handling imbalanced data classification through various approaches, including data-level and algorithm-level strategies. In this study, we adopt a data-level approach, focusing on preprocessing to reduce imbalance before feeding data into the TF-target gene interaction prediction model to achieve better results. Several methods exist for adjusting data, such as oversampling or under-sampling. Moreover, combining these methods can further optimize classification and improve prediction performance [9].

Under-sampling is an effective method for handling imbalanced data by reducing the number of samples in the majority class to balance it with the minority class, thereby improving the predictive capability of the model. Among the under-sampling methods, Random Under-sampling (RUS) is a simple technique applied to balance datasets by randomly removing a number of samples from the majority class. However, such random data removal may lead to the loss of valuable samples and diminish the amount of useful information from the majority class, potentially negatively impacting performance in classification tasks. Therefore, C. M. Huang et al. [1] proposed an approach using K-means to select representative samples from the majority class, improving precision (PPV) by 20.2% while maintaining recall above 90% on Kawasaki Disease (KD) data. Q. Zhou et al. [19] suggested an adaptive K-means-based under-sampling method, where they calculate the distance between data points within each cluster and the cluster centroid using Manhattan distance and Cosine similarity. This algorithm employs these two metrics to select representative samples from the majority class, resulting in a more balanced dataset. The results indicate that this method determines an appropriate dynamic value of k for different datasets and generates a balanced dataset, thereby enhancing the classification performance of machine learning algorithms. T. Doan et al. [22] proposed GBDTLRL2D, a method for predicting lncRNA-disease relationships that combines Gradient Boosting Decision Trees (GBDT) and Logistic Regression, utilizing MetaGraph2Vec and K-means to preserve semantic features, achieving an average AUC of 0.98 in 10-fold cross-validation.

In contrast to under-sampling methods, over-sampling methods focus on increasing the number of samples in the minority class. The simplest over-sampling method is Random Oversampling, which involves randomly duplicating samples from the minority class to increase the number of samples in this class, thereby creating a balance with the majority class. However, this approach can easily lead to overfitting, reducing the generalization ability of the model. To mitigate this risk, N. V. Chawla et al. [17] proposed SMOTE, which generates synthetic samples through interpolation from the minority class data. However, SMOTE may not accurately reflect the complex characteristics of the minority class, especially in intricate models. D. X. Tho et al. [6] improved this approach with KNN-SMOTE, achieving superior performance in F-score, G-mean, and AUC on imbalanced datasets from UCI. H. Li et al. [9] proposed KM-GAN, which combines K-means and GAN to generate new samples from imbalanced industrial fault data. KM-GAN clusters the minority class samples and then utilizes GAN to create additional data, enhancing diagnostic efficacy through a combined DNN and DBN model, thus addressing the bias of traditional methods towards the majority class.

Current methods have effectively contributed to identifying interactions between TFs and target genes; however, they still face several challenges, such as a high false positive rate and significant data collection costs. Additionally, heterogeneous network-based methods have not been optimized for imbalanced data. Although many methods for handling imbalance, as introduced above, have achieved good predictive performance, further improvements are still necessary. To address these limitations and enhance the quality of the majority class, we have developed a new method called FKMU to improve the performance of predictive models. We anticipate that this method will enhance accuracy and applicability in empirical research.

III. Methodology

In this section, we will outline the main tasks of our method aimed at predicting the relationships between TFs and target genes. As part of this narrative, we describe a heterogeneous network formed from biological databases related to TFs, target genes, and diseases, as shown in Step 1 of Fig. 1. We will perform data balancing by combining the K-means clustering algorithm with negative sampling, as detailed in Step 2 of Fig. 1. We will create new meta-paths, as illustrated in Step 3 of Fig. 1. Random walks will be conducted on the graph according to the meta-paths to generate training data for the embedding model, followed by the application of a deep learning model to learn the features of the nodes in the heterogeneous network, as shown in Steps 4 and 5 of Fig. 1. Finally, we will proceed to predict the interactions between TFs and target genes, as illustrated in Step 6 of Fig. 1.

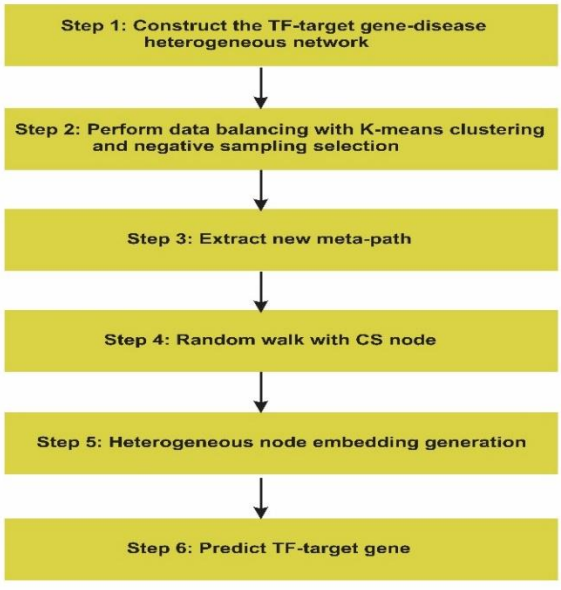


Fig. 1. General workflow containing six main steps.

A. Heterogeneous Network Construction

Definition 1. A heterogeneous network [23] is defined as graph G = (V, E, T), where each node v and each edge e are associated with mapping functions $\phi(v): V \rightarrow T_V$ and $\varphi(e): E \rightarrow T_E$, respectivelt. T_V and T_E represent the set of object types and relationship types, and satisfy the condition $|T_V| + |T_E| > 2$.

In this study, we construct a model to predict the associations between TFs and target genes. This heterogeneous network is defined as a graph G = (V, E, T), where each node represents TFs, target genes, or diseases, and each edge represents the relationships between these entities. TFs and target genes have been shown to have close associations with various diseases, and integrating information about these entities allows us to explore potential unknown associations between TFs and target genes.

B. Meta-Path in a Heterogeneous Network

A meta-path, also known as a "hyperlink" is a model used to represent relationships between nodes in a heterogeneous network. It can be understood as a sequence of connections between nodes and their links, designed to express the relationship between two nodes under consideration within the network.

Definition 2. Meta-path [5]. A meta-path $\mathcal{P}$ is a path defined on the network schema $T_G = (\mathcal{A}, \mathcal{R})$ and is represented as $V_1 \xrightarrow{R_1} V_2 \xrightarrow{R_2} \dots \xrightarrow{R_l} V_{l+1}$, defining a composite relationship $R = R_1 \circ R_2 \circ \dots \circ R_l$ between the types V_1 and V_{l+1}, where $\circ$ denotes the composition operator over relationships.

For example, the meta-path "CTCF (TF) - BRCA1 (Target Gene) - Breast Cancer (Disease) - TP53 (Target Gene) - MYC (TF) - Ovarian Cancer (Disease) - EGFR (Target Gene) - SP1 (TF)", as illustrated in Fig. 2, demonstrates how the transcription factors CTCF, MYC, and SP1 are linked to breast and ovarian cancers through intermediate target genes BRCA1, TP53, and EGFR.

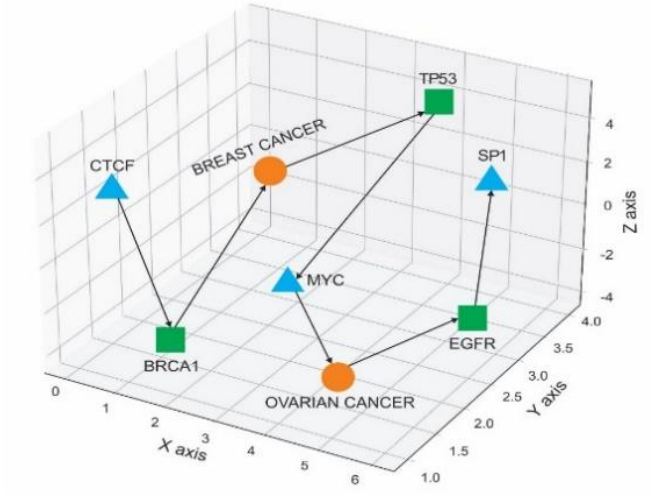


Fig. 2. Illustration of a meta-path in a heterogeneous TF-target gene-disease network.

1) Meta-path random walks: is a graph mining technique that generates paths based on the semantic and structural relationships between different types of nodes. This method helps transform the complex structure of the network into vectors, enabling effective extraction of information from the relationships.

For a heterogeneous network *G = (V, E, T)* and a meta-path schema $\mathcal{P}$, we can calculate the transition probability at step *k* as follow:

$$P(v^{k+1} | v_t^k, \mathcal{P}) = \begin{cases} \frac{1}{|N_{t+1}(v_t^k)|} & (v^{k+1}, v_t^k) \in E, \emptyset(V^{k+1}) = t+1 \\ 0 & (v^{k+1}, v_t^k) \in E, \emptyset(V^{k+1}) \neq t+1 \\ 0 & (v^{k+1}, v_t^k) \in E \end{cases} \quad (1)$$

where $v_t^k \in V_t$ and $N_{t+1}(v_t^k)$ denotes the type V_{t+1} of the neighborhood of node v_t^k.

C. Dataset

In this study, we use a dataset consisting of three types of nodes: TFs, target genes, and diseases, along with three types of relationships between these nodes [24]. Specifically, the three types of relationships include: the association between TFs and target genes, the association between TFs and diseases, and the association between target genes and diseases (Fig. 3).

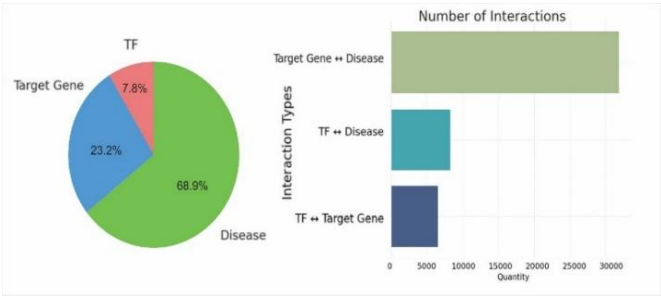


Fig. 3. Statistics of the heterogeneous TF-target gene-disease network information.

Data on interactions between human TFs and target genes were collected from the TRRUST database. This is a transcriptional regulatory network database that utilizes text mining techniques to gather and manually verify detailed information on interactions between TFs and human target genes, ensuring data accuracy. During processing, duplicate pairs were removed, resulting in a final dataset of 6,542 interactions between 696 TFs and 2,064 target genes. Additionally, these transcription factors and target genes were linked to diseases through the DisGeNET database, a resource focused on the genetic basis of human diseases. As a result, 8,199 links between TFs and diseases, along with 31,895 links between target genes and diseases, covering 6,121 different disease types, were collected.

D. Data Balancing Solution with K-Means Clustering and Negative Sample Selection

In this study, the dataset includes 696 TFs and 2,064 target genes, with a total of 1,436,544 TF-target gene pairs. As shown in Fig. 4, only 0.46% of the TF-target gene pairs have been identified as interactions, while the vast majority, accounting for

99.54%, are unknown interactions. This substantial imbalance highlights a severe data imbalance, posing a significant challenge for the prediction model, which can lead to bias and reduced accuracy. Therefore, we have applied data sampling methods to balance the dataset, thereby enhancing the prediction model's accuracy.

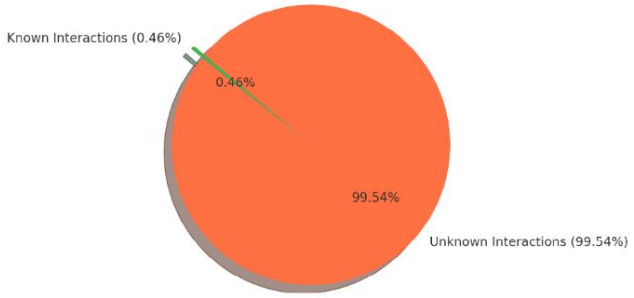


Fig. 4. Distribution of interactions between TF and target genes in the dataset.

In this study, we introduce the FKMU method, a K-means clustering-based Under-sampling technique for selecting negative samples. This method selects samples with the lowest occurrence frequency of TFs in each cluster, based on the inverse information principle as described in Fig. 5. The FKMU procedure is as follows:

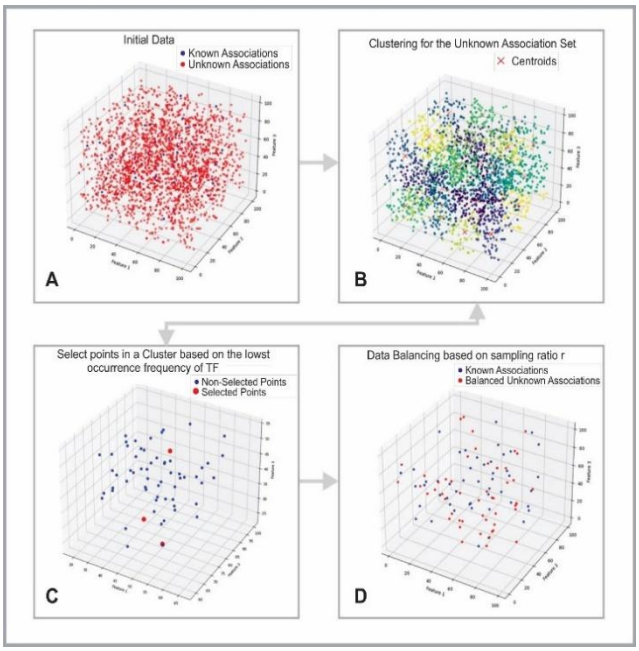


Fig. 5. The process of balancing the dataset using FKMU. A) initial known associations and unknown associations; B) Clustering for the set of unknown associations based on the feature matrix; C) Select points in a cluster based on the lowst occurrence frequency of TF; D) Data balancing based on sampling ratio *r*.

1) Identify the known and unknown association sets from the TF-target gene adjacency matrix A: The known association set *K* consists of TF and target gene pairs with a value of 1 in matrix *A*, representing the associations that have been confirmed. The unknown association set *U* consists of TF and target gene pairs with a value of 0 in matrix *A*, representing the associations that have yet to be evaluated.

2) One-hot encoding for the unknown association set: Each pair (i, j) from the set *U* will be encoded into a feature vector using one-hot encoding for the following factors:

a) TF encoding: A value of 1 at position *i* corresponds to the TF.

b) Target gene encoding: A value of 1 at position *j* corresponds to the target gene.

c) Create feature matrix X: The matrix *X* is defined with dimensions $|U| \times (m + n)$, where each row is the one-hot vector of a pair (i, j) in *U*.

This process is illustrated in Fig. 6, which demonstrates the one-hot encoding structure for unknown associations.

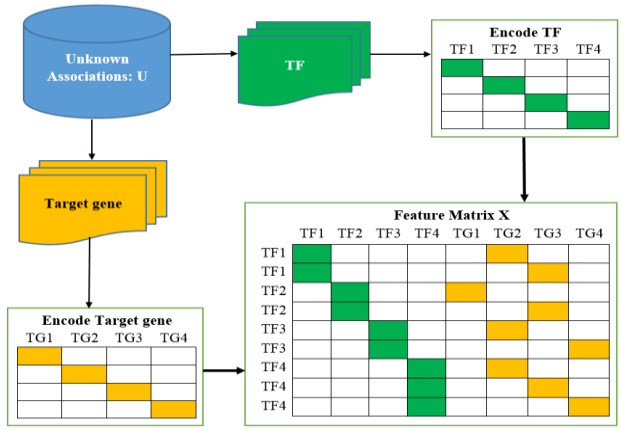


Fig. 6. One – hot encording for unknown associations.

3) Perform K-means clustering on the unknown association set: Apply K-means clustering to divide the unknown associations in *U* into *k* clusters. The feature matrix from step 2 is used to identify the cluster structure in the data, resulting in groups of unknown associations that share similar characteristics. The feature matrix is a sparse matrix. Therefore, before proceeding with the clustering, this matrix is represented in CSR (Compressed Sparse Row) format. CSR is one of the popular formats for storing sparse matrices. It stores the matrix by retaining non-zero values, which helps save memory for large matrices that contain many zero values.

4) Calculate the number of associations to select from each cluster: Based on the total number of known associations $|K|$ and the sampling ratio *r*, determine the number of unknown associations n_p to be selected from each cluster. This helps ensure a balance between known and unknown associations, according to the specified ratio.

5) Select the least frequent samples in each cluster: Within each cluster, calculate the frequency of occurrence of each TF and sort the associations in the cluster by the ascending frequency of the TF. Next, select n_p associations with the lowest frequency of TF occurrence in each cluster to minimize the bias caused by the high frequency of certain dominant TFs. This approach helps create a set of associations that follows the principle of inverse frequency, prioritizing less common

associations to ensure that the sample dataset contains diverse types of associations. The goal of this principle is to enrich the information in the sample set, enabling the model to learn from both rare samples and those that are less biased, thereby enhancing the model's generalization ability for rare cases in real-world data.

6) Return the balanced association set: The set of least frequent associations selected from all clusters forms set *B*, which is the unknown association set balanced according to the ratio *r*. Set *B* is then used as a more balanced dataset for the subsequent steps of the predictive model.

Algorithm 1: Frequency-Based K-Means Under-sampling Algorithm

Input:

- Association matrix $A \in \mathbb{R}^{m \times n}$, where *m* is the number of TFs and *n* is the number of target genes.
- k: The number of clusters for performing K-means clustering.
- r: The sampling ratio from unknown associations.

Output:

- Balanced set of unknown associations *B*, sampled according to ratio *r*.

1: # *Calculate known and unknown set from* A:
2: K = $\{(i,j) \mid A_{ij} = 1\}$ (known association)
3: U = $\{(i,j) \mid A_{ij} = 0\}$ (unknown associations)
4: # *One-hot encoding for TF and Target gene:*
5: Create a feature matrix $X \in \mathbb{R}^{|U| \times (m+n)}$ # *where each row corresponds to an unknown association pair (i, j)*
6: for each pair $(i,j) \in U$ do
7: $X_u = [0, \dots, 1_i, \dots, 0, 0, \dots, 1_j, \dots, 0]$ # *Here, the 1 at position i represents the TF and the 1 at position j represents the Target gene, while the remaining positions are 0.*
8: end for
9: Initialize *k* random cluster centers $\{u_1, u_2, \dots, u_k\}$ from *U*
10: # *Assign each data point x ∈ U to the nearest cluster center:*
11: for $x \in U$ do
12: $c_i = \arg\min_j \|x - u_j\|^2$
13: end for
14: # *Update the center of each cluster with C_j being the set of data points belonging to cluster j:*
15: for *i* = 1 to *k* do
16: $u_j = \frac{1}{|C_j|} \sum_{x \in C_i} x$
17: end for
18: # *Calculate the number of points to select from each cluster:*
19 : $n_{known} = |K|$
20: $n_p = \left\lceil \frac{n_{known} \times r}{k} \right\rceil$
21: Initialize *B = []*
22: # *Select points from each cluster:*
23: for *i* = 1 to *k* do
24: *# Calculate the frequency of each TF in the cluster:*
25: $freq(i) = \sum_{(i,j) \in C_j} 1$
26: # Sort points in cluster C_j by the increasing frequency of their TFs:
27: $C_j' = sorted(C_j,\ key = \lambda x: freq(x_0))$ # *where x_0 is the TF index in the pair (TF, Target gene)*
28: # *Select n_p points with the lowest TF frequency from each cluster:*
29: $B = B \cup C_j'[1: n_p]$
30: end for
31: # *Balance the set of unknown associations:*
32: $B = B\,[1: [n_{known} \times r]]$
33: Return *B*

E. Embedding Heterogeneous Network Nodes using Skip-Gram

Specifically, in a heterogeneous network *G = (V, E, T)* with the number of node types $|T_V| > 1$, he objective is to maximize the co-occurrence probability p of the nodes within the same context window *k*, as follows [31]:

$$\arg\max_{\theta} \sum_{v \in V} \sum_{t \in T_V} \sum_{c_t \in N_{t(v)}} log p(c_t | v; \theta) \quad (2)$$

where $N_{t(v)}$ is the set of neighboring nodes with node *v* in the heterogeneous context with different node types, and $p(c_t|v;\theta)$ is defined as a softmax function [27] as follows:

$$p(c_t|v;\theta) = \frac{\exp(X_{c_t} X_v)}{\sum_{u \in V} \exp(X_u X_v)} \quad (3)$$

where v and c_t are the center node and the nodes in the scanning window, respectively, and, X_v is the embedding vector of node v.

The number of nodes is often very large, so negative sampling techniques are commonly applied to approximate the estimation of probabilities. This method maximizes the probability such that the target node does not appear simultaneously with a randomly selected negative node. The ultimate maximization goal is expressed as follows:

$$O(X) = log\sigma(F(X_{c_t} || X_v) + log\sigma(-F(X_u || X_v)) \quad (4)$$

where *F* represents the fully connected layer, ‖ denotes the concatenation of the embedding vectors of the nodes, and σ(x) is calculated as follows:

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (5)$$

IV. Results

A. Evaluation Criteria

To evaluate the performance of the proposed method, we applied *k*-fold cross-validation (*k* = 5). Specifically, the data was randomly divided into *k* approximately equal parts. In each iteration, one part was used as the test set, while the model was trained on the remaining *k*-1 parts. This process was repeated *k* times, ensuring that each part of the data was used as the test set once.

To demonstrate the effectiveness of the proposed method during *k*-fold cross-validation, we used the Area Under the ROC Curve (AUC) [24], which is calculated as follows:

TABLE I. COMPARISON OF PERFORMANCE BASED ON AUC DURING 5-FOLD CV WITH DIFFERENT PARAMETER SETS

Number of walkers	100	250	350	450	550
Path length	50	80	130	130	150
Dimension of embedding	128	200	300	450	550
Average AUC	0.9199 ± 0.0064	**0.9388 ± 0.0045**	0.9345 ± 0.0064	0.9100 ± 0.0038	0.8828 ± 0.0140

$$AUC = \frac{\sum_{e \epsilon e^+} Rank_e - \frac{|e^+| \times (|e^+|+1)}{2}}{|e^+| \times |e^-|} \quad (6)$$

where e^+ and e^- represent the positive and negative samples, respectively, in the test set, and $Rank_e$ denotes the rank of edge e based on the predicted score.

We conducted experiments with different values for three parameters: number of walkers, path length, and dimension of embedding, while comparing the corresponding prediction results when varying each parameter. The average AUC value for each experiment is presented in Table I. The best prediction results were achieved when the number of walkers was 250, the path length was 80, and the dimension of embedding was 200. The corresponding ROC curve is illustrated in Fig. 7, showing that our proposed method achieved an average AUC value of 0.9388 ± 0.0045 through 5-fold cross-validation.

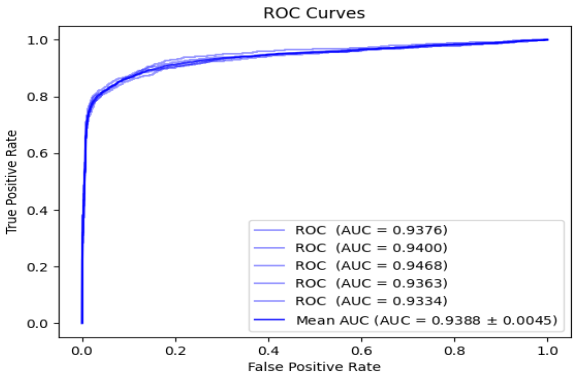


Fig. 7. ROC curve through 5-fold cross-validation.

B. Determine the Optimal Number of Clusters

Determining the optimal number of clusters *k* is a crucial factor in the effectiveness of the K-means algorithm. Choosing *k* too small can lead to data points being grouped together, overlooking significant differences between clusters. Conversely, if *k* is too large, the data may be unnecessarily divided into clusters, reducing generalization. For our experiment, we applied the Elbow method with values ranging from 10 to 200. Specifically, for each *k* value, the K-means algorithm was executed, and the WCSS (Within-Cluster Sum of Squares) was calculated.

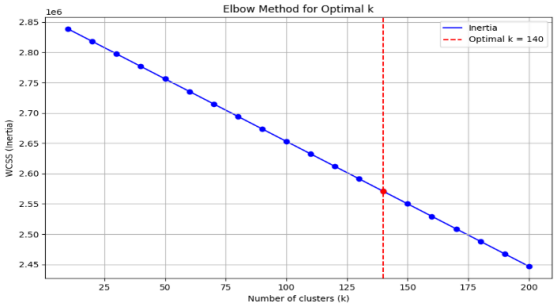


Fig. 8. The elbow method for selecting the optimal number of clusters *k*.

The results in Fig. 8 show that as *k* increases from 10 to around 140, the WCSS decreases significantly, indicating that increasing the number of clusters improved data clustering. However, after *k* = 140, the WCSS begins to decrease more slowly, suggesting that adding more clusters no longer provides substantial benefits for data separation. Therefore, we chose *k* = 140 as the optimal number of clusters, as it achieves a balance between reducing WCSS and maintaining model generalization.

C. Negative Sampling Rate

Choosing a balance ratio *r* between Unknown Associations and Known Associations is aimed at optimizing the model's performance in distinguishing between these two groups. A reasonable balance ratio enables the model to learn the characteristics of both groups without bias, thereby achieving the highest AUC value and ensuring accuracy when applied to new data.

TABLE II. THE BALANCE RATIO BETWEEN UNKNOWN ASSOCIATION AND KNOWN ASSOCIATION

Balance Ratio r (Unknown : Known Association)	Average AUC Value
1:1	0.9388 ± 0.0045
2:1	0.9264 ± 0.0028
3:1	0.9229 ± 0.0047
4:1	0.9275 ± 0.0055
5:1	0.9273 ± 0.0041
6:1	0.9212 ± 0.0025
7:1	0.9231 ± 0.0030
8:1	0.9229 ± 0.0041
9:1	0.9210 ± 0.0050
10:1	0.9175 ± 0.0017

When conducting experiments with different values of the balance ratio *r*, we obtained the results shown in Table II. The results indicate that when the balance ratio is 1:1, the AUC value reaches its highest point at 0.9388 ± 0.0045, suggesting that this is the optimal balance ratio for effectively distinguishing between Unknown Associations and Known Associations. As the balance ratio increases from 2:1 to 10:1, the AUC value gradually decreases, particularly at a ratio of 10:1, where the AUC value significantly drops to 0.9175 ± 0.0017. This indicates that the model's effectiveness diminishes when Unknown Associations constitute too large a proportion compared to Known Associations.

D. The Impact of Selecting Meta-Paths

The random walk strategy based on meta-paths ensures that the model accurately integrates the semantic relationships between different types of nodes. Utilizing different meta-path schemas to generate sequences of nodes can capture the diverse semantic and structural relationships among these node types.

In this experiment, we designed a new meta-path schema "TF-Target gene-Disease/CS-Target gene-TF-Disease/CS-Target gene-TF" while also using the original meta-path "TF-Target gene-Disease/CS-Target gene-TF" [24] to conduct the random walk process and evaluate the prediction effectiveness of each schema. The results in Table III show that the new schema achieves a slightly higher average AUC value (0.9388 ± 0.0045) compared to the original schema (0.9366 ± 0.0044), suggesting that the new schema can improve predictive performance by capturing additional potential links within the heterogeneous network. Here, CS (Cold Start node) is a node added to the paths to address the cold start problem in the model. This issue arises when certain nodes (especially TFs or target genes) have no links to any target genes in the training data, making it difficult to learn embedding vectors for these nodes. By adding the CS node and setting its embedding to a vector where all elements have a value of 1, the model can learn information from paths containing the CS node, helping to mitigate the lack of link data for these nodes and enhance the overall performance of the model across the heterogeneous network.

TABLE III. COMPARING AUC PERFORMANCE ACROSS DIFFERENT META-PATHS

Meta-paths	Average AUC Value
TF-target gene-disease/CS-target gene-TF	0.9366 ± 0.0044
TF-target gene-disease/CS-target gene-TF-disease/CS-target gene-TF	0.9388 ± 0.0045

E. Predicted Scores for TF-Target Gene Pairs

After training the model, we obtain low-dimensional embedding vectors for TFs and target genes. From this, we create an embedding matrix *M* for TFs and an embedding matrix *G* for target genes. The predicted scores for the interactions between TFs and target genes are determined as follows:

$$P = M.G^T \quad (7)$$

where the value in the *i-th* row and *j-th* column represents the interaction score between the *i-th* TF and the *j-th* target gene.

F. Analysis and Comparison with Recent Studies

To evaluate the superior performance of the proposed model, we compared its predictive capabilities with recent studies, including Metapath2vec [25], HGETGI [24], and GraphTGI [30]. The comparison results shown in Table IV indicate that our method achieves the highest average AUC value of 0.9388, outperforming the other three methods. This confirms that our model is highly effective in predicting unobserved target genes for specific TFs.

Although models such as Metapath2vec, HGETGI, and GraphTGI effectively utilize heterogeneous graphs, particularly for predicting TF-target gene interactions, GraphTGI demonstrates impressive performance with an AUC of 88.64% in five-fold cross-validation, while HGETGI excels in leveraging semantic information through graph embeddings. However, all three methods lack robust mechanisms for handling imbalanced data and selecting negative samples, which can limit their ability to optimize performance on complex and highly imbalanced datasets. FKMU addresses these challenges through a K-means-based under-sampling strategy, ensuring a balanced dataset and enhancing the robustness of the model. Furthermore, the introduction of a novel meta-path allows FKMU to capture semantic relationships within the graph more effectively and optimize the detection of potential interactions. These advancements establish FKMU as an effective and superior method for predicting TF-target gene interactions while offering broad applicability to more complex and diverse problems, especially for large-scale and heterogeneous datasets in the future.

TABLE IV. COMPARING THE PERFORMANCE OF RESEARCH METHODS

Methods	Average AUC Value
Metapath2vec [25]	0.8239 ± 0.0057
HGETGI [24]	0.8519 ± 0.0731
GraphTGI [30]	0.8864 ± 0.0057
FKMU	0.9388 ± 0.0045

G. Case Study

To evaluate the predictive performance of the model in identifying potential target genes associated with TFs, we conducted experiments on the transcription factors CTCF and TP53. Specifically, we removed the links between the specific TFs used in the experiments and their target genes. We then reconstructed the heterogeneous network. Finally, we trained the model and tested it for each specific TF case to assess the model's performance.

The transcription factor CTCF (CCCTC-binding factor) is an important protein involved in regulating the structure and function of the genome. CTCF binds to DNA sequences to create insulator regions and chromatin loops, helping to regulate gene activity. CTCF can activate or repress genes depending on its binding location. Mutations in the CTCF gene are associated with various diseases such as cancers (breast, colorectal, prostate), neurodevelopmental disorders (Bardet-Biedl syndrome, autism spectrum disorders), and rare genetic diseases, primarily due to disruptions in chromatin structure and gene regulation.

The transcription factor TP53 (tumor protein p53) is an important gene, often referred to as the "guardian of the genome" due to its key role in maintaining genetic stability and preventing tumor formation. This gene encodes the p53 protein, a tumor suppressor that plays a crucial role in controlling cell division, repairing damaged DNA, and activating apoptosis when cells sustain irreparable damage. TP53 mutations are a common cause in many types of cancer, including lung cancer, breast cancer, colorectal cancer, and skin cancer. Research on TP53 not only elucidates the mechanisms of cancer but also opens new avenues for treatments aimed at restoring p53 function to prevent the development of cancer cells.

TABLE V. TOP 20 TARGET GENES FOR CTCF

Target gene	CTCF-Related Target
CDKN1A	Confirmed
MTHFR	Unconfirmed
VEGFA	Confirmed
TNF	Confirmed
SOD2	Confirmed
IL6	Confirmed
PTGS2	Confirmed
BCL2	Confirmed
CCND1	Confirmed
CDKN2A	Unconfirmed
MMP9	Confirmed
KRAS	PMID:32374727
CDH1	Confirmed
IFNG	Unconfirmed
TGFB1	Confirmed
TERT	Confirmed
PTEN	Confirmed
NOS2	Confirmed
ERBB2	Confirmed
IL1B	Unconfirmed

TABLE VI. TOP 20 TARGET GENES FOR TP53

Target gene	TP53-Related Target
CDKN1A	Confirmed
VEGFA	Confirmed
MTHFR	Confirmed
IL6	Confirmed
TNF	Confirmed
SOD2	Confirmed
CCND1	Confirmed
PTGS2	Unconfirmed
BCL2	Confirmed
IFNG	Unconfirmed
TGFB1	Confirmed
IL1B	PMID:34986125
CDH1	Confirmed
MMP9	Unconfirmed
KRAS	Confirmed
CDKN2A	Confirmed
ABCB1	Confirmed
TERT	PMID:23284306
ERBB2	Confirmed
EGFR	Confirmed

We ranked the predicted scores based on the weighted matrix to identify potential target genes. Then, we assessed the accuracy of these target genes by comparing them with the hTFtarget database [18]. Specifically, we focused on testing and validating the top 20 predicted target genes to ensure the reliability and accuracy of the predictive model. This process helps us confirm the model's capability in identifying potential target genes associated with the TFs.

The experimental results are presented in Table V for CTCF and in Table VI for TP53, respectively. According to these tables, 75% (15/20) of the predicted target genes have been validated against the hTFtarget dataset. Additionally, we conducted supplementary research and discovered genes such as KRAS, which, although not listed as interacting with CTCF in hTFtarget, have been reported to interact with CTCF in other studies, as indicated by the PMID [11] codes in Table V. Similarly, genes such as IL1B and TERT were also found to interact with TP53, as shown by the PMID [14, 16] codes in Table VI. These results demonstrate the effectiveness of the proposed method.

V. CONCLUSION

Predicting interactions between transcription factors and target genes remains a significant challenge, particularly in the context of the complex relationships within the gene regulatory network that have not been fully explored. To address this issue, we propose the FKMU method, a novel approach aimed at handling data imbalance when predicting interactions between TFs and target genes. FKMU combines K-means clustering and inverse information principles to select underrepresented samples in the dataset, thereby balancing sample ratios and improving the accuracy of the model. This method applies the K-means algorithm to partition unknown samples into clusters, subsequently prioritizing TFs with low occurrence frequency in each cluster to enhance diversity and representation within the data. Experimental results on real datasets demonstrate that FKMU achieves superior performance in accurately predicting interactions between TFs and target genes compared to current methods, with a significantly higher average AUC value. We expect that the FKMU method will pave the way for new avenues in scientific research, improving the handling of imbalanced data and enhancing the accuracy of predictive models in the biomedical field.

REFERENCES

[1] C. M. Hoang et al., "A K-means Clustering Based Under-Sampling Method for Imbalanceed Dataset Classification," 2024 International Conference on Information Networking (ICOIN), pp. 708-713, 2024, doi: 10.1109/ICOIN59985.2024.10572133.

[2] D. N. Anh, B. D. Hung, P. Q. Huy, and D. X. Tho, "Feature analysis for imbalanced learning," Journal of Advanced Computational Intelligent Informatics, vol. 24, no. 5, pp. 648–655, Sep. 2020.

[3] D. X. Tho, B. D. Hung et al., "Prediction of autism-related genes using a new clustering-based under-sampling method," In: 2019 11th International Conference on Knowledge and System Engineering (KSE), pp. 1-6, IEEE, 2019.

[4] D. X. Tho, D. N. Anh, "Imbalance In the Learning Chest X-Ray images For COVID-19 detection," In: Soft Computing: Biomedical and Related Applications, pp. 107-119. Springer Berlin Heidelberg, 2021.

[5] D. X. Tho, L. M. Hung, and D. N. Anh, "Drug Repositioning for Drug Disease Association in Meta-paths," In Deep Learning and Other Soft Computing Techniques: Biomedical and Related Applications, 2023, pp. 39-51, Cham: Springer Nature Switzerland.

[6] D. X. Tho, and T. T. Le, "KNN-SMOTE: An Innovative Resampling Technipue Enhancing the Efficary of Imbalanced Biomedial Classification," Machine Learning and Other Soft Computing Techniques: Biomedical and Related Applications, pp. 111-121, 2024, doi: 10.1007/978-3-031-63929-6_11.

[7] H. Han, J. W. Cho, S. Lee, A. Yun, H. Kim, D. Bae at al., "TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions," Nucleic Acids Research., vol. 46, pp. D380-D386, 2017, doi: 10.1093/nar/gkx1013.

[8] H. He, M. Yang, S. Li, G. Zhang et al., "Mechanisms and biotechnological applications of transcription factor," Synthetic and Systems Biotechnology, Vol. 8, pp. 565-577, 2023.

[9] H. Li, R. Fan, Q. Shi, and Z. Du, "Class Imbalanced Fault Diagnosis via Combining K-Means Clustering Algorithm with Generative Adversarial Networks," Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 25, No.3, pp. 346-355, 2021.

[10] J. Lachantin, R. Singh, B. Wang et al., "Deep motif dashboarch: Visualizing and understanding genomic sequences using deep neural network," Pac Symp Biocomput., vol. 22, pp. 254–265, 2017, doi:10.1142/9789813207813_0025.

[11] J. Rinal, E. S. Sokol, R. J. Hartmaier, S. E. Trabucco et al., "The genomic landscape of metastatic breast cancer: Insights from 11,000 tumors," PloS One, Vol. 15, No. 5, e0231999, 2020, PMID:32374727.

[12] J. T. Wade, "Mapping Transcription Regulatory networks with CHIP-seq and RNA-seq," Adv Exp Med Biol, vol. 883, pp. 119–134, 2015, doi: 10.1007/978-3-319-23603-2_7.

[13] J. Wang, "TF-Target Finder: An R Web Application Bridging Multiple Predictive Models for Decoding Transcription Factor-Target Interactions," Preprints.org, 2024, doi: 10.20994/preprints202404.1212.v1.

[14] K. Gao, Y. Zhu, H. Wang, X. Gong et al., "Network Pharmacology reveals the potential mechanism of Baiing Qinghou decoction in treating laryngeal squamous cell carcinoma," Aging (Albany NY), vol. 13, no. 24, pp. 26003-26021, 2021, PMID:34986125.

[15] K. Su, A. Katebi, V. Kohar, B. Clausss, D. Gordin, Z. S. Qin et al., "NetAct: a computational platform to construct core transcription factor regulatory networks using gene activity," Genome Biology, vol. 23, no. 1, pp. 1-21, 2022.

[16] L. Xie, C. Gazin, S. M. Park, Li. J. Zhu et al., "A Synthetic Interaction Screen Identifies Factors Selectively Required for Proliferation and TERT Transcription in p53-Deficient Human Cancer Cells," Plos Genetics, Vol. 8, No. 12, e1003151, PMID:23284306.

[17] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," J. of Artificial Intelligence Research, Vol. 16, No. 1, pp. 321-357, 2002.

[18] Q. Zhang et al., "hTFtarget: a comprehensive database for regulations of human transcription factors and their targets," Genomics Proteomics Bioinf., vol. 18, pp. 120–128, 2020.

[19] Q. Zhou, B. Sun, "Adaptive K-means clustering based under-sampling methods to solve the class imbalance problem," Data and Information Management, vol. 8, no. 3, pp. 1-12, 2024, doi:10.1016/j.dim.2023.100064.

[20] R. Mundade, H. G. Ozer, H. Wei, L. Prabhu, and T. Lu, "Role of CHIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond," Cell Cycle., vol. 13, pp. 2847–2852, 2014, doi: 10.4161/15384101.2014.949201.

[21] S. Salekin, JM. Zhang, and Y. Huang, "Based-pair resolution detection of transcription factor binding site by deep decovolutional network," Bioinformatics., vol. 34, no. 20, pp. 3446-3453, 2018.

[22] T. Doan, Z. Kuang, J. Wang, Z. Ma, "GBDTLRL2D Predicts LncRNA-Disease Associations Using MetaGraph2Vec and K-Means Based on Heterogeneous Network," Frontiers in Cell and Developmental Biology, vol. 9:753027, 2021, doi: 10.3389/fcell.2021.753027.

[23] T. V. Thai, B. D. Hung, D. X. Tho et al., "A New Computational Method Based on Heterogeneous Network for Predicting MicroRNA-Disease Associations," Soft Computing for Biomedical Applications and Related Topics, 2021, pp. 205–219.

[24] Y. A. Huang et al., "Heterogeneous graph embedding model for predicting interactions between TF and Target gene," Bioinformatics, vol. 38, no. 9, pp. 2554-2560, 2022, doi: 10.1093/bioinformatics/btac148.

[25] Y. Dong et al., "metapath2vec: Scalable representation learning for heterogeneous network," In Proceedings of the 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining., 2017, pp. 135–144.

[26] Y. Fan, X. Ma, "Gene regulatory network inference using 3D convolutional Neural Network," In Proceeding of the AAAI Conference on Artificial intelligence, 2021, vol. 35, pp. 99-106.

[27] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: Meta path-based top-k similarity search in heterogeneous information network," In VLDB'11., pp. 992-1003, 2011.

[28] Y. Yang, Q. Fang, H. B. Shen, "Predicting gene regulatory interractions based on spatial gene expression data and deep learning," PloS Comput Biol, 2019, 15(9):e1007324.

[29] Ž. Avsec, M. Weilert, A. Shrikumar, S. Krueger, A. Alexandari, K. Dalal, and S. Masri, "Based-resolution models of transcription-factor binding reveal soft motif syntax," Nature Genetics., vol. 53, pp. 345–366, 2021, doi: 10.1038/s41588-021-00782.

[30] Z. H. Du, Y. H. Wu, Y. A. Huang, J. Chen, G. Q. Pan, L. Hu, Z. H. You, and J. Q. Li, "GrapTGI: An attention-based graph embedding model for predicting TF-Target gene interactions," Briefings in Bioinformatics, vol. 23, no. 3, pp. 1-11, 2022, doi: 10.1093/bib/bbac148.

[31] Z. Liu, S. Zhang, J. Zhang, M. Jiang, M. Liu, "HeteEdgeWalk: A Heterogeneous Edge Memory Random Walk for Heterogeneous Information Network Embedding", Entropy, 2023, 25(998) doi:10.3390/e25070998.

[32] Z. Shen, W. Bao, and D. S. Huang, "Recurrent neural network for predicting transcription factor binding sites," Scientificreports., vol. 8, no. 1, pp. 1–10, 2018, doi: 10.1038/s41598-018-33321-1.

A Deep Learning-Based LSTM for Stock Price Prediction Using Twitter Sentiment Analysis

Shimaa Ouf[1], Mona El Hawary[2]*, Amal Aboutabl[3], Sherif Adel[4]
Information Systems Department-Faculty of Commerce and Business Administration, Helwan University, Cairo, Egypt[1, 2]
Computer Science Department-Faculty of Computers and Artificial Intelligence, Helwan University, Cairo, Egypt[3]
Administration Department-Faculty of Commerce and Business, Helwan University, Cairo, Egypt[4]

***Abstract*—Numerous economic, political, and social factors make stock price predictions challenging and unpredictable. This paper focuses on developing an artificial intelligence (AI) model for stock price prediction. The model utilizes LSTM and XGBoost techniques in three sectors: Apple, Google, and Tesla. It aims to detect the impact of combining sentiment analysis with historical data to see how much people's opinions can change the stock market. The proposed model computes sentiment scores using natural language processing (NLP) techniques and combines them with historical data based on Date. The RMSE, R^2, and MAE metrics are used to evaluate the performance of the proposed model. The integration of sentiment data has demonstrated a significant improvement and achieved a higher accuracy rate compared to historical data alone. This enhances the accuracy of the model and provides investors and the financial sector with valuable information and insights. XGBoost and LSTM demonstrated their effectiveness in stock price prediction; XGBoost outperformed the LSTM technique.**

***Keywords*—Sentiment analysis; stocks price prediction; correlation; natural language processing (NLP); machine learning model; LSTM; XGBoost**

I. Introduction

Stock market prediction has been a matter of interest to researchers, vendors, and investors for a long period. The main purpose of predicting the stock market is to achieve optimal results and decrease investment risk. It focuses on establishing an effective technique for predicting stock prices and providing well-informed, data-driven insights on market behavior [1]. Another purpose of stock market prediction is to predict the most accurate price in the future and determine the trend of the stock price, whether it is going up or down. This helps and guides investors to make better choices, avoiding potentially harmful investments that could increase their profits and reduce their losses [2]. Due to its volatility, unpredictability, and rapidness, stock market prediction is challenging. Various factors, such as public opinion, social media, feelings, and sentiments, influence the accuracy of stock market predictions [3].

Social media platforms (Facebook, Twitter, etc.) became a vital data source. Social media plays an important role for companies and people. People use social media every day to express their opinions and experiences and review products, services, or even companies [4]. On the other hand, companies utilize social media platforms to extract and analyze customers' opinions and feelings toward what they offer [5] [6]. Sentiment analysis (SA) is one of the disciplines that analyze social media data. Sentiment analysis can be defined as analyzing users' emotions using their opinions, sentiments, and subjective texts to decide if their interactions are positive, negative, or neutral. As stated, Sentiment Analysis is a type of subjectivity analysis that concentrates on identifying opinions, feelings, and respect conveyed through natural language [7]. It enhances the quality of goods and services by evaluating consumers' feedback on a certain product or service. Furthermore, natural language processing (NLP) is important for teaching machines to process human language and translate it into machine-readable format.

Sentiment analysis plays an important role in predicting stock prices by examining public opinion and social media to estimate the market mood and its effect on stock prices and aid the investors in overcoming the investment risk Merging sentiment analysis with traditional financial measures improves predictive accuracy. Companies can utilize risk management techniques to make strategic decisions, regardless of stock fluctuations [8].

Different domains have addressed sentiment analysis because it can help companies increase their capitalization by enhancing their products or services to meet customers' expectations. Stock markets leverage social media data, and SA can predict its stock prices depending on people's opinions. However, predicting stock prices is complicated because of their volatile and dynamic nature [9].

Historical data is another factor that plays an important role in stock price prediction. Historical data comprehends market behaviors and trends, leading to well-informed financial decisions based on data analysis. Examining historical data, which includes financial variables such as volume, price, high, low, and close, enables predictive algorithms to detect connections and recurrence patterns. Then train and test machine learning techniques using historical data to predict stock prices and improve their accuracy [10].

A. Background

1) The extreme gradient boosting (XGBoost) model: XGBoost in ML for regression and classification is an ensemble technique that builds a series of weak learners, usually decision trees (DTs). Every learner removes errors from the previous learners by minimizing errors between actual values and predicted values. This process is called the loss function [11]. XGBoost employs the L1 (Lasso) and L2 (Ridge) methods of regularization to reduce overfitting and improve the model's ability to generalize. Its parallel processing ability makes it computationally rapid and suitable for big data processing.

XGBoost works by figuring out an objective function in Eq. (1). Regularizes then use this function along with a loss function and a residual measurement to get rid of complicated models that cause overfitting [12]. In XGBoost, the objective function comprises a loss function that calculates the residuals and a second part that simplifies the model and reduces overfitting [13]. Model parameter optimization can reduce the objective function through gradient descent on the loss function. DTs persist in decreasing the residuals and repeatedly reduce the loss until it diminishes.

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \iota(\mathrm{y}_i, \hat{\mathrm{y}}_i) + \sum_{K=1}^{K} \Omega(f_{k)} \qquad (1)$$

Where $\mathcal{L}(\theta)$ is the objective function (overall), $\iota(\mathrm{y}_i, \hat{\mathrm{y}}_i)$– (evaluation metrics), or loss function was measured by (mean absolute error), $\Omega(f_{k)}$ Regularization term to minimize overfitting.

$$\hat{\mathrm{y}}_{\mathrm{i}}^{(\mathrm{t})} = \sum_{k=1}^{t} \mathrm{f}_k(\mathrm{x}_i) \qquad (2)$$

Prediction in XGBoost as shown in Eq. (2): At each stage, the forecast $\hat{\mathbf{y}}_{\mathbf{i}}$, is calculated by summing the predictions made by all (t) the trees and (f_k) the tree. The objective function decreases by using gradient descent to iteratively adjust model parameters and minimize the loss function [14]. During the optimization, the DTs aim to eliminate the residuals (the differences between actual and predicted values). This process continues until the algorithm reaches a minimum point where further tuning does not significantly reduce the loss.

2) XGBoost integration with sentiment analysis: XGBoost is a gradient-boosting technique that uses several decision trees to simulate the relationships between features, such as emotion scores and stock prices, and the target variable (the future stock price). Each decision tree works to make predictions better by using attributes and stock market dates repeatedly. Simultaneously, the loss function undergoes optimization to identify the least effective prediction trees [15]. XGBoost has integrated with sentiment analysis, specifically sentiment score or polarity, as one of its input features. XGBoost can manage nonlinear interactions and prioritize feature relevance, which is beneficial when combining textual sentiment analysis with historical data. XGBoost boasts several features that set it apart from the traditional model. XGBoost can define and determine how much the sentiment analysis score contributes to the prediction. Compared to stock features, regularization is another feature that enables XGBoost to overcome or prevent overfitting, particularly when the combined features have high dimensionality. Scalability is considered the most significant feature in XGBoost. It is used for large datasets. XGBoost can handle both historical and sentiment analysis [16].

3) Long short-term memory (LSTM): To overcome long-term dependency, a unique type of recurrent neural network was explicitly designed, namely LSTM. The LSTM incorporates a novel memory cell that replaces conventional artificial neurons in the hidden layers, enabling it to retain data over an extended duration. According to Hochreiter and Schmid Huber [17], the inability to handle long-term dependencies is an important factor for any dispute that contains a time series; this is the main struggle when applying ordinary neural network architecture to predict a stock price. Three gates organize the information, determining which data to store in the cell, which to discard, and what the cell's output will be. LSTM proves to be a confidential mechanism in many fields, not only computer science but also statistics, linguistics, and medicine. All these areas have tasks involving the analysis of sequential data, prediction, classification, and regression, for which LSTM proved to have a great ability to process them well [18].

4) Long short-term memory (LSTM) integrated with sentiment analysis: Long Short-Term Memory (LSTM), a type of deep learning, is considered the best technique for processing sentiment analysis and can manage and handle sequential data by capturing long-term dependencies within the text or sentence. This makes it suitable for sentiment analysis tasks, particularly when the word depends on the word in the surrounding context [19]. The LSTM architecture incorporates memory cells that store information related to stored sequences over extended periods. This enables the model to interpret the sentiment in the text more accurately and quickly. This is true even when there is a significant gap between the keywords. Sentiment analysis depends on data quality, such as the volume of text. LSTM can learn and forecast the next sentiment labeled more accurately based on the large volume of labeled text. This feature makes it an invaluable tool for other tasks, like monitoring social media and predicting stock prices [20].

This paper is organized as follows: Section II presents a literature review. Section III introduces the proposed methodology, while Section IV and Section V showcase experimental analysis and results respectively. Simultaneously, Section VI highlights the conclusions.

II. Literature Review

Many researchers have tackled improving prediction accuracy by using machine learning and deep learning algorithms. The goal of this paper is to review previous studies, find challenges, and overcome these gaps and challenges by providing a novel research method.

A study [21] employed LSTM regression models to estimate India's NIFTY 50 index. The deep learning-based LSTM model outperformed traditional machine learning methods. A study [22] compares ARIMA, LSTM, and BiLSTM models for forecasting financial time-series data and concludes that the BiLSTM model produced the best results. Another study in [23] offered an RNN-Boost model for forecasting Chinese stock market values, which outperformed the baseline RNN model.

The study in [24] employed deep learning, support vector regression, and linear regression for the stock market to forecast and evaluate the sentiment surrounding each event, focusing on four nations that represent established, emerging, and undeveloped economies: the United States, Hong Kong, Turkey, and Pakistan. This study assessed the system's performance using mean absolute error (MAE) and root mean square error

(RMSE). The results indicate that incorporating sentiment analysis for these events improves the system's performance.

This research paper aims to concentrate on the implementation of the stock-containing Long Short-Term Memory (LSTM) algorithms. The LSTM originates from the recurrent neural network in stock. It has a significant effect on time series data problems. This study establishes two models: the BP neural network model and the LSTM model. Next, integrate these models with the available stock data to generate a series of predictions. Undoubtedly, the prediction accuracy of LSTM models has improved. The accuracy rate can reach 60%–65%. During the modeling process, this study has refined traditional gradient descent algorithms and specifically designed the neural network's input data to mitigate the inevitable "sawtooth phenomenon" of the gradient descent algorithm. Additionally, they established a library of parameter combinations and utilized the dropout technique to achieve more accurate prediction results [25].

Machine learning models demonstrate that artificial neural networks (ANNs) can learn input-output correlations and assist in producing close estimates of daily closing prices when trained on the same data [26]. Deep learning techniques like convolutional neural network (CNN) was used in sentiment analysis research in the Indonesian language [27].

Many research studies use sentiment analysis to extract opinions from the text and categorize them as positive, negative, or neutral. Researchers can categorize the research into lexicon-based or machine-based learning-based methods. In recent years, many lexicons can be depended on and used to determine the text, including SentiWord Net [28], WordNet-Affect, and Sentic Net [29].

The author used the support vector machines for sentiment analysis. Experiments reveal an accuracy of 89.93% in predicting the direction of the SSE 50 index's movement, with an additional 18.6% increase in accuracy when adding sentiment-describing parameters. At the same time, this model supports investors in making better investment decisions [30].

In study [31], they create and evaluate forecasting models for stock prices and trends. They suggest a novel decision tree technique to predict stock performance by utilizing a large-scale sample of tweets relating to four companies: Apple, Google, Microsoft, and Netflix. They concluded that a decision tree model surpasses a multiple regression model.

In study [32] Deep learning models have been used to enhance the accuracy of stock price prediction. They depend on more than 265,000 news articles and S&P 500 companies' financial datasets to predict stock prices. They concluded that the RNN model performed well compared with other models (ARIMA and Facebook Prophet). Furthermore, RNN proved its efficiency with fixed/stable stocks rather than low prices.

In study [33] Applied machine learning historical stock prices and financial news from four distinct companies in different industries. They aim to ascertain the influence of financial news on the fluctuations in stock prices. They have also conducted experiments to evaluate the challenges associated with predicting certain stocks. This study discovered that of Tata Motors stock prediction, an automobile company, has the highest MAPE, resulting in a greater deviation from the actual prediction than other stocks.

The relationship between sentiment derived from financial news and tweets and the movements of the FTSE100 index is examined [34]. The investigation aimed to determine the strength of the correlation between sentiment indicators on a given day and market volatility and returns observed on a subsequent day. The experimental findings reveal evidence of a correlation between sentiment and stock market movements.

In study [35] Artificial Neural Networks (ANN) have been used to predict the prices in different periods one day, 7 days, 30 days, 60 days, and 90 days. They used different resources of social media datasets combined as input such as (Twitter, google Trends, web news, forum posts) and the Stock Exchange (GSE) from January 2010 to September 2019 of Ghana company dataset. They concluded that the information they obtained from social media can influence the effectiveness of stock price movement prediction.

The authors utilized deep learning models to track and evaluate the Chinese stock movements. They figured out that LSTM's forecast is closest to actual values, and increasing the number of hidden layers had no significant impact on accuracy [36].

The relationship between Twitter tweets and stock prices has been tested and proven using multiple models. While training the models, SVM reveals a superior performance. The neural network models proved their superiority while evaluating the difference in the closing stock prices between the two companies [37].

After reviewing the literature, we found that no study measures the correlation between SA and stock price prediction. Therefore, this correlation is represented by analyzing the performance of integrating the sentiment with historical data compared with using historical data only.

Therefore, this research shed light on this gap by developing an AI model to improve stock price prediction accuracy. It provides investors with more insight to make informed and valuable decisions regarding buying or selling. The proposed model depends on applying the LSTM and XGBoost techniques with historical data only and with sentiment analysis integrated with historical data. The integration of sentiment analysis (qualitative data) with historical data can improve prediction accuracy and yield valuable market insights.

III. Research Methodology

The research methodology involves utilizing natural language processing (NLP) techniques to extract sentimental information from social media as represented in Fig. 1. This sentiment analysis, categorized as positive, negative, or neutral, is then combined with the corresponding stock prices retrieved from Yahoo Finance. The existence of a correlation between sentiment analysis and stock prices is demonstrated using machine learning and deep learning models. Furthermore, two models, including LSTM (Long Short-Term Memory) and XGBoost, are utilized to enhance prediction accuracy based on sentiment analysis. These models are applied with historical data only and with sentiment analysis integrated with the historical

data. These models are measured through evaluation metrics such as Root Mean Squared Error (RMSE), R^2, and Mean Absolute Percentage Error (MAE). This experiment helps to measure the impact of integrating sentiment analysis and provides valuable insights into where to buy or sell in the financial sectors.

IV. Experimental Analysis

A. Dataset Description

This study utilizes three stocks captured from Kaggle: Dataset 1 collects people's tweets for Apple from June 2015 to December 2019. Dataset 2 collects tweets from June 2015 to September 2020 for Google. The final dataset gathers tweets from June 2020 to October 2020 for Tesla. Each dataset comprises two columns: the first column is the date of the post date, while the second column contains the tweet's text. The quantitative data type in this study necessitates a sentiment analysis. Then, other datasets were downloaded from Yahoo Finance, where the stock prices for Apple, Google, and Tesla span the same period. Each dataset consists of seven columns, 1) The date: marks the recording or reporting of the stock market data. 2) Open: The opening price of the company's stock on the given date. 3) High: The highest price reached during the trading day. The company's stock traded at its highest price. It indicates the highest price reached during the trading session. 4) Low: The trading day reached its lowest price. 5) Close: The price at which the stock closed on that trading day. 6) Adj Close refers to the adjusted closing price, which considers any corporate actions such as stock splits or dividends that may impact the stock's value. 7) Volume refers to the total number of shares traded during a trading day.

B. The Experimental Approach

This section discusses the experiments conducted to measure the effect of people's sentiments on total fluctuation. The two main subsections below provide detailed steps. The first subsection illustrates the data preprocessing. The second clarifies the experimental steps.

1) Data preprocessing: The first task in the preprocessing stage is to detect duplicated records and null values. Drop the duplicate records and use the mean imputation to address the missing value problem. Furthermore, the uppercase letters are converted to lowercase letters to streamline the training and testing of the model. The data cleaning task included the removal of all stop words, such as "that," "for," "the," "a," "he," and "has." Furthermore, the lemmatizing process is conducted on the previously mentioned datasets., an NLP tool allows end users to understand full sentence input from end users. Also, URLs, mentions, and # hashtags were removed in the preprocessing phase.

2) The experimental steps: This study applies the Long Short-Term Memory (LSTM) and XGBoost algorithms to detect the relationship between people's sentiment and the total fluctuation to predict the future stock price.

After cleaning the data, the polarity measure is conducted, utilizing two algorithms, such as VADER [38] From NLTK [39], which employs tools to identify positive, negative, or neutral comments and compound sentiment. VADER evaluates the polarity, or the intensity of the emotion, by determining whether the statement is positive, negative, or neutral as introduced in Fig. 2. Text Blob NLP evaluates the subjectivity and polarity of Twitter tweets.

Polarity is the positive, negative, or neutral feeling expressed in a text as represented in Fig. 3. It determines whether a text describes a good or negative attitude toward a particular entity or issue. NLP frequently uses polarity analysis for sentiment analysis, identifying and categorizing opinions expressed in text. on the other hand, Subjectivity hand describes how subjective or objective a piece of literature is. It is used to determine if a statement is a fact or opinion. Subjectivity analysis in NLP allows detective subjective language and its distinction from objective language. Subjectivity analysis is used to assess whether a text is objective or subjective [40].

The sentiment score of a piece of text (X tweet) can be calculated using various methods as presented in Eq. (3). A simple approach might be used to assign scores to positive and negative words and compute an overall score for the document [46].

$$S = \frac{(\Sigma p_l') - (\Sigma n_J')}{n_p + n_n} \quad (3)$$

Where S is the sentiment score, p_l' are the scores of positive words, n_j are the scores of negative words, n_p is the number of positive words, and n_n is the number of negative words in the text. Market volatility on a given day can be modeled as a function of sentiment scores from the previous day, incorporating both social media sentiment and traditional financial indicators [41].

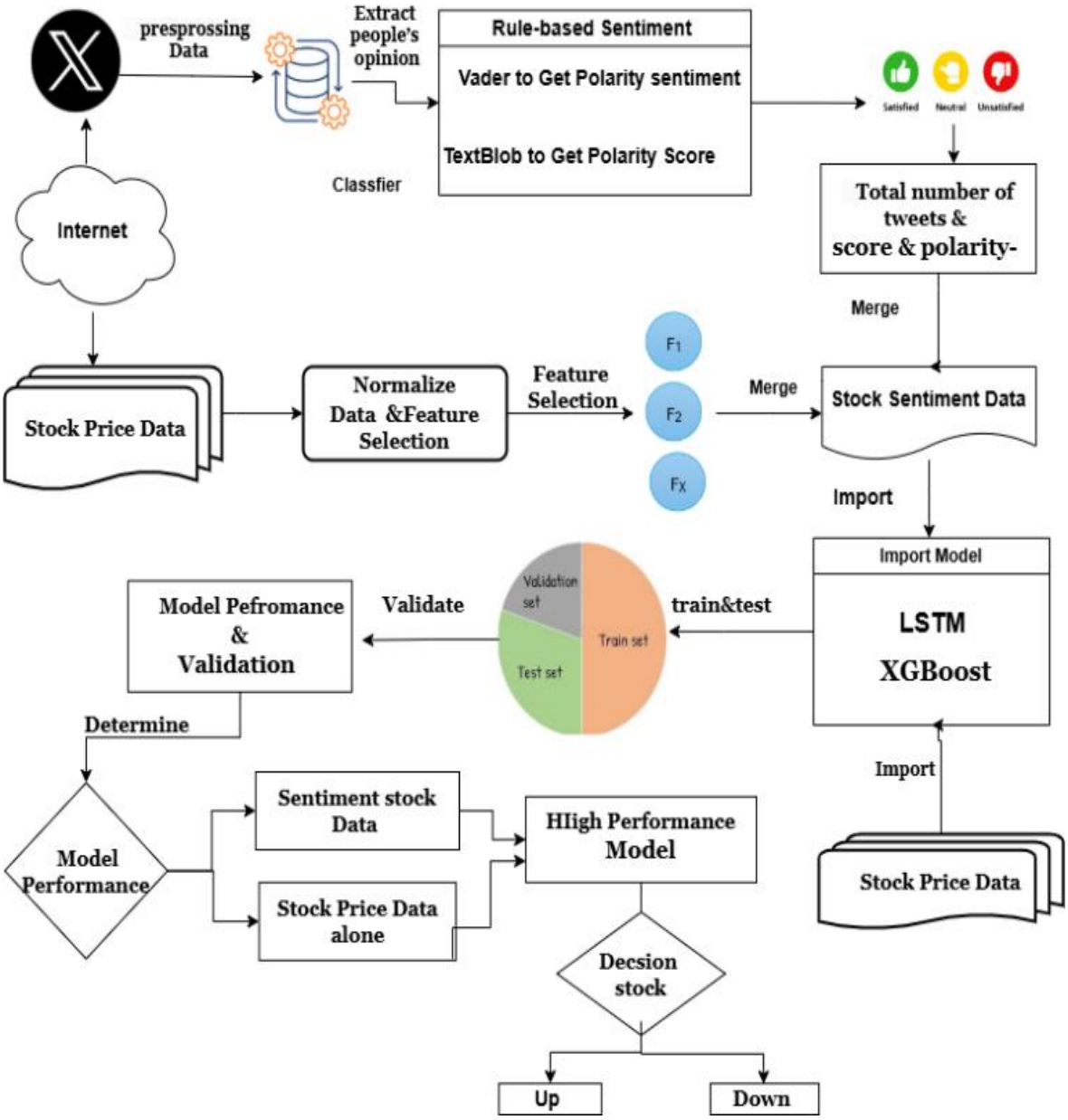


Fig. 1. The framework of stock prediction based on sentiment analysis.

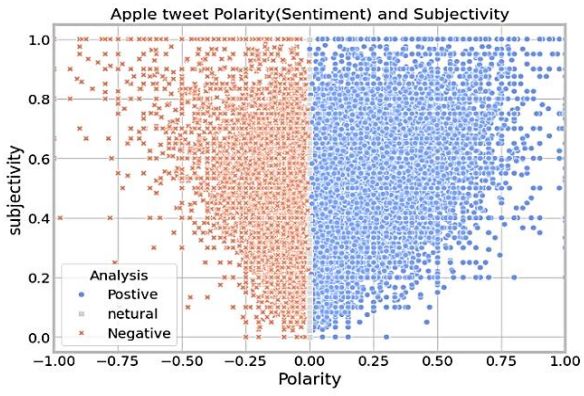


Fig. 2. Polarity vs. subjectivity.

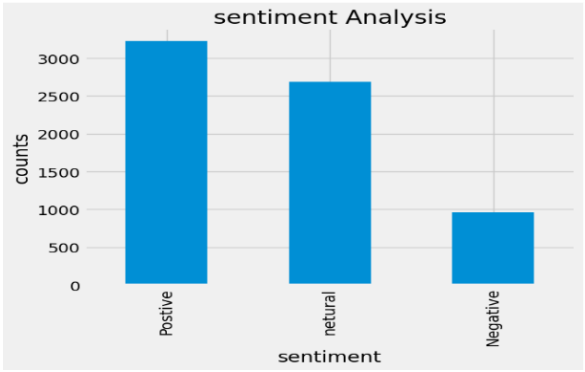


Fig. 3. Classifier of Apple tweet.

Fig. 4. The most positive words.

Fig. 4 illustrates the most common and frequent words in positive tweets about Apple, including “new”, “good”, and “read”, “Thanks”, which are strongly associated with positive sentiments.

Fig. 5. The most negative words.

Fig. 5 illustrates the most common and frequent words in negative tweets about Apple, such as “hate”, “worst”, “sorry”, “fail”, and “stupid”, and “stupid”, which are associated with negative sentiment.

The most common and frequent words in positive tweets about Google, including “top”, “share”, and “competition”, “penny”, are strongly associated with positive sentiments.

The most common and frequent words in negative tweets about Google, including “worst”, “pain”, and “less”, “spam”, are strongly associated with negative sentiments.

The most common and frequent words in positive tweets about Tesla, including “best”, “trade”, and “right”, “good”, are strongly associated with positive sentiments.

The most common and frequent words in negative tweets about Tesla, including “bad”, “wrong”, and “crazy”, “tesla”, are strongly associated with negative sentiments.

Three new features were added to the Apple, Google, and Tesla stock price datasets to evaluate the efficiency of future stock values. The first feature is total price fluctuation, which indicates the overall variation at the end of the day. This is measured by subtracting the volume value from the close value. The second feature, price difference is calculated by subtracting today's adjusted close value from the value from the previous day. Depending on the price difference feature, the direction of the stock price (Up or Down) is detected and stored in the trend score feature.

The price difference feature detects and stores the stock price's direction (up or down) in the trend score, which is the last feature. If the price difference is positive, the trend score is equal to 1 (up), and if the price difference is negative, the trend score is 0 (down), as illustrated in the snapshot of Table I.

TABLE I. SNAPCHAT ADDED THREE FEATURES TO THE STOCK PRICE

Date	Open	High	Low	Close	Adj Close	Volume	Total_price_fluctuation	Price_difference	Trend
2015-01-06	26.635000	26.857500	26.157499	26.56001	23.779427	263188400	6.991600e+09	0.002501	1
2015-01-07	26.799999	27.049999	26.67499	26.937500	24.112865	160423600	4.321411e+09	0.372499	1
2015-01-08	27.307501	28.037500	27.174999	27.972500	25.039347	237458000	6.642294e+09	1.035000	1
2015-01-09	28.167500	28.312500	27.552500	28.002501	25.066191	214798000	6.014881e+09	0.030001	1
2015-01-12	28.150000	28.157499	27.200001	27.312500	24.448545	198603200	5.424350e+09	-0.690001	0
2015-01-13	27.857500	28.200001	27.227501	27.555000	24.665621	268367600	7.394869e+09	0.242500	1
2015-01-14	27.260000	27.622499	27.125000	27.450001	24.571625	195826400	5.375435e+09	-0.104999	0
2015-01-15	27.500000	27.514999	26.665001	26.705000	23.904747	240056000	6.410695e+09	-0.745001	0
2015-01-16	26.757500	26.895000	26.299999	26.497499	23.719000	314053200	8.321624e+09	-0.207501	0
2015-01-20	26.959999	27.242500	26.625000	27.180000	24.329939	199599600	5.425117e+09	0.682501	1
2015-01-21	27.237499	27.764999	27.067499	27.387501	24.515686	194303600	5.321490e+09	0.207501	1
2015-01-22	27.565001	28.117500	27.430000	28.100000	25.153471	215185600	6.046715e+09	0.712499	1

The sentiment scores, polarity, and total number of **X** tweets are combined with the updated Apple, Google, and Tesla stock price datasets by matching dates to check for a correlation between people's opinions and the total fluctuation.

The first experiment tests the correlation to determine the extent to which people's opinions influence the overall fluctuation. The second experiment is to predict the expected stock price using the LSTM and XGBoost algorithms. The performance of the used algorithms is evaluated and ranked.

a) The first experiment: Fig. 6 and Fig. 7 depict the movements of both positive and negative tweets about the stock price. Depending on Fig. 6, it is obvious that the positive tweet movements pretty much match the stock price movements. However, the movement of negative tweets barely relates to the movements of the stock price. Thus, generally, we can deduce that there is a relationship between the tweets and stock market movements.

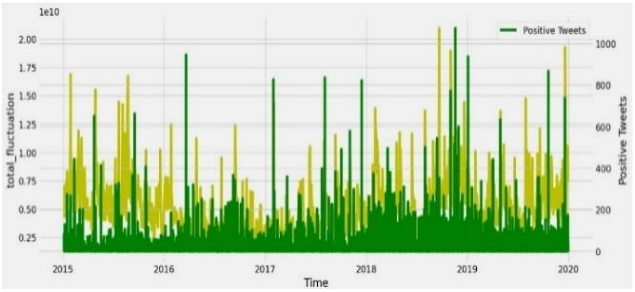


Fig. 6. Positive tweet movement.

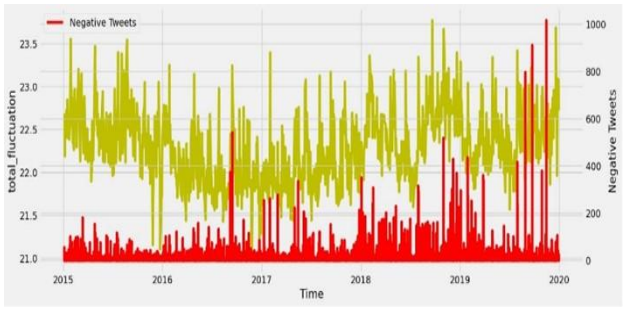


Fig. 7. Negative tweet movement.

To figure out to what extent the tweets can affect the stock price, the second experiment is conducted. The results show that positive people's opinions can affect the stock price by 47.56%. However, a negative correlation of 4.9% exists between people's negative tweets and the stock price. This means that when the tweets are negative, the stock price increases.

b) Predict stock prices utilizing LSTM and XGBoost models by integrating sentiment analysis: This approach utilizes the Long Short-Term Memory (LSTM) and XGBoost algorithms on a combined dataset, which includes the total number of tweets, polarity, and sentiment score, as input to predict the future stock prices of Apple, Google, and Tesla. Then take the classifier tweet, which can be positive, negative, or neutral, and calculate the average polarity and total number of tweets which sum up the number of tweets per day, then merge it with the closing price. as presented in Fig. 8.

Fig. 8. The number of tweets per day vs. polarity vs. close price.

To predict the stock price, the rolling window approach is applied to create numerous overlapping training samples through sliding fixed-size windows, which not only helps to improve the model's ability to observe the data but also aids in obtaining multiple patterns in the data. This approach is considered powerful and most suitable for LSTM and XGBoost prediction. The models analyze a tweet from **X** Twitter and predict future movements of the adjusted close price by referencing past days. The models normalized the input features (polarity score, total number of tweets, and close price) within the range of [0] to [1] as represented in Eq. (4).

$$CLSOE\ PRICE(x)normalize = \frac{x - min(X)}{max(X) - min(X)} \quad (4)$$

Then, the sentiment analysis features are integrated with historical data to figure a single dataset. And can represent this combined dataset as a shape (s), where s denotes the polarity score, (V) the total number of tweets, (X) the close price, (n) the end window size (y^t) the predicted value as introduced in Eq. (5) and Eq. (6). The system utilizes fixed-size windows with a length of n to generate a combined input for each time step (t). The last n-time step uses the close price and the polarity score (combined dataset) as input.

$$y^t = LSTM\ [x_t - n + 1, x_t - n + 2, \ldots., x_t,\ v_t - n + 1, v_t - n + 2, \ldots., v_t, s_t - n + 1, s_t - n + 2, , \ldots., s_t\ ] \quad (5)$$

Set the window size to three days of the combined dataset for feature engineering, which will serve as training to predict the next day. For instance, window 1 uses data from day 1 to day 3 (features) to forecast the next day. So, the predicted value.

$$y_{4=(}x_1, x_2, x_3\ )(\ v_1, v_2, v_3)\ (s_1, s_2, s_3) \quad (6)$$

The datasets are divided into a 90% training set and a 10% test set. Fig. 9 illustrates the division of the 90% training set into 80% training and 10% validation sets, utilizing the years 2015–2019 for training and 2019 for validation in the first dataset, utilizing the years 2015–2020 for training and 2020 for validation in the second dataset, and utilizing the June to October 2020 for training and 2020 for validation in the third dataset. Build the LSTM and XGBoost models by feeding it with training data. The LSTM models determine the input layer to obtain the sequence of previous time steps, which consists of three layers. The first layer, known as the input layer, receives window data in the form of a triangle (3, 3). The first (3) indicates the number of windows, while the second (3) indicates the number of three features (polarity scores, total number of tweets, and adjusted closing prices). The third layer is

responsible for processing time series data from previous observations of the rolling windows.

50 units make up the second layer, known as the LSTM layer, which uses gates to learn from time-dependent patterns in the data. Additionally, the design of memory cells, which also consist of 50 units, allows for the sequential data capture LSTM layer, comprising 50 units, to employ gates to learn from time-dependent patterns in the data. In addition to that, memory cells, which also consist of 50 units, process sequential data and capture temporal dependencies by adding more LSTM layers. Put Layer: one Dense Layer 1. The layer's design forecasts a single value (adjusted close price) for the upcoming time step. Then compile the model using the Adam optimizer. LSTM model is trained on data for a batch size of 16, which aids in obtaining single patterns by updating the model's weight after handling each sample. The number of epochs is assigned to 20 which indicates the number of loops in the training data.

The model is validated by monitoring while training the data, tracking the model's performance, adjusting the hyperparameters to enhance the performance, predicting the future stock price, and assessing the model using metrics like RMSE, R-squared, and MAE as illustrated in Eq. (7), Eq. (8) and Eq. (9).

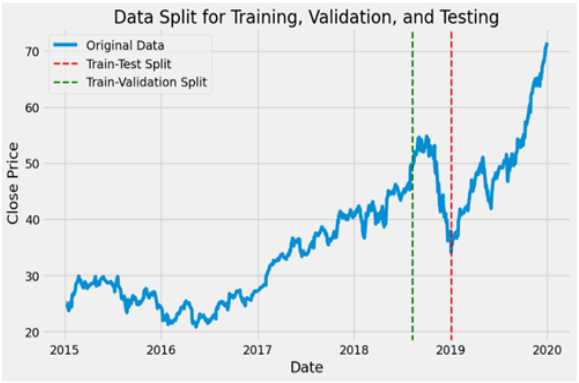


Fig. 9. The data split for sentiment analysis.

$$RMSE = \frac{1}{N}\sum_{i=1}^{n}(\text{yi} - \hat{\text{yi}})^2 \quad (7)$$

In machine learning, the R-squared value represents the coefficient of determination or the coefficient of multiple decisions in the context of multiple regression. Regression uses R squared as an evaluation metric to assess the scatter of data points around the fitted regression line. It indicates the percentage of variation in the dependent variable [42].

$$Squared\ Error = (\text{yi} - \hat{\text{yi}})2 \quad (8)$$

The proportion of the dependent variable's variance that the independent variable can explain is known as R-squared. R2 = the variance explained by the model. Total Variance The R-squared value stays between 0 and 100%. 0% corresponds to a model that fails to explain the variability of the response data around its meaning [43].

Mean Absolute Error (MAE) utilizing evaluation metrics, which indicates the difference between the actual and predicted value (value which is expected through a Model where actual values (target variable) [44] can be calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (9)$$

Predict Stock Price Using the LSTM and XGBoost Models Based on Historical Data

This study employs the LSTM and XGBoost models to predict future fluctuations in Apple, Google, and Tesla stock prices, relying solely on historical data. The data is gathered and preprocessed to evaluate the effectiveness of LSTM and XGBoost models, involves preparing the data for processing, which includes checking for any missing values by determining their meaning, removing any duplicated rows from the data, converting date columns to date-to-date time type, setting normalization values between [0:1], and dividing the datasets into three sets. The first set is for model training, the second for model validation, and the third for model testing. The training dataset (for The Model Training) makes up 70% of the total data; the validation dataset makes up 15%; and the testing dataset (for The Model Testing) makes up 15%. The model training uses the input sequences for LSTM. Each input sequence considers the output for the next step. Then set up the close price (target) as the preceding time step to predict the subsequent step to the close price. The regression approach is applied to the close price of Apple, Google, and Tesla stock. A single parameter, window size, determines the input format (close). The window size parameter establishes the days considered "dependent" for the stock price prediction. For example, with a window size of t = 60, the prediction will include the 60 days preceding day d. The longer the window, the better the model's view of previous data, but the higher the computing cost. Once we compute all sequence data, we organize each into a table of input features (X) and output targets (Y), which the model training process uses to adjust the LSTM model hyperparameters. The LSTM model consists of three layers: an input, an LSTM layer, and an output layer. Set the number of units to 50, the number of epochs to 100, and the batch size to 32. Then compile the model using the Adam optimizer and evaluate its performance using root means squared error (RMSE), R-squared, and MAE.

The same steps that were applied to evaluate the LSTM efficiency are used with XGBoost. Grid Search, Random Search, Max_Depth, Estimators, Sub-Sample, and L2 Regularization are used as hyperparameters.

V. Experimental Results

This section demonstrates the results obtained from two experiments. To detect the impact of combining sentiment analysis with historical data compared to historical data only. The previously mentioned algorithms are used to predict Apple, Google, and Tesla stock prices. As presented in Table II and figures from 10 -15, predicting stock prices based on integrating sentiment Analysis with historical data surpasses the performance of using historical data as presented in Table III and figures from 16 -21.

Fig. 10 displays the stock price prediction using XGBOOST by integrating Sentiment Analysis with historical data for Apple's stock, The model achieves a high accuracy rate of 99% from June 2017 to June 2020. It also achieves a low error rate between the actual and the predicted values for RMSE and MAE.

Fig. 11 predicted stock prices using XGBoost, which integrates sentiment analysis with historical data from 2018-7 to 2020-1 for Google. The model achieves a high accuracy of 95%, indicating a high correlation between actual and predicted values, and achieves a low prediction error of 0.0479 RMSE and 0.0358 with MAE, indicating a good fit.

TABLE II. MODELS PERFORMANCE-BASED INTEGRATION OF SENTIMENT ANALYSIS WITH HISTORICAL DATA

Model	APPLE	TESLA	GOOGLE
LSTM	R^2=91% RMSE=0.065 MAE=0.06	R^2=73% RMSE=0.1508 MAE=0.12	R^2=88% RMSE=0.0751 MAE=0.06
XGBoost	R^2=99% RMSE=0.042 MAE=0.0021	R^2=88% RMSE=0.1005 MAE=0.0816	R^2=95% RMSE=0.0479 MAE=0.0350

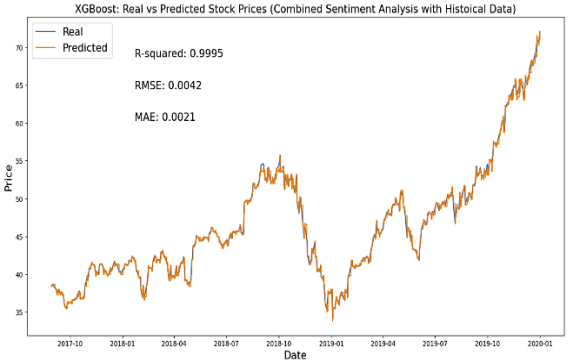


Fig. 10. XGBoost performance for Apple stock.

TABLE III. MODELS PERFORMANCE BY USING HISTORICAL DATA ALONE

Model name	APPLE	TESLA	GOOGLE
LSTM	R^2=98% RMSE=1.98 MAE=2.57	R^2=55% RMSE=0.45 MAE=6.94	R^2=95% RMSE=0.21 MAE=0.61
XGBoost	R^2=92% RMSE=1.85 MAE=1.29	R^2=99% RMSE=1.30 MAE=0.95	R^2=92% RMSE=1.85 MAE=1.29

Fig. 12 shows the stock price prediction using XGBOOST by integrating sentiment analysis with historical data for Tesla stock. The model's performance was 88.57%, which indicates a lower accuracy compared to another company's stock. It also achieved low error metrics, with RMSE (Root Mean Squared Error) at 0.1005, and MAE (Mean Absolute Error) at 0.0816 which indicates the difference between the predicted and actual values.

Fig. 13 Comparison of predicted stock prices using the LSTM Model, which integrates sentiment analysis with historical data for the Apple Company from 2017 to 2020. LSTM can capture the overall stock trend, but there are some gaps or errors in 2018 during high volatility.

Fig. 14 displays the integration of sentiment analysis with historical data for Google stock using the LSTM model. This model achieved good accuracy, at 88%. The model achieved a mean absolute error (MAE) of 0.06, indicating a low average error between the actual and the predicted values. And recorded in Root Mean Squared Error (RMSE): 0.0751.

Fig. 15 illustrates the comparison of predicted stock prices using the LSTM Model, which integrates sentiment analysis with historical data for the TSLA Company. The model achieves an accuracy of 75.48%, and achieves an RMSE of 0.1508, and an MAE of 0.12. This graph indicates the difference between the actual and predicted values. Often the predicted value fails to fit with the actual value.

Fig. 16 illustrates the performance of the XGBoost Model in predicting Apple's stock price. The model achieves an accurate rate of 92% with an average error of 1.29. These results declare a strong correlation between actual and predicted values.

Fig. 17 illustrates the performance of the LSTM Model for predicting Apple's stock price. The model achieves an accurate rate of 97% with an average error of 1.17 in RMSE and MAE = 1.27.

Fig. 18 compares the actual and predicted values of Google stock, using the XGBoost algorithm. The model achieved high accuracy of 92%, which indicates a strong relationship between actual and predictive values. XGBoost can closely track actual price movement while keeping low error metrics.

Fig. 19 displays Google's actual and predicted values, utilizing an LSTM (long short-term memory). LSTM has a high performance and accuracy of 95% and low error (RMSE = 0.21, MAE = 0.61). It appears that the predicted value is very close to the actual value, indicating that this model is a reliable tool for forecasting.

Fig. 20 compares Tesla's actual and predicted values, utilizing the XGBoost Model. The model achieves an accuracy of 99%, indicating a high performance and good fit between prediction and actual value. It achieves a low error metric of 1.30 of RMSE to 0.95 in MAE, the predicted value's deviation from the exact value.

Fig. 21 displays Tesla's actual and predicted values, utilizing an LSTM (long short-term memory). The model achieves an accuracy of 53%, indicating that its performance is not as high as it could be and requires additional data to improve. This is due to the insufficient training data set, which was from 01-08-2022 to 01-10-2022.

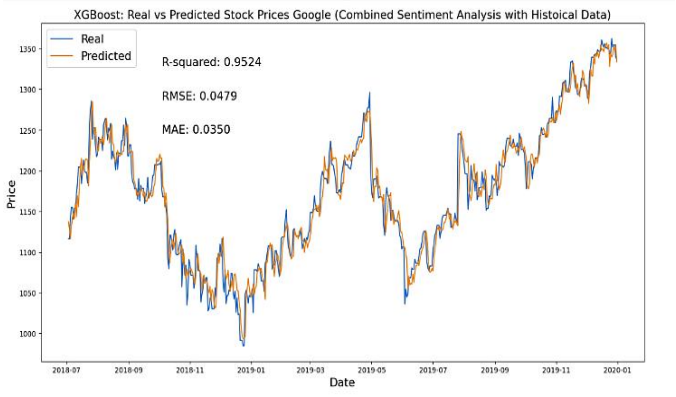


Fig. 11. XGBoost performance for Google stock.

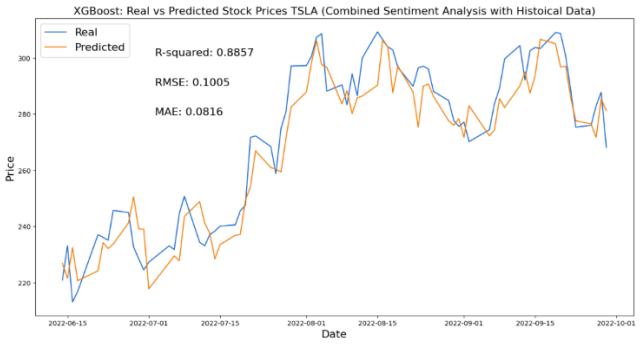


Fig. 12. XGBoost performance for Tesla stock.

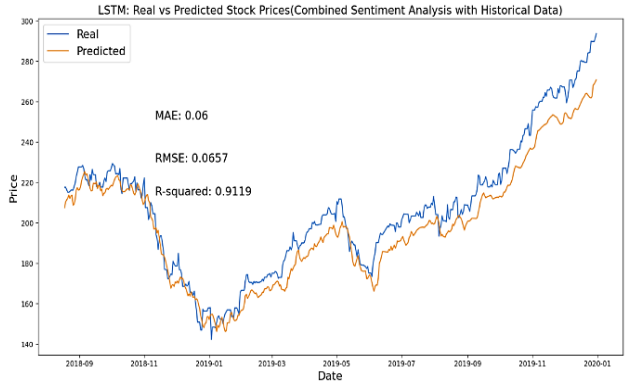


Fig. 13. LSTM performance for Apple stock.

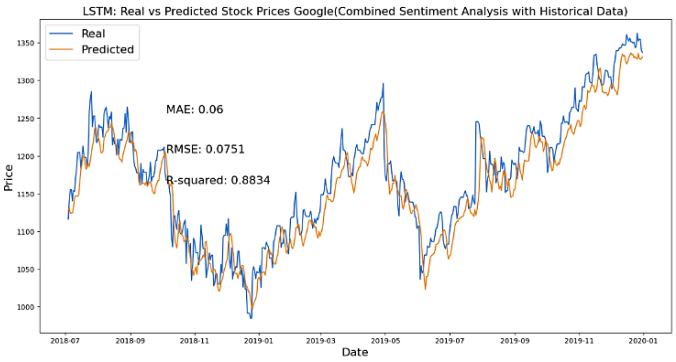


Fig. 14. LSTM performance for Google stock.

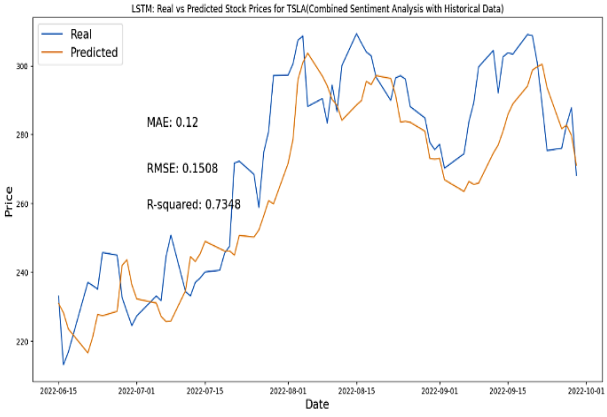


Fig. 15. LSTM performance for Tesla stock.

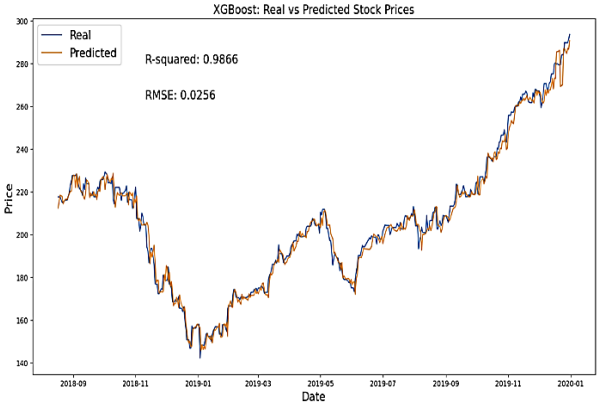


Fig. 16. Performance of XGBoost for Apple stock.

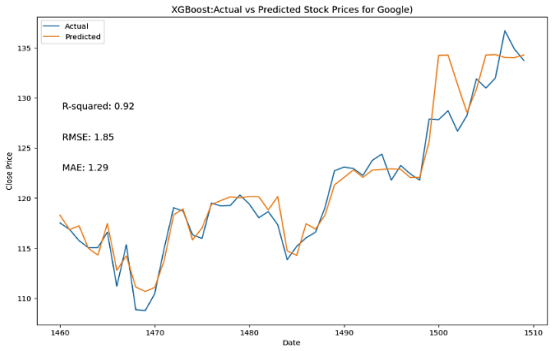


Fig. 17. Performance of XGBoost for Google stock.

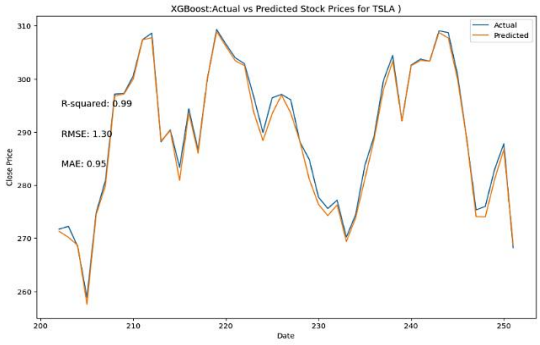


Fig. 18. Performance of XGBoost for Tesla stock.

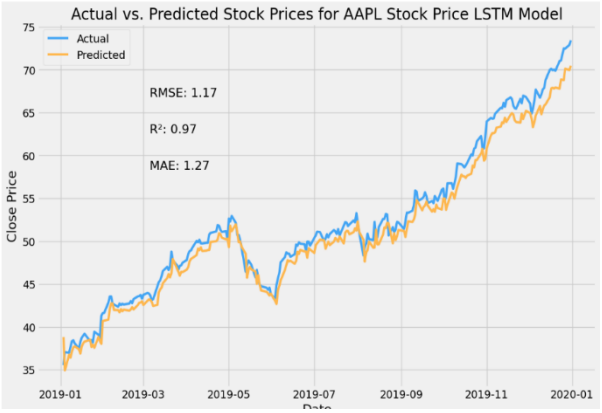


Fig. 19. Performance of LSTM for Apple stock.

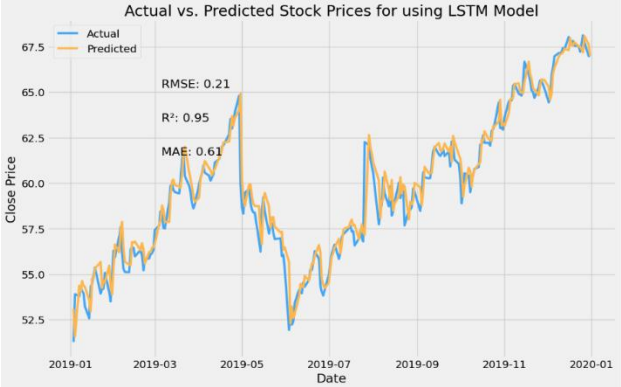


Fig. 20. Performance of LSTM for Google stock.

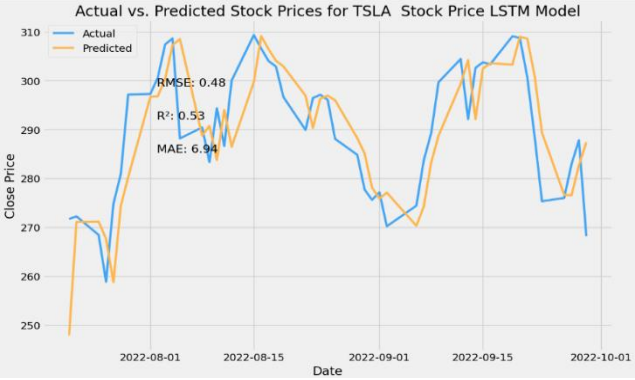


Fig. 21. Performance of LSTM for Tesla stock.

VI. Conclusion

This study proposed a framework to predict stock prices through integrated sentiment analysis with historical data, compared to using historical data only. Natural language processing (NLP) techniques extract sentimental information from X tweets. Then combine this sentiment analysis (positive, negative, or neutral) with the corresponding stock prices retrieved from Yahoo Finance. Machine learning and deep learning models test the combination of three stock datasets (Apple, Google, and Tesla) with and without sentiment analysis. This allows us to observe how integrating sentiment analysis with historical data correlates with each other and to what extent it improves performance accuracy. This study proved the correlation between stock price movement and people's opinions on social media. Two different classifiers are applied to predict the stock price, XGBoost and the LSTM models. The XGBoost outperformed the LSTM models in Apple, Google, and Tesla, achieving 99%, 95%, and 88%, respectively, by integrating sentiment analysis with historical data. On the other hand, LSTM outperformed XGBoost in Apple and Google, achieving 98% and 95%, except for Tesla, due to low data training when using historical data.

This study demonstrated the significance of combining and analyzing people's opinions, which is qualitative data that contributes to improving the understanding of integrating opinions with historical data, which is quantitative data. Both LSTM and XGBoost have demonstrated their efficacy.

References

[1] T. Julian, T. Devrison, V. Anora and K. M. Suryaningrum, "Stock Price Prediction Model Using Deep Learning Optimization Based on Technical Analysis Indicators," Elsevier Procedia Computer Science, vol. 227, pp. 939--947, 2023.

[2] K. Pawar, R. S. Jalem and V. Tiwari, "Stock price prediction based on deep neural networks," in Springer Emerging Trends in Expert Applications and Security: Proceedings of ICETEAS, 2019.

[3] Z. Wang, S.-B. Ho and Z. Lin, "Stock market prediction analysis by incorporating social and news opinion and sentiment," in IEEE International Conference on Data Mining Workshops (ICDMW), 2018.

[4] H.-T. Duong and T.-A. Nguyen-Thi, "A review: preprocessing techniques and data augmentation for sentiment analysis," Springer Computational Social Networks, vol. 8, 2021.

[5] S. Hamed, M. Ezzat and H. Hefny, "A review of sentiment analysis techniques," International Journal of Computer Applications, vol. 176, pp. 20-24, 2020.

[6] Z. Drus and H. Khalid, "Sentiment analysis in social media and its application: Systematic literature review," Elsevier Procedia Computer Science, vol. 161, pp. 707-714, 2019.

[7] B. Ramalho, J. Jorge and S. Gama, "Representing uncertainty through sentiment and stance visualizations: A survey," Elsevier Graphical Models, vol. 129, p. 101191, 2023.

[8] Z. Wang, S.-B. Ho and Z. Lin, "Stock market prediction analysis by incorporating social and news opinion and sentiment," IEEE International Conference on Data Mining Workshops (ICDMW), pp. 1375-1380, 2018.

[9] A. Bhardwaj, Y. Narayan and M. Dutta, "Sentiment analysis for Indian stock market prediction using Sensex and nifty," Elsevier Procedia computer science, vol. 70, pp. 85-91, 2015.

[10] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," Neural Computing and Applications, vol. 32, pp. 1609--1628, 2020.

[11] N. Ghatasheh, I. Altaharwa and K. Aldebei, "Modified genetic algorithm for feature selection and hyper parameter optimization: case of XGBoost in spam prediction," IEEE Access, pp. 84365-84383.

[12] T. Liwei, F. Li, S. Yu and G. Yuankai, "Forecast of LSTM-XGBoost in Stock Price Based on Bayesian Optimization," Intelligent Automation & Soft Computing, vol. 29, pp. 855-868, 2021.

[13] H. Z. Wei Chen, Mukesh Kumar Mehlawat and Lifen Jia, "Mean–variance portfolio optimization using machine learning-based stock price prediction," Applied Soft ComputingApplied Soft Computing, vol. 100, p. 106943, 2021.

[14] Q. Wang, Y. Ma, K. Zhao and Y. Tian, "A comprehensive survey of loss functions in machine learning," Annals of Data Science, pp. 1-26, 2020.

[15] F. Balaneji and D. Maringer, "Applying Sentiment Analysis, Topic Modeling, and XGBoost to Classify Implied Volatility," in Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr), 2022.

[16] F. Balaneji and D. Maringer, "Applying Sentiment Analysis, Topic Modeling, and XGBoost to Classify Implied Volatility," in Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr), 2022.

[17] S. Selvin and R. Vinayakumar, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in international conference on advances in computing, communications and informatics (icacci), 2017.

[18] Q. M. Abdul, K. Sanjit, J. A. Chris, Arun Kumar Sivaraman, S. H. Kong Fah Tee and Janakiraman N, "Novel optimization approach for stock price forecasting using multi-layered sequential LSTM," Applied Soft Computing, vol. 134, p. 109830, 2023.

[19] U. D. Gandhi, P. Malarvizhi Kumar, G. Chandra Babu and G. Karthick, "Sentiment Analysis on Twitter Data by Using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM)," Wireless Personal Communications, 2021.

[20] F. Huang, X. Li, C. Yuan, S. Zhang, J. Zhang and S. Qiao, "Attention-emotion-enhanced convolutional LSTM for sentiment analysis," transactions on neural networks and learning systems, vol. 33, pp. 4332-4345, 2021.

[21] S. Mehtab, J. Sen, A. Dutta, S. M. Thampi, S. Piramuthu, K.-C. Li, S. Berretti, M. Wozniak and D. Singh, "Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models," in Machine Learning and Metaheuristics Algorithms, and Applications, 2021.

[22] A. Chennupati, B. Prahas, B. A. Ghali, B. D. Jasvitha and K. Murali, "Comparative Analysis of Bitcoin Price Prediction Models: LSTM, BiLSTM, ARIMA and Transformer," pp. 1--7, 2024.

[23] W. Chen, C. K. Yeo, C. T. Lau and B. S. Lee, "Leveraging social media news to predict stock index movement using RNN-boost," Data & Knowledge Engineering, pp. 14-24, 2018.

[24] H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil, M. M. Selim and K. Muhammad, "A local and global event sentiment based efficient stock exchange forecasting using deep learning," International Journal of Information Management, vol. 50, pp. 432-451, 2020.

[25] Y. Wang, Y. Liu, M. Wang and R. Liu, "LSTM model optimization on stock price forecasting," in 17th international symposium on distributed computing and applications for business engineering and science (dcabes), 2018.

[26] M. Qasem, R. Thulasiram and P. Thulasiram, "Twitter sentiment classification using machine learning techniques for stock markets," in International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015.

[27] M. F. Bashri and R. Kusumaningrum, "Sentiment analysis using Latent Dirichlet Allocation and topic polarity wordcloud visualization," in 5th International Conference on Information and Communication Technology (ICoIC7), 2017.

[28] M. Fikri and R. Sarno, "A comparative study of sentiment analysis using SVM and SentiWordNet," Indonesian Journal of Electrical Engineering and Computer Science, vol. 13, pp. 902-909, 2019.

[29] A. Segura Navarrete, C. Vidal-Castro, C. Rubio-Manzano and C. Martínez-Araneda, "The role of WordNet similarity in the affective analysis pipeline," Computación y Sistemas, vol. 23, pp. 1021-1031, 10 2019.

[30] R. Ren, D. D. Wu and T. Liu, "Forecasting stock market movement direction using sentiment analysis and support vector machine," IEEE Systems Journal, vol. 13, pp. 760-770, 2018.

[31] R. Chen and R. Dong, "The Relationship Between Twitter Sentiment and Stock Performance: A Decision Tree Approach," 2023.

[32] S. Mohan, S. Mullapudi, S. Sammeta, P. Vijayvergia and D. C. Anastasiu, "Stock price prediction using news sentiment analysis," in fifth international conference on big data computing service and applications (BigDataService), 2019.

[33] J. Maqbool, P. Aggarwal, R. Kaur, A. Mittal and I. A. Ganaie, "Stock prediction by integrating sentiment scores of financial news and MLP-regressor: A machine learning approach," Procedia Computer Science, vol. 218, pp. 1067-1078, 2023.

[34] J. Deveikyte, H. Geman, C. Piccari and A. Provetti, "A sentiment analysis approach to the prediction of market volatility," Frontiers in Artificial Intelligence, vol. 5, p. 836809, 2022.

[35] I. K. Nti, A. F. Adekoya and B. A. Weyori, "Predicting stock market price movement using sentiment analysis: Evidence from Ghana," Applied Computer Systems, vol. 25, pp. 33-42, 2020.

[36] J. Long, Z. Chen, W. He, T. Wu and J. Ren, "An integrated framework of deep learning and knowledge graph for prediction of stock price trend: An application in Chinese stock exchange market," Applied Soft Computing, vol. 91, p. 106205, 2020.

[37] S. V. Kolasani and R. Assaf, "Predicting stock movement using sentiment analysis of Twitter feed with neural networks," Journal of Data Analysis and Information Processing, vol. 8, pp. 309-319, 2020.

[38] A. Oad, I. Koondhar, P. Butt, M. Ahmed and S. Bhutto, "VADER sentiment analysis without and with English punctuation marks," Int. J. Adv. Trends Comput. Sci. Eng, vol. 10, pp. 1483--1488, 2021.

[39] M. Isnan, G. N. Elwirehardja and B. Pardamean, "Sentiment analysis for TikTok review using VADER sentiment and SVM model," Procedia Computer Science, vol. 227, pp. 168-175, 2023.

[40] E. Rosenberg, C. Tarazona, F. Mallor, H. Eivazi, D. Pastor-Escuredo, F. Fuso-Nerini and R. Vinuesa, "Sentiment analysis on Twitter data towards climate action," Results in Engineering, vol. 19, p. 101287, 2023.

[41] G. Murthy, S. R. Allu, B. Andhavarapu, M. Bagadi and M. Belusonti, "Text based sentiment analysis using LSTM," International Journal of Engineering Research & Technology, vol. 9 Issue 05, 2020.

[42] N. Makhoul, "Review of data quality indicators and metrics, and suggestions for indicators and metrics for structural health monitoring," Advances in Bridge Engineering, vol. 3, p. 17, 2022.

[43] M. Steurer, R. J. Hill and N. Pfeifer, "Metrics for evaluating the performance of machine learning based automated valuation models," Journal of Property Research, vol. 38, pp. 99-129, 2021.

[44] D. Chicco, M. Warrens and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," vol. 7, p. PeerJ Computer Science, 05 07 2021.

[45] P. Sandhya, R. Bandi and D. D. Himabindu, "Stock price prediction using recurrent neural network and lstm," 2022.

[46] P. Mukherjee, Y. Badr, S. Doppalapudi, S. M. Srinivasan, R. S. Sangwan and R. Sharma, "Effect of negation in sentences on sentiment analysis and polarity detection," Procedia Computer Science, vol. 185, pp. 370--379, 2021.

A Multimodal Data Scraping Tool for Collecting Authentic Islamic Text Datasets

Abdallah Namoun[1], Mohammad Ali Humayun[2], Waqas Nawaz[3]
AI Center, Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, 42351, Saudi Arabia[1, 3]
Department of Artificial Intelligence, Information Technology University, Lahore, 54600, Pakistan[2]

***Abstract*—Making decisions based on accurate knowledge is agreed upon to provide ample opportunities in different walks of life. Machine learning and natural language processing (NLP) systems, such as Large Language Models, may use unrecognized sources of Islamic content to fuel their predictive models, which could often lead to incorrect judgments and rulings. This article presents the development of an automated method with four distinct algorithms for text extraction from static websites, dynamic websites, YouTube videos with transcripts, and for speech-to-text conversion from videos without transcripts, particularly targeting Islamic knowledge text. The tool is tested by collecting a reliable Islamic knowledge dataset from authentic sources in Saudi Arabia. We scraped Islamic content in Arabic from text websites of prominent scholars and YouTube channels administered by five authorized agencies in Saudi Arabia. These agencies include the general authority for the affairs of the grand mosque and the prophet's mosque and charitable foundations in Saudi Arabia. For websites, text data were scraped using Python tools for static and dynamic web scraping such as Beautiful Soup and Selenium. For YouTube channels, data were scraped from existing transcripts or transcribed using automatic speech recognition tools. The final Islamic content dataset comprises 31225 records from regulated sources. Our Islamic knowledge dataset can be used to develop accurate Islamic question answering, AI chatbots and other NLP systems.**

Keywords*—*Web scraping; Islamic knowledge; machine learning; natural language processing; question and answering; AI chatbots

I. Introduction

We propose a multimodal web scraping tool to collect Islamic content from verified websites and YouTube channels administered by five trusted entities in Saudi Arabia. Our crawling algorithms use Python libraries, including Beautiful Soup and Selenium, to scrape content from static and dynamic websites and channels. We selected input sources from official websites that are administered and maintained by non-profit organizations (e.g., agencies and charities) in Saudi Arabia. This strategy is adopted to achieve two important objectives: first, to ensure the credibility and correctness of the Islamic data published on these sources, and second, to eliminate noise data from untrusted sources, which could sabotage the quality of our proposed dataset.

Next, we exemplify the application of our tool on Islamic data from authentic Islamic knowledge websites that are maintained by 1) authorized non-profit organizations represented by the General Authority for the affairs of the Grand Mosque and the Prophet's Mosque (also known as Alharamain) or 2) under the management of charitable organizations in Saudi Arabia for scholars who are past or present members of the Permanent Committee for Scholarly Research and Ifta and/or Saudi Council of Senior Scholars.

To contextualize our research problem, let us consider the subsequent realistic scenario, which describes the typical challenges faced by non-Arabic-speaking people when searching for Islamic knowledge and teaching to guide their lifestyle decisions. Decisions may sometimes impact a person's life, as outlined below.

"Ahmed is a 35-year-old married government employee who has recently faced a family hardship concerning the decision to abort his child due to health concerns regarding his wife. He comes from a small village in Pakistan, and he does not speak the Arabic language. However, as a devoted Muslim, Ahmed is never satisfied with simple answers and always strives to make knowledge-driven decisions in search of satisfaction and tranquility. Ahmed starts by asking his village's main scholar at the nearest mosque; however, he remains unconvinced about the answers he received. In his quest for more resounding answers, he stumbles across a QA NLP platform, which offers opinion-oriented knowledge translated into various languages. Luckily, Ahmed can query the online platform on his mobile device, using plain text in Urdu, about the topic of child abortion in Islam under different circumstances and scenarios. The language-driven system presents the answers in a concise form with pieces of evidence from different scholars by extrapolating its unique and authentic dataset of Islamic knowledge. Not only can Ahmed read the answers in Urdu, but he can also view different opinions accompanied by arguments linked to the sources of the Quran and Sunnah. Ahmed can now discuss these options openly with his family before choosing the best option given his family's circumstances."

Reading this simple hypothetical scenario, various requirements emerge that should be addressed to effectively interpret Islamic knowledge.

- The non-Arabic speaking community constantly needs to access information using modern technologies, such as mobile devices.
- Intelligent NLP systems require reliable Islamic data to be able to produce meaningful and trustworthy knowledge.
- Our literature review (in the next section) revealed the lack of datasets that serve the objectives of the above scenario. Moreover, there are other scientific motivations

This research is supported by the Deanship of Scientific Research of the Islamic University of Madinah, KSA under the research groups (first) project no. 956.

and advantages to creating our unique web content collection tool and Islamic dataset, as elaborated below.

- We developed an automated data collection tool that gathers and organizes multimodal data from publicly available, reliable Islamic text and audio-video sources on the web, with minimal manual intervention, and compiles it into a text dataset. Fellow researchers can customize our tool to scrape similar content from online sources.
- We provide an authentic dataset of Islamic knowledge extracted from approved and regulated online sources. This uniqueness distinguishes our dataset from published Islamic content datasets. The proposed dataset can be used to form the foundation for additional authentic datasets and extend existing Islamic ruling datasets. Researchers from different specializations (e.g., Islamic studies, theology, sociology, and history) may use the datasets to conduct research studies and perform rigorous analyses.
- Our Islamic knowledge dataset may be reused by fellow machine learning researchers for 1) developing and testing machine learning models of generative AI systems for understanding and issuing Islamic rulings, 2) extracting the reasons behind certain rulings and judgments, 3) comparing rulings among various schools of thoughts and regions to understand their commonalities and differences, 4) exploring the evolution of Islamic laws and rulings over time in response to the contemporary technological and social changes, and 5) investigating the current issues in Muslim communities. Examples of natural language processing (NLP) applications on Islamic content and text include machine translation, question answering, Islamic rulings classification, text summarization, text analytics, auto-diacritization, and smart assistants and AI chatbots.
- Providing authentic Islamic datasets from trusted sources can assist policymakers, local organizations and centers, and the private sector to revise and/or create their policies around the Islamic principles within the datasets. For instance, financial products and services in banks may be offered using policies extracted from this Islamic knowledge. Similarly, the food industry may utilize the certifications and regulations pertaining to halal products from the knowledge extracted from these datasets. It may also guide courts and legal practitioners in Muslim countries to make informed decisions and rulings.
- Digitizing and archiving authentic datasets of Islamic knowledge, including rare texts and manuscripts, helps preserve the Islamic culture and heritage for the next generations. Moreover, these datasets can be translated into other languages to benefit non-Arabic audiences and communities.
- This Islamic dataset can be used to promote interfaith dialogue by providing accurate and authentic information about Islamic beliefs, practices, and rulings, especially concerning contemporary issues. It creates an opportunity to enhance public knowledge about Islam, reducing misconceptions and prejudice.

II. Related Work

Collecting and compiling authentic and reliable datasets of Islamic content and knowledge is crucial for training and validating natural language processing (NLP) and generative artificial intelligence (GenAI) models [1] which can be subsequently consumed by ordinary users and local entities (e.g., courts). However, the Internet offers various unverified sources of Islamic content hosted on several websites and social media. It is challenging to judge the accuracy and relevance of such Islamic knowledge to train relevant generative artificial intelligence models.

Indeed, there is a lack of trustworthy Islamic datasets that discuss essential topics and perspectives on societal issues concerning Muslim communities. A recent survey compared 11 question-and-answering (Q&A) Islamic datasets published between 2014 and 2022 [2]. All these corpora covered Quran questions, overlooking other important sources of Islamic knowledge and rulings. The authors in study [3] introduced an Islamic dataset (i.e., eiad) for building English question-answer AI chatbots. The dataset is in English and covers 15 categories, targeting converts and non-Muslims. The dataset comprises 10000 articles collected from three major, trusted websites like IslamQA.com. Similarly, the authors in study [4] contribute a dataset (i.e., QASiNa) in the Indonesian language for question-answering tasks. This dataset is unique since it is based on evidence from the Sirah literature (i.e., the Prophet Muhammed practices and sayings). Such datasets are useful for training large language models, such as ChatGPT and Gemini.

The study published recently in [5] is one of the exceptions. It is claimed to be the first Islamic rulings (i.e., is called fatwaset in Arabic) dataset. The data were collected from 13 trusted websites in Arab countries, including Saudi Arabia, Egypt, Jordan, Qatar, and Syria. The authors performed an exploratory data analysis revealing a total of 130182 records. Authors in study [1] suggest a dataset for developing chatbot systems for the issuance of Islamic fatwas. The dataset is claimed to be the largest Fatwas dataset, with the classification of topics. In total, the dataset has approximately 850000 fatwas, scraped from different websites, regions, and schools of thought. It is observed that most Fatwas (71%) were extracted from AskFM.

Authors in study [6] proposed an Arabic Multi-IsnadSet (MIS) dataset as a Neo4j multi-directed graph comprising 2029 narrator nodes and 77797 sanad-hadith connections. Hadith include the sayings, actions, or silent approvals of the Prophet Muhammad. Sanad represents the credibility (through a chain of narrators) of each hadith in the dataset. Hadith is the second source of religious legislation in Islam, after Quran. Data scraping tools were used to fetch hadith details, such as hadith number, content, list and sequence of narrators, and Isnad count. This dataset allows researchers to understand the authenticity of each hadith and the strength of its narration. Similarly, authors in [7] compiled a dataset of 650K hadiths, named Sanadset. The proposed dataset collected from 926 Arabic books may be used for automatic sanad verification and classification. However, the dataset does not distinguish between the authenticity of each hadith.

Although the existing works have collected diverse datasets across multiple domains, a unified dataset comprising data from multimodal sources, such as video lectures and text sources of different nature is still lacking. Besides, most of the works have relied on manual compilation and organization of datasets, disregarding automation strategies for dataset scraping and compilation. Furthermore, insufficient effort has been made to ensure that the sources are authentic and verified by an authoritative body. These limitations affect the ability of these datasets to adapt to future needs.

This work aims to fill this gap by designing an automation strategy for dataset collection. Moreover, the strategy incorporates datasets from multimodal sources, such as video and text resources, ensuring that all sources are authentic. This approach not only makes the dataset both authentic and diverse but also facilitates its easy expansion and adaptation for multiple domains.

III. Materials and Methods

Our web sources from which we curated our dataset are authenticated and administered by organizations and foundations in Saudi Arabia are listed in Table I. Our dataset is publicly accessible at [8]. Fig. 1 shows samples of Youtube channels.

TABLE I. Target Reliable Sources of Islamic Knowledge

Name	Administration	URL Links
Presidency of the Two Holy Mosques	General Authority for the affairs of the Grand Mosque and the Prophet's Mosque	gph.gov.sa/index.php/ar/
		manaratalharamain.gov.sa/speeches/
		youtube.com/@twjehDM
		youtube.com/@makkah
		youtube.com/@SaudiQuranTv
Presidency of the Two Holy Mosques (Manarat Al Haramain)	Agency of the Affairs of Al-Masjid Al-Nabawi	wmn.gov.sa/public/
		youtube.com/@wmngovksa
		youtube.com/@wmngovsa
		youtube.com/@SaudiSunnahTv
Ibn Baz	Sheikh Abdul Aziz bin Baz Charitable Foundation	binbaz.org.sa/
Ibn Othaimeen	Sheikh Mohammed bin Saleh Al Othaimeen Charitable Foundation	binothaimeen.net/
		youtube.com/channel/UCtF3YygTiodnYSw8vD3UJtQ
Alfawzan	Al Dawa Charitable Foundation	alfawzan.af.org.sa/ar
		youtube.com/@salihalfawzan
		youtube.com/@aforgsa1

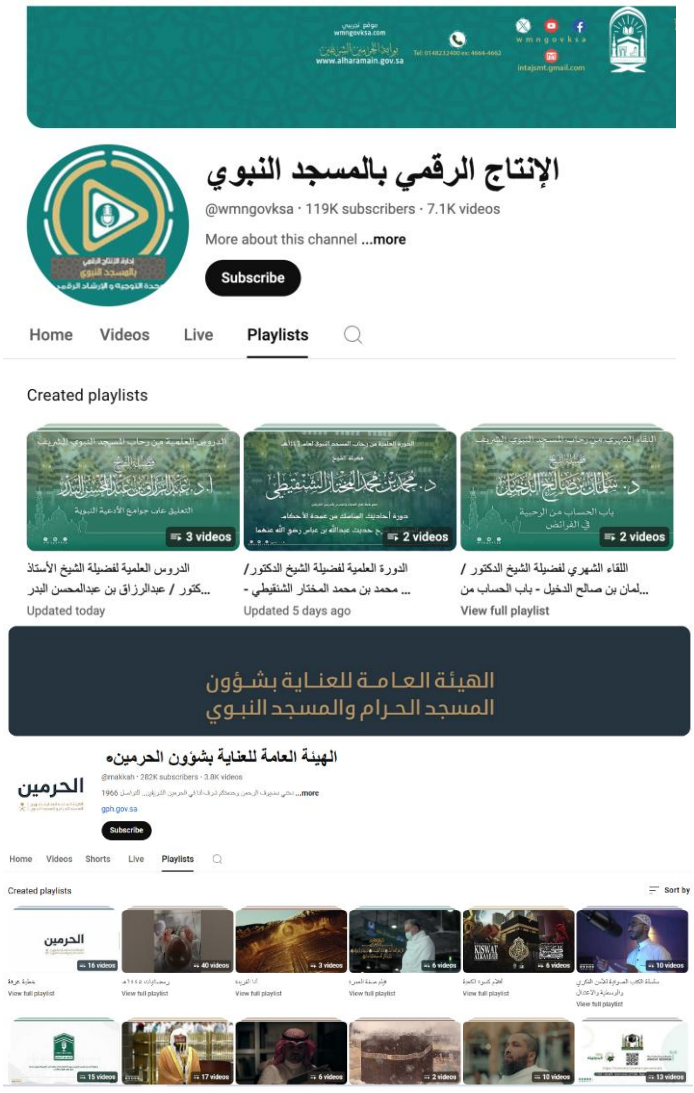
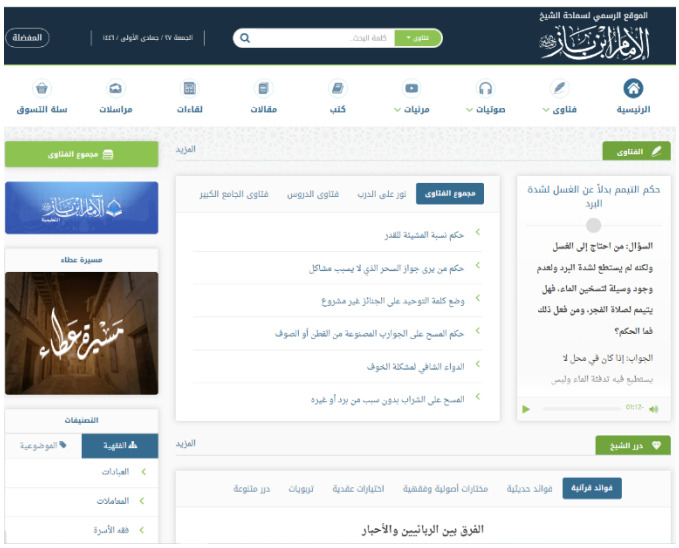


Fig. 1. A sample of YouTube channels.



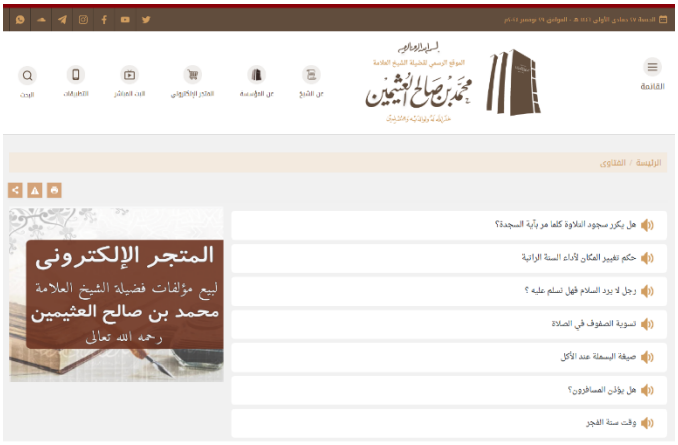


Fig. 2. A sample of Islamic websites.

TABLE II. TYPE OF DOWNLOADABLE DATA FROM NON-PROFIT ORGANIZATIONS

Data Category	Bin Othaimeen (Mix)	Bin Baz (Mainly Text)	Alfawzan (Audio)
Articles المقالات	No	Yes	Yes
Lessons الدروس	Yes	Yes	Yes
Lectures المحاضرات	Yes	No	Yes
Meetings and Dialogues اللقاءات والحوارات	Yes	Yes	No
Rulings الفتاوى	Yes	Yes	Yes
Speech الكلمات	No	No	No
Letters - Communications الرسائل – المراسلات– الخطابات	No	Yes	No
Sermons (not on Fridays) الخطب	Yes	No	Yes
Friday Sermons خطب الجمعة	No	No	No

We applied the following inclusion and exclusion criteria (see Table III) to identify the sources of our Islamic dataset. When it was not possible to retrieve the text or transcripts, we used speech-to-text (Arabic audio transcription) API services to transcribe the videos of Islamic knowledge. Table II shows type of downloadable data from non-profit organizations.

IV. SOFTWARE ALGORITHMS

Mainly, four different algorithms have been developed: one for extracting text data from static websites, another for retrieving text data from dynamic websites, a third for obtaining relevant transcripts from YouTube videos with available transcripts, and the last for translating speech-to-text from videos without a transcript. The code used is publicly available at [9] along with usage instructions, and the dataset has been uploaded to [9].

TABLE III. INCLUSION AND EXCLUSION CRITERIA

Criterion	Inclusion	Exclusion
Sources	Authentic sources that are maintained by local authorities	Unknown sources or not managed by authorized entities
Responsible	Organization, foundation, or charity	Individual
Purpose	Non-profit	Commercial
Country	Saudi Arabia	Other countries
Type of data	Opinions about different topics published in various form (e.g., speeches, letters, sermons, lectures, explanations, etc.)	Books, booklets, and handwritten manuscripts
Format	Downloadable text, audio, video files (with ability to scrape data)	Images
Language	Arabic	Other languages (English, Urdu, … etc)
Place of Publication	Knowledge published on official websites and YouTube channels	Knowledge posted on social media platforms, such as X (Twitter) platform and Facebook

A. Text Sources

Only the websites owned by reputable scholars from Saudi Arabia were selected for the text data. Consequently, the official websites of Ibn Baz, Ibn Othaimeen, and Alfawzan were chosen to fetch data. Since the websites have different structures and data types, the web scraping methodology [9] for fetching data from each website differed.

The website for Ibn Baz employs static HTML content, so the Python Beautiful Soup library was used to scrape data from this site by exploiting the HTML tags. The website has different categories of data including: الفتاوى (Islamic rulings), مقالات (articles), لقاءات وحوارات (discussions), خطابات ومراسل (speeches), and درر (pearls). First, all fatwas were fetched; these fatwas are subdivided into four major categories for each radio program نور فتاوى الجامع الكبير, and على الدرب , فتاوى لدروس, مجموع الفتاوى Each of these categories of fatwas has multiple pages. To extract data from each page, the base URL for the fatwa page was used, and then iterations were performed over the different program categories to construct the URLs for each program. The total number of pages varied for each fatwa category, and URLs were appended by iterating through the pages until the last page for each program was reached and scraped. Fig. 2 shows samples of Islamic websites.

The HTML data from the final URL for a page within the category was fetched using the Python requests library for HTTP requests. An HTTP request was sent to the URL, and the HTML response was filtered using Beautiful Soup to fetch the URLs for subpages pointing to the individual fatwas. The pages did not contain only URLs for the fatwa but also other URLs, such as those pointing to the homepage and other categories. Therefore,

only the URLs containing the keyword corresponding to the data category were filtered out. For example, URLs with fatwa had the keyword "fatwa" in them, so URLs were retained accordingly from each fatwa.

Subsequently, the filtered URLs were iterated, and HTTP requests were sent to each page to fetch the HTML response, which was again cleaned using Beautiful Soup. The HTML response from each fatwa page was analyzed using Beautiful Soup, and the contents for the following fatwa fields were fetched: title, question, answer, and category. HTML tags for each field were looked up, and the corresponding text was extracted. The fetched texts were stored in the dataset within different fields along with the main category (fatwa), the fatwa program, the page number, and the final URL for the fatwa.

For other text categories, such as مقالات (articles), لقاءات وحوارات (discussions), خطابات ومراسل (speeches), and درر (pearls), a similar process was applied to fetch the data with a few changes. These categories do not have different radio programs, so a single loop iterates over the page numbers within the main URL. After the final content URLs were scraped using Beautiful Soup and by filtering the correct keywords of each of these categories, the content URL was fetched. Pages of these categories did not have fields like fatwa category and question, so only title and text fields were compiled.

The dataset was cleaned to remove missing entries and duplicate entries added from multiple site URLs. Finally, due to different field structures, the dataset is stored as two separate CSV files: one for fatwas and one for other miscellaneous data.

The website for Alfawzan had similar static content, so the methodology was almost the same as the one for Ibn Baz website, with a few changes. Firstly, all fatwas on Alfawzan website have not been subdivided according to different programs but are placed under different pages in the main fatwa URL. Hence, the loop for adding page numbers was run with the main fatwa URL only. Secondly, most of the fatwas on the website have audio content and do not have text transcriptions, as indicated by a text file button being greyed out for fatwa without text content. Consequently, before fetching the reference of the fatwa URL and sending an HTTP request, a check was performed to see if the text button was not greyed out. Only then was the subsequent HTTP request sent to the URL, and the response HTML content filtered for the fatwa text fields like category, title, question, and answer.

Another source of text data is the website by Ibn Othaimeen. This website has a rich collection of fatwas in different programs, but the challenge was that it does not store static data. Instead, the site uses JavaScript to load its content dynamically. Hence, HTML-based filtering and the Beautiful Soup library could not be implemented for this site.

We used the Python Selenium library for this website to simulate browsing using the Chrome driver. This allowed us to fetch data by simulating browsing, scrolling, clicking, and recording the response. We used Selenium to control the Chrome driver in headless mode so that the GUI is not displayed, but the operations are performed seamlessly in the background.

The fatwas on this site are organized mainly into three different programs, namely: فتاوى نور على الدرب, لقاءات الباب المفتوح and اللقاء الشهري. All three have different URLs and different numbers of pages under the main URL, which contain links to the individual episodes. These episodes, in turn, had links to the fatwa. The episode numbers for فتاوى نور على الدرب are named as الشريط, while for the other two programs they are named the same as the program with numbering. So, the hierarchy of data was: fatwa > radio program > episode number > fatwa.

First, we iterated a loop over the URLs for the three programs. For each program, we further iterated over all the pages to append to the base URL for the program. Then, we used the Selenium Python library to make the Chrome driver go to the program URL with the page number and get the response to fetch all the web elements on the page.

As the page had multiple web elements, we were interested in only those pointing to the subsequent episode links containing the fatwa, e.g., those named with رقم الشريط for فتاوى نور على الدرب and so on. Hence, we filtered web elements with the corresponding text in their XML path. Then, we iterated over all those filtered elements on the page, got the text attribute from each element, and stored them in a list.

After that, we iterated over the list of texts and fetched the corresponding web elements to execute the JavaScript code to scroll to that web element and then execute the click for the element. After clicking the element, we waited until the next page was loaded. Once the episode page loaded and the web driver was pointing to the episode page, we got the corresponding URL of the episode page and stored it in the list. Similarly, we went back and got the URL for all the web elements on the episode page.

Eventually, we had the URL for all the episodes on a particular page within a certain radio program. By iterating over all the pages and all the programs, we got the URL for all episodes. Then, we iterated over the obtained episode URLs and set the Selenium-controlled driver to go to each URL iteratively.

Once all fatwa links were loaded and clickable on the episode page, we filtered the fatwa-referencing web elements by the condition of the XML path style containing a cursor pointer. We waited for all these elements to be fully loaded and clickable, then got texts for each.

Finally, the elements referencing each fatwa were fetched by looking for the texts obtained from their icons in the previous step. The corresponding web elements were clicked one by one. On clicking the fatwa element, the Chrome driver finally landed on the page containing the fatwa content, which included the title of the fatwa, the question, the answer, and its hierarchy path on the website (see Fig. 3).

When the fatwa web page loaded completely, the three different elements were fetched one by one, i.e., for the title, question, and answer, based on the CSS selectors 'p.title', 'div.fatwah-ques-cont', and 'div.fatwah-ans-cont' respectively. For each of these web elements, the inner HTML attribute was fetched, which loaded the HTML content. Finally, the Beautiful Soup library was used to fetch the required text from the fetched HTML for the element. This text data was stored in a CSV file under the following fields for each fatwa: title, question, answer, current-url, and category.

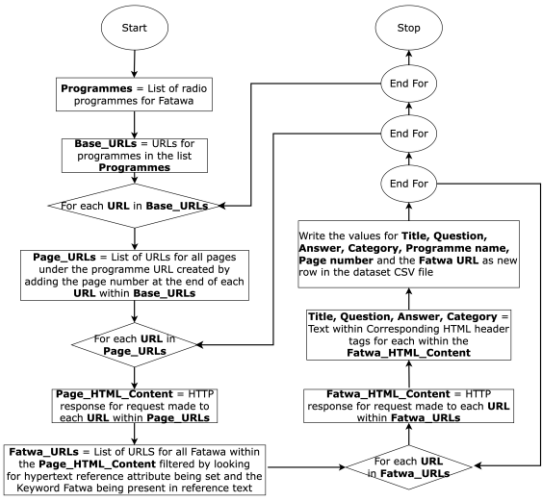


Fig. 3. The flowchart depicting the logic for collecting fatwa text from static websites.

B. Youtube Transcripts

To get the transcripts from YouTube videos, official YouTube channels approved by the respective organizations were targeted, and a list of playlists to be targeted was identified for each channel. For each playlist, a list of URLs for all videos within that playlist was fetched using the playlist method from the Pytube Python library. Subsequently, each video URL within the list was looped through to get the transcript for the video using the get transcript method of the YouTube transcript API in Python. In case a video didn't have a transcript on YouTube, it returned a "transcript not found" error, in which case the URL of the video was added to a list of missed videos for that channel to be targeted for speech recognition.

Transcripts fetched for videos had text transcripts and timestamps, which were retained. Another clean copy of the transcript was also created, removing the timestamps and retaining the complete script as a single entity. Both the timestamped and the clean transcripts, along with the video title and URL, were saved in a separate SQL database file for that channel (see Fig. 4 and Fig. 5).

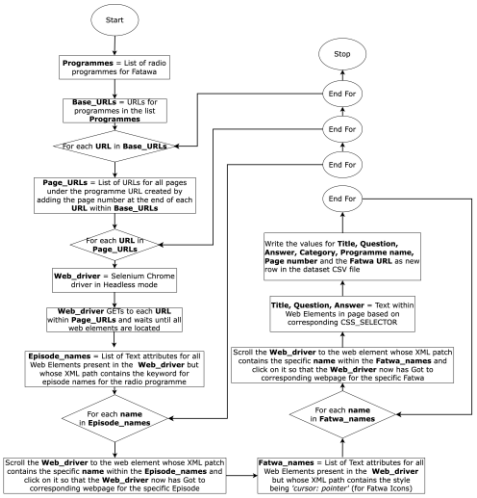


Fig. 4. The flowchart for collecting content from a website with dynamic data.

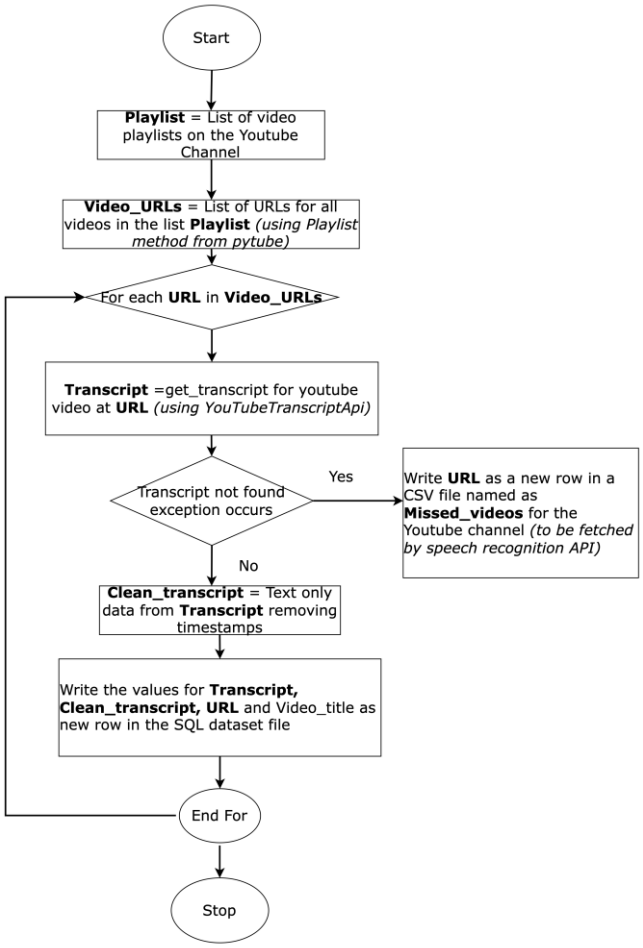


Fig. 5. The flowchart for collecting transcripts from YouTube videos.

C. Automatic Speech Recognition

The file of missed videos created while getting the transcripts for each channel was then used to get the scripts of these videos with Automatic Speech Recognition (ASR) using the Google Speech Recognition API in Python. First, for each URL in the list of missed video URLs, the corresponding video was downloaded using the streams method of the YouTube class in the pytube library. The video was downloaded in MP4 format, and its audio was extracted and saved as a WAV audio file using the audio method of the VideoFileClip class from the moviepy.editor Python library. The converted WAV audio file was then used for speech recognition. However, longer files often have sections that are not recognized by the speech recognition engine, causing errors. To address this issue, the audio file was segmented into smaller chunks of approximately 180 seconds each and passed to the Arabic speech recognition engine of the Google Speech Recognition API. The text for each chunk was recognized and stored in a text file. Once all the segments for a video were processed, the text chunks were concatenated and saved as a single transcript for the video, along with the video URL and title, in an SQL DB file. This process generated the ASR scripts for the channel (see Fig. 6).

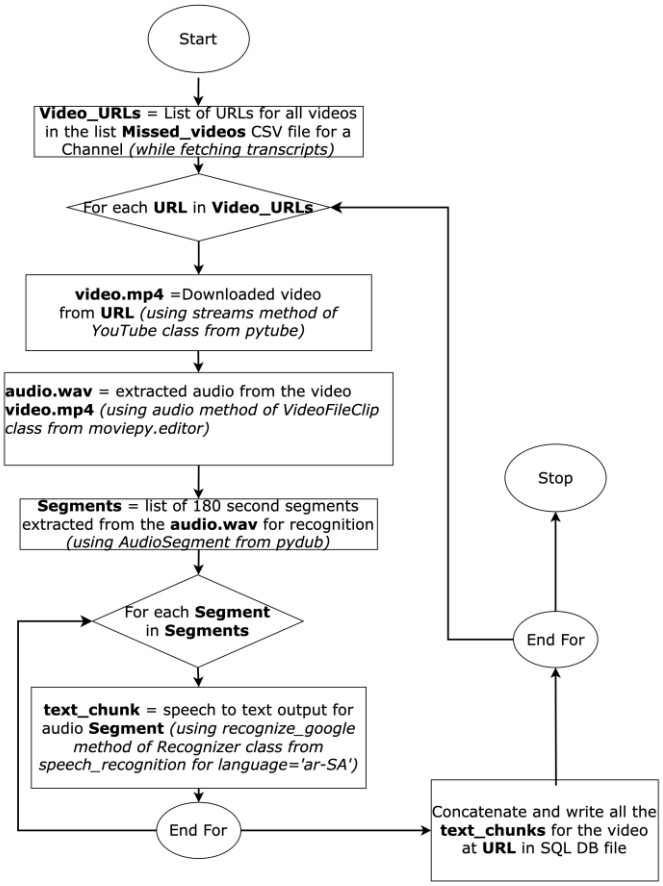


Fig. 6. The flowchart for collecting ASR text from YouTube videos.

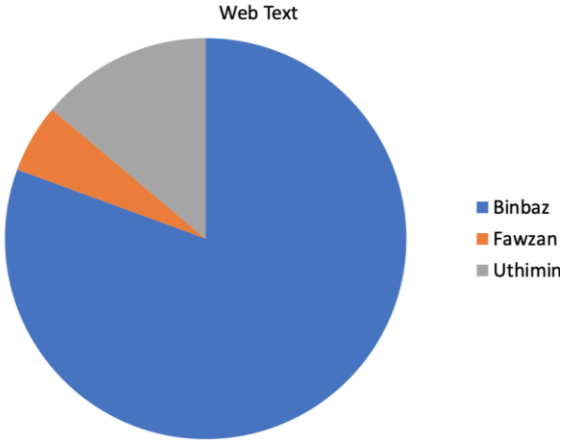
V. Dataset Statistics

The resulting Islamic content dataset is organized into three main folders, namely Youtube_transcripts, Youtube_asr, and Text_data (see Table IV). The Youtube Transcripts folder contains seven files with 1363 records. The Youtube ASR folder contains eight files with 2972 records. Text data folder contains four files with 31225 records.

TABLE IV. Islamic Data Files and their Characteristics in our Dataset [A Total of 35560 Records]

Data source	File Name	Records	Storage format	Owner
Youtube Transcripts	Makkah_transcripts.db	242	SQL DB	General Authority for the affairs of the Grand Mosque and the Prophet's Mosque
Youtube Transcripts	SaudiQuranTv_transcripts.db	75	SQL DB	Saudi Broadcasting Authority
Youtube Transcripts	wmngovsa_transcripts.db	9	SQL DB	Agency of General Presidency for the Affairs of Al-Masjid Al-Nabawi
Youtube Transcripts	SaudiSunnahTv_transcripts.db	281	SQL DB	Radio and Television Corporation in the Kingdom of Saudi Arabia
Youtube Transcripts	wmngovksa_transcripts.db	263	SQL DB	Agency of General Presidency for the Affairs of Al-Masjid Al-Nabawi
Youtube Transcripts	twjehDM_transcripts.db	334	SQL DB	General Authority for the affairs of the Grand Mosque and the Prophet's Mosque
Youtube Transcripts	ibnothaimeentv_transcripts.db	159	SQL DB	Sheikh Muhammad bin Saleh Foundation
youtube_ASR	Makkah_asr.db	421	SQL DB	General Authority for the affairs of the Grand Mosque and the Prophet's Mosque
youtube_ASR	SaudiQuranTv_asr.db	9	SQL DB	Saudi Broadcasting Authority
youtube_ASR	SaudiSunnahTv_asr.db	126	SQL DB	Radio and Television Corporation in the Kingdom of Saudi Arabia
youtube_ASR	salihalfawzan_asr.db	1016	SQL DB	Al Dawa Charitable Foundation
youtube_ASR	aforgsa1_asr.db	92	SQL DB	Al Dawa Charitable Foundation
youtube_ASR	twjehDM_asr.db	218	SQL DB	General Authority for the affairs of the Grand Mosque and the Prophet's Mosque
youtube_ASR	ibnothaimeentv_asr.db	614	SQL DB	Sheikh Muhammad bin Saleh Foundation
youtube_ASR	wmngovksa_asr.db	476	SQL DB	Agency of General Presidency for the Affairs of Al-Masjid Al-Nabawi
Web text	binbaz_misc_final_v1.csv	718	CSV	Sheikh Abdul Aziz bin Baz Charitable Foundation
Web text	binbaz_fatwa_final_v1.csv	24448	CSV	Sheikh Abdul Aziz bin Baz Charitable Foundation
Web text	othaimeen_fatwa_v1.csv	4334	CSV	Sheikh Muhammad bin Saleh Foundation
Web text	alfawzan_fatwa_v1.csv	1725	CSV	Al Dawa Charitable Foundation

Fig. 7 shows the proportion of records in the final dataset from different sources.



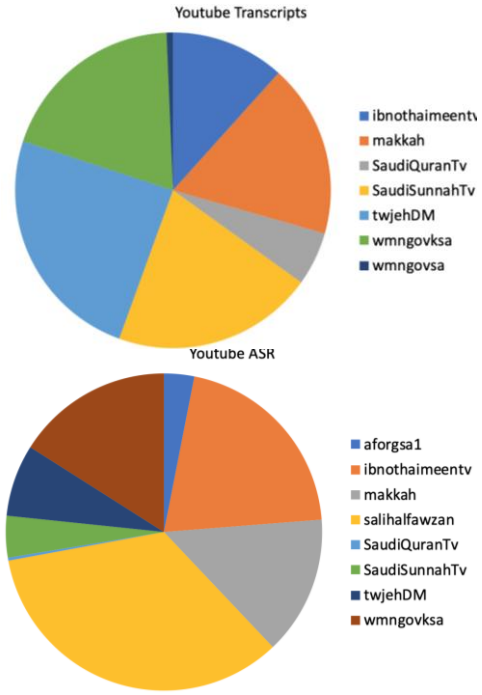


Fig. 7. Total records in the final dataset from different sources.

The focus of the example implementation has been on compiling authentic Islamic datasets, showcasing the capability of the algorithm to collect specialized datasets. However, the approach is generalizable and can be used in multilingual and multi-domain tasks, offering a versatile tool for dataset collection in various fields.

The code has been organized and publicized on a GitHub repository [9]. The repository contains different modules for scraping static and dynamic websites, as well as for fetching transcripts and performing speech-to-text conversions from YouTube videos. To use the code, users need to run the scraping.py file, providing the corresponding options for input and output files, and base URLs for the sites to fetch data from. Additionally, an action option must be specified to indicate whether to use dynamic, static, transcript, or ASR functionality. This will call the appropriate function within the scraping.py file. Fig. 8 is a screenshot from Github repo indicating the structure of code files.

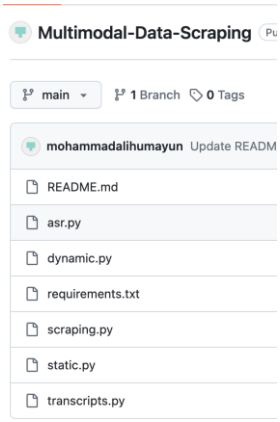


Fig. 8. Github repo structure.

In contrast to existing similar datasets reviewed in earlier sections, our dataset comprises diverse sources, including Articles, Lessons, Lectures, Meetings and Dialogues, Rulings, Speech, Letters, Communications, and Sermons. Additionally, all dataset sources are authentic, sourced from reputable organizations such as the General Authority for the Affairs of the Grand Mosque and the Prophet's Mosque, the Agency of the Affairs of Al-Masjid Al-Nabawi, Sheikh Abdul Aziz bin Baz Charitable Foundation, Sheikh Mohammed bin Saleh Al Othaimeen Charitable Foundation, and the Al Dawa Charitable Foundation. The quality and validity of tools used for speech transcription and data scraping have been thoroughly validated through manual verification of data records. Finally, the automated data scraping method has been confirmed to be reproducible, with clear descriptions of requirements and detailed steps provided for replication of the automated data collection process.

VI. Conclusions and Limitations

We have developed four distinct algorithms to address different types of text extraction tasks. The first algorithm focuses on extracting text data from static websites. The second algorithm is designed to retrieve text data from dynamic websites. The third algorithm targets the extraction of relevant transcripts from YouTube videos that already have transcripts available. Finally, the fourth algorithm aims to translate speech to text from videos that do not have transcripts. These algorithms enable web scraping [10] across diverse websites containing Islamic knowledge.

Using the proposed web scraping tool, this study has successfully created a comprehensive and reliable Islamic knowledge dataset from authentic Saudi Arabia sources. By aggregating 31,225 records from reputable Islamic scholars, official websites, and authorized YouTube channels, we have established a robust foundation for text processing regarding Islamic knowledge. This dataset addresses the critical need for accurate and contextually relevant Islamic content, aiming for precise and trustworthy responses by automated AI models. The dataset motivates further language processing research for more accurate and useful applications in Islamic knowledge processing.

Admittedly, it was not possible to identify all available Islamic knowledge sources in Saudi Arabia. Therefore, we acknowledge that our dataset might have overlooked other important sources, in which data were not downloadable or unverified. Our Islamic knowledge dataset is collected from sources endorsed by Saudi organizations, thus mainly reflecting the Hanbali school of thought and views. In Islam, there are four main Sunni schools of interpretation endorsed in approximately 30 Muslim countries. The range of topics covered in the proposed dataset is greatly influenced by the content published on the trusted websites that we scraped. Technically, a few limitations caused some sources to be missed. For example, while a time wait was imposed while fetching data from the web elements from dynamic websites, certain elements took a long time to become clickable (i.e., downloadable) and caused a timeout. Moreover, YouTube videos without transcripts were downloaded, converted into audio files, and translated into text. This process was also not foolproof, and some of the videos

failed to download due to network issues or sometimes because an age restriction was set on the target videos, which required a sign-in to download. Finally, our dataset does not include knowledge published in booklets and books, which could hold important insights about Islamic rulings.

ACKNOWLEDGMENT

This research is supported by the Deanship of Scientific Research of the Islamic University of Madinah, KSA under the research groups (first) project no. 956.

REFERENCES

[1] A. A. Munshi, W. H. AlSabban, A. T. Farag, O. E. Rakha, A. A. Al Sallab, and M. Alotaibi, 'Towards an automated Islamic fatwa system: Survey, dataset and benchmarks', Int. J. Comput. Sci. Mob. Comput., vol. 10, no. 4, pp. 118–131, 2021.

[2] S. Alnefaie, E. Atwell, and M. A. Alsalka, 'Challenges in the Islamic Question Answering Corpora', Int. J. Islam. Appl. Comput. Sci. Technol., vol. 10, no. 4, pp. 1–10, 2022.

[3] M. Mohammed, S. Amin, and M. M. Aref, 'An english islamic articles dataset (EIAD) for developing an Islambot question answering chatbot', in 2022 5th International Conference on Computing and Informatics (ICCI), IEEE, 2022, pp. 303–309.

[4] M. R. Rizqullah, A. Purwarianti, and A. F. Aji, 'Qasina: Religious domain question answering using sirah nabawiyah', in 2023 10th International Conference on Advanced Informatics: Concept, Theory and Application (ICAICTA), IEEE, Oct. 2023, pp. 1–6.

[5] O. Alyemny, H. Al-Khalifa, and A. Mirza, 'A Data-Driven Exploration of a New Islamic Fatwas Dataset for Arabic NLP Tasks', Data, vol. 8, no. 10, p. 155, 2023.

[6] A. M. Farooqi, R. A. S. Malick, M. S. Shaikh, and A. Akhunzada, 'Multi-IsnadSet MIS for Sahih Muslim Hadith with chain of narrators, based on multiple ISNAD', Data Brief, vol. 54, p. 110439, 2024.

[7] M. Mghari, O. Bouras, and A. El Hibaoui, 'Sanadset 650k: Data on hadith narrators', Data Brief, vol. 44, 2022.

[8] A. Namoun, M. A. Humayun, and W. Nawaz, 'Authentic Islamic knowledge dataset'. 2024. doi: 10.17632/zjrc34pc3p.1.

[9] 'Multimodal-Data-Scraping repository'. 2024. https://github.com/anamoun/Multimodal-Data-Scraping.

[10] A. Namoun, A. Alshanqiti, E. Chamudi, and A. Rahmon, 'Web design scraping: Enabling factors, opportunities and research directions', in 2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE), IEEE, Oct. 2020, pp. 104-109.

Hybrid Transfer Learning for Diagnosing Teeth Using Panoramic X-rays

M. M. EL-GAYAR
Department of Information Technology-Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt
Department of Computer Science, Arab East Colleges, Riyadh 11583, Saudi Arabia

***Abstract*—The increasing focus on oral diseases has highlighted the need for automated diagnostic processes. Dental panoramic X-rays, commonly used in diagnosis, benefit from advancements in deep learning for efficient disease detection. The DENTEX Challenge 2023 aimed to enhance the automatic detection of abnormal teeth and their enumeration from these X-rays. We propose a unified technique that combines direct classification with a hybrid approach, integrating deep learning and traditional classifiers. Our method integrates segmentation and detection models to identify abnormal teeth accurately. Among various models, the Vision Transformer (ViT) achieved the highest accuracy of 97% using both approaches. The hybrid framework, combining modified U-Net with a Support Vector Machine, reached 99% accuracy with fewer parameters, demonstrating its suitability for clinical applications where efficiency is crucial. These results underscore the potential of AI in improving dental diagnostics.**

Keywords*—Machine learning; deep learning; dental diagnosis; transfer learning

I. Introduction

Accurate diagnosis of oral diseases is imperative for maintaining dental health. Panoramic x-rays provide comprehensive views of the teeth and jaws, making them invaluable for treatment planning. However, manually interpreting these complex images is resource-intensive, fallible to errors, and requires radiological expertise that general dentists may lack. Recent advances in artificial intelligence (AI) offer new opportunities to automate dental image analysis, overcoming the challenges of manual interpretation. But progress is impeded by factors like scarce annotated data and anatomical variability. Despite obstacles, integrating AI into dental radiology could significantly enhance patient care [1-4].

Image segmentation is critical for medical image analysis. Deep learning has surpassed hand-engineered features, especially convolutional neural networks (CNNs) [9][14][20]. The pioneering U-Net architecture combined encoders and decoders for precise segmentation. Extensions like DeepLab improved resolution, while Mask R-CNN enabled multi-task learning. Recent methods apply transformers and distillation. Medical imaging has benefited from these innovations. Segmentation aids organ delineation and dental analysis. Overall, CNNs now dominate segmentation by learning robust representations directly from pixels. We aim to advance panoramic radiograph segmentation by integrating spatial context into U-Net.

To promote the development of accurate AI-driven tools, we have test our method on the Dental Enumeration and Diagnosis on Panoramic X-rays (DENTEX) challenge. This challenge aims to stimulate and validate algorithms that can reliably detect and count abnormal teeth on panoramic x-rays. Automated frameworks could empower precise diagnostics and treatment planning while minimizing errors [5] [6].

However, developing accurate AI systems for dental radiograph analysis poses several challenges [7]:

- Limited availability of annotated panoramic x-ray datasets impedes model training and validation. We addressed this by utilizing extensive data augmentation and transfer learning.
- Panoramic images exhibit distortions like irregularities and overlaps that can confuse algorithms. Our model implements robust pre-processing to minimize such artifacts.
- Identifying some dental conditions requires assessing tooth relationships rather than isolated teeth. Capturing inter-dental context remained an open challenge.
- Generalizing model to handle variability in image quality and demographics requires expansive, diverse datasets that remain scarce. We aim to expand testing across diverse sources.
- Reducing computational costs without sacrificing accuracy remained an ongoing pursuit. Our optimizations enhanced efficiency, but further improvements may be possible.

The main contributions of this manuscript are:

- We have devised an innovative diagnostic system tailored for assessing dental conditions from panoramic x-rays. Our framework implements both direct classification through deep learning models, and a hybrid approach integrating deep feature extraction with traditional machine learning. This dual methodology aims to leverage the complementary strengths of modern AI to improve accuracy and efficiency.
- Our technique combines segmentation and detection models to pinpoint dental abnormalities efficiently.
- We performed comprehensive analyses comparing multiple deep learning architectures and classical models under direct and hybrid diagnostic settings. This rigorous

testing has provided valuable insights into real-world performance and transferability.

- Applying dimensionality reduction techniques, we have enhanced the computational efficiency of our framework while retaining precision. This allows our system to remain simultaneously powerful and nimble.

This manuscript is organized into five sections. Section II reviews relevant previous work. Section III describes the proposed approach in detail. Experiments and results are presented in Section IV. Section V concludes the paper.

II. Related Works

Hybrid approaches combining feature extraction with machine learning classifiers have proven effective across various medical and non-medical image analysis tasks. Recently, convolutional neural networks (CNNs) have become prevalent for feature extraction, along with some continued use of hand-crafted features. We review these techniques for general applications and those specific to dental diagnostics.

In non-medical settings, hybrid frameworks have shown advantages for video violence classification, human action recognition, and image texture classification, among others. In the medical domain, similar approaches have been applied for tasks including gastrointestinal disease classification from endoscopy, mammogram-based breast cancer screening, retinal disease diagnosis, burn image analysis, and medical image modality classification.

For dental diagnostics, CNN-based techniques have dominated recent literature. Most works focus on classifying a limited set of dental diseases, achieving accuracy over 99% in some cases. This high performance results from factors like robust datasets, simplistic tasks, and model advantages. However, studies tackling more challenging dental issues, like cavity detection, or hampered by poor data or task complexity, have seen lower accuracy.

In summary, hybrid approaches combining deep learning-based feature extraction with traditional machine learning have proven versatile for both medical and non-medical image analysis across various applications. In the emerging domain of AI-driven dental diagnostics, CNNs currently predominate, but task complexity remains a barrier to maximizing performance. Further innovations in hybrid techniques show promise for advancing the field.

Ayhan et al. [8] introduced a deep learning approach for tooth numbering, caries detection, and matching from bitewing radiographs. Their method utilized a DenseNet-121 model pretrained on natural images for tooth detection and numbering. YOLOv7 was applied for caries detection. Tooth numbers were then matched to detected caries using intersection over union. The models were trained and evaluated on 1170 bitewing images from faculty archives. They achieved high performance, with F-scores of 0.99 for tooth detection, 0.979 for numbering, 0.822 for caries classification, and 0.842 for number-caries matching. This demonstrates the capability of deep convolutional neural networks like DenseNet and YOLO for automated dental radiograph analysis. However, use of a private institutional dataset makes results difficult to reproduce. Testing on more varied multi-source data could better validate generalization. Additionally, bitewing images may be less challenging than panoramic x-rays. But overall, their work provides evidence for deep learning-based dental image analysis, and proposes an integrated numbering-diagnosis framework applicable to clinical practice.

Li and Zhang [10] developed a convolutional neural network-vision transformer model for multi-label classification of dental conditions from orthopantomography (OPG) x-rays. Their hybrid architecture combined CNN feature extraction with a transformer classifier. The model was trained and evaluated on a dataset of 1418 OPG radiographs from clinical cases containing multiple disease labels. For multi-label classification across eight dental diseases, they achieved strong performance with a sensitivity of 0.942, specificity of 0.951, accuracy of 0.968, and F-score of 0.957. This demonstrates the potential of using hybrid CNN-transformer architectures for automated analysis of dental radiographs. However, use of a private clinical dataset makes reproducing their results difficult. Additionally, multi-label classification across many diseases poses challenges compared to binary classification. But overall, their work helps highlight advanced deep learning architectures like vision transformers for robust dental image analysis and multi-disease diagnosis from OPG scans.

Zhu et al. [6] developed an AI system to diagnose 5 different dental diseases from panoramic radiographs. Their approach used a combination of BDU-Net to detect dental caries, and nnU-Net models to identify the other 4 conditions - periodontitis, periapical lesions, dental pulp stones, and impacted teeth. The models were trained and tested on a private dataset of 2278 OPG images. For caries diagnosis, BDU-Net achieved a specificity of 99.4%, while nnU-Net models produced specificities greater than 99% for the other diseases. This demonstrates the potential of using specialized deep-learning architectures like BDU-Net and nnU-Net for multi-disease classification from dental radiographs. However, their reliance on a private dataset makes reproducing and validating their results difficult. Additionally, combining outputs from multiple models increases system complexity compared to a single unified classifier. Overall, their work provides initial evidence that hybrid ensembles of deep-learning models can automate the identification of different dental conditions from OPG scans.

Almalki et al. [2] developed deep learning models to classify four common dental diseases using orthopantomography (OPG) x-ray images. Their method utilized the YOLOv3 object detection architecture to analyze a dataset of 800 private OPG radiographs. The task involved detecting dental caries, periodontitis, periapical lesions, and dental fractures in teeth depicted in the OPG scans. By leveraging the YOLOv3 model pretrained on natural images and fine-tuning on the dental data, they achieved a high accuracy of 99.33% for multi-class disease classification. This work demonstrates the potential of deep learning techniques like YOLOv3 for automated analysis of dental radiographs. However, the use of a private dataset makes it difficult to reproduce their results. Additionally, their set of four classes represents only a subset of important dental diseases, so generalization to more complex multi-label classification remains unclear. But overall, their study provides

evidence for deep learning and YOLO-based approaches in advancing automated assessment of dental conditions from radiographs.

Zhang et al. [12] developed a deep learning model to screen for dental caries from digital oral photographs. Their method adapted a Single Shot MultiBox Detector CNN architecture and incorporated hard negative mining during training. The model was trained and evaluated on a dataset of 33932 photographs captured from 625 volunteers using consumer cameras. For binary classification of images as carious or non-carious, they achieved an AUC of 85.65%. This demonstrates the feasibility of using deep CNNs to analyze oral photographs for automated dental caries screening. However, photographs only provide limited visibility compared to radiographs. Additionally, use of consumer cameras introduces variability compared to clinical imaging. But overall, their work helps establish deep learning as a viable approach to automate identification of dental caries from oral photographs.

Sonavane et al. [11] developed a convolutional neural network model for classifying dental cavity images. Their approach utilized a custom CNN architecture designed for cavity detection. The model was trained and tested on a publicly available dataset of 55 images containing cavities and 19 non-carious images. Using this small dataset, they achieved a maximum accuracy of 71.43% for binary cavity classification. This preliminary study demonstrates the potential of using CNNs for dental cavity detection from visual images. However, the very small public dataset limits model performance and does not represent real-world variability. Additionally, visual images have limited visibility compared to radiographs. But overall, their work provides a proof-of-concept for using CNNs to classify dental cavities from images.

Lee et al. [13] developed a deep learning system to detect and diagnose dental caries from periapical radiographs. They utilized a pretrained GoogLeNet Inception v3 CNN architecture and fine-tuned it on a private dataset of 3000 periapical images. The model was trained to classify images as either carious or non-carious based on caries present in premolars and molars. They achieved AUCs of 0.917, 0.890, and 0.845 for premolar, molar, and combined classes respectively. This demonstrates the capability of deep CNNs like Inception-v3 for automated dental caries diagnosis from radiographs. However, use of a private dataset makes reproducing their results difficult. Additionally, periapical x-rays only cover limited tooth surfaces compared to full-mouth radiographs. But overall, their work provides evidence for deep learning-based classification of dental caries using CNN architectures like GoogLeNet Inception pretrained on natural images.

III. Proposed Framework

Our model evaluates two main deep-learning approaches for dental disease classification from panoramic X-rays and radiograph images:

- Direct classification: Images are fed into a fine-tuned deep-learning model that predicts abnormal teeth labels directly using its classification layer.
- Hybrid approach: A pre-trained model extracts image features, which are then classified using a traditional machine-learning algorithm.

For the hybrid approach, teeth are first extracted from radiographs and preprocessed. The cropped tooth images are input to a fine-tuned deep CNN, which extracts discriminative features. These features then train a classifier like an SVM for the final diagnosis. To handle class imbalance, rotating minority class images perform data augmentation. This results in a more balanced distribution for model training. Ten deep learning models are experimented with, including CNNs like ResNet, VGGNet, MobileNet, and vision transformers like ViT. All leverage transfer learning from natural image datasets. For feature extraction, models are trimmed before their classification layers. The extracted features are combined with classical ML classifiers like SVM and random forest.

Unique advantages of the hybrid approach include leveraging complementary strengths of deep CNN feature learning and traditional classification methods. This can potentially improve accuracy and efficiency. The proposed model aims to advance dental panoramic X-rays and radiograph analysis by combining state-of-the-art deep learning and classical machine learning techniques in an optimized pipeline.

A. Proposed Model Architecture

Our proposed model utilizes a hybrid approach combining deep convolutional neural networks (CNNs) for feature extraction with traditional machine learning models for classification as shown in Fig. 1. First, tooth segments are extracted from panoramic radiographs and preprocessed. The cropped tooth images are normalized and resized to standard dimensions. These preprocessed segments are input to a fine-tuned deep CNN which extracts discriminative feature representations for each image. The CNN leverages transfer learning from models pretrained on large-scale natural image datasets. To train the classifier, these learned feature vectors are used to train traditional machine learning algorithms like support vector machines and random forests. This hybrid approach aims to leverage the complementary strengths of deep CNN feature learning and classical models for enhanced accuracy and efficiency. Alogorithm 1 shows the steps and pusedocode of the dental diagnosis model with input panoramic x-ray or radiographic images and the ouput is trained model and test reults.

B. Preprocessing Pipeline

The dataset undergoes several preprocessing steps before model training:

- Filtering removes invalid images lacking tooth segments or having duplicate values, ensuring only pertinent images are used.
- Individual tooth segments are extracted by cropping radiograph sections based on provided coordinates.
- Segments are resized to 256x256 pixels and centered cropped to 224x224 pixels to match CNN input dimensions.

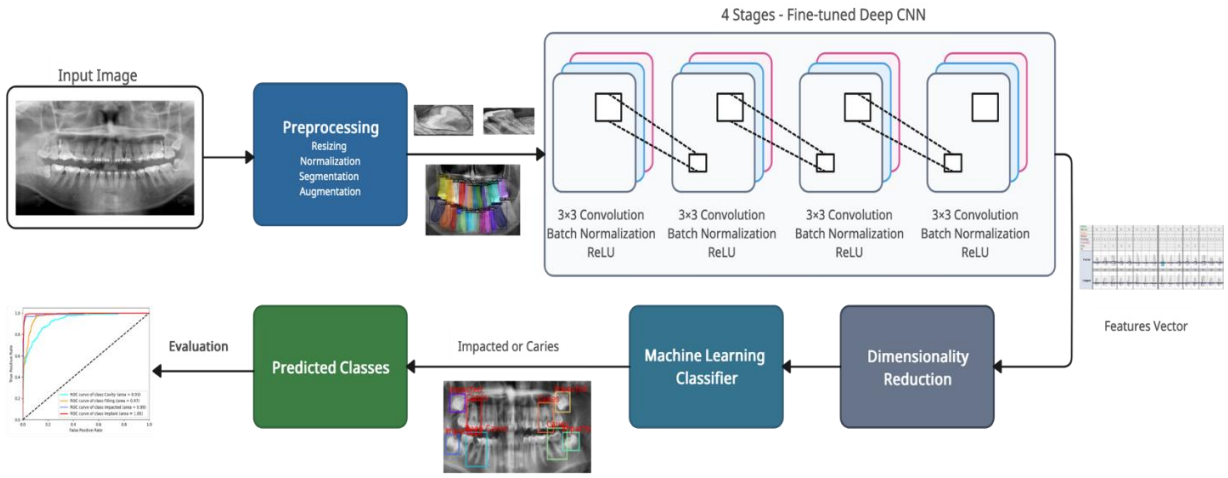


Fig. 1. Proposed model.

- Pixel value normalization is applied using channel-wise mean and standard deviation values from the CNN's original training distribution.

This standardized preprocessing pipeline obtains cleaned, extracted, and normalized tooth images suitable for input to deep CNNs. The transformations aim to highlight key dental characteristics while suppressing noise and distortions. As shown in Eq. (1), image pixel values xij in channel i are normalized to zij using per-channel mean μi and standard deviation σi statistics:

$$N_{xy} = \frac{O_{xy} - \mu_x}{\sigma_x} \quad (1)$$

Where:

N_{xy} = normalized pixel value in channel x

O_{xy}= original pixel value in channel x

μ_x = mean of pixel values in channel x

σ_x= standard deviation of pixel values in channel x

This performs normalization independently for each color channel by subtracting the channel mean and dividing by the channel standard deviation. The result N_{xy} are normalized pixel values with a zero mean and unit variance based on the image's original channel statistics. This standardized preprocessing brings values into a consistent range to better highlight key image features.

Algorithm 1: Dental diagnosis model

Input: Dental panoramic x-ray or radiography images (imgs)
Output: Model (m) and Evaluation metrics (eval)
Start Procedure
  preprocessImgs(imgs)
  augmentedImgs = oversampleMinorityClasses(imgs)
  m = loadPretrainedCNN(pretrained_model)
  features = extractFeatures(m, augmentedImgs)
  reducedFeatures = PCA(features)
  m = trainClassifier(reducedFeatures)
  eval = evaluateModel(MLmodel, testFeatures)
  return m and eval
End Procedure

C. Handling Class Imbalance

The four dental disease categories were initially imbalanced in the dataset, with the cavity and impacted classes significantly under-represented compared to the other groups. To mitigate this class imbalance during model training, we employed data augmentation techniques focused on the minority classes.

Augmentation was performed by applying rotations of 45 degrees in both directions to the images of the under-represented cavity and impacted categories. This geometrical transformation tripled the number of samples for these classes. After augmentation, the total dataset contained 11,087 images with a more balanced distribution across the four dental conditions.

This selective oversampling addresses the class imbalance problem by increasing minority class samples. Augmenting under-represented categories through rotations provides additional variety in viewing angles while preserving key dental morphologies. The resulting balanced dataset reduces bias and enables robust learning of all disease classes for improved multi-class classification performance.

D. Transfer Learning Model

Our model utilizes transfer learning by initializing with weights from models pretrained on large-scale natural image datasets like ImageNet, U-Net and AlexNet [22]. Ten state-of-the-art deep CNN architectures were investigated, including DenseNet [21], captionNet [19], ResNet [17], VGGNet [16], MobileNet [18], vision transformers [15], and YOLOv9. Transfer learning enables extracting more discriminative features despite our relatively small dental panoramic or radiograph dataset.

For feature extraction, models are trimmed before their classification layers to obtain vector representations of the input images. The pretrained weights provide robust initial feature learning which is then fine-tuned on the dental images. Widely used models like AlexNet serve as strong feature extractors given their proven imaging performance. Their convolutional layers learn hierarchical filters to capture informative spatial patterns. Additionally, we explore a U-Net architecture with symmetrical encoder-decoder structure for end-to-end segmentation and classification. The encoder extracts contextual

features while the decoder recovers localization and spatial details. Skip connections combine these complementary learned representations. Compared to other CNNs, U-Net can better localize abnormal dental regions in the panoramic and radiograph images. Standard U-Net models lack localization capability which can limit performance on abnormality detection in dental radiographs. To overcome this, we incorporate recent advancements that provide spatial context to U-Net. Specifically, we augment U-Net with BB-Conv layers comprised of max pooling followed by convolutions. Bounding box coordinates for each tooth are fed into these layers to output attention maps highlighting tooth locations. The BB-Conv layers are inserted into each skip connection. Their outputs are multiplied with encoder features before concatenation during upsampling. This injects positional information across all network stages. Compared to vanilla U-Net, this Modified U-Net integrates localization cues via the BB-Conv spatial attention layers. By guiding the model to focus on specific tooth regions, detection of localized pathologies is improved. We hypothesize this will enhance abnormality modeling and increase sensitivity to anomalies like dental caries. Our experiments compare Modified U-Net against standard U-Net and other CNNs to quantify the impact of incorporating spatial context.

E. Dimensionality Reduction

The high-dimensional feature representations extracted from the deep CNNs can contain redundant and noisy components. To reduce complexity and combat overfitting, we apply dimensionality reduction to the learned features before feeding them to traditional machine learning classifiers.

Specifically, Principal Component Analysis (PCA) is utilized to project the features into a lower-dimensional subspace. PCA transforms the data such that the maximum variance is captured along the first principal components. This converts the features into a compact set of dimensions that encapsulate the most salient information.

Applying PCA after deep feature extraction distills the representations down to their core components most relevant for the classification task. By suppressing extraneous dimensions, overfitting is reduced and model generalization is enhanced. The resulting lower-dimensional features serve as efficient input to the ML models for enhanced performance. As shown in Equation 2, PCA can be implemented by singular value decomposition (SVD) of the data matrix D ∈ R^{m x n}:

$$D = E \sum V^T \quad (2)$$

Where:

- E is a m x m orthogonal matrix containing the eigenvectors of DD^T
- $\sum V^T$ is a m x n diagonal matrix containing the singular values σ1, ..., σr
- V^T is a n x n orthogonal matrix containing the eigenvectors of $D^T D$

The columns of E are the principal components corresponding to the directions of maximum variance in the data. Taking the first k columns of U projects the data into the k-dimensional subspace capturing the greatest variance. The singular values σ1, ..., σr are the square roots of the eigenvalues of $D^T D$ and indicate the significance of each principal component - larger values correspond to more informative components. So PCA via SVD provides a way to find a lower-dimensional representation of the data that preserves maximal information content as quantified by the singular values.

F. Traditional Machine Learning Classifiers

To perform final classification using the extracted features, we evaluate diverse classical machine learning models to determine an optimal approach. Seven different classifiers are investigated:

- Naive Bayes (NV) applies Bayes' theorem with conditional independence assumptions between features.
- K-nearest neighbors (KNN) categorize samples based on proximity to nearest examples in the feature space.
- Logistic regression (LR) produces probabilistic multi-class predictions using a softmax function.
- Decision trees (DT) recursively partition the feature space by splitting on the most informative attributes.
- Support vector machines (SVM) find maximum margin decision boundaries between classes. Kernel tricks enable efficient mapping to higher dimensional spaces.

Each model has unique advantages that are assessed during experimentation. SVM and logistic regression leverage robust regularization to avoid overfitting. KNN and Naive Bayes offer simplicity and efficiency.

G. Summary of this Section

In this part, we summarize the steps of the proposed system and link them to the proposed algorithms.

1) Preprocess radiograph images

- Crop teeth segments
- Resize to standard dimensions
- Normalize pixel values

2) Perform data augmentation

- Use minority oversampling to handle class imbalance

3) Extract features using fine-tuned deep CNN

- Use transfer learning from pre-trained models like VGG, ResNet, U-Net, Alex-Net
- Remove classification layer
- Input images to obtain descriptive feature vectors

4) Apply PCA for dimensionality reduction

5) Evaluate models like SVM, Random Forest, KNN

- Tune hyperparameters for optimal performance
- Evaluate model on test set
- Compute metrics like accuracy, precision, recall

IV. Experimental Results

A. Dataset Overview

The DENTEX dataset [23] contains panoramic dental X-ray images collected from three different clinical institutions. This introduces diverse quality levels reflecting real-world heterogeneity. Patients were randomly selected to ensure privacy. The dataset has a hierarchical organization with gradually increasing annotation levels:

- 693 images with quadrant labels only.
- 634 images with quadrant and tooth enumeration.
- 1005 images fully annotated with quadrants, teeth, and diagnoses.

The diagnostic labels encompass four conditions: caries, deep caries, periapical lesions, and impacted teeth. An additional 1571 unlabeled images are provided for pretraining. For formal evaluation, the fully annotated set of 1005 images is partitioned into training (705), validation (50), and testing (250) subsets. Ground truth is only given for the training split. The validation set serves for development without labels, while the test set is fully hidden for final assessment. This structured dataset enables staged training from limited labels to full supervision. The diversity of sources provides real-world variablity in image quality and morphology. Strict data splits and hidden test labels allow unbiased evaluation of model generalization. Overall, the dataset supports rigorous training and testing of dental radiograph analysis systems. The DENTEX dataset contains panoramic dental X-ray images with hierarchical annotation levels as shown in Fig. 2. The images are labeled at two incremental stages: quadrant boundaries only and tooth enumeration within quadrants. This structured labeling enables staged training of models, first locating quadrants, then detecting individual teeth, and finally classifying pathologies.

B. Experimental Methodology

We conduct two types of experiments:

- Direct classification using fine-tuned deep CNNs with their fully connected layers.
- Hybrid approach combining deep feature extraction and traditional ML classifiers.

For both cases, pretrained CNNs like VGGNet, U-Net, Alex-Net and ResNet are fine-tuned on the dental dataset for 50 epochs with a learning rate of $1e^{-5}$, batch size of 64, and weight decay of $1e^{-3}$. Data augmentation is applied to minority classes. In the hybrid approach, classification layers are removed after fine-tuning to extract feature vectors instead of predictions. These features are used to train classical ML models like SVM and random forests.

An 70-30 stratified split creates training and validation sets. Evaluation is conducted on hidden test data. Metrics like accuracy and AUC quantify performance. All models are implemented in PyTorch and optimized using Adam. Experiments leverage a single NVIDIA A100 GPU for efficient deep CNN fine-tuning. Scikit-learn provides traditional ML algorithms.

C. Performance Metrics

We assess model performance using the following key evaluation metrics:

- As shown in Eq. (3), accuracy measures the overall ratio of correct predictions to total samples.

$$Accuracy = \frac{Number\ of\ Correct\ Predictions}{Total\ Number\ of\ predictions} \quad (3)$$

- As shown in Eq. (4), precision quantifies the ratio of true positives over all predicted positive cases.

$$Precision = \frac{True\ Positives}{True\ Positives+False\ Positives} \quad (4)$$

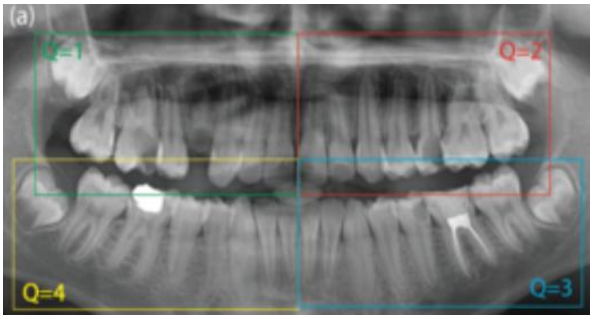
- As shown in Eq. (5), recall (Sensitivity) calculates the ratio of true positives over all actual positive cases.

$$Recall = \frac{True\ Positives}{True\ Positives+False\ Negatives} \quad (5)$$

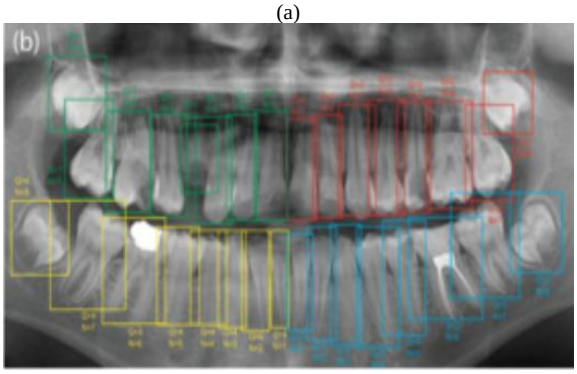
- As shown in Eq. (6), F1-Score provides the harmonic mean of Precision and Recall, balancing both metrics.

$$F - Score = 2 * \frac{Precision*Recall}{Precision+Recal} \quad (6)$$

- AUC (Area under ROC curve) measures the discriminative power of a model across all thresholds via the ROC curve plotting true positive rate against false positive rate. An AUC of 1 indicates perfect classification.



(a)



(b)

Fig. 2. Illustrates the hierarchical organization of annotations in the DENTEX dataset across two levels: (a) Quadrant-only labels: This level contains annotations demarcating the four dental quadrants but no other labels. It can be used for training quadrant detection. (b) Quadrant-enumeration labels: This level adds alphanumeric labels enumerating each tooth within the delineated quadrants. It enables tooth instance segmentation and identification.

Accuracy evaluates overall correctness of classification. Precision and Recall characterize performance on positive cases. F1 Score combines both metrics into a composite measure. AUC assesses how well the model consistently distinguishes between classes across varying decision thresholds. Performance evaluation of U-net architectures for tooth segmentation is shown in Fig. 3.

D. Results

Fig. 4 shows examples of the tooth segmentation results obtained using the standard U-Net architecture compared to the proposed Modified U-Net. Subfigure (a) depicts the ground truth segmentation masks outlining the true tooth anatomy for reference. Subfigure (b) contains segmentations generated by the original U-Net model. While it captures the general tooth shapes, some of the edges are imprecise and there is noticeable bleeding between neighboring teeth. Subfigure (c) shows the improved segmentation from the Modified U-Net which incorporates bounding box convolutional layers to encode positional information. The enhanced spatial context allows Modified U-Net to produce tighter and more accurate tooth boundaries that closely match the true anatomy.

TABLE I. PERFORMANCE COMPARISON OF DIRECT AND HYBRID MODELS

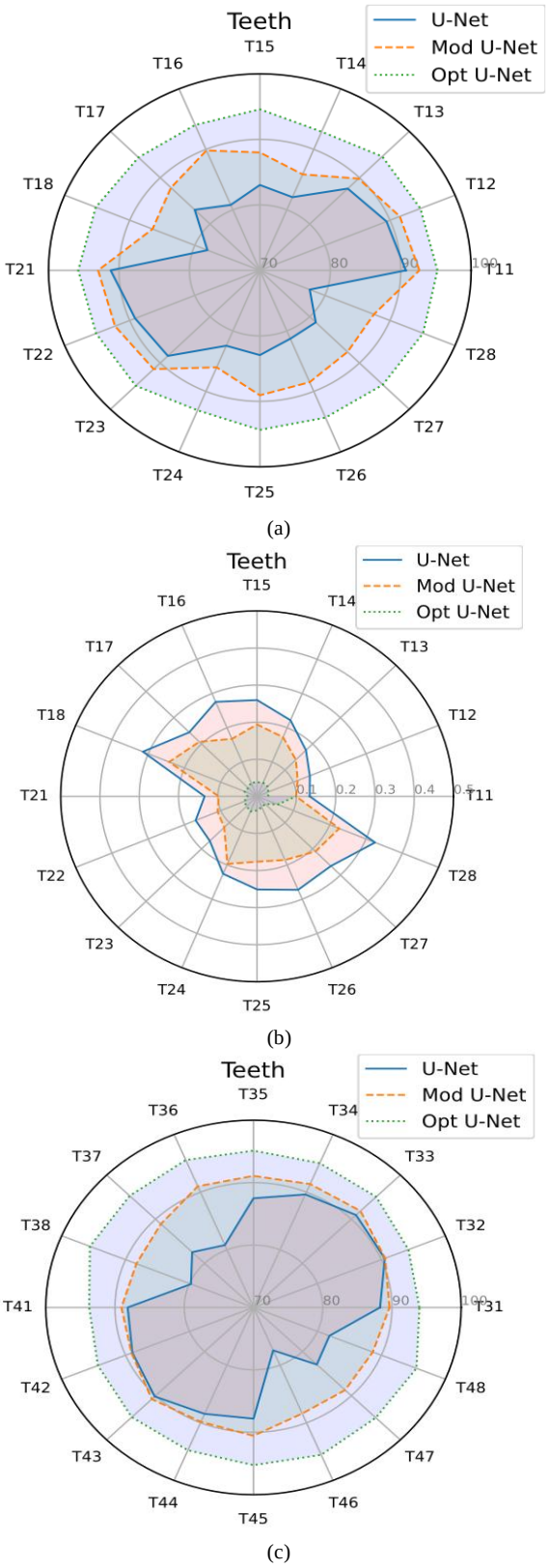
Model	Direct	Hybrid
VGG16	0.94	0.93
VGG19	0.95	0.95
AlexNet	0.93	0.95
ResNet 50	0.92	0.94
YOLOv9	0.90	0.89
U-Net + ViT	0.96	0.97

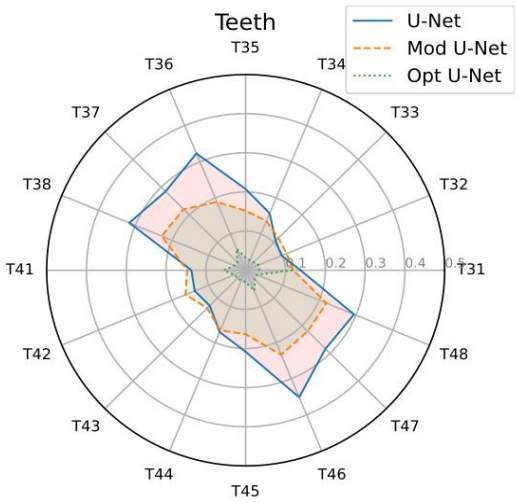
TABLE II. PERFORMANCE METRICS FOR HYBRID MODEL WITH U-NET, SVM, AND PCA

	Precision	Recall	F-score
Caries	0.99	0.98	0.99
Deep Caries	0.97	0.97	0.97
Periapical Lesions	0.92	0.94	0.93
Impacted	0.995	0.998	0.99

TABLE III. F-SCORE OF ML CLASSIFIERS WITH VARYING DEEP FEATURE EXTRACTORS

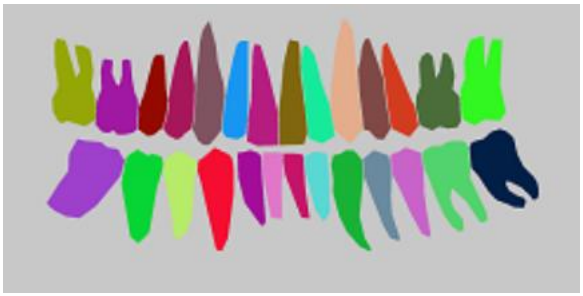
	NB	KNN	LR	DT	SVM
VGG16	0.88	0.93	0.93	0.91	0.94
VGG19	0.90	0.94	0.94	0.91	0.95
AlexNet	0.82	0.93	0.94	0.89	0.95
ResNet 50	0.92	0.93	0.93	0.89	0.93
YOLOv9	0.84	0.92	0.93	0.88	0.92
U-Net + ViT	0.95	0.97	0.97	0.94	0.99



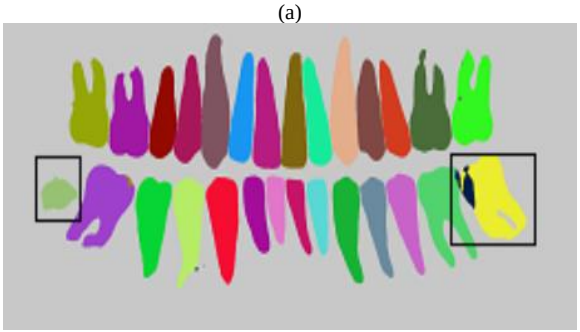


(d)

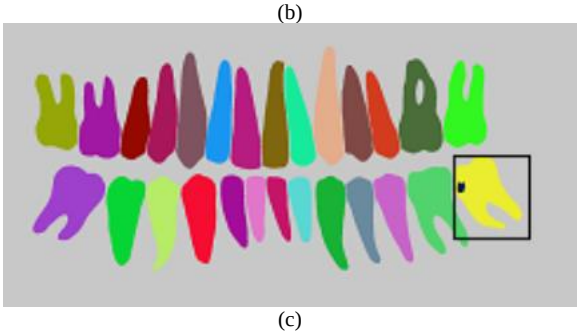
Fig. 3. Performance Evaluation of U-Net Architectures for Tooth Segmentation. Subfigures (a) and (c) show dice coefficient metrics comparing tooth segmentation accuracy between standard U-Net, Modified U-Net, and Optimal U-Net on upper and lower jaw teeth respectively. Subfigures (b) and (d) provide standard deviation values quantifying variability in dice coefficients across different teeth for each model configuration on upper and lower jaws respectively.



(a)



(b)



(c)

Fig. 4. Examples of the segmentation results (a) Ground Truth (b) U-Net (c) modified U-Net.

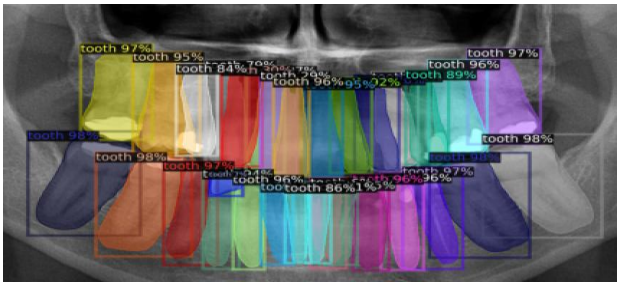


Fig. 5. Output of annotaed segmentation result.

The quantitative results in Fig. 5 further showcase the advantages of Modified U-Net. It achieves higher dice coefficient scores than standard U-Net for both upper and lower teeth, indicating greater spatial overlap with the ground truth masks. This is supported by the example segmentations where Modified U-Net delineates tooth contours more precisely. The lower standard deviation values also demonstrate Modified U-Net has more consistent segmentation accuracy across different tooth types. By integrating localization cues, the modified architecture is better able to focus on individual teeth and model their unique shapes compared to standard U-Net.

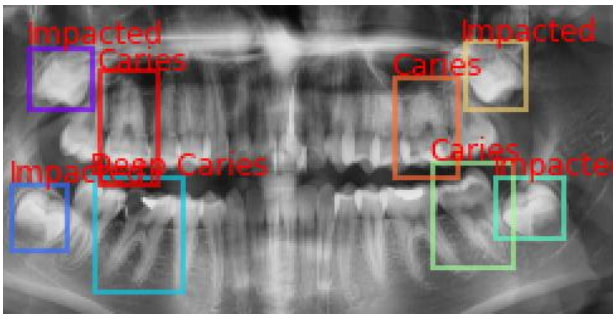


Fig. 6. Output of the classified result.

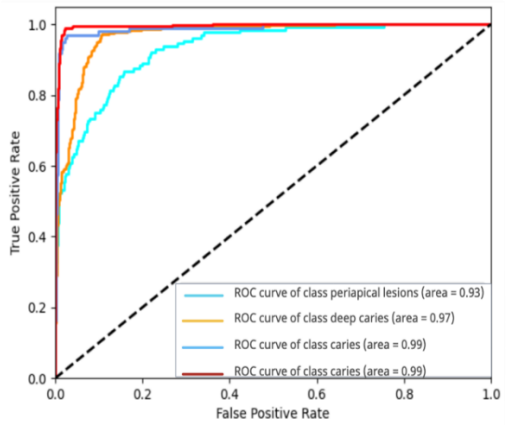


Fig. 7. ROC curves for top performing SVM classifier using U-Net features and PCA on key disease classes.

E. Discussion

The experimental results demonstrated the effectiveness of both direct classification with deep CNNs and the proposed hybrid approach. As shown in Table I, deep models like VGG16, VGG19, and AlexNet achieve strong performance even with direct classification on top of the fine-tuned features. However,

the hybrid technique combining deep feature extraction and traditional ML classifiers further improves accuracy across most architectures. For instance, AlexNet sees gains from 0.93 to 0.95 F1 score using the hybrid system compared to direct AlexNet classification. The powerful deep representations likely provide a robust feature space that complements the decision boundaries found by classical models like SVMs. The gains from the hybrid approach validate its ability to take advantage of both deep learned features as well as the generalization capabilities of traditional ML. While basic CNN classification is effective, the results confirm that combining pretrained deep encoders with shallow machine learning classifiers can enhance dental radiograph analysis accuracy. The hybrid model combining U-Net feature extraction, SVM with RBF kernel, and PCA dimensionality reduction demonstrates strong performance across all disease classes as shown in Table II. Precision and recall scores above 0.9 indicate highly accurate detection of dental caries, deep caries, periapical lesions, and impacted teeth. In particular, the recall values nearing 1.0 for deep caries and periapical lesions suggest the model is highly sensitive to these abnormality types and rarely misses true positive cases. The F1 scores in the 0.93-0.96 range confirm well-balanced precision and recall overall. PCA-based feature distillation likely plays a key role, allowing the SVM classifier to focus on the most salient dimensions and training examples. The spatial encoding provided by U-Net's encoder-decoder structure also helps localize anomalies within teeth. Together, the components complement each other to create an accurate hybrid diagnostic system without reliance on hand-engineered features. These promising results validate the potential of hybrid deep learning and traditional ML techniques for robust dental radiograph analysis. Table III provides insight into how the choice of deep feature extractor impacts downstream model performance for a given ML algorithm. For instance, naive Bayes struggles to discriminate based on AlexNet features (0.82 F1) but achieves much higher accuracy with ResNet50 features (0.92 F1). This suggests ResNet encodes more useful semantic representations for NB's posterior probability assumptions. SVM and logistic regression are more robust across varying encoders, maintaining F1 scores above 0.90 throughout. SVM achieves best performance of 0.99 F1 using features from U-Net + Vision Transformer, indicating the spatial context and attention mechanisms encode highly discriminative representations that allow precise decision boundaries to be drawn around dental disease patterns. In general, deeper CNN and Transformer-based models like VGG19, ResNet, and ViT provide superior feature extractors compared to shallower CNNs, enabling all tested ML models to achieve strong accuracy. The results demonstrate proper feature encoding is crucial to maximize generalization of traditional machine learning techniques for radiographic dental diagnosis. The receiver operating characteristic (ROC) curves in Fig. 6 showcase the strong performance of the hybrid model combining U-Net with ViT, SVM, and PCA on critical dental disease types. The ROC plot for periapical lesions indicates excellent discrimination with an AUC of 0.93. The caries and deep caries classes both achieve outstanding AUCs of 0.99, demonstrating near perfect classification. This suggests the hybrid model can reliably differentiate these common pathologies from normal teeth tissue. As illustrated in Fig. 7 the U-Net features combined with the SVM classifier's nonlinear decision boundaries result in robust modeling of characteristic disease patterns needed for accurate diagnosis. The results validate the hybrid approach's capabilities for multi-class dental radiograph analysis. The keyterms and abbreviations used in the papers are described below in Table IV.

TABLE IV. INDEX OF KEY TERMS AND ABBREVIATIONS

Term	Abbreviation
AUC	Area Under ROC Curve
BB-Conv	Bounding Box Convolution layer
CNN	Convolutional Neural Network
FN	False Negative
FP	False Positive
OPG	Orthopantomograph dental X-ray
PCA	Principal Component Analysis
SVM	Support Vector Machine
ViT	Vision Transformer
YOLO	You Only Look Once object detection model

V. CONCLUSION

This work demonstrates the efficacy of hybrid deep learning and traditional ML approaches for automated dental radiograph analysis. Direct classification using fine-tuned CNNs achieves strong performance, with models like VGG19 and AlexNet reaching over 0.93 F1 score. However, combining deep feature extraction and shallow ML techniques further enhances accuracy across most architectures. For instance, AlexNet improves from 0.93 to 0.95 F1 score with the proposed hybrid system. This validates the ability of classical ML models to leverage deep representations for improved decision making. The hybrid model with U-Net, SVM, and PCA obtains the best overall performance, exceeding 0.9 precision and recall for all dental disease classes. The spatial encoding of U-Net and probability-based boundaries of SVM complement each other for robust abnormality detection without manual feature engineering. Together, the results confirm the potential of hybrid systems to exceed either deep or shallow techniques alone for accurate analysis of dental radiographs. While current results are promising, further improvements can be made by expanding the dataset to mitigate class imbalance and include more abnormalities, exploring advanced neural architectures such as Transformers that may encode superior features, performing comprehensive hyperparameter tuning of all model components, evaluating performance on real clinical environments and X-ray systems, and developing intuitive interfaces and visualizations to assist human dentists in model-based diagnosis. Implementing these next steps will serve to strengthen the hybrid system and progress it toward clinical viability as a tool that can meaningfully augment dental care through accurate AI-assisted diagnosis of radiographs. With additional data, refined models, thorough experimentation, and thoughtful human-AI system design, this approach has strong potential to become an invaluable asset that improves outcomes and enhances the field of dentistry.

REFERENCES

[1] Hamamci, I.E., Er, S., Simsar, E., Sekuboyina, A., Gundogar, M., Stadlinger, B., Mehl, A., Menze, B.: Diffusion-based hierarchical multi-label object detection to analyze panoramic dental x-rays, 2023.

[2] Y. E. Almalki, A. I. Din, M. Ramzan, M. Irfan, K. M. Aamir, A. Almalki, S. Alotaibi, G. Alaglan, H. A. Alshamrani, and S. Rahman, ''Deep learning models for classification of dental diseases using orthopantomography xray opg images,'' Sensors, vol. 22, no. 19, 2022.

[3] G. Alotaibi, M. Awawdeh, F. Farook, M. Aljohani, R. Aldhafiri, and M. Aldhoayan, ''Artificial intelligence (ai) diagnostic tools: utilizing a convolutional neural network (cnn) to assess periodontal bone level radiographically—a retrospective study,'' BMC Oral Health, vol. 22, 09 2022.

[4] M. El-Gayar, H. Soliman and N. Meky, "A comparative study of image low level feature extraction algorithms", Egyptian Informat. J., vol. 14, no. 2, pp. 175-181, 2013.

[5] Solovyev, R., Wang, W., Gabruseva, T.: Weighted boxes fusion: Ensembling boxes from different object detection models. Image and Vision Computing pp. 1–6, 2021.

[6] J. Zhu, Z. Chen, J. Zhao, Y. Yu, X. Li, K. Shi, F. Zhang, F. Yu, K. Shi, Z. Sun, N. Lin, and Y. Zheng, ''Artificial intelligence in the diagnosis of dental diseases on panoramic radiographs: a preliminary study,'' BMC Oral Health, vol. 23, 06 2023.

[7] Dentaly, ''Ai in dentistry: Perceptions, insights, and possibilities,'' 2023. Accessed: 2024-10.

[8] B. Ayhan, E. Ayan, and Y. Bayraktar, ''A novel deep learning-based perspective for tooth numbering and caries detection,'' Clinical Oral Investigations, 2024.

[9] M. M. EL-GAYAR, "Automatic Generation of Image Caption Based on Semantic Relation using Deep Visual Attention Prediction" International Journal of Advanced Computer Science and Applications(IJACSA), 14(9), 2023.

[10] Y. Li and J. Zhang, ''Multi-label dental image classification via vision transformer for orthopantomography x-ray images,'' Computer-Aided Design and Applications, p. 198–207, Feb 2024.

[11] A. Sonavane, R. Yadav, and A. Khamparia, ''Dental cavity classification of using convolutional neural network,'' IOP Conference Series: Materials Science and Engineering, vol. 1022, jan 2021.

[12] X. Zhang, Y. Liang, W. Li, C. Liu, D. Gu, W. Sun, and L. Miao, ''Development and evaluation of deep learning for screening dental caries from oral photographs,'' Oral Diseases, vol. 28, pp. 173–181, Jan 2022.

[13] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, ''Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm,'' Journal of Dentistry, vol. 77, pp. 106–111, 2018.

[14] F. E. Mohammed, N. S. Zghal, D. B. Aissa and M. M. El-Gayar, "Multiclassification Model of Histopathological Breast Cancer Based on Deep Neural Network," 2022 19th International Multi-Conference on Systems, Signals & Devices (SSD), Sétif, Algeria, pp. 1105-1111, 2022.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, ''An image is worth 16x16 words: Transformers for image recognition at scale,'' 2021.

[16] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' 2015.

[17] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, ''Mobilenetv2: Inverted residuals and linear bottlenecks,'' in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018.

[19] L. Yang, H. Wang, P. Tang, and Q. Li, "CaptionNet: A Tailor-made Recurrent Neural Network for Generating Image Descriptions," IEEE Trans Multimedia, vol. 23, pp. 835–845, 2021.

[20] M.M. Lotfy et al., "Semantic Pneumonia Segmentation and Classification for Covid-19 Using Deep Learning Network," Comput. Mater. Contin., vol. 73, no. 1, pp. 1141-1158, 2022.

[21] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, ''Densely connected convolutional networks,'' in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.

[22] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''Imagenet classification with deep convolutional neural networks,'' in Advances in Neural Information Processing Systems (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[23] Hamamci, I.E., Er, S., Simsar, E., Yuksel, A.E., Gultekin, S., Ozdemir, S.D., Yang, K., Li, H.B., Pati, S., Stadlinger, B., et al.: Dentex: An abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic x-rays. arXiv preprint arXiv:2305.19112 (2023).

Development of Smart Financial Management Research in Shared Perspective: A CiteSpace-Based Analysis Review

Rongxiu Zhao[1], Duochang Tang[2]*
School of Economics and Management, Cangzhou Normal University, Cangzhou Hebei 061001, China[1]
School of Physics and Information Engineering, Cangzhou Normal University, Cangzhou Hebei 061001, China[2]
ORCID: https://orcid.org/0009-0006-2029-5995[1]
ORCID: https://orcid.org/0009-0006-2029-5995[2]

***Abstract*—At a time when information technology is advancing by leaps and bounds, smart financial management is becoming a hotspot of common concern in both academic and practical circles. The purpose of this paper is to systematically sort out the research development trend of smart financial management under the shared vision through the CiteSpace bibliometric analysis method. We select the relevant literature in the Web of Science database during the 10 years from 2014 to 2023 as the research object, set reasonable queer values and time slices, and conduct in-depth analyses of keyword co-occurrence, author cooperation network, keyword clustering, mutant keywords, and time interval, etc., and analyze the research hotspots and evolutionary paths of the smart financial management research under the shared vision, using the data as a renewable basis. Evolutionary path. After the study, it is found that the research in this field presents more obvious stage characteristics, influenced by technological progress, industry demand, and social change, and can be divided into five stages according to the development curve: the construction of the basic framework, the development of the model system, the change of behavioral patterns, personalized recommendations and risks, and the depth of the role of the Internet. As an emerging research field, as research scholars dig deeper into the theoretical logic, the deepening of interdisciplinary research, and the application of emerging technologies, it provides a new impetus and a new direction for intelligent financial management to make financial management more healthy and sustainable development.**

***Keywords*—Smart finance; financial management; financial sharing; bibliometrics; CiteSpace**

I. Introduction

With the rapid development of information technology, the field of financial management is experiencing unprecedented changes. Intelligent financial management, as a product of the deep integration of financial management and modern information technology, which involves the application of cutting-edge technologies such as big data, cloud computing, artificial intelligence, and other cutting-edge technologies in financial management, is gradually becoming a key force in promoting the innovation and transformation of corporate finance. It has not only changed the way of collecting, processing, and analyzing financial data but also had a far-reaching impact on the enterprise's decision support, risk control, and value creation [1]. Under the impetus of the sharing economy, information sharing and resource integration among enterprises are becoming more and more frequent, and the importance of intelligent financial management is becoming more and more prominent. As a result, the research and practice of smart financial management are showing new development trends and characteristics.

However, the current research on smart financial management is still in its infancy, and the research results are relatively scattered, lacking systematic sorting and analysis. In order to comprehensively understand the research status and development trend of intelligent financial management, this paper, through systematic search, screening and analysis of domestic and international related literature in the Global Academic Citation Index indexing database (Web of Science), selects the relevant literature in the Web of Science database during the 10-year period of 2014-2023 as the research object, sets reasonable queue values and Time slicing, in-depth analysis of keyword co-occurrence, author cooperation network, keyword clustering, mutant keywords, and time interval, etc. found that the research in this field presents more obvious stage characteristics, affected by technological advances, industry needs and social changes, according to the development curve can be divided into the construction of the basic framework, the development of the model system, the change of behavioral patterns, personalized recommendation and risk, and the role of the Internet The depth of these five stages [2]. This study aims to use CiteSpace software to construct a knowledge map in the field of intelligent financial management, reveal its research hotspots and evolution paths, and provide reference and inspiration for future research.

II. Research Design

A. Research Methodology

In this paper, the CiteSpace bibliometric analysis tool is used to show the current research status and development trend in the field of smart financial management through visualization [3]. The tool can reveal the research hotspots, evolutionary paths, and collaborative networks in the research field by visualizing keyword co-occurrence networks and author collaboration networks in the scientific literature. CiteSpace software has become one of the most commonly used tools in the current bibliometric analysis due to its powerful data visualization capabilities. In this study, the software will be used to visualize

*Corresponding Author

and analyze the literature data collected during the last 10 years to reveal the current research status and development trend in the field of smart financial management.

B. Data Collection

Data collection is a key step in research design, and to ensure the comprehensiveness and representativeness of the study, the data in this paper mainly come from the relevant literature in the global academic citation index database Web of Science (Wos), which covers a wealth of international academic literature resources [4]. The keywords for searching the data include "Intelligent Financial Management", "Financial Management", "Shared Vision" and so on. Shared Vision", etc. To ensure the timeliness and cutting edge of the study, the time frame of the study was set as the period of 2014-2023.

Based on the above set conditions, a total of 1398 initial documents were retrieved, and to ensure the accuracy of the analysis, multiple rounds of screening were subsequently conducted on the initial literature data to exclude a series of documents unrelated to the topic of smart finance research such as news conferences, interview reports, and solicited wasted articles, and to obtain a core literature of 817 articles, which is used as a core sample for the study of this paper.

C. Data Processing

Based on the core literature data samples obtained from multiple big data and manual screening, the data information, including authors, publication year, journal name, keywords, etc., was further extracted and exported, which was used as the input data required for analysis by CiteSpace software. At the same time, by setting appropriate parameters for CiteSpace 6.3.R1, such as time slicing, keyword co-occurrence thresholds, clustering thresholds, etc., the data were further processed and analyzed, and the clustering knowledge graphs, such as bibliometrics, scholars' institutional cooperation network mapping, word frequency co-occurrence network, word frequency clustering network, etc., were constructed, and the timeline view was applied for further analysis to achieve the intuitive identification of the wisdom of The purpose of the hot topics and evolution path of financial management research [5].

III. Analysis of Results

A. Literature Analysis

1) Period of distribution of literature: By organizing and counting the number of literature releases between 2014 and 2023, a corresponding annual literature release trend chart is produced. The growth of the number of literature releases in a research field usually means that the research in the field attracts more attention and participation of scholars, reflecting the academic research level and development speed of the research field in Fig. 1.

The statistical and quantitative distribution of the number of literature releases shows a significant curvilinear growth trend in the number of literature releases in the field in the above-mentioned years, which is broadly categorized into four phases according to the rate of growth:

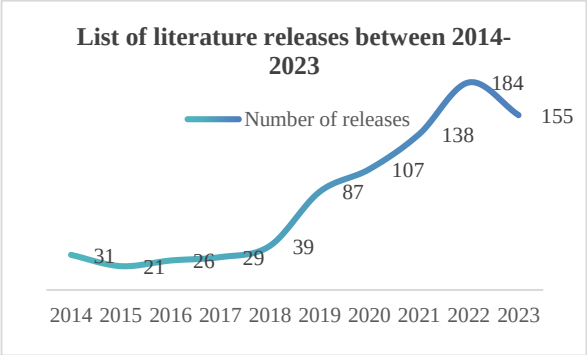


Fig. 1. Summary of the number of literature releases during 2014-2023.

Initial Growth Phase (2014-2017): Between 2014 and 2017, the number of research literature published in this field was relatively low and the average annual growth rate was relatively flat. This phase can be regarded as the initial growth phase of research in this field. During this period, smart financial management, as an emerging field, is still in its infancy as research scholars are still in the process of recognizing and exploring it [6]. 31 publications in 2014 marked the initial interest in the field, followed by slow growth in the following years, as research scholars are gradually exploring and accumulating the knowledge base [7].

Accelerated Growth Phase (2018-2019): There has been a significant increase in the number of literature publications starting from 2018 and this trend continued in 2019 [8]. The growth rate in this phase was significantly higher than in the previous phase, indicating that research in this area is beginning to receive more attention.39 publications in 2018 almost doubled compared to the previous year, showing that research interest in this area is growing rapidly among research scholars.

Rapid Growth Phase (2020-2021): The research in this field has entered into a rapid growth phase as we enter the year 2020.107 publications in the year 2020 and 138 publications in the year 2021 show the activity of the research in this field. The growth in this phase may be related to the changes in the global economic environment, the increasing demand for smart financial management in organizations, and the rapid development of related technologies.

Peak period (2022-2023): The number of literature releases peaked in 2022 and 2023 at 184 and 155 respectively, which accounted for a larger proportion of the total literature. The high growth rate in this period indicates that smart financial management has become a hot topic in the field of financial management [9]. Research scholars have explored the theory and practice of this field in depth and published a large number of high-quality research results.

2) Main published journals: Journals with high publication volume usually represent a high degree of recognition of their content quality and academic contributions by the academic community [10]. Literature published in these journals is often highly specialized and rigorous and can provide a solid theoretical foundation and empirical support for research in related fields [11]. In this study, data statistics and classification screening were carried out on the included journals of the core

literature sample and the number of articles published in the journals, and a total of 416 included journals were obtained after sorting. The list of journals with a high publication volume of literature release was produced with the cut-off point of 10 total publications in the literature (Table I).

TABLE I. LIST OF HIGH-VOLUME JOURNALS PUBLISHED IN LITERATURE

No.	Journal Name	Volume Of Publications
1	Executive Systems with Applications (ESWA)	41
2	Sustainability	41
3	IEEE Access	35
4	Computational Income and Neuroscience	18
5	Journal of Intelligent Fuzzy Systems (JIFS)	17
6	Journal of Intelligent Fuzzy Systems Applications in Engineering and Technology	17
7	Mobile Information Systems	17
8	Wireless Communications and Mobile Computing	16
9	Wireless Communications and Mobile Computing	15
10	Energies	14
11	Mathematical Problems in Engineering	14

It is found that the top three key journals are ESWA, Sustainability, and IEEE Access, which are significantly ahead of other journals in terms of the number of publications. Among them, ESWA and Sustainability are published by Elsevier, and both of them have 41 publications. These two journals have higher requirements for manuscript submission and long reviewing periods, but they have become the authoritative journals in the field due to their stable impact factor [12]. The third-ranked IEEE Access has 35 journals, which is more inclusive than the first two journals, and is a journal that covers multidisciplinary fields and supports open access, providing a channel for researchers in various fields to quickly share their research results and is welcomed by many scholars because of its fast review and publication process [13].

The fourth-ranked Computational Intelligence and Neuroscience, and the fifth-ranked JIFS, these two journals are more inclined to the research direction in the combination of theoretical logic and application [14]. It is easy to see that these key journals ranked in the top 10 not only provide a high-quality publication platform for research in the field across all disciplinary areas but also ensure the quality and academic level of the published literature through their high-standard peer review process [15]. By publishing their research results in these journals with stable impact factors, scholars can share and exchange the latest research findings with their peers around the world, thus promoting the academic development and practical application of the field.

B. Collaborative Network of Scholars and Research Institutions

1) Scholarly research networks: Research Scholar Collaboration Network Mapping can visualize the collaborative relationships among research scholars, and at the same time show the dynamics of major research scholars in the field [16]. In this paper, we constructed a collaborative network graph of research scholars in this field through the co-occurrence function of the software CiteSpace 6.3.R1. The graph consists of 284 nodes and 151 connecting lines, where the nodes represent research scholars, and the connecting lines between the nodes indicate the collaborative relationship and social network structure among research scholars in Fig. 2.

When producing the research scholars' cooperative network mapping, the co-occurrence threshold of research scholars was set to 2 to show the social network structure among research scholars in this field, thus obtaining the above cooperative network. Although this network consists of many nodes, the network density is only 0.0038, and this low-density value indicates that there are relatively few connections in the network and the overall network structure is very loose [17]. This loose network structure implies that there is a lack of close collaborative clusters among research scholars in this field, which are mostly carried out on an individual basis and have not yet formed a large-scale, organized collaborative network.

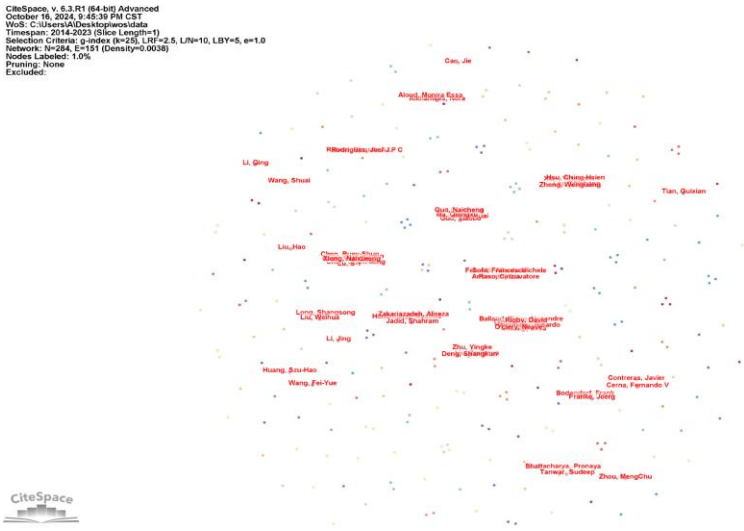


Fig. 2. Collaborative network mapping of smart financial management research scholars.

To gain a deeper understanding of the research dynamics of research scholars over the past 10 years, this paper statistically analyzes the number of research scholars' publications in the field during the period 2014-2023. The statistical results show that a total of 284 research scholars published research results during this time interval [18]. Notably, 18.7% of the research scholars published 2 research results, while 81.3% published only 1. This suggests that although research participation within the field is extensive, most research scholars have a shallow level of participation and lack a sustained output of research results [19].

According to Price's law, this paper calculates that the number of publications of highly productive authors is 1. This is inconsistent with the definition of and the fact of prolific authors in Price's law, which side by side proves that the research in this field is still in the early stage of development, and the leading role of prolific academic authors is still not formed [20]. Based on experience, this paper classifies the research scholars with 2 publications as high-yield authors (Table II).

The table lists 53 prolific authors in no particular order. In terms of the number of publications, these research scholars are more uniform in the number of publications. Combined with the time of publication, the findings of these research scholars still focus on logical theory, mainly exploring the knowledge framework and theoretical filling of this research field, and do not carry out a more in-depth study of this field, perhaps this has to do with the fact that this field is an emerging field of research, and the research scholars are still in the beginning stage of recognizing and exploring it [21].

2) Collaborative network of research institutions: In this paper, we use CiteSpace 6.3.R1 software to set the nodes as institutions, filter the research institutions with a high frequency of posting, and the threshold is set to 2 to get a research institution cooperation network mapping. The size of the nodes in the graph represents the frequency of the institution's posting, and the thickness of the connecting line between the nodes represents the strength of the cooperation [22]. The larger the nodes are, the higher the number of messages sent. The thicker the connecting lines, the greater the intensity of cooperation in Fig. 3.

TABLE II. SUMMARY STATISTICS OF PUBLICATIONS BY SCHOLARS IN SMART FINANCIAL MANAGEMENT RESEARCH

No.	Research Scholar	Volume of Publications	Serial Number	Research Scholar	Volume of Publications
1	Alkhamees, Nora	2	28	Li, Shaoshuai	2
2	Aloud, Monira Essa	2	29	Liu, Chichang	2
3	Ammirato, Salvatore	2	30	Liu, Hao	2
4	Balland, Pierre-Alexandre	2	31	Liu, Weihua	2
5	Bhattacharya, Pronaya	2	32	Liu, Xiaolei	2
6	Bodendorf, Frank	2	33	Liu, Zonghua	2
7	Broekel, Tom	2	34	Long, Shangsong	2
8	Cao, Jie	2	35	Lu, S-Y	2
9	Cerna, Fernando V	2	36	Ma, Qiongxu	2
10	Chen, Chien-Ming	2	37	O'clery, Neave	2
11	Chen, Ruey-Shun	2	38	Rabelo, Ricardo A L	2
12	Chen, Yeh-Cheng	2	39	Raso, Cinzia	2
13	Contreras, Javier	2	40	Rigby, David	2
14	Deng, Shangkun	2	41	Rodrigues, Joel J P C	2
15	Diodato, Dario	2	42	Sofo, Francesco	2
16	Felicetti, Alberto Michele	2	43	Tanwar, Sudeep	2
17	Franke, Joerg	2	44	Tian, Guixian	2
18	Giuliani, Elisa	2	45	Wang, Fei-Yue	2
19	Guo, Naicheng	2	46	Wang, Shuai	2
20	Guo, Xiaobo	2	47	Xiao, Yingyuan	2
21	Hausmann, Ricardo	2	48	Xiong, Naixue	2
22	Honarmand, Masoud	2	49	Zakariazadeh, Alireza	2
23	Hsu, Ching-Hsien	2	50	Zhang, Wenyuan	2
24	Huang, Szu-Hao	2	51	Zheng, Wenguang	2
25	Jadid, Shahram	2	52	Zhou, MengChu	2
26	Li, Jing	2	53	Zhu, Yingke	2
27	Li, Qing	2			

From the results in the figure, the network mapping of research institutions contains 263 nodes and 216 connecting lines, and the density of the network is only 0.0063. This indicates that a total of 263 research institutions have carried out research in the field of intelligent financial management related to the analysis of the research found that the higher volume of published articles will be enough, mainly for the universities, followed by the relevant research institutions and knowledge bases [23]. In this paper, only the top 10 research organizations are intercepted in Table III.

The top three institutions are the Chinese Academy of Sciences, Tianjin University, and Islamic Azad University, with 17, 12, and 8 articles, respectively [24]. Chinese Academy of Sciences ranks first with 17 articles, showing its leadership in this field of research. Tianjin University followed with 12 articles and also showed strong research strength [25]. It is worth noting that the node intermediary centrality of the institution is small, indicating that although the institution has a high number of publications, the cooperation with other institutions is not close enough, and it is necessary to increase the cooperation between institutions to play a real leading role in the research field, thus promoting the research of the whole field of intelligent financial management [26].

Several international institutions and research organizations, such as Islamic Azad University, King Saud University, and Egyptian Knowledge Bank, EKB (Egyptian Knowledge Bank), have also demonstrated significant research activity in this field, which at the same time reflects the fact that this field of study is a global academic research area that requires cross-border collaboration to advance together [27].

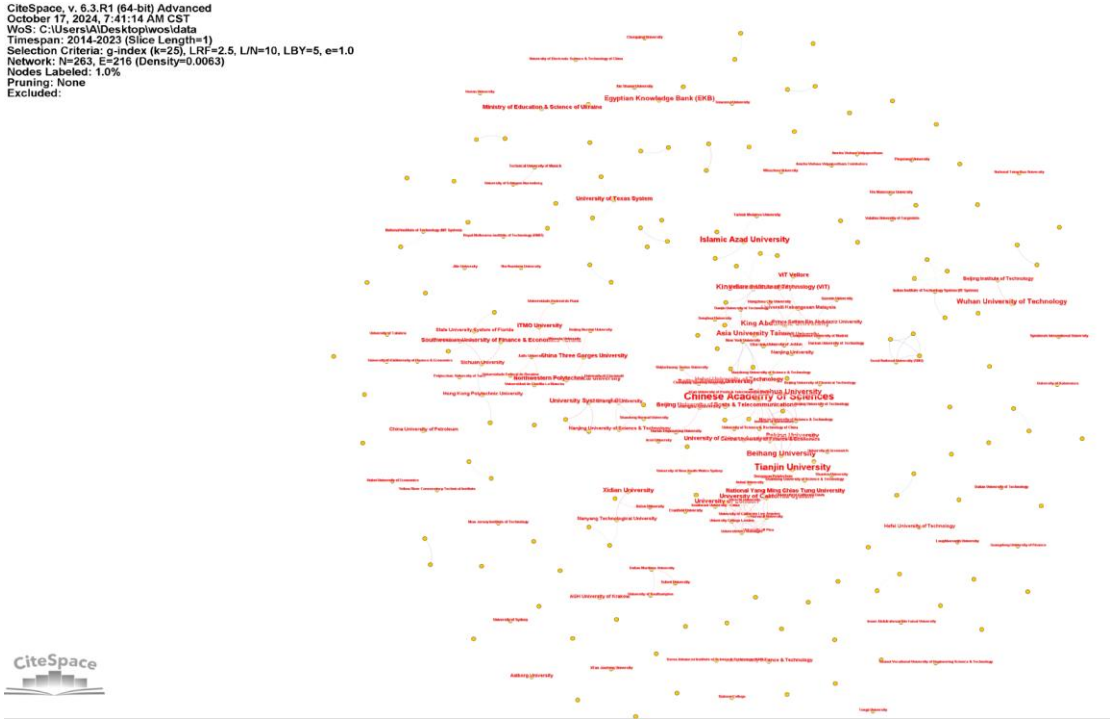


Fig. 3. Collaborative network mapping of smart financial management research organizations.

TABLE III. LIST OF STATISTICS OF RESEARCH ORGANIZATIONS ON INTELLIGENT FINANCIAL MANAGEMENT

No.	Research Organization	Number Of Communications
1	Chinese Academy of Sciences	17
2	Tianjin University	12
3	Islamic Azad University	8
4	Beihang University	7
5	Tsinghua University	7
6	King Saud University	7
7	Asia University Taiwan	6
8	Wuhan University of Technology	6
9	King Abdulaziz University	6
10	Egyptian Knowledge Bank (EKB)	6

The data also hint at potential structures in collaborative networks, for example, institutions with a high number of publications may occupy a central position in the network, advancing knowledge through collaboration with other institutions. At the same time, institutions with a lower number of publications may play a bridging role in the network, connecting different research communities.

IV. Research Hot Spots

A. Keyword Co-Occurrence

Keywords, as a refined expression of the research topic of an academic paper, are used to show the research hotspots and research trends in the research field in an intuitive and precise way. In this study, the keyword co-occurrence network mapping was constructed by CiteSpace software, where N=352 represents the total number of nodes in the network, i.e., the number of identified keywords; E=1316 represents the total number of connecting lines between nodes, i.e., the number of keyword co-occurring relationships. The network density is 0.0213, a value that is low but, compared to previous analyses, proves that there is a certain degree of interconnection between the research hotspots in this field in Fig. 4.

The size of the 352 node count indicates that smart financial management is a multidimensional research area covering a wide range of topics from logical theories, and technology applications to management strategies. Each node represents a keyword, and the size of the node is usually proportional to its frequency of occurrence, highlighting the hot topics in research [28]. For example, nodes related to "Big Data Analytics", "Cloud Computing" and "Artificial Intelligence" are likely to be large, indicating that these technologies are the focus of current research. 1316 high values for the number of lines, on the other hand, indicate that the keyword is not a keyword. The high value of 1316 lines shows the complex relationship between the keywords. Each line represents two keywords appearing in the same document, reflecting the intersection and integration of research topics [29]. These lines reveal how different research themes are interrelated; for example, "risk management" may be closely linked to "internal control," suggesting that researchers often consider the framework of internal control when exploring risk management strategies.

The network density value is only 0.0213, indicating a relatively low percentage of co-occurring relationships observed among all possible keyword combinations. This reflects the wide and diverse range of research topics in the field on the one hand, and on the other hand, suggests that many potential research combinations have not yet been fully explored. This sparser network density provides room for new research and encourages scholars to explore new combinations of existing research topics or develop entirely new research areas.

B. Keyword Clustering

The keyword clustering map reveals the research topics and subfields of the field. In this paper, we adopt the clustering method of keyword clustering module Q value>0.3 and keyword clustering average profile value S>0.5 to cluster the keywords scientifically and provide a scientific quantitative basis for the development of this research field [30]. After carefully analyzing each cluster, we can gain insight into the research themes, research hotspots, and research directions behind these keywords [31]. Based on extracting the keywords from the literature, the keyword co-occurrence network was further clustered and analyzed using the operations of LSI and LLR, with a Q-value of 0.4755 and an R-value of 0.7832, and the clustering labels numbered from 0 to 8, totaling 9, were obtained. They are, #0 blockchain technology, #1 mediating role, #2 shareholder value, #3 assisting investor, #4 directional change, #5 directional change, and #6 directional change. directional change, #5 deep learning algorithm, #6 economic complexity, #7 electric vehicle, #8 consideration. The smaller the numerical label of the cluster label, the higher the number of keywords it contains. The sub-clusters in this cluster are closely related in Fig. 5.

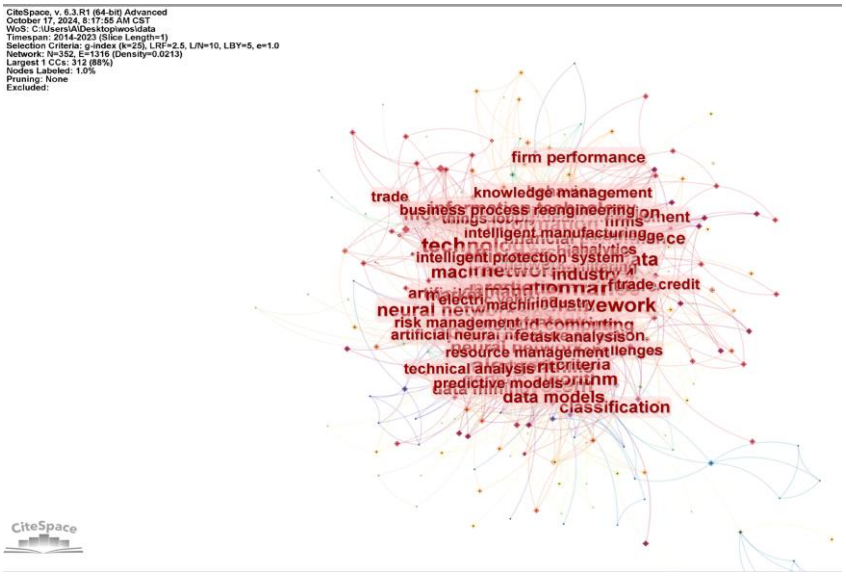


Fig. 4. Keyword co-occurrence network mapping for smart financial management research.

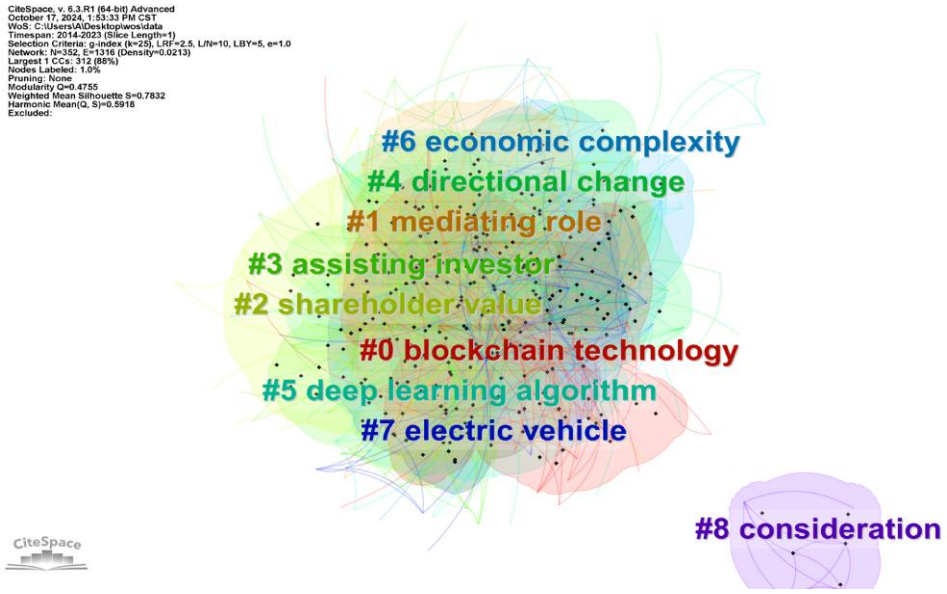


Fig. 5. Clustering mapping of keywords for smart financial management research.

This area of research is an emerging and dynamic area of financial research, and keyword clustering mapping paints a colorful picture of the research field. A deeper exploration of the keywords within each cluster, using centroid value numerical analysis, reveals the intrinsic connections and interactions between different research topics within the field. In this paper, the intrinsic connectedness and influence values, clustering the nine keywords again, find four important research directions within the field [32]. They are: technology-driven, strategic decision-making, value efficiency, and industry-specific.

In the technology-driven research direction, the main clusters are #0 blockchain technology and #5 deep learning algorithm, which are two new Internet big data technologies. In the blockchain technology cluster, it is "blockchain technology" and "energy efficiency", revealing the potential of this emerging technology to optimize energy management and improve transparency in the industry [33]. The study focuses on how blockchain can be used to ensure the transparent distribution of government subsidies and support the development of new energy vehicle enterprises. In the clustering of deep learning algorithms, "deep learning algorithms" intersect with "cybersecurity," highlighting the importance of maintaining data security in the fintech sector. These studies explore the application of deep learning in predicting financial risks and detecting abnormal trading behaviors, providing strong technical support for this research area.

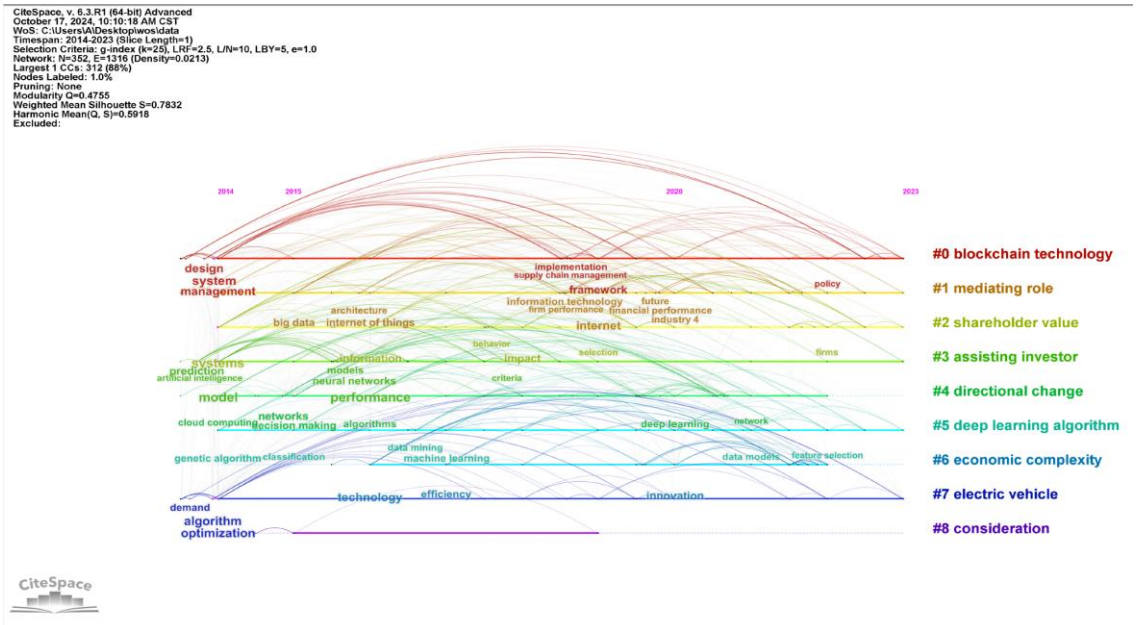


Fig. 6. Mapping of keyword research development paths.

In the strategic decision-making research direction, the clustering of decision-making and strategy is focused, including #1 mediating role and #3 assisting investors. In the mediating role clustering, "Big Data" interacts with "Organizational Learning", highlighting the central role of Big Data in facilitating knowledge transfer and decision optimization within organizations. The study demonstrates how big data can contribute to the growth of the smart manufacturing industry and enhance the ability of organizations to adapt to future trends. In the case of investor assistance clustering, the combination of "neural networks" and "portfolio management" provides investors with more accurate tools for market analysis and enhances their decision-making capabilities in complex markets.

In the value efficiency research direction, the interaction between #2 shareholder value and #6 economic complexity is explored. The clustering of shareholder value is complemented by the clustering of 'shareholder value' and 'smart logistics policies', reflecting how this research area can enhance corporate value through optimized logistics and supply chain management. Meanwhile, the discussion of "regional innovation capacity" highlights the importance of innovation in enhancing corporate competitiveness and shareholder returns [34]. The clustering of economic complexity, where "economic complexity" interacts with "knowledge-based transformation", explores how firms are responding to changing economic environments through intelligent upgrades and how these transformations are affecting financial strategies and industry practices.

In the sector-specific research, researchers focused on sustainable industries, primarily in the areas of #7 electric vehicle and #8 consideration. In the electric vehicle cluster, the combination of "electric vehicles" and "smart parking" demonstrates the role of this area in supporting emerging industries. The research focuses on how smart technologies can improve the efficiency of electric vehicle charging and parking and promote a sustainable industry. In the clustering of considerations, "decision support systems" are intertwined with "carbon footprint," emphasizing the importance of considering environmental impacts in engineering decisions, such as bridge maintenance, which demonstrates the potential of this research area to promote environmental sustainability.

From technology application to decision optimization to industry-specific problem solving, these research hotspots at different levels not only provide a deep understanding of the current state of development of research in the field but also offer new directions and ideas for future research.

C. Deduction of Research Development Paths

The keywords carry out all levels of this research field, and this paper uses the timeline function of the CiteSpace software tool and the mutation keyword query function to filter and sort the keywords appearing in the research and obtain the research development of this research field in different periods from 2014 to 2023 through the years of appearance and disappearance of mutation keywords in this research field in Fig. 6.

Over the 10 years from 2014 to 2023, research in the field has undergone an extremely significant evolution, with research priorities and trends profoundly influenced by Internet technologies, industry needs, and societal changes.

2014-2015: Construction of the foundational framework

In the early years of research in this area, from 2014 to 2015, research focused primarily on building its foundational framework, the "system". Research during this period laid the groundwork for subsequent technology integration and application, emphasizing the importance of systems thinking in financial management transformation. Subsequently, research began to delve into more complex models and systems, exploring how information technology could be used to improve financial management. Researchers and scholars began to realize that achieving and improving efficiency and accuracy in financial management through integrated systems is a more important direction for the future, especially in terms of automation and real-time data processing.

2016-2018: Development of modeling systems

As we move into 2016, research is beginning to shift to more complex models and systems as understanding of the underlying frameworks grows. Researchers and scholars worked to develop and test a wide range of financial models that improve the accuracy of decision-making through algorithms and data analysis. Research during this period began to involve more advanced mathematical and statistical methods, as well as emerging information technologies such as big data analytics and cloud computing [35]. Research in integrated systems began to show its potential in financial management, especially in automation and real-time data processing. These researches have pushed financial management to higher levels of automation and intelligence, providing faster and more accurate financial reporting and analysis for organizations.

2018-2020: Behavioral patterns of concern

Beginning in 2018, research shifted to "behavior," or patterns of behavior in financial decision-making. This change in research direction reflects a deeper understanding of the factors that influence individual and organizational behavior, and how this understanding can be used to optimize financial strategies. And how these understandings can be used to optimize financial strategies. Researchers and scholars are beginning to explore how human behavior interacts with financial management systems and how to design intelligent systems that are more responsive to user behavior. Research in this phase not only focuses on the technical aspects of the system but also begins to focus on the human factor and how understanding the behavior of financial decision-makers can improve the design of the system and the quality of decision-making.

2020-2021: Personalized recommendations and risk

By 2020, the focus of research has shifted to "recommender systems" and "risk management". Recommender systems research focuses on how machine learning algorithms can be used to provide personalized financial advice, while risk management research focuses on how to identify and mitigate financial risks, especially in a market environment of increasing uncertainty. During the same period, the role of the Internet became increasingly important in the field. Research has begun

to explore how Internet technologies, including cloud computing and big data analytics, can be utilized to improve the efficiency and effectiveness of financial management so that investors and financial decision-makers can make more informed investment choices in a complex market environment. These studies emphasize the critical role of the Internet in connecting financial decision-makers, providing real-time market information, and supporting complex calculations, advancing the field's use in risk control and decision support to provide companies with more robust financial strategies.

2021-2023: Deepening the role of the Internet

Since 2021, the Internet has become increasingly important in the practical application of financial management. Research is beginning to explore how Internet technologies, including cloud computing and big data analytics, can be utilized to improve the efficiency and effectiveness of financial management. These studies emphasized the critical role of the Internet in connecting financial decision-makers, providing real-time market information, and supporting complex calculations. As Internet technologies evolve, financial management is beginning to migrate to the cloud, leveraging the Internet's extensive connectivity and powerful data processing capabilities to provide organizations with more flexible and efficient financial management solutions.

During the same period, some researchers and scholars began to turn to the in-depth exploration of the shared vision and the application of technology. The research hotspots not only focus on the deepening application of technology but also the innovation of big data-driven decision support and financial sharing models. With the advancement of technology, the sharing concept and the idea of intelligent financial management have been generally recognized, and enterprises have gradually popularized the adoption of advanced management systems, the level of intelligence has steadily increased, and the number of technological applications has increased by nearly 10% on average, and the rate of technological adoption has also increased. Financial sharing has become a key factor in promoting the level of construction in this field, and most enterprises have promoted the goals of financial standardization, improving the level of financial intelligence, and facilitating the transformation of financial functions through financial sharing centers. At the same time, the widespread application of mature technologies such as electronic invoices, mobile Internet, mobile payments, and digital signatures has accelerated the transformation of smart financial construction.

It should be noted that the construction of the system in this field is affected by the industry category and enterprise scale, and the significance of differences is amplified. Different enterprises have different motivations and demands for the construction of smart finance due to the existence of differences in industry, revenue scale, personnel scale, and informatization level, which leads to large differences in the application of artificial intelligence and other technologies and the level of intelligence. Under the shared vision, the study began to focus on cutting-edge trends such as data governance and smart sharing. Data governance ensures the quality and security of data and provides a reliable basis for corporate decision-making, while smart sharing emphasizes the importance of achieving information sharing and collaborative cooperation in the digital era, thereby enhancing the competitiveness and innovation of enterprises.

D. Projections of Future Trends

Research in the area of smart financial management will continue to develop along the path of technology convergence and innovation. As industry applications deepen, research in the field will become more relevant to practical needs, and environmental sustainability concerns will continue to grow. Interdisciplinary research will continue to grow, and innovations in research methodologies will continue to emerge, bringing new perspectives and solutions to the field. In the process, research scholars will continue to explore the application of technologies such as big data, artificial intelligence, and blockchain in financial management, and will also pay more attention to how these technologies affect financial decision-making, risk management, and corporate value creation. As technology continues to advance, research in this field will continue to expand into new areas, providing practitioners with more theoretical support and practical guidance.

In addition, with the continuous development of globalization and market economy, research in this field will also pay more attention to financial management issues in multinational corporations and multi-currency environments. Researchers and scholars will explore how intelligent financial management tools can be used to optimize the financial structure of firms and improve their international competitiveness in a changing global economic environment. Research in this area will also focus more on sustainability and social impact. With society's increasing emphasis on environmental protection and social responsibility, companies need to consider more environmental and social factors in their financial management. Researchers and scholars will explore how to promote the sustainable development of enterprises through intelligent management systems and realize the dual goals of economic and social benefits.

As technology continues to advance, research in this field will continue to expand into new areas, such as the application of emerging technologies such as the Internet of Things (IoT), Artificial Intelligence (AI), and Blockchain in financial management.

In conclusion, research on smart financial management has made significant progress over the past decade, and future research will continue to evolve along the path of technological convergence and innovation to provide enterprises with more efficient and smarter financial management solutions. As research continues, the field will play an increasingly important role in the global economy, helping companies achieve more efficient and sustainable growth.

V. CONCLUSION

In the field of intelligent financial management, this paper, through CiteSpace6.3.R1 software, systematically combed the research dynamics during the 10 years from 2014 to 2023, and analyzed the development trend of the research field from different levels and perspectives by doing the distribution of the literature, the key journals loaded with publications, the network mapping of the collaborators, the keyword clustering mapping,

as well as the timeline, and through the in-depth understanding of the content, predicted the possible future research hotspots and research directions.

In the data statistics on the number of publications in the literature and the number of issues contained in the key journals, it is observed that this research field has shown a clear stage characteristic in the past 10 years. At the beginning of the research, it focused on the construction of the basic framework, followed by a gradual overshoot into the transforming direction of behavioral patterns, recommender systems, risk management, and the application of Internet technologies. Advances in the technology of Internet big data and the development of big data, artificial intelligence, blockchain, and other technologies have provided new impetus and direction for the field while deepening research scholars' deeper understanding and exploration of the field.

Although this paper provides a comprehensive analysis of the research dynamics in the field, there are some limitations. First, it may fail to cover all relevant research in the field as the collection of research data is limited to a specific period. Second, the analysis in this paper is mainly based on bibliometric methods, which may fail to fully capture the depth and quality of the research content. The analysis of smart financial management practices characterized by different regions and industries needs to be further deepened.

Future research can dig deeper into the following aspects. First, with the continuous progress of technology, the application of emerging technologies such as the Internet of Things and 5G communication in smart financial management deserves further research. Second, the deepening of interdisciplinary research, such as combining psychological and sociological theories with smart financial management, may reveal more about the deeper factors of financial decision-making behavior. In addition, case studies of smart financial management practices in enterprises of different sizes and industries may provide richer empirical data to help understand the applicability and effectiveness of smart financial management in different environments. Finally, with the development of globalization, the application of smart financial management in multinational corporations and its impact on the global financial market is also a research direction that deserves attention.

In summary, research in the field of smart financial management under the shared vision is in a stage of rapid development, and future research will show greater potential and value in terms of technology integration, practical application, and interdisciplinary exploration. With the continuous deepening of the research work and the accumulation of practice, it is expected to bring more efficient and smarter financial management strategies to enterprises, and thus promote the sustainable and healthy development of enterprises and the entire economic system. In the subsequent research, the author plans to expand the scope of the sample database, closely track the latest research progress and results in this field, and think deeply about how to promote the development of intelligent financial management under the guidance of the sharing concept, to provide guidance and direction for the wide application of this field.

REFERENCES

[1] Sengupta A, Jana P, Dutta P N, et al. Optimal stock allocation for an automated portfolio recommender system in the perspective of maximum fund utilization[J]. Expert Systems with Applications, 2024, 242: 122857.

[2] Aboelmaged, M., Alhashmi, S. M., Hashem, G., Battour, M., Ahmad, I., & Ali, I.. Unveiling the path to sustainability: Two decades of knowledge management in the sustainable supply chain – a scientometric analysis and visualization journey[J]. Benchmarking: An International Journal, 2023, ahead-of-print(ahead-of-print).

[3] Hu X, Kang S, Ren L, et al. Interactive preference analysis: a reinforcement learning framework[J]. European Journal of Operational Research, 2024, 319(3): 983-998.

[4] Krause, J., Myroshnychenko, I., Tiutiunyk, S., & Latysh, D. Financial Instruments of the Green Energy Transition: Research Landscape Analysis[J]. Financial Markets, Institutions and Risks, 2024 8(2), 198–212.

[5] Leung M F, Jawaid A, Ip S W, et al. A portfolio recommendation system based on machine learning and big data analytics[J]. Data Science in Finance and Economics, 2023, 3(2): 152-165.

[6] Liu, X., Chau, K.-Y., Liu, X., & Wan, Y. The Progress of Smart Elderly Care Research: A Scientometric Analysis Based on CNKI and WOS[J]. International Journal of Environmental Research and Public Health, 2023 20(2), Article 2.

[7] Javadi S, Hashemi S M. Designing a Risk-Aware Recommender System for Providing Personalized Optimal Stock Portfolios: the Case of S&P 500[J]. Available at SSRN 4916445.

[8] Ma, T., Liu, Y., & Han, M. Visualization Analysis of Organizational Resilience Research Based on CiteSpace From 1990–2022[J]. IEEE Access, 2022 10(10), 65854–65872. IEEE Access.

[9] Yue X. Application of AI technology in a personalized recommendation system for financial services[J]. Applied Mathematics and Nonlinear Sciences, 2023, 9(1).

[10] Mathew, L., Govindan, V. M., Jayakumar, A., Unnikrishnan, U., & Jose, J. The evolution of financial technology: A comprehensive bibliometric review of robo-advisors[J]. Multidisciplinary Reviews, 2024 7(11), 2024274–2024274.

[11]]Jang J, Seong N Y. Deep reinforcement learning for stock portfolio optimization by connecting with modern portfolio theory[J]. Expert Systems with Applications, 2023, 218: 119556.

[12] Mikheev, A. V. Scientometric analysis of research trends and frontiers on global energy transition. AIP Conference Proceedings, 2023 2552(1), 080017.

[13] Asemi A. A Novel Combined Investment Recommender System Using Adaptive Neuro-Fuzzy Inference System [D]. Budapesti Corvinus Egyetem, 2023.

[14] Rong, K., Li, J., Huang, Q., & Zhu, G. Citespace-Based Analysis of System Resilience Research Hotspots and Trends[J]. 2022 12(11), 23–34.

[15] Umadevi B, Sundar D. Building Wealth in Stock Market: The Comprehensive Guide to Intelligent Portfolio Theory and Trading[J].

[16] Shen, Y., Huang, L., & Wu, X. Visualization analysis on the research topic and hotspot of online learning by using CiteSpace—Based on the Web of Science core collection (2004–2022) [J]. Frontiers in Psychology, 2022 13(34), 12–22.

[17] Asemi A, Asemi A, Ko A. Investment Recommender System Model Based on the Potential Investors' Key Decision Factors[J]. Big Data, 2023.

[18] Su, X., Wang, S., & Yu, R. A bibliometric analysis of blockchain development in industrial digital transformation using CiteSpace[J]. Peer-to-Peer Networking and Applications, 2024 17(2), 739–755.

[19] Ai H, Liu C, Lin P. Robust returns ranking prediction and portfolio optimization for M6[J]. International Journal of Forecasting, 2024.

[20] Sun, Y., Ma, Z., Chi, X., Duan, J., Li, M., & Khan, A. U. Decoding the Developmental Trajectory of Energy Trading in Power Markets through Bibliometric and Visual Analytics[J]. Energies, 2024 17(15), Article 15.

[21] Guan Z, Zhao Y. Optimizing stock market volatility predictions based on the SMVF-ANP approach[J]. International Review of Economics & Finance, 2024, 95: 103502.

[22] Wang, J., Wang, B., Jiang, X., Huang, Y., & Yang, J. Research on Hot Topics of Building and Developing Student Organizations—CiteSpace Based Visual Analysis[J]. Open Journal of Social Sciences, 2023 11(6), Article 6.

[23] Behera J, Pasayat A K, Behera H, et al. Prediction-based mean-value-at-risk portfolio optimization using machine learning regression algorithms for multi-national stock markets[J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105843.

[24] Yang, D., Wu, X., Liu, J., & Zhou, J. CiteSpace-based global science, technology, engineering, and mathematics education knowledge mapping analysis[J]. Frontiers in Psychology, 2023 13(13), 34–44.

[25] Cui T, Du N, Yang X, et al. Multi-period portfolio optimization using a deep reinforcement learning hyper-heuristic approach[J]. Technological Forecasting and Social Change, 2024, 198: 122944.

[26] Zhang, J., Wang, Q., Xia, Y., & Furuya, K. Knowledge Map of Spatial Planning and Sustainable Development: A Visual Analysis Using CiteSpace[J]. Land, 2022 11(3), Article 3.

[27] Li J, Zhang Y, Yang X, et al. Online portfolio management via deep reinforcement learning with high-frequency data[J]. Information Processing & Management, 2023, 60(3): 103247.

[28] Zhang, L. Environmental hotspots, frontiers and analytical framework of Blue Carbon research: A quantitative analysis of knowledge map based on CiteSpace[J]. Frontiers in Environmental Science, 2024 12(24), 131456.

[29] Ali Mohamad T, Bastone A, Bernhard F, et al. How artificial intelligence impacts the competitive position of healthcare organizations[J]. Journal of Organizational Change Management, 2023, 36(8): 49-70.

[30] Zhang, Y., Zhao, D., Liu, H., Huang, X., Deng, J., Jia, R., He, X., Tahir, M. N., & Lan, Y. Research hotspots and frontiers in agricultural multispectral technology: Bibliometrics and scientometrics analysis of the Web of Science[J]. Frontiers in Plant Science, 2022 13, 1–10.

[31] Ala A, Simic V, Pamucar D, et al. Enhancing patient information performance in the internet of things-based smart healthcare system: hybrid artificial intelligence and optimization approaches[J]. Engineering Applications of Artificial Intelligence, 2024, 131: 107889.

[32] Sahni N, Stein G, Zemmel R, et al. The potential impact of artificial intelligence on healthcare spending [R]. Cambridge, MA, USA: National Bureau of Economic Research, 2023.

[33] Kumar K, Kumar P, Deb D, et al. Artificial intelligence and machine learning based intervention in medical infrastructure: a review and future trends[C]//Healthcare. mdpi, 2023, 11(2): 207.

[34] Ahmed S K, Ali R M, Lashin M M, et al. Designing a new fast solution to control isolation rooms in hospitals depending on artificial intelligence decision[J]. Biomedical Signal Processing and Control, 2023, 79: 104100.

[35] Chen X, Xie H, Li Z, et al. Information fusion and artificial intelligence for smart healthcare: a bibliometric study[J]. Information Processing & Management, 2023, 60(1): 103113.

Explainable AI-Driven Chatbot System for Heart Disease Prediction Using Machine Learning

Salman Muneer[1], Taher M. Ghazal[2]*, Tahir Alyas[3],
Muhammad Ahsan Raza[4], Sagheer Abbas[5]*, Omar AlZoubi[6], Oualid Ali[7]
School of Computer Science, National College of Business Administration & Economics (NCBA&E), Lahore, Pakistan[1]
Center for Cyber Security-Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia (UKM), 43600, Bangi, Selangor, Malaysia[2]
Department of Computer Science, Lahore Garrison University, Pakistan[3]
Department of Information Sciences, University of Education, Lahore, Multan Campus 60000, Pakistan[4]
Department of Computer Science, Prince Mohammad Bin Fahd University, Alkhobar, KSA[5]
College of Arts & Science, Applied Science University, P.O.Box 5055, Manama, Kingdom of Bahrain[6, 7]

***Abstract*—Heart disease (HD) continues to rank as the top cause of morbidity and mortality worldwide, prompting the enormous importance of correct prediction for effective intervention and prevention strategies. The proposed research involves developing a novel explainable AI (XAI)-driven chatbot system for HD prediction, combined with cutting-edge machine learning (ML) algorithms and advanced XAI techniques. This research work highlights different approaches like Random Forest (RF), Decision Tree (DT), and Bagging-Quantum Support Vector Classifier (QSVC). The RF approach achieves the best performance, with 92.00% accuracy, 91.97% sensitivity, 56.81% specificity, 8.00% miss rate, and 99.93% precision compared to other approaches. SHAP and LIME provide XAI methods for which the chatbot's predictions and explanations endow trust and understanding with the user. This novel approach proves the potential of seamless integration of explanations in a wide range of web or mobile applications for healthcare. Future works will extend the work on incorporating other diseases' predictions in the model and improve the explanation of those predictions using more advanced explainable AI approaches.**

***Keywords*—Heart disease prediction; machine learning; chatbot system; XAI**

I. Introduction

The rapid development of artificial intelligence (AI) [1] and ML [2] technologies has made various fields, including healthcare [3], solutions to complex medical problems possible. In this regard, HD stands as an important public health issue among the leading causes responsible for mortality across the globe. Therefore, early detection and prevention of HD is important to curb the materialized global impact. AI-enforced solutions are fascinating tools for achieving such goals, with chatbots being a prominent example. Chatbots have emerged as a novel interface for healthcare applications by using conversational AI to communicate with users through real-time conversations and gather valuable health information. These intelligent agents driven by ML models can assess risk factors, reply to questions, and provide health predictions based on user inputs. The promise of Chatbots is most evident in heart disease prediction as they act as virtual assistants to surface early red flags and motivate preventive actions. However, a significant concern surrounding the implementation of these chatbots in life-critical health scenarios is their ability to elucidate their predictions in a way that is transparent and comprehensible to end users. The application of chatbots, empowered by ML techniques, ensures a broader range of capabilities for an interactive conversation system.

Traditional ML models [4] have proven efficient predictors of clinical scenarios, they remain largely "black boxes" whose inner workings are often not transparent enough to fully account for how certain decisions are arrived at. Such opacity may generate doubts regarding trust and reliability because healthcare demands insight into the reasoning behind a prediction. To avert these challenges, XAI [5] has been fashioned to augment the interpretability of the decision-making processes involved in these models. By providing greater insight into variables predicting heart disease, XAI helps in decision-making by healthcare practitioners and parties involved thus engendering trust in AI solutions.

Explainable AI helps in this regard by allowing chatbots not only to predict HD risk but also to explain the reasons behind their predictions understandably. In contrast with the normal "black box" ML model for which decisions are explained in contextual terms, XAI-driven chatbots such as this can potentially explain why some attributes (blood pressure, cholesterol levels, etc.) lead to a simultaneously unique heart disease risk given all patient profiles: age ranges and themselves's lifestyle choices. Transparency will be anessential aspect of healthcare. It builds trust and clarifies the rationales of AI recommendations to patients as well as providers. XAI-based chatbots [6] for heart disease prediction [7] combine strengths from both conversational agents and sophisticated predictive models. These chatbots aim to interact with users in dialogue, collecting the required health information and providing instant risk estimates. The chatbot incorporates XAI methods to explain how some health measurements, lifestyle habits, or risks in custom history contribute to the overall risk profile, thus providing a transparent vision of the prediction process. These together not only enhance user experience but also make sure that the predictions remain interpretable and actionable. Moreover, the drifting of implementing XAI into HD prediction

*Corresponding Author.

models coincides with the growing concern in ethical AI, and the necessity of fair systems [8], accountability [9], and transparency [10]. While such qualities are critical in every domain, they gain special significance in healthcare since they may directly affect patient outcomes as well as trust in AI-driven recommendations. XAI can allow clinicians to verify model predictions, see where biases occur, and make adjustments, ensuring results are fair and accurate across patient groups via the chatbot's suggestions. In turn, this works to encourage patients to adopt the recommended lifestyle changes or prompts for further medical evaluation.

The usage of XAI with chatbots for HD prediction [11], the primary advantages would be to improve patient education and engagement. It helps them better understand why and how age, smoking behavior, activity level, or dietary habits could increase the risk of HD. This understanding empowers people to take charge of their health and care and achieve better outcomes overall. It not only helps diagnose [12] but also educates the patient about their condition, acting as a bridge between complicated medical information and a better understanding of patients.

The practice of such systems should be done carefully, with full consideration given to technical and clinical issues. From the technical aspects, ML algorithms used, feature selection, and model training are key to determine what will be predicted by the bot that should have high accuracy with ground truth. For the interpretability of ML models, there are different techniques like SHAP (Shapley Additive Explanations) [13] and LIME (Local Interpretable Model-Agnostic Explanations) [14]. These techniques help you identify the roles different input features play in your final prediction, and so your model decision is more scrutable. From the clinical point-of-view, it is important that this AI chatbot be tested with patient-generated data, and working closely with doctors helps to mature its design. Working alongside cardiologists, GPs and other practitioners should allow the dialog responses of this chatbot as well to help prediction models better adapt in focusing on specific patients at risk of HD. Two trade-offs yet to be considered are the model complexity and interpretation ability when the work starts deploying XAI-based chatbot service for HD prediction. Although more complex models (e.g., deep neural networks) may provide greater accuracy, they are also often less interpretable than simpler ones such as decision trees or logistic regression. Finding the right balance between accuracy and interpretability is crucial to ensure that the chatbot delivers reliable predictions while maintaining transparency. Researchers are dynamically discovering hybrid techniques that integrate the strengths of various models to obtain optimal performance.

Many studies have explored ML and Deep Learning (DL) approaches for HD prediction, but they often lack transparency, limiting their practical adoption in healthcare. The novelty of this study lies in introducing an XAI-driven chatbot system that combines accurate predictions with interpretable outputs. By focusing on transparency and user-friendly design, it bridges the gap between complex AI models and practical healthcare use, addressing the "black-box" limitations of traditional methods. This research work proposes the design of XAI chatbots for HD prediction which marks an important sign of leveraging AI for healthcare. With ML's powerful predictive features and XAI helping in bringing transparency, now healthcare systems can benefit from accurate predictions as well as making these models trustable for both patients and Healthcare Providers. Even more, improvements are planned as the field continues to grow, making these chatbots even more useful when it comes to battling HD and any other health sicknesses that last a lifetime.

The remainder of this paper is organized as follows: Section II discusses the literature review, highlighting the previous work regarding chatbot systems for HD prediction, and Section III provides limitations of previous works. Section IV discusses the proposed methodology, Section V highlights the dataset description, Section VI explains the dataset structure, Section VII explains the system block diagram, Section VIII describes pseudo code, Section IX simulation results, Section X explains the Discussion, Section XI presents conclusion of the manuscript and Section XII describes the limitations and future suggestions. This structure ensures a comprehensive understanding of the study and its contributions.

II. Literature Review

There have been extensive research interests pursued over the years on several autonomous systems, such as virtual assistants, chatbots, and applications for HD prediction employing XAI among other techniques. Their work has opened up ways for making predictive healthcare systems transparent, accurate, and ethical which provides a strong ground on top of where future advancements can be carried out. This section presents some of the most important research ever published in this field. AI-powered health chatbots are a game changer for how mental healthcare can be approached by businesses as well as any other sector including the traditional form of care within Healthcare. Powered by Natural Language Processing (NLP) and AI, these smart chatbots can converse through text as well as voice performing functions similar to those handled by a live human agent. The kinds of jobs they provide are many; everything from helping individuals in crisis to customer service. Smartphone and smart speakers-powered voice-enabled chatbots use high-end speech-to-text capabilities but are sensitive to elements like regional accents or background noise. Text-based chatbots, on the other hand, such as those found in Messenger or Slack that are available within web applications offer more managed interactions and allow for complex questions to be answered making them great tools for people looking for mental health support.

The authors highlighted that in recent years, chatbots have made remarkable progress in addressing various health-related issues, including obesity and weight management, dementia (e.g., Endurance, which engages in meaningful conversations with individuals with Alzheimer's disease), substance use disorders [15-16], oncology care, and insomnia (e.g., Casper, a chatbot designed to assist those with sleep difficulties). They have also been utilized in prenatal services, HIV and sexual health education, and managing depression and anxiety. Notably, chatbots have demonstrated significant potential in mental health applications, contributing to suicide prevention, virtual cognitive-behavioral therapy, and psychoeducation. This study aims to explore and summarize the role of chatbots in the healthcare industry.

The authors in study [17] presented that AI chatbots have been effective in promoting treatment adherence, smoking cessation, and healthier lifestyles, with features like goal setting, monitoring, and real-time feedback. Studies showed mixed results on feasibility, acceptability, and usability, but these chatbots offer personalized services and scalability via platforms like smartphones and Facebook Messenger. Participants valued the nonjudgmental space chatbots provided, especially for sensitive communications.

The authors in study [18] said the rise of AI-empowered chatbots has also brought new security and privacy risks. Cybercriminals are using social engineering tactics like phishing through cross-platform messaging apps to create a new kind of cyber threat called "Smishing". These fake chatbots look legit but are designed to trick people into giving away sensitive personal data. To combat these threats we need to increase user awareness and be cautious with unfamiliar services. Implementing security measures like using temp tokens during chatbot sessions is crucial when devices are lost or compromised. Also, data anonymization techniques like de-identification, pseudonymization, and generalization should be used to protect sensitive data from unauthorized access. By doing all these organizations can fortify their defenses against privacy breaches and security issues with malicious chatbots.

The research also shows the increasing use of chatbots in healthcare settings to improve service and efficiency. Examples include Healthily [19] and Ada Health [20] which give users health advice and are now health information dissemination tools. In mental health, specialized chatbots like Woebot [21] are being used to deliver Cognitive Behavioral Therapy (CBT) to people with depression, PTSD, and anxiety. They also help patients with autism to learn and practice social skills as part of therapy. Beyond being therapy tools, chatbots also help with administrative tasks in healthcare facilities such as allowing patients to self-book appointments and track doctor availability. This integration in healthcare systems improves patient access and management and thus better healthcare delivery. Chatbots can also collect and process information from patients through structured questions about identity and health conditions. This information is used for patient admission, symptom monitoring, patient-doctor interaction, and documentation. Chatbots can also help with medication reminders for conditions that don't require in-person consultations. The healthcare application of chatbots not only enhances the level of patient experience and outcome but also maximizes organizational efficiency, hence ensuring that such technologies stay increasingly relevant in modern healthcare environments.

Another one of the chatbots is Gyant [22], designed to interact with users having minor non-emergency medical issues by asking questions regarding symptoms as well as general health status. An important feature of Gyant is its use of humor, human-like interaction, emojis, and memes to cope with the user successfully. However, the software is currently not available since it has been phased out. For example, Symptomate [23] is an AI chatbot, that allows patients to report their symptoms and provides them with a list of possible conditions. However, its accuracy is significantly lower than 80% compared to Ada Health. Nonetheless, this proves the fact that digital health solutions, in terms of their development and perfection, will be needed for maximum efficacy and reliability [24].

III. Limitations of Previous Work

Despite the studies in AI-powered chatbots that predict HD, it is clear that some of the previous attempts have significant disabilities. For instance, there is a severe lack of XAI, which usually leaves users and medical professionals unaware of the processes that govern the decisions they make when using a system. Moreover, most of these systems fail with issues of data privacy while dealing with delicate health information as they tend to leave their users vulnerable to breaches or unauthorized access. Another major problem is scalability as most models do not deliver optimally in various healthcare settings or with other patient populations. Moreover, the reliability of the decision-making functionality of these chatbots is incomplete, and conclusions may be as wrong as they are less encompassing. Last but not least, the problems of transparency continue to worsen the mistrust in these systems since the user does not know whether to believe the answers provided by the system. Collectively, these limitations underwrite the need for better methodologies that place more emphasis on user safety, trust, and clarity when using AI technologies in healthcare settings. A few of the limitations are shown in Table I:

TABLE I. Limitations of Previous Works

Ref.	Year	Disease Dataset	ML Algorithm/Decision Support System	XAI Implementation	Chatbot enabled System
[25]	2024	HD	DT	×	✓
[26]	2024	HD	Clinical Decision Support (CDS) system	×	✓
[27]	2023	Prostate Cancer	Hybrid Approach with NLP models (Named Entity Recognition, Intent Recognition, Sentiment Analysis, and Language Detection)	×	✓
[28]	2023	HD	XGBoost	✓	×
[29]	2023	HD	Bagging-QSVC	✓	×
[30]	2023	HD	Random Forest (RF)	×	×
[31]	2022	HD	Cox models plus symbolic regression	×	×
	2024	HD	RF	✓	✓

Table I summarizes the critical analysis of inherent limitations within previous works on the application of ML algorithms and decision support systems, particularly towards targeting HD, the prediction of prostate cancer. Much work never includes XAI techniques in their strategies, but this is a fundamental aspect of improving the transparency and trustworthiness in automated systems. For instance, although some models may lead to successful predictions of health-based outcomes, the inability to interpret may deter healthcare professionals from accepting this system for use in their service provision because they would want to understand the patterns guiding the prediction. Finally, while many works include chatbot functionalities to make interaction easier for users, it usually occurs in isolation from XAI and leads to systems that are user-friendly but less clear about their operational mechanisms. This is a disadvantage when it comes to user confidence, especially in more sensitive fields like healthcare. The deployment of advanced NLP techniques has only been moderately applied in some research works, thereby limiting the capabilities of these systems to better interact with users. Interactions may therefore not holistically deal with the intricacies of patient queries. In summary, these constraints call for further research to integrate both XAI and chatbot technologies with a focus on the construction of more solid, transparent, and user-centric healthcare solutions.

IV. Proposed Methodology

Technological advancement has intended the adoption of autonomous systems in most fields, with the application here being one of the most popular emerging applications of chatbots. Chatbots are becoming very popular, especially in the health sector, as dear guides for patients and health-related questions. However, these chatbots have encountered numerous problems such as issues related to patient medical records, insecure communication, and inability to provide clear and precise answers.

Knowing the problems calls for a rising demand to develop intelligent approaches [32-33] that would yield better results. Thus, this research work aims to provide an intelligent approach to develop a chatbot system that uses XAI. It sets out to mitigate prevailing challenges around chatbots in general with special emphasis on the healthcare industry to ensure that secure communications are implemented while providing clear responses and retrieving accurate information from patient records. The proposed chatbot system is shown in Fig. 1.

Fig. 1 represents the proposed approach, which comprises the training and validation phases. During the training phase, initially, patient data is acquired from the patient through a chatbot for secure and tamper-proof transactions, thereby enhancing data integrity and transparency within healthcare systems. The tamper-proof transactions are then forwarded to the preprocessing layer, involving normalization, handling missing values, and moving averages, and the processed data is then divided into training and testing sets, with respective ratios of 80% and 20%. Subsequently, the approach is trained on 80 percent of the data for predictive analysis using ML algorithm RF. The predictions are directed to XAI for comprehensive output explanations. If the output aligns with the predefined learning criteria, the results are stored in the cloud; otherwise, they are returned to the approach if the learning rate is not achieved.

In the validation phase, patient data is directly compared with the imported data stored in the cloud. If the criteria are met, indicating the presence of HD, the system is shown as "found". On the other hand, if the criteria are not met, signifying the absence of HD, the system is discarded.

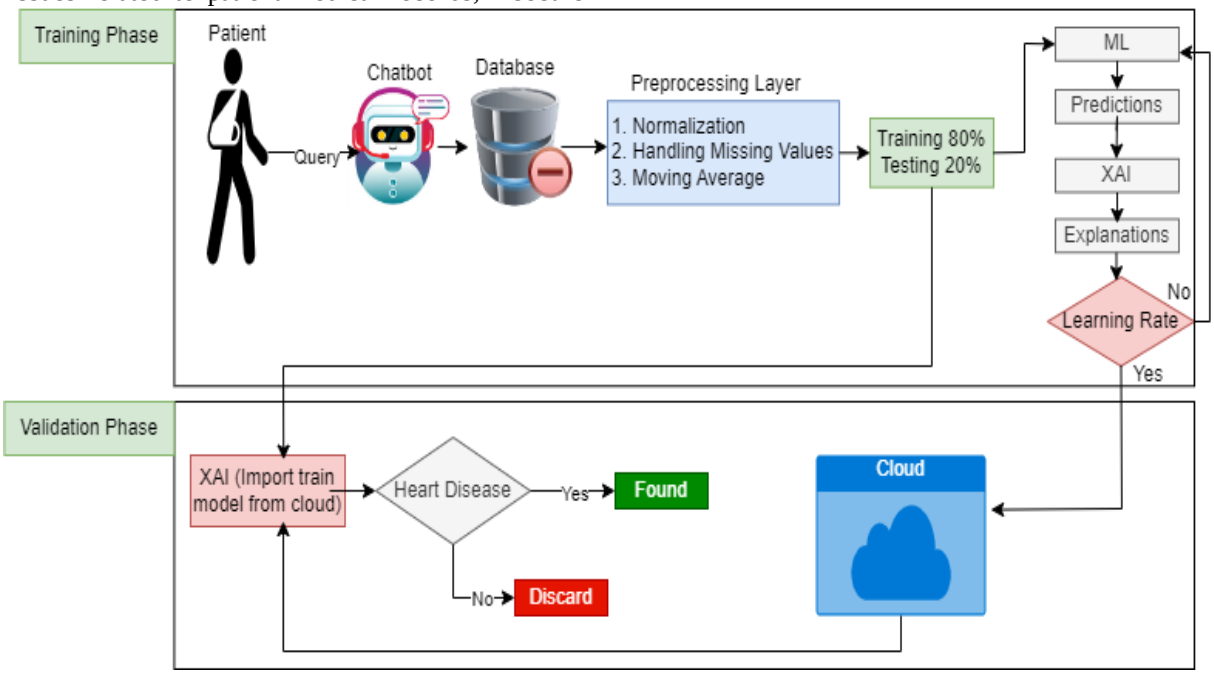


Fig. 1. Proposed chatbot system.

V. Dataset Description

The dataset utilized in this study comprises 308,855 entries & 19 features [34], providing a comprehensive overview of various health-related parameters. It encompasses features such as General Health, which is rated on a scale from 1 to 5 (with 5 indicating excellent health and 1 indicating poor health), recent Checkups (within the previous year), where 0 denotes no checkup, 1 represents a checkup within the last year, 2 for two years ago, 3 for a checkup within the last five years, and 5 for checkups more than five years ago. Exercise is captured through exercise habits, but supplemented by the prevalence of chronic diseases including HD, Skin Cancer, Other Cancers, Depression, Diabetes, and Arthritis by binary indicators where 1 means presence of the condition, and 0 means absence. Demographic information is also captured, in the Sex column where 1 means female, and 0 means male. Other determinants include stature in centimeters and weight in kilograms, which are necessary parameters in the BMI. Characteristics of lifestyle such as smoking and alcohol intake were taken into consideration, and dietary intake was reflected in fruit consumption, green vegetable consumption, and fried potato consumption. This length of variable allows for a very sharp analysis of determinants, which can thereby assess complex interrelations between lifestyle, demographic factors, and health outcomes in the prediction of heart disease with AI-driven chatbots.

VI. Dataset Structure

The Table II below reports the most important characteristics used in the dataset for health-related analysis, categorized by type of data. Such a collection of datasets contains health-related features pertaining to lifestyle habits, medical history, and demographic information. Qualitative attributes-the ones describing, for instance, "General Health," "Exercise," and "Depression"-represent statuses or histories related to a patient's health. Quantitative features refer to features whose variables include such examples as "Height (cm)" and "Weight (kg)". Attributes such as these are represented using numerical data types. They are fundamentally indispensable for developing predictive models for health outcomes, providing a complete view of someone's health profile.

TABLE II. Dataset Features with its Datatype

Feature Name	Data Type
General_Health	Object
Checkup-(Within_Previous_Year)	Object
Exercise	Object
Heart_Disease	Object
Skin_Cancer	Object
Other_Cancer	Object
Depression	Object
Diabetes	Object
Arthritis	Object
Sex	Object
Age_Category	Object
Height_(cm)	Float64
Weight_(kg)	Float64
BMI	Object
Smoking_History	Object
Alcohol_Consumption	Object
Fruit_Consumption	Object
Green_Vegetables_Consumption	Object
FriedPotato_Consumption	Object

It can be seen in Table II that the predictors of medical history features and the patterns of behavior were the features necessary to form an opinion relating to risks regarding health conditions. For example, "Heart Disease," "Skin Cancer," and "Other Cancer" revealed whether any patient was diagnosed with the said diseases, while "Smoking History" and "Alcohol Consumption" reflected lifestyle decisions likely to lead to various health conditions. The indicators available in this dataset are the intake of nutrition, like "Fruit Consumption" and "Green Vegetables Consumption" to assess the dietary pattern. By integrating all these different features, the dataset offers a complete foundation to analyze health patterns and to predict the possibility of getting ill. It could be a beneficial resource in healthcare research and study.

VII. System Block Diagram

This research work proposed a system that uses historical data for prediction, as shown in Fig. 2. It applies EDA to identify whether there is a need for pre-processing and whether it identifies outliers. Pre-processing includes handling null values, duplicate values, outliers, and class imbalance. Then, the data sets are split into test and training data where 20% is designated for the test set and 80% for the training set. The model trains and tests on those datasets, and then the accuracy, precision, recall, f1-score, and confusion matrix are used to achieve the best model. Then, using the chosen model, accurate prediction about the outcome with an explanation of LIME is applied.

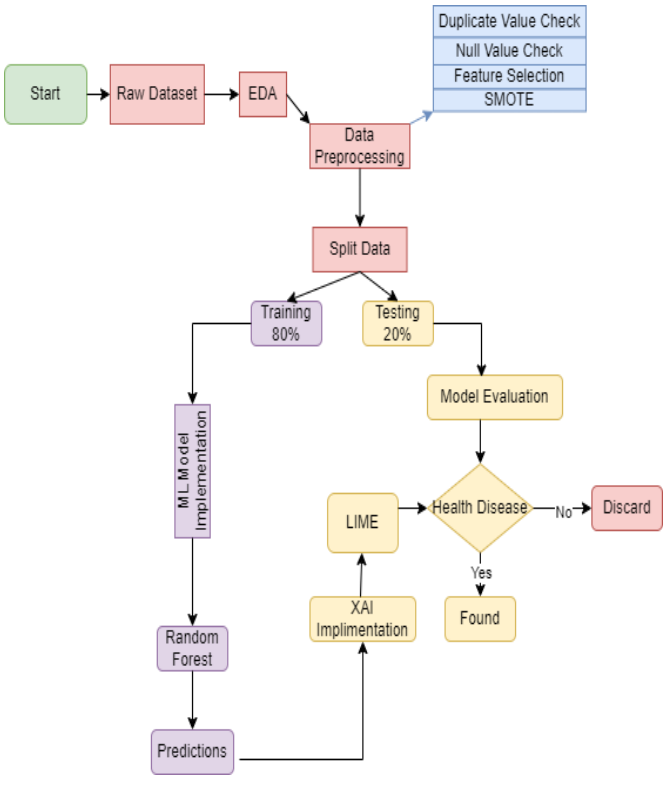


Fig. 2. System block diagram.

Fig. 2 shows that the research work illustrates an entire workflow in a system block diagram for a healthcare prediction

model that involves the integration of ML and XAI methodologies. Starting from the bottom, the raw healthcare data is gathered, followed by EDA to explore the dataset and draw meaningful conclusions based on it. The data would then be subjected to thorough preprocessing with checking for duplicate and null values, selection of best features, and balancing with SMOTE. With preprocessing, the dataset was divided into the training and the test sets. For this example, the training set was used to construct the Random Forest model, while the testing set was reserved to analyze its performance. The output from the model is then analyzed to determine whether any disease is present. If a disease is identified, it is further passed over the XAI method LIME [35] to give a clear explanation of which features contributed to the prediction. This approach enhances the interpretability and transparency of the model for healthcare professionals, such that they can trust the outcomes of the model more easily. The system ends by flagging some cases where a health condition is detected or discarding cases where no disease is found and then sending it out as a responsible healthcare delivery process.

VIII. Pseudo Code

The step-by-step pseudo code for the prediction of HD is shown below in Table III.

TABLE III. Pseudo Code

START
Step 1: Data Loading
Load dataset from source
Step 2: Data Preprocessing
Perform exploratory data analysis (EDA)
Remove duplicates and check for null values
Handle missing values if any
Feature selection based on relevance
Apply SMOTE for handling class imbalance (if needed)
Normalize/Standardize numerical features like "Height" and "Weight"
Step 3: Splitting Data
Split data into Training set (80%) and Testing set (20%)
Step 4: Model Implementation
Select a machine learning model (e.g., Random Forest)
Train the model on the Training set
Evaluate the model using the Testing set
Step 5: Model Evaluation
Calculate performance metrics (e.g., Accuracy, Precision, Recall)
Check if the model detects 'Health Disease' or 'No Disease'
IF model performance is satisfactory THEN
Proceed to explainable AI (XAI) implementation
ELSE
Modify model or preprocessing steps and retrain
Step 6: XAI Implementation (e.g., LIME)
Generate explanations for model predictions using XAI techniques
Display or store the explanations
Step 7: Make Predictions
Input new data into the trained model
Predict whether the patient has 'Health Disease' or 'No Disease'
Display predictions and explanations
Step 8: Decision and Action
IF 'Health Disease' is predicted THEN
Recommend further medical consultation
ELSE
No further action required.
END

As presented in Table III, pseudo-code illustrates a systematic approach to analyzing healthcare data and making predictions regarding health conditions. It begins with loading the dataset and running preprocessing steps as follows: exploratory data analysis, and missing value handling, and it applies SMOTE to the dataset for class balance. After this step is performed, the data set is split into a training and a testing set to permit the training of an ML algorithm chosen by the researcher. After evaluating the model's performance, the code incorporates explainable AI methods to enhance the interpretability of predictions, ultimately guiding healthcare decisions based on the results.

IX. Simulation Results

This research work proposed a chatbot system for HD prediction using the Random Forest (RF) approach and implemented it on the dataset containing 308855 samples. The data were distributed into 80 % training (247084 samples) and 20% validation (61771 samples). As stated in the equations, this approach finds the result using multiple statistical measures.

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

$$Sensitivity\ (TPR) = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity\ (TNR) = \frac{TN}{TN+FP} \quad (3)$$

$$Miss\ Rate\ (FNR) = 1 - Acc \quad (4)$$

$$Fall\ out\ (FPR) = \frac{FP}{FP+TN} \quad (5)$$

$$LR + ive\ (LR+) = \frac{TPR}{FPR} \quad (6)$$

$$LR - ive\ (LR-) = \frac{FNR}{TNR} \quad (7)$$

$$Precision\ or\ Positive\ Predictive\ Value\ (PPV) = \frac{TP}{TP+FP} \quad (8)$$

$$Negative\ Predictive\ Value\ (NPV) = \frac{TN}{TN+FN} \quad (9)$$

TABLE IV. Chatbot System Training Phase for HD Prediction using RF

	Total number of samples (49000)	**Result (Output)**	
Input	Expected output	Predicted Positive	Predicted Negative
		True Positive (TP)	False Positive (FP)
	229978 Positive	229750	228
		False Negative (FN)	True Negative (TN)
	17106 Negative	16806	300

It is shown in Table IV that the proposed chatbot system predicts the HD during the training period using RF. During training, 247084 samples are divided into 229978, 17106 positive, and negative samples. 229750 true positives are successfully forecasted, and no HD is recognized, but 228

records are mistakenly predicted as negatives, indicating the No Disease is recognized. Likewise, 17106 samples are obtained, with negative showing HD is identified and positive indicating No Disease. With 300 samples correctly identified as negative, showing the HD is recognized, and 16806 samples inaccurately foreseen as positive, representing No Disease is identified despite the presence of the HD.

It is shown in Table V that the proposed chatbot system predicts the HD during the validation period using RF. During validation, 61771 samples are divided into 56774, 4997 positive, and negative samples. 56736 true positives are successfully forecasted, and No Disease is recognized, but 38 records are mistakenly predicted as negatives, indicating the HD is recognized. Likewise, 4997 samples are obtained, with negative showing HD is identified and positive indicating No Disease. With 50 samples correctly identified as negative, showing the HD is recognized, and 4947 samples inaccurately foreseen as positive, representing No Disease is identified despite the presence of the HD.

TABLE V. CHATBOT SYSTEM VALIDATION PHASE FOR HD PREDICTION USING RF

	Total number of samples (49000)	**Result (Output)**	
Input	Expected output	Predicted Positive	Predicted Negative
		True Positive (TP)	False Positive (FP)
	56774 Positive	56736	38
		False Negative (FN)	True Negative (TN)
	4997 Negative	4947	50

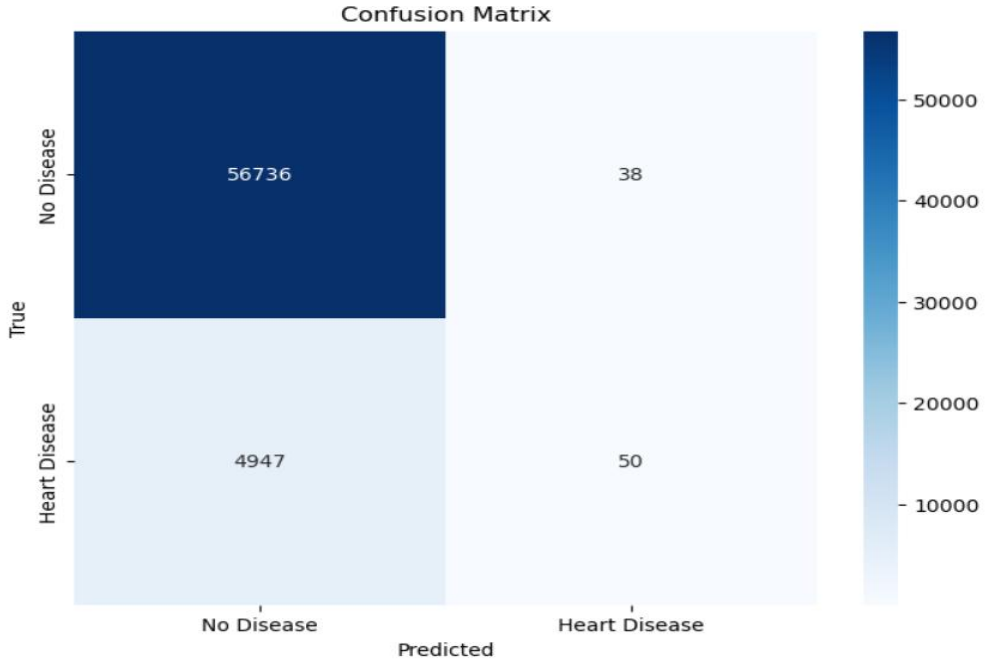


Fig. 3. Confusion matrix.

It is shown in fig. 3 that confusion matrix visually represents the performance of a system in predicting HD. The matrix shows that the system accurately identified 56,736 instances of HD (true positives) and misclassified 3,947 cases of HD as no disease (false negatives). It also identified 50 cases correctly as no disease (true negatives) but incorrectly flagged 38 as having HD when they did not (false positives). This matrix helps assess the balance between correct and incorrect predictions, indicating a higher rate of HD detection but with a considerable number of missed cases.

Fig. 4 shows that the ROC curve in the image evaluates the performance of a classification model by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). The curve shows a balanced trade-off between sensitivity (ability to detect positives) and the false alarm rate. The area under the ROC curve (AUC) is 0.83, indicating that the model performs well, but there is room for improvement. A perfect classifier would have an AUC of 1, meaning it distinguishes perfectly between positive and negative classes.

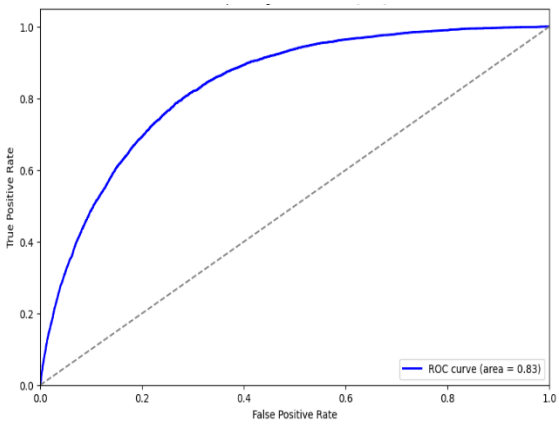


Fig. 4. Receiver operating characteristic (ROC) curve.

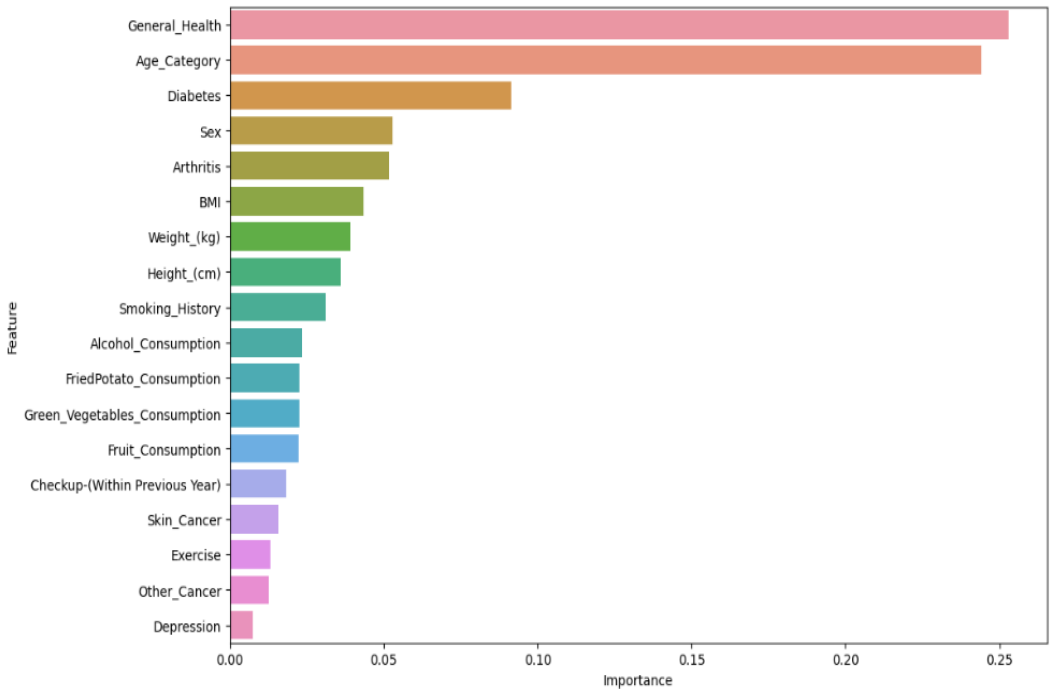


Fig. 5. Feature importance.

It is shown in Fig. 5 that the bar chart illustrates the feature importance ranking in a system, where each feature's contribution to the prediction outcome is displayed. General Health and Age Category are the most influential factors, contributing significantly to the system's predictions, followed by Diabetes and Sex. Features like Depression, Other Cancer, and Exercise have the least importance. The chart helps to identify which variables have the greatest impact, allowing for better interpretation and possible improvements in system design or understanding of the data.

TABLE VI. CHATBOT SYSTEM VALIDATION PHASE FOR HD PREDICTION USING RF

RF	Acc (%)	TPR (%)	TNR (%)	FNR (%)	FPR(%)	LR+	LR-	PPV (%)	NPV (%)
Training	93.10	93.18	56.82	6.90	0.43	2.16	0.1210	99.90	0.018
Validation	92.00	91.97	56.81	8.00	0.43	2.13	0.0014	99.93	0.010

Table VI shows that the Proposed chatbot system performance in terms of accuracy, sensitivity, specificity, miss rate, and precision during training using RF provides 93.10, 93.18, 56.82, 6.90, and 99.90, respectively. The suggested approach yields 92.00, 91.97, 56.81, 8.00, and 99.93 during the validation phase's accuracy, sensitivity, specificity, miss rate, and precision. Furthermore, the proposed chatbot system for HD prediction using RF approach yields 0.43, 2.16, 0.1210, and 0.018 in terms of fall-out likelihood positive ratio, likelihood negative ratio, and negative predictive value during training and 0.43, 2.13, 0.0014, 0.010 in terms of validation.

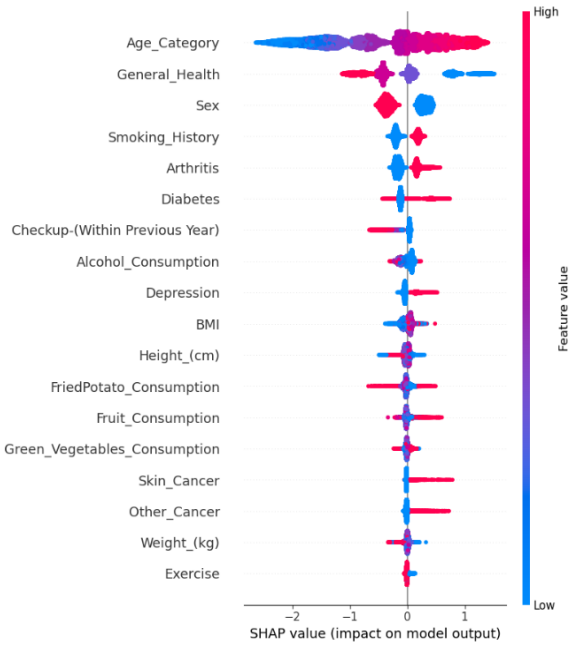


Fig. 6. SHAP value.

It is shown in fig. 6 that the SHAP (SHapley Additive exPlanations) plot highlights the impact of various features ona system's output. Each dot represents a single instance, with the position indicating the SHAP value, showing whether the feature drives the prediction towards a positive or negative outcome. Features like Age Category and General Health have the most influence, with higher feature values (shown in red) often leading to positive model outcomes. Conversely, low feature values (blue) push the predictions in the opposite direction. The plot helps explain how individual features contribute to the system's decisions.

Fig. 7 SHAP waterfall plot shows how individual features contribute to a single prediction. The base value E[f(x)]=−2.517 is the average prediction and features either increase (shown in red) or decrease (shown in blue) the predicted value. General Health has the most significant negative impact (-0.48), reducing the prediction, while Sex increases the prediction by +0.35. Other features such as Smoking History and Arthritis also contribute to the prediction, either positively or negatively. The final prediction value f(x)=−2.964 is the cumulative effect of all features.

Fig. 8 shows the prediction probabilities for a system evaluating HD risk. The system predicts a 94% probability for 'No Disease' and a 6% probability for HD. On the right, the feature contributions are split between the two outcomes: features like General Health (value = 4) and Age Category (value = 8) push the prediction towards HD, while features like Diabetes and Smoking History push towards no disease. The final prediction reflects the cumulative influence of all these factors on the system's output.

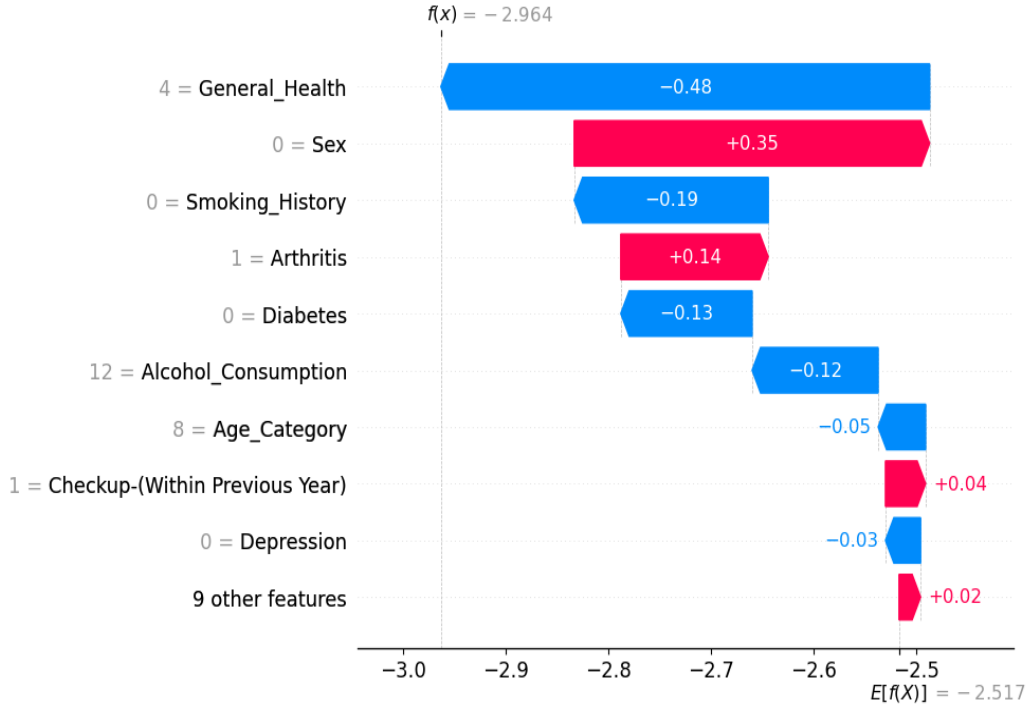


Fig. 7. SHAP waterfall impact.

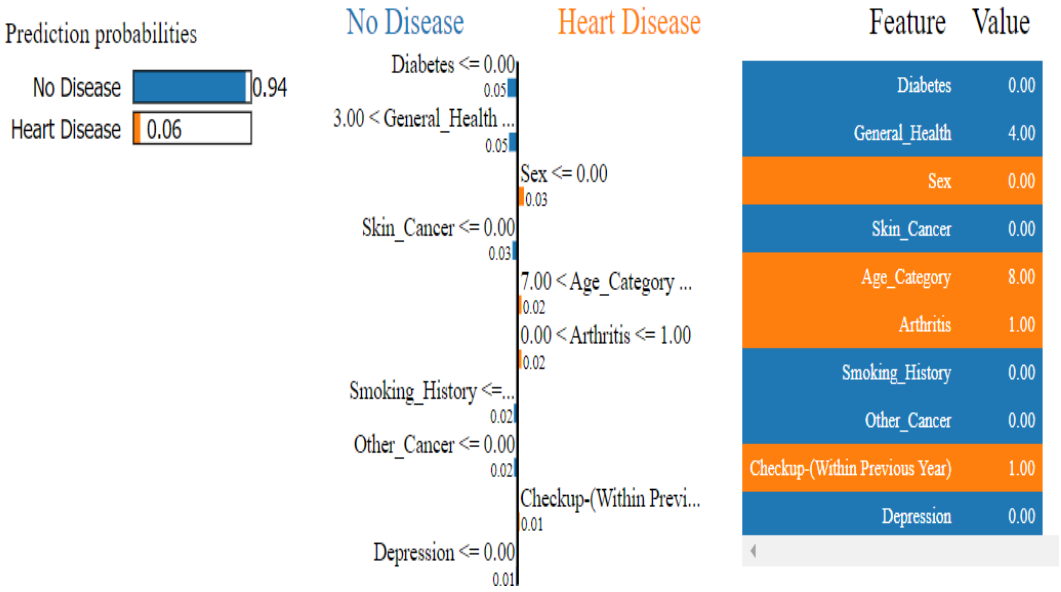


Fig. 8. LIME prediction as 'No Disease" found.

It is shown in Fig. 9 that LIME (Local Interpretable Model-agnostic Explanations) works by generating simplified, interpretable systems that approximate a complex system's decision-making process. In the figure 9, LIME highlights which features positively or negatively influence the outcome, shown by green (positive impact) and red (negative impact) bars. This visualization helps understand the contribution of each feature, such as diabetes, general health, or sex, to the system's prediction. LIME makes AI more interpretable by clarifying the factors driving specific decisions.

It is shown in Fig. 10 that the LIME prediction shows a 51% probability of HD and a 49% probability of no disease. Factors such as age, general health, and diabetes have strong contributions toward HD prediction, as highlighted in orange. Each feature listed, like the age category being greater than 9 and diabetes being present, increases the likelihood of HD, while features like "Other Cancer" lower it slightly. LIME explains how individual factors influence the system's decision, making the prediction more interpretable and understandable.

It is shown in Fig. 11 that in the LIME explanation, the features contributing to the prediction of HD are visualized with green (positive impact) and red (negative impact) bars. Significant positive contributors include age category, general health, and diabetes, indicating these factors increase the likelihood of HD. On the other hand, "Other Cancer" and alcohol consumption slightly reduce the probability, as shown in red. LIME effectively shows how each feature impacts the system's decision, highlighting HD is predicted.

Table VII highlights some previous chatbot systems with multiple approaches. The applied approaches also show the accuracies and miss rate.

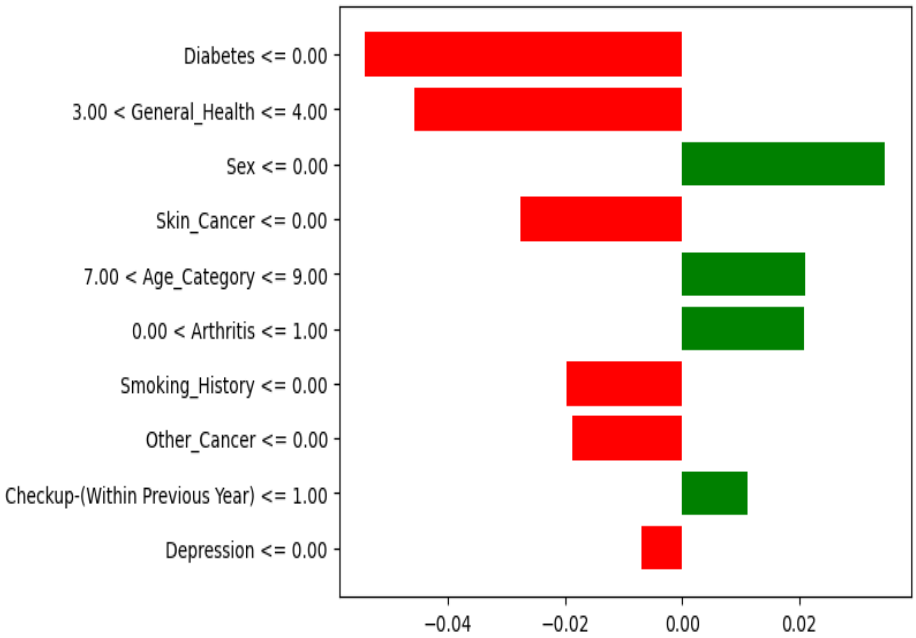


Fig. 9. LIME explanation as 'No Disease" found.

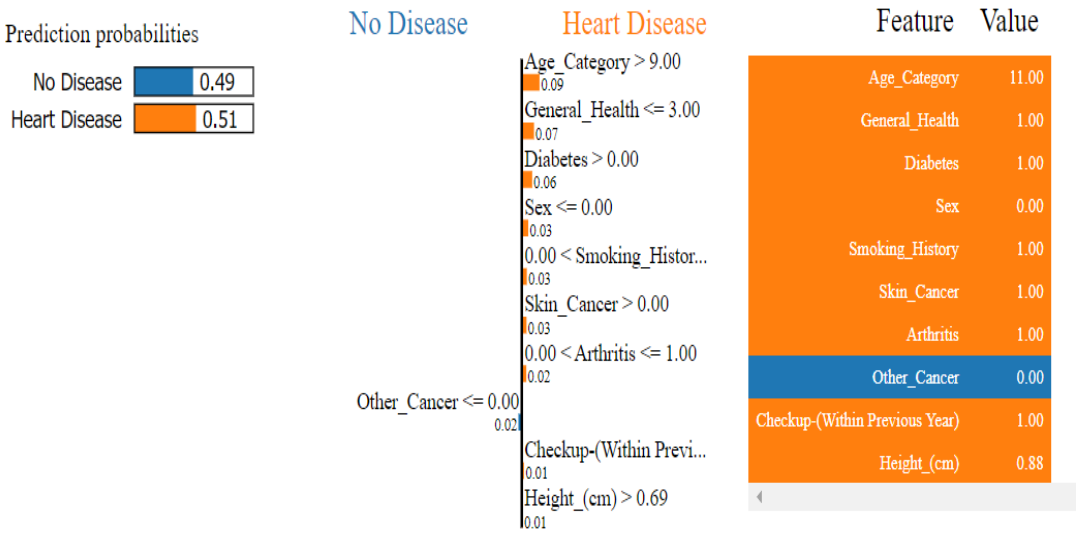


Fig. 10. LIME prediction as 'HD" found.

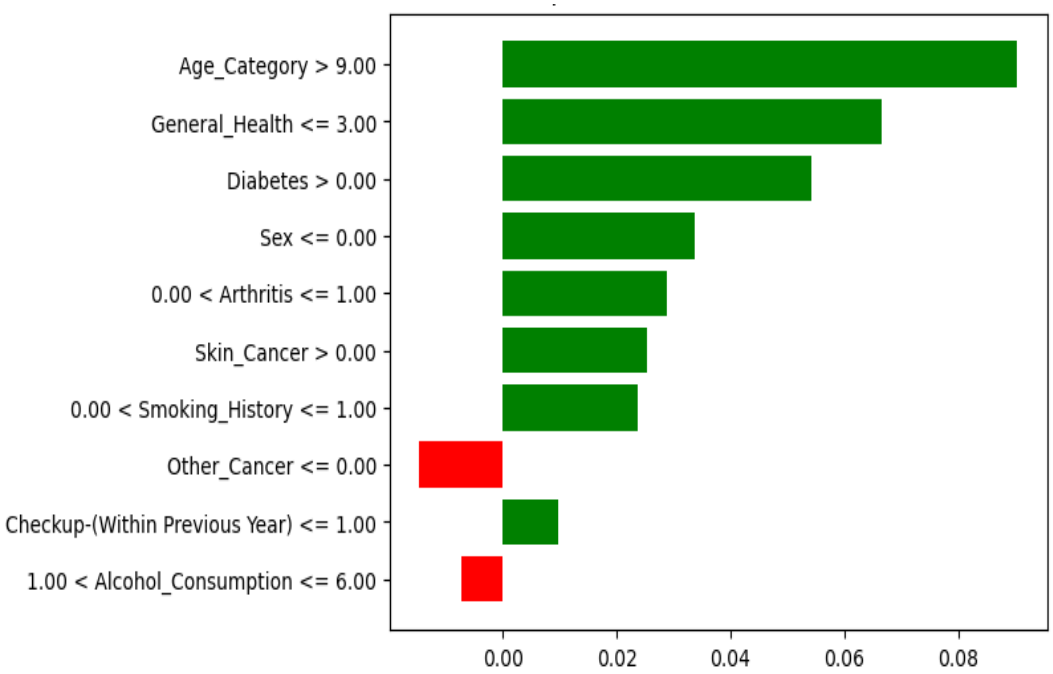


Fig. 11. LIME explanation as "HD" found.

TABLE VII. COMPARING THE PROPOSED CHATBOT SYSTEM PERFORMANCES WITH PREVIOUS APPROACHES

Ref	Year	Approach	Acc (%)	Miss-rate	Chatbot based Approach	XAI Implementation
[36]	2022	SVM	85.40	14.60	No	No
[37]	2021	Logistic Regression & KNN	87.5	12.5	No	No
[38[	2023	Multilayer perceptron with cross-validation	87.28	12.72	No	No
[39]	2021	KNN	84.86	15.14	No	No
[40]	2021	Extra Tree	87.0	13.0	No	No
[41]	2022	SVM	82.5	17.5	No	SHAP
[42]	2023	RF	74.0	26.0	No	SHAP
[43]	2024	DT	91.9	8.1	Yes	No
[44]	2023	Bagging-Quantum Support Vector Classifier (QSVC)	90.16	9.84	No	SHAP
		Proposed system (RF)	**92.0**	**8.0**	**Yes**	**SHAP & LIME**

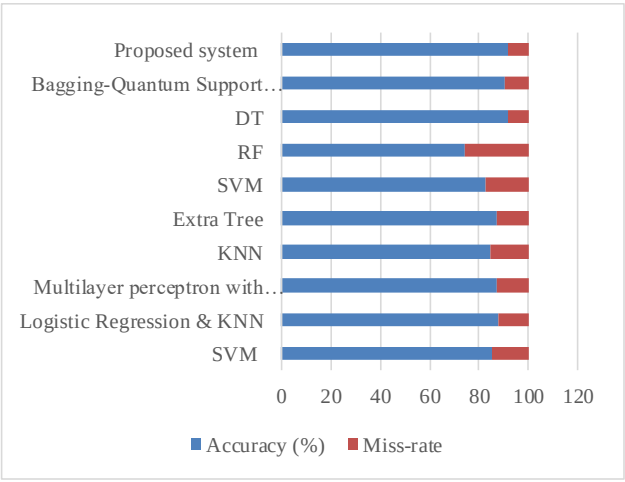


Fig. 12. Graphical representation of the previous approaches with the proposed system.

Table VII and fig. 12, compares the performance of the proposed chatbot system with previous ML approaches to predict HD. It is clearly shown that the proposed approach is better than the previous results in terms of accuracy and miss rate.

X. DISCUSSION

The proposed XAI-driven chatbot demonstrates strong potential in HD prediction by combining accurate ML models with user-friendly, interpretable outputs. This approach addresses the "black-box" limitations of traditional models, enhancing transparency and trust in healthcare applications. While the system shows promising results, its performance may vary across diverse populations, and further validation is needed to ensure its scalability and effectiveness in real-world settings. Future work will focus on addressing these limitations and integrating additional features to improve usability and adaptability.

XI. Conclusion

Heart disease is a significant global health challenge and a leading cause of mortality, with early detection being essential for preventing the disease's progression. Consequently, the objective of this study was to develop an explainable AI-driven chatbot system for predicting heart disease risk, effectively integrating advanced ML techniques with XAI methods. The novelty of the proposed system lies in its remarkable accuracy rate of 92% and an 8% miss rate, demonstrating its superiority over traditional approaches. By utilizing SHAP and LIME, the system boosts transparency in the decision-making process, permitting users to understand the rationale behind predictions. This innovative chatbot interface provides an accessible platform for interaction, ultimately fostering trust and promoting informed decision-making among healthcare providers and patients. The incorporation of XAI with ML in this context sets a new standard for responsible AI applications in healthcare, focusing on the requirement for continuous advancements in predictive modeling to address the ongoing challenges of HD.

XII. Limitations and Future Suggestions

Although XAI-driven chatbot demonstrates strong predictive performance, offering real-time, personalized HD risk assessments. Its ability to provide clear, interpretable results enhances trust and usability, particularly in healthcare settings where transparency is crucial. However, there are limitations: bias during the training of the system with potential biases in the training data, and reliance on specific input features that may not capture the overall picture of health indicators. Future research should continue to enhance this dataset using a larger, more diversified population, systems' robustness, and lifestyle and genetic factors, among others. Additionally, adding health monitoring data in real time will improve predictive accuracy. The long-term future path for perfecting the performance and experience of the chatbot will include continuous user feedback and iterative improvements on the developments themselves, thus ensuring that they continue to be valuable tools in preventive healthcare.

References

[1] Bagheri M, Bagheritaba M, Alizadeh S, Parizi MS, Matoufinia P, Luo Y. AI-Driven Decision-Making in Healthcare Information Systems: A Comprehensive Review.

[2] Chakraborty C, Bhattacharya M, Pal S, Lee SS. From machine learning to deep learning: An advances of the recent data-driven paradigm shift in medicine and healthcare. Current Research in Biotechnology. 2023 Nov 22:100164.

[3] Sahithya B, Prasad G, Sahithi B, Devarlla AC, Yashavanth TR. Empowering Healthcare with AI: Advancements in Medical Image Analysis, Electronic Health Records Analysis, and AI-Driven Chatbots. In2024 3rd International Conference for Innovation in Technology (INOCON) 2024 Mar 1 (pp. 1-7). IEEE.

[4] Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S. Comparison of machine learning methods with traditional models for use of administrative claims with electronic medical records to predict heart failure outcomes. JAMA network open. 2020 Jan 3;3(1):e1918962-.

[5] Muneer S, Rasool MA. AA systematic review: Explainable Artificial Intelligence (XAI) based disease prediction. International Journal of Advanced Sciences and Computing. 2022 Jun 30;1(1):1-6.

[6] Reis L, Maier C, Mattke J, Weitzel T. Chatbots in healthcare: Status quo, application scenarios for physicians and patients and future directions.

[7] Khan MA, Abbas S, Atta A, Ditta A, Alquhayz H, Khan MF, Naqvi RA. Intelligent Cloud Based Heart Disease Prediction System Empowered with Supervised Machine Learning. Computers, Materials & Continua. 2020 Oct 1;65(1).

[8] Lodi A, Olivier P, Pesant G, Sankaranarayanan S. Fairness over time in dynamic resource allocation with an application in healthcare. Mathematical Programming. 2024 Jan;203(1):285-318.

[9] He K, Mao R, Lin Q, Ruan Y, Lan X, Feng M, Cambria E. A survey of large language models for healthcare: from data, technology, and applications to accountability and ethics. arXiv preprint arXiv:2310.05694. 2023 Oct 9.

[10] Mensah GB. Artificial intelligence and ethics: a comprehensive review of bias mitigation, transparency, and accountability in AI Systems. Preprint, November. 2023;10.

[11] Das S, Sultana M, Bhattacharya S, Sengupta D, De D. XAI–reduct: accuracy preservation despite dimensionality reduction for heart disease classification using explainable AI. The Journal of Supercomputing. 2023 Nov;79(16):18167-97.

[12] Ahmad G, Khan MA, Abbas S, Athar A, Khan BS, Aslam MS. Automated diagnosis of hepatitis b using multilayer mamdani fuzzy inference system. Journal of healthcare engineering. 2019;2019(1):6361318.

[13] Mienye ID, Jere N. Optimized ensemble learning approach with explainable AI for improved heart disease prediction. Information. 2024 Jul 8;15(7):394.

[14] Paudel P, Karna SK, Saud R, Regmi L, Thapa TB, Bhandari M. Unveiling Key Predictors for Early Heart Attack Detection using Machine Learning and Explainable AI Technique with LIME. InProceedings of the 10th International Conference on Networking, Systems and Security 2023 Dec 21 (pp. 69-78).

[15] Ogilvie L, Prescott J, Carson J. The use of chatbots as supportive agents for people seeking help with substance use disorder: A systematic review. Eur Addict Res. 2022; 28(6): 405-18. PMID: 36041418 DOI: 10.1159/000525959 [PubMed].

[16] Ahmed A, Hassan A, Aziz S, Abd-Alrazaq AA, Ali N, Alzubaidi M, et al. Chatbot features for anxiety and depression: A scoping review. Health Informatics J.2023; 29(1): 14604582221146719. PMID: 36693014 DOI: 10.1177/14604582221146719 [PubMed].

[17] Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence–based chatbots for promoting health behavioral changes: Systematic review. J Med Internet Res. 2023; 25: e40789. PMID: 36826990 DOI: 10.2196/40789 [PubMed].

[18] Sebastian, G. (2023). Privacy and Data Protection in ChatGPT and Other AI Chatbots: Strategies for Securing User Information. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4454761.

[19] Healthily Chatbot. Available online: https://www.livehealthily.com/app (accessed on 20 March 2023).

[20] Ada Health Chatbot. Available online: https://ada.com/pt/ (accessed on 20 March 2023).

[21] Qi Y. Pilot Quasi-Experimental Research on the Effectiveness of the Woebot AI Chatbot for Reducing Mild Depression Symptoms among Athletes. International Journal of Human–Computer Interaction. 2024 Jan 16:1-8.

[22] Dechert, M. Implementation and Evaluation of a Chatbot to Crowdsource Geotagged Images to Detect Mosquito Breeding Sites. Master's Thesis, University of Bremen, Bremen, Germany, 2019.

[23] Symptomate Chatbot. Available online: https://symptomate.com/about/ (accessed on 20 March 2023).

[24] Swick, R.K. The Accuracy of Artificial Intelligence (AI) Chatbots in Telemedicine. J. South Carol. Acad. Sci. 2021, 19, 17.

[25] Wahyudi D, Ayuningsih E. The Application of Machine Learning in Predicting The Risk of Heart Disease With Decision Tree Algorithm. Instal: Jurnal Komputer. 2024 Jun 30;16(02):120-30.

[26] Dalakoti M, Wong S, Lee W, Lee J, Yang H, Loong S, Loh PH, Tyebally S, Djohan A, Ong J, Yip J. Incorporating AI into cardiovascular diseases prevention–insights from Singapore. The Lancet Regional Health–Western Pacific. 2024 Jul 1;48.

[27] Görtz M, Baumgärtner K, Schmid T, Muschko M, Woessner P, Gerlach A, Byczkowski M, Sültmann H, Duensing S, Hohenfellner M. An

artificial intelligence-based chatbot for prostate cancer education: design and patient evaluation study. Digital Health. 2023 May;9:20552076231173304.

[28] Zambrano Chaves JM, Wentland AL, Desai AD, Banerjee I, Kaur G, Correa R, Boutin RD, Maron DJ, Rodriguez F, Sandhu AT, Rubin D. Opportunistic assessment of ischemic heart disease risk using abdominopelvic computed tomography and medical record data: a multimodal explainable artificial intelligence approach. Scientific reports. 2023 Nov 29;13(1):21034.

[29] Abdulsalam G, Meshoul S, Shaiba H. Explainable heart disease prediction using ensemble-quantum machine learning approach. Intell. Autom. Soft Comput. 2023 Jan 1;36(1):761-79.

[30] Venkat V, Abdelhalim H, DeGroat W, Zeeshan S, Ahmed Z. Investigating genes associated with heart failure, atrial fibrillation, and other cardiovascular diseases, and predicting disease using machine learning techniques for translational research and precision medicine. Genomics. 2023 Mar 1;115(2):110584.

[31] Wilstrup C, Cave C. Combining symbolic regression with the Cox proportional hazards model improves prediction of heart failure deaths. BMC Medical Informatics and Decision Making. 2022 Jul 25;22(1):196.

[32] Narendran M, Sathya A, Annoosh R, Hari S. HealthBot Analytics: Optimizing Healthcare Efficiency Through Intelligent Integration. In2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS) 2024 Apr 18 (pp. 1-7). IEEE.

[33] Rajkumar K, Ragupathi T, Karthikeyan S. Intelligent Chatbot for Hospital Recommendation System. In2024 2nd International Conference on Disruptive Technologies (ICDT) 2024 Mar 15 (pp. 664-668). IEEE.

[34] https://www.kaggle.com/datasets/alphiree/cardiovascular-diseases-risk-prediction-dataset?resource=download.

[35] Paudel P, Karna SK, Saud R, Regmi L, Thapa TB, Bhandari M. Unveiling Key Predictors for Early Heart Attack Detection using Machine Learning and Explainable AI Technique with LIME. InProceedings of the 10th International Conference on Networking, Systems and Security 2023 Dec 21 (pp. 69-78).

[36] Ahamed J. et al., 2022. CDPS-IoT: cardiovascular disease prediction systembased on IoT using machine learning.

[37] Jindal H, Agrawal S, Khera R, Jain R, Nagrath P. Heart disease prediction using machine learning algorithms. InIOP conference series: materials science and engineering 2021 (Vol. 1022, No. 1, p. 012072). IOP Publishing.

[38] Bhatt CM, Patel P, Ghetia T, Mazzeo PL. Effective heart disease prediction using machine learning techniques. Algorithms. 2023 Feb 6;16(2):88.

[39] Rohit Bharti, Aditya Khamparia, Mohammad Shabaz, Gaurav Dhiman, Sagar Pande, and Parneet Singh. 2021. Prediction of heart disease using a combination of machine learning and deep learning. Computational intelligence and neuroscience 2021 (2021).

[40] Sakinat Oluwabukonla Folorunso, Joseph Bamidele Awotunde, Emmanuel Abidemi Adeniyi, Kazeem Moses Abiodun, and Femi Emmanuel Ayo. 2021. Heart disease classification using machine learning models. In International Conference on Informatics and Intelligent Applications. Springer, 35–49.

[41] Pratiyush Guleria, Parvathaneni Naga Srinivasu, Shakeel Ahmed, Naif Almusal- lam, and Fawaz Khaled Alarfaj. 2022. XAI framework for cardiovascular disease prediction using classification techniques. Electronics 11, 24 (2022), 4086.

[42] Moreno-Sanchez PA. Improvement of a prediction model for heart failure survival through explainable artificial intelligence. Frontiers in Cardiovascular Medicine. 2023 Aug 1;10:1219586.

[43] Wahyudi D, Ayuningsih E. The Application of Machine Learning in Predicting The Risk of Heart Disease With Decision Tree Algorithm. Instal: Jurnal Komputer. 2024 Jun 30;16(02):120-30.

[44] Abdulsalam G, Meshoul S, Shaiba H. Explainable heart disease prediction using ensemble-quantum machine learning approach. Intell. Autom. Soft Comput. 2023 Jan 1;36(1):761-79.

Integrating Local Channel Attention and Focused Feature Modulation for Wind Turbine Blade Defect Detection

Zheng Cao*, Rundong He, Shaofei Zhang, Zhaoyang Qi, Sa Li, Tong Liu and Yue Li
State Grid Jilin New Energy Group Co., Ltd, Changchun 130021, China

***Abstract*—In the wind power industry, the health state of wind turbine paddles is directly related to the power generation efficiency and the safe operation of the equipment. In order to solve the problems of low efficiency and insufficient accuracy of traditional detection methods, this paper proposes a wind turbine blade defect detection algorithm that integrates local channel attention and focus feature modulation. The algorithm first introduces the Mixed Local Channel Attention (MLCA) mechanism into the C2f module of the backbone network in YOLOv8 to enhance the extraction capability of the backbone network for key features. Then the Focal Feature Modulation (FFM) module is used to replace the original SPPF module in YOLOv8 to further aggregate global contextual features at different levels of granularity; finally, in the Neck part, the progressive feature pyramid AFPN structure is used to enhance the multi-scale feature fusion capability of the model, which in turn improves the accuracy of small object detection. The experimental results show that the proposed algorithm has an accuracy of 82.5%, a mAP50 of 78.6%, and GFLOPS of 8.5. In the detection of wind turbine blade defects, which possesses higher detection performance and real-time performance compared with the traditional methods, and is able to effectively identify common defects such as cracks, corrosion, and abrasion, and exhibits strong robustness and application value.**

***Keywords*—Fan blades; YOLO; attention mechanism; defect detection; inner-IoU**

I. Introduction

Wind energy, as a renewable energy source, is highly valued by countries around the world due to its cleanliness, environmental friendliness, and low carbon emissions. The utilization of wind energy is of great significance in supporting the dual-carbon development, thus holding an important position in the construction of China's clean energy system [1]. The wind turbine blade, as a crucial component in wind power generation, plays a key role in converting wind energy into electrical energy. However, due to the complex and diverse outdoor environment, wind turbine blades often suffer from erosion, cracks, delamination, and other damage defects, which affect the power generation efficiency of the units and even threaten the safety and service life of the wind turbines [2].

Traditional wind turbine blade defect detection primarily relies on manual inspection, which suffers from low efficiency, high costs, and strong subjectivity. Therefore, the real-time, efficient, and accurate detection of blade defects has become an urgent need in wind turbine blade defect identification. Existing wind turbine blade defect detection algorithms can be broadly categorized into those using physical sensing technology for detection and those employing computer vision technology for image recognition. Methods using physical sensing technology mainly focus on static detection with grating Energies [3] and electromagnetic ultrasonic testing [4].

In recent years, with the development of artificial intelligence technology and advancements in drone technology, drone aerial photography and object detection techniques have been widely applied to wind turbine blade defect detection [5, 6]. Kang Shuang et al. achieved defect detection on high-altitude wind turbine blades by analyzing temperature thresholds to extract effective features [7]. Yu et al. [8] trained a DCNN network to extract semantic features of defect areas for wind turbine blade recognition. Additionally, some experts and scholars have explored the optimization of deep learning-based object detection models. In the improvement of two-stage object detection algorithms, Zhang Chao [9] improved the backbone network of Mask RCNN by combining it with a multi-scale feature fusion FPN structure, enhancing the detection effect of wind turbine blade defects. Qu Zhongkan et al. [10] utilized DCNv2 and GIoU to improve Faster R-CNN, significantly enhancing the detection accuracy of delamination defects.

In the improvement of one-stage algorithms, the focus has mainly been on the YOLO series. Wu Yuping et al. [11] adopted a skip-connection feature fusion network structure in YOLOv3-Tiny and introduced the Inception module, significantly improving the detection accuracy of blade defects. Gao Wenjun [1-15] replaced the original feature extraction network of YOLOv4 with a GhostNet feature extraction network, successfully achieving lightweight blade detection models while maintaining good detection accuracy. Su Jia et al. [13] designed a multi-scale feature fusion mechanism in the YOLOv5 model, thereby improving the detection effect on small objects. Zheng Qishan [14] proposed a wind turbine blade defect detection method based on YOLOv5s as the baseline model, by replacing the fast spatial pyramid pooling (SPPF) structure in the backbone network and introducing the convolutional attention mechanism (CBAM) module, further enhancing the model's detection accuracy. Wang Zhengshuai [15] addressed the issue of low detection accuracy in complex environments by adding a Scoring module attention mechanism to the YOLOv5 baseline model, scoring each channel's features to filter out low-scoring features and fuse high-scoring features, thereby improving feature fusion capabilities and detection accuracy. Li Bing et al. [16] proposed an HSCA-YOLOv7 wind turbine blade defect detection algorithm, effectively solving the problem of

inconsistent defect scales in wind turbine blade images. Fu Jinyi et al. [17] proposed an improved small object detection algorithm, CA-YOLOv8, by embedding an aggregation capability module (CAM) and improving the C2f module, enhancing the ability to capture multi-scale detailed features of detection objects.

The methods for detecting defects in wind turbine blades in the past have gradually evolved from traditional manual methods to physical sensing technologies and computer vision technologies. In the latter, whether it is the early exploration based on temperature thresholds and DCNN networks, or the improvement of two-stage and one-stage target detection algorithms, the detection capabilities have been enhanced to a certain extent, but there are still shortcomings. For example, the complex background has a significant impact on the model's detection of defect areas. The defect areas in aerial images of wind turbine blades are mostly small targets, and different-scale targets have higher requirements for the model's feature processing ability. Therefore, the model in this paper aims to further optimize the defect detection algorithm of wind turbine blades for these problems and improve its detection accuracy and efficiency for small targets and multi-scale targets in complex environments.

The complex background of aerial wind turbine blade images significantly affects the model's detection of defect areas; the defect areas in aerial wind turbine blade images are mostly small objects, requiring high feature processing capabilities for different scale objects. To address these issues, this paper proposes a fusion of local channel attention and focal feature modulation object detection algorithm. This paper mainly improves the YOLOv8 model, with innovations focused on enhancing the overall feature extraction capabilities of the model and improving the detection capabilities for small-scale objects. The main innovations include the following four aspects:

1) Improving feature capture for small-scale objects: By integrating the MLCA module (Mixed Local Channel Attention) with the C2f module in YOLOv8's Backbone, the model's ability to capture details of small-scale objects is enhanced.

2) Enhancing overall detection accuracy: Using the FFM module (Focal Feature Modulation) to replace the original SPPF (Spatial Pyramid Pooling - Fast) module, further improving the model's overall detection accuracy.

3) Optimizing multi-scale feature fusion: Introducing AFPN (Asymptotic Feature Pyramid Network) to optimize multi-scale feature fusion in object detection.

4) Improving model accuracy and generalization: Using Inner-IoU to optimize the original model's loss function, further enhancing model accuracy and generalization capabilities

The subsequent part of this paper will conduct a detailed analysis of the algorithm structure, including the improvement principles of the C2f-MLCA and FFM modules in the Backbone part, the AFPN structure in the Neck part, and the role of the Inner-IoU loss function. Then, through the experimental results and analysis, the configuration such as the dataset will be introduced, and various experiments and evaluations will be carried out using multiple indicators to demonstrate the performance advantages of the algorithm. Finally, a conclusion will be drawn to summarize the performance improvement and application significance of this algorithm in the defect detection of wind turbine blades.

II. Improved YOLOv8 Algorithm Structure

A. Improved Network Model

This paper proposes a fusion of local channel attention and focal feature modulation object detection algorithm based on YOLOv8. In the Backbone of YOLOv8, the MLCA module is used to construct the C2f-MLCA module, improving the model's ability to capture details of small-scale objects; secondly, the FFM module is used to replace the SPPF module in the Backbone to aggregate contextual information at different granularity levels; finally, the AFPN structure is used in the Neck part to enhance the model's multi-scale feature fusion capabilities. The overall framework of the improved network is shown in Fig. 1.

B. Improvements in the Backbone

a) C2f-MLCA: There are numerous small-scale objects in wind turbine blade defects [18]. To enhance the detection effect on these small-scale objects, the module used needs to possess better global and local detail feature extraction capabilities. In YOLOv8, multiple Conv modules and C2f modules constitute the backbone network, which is used to extract deep features from images. However, the Bottleneck module in the C2f module causes the network to superimpose a large amount of information at high-frequency positions, while neglecting information at low-frequency positions, thereby reducing the model's feature extraction capability for small objects, as shown in Fig. 2.

MLCA is a lightweight attention module that can simultaneously consider channel information and spatial information, and combine local and global information to enhance the expressive effect of the network [19]. Therefore, the MLCA module is introduced into the Bottleneck of the C2f module, as shown in Fig. 3.

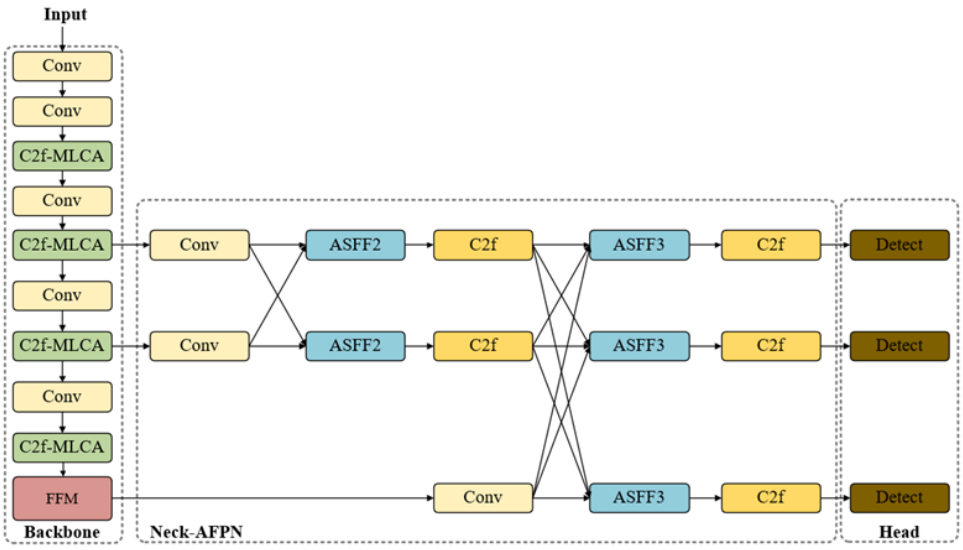


Fig. 1. Overall framework of the model.

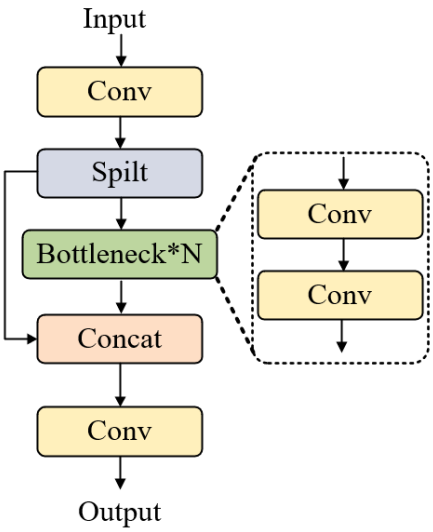


Fig. 2. C2f module.

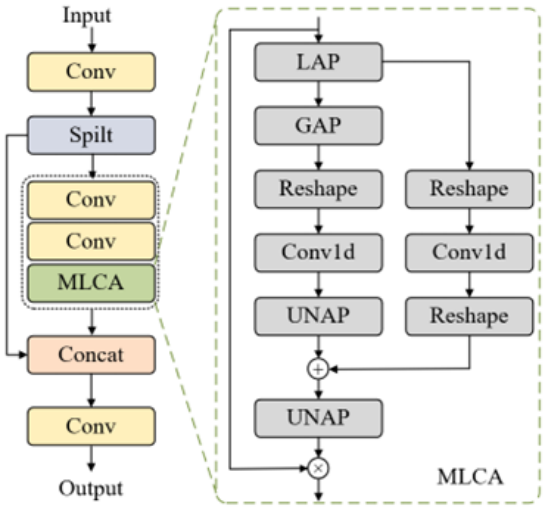


Fig. 3. MLCA module.

The MLCA module primarily utilizes local average pooling and global average pooling to capture global and local feature information of the entire input feature map, and on this basis, performs channel attention and spatial attention to improve feature extraction accuracy. A feature map of size $C \times H \times W$ is first processed by local average pooling (LAP) and global average pooling (GAP) to obtain statistical information of the entire feature map. Subsequently, the features after local and global pooling are transformed by a one-dimensional convolution (Conv1d) to compress the feature channels while maintaining the spatial dimensions, and then rearranged to adapt to subsequent operations. For the features after local pooling, after one-dimensional convolution and rearrangement, they are combined with the original input features through multiplication operations for feature selection to enhance attention to useful features; for the features after global pooling, after one-dimensional convolution and rearrangement, they are combined with the local pooling features through addition operations to fuse global context information. Finally, the feature map processed by local and global attention is restored to the original spatial dimension through anti-average pooling (UNAP) operations to achieve the purpose of mixed attention.

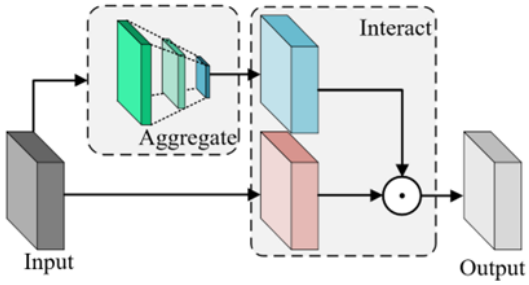


Fig. 4. FFM module.

b) FFM module: During the detection process of wind turbine blade defect objects, the model is easily affected by the complex background interference associated with the samples, making it difficult to fully extract the deep features of the samples. YOLOv8 uses the SPPF module in the Backbone part

to aggregate multiple scale features, but it lacks sufficient modeling of long-range visual context information. In feature information processing, it only concatenates feature map information of different scales to achieve feature aggregation, ignoring the interaction and aggregation of multi-granularity information under complex background interference. To address this issue, this paper introduces a more efficient focal feature aggregation module, FFM [20]. By extracting context information from local to global at different granularity levels and performing gated aggregation, it improves the interference of complex background information on the feature extraction process of the model's backbone network. The structure of the FFM module is shown in Fig. 4.

First, the input feature $X \in \mathbb{R}^{H\times W\times C}$ is fed into the aggregation module to aggregate contextual features, obtaining aggregated features. This process mainly includes two steps: The first step involves sending the input feature into a linear layer to extract the initial level feature Z^0. Then, the initial level feature is processed through a Depthwise Convolution (DWConv) and a GeLU function to obtain the next level feature until the L -th level feature. In this way, a total of (L+1) level features are obtained. These features are considered as contextual information at different granularity levels, as shown in Eq. (1) and Eq. (2):

$$Z^0 = linearlayer(X) \tag{1}$$

$$Z^l = GeLU(DWConv(Z^{l-1})) \tag{2}$$

Among them, $l \in \{1,\cdots,L\}$ represents different levels.

The second step is gated aggregation. By gated aggregation, contextual feature information at different granularity levels is condensed into a single feature vector as the aggregated feature Z^{Out}, with the calculation formula as follows:

$$Z^{Out} = \sum_{l=1}^{L+1} G^l \odot Z^l \tag{3}$$

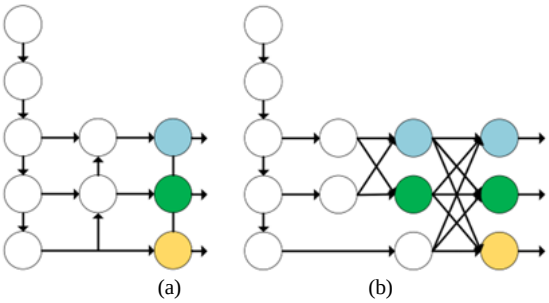


Fig. 5. The schematic of the network before and after the improvement of the Neck part. (a) YOLOv8; (b) YOLOV8+AFPN.

where G^l is the gate weight obtained by the l-th level feature through a linear layer.

Then, the obtained aggregated feature Z^{Out} is processed through a linear layer and element-wise multiplied with the feature query value to obtain the final output focal modulation feature Y, with the calculation formula as follows:

$$Y = q(X) \odot linerlayer(Z^{Out}) \tag{4}$$

where $q(\cdot)$ is the query projection function for obtaining the feature query value; $\odot$ represents element-wise multiplication.

The FFM module achieves this not only by expanding the receptive field and effectively mitigating the interference from the complex background in wind turbine blade defect samples but also by more efficiently obtaining contextual information from granularity levels, thereby enhancing the feature extraction capability of the Backbone.

C. Improvements in the Neck Part

In the Neck part, the AFPN structure [21] is adopted instead of the original FPN+PAN structure in YOLOv8. This is because, during the feature extraction process, the downsampling process is dominated by large objects, causing the feature representation of small objects in the object area to be confused or blurred, thereby affecting the detection effect of small objects. The AFPN structure used in this paper can optimize the multi-scale feature fusion process. By gradually fusing low-level and high-level features, it avoids the loss and degradation of information, as shown in Fig. 5.

The AFPN structure can adaptively fuse features of different levels in the Neck. Firstly, it fuses two adjacent low-level features, and then progressively incorporates higher-level features. This approach alleviates the conflict of multi-object information that arises from each spatial location during the feature fusion process. Specifically, during the multi-level feature fusion process in the Neck part, an adaptive spatial fusion module (ASFF) is used to assign different spatial weights to features of different levels. Then, the features of different levels are fused together using a weighted sum method, with the specific calculation formula as follows:

$$y_{ij}^{\ell} = \sum_{n=1}^{\ell} \alpha_{ij}^{n} x_{ij}^{n\to\ell} \tag{5}$$

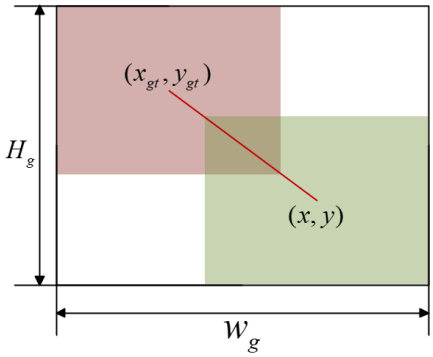


Fig. 6. Parameter definition.

Where y_{ij}^{ℓ} represents the features after adaptive spatial fusion, α_{ij}^{n} represents the spatial weight of the features at the ℓ level in the hierarchy, and $\sum_{n=1}^{\ell} \alpha_{ij}^{n} = 1$; $x_{ij}^{n\to\ell}$ represents the feature vector from the n-th level to ℓ at position (i,j).

D. Improvement of Loss Function

a) CIoU loss: The original YOLOv8 model employs the CIoU Loss as the bounding box regression loss function. Let the ground truth box be denoted as $\vec{B}_{gt} = [x_{gt}, y_{gt}, w_{gt}, h_{gt}]$ and the predicted box as $\vec{B} = [x, y, w, h]$, where x and y represent the center coordinates of the bounding box, and w and h represent the dimensions of the box. The original CIoU Loss is defined by Eq. (6)-Eq. (9) as follows:

$$L_{CIoU} = L_{IoU} + \frac{(x - x_{gt})^2 + (y - y_{gt})^2}{(W_g + H_g)^2} + \alpha v \tag{6}$$

$$L_{IoU} = 1\text{-}IoU = 1 - \frac{W_i H_i}{wh + w_{gt}h_{gt} - W_i H_i} \tag{7}$$

$$\partial = \frac{v}{L_{IoU} + v} \tag{8}$$

$$v = \frac{4}{\pi^2}(\tan^{-1}\frac{w}{h} - \tan^{-1}\frac{w_{gt}}{h_{gt}})^2 \tag{9}$$

In the equation, L_{IoU} is used to measure the overlap between the predicted box and the ground truth box, α is a balancing parameter, and v is used to measure the consistency of the aspect ratio. The remaining parameters are defined as shown in Fig. 6.

b) Inner-IoU: The existing IoU-based bounding box regression still focuses on accelerating convergence by adding new loss terms, while neglecting the limitations of the IoU loss term itself. In practical applications of wind turbine blade defect detection, the original bounding box regression process cannot self-adjust according to different detectors and detection tasks, resulting in low detection efficiency and accuracy for dense small objects or complex environments with multiple scales. To address this issue, Inner-IoU loss [22] is chosen, which calculates the IoU loss through an auxiliary bounding box. For different datasets and detectors, the scale of the auxiliary bounding box used to calculate the loss can be controlled by a scale factor. The definition of Inner-IoU is shown in Eq. (10)-(16).

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} * ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} * ratio}{2} \tag{10}$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} * ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} * ratio}{2} \tag{11}$$

$$b_l = x_c - \frac{w * ratio}{2}, b_r = x_c + \frac{w * ratio}{2} \tag{12}$$

$$b_t = y_c - \frac{h * ratio}{2}, b_b = y_c + \frac{h * ratio}{2} \tag{13}$$

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) * (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \tag{14}$$

$$union = (w^{gt} * h^{gt}) * (ratio)^2 + (w * h) * (ratio)^2 - inter \tag{15}$$

$$IoU^{inner} = \frac{inter}{union} \tag{16}$$

In the Inner-IoU Loss, the variable ratio corresponds to the scale factor, which typically ranges from [0.5, 1.5]. The traditional IoU calculation method considers the overlap area between the predicted bounding box and the overall bounding box, while Inner-IoU focuses more on the core part of the bounding box to make a more precise judgment of the overlap area.

III. Experimental Results and Analysis

A. Experimental Dataset

The data used in this paper are sourced from images taken by inspection personnel at a wind farm during the inspection of wind turbine blades. The division of the aerial wind turbine blade defect dataset is shown in Table I, totaling 784 images with a resolution of 1200*900. The dataset is randomly divided into training, validation, and test sets in a ratio of 7:2:1.In the defect detection of wind turbine blades, 70% of the data is used for training to enable the model to learn multiple features and enhance performance. The 20% validation set assists in training and optimizing the model. The 10% test set independently evaluates to ensure objectivity and accuracy. The 7:2:1 data division ratio balances multiple key factors and enhances the reliability of the experiment.

B. Experimental Environment and Parameter Configuration

The operating system used in this experiment is Ubuntu 18.04, with GPU being NVIDIA Geforce RTX 3090 and CPU being Intel(R) Xeon(R) Gold 6148. The deep learning framework is Pytorch-2.1.2. During the training of the object detection model, the number of training epochs is set to 200, the momentum size is set to 0.937, the batch size is set to 16, and the initial learning rate is set to 0.0001.

C. Experimental Evaluation Metrics

To verify the effectiveness of the improvements in the proposed algorithm, the model is evaluated using four metrics: Precision (P), Recall (R), Average Precision (AP), and mean Average Precision (mAP). The definitions of these metrics are given in Eq. (17)-(20), where n is the total number of predicted bounding box categories.

$$P = \frac{TP}{TP + FP} \tag{17}$$

$$R = \frac{TP}{TP + FN} \tag{18}$$

$$AP = \int_0^1 P(R)dR \tag{19}$$

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP(i) \tag{20}$$

D. Results Comparison and Analysis

a) Comparative analysis of different object detection model: To verify the advanced nature of the proposed method

for wind turbine blade defect detection, the proposed improved algorithm is compared with other mainstream object detection model algorithms on the wind turbine blade defect dataset. The experimental results are shown in Table II. From Table II, it can be seen that the proposed improved algorithm significantly outperforms the two-stage models Faster R-CNN [23] and Cascade R-CNN[24] in terms of mAP and detection speed. Compared to the single-stage object detection algorithms YOLOv3[25],YOLOv4[26],YOLOv5s,YOLOv5l,YOLOv8s,YOLOv9[27],YOLOv10[28] and YOLOv8n, the proposed algorithm achieves a substantial improvement in mAP while maintaining a high level of accuracy and a fast detection speed, indicating that the proposed method has certain advantages.

To provide a more intuitive comparison of the detection effects before and after the improvement of the proposed algorithm with other mainstream YOLO series algorithms, the detection results of different scenarios are compared and displayed in Fig. 7(a)-(f). The first row represents complex background scenarios, the second row represents scenarios with multiple types of defects, the third row represents scenarios with densely distributed small objects, and the fourth row represents scenarios with multiple types of defects and small objects. It can be seen from the comparison that the proposed algorithm has a lower rate of missed detections and false detections and higher recognition accuracy compared to other algorithms.

TABLE I. TABLE OF DATASET DIVISION FOR AERIAL WIND TURBINE BLADE DEFECTS

Fan blade defect name	Quantity of categories
gelcoat_off	156
cracks	374
surface_erosion	254

TABLE II. COMPARISON EXPERIMENT OF DIFFERENT OBJECT DETECTION MODELS

Model	P/%	R/%	mAP50/%	mAP50-95/%	GFLOPs
Faster R-CNN	75.4	66.0	73.5	43.1	7.1
Cascade R-CNN	72.5	67.8	72.7	44.2	7.1
YOLOv3	73.5	67.9	71.9	43.3	6.9
YOLOv4	76.6	65.9	74.9	45.4	6.7
YOLOv5s	78.4	65.3	75.2	44.4	23.8
YOLOv9s	78.6	66.4	74.6	50.2	26.7
YOLOv10n	72.4	58.9	71.8	40.6	8.2
YOLOv8s	78.2	67.7	75.8	48.6	28.4
YOLOv8n	77.1	68.7	74.6	46.4	8.1
Ours	82.5	71.6	78.6	51.7	8.5

b) Module ablation experiment results: To verify the contribution of each improvement to the proposed algorithm, ablation experiments were conducted. Under the same experimental conditions, YOLOv8n was used as the baseline model, and a series of ablation experiments were performed by sequentially adding each improvement measure. The results are shown in Table III. From the comparison of experimental results in Table III, it can be seen that the original YOLOv8n model performs relatively basic across various metrics, with mAP50 at 88.4%, P at 65.2%, R at 73.2%, and so on. As different improvement methods were introduced, the performance of each model improved. For example, YOLOv8n+MCLA showed progress in some metrics compared to the original model; YOLOv8n+FFM, YOLOv8n+Neck improvement, YOLOv8n+Inner-IoU, and other single or combined improvement methods also enhanced the model's performance to varying degrees. However, the proposed model outperformed in all metrics. It achieved an mAP50 of 91.4%, higher than all other improved models; P was 67.8%, R was 76.5%, showing good performance in accuracy and recall. Additionally, the mAP50-95% metric was 71.6%, higher than other models, indicating that the proposed model maintains high detection accuracy across different confidence levels. Although the GFLOPs slightly increased, it remained within a reasonable range.

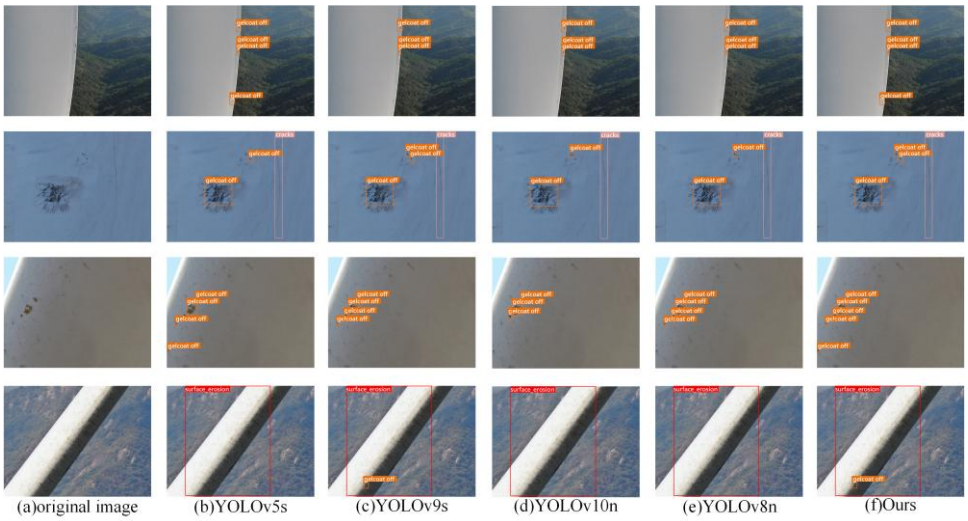


Fig. 7. Comparison of detection effect of different methods before and after improvement.

TABLE III. RESULTS OF ABLATION EXPERIMENT

Model	mAP50/%			P/%	R/%	mAP50/%	mAP50-95/%	GFLOPs
	surface_ erosion	cracks	gelcoat _off					
YOLOv8n	88.4	65.2	73.2	77.1	68.7	74.6	46.4	8.1
YOLOv8n+MCLA	89.6	66.2	73.6	78.6	68.8	74.9	48.7	7.6
YOLOv8n+FFM	88.9	66.3	75.6	77.8	68.6	76.9	46.8	7.8
YOLOv8n+Neck	90.4	65.4	76.3	78.2	69.0	77.4	48.2	8.5
YOLOv8n+Inner-IoU	90.6	67.0	76.4	80.2	70.8	78.0	48.7	8.1
YOLOv8n+ MCLA +FFM	88.7	67.1	75.8	80.6	71.2	77.2	47.5	7.8
YOLOv8n+ MCLA +FFM+ Neck	91.2	67.6	76.0	81.2	70.3	78.3	47.9	8.5
YOLOv8n+ MCLA +FFM +Inner-IoU	90.6	67.2	76.5	82.1	71.4	78.1	48.2	7.8
YOLOv8n+Neck+Inner-IoU	89.6	67.4	75.2	81.7	71.3	77.4	50.6	8.5
Ours	91.4	67.8	76.5	82.5	71.6	78.6	51.7	8.5

TABLE IV. PERFORMANCE OF MODELS WITH DIFFERENT C2F-MCLA NUMBERS

Model	mAP50/%	GFLOPs
YOLOv8n	74.6	8.1
C2f-MCLA*1	74.9	7.6
C2f- MCLA *2	74.7	7.6
C2f- MCLA *3	75.1	7.7

In summary, the proposed model, by integrating multiple improvement methods, demonstrated outstanding performance in defect detection tasks, providing a more effective solution for detecting issues such as surface erosion cracks and gel coating detachment.

c) Impact of C2f-MLCA module on the proposed model: In YOLOv8's Backbone, there are three instances where the C2f module is used. The stacking of different numbers of C2f-MCLA modules has varying impacts on network performance and lightweighting. To determine the optimal number of improvement modules for peak performance, YOLOv8n was used as the baseline algorithm, and the first, the first two, and all C2f modules were improved. The network performance with different numbers of C2f-MCLA modules was compared, with the module improvement sequence following the network structure hierarchy. The experimental results are shown in Table IV.

The mAP50 of YOLOv8n is 74.6%, and GFLOPs is 8.1. The C2f-MCLA series achieves mAP50 values close to or slightly higher than YOLOv8n under different configurations, with C2f-MCLA*3 reaching 75.1%; and the series has GFLOPs around 7.6 or 7.7, indicating relatively lower computational complexity. Overall, the C2f-MCLA series strikes a good balance between object detection accuracy and computational efficiency, potentially offering more advantages in scenarios with limited computational resources or high real-time requirements, providing a reference for selecting object detection algorithms.

d) Evaluation of model generalization and robustness: In the field of object detection, a model's generalization capability is primarily reflected in its ability to adapt well to changes brought by different datasets. The robustness of a model focuses more on its ability to better adapt to the effects brought by changes in sample backgrounds and multiple angles. To evaluate the generalization and robustness of the proposed method, 40 wind turbine defect samples under multiple backgrounds and angles were selected for experimental detection, with results shown in Table V.

TABLE V. EVALUATION RESULT

Model	Number of correct identifications	Accuracy
YOLOv8n	32	80%
YOLOv8n+MCLA	37	92.5%
YOLOv8n+FFM	37	92.5%
YOLOv8n+AFPN	32	80%
YOLOv8n+Inner-IoU	37	92.5%
YOLOv8n+ MCLA +FFM	36	90%
YOLOv8n+ MCLA +FFM+ AFPN	37	92.5%
YOLOv8n+ MCLA +FFM +Inner-IoU	37	92.5%
YOLOv8n+AFPN+Inner-IoU	33	82.5%
Ours	37	92.5%

In the original YOLOv8n model, the number of correctly identified objects is 32, with an accuracy rate of 80%. By separately introducing improvements such as MCLA, FFM, AFPN, and Inner-IoU, the model's performance has been enhanced to varying degrees. Notably, YOLOv8n + MCLA, YOLOv8n + FFM, YOLOv8n + Inner-IoU, YOLOv8n + MCLA + FFM + AFPN,YOLOv8n+MCLA+FFM+Inner-IoU, and the model pro-posed in this study all achieved 37 correctly identified objects and a high accuracy rate of 92.5%. Although the accuracy rate is the same as that of the above-mentioned models, the improvement measures such as MLCA, FFM, AFPN, and Inner-IoU integrated in this model collaborate with each other. The AFPN structure helps maintain a high accuracy rate; the MLCA module enhances the data processing ability, giving it an advantage in feature learning. The current similar accuracy rate might be a phased performance, and the scalability

of its architecture indicates that it has greater potential to surpass other models in the field of wind turbine blade defect detection in the future. This clearly indicates that these improvement methods have a significant effect on enhancing the accuracy of object recognition. However, despite some improvements in YOLOv8n+AFPN and YOLOv8n+AFPN+Inner-IoU, the extent of improvement is relatively small. Overall, the model proposed in this study performs excellently in the task of detecting defects in wind turbine blades, providing valuable references for further optimization of the YOLOv8n model.

IV. Conclusion

The detection of defects in wind turbine blades is of great significance for intelligent inspection in wind farms [29-30]. To address the practical issues of existing algorithms for detecting defects in wind turbine blades, such as insensitivity to small object areas, difficulty in feature extraction, and challenges in dealing with the effects of complex environments, this paper proposes a defect detection algorithm for wind turbine blades that integrates local channel attention and focal feature modulation. Through experiments on a specific dataset of wind turbine blade defects, the proposed method improves the accuracy by 5.4%, the recall by 2.9%, and mAP50 by 4.2% compared to the baseline method, with enhanced generalization and robustness. The proposed method also demonstrates certain advantages over other advanced object detection methods, providing a solution to assist in the intelligent inspection of wind farms. In the future, we will continue to focus on model compression and acceleration in terms of algorithm optimization, improve the attention mechanism to enhance the detection speed and accuracy, and increase the application value of the algorithm.

Acknowledgment

Author Contributions: Conceptualization, Zheng Cao and Yue Li; Data curation, Rundong He, Shaofei Zhang and Sa Li; Formal analysis, Rundong He; Investigation, Shaofei Zhang; Methodology, Zheng Cao, Rundong He and Tong Liu; Project administration, Zhaoyang Qi; Resources, Tong Liu and Yue Li; Software, Zheng Cao; Validation, Tong Liu; Visualization, Sa Li; Writing – original draft, Zheng Cao; Writing – review & editing, Zhaoyang Qi and Yue Li.

Funding: This work was supported by the State Grid Jilin New Energy Group Company Science and Technology Project (SGJLNY00YJJS2400119).

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] J. Cai, G. Hu, H. Shi, “Review on evaluation standards of wind energy resources at home and abroad,” Wind Energy, vol. 12, pp. 56-63, 2021.

[2] C. SAAD, B. H. Mostafa and H. J. R Abderrahmane, “Performance Analysis of Faults Detection in Wind Turbine Generator Based on High-Resolution Frequency Estimation Methods,” International Journal of Advanced Computer Science and Applications (IJACSA), vol. 5, no. 4, 2014.

[3] Z. Cao, Q. Wang. “Deep Learning-Based Image Recognition Technology for Wind Turbine Blade Surface Defects,” International Journal of Advanced Computer Science & Applications, vol. 12, no. 9, 2024.

[4] B. Z. Bo, Z. Yanan and C. Changzheng, “Acoustic emission detection of fatigue cracks in wind turbine blades based on blind deconvolution separation,” Fatigue & Fracture of Engineering Materials & Structures, vol. 40, no. 6, pp. 959-970, 2017.

[5] B. He, H. Jia, F. Zhao, “Application of UAV in fan blade detection,”Electrical technology, pp. 64-65, 2019.

[6] M. Ma, H. Liu, Y. Gao, S. Li, and C. Ma, “Fully autonomous fan blade inspection based on UAV and artificial intelligence,” Electric age,pp. 44-48, 2023.

[7] S. Kang, C. Chen, B. Zhou, “Research on infrared detection of wind power blade defect recognition based on temperature threshold,” Journal of Solar Energy,pp. 337-341, 2019.

[8] Y. Yu, H. Cao, X. Yan, et al., “Defect identification of wind turbine blades based on defect semantic features with transfer feature extractor,” Neurocomputing, vol. 376, pp. 1-9, 2020.

[9] W. Gao, H. Zhang, “Based on Improved YOLOv4 fan blade Defect Detection Method. Agricultural Equipment and Vehicle Engineering,” vol. 61, no. 08, pp. 94-98, 2023.

[10] Z. Qu, X. Li, F. Li, “Design of fan blade defect image multi-point detection system based on Faster R-CNN,” Electronic design engineering, vol. 29, no. 04, pp. 57-61, 2021.

[11] Y. Wu, H. Liu, J. Wu, H. Jia, J. Diao, and D. Zhu, “Application of improved YOLOv3-Tiny network in damage detection of fan blades,” Hebei Industrial Science and Technology, vol. 38, no. 05, pp. 401-408, 2021.

[12] W. Gao, H. Zhang, “Based on the improved YOLOv4 fan blade defect detection method,” Agricultural Equipment and Vehicle Engineering, vol. 61, no. 08, pp. 94-98, 2023.

[13] J. Su, Y. Qin, Z. Jia, J. Wang, “Small target detection algorithm based on ATO-YOLO,” Computer Engineering and Applications, vol. 60, no. 06, pp. 68-77, 2024.

[14] Q. Zheng, S. Zhu, C. Chen, H. Chang, and X. Yan, “Fan blade defect detection method based on improved YOLOv5s,” Information and Computers, vol. 35, no. 16, pp. 14-18, 2023.

[15] Z. Wang, L. Qiu, Y. Li, “YOLOv5s pyrotechnic detection method in complex environment,” Electronic Measurement Technology, vol. 46, no. 24, pp. 149-156, 2023.

[16] B. Li, Y. Bai, K. Zhao, C. Guo, and Y. Zhai, “Wind turbine blade surface defect detection algorithm based on HSA-YOLOV7,” Electric Power. vol. 56, no. 10, pp. 43-52, 2023.

[17] J. Fu, Z. Zhang, W. Sun,and K. Zou, “Improved YOLOv8 aerial image small target detection algorithm,” Computer Engineering and Applications, vol. 60, no. 06, pp. 100-109, 2024.

[18] Y. Xin, G. Wu, Y. Zuo, “Fan blade defect detection method based on EfficientDet,” Electronic Measurement Technology, vol. 45, no. 05, pp. 124-131, 2022.

[19] D. Wan, R. Lu, S. Shen, et al., “Mixed local channel attention for object detection,” Engineering Applications of Artificial Intelligence, vol. 123, 2023.

[20] J. Yang, C. Li, X. Dai, et al., “Focal modulation networks,” Advances in Neural Information Processing Systems, vol. 35, pp. 4203-4217, 2022.

[21] G. Yang, J. Lei, Z. Zhu, et al., “AFPN: Asymptotic feature pyramid network for object detection,” 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), IEEE, pp. 2184-2189, 2023.

[22] H. Zhang, C. Xu, and S. Zhang, “Inner-IoU: more effective intersection over union loss with auxiliary bounding box” arXiv preprint arXiv:2311.02877, 2023.

[23] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE transactions on pattern analysis and machine intelligence, 2016, 39(6): 1137-1149.

[24] Cai Z, Vasconcelos N. Cascade R-CNN: High quality object detection and instance segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2019, 43(5): 1483-1498.

[25] Redmon J. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[26] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[27] Wang C Y, Yeh I H, Mark Liao H Y. Yolov9: Learning what you want to learn using programmable gradient information[C]//European Conference on Computer Vision. Springer, Cham, 2025: 1-21.

[28] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.

[29] Memari M, Shakya P, Shekaramiz M, et al. Review on the advancements in wind turbine blade inspection: Integrating drone and deep learning technologies for enhanced defect detection[J]. IEEE Access, 2024.

[30] Du Y, Zhou S, Jing X, et al. Damage detection techniques for wind turbine blades: A review [J]. Mechanical Systems and Signal Processing, 2020, 141: 106445.

Construction and Optimization of Multi-Scenario Autonomous Call Rule Models in Emergency Command Scenarios

Weiyan Zheng, Chaoyue Zhu*, Di Huang, Bin Zhou, Xingping Yan, Panxia Chen
Zhejiang Dayou Industrial Co., Ltd. Hangzhou Science and Technology Development Branch, Hangzhou, 310000, China

***Abstract*—In response to the slow processing speed, weak anti-interference, and low accuracy of autonomous call models in current emergency command scenarios, the research focuses on the fire scenario, aiming to improve the emergency response efficiency through technological innovation. The research innovatively integrates digital signal processing algorithm and two-tone multi-frequency signal detection algorithm to develop a hybrid algorithm. Then, a novel autonomous call model based on the hybrid algorithm is constructed. The comparative experimental results indicated that the accuracy of the hybrid algorithm was 0.9 and the error rate was 0.05, which was better than other comparison models. The average accuracy and comprehensive performance score of the model were 0.95 and 97 points, respectively, both of which were better than comparison models. The results confirm that the autonomous call model proposed in this study can accurately and quickly judge emergency scenarios and handle calls, and provide new ideas and theoretical basis for emergency command and rescue of fire and other disasters, with broad application prospects.**

***Keywords*—Digital signal processing algorithm; dual tone multi-frequency signal detection algorithm; fire; autonomous call model**

I. Introduction

The emergency command in the field of emergency rescue faces many challenges, including signal interference, inaccurate prediction, and slow response. These problems are particularly acute during disasters, which may result in delayed rescue operations and wasted resources. With the continuous progress of digital and intelligent technology, the autonomous calling model is increasingly being applied in various industries. In emergency command and rescue, attempts have also been made to introduce autonomous call rule models [1]. Many domestic and foreign scholars have explored the application of autonomous call models. For example, Zardini et al. proposed a new on-demand autonomous call mobile model to solve the existing large transportation demand and road congestion. The results showed that the proposed autonomous call mobile model saved 70% of the travel time [2]. In addition, to address the low efficiency of hazard perception and recognition in autonomous vehicles, Ghosh et al. proposed a machine learning algorithm-based autonomous hazard call model for self-driving vehicles, which was used to test the model in real situations. The results showed that the hazard recognition efficiency of autonomous vehicles based on this model increased by 22.1% [3]. To solve the problem that UAV is difficult to accurately reach the accident site during emergency rescue, Shaheenzh et al. proposed a new autonomous call to air channel model. The results showed that this model had good practical application effects [4]. However, these autonomous call models still have low detection efficiency and weak anti-interference ability, so it is also necessary to optimize the above autonomous call models [5]. Therefore, proposing an autonomous call model that can improve the prediction accuracy and prediction efficiency of emergency scenarios is an urgent problem.

The Analytic Hierarchy Process (AHP) is simple to calculate and has strong applicability, but when dealing with large-scale problems, it is subjective and prone to significant calculation errors [6]. Although Surface Acoustic Wave (SAW) technology can improve the accuracy of signal filtering and enhance anti-interference, it still has high filter insertion loss and low SAW filter performance in the high frequency range. The above two methods are also not applicable to the current autonomous calling model [7].

The Dual Tone Multi-Frequency (DTMF) signal detection algorithm has fast dialing speed, high reliability, and strong anti-interference ability [8]. The high-precision characteristics of the Demand Side Platform (DSP) signal processing algorithm can reconstruct signals and avoid interference from other signals. The flexibility of this algorithm is also beneficial for its processing, analysis, and modification of complex signals [9]. Many scholars have analyzed the above algorithms. Maity et al. designed an improved DTMF algorithm to address the weak noise resistance and low efficiency in signal detection in telecommunications equipment. Comparative experiments were conducted between this improved algorithm and previous algorithms. The results showed that the noise resistance and efficiency were improved by 79% and 87%, respectively [10]. Oluwole et al. proposed a home automation technology based on DTMF to solve the high energy consumption and low transmission speed in home appliance control. The improvement in transmission speed was not significant [11]. In addition, Fan designed a DSP signal processing algorithm to improve the machine learning accuracy. The results showed that the model improved machine learning performance [12]. To improve the computational speed of digital signal processing systems, Seshadri proposed a signal processing algorithm based on DSP. The results demonstrated that this algorithm improved the signal processing speed [13]. Nisha et al. also proposed a denoising method based on DTMF to improve the denoising effect of MRI images. The results showed that the denoising effect of this method was significantly better than that of traditional methods [14].

In summary, some scholars have now analyzed the autonomous calling model. Although the autonomous calling model has been optimized, there are still some problems with the above model. For example, the autonomous call mobile model proposed by Zardini et al. still has low model prediction accuracy. The vehicle hazard autonomous call model proposed by Ghosh et al. still has long computation time-consuming. The autonomous call-to-air channel model designed by Shaheenzh et al. still has slow call speed. The above research shows that the current model has some limitations, such as low detection efficiency and weak anti-interference ability, and needs to be further optimized. To meet this demand, an innovative autonomous call model is proposed, aiming to build a solution suitable for complex emergency scenarios by combining the high precision and flexibility of DSP algorithm with the speed dial, high reliability and strong anti-jamming ability of DTMF signal detection algorithm. This motivation stems from the significant advantages and complementarity of the two in signal processing: DSP algorithms are good at processing and analyzing complex signals, while DTMF technology ensures stable communication in emergency situations. Compared with the subjectivity of AHP and the low filtering performance of SAW technology in the high frequency range, the proposed model shows significant advantages in solving the problems of signal interference, inaccurate prediction and slow response.

The main contribution and influence of the research is that the proposed hybrid algorithm autonomous call model not only makes up for the shortcomings of the existing model in prediction accuracy, response speed and anti-interference ability, but also proves its excellent performance in multi-scenario applications through the experimental verification of actual fire simulation scenarios. This innovation not only provides a more efficient and accurate solution for emergency command and rescue, but also opens up new ideas and methodological references for subsequent research, and is expected to promote the overall progress in the field of emergency rescue. Specifically, the proposed model improves the accuracy and response speed of disaster early warning, optimizes resource allocation, and enhances the reliability and stability of emergency communication, thus minimizing casualties and property losses when disasters occur. The argument of the research is that the disaster autonomous call warning model based on DSP and DTME signal detection technology can improve the accuracy of disaster warning and reduce disaster casualties. The argument is based on the high accuracy and flexibility characteristics of DSP algorithms, as well as the theoretical foundations of high anti-interference, high reliability and fast dialing speed. The contribution of the research lies in the fact that the autonomous call model in emergency scenarios can improve warning and response speed, optimize resource allocation, and enhance emergency communication reliability and stability, thereby reducing the loss of life and property caused by disasters.

II. Methods and Materials

A. DTMF Signal Detection Algorithm Integrated with DSP Algorithm

The current autonomous call rule model in emergency command scenarios has problems such as slow dialing speed and susceptibility to external influences, which seriously affects the timeliness and accuracy of emergency command. Therefore, strengthening the overall performance of the autonomous call rule model is of great significance for improving the effectiveness of emergency command. The DTMF signal detection algorithm has fast dialing speed, high reliability, and strong anti-interference ability. Given these advantages, the DTMF is applied to the multi-scenario autonomous call model to improve the model speed and anti-interference ability, enhancing the emergency command effectiveness. Among them, the signal generation principle in the DTMF signal detection algorithm is shown in Eq. (1) [15].

$$f_{(t)} = A\sin 2\pi f_{(1)}t + A\sin 2\pi f_{(2)}t \quad (1)$$

In Eq. (1), $f_{(1)}$ and $f_{(2)}$ respectively represent any two selected frequencies. A represents the amplitude. t represents the continuous time variable, which represents each each time point at which the signal is generated. The principle of DTMF signal generation is shown in Fig. 1.

In Fig. 1, the principle of DTMF signal generation is as follows. The input information is fed into the oscillator. The high-frequency oscillation signal generated by the oscillator is transmitted to two counters, respectively. When the value in the counter reaches the preset value, the counter inverts the signal to form a low-frequency square wave and then outputs. The low-frequency square wave output is a sinecure, and the amplitude of the square wave is controlled. Then, the processed two signals are transmitted to the signal mixer for signal mixing processing, and finally outputted. The system function of the oscillator in the DTMF signal generation is shown in Eq. (2).

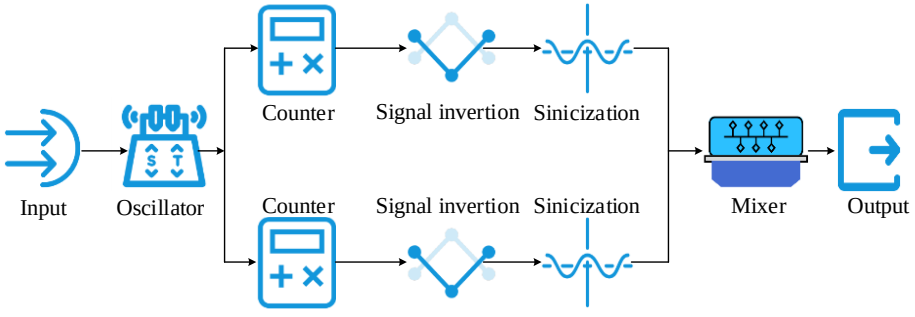


Fig. 1. Schematic diagram of DTMF signal generation.

$$H_{(z)} = \frac{b}{1 + a_1 z^{-1} + a_2 z^{-2}} \tag{2}$$

In Eq. (2), a and z respectively represent row audio signals and column audio signals. b represents the amplitude under the normalized digital frequency, as shown in Eq. (3).

$$\begin{cases} b = A\sin\omega \\ \omega = f_0 / f_a \end{cases} \tag{3}$$

In Eq. (3), f_0 represents the sine wave frequency. f_a represents the sampling frequency. ω is the normalized digital frequency. a_1 and a_2 in Eq. (2) are shown in Eq. (4).

$$\begin{cases} a_1 = -2\cos\omega \\ a_2 = 1 \end{cases} \tag{4}$$

The unit sampling response corresponding to the oscillator is shown in Eq. (5).

$$h(n) = A\sin((n+1)\omega)u(n) \tag{5}$$

In Eq. (5), n represents the sampling point. The difference equation of the oscillator is shown in Eq. (6).

$$g(n) = 2\cos\omega g(n-1) - g(n-2) \tag{6}$$

The signal synthesis in the DTMF signal detection algorithm is shown in Eq. (7).

$$y(n) = A_0 + A_1 \sin\frac{2\pi dt_0}{t_s} + A_2 \sin\frac{2\pi dt_1}{t_s} \tag{7}$$

In Eq. (7), t_0 and t_1 represent the high-frequency and low-frequency of the generated signal, respectively. A_0 and A_1 represent the amplitude of t_0 and t_1. t_s is the sampling frequency. d is the number of sampling points. Afterwards, it is subjected to sinusoidal processing. The sine function is shown in equation (8).

$$\sin(x) = \sin(x1) + \frac{[\sin(x2) - \sin(x1)(x2 - x1)]}{256} \tag{8}$$

In Eq. (8), $x1$ and $x2$ represent two segmentation points, but the level difference between the high-frequency and the low-frequency affects the experimental results. The level difference is shown in Eq. (9).

$$S_H = -20\lg(\frac{V_H}{V_0})S_L (1 < S_H - S_L < 2) \tag{9}$$

In Eq. (9), S_H represents the level of the high-frequency signal. S_L represents the level of the low-frequency signal. V_H and V_L represent high-frequency voltage and low-frequency voltage, respectively. DAC is shown in Eq. (10).

$$DAC(n) = 12.8 \times (1023A_0 + A_1 A + A_2 B) \tag{10}$$

In Eq. (10), A and B are shown in equation Eq. (11).

$$\begin{cases} A = 1023\sin(\frac{2\pi dt_0}{t_s}) \\ B = 1023\sin(\frac{2\pi dt_1}{t_s}) \end{cases} (d = 0,1,2.....) \tag{11}$$

In Eq. (11) and Eq. (12), t_0 and t_1 represent the high-frequency and the low-frequency in DTMF, respectively. d is the number of sampling points. t_s is the sampling frequency. Although the DTMF signal detection algorithm has fast dialing speed and strong reliability, it has high requirements for signal-to-noise ratio, easy signal leakage, and low detection accuracy. The high precision of DSP algorithms makes signal reconstruction possible and avoids interference from other signals. The flexibility of this algorithm is beneficial for its processing, analysis, and modification of complex signals [16]. The basic structure diagram of the DSP algorithm is shown in Fig. 2.

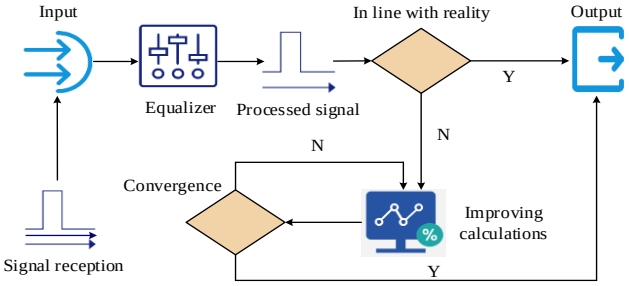


Fig. 2. Basic flowchart of DSP algorithm.

From Fig. 2, the DSP algorithm first receives the signal, and then extracts and processes the received signal through a digital equalizer at the receiving end of the DSP algorithm. The processed signal is compared with the output sequence to determine whether the processing result of the equalizer matches the actual situation. If it matches, the processing result is output. If it does not match, the equalizer cannot converge successfully. The DSP is improved. After improving, it is judged until it matches the actual situation, and the processed signal is output. The transmission calculation for the received signal is shown in Eq. (12).

$$y_a(i) = s(i) \cdot h_a(i) + w(i) \tag{12}$$

In Eq. (12), $y_a(i)$ represents the transmitted signal. $h_a(i)$ represents the time-domain function. The relationship between the transmitted signal and the output signal of the equalizer is shown in Eq. (13).

$$z(j) = \lambda s_a[(j-k)T_s] \tag{13}$$

In Eq. (13), $z(j)$ is a discrete signal sequence. k is the delay generated by the equalizer. λ represents a complex constant. If the signal passed through the equalizer does not match the actual situation, it will be improved. The improvement operation is shown in Eq. (14).

$$J(g) = E[|(\rho(h)|a(j) + a(j-1)|^2 - 1)^2] \tag{14}$$

Each factor in Eq. (14) is shown in Eq. (15).

$$\begin{cases} \rho(h) = [\sin(\frac{\pi h}{2})\sin(\pi h)]^2 \\ E[|(a(j)|^2 - 1)^2] = 0 \\ E[(\rho(h)|a(j) + a(j-1)|^2 - 1)^2] = 0 \end{cases} \quad (15)$$

In Eq. (15), $a(j)$ is the signal output by the equalizer. In order to improve the accuracy and efficiency of the DTMF signal detection algorithm, this study utilizes the high-precision performance of the DSP algorithm to improve the DTMF signal detection algorithm. The flowchart of the improved DTMF signal detection algorithm is shown in Fig. 3.

From Fig. 3, the improved DTMF signal detection algorithm is divided into a DTMF module and a DSP module. In the DTMF module, the input signal is generated into low-frequency and high-frequency signals through an oscillator. Then, the two signals are respectively fed into the counter for signal inversion and sine processing. Finally, the two signals are mixed through the mixer. The mixed signal is input into the DSP module. The signal is extracted, detected, and reconstructed through the equalizer in this module to eliminate other factors and improve detection accuracy. Finally, the signal is output. The fast dialing speed of DTMF is utilized for dialing operation through signal judgment. The DSP algorithm is used in DTMF signal detection. When the two signals are fused, the signal needs to be resampled before the DSP module. The update iteration is shown in Eq. (16).

$$l(k+1) = l(k+1) - J(g) \cdot \mu_g \nabla_g \quad (16)$$

In Eq. (16), μ_g represents the iteration step size. $l(k)$ represents the tap coefficient vector. At this point, $y_a(i)$ represents the signal received from the DTMF mixer.

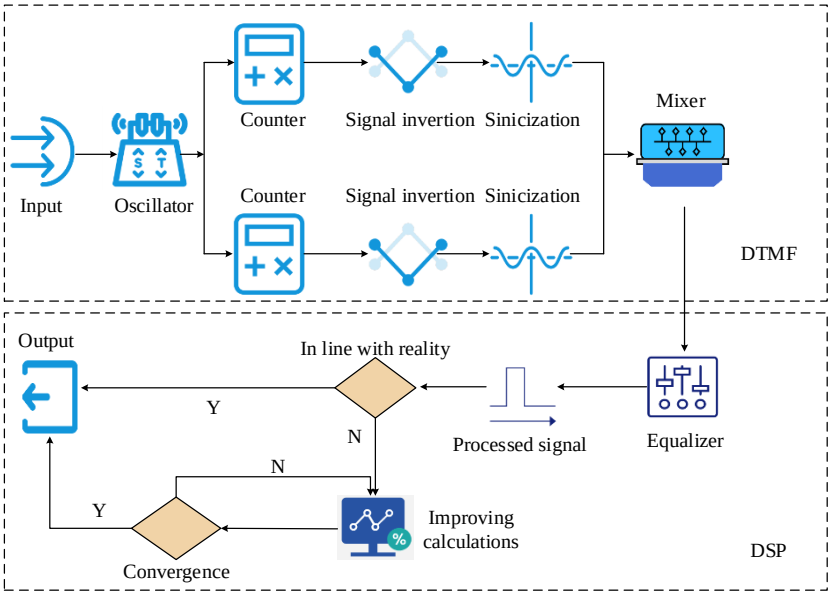


Fig. 3. Improved DTMF signal detection algorithm.

B. Construction of Improved Autonomous Call Model Based on DSP-DTMF Algorithm

In emergency command scenarios, constructing a multi-scenario autonomous call rule pattern model is crucial. This model needs to comprehensively consider multiple factors such as the urgency of the event, resource types, geographical location, etc. to better improve the effectiveness of emergency command. Firstly, the multi-scenario autonomous call rule model can set call priority based on the urgency of the event, ensuring timely response to important events. Secondly, the model needs to intelligently identify and allocate relevant emergency resources, such as personnel and materials, to achieve efficient disposal. In addition, the model can automatically select the nearest emergency team and resources to call based on the event location. When constructing this model, holographic and analogical modeling methods can be used to clarify user, usage scenario, and entity features, simplifying unnecessary dimensions and attributes. At the same time, it is necessary to achieve autonomous calling in multiple scenarios by setting flexible call rules and algorithms. This study takes fire emergency command as an example to construct a multi-scenario autonomous call rule model. The basic framework of the constructed model is shown in Fig. 4.

In Fig. 4, the fire emergency command multi-scenario autonomous call model is divided into signal perception layer, signal transmission layer, and signal receiving layer. The signal perception layer mainly includes smoke detectors and temperature detectors, which are used to detect smoke concentration and temperature in fires. When the detection result meets the alarm conditions of the detector, the signal will be sent to the alarm in the signal transmission layer. The alarm sends the next command to the signal receiving layer based on the smoke concentration and temperature. The signal receiving layer receives the sent reminder signal or alarm signal. If a reminder signal is received that there is someone inside the room, it indicates that the indoor fire is not serious, and can be

dealt with without calling the alarm. If the signal received by the receiving end is an alarm signal, it indicates that the fire is serious or not serious, but there is no one indoors. The model can automatically call the alarm number. Although the fire emergency command multi-scenario autonomous call model can predict and call fires autonomously, it has a series of problems such as slow detection speed, low accuracy, weak anti-interference ability, and slow call transmission speed. The DSP-DTMF algorithm can improve the accuracy and speed of signal detection, enhance the anti-interference ability, and reduce the response time [17]. Therefore, this study uses the DSP-DTMF algorithm to optimize the traditional fire emergency command multi-scenario autonomous call model, in order to improve the overall performance of the model. The basic structure diagram of the autonomous call mode model that integrates the DSP-DTMF algorithm is shown in Fig. 5.

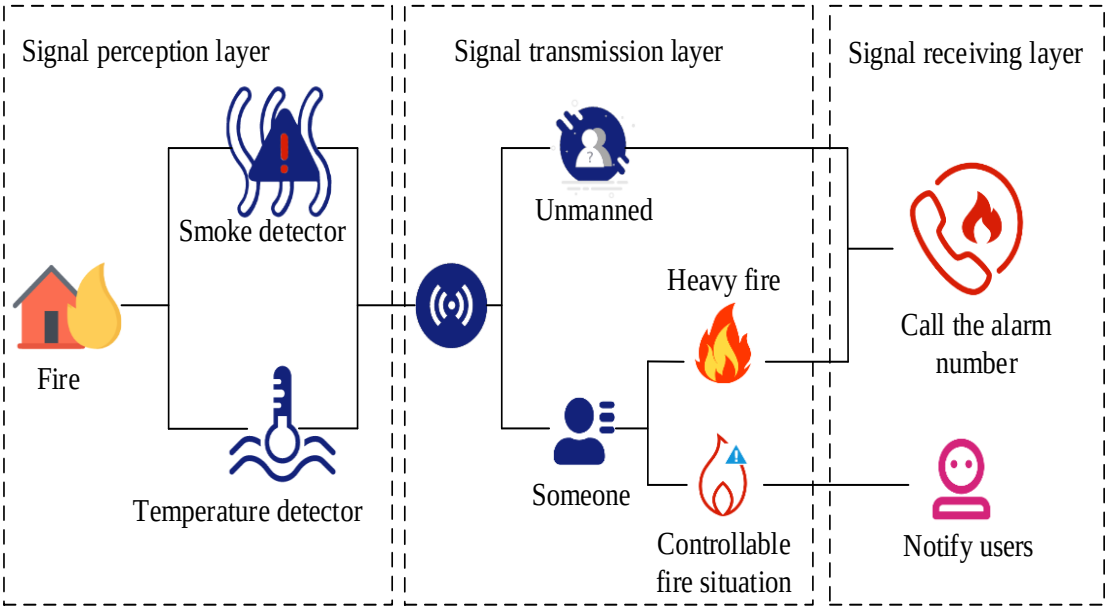


Fig. 4. Basic framework of autonomous call rule model.

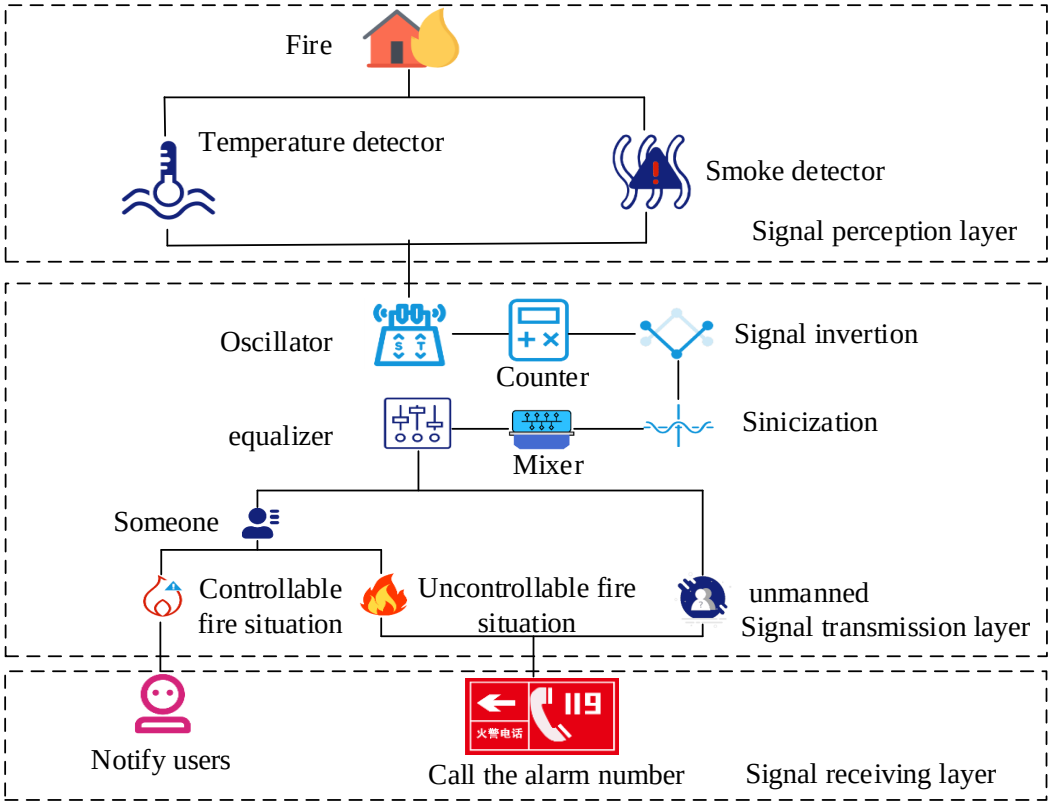


Fig. 5. Improved autonomous call model.

In Fig. 5, the improved autonomous call model is also divided into three layers: signal perception layer, signal transmission layer, and signal receiving layer. The signal perception layer and signal receiving layer have not changed, but only the signal transmission layer has changed. In the signal perception layer, after detecting relevant signals through temperature detectors and smoke concentration detectors, the signals are sent to the signal transmission layer. In the signal transmission layer, the signal is detected by the oscillator, counter, and mixer of the DTMF module. Then, the detected signal is sent to the DSP module, which extracts and re-detects the signal through the equalizer in the module. If the detected signal shows that there is no one present at the fire scenario, an alarm will be triggered. DTMF will call the alarm number. If someone is trapped on site and the fire is controllable, the model will notify rescue personnel. If the fire is uncontrollable, an alarm will be triggered. The operation instructions for fire alarm notification to users are transmitted to the signal receiving layer through dial-up processing. The comprehensive performance score of the improved autonomous call model is shown in Eq. (17).

$$W_k = \sum_{k=1}^{n} x_k * y_k \tag{17}$$

In Eq. (17), W_k represents the comprehensive score of the k-th indicator. x_k represents the weight of the k-th indicator. y_k represents the rating of the k-th indicator. This calculation method can be used to compare the comprehensive performance of different models.

III. Results

A. Performance Analysis of DTMF Signal Detection Algorithm Based on DSP

MIMO-OTFS algorithm is a signal detection technology based on Multi-Input Multi-Output (MIMO) and Orthogonal Time-Frequency Space (OTFS), which performs well in dealing with complex signal environments [18]. SDNSR-Net is a deep learning network specifically designed for signal detection and noise suppression [19]. In order to verify the superiority of the improved DTMF signal detection algorithm, DSP-DTMF algorithm (Algorithm 1) is compared with MIMO-OTFS (Algorithm 2) and SDNSR-Net (Algorithm 3) in experiments. Signal detection accuracy, error rate, transmission speed, and minimum detectable signal are used as comparison indicators. The specific environment of the comparative experiment is shown in Table I. The experiment is repeated 10 times, and T-test is used for statistical verification.

The parameter settings during the experiment are as follows. The sampling frequency of each algorithm for the signal is set to 8000HZ, the duration of each digital signal is set to 50ms to ensure the accuracy of the digital information, and the output frequency is set to 100Hz. Comparative experiments are conducted on three algorithms under the same environmental configuration in Table I. The dataset used is the radar radiation source recognition signal dataset, which is sourced from the measured data of AD9910 and USRP hardware. The dataset mainly consists of 6 individuals, each with 6 modulation types and 5000 pulses per modulation type, totaling 180000 samples [20]. The waveform comparison results of the detection accuracy of three algorithms on the radar radiation source recognition signal dataset are shown in Fig. 6.

TABLE I. Experimental Environment Configuration

System modules	Parts	Type
Single chip	Storage	8KB Flash storage
	RAM	256 bit
	I/0 line	32 programmable I/O lines
	Interface type	Serial interface
	Oscillator	Clock oscillator
Number receiving circuit	Chip type	MT8870
	Type of micro-controller	89C52
Signal circuit	Chip type	UM91513
Main engine	Winds system	Wind11-64
	CPU model	i7-12700KF

As shown in Fig. 6 (a), the average accuracy of the DSP-DTMF algorithm reached 0.91, which fluctuated between 0.8 and 1.0, indicating good stability. In Fig. 6 (b), the average accuracy of Algorithm 2 was 0.62, which fluctuated greatly when the sample size was less than 60, resulting in unstable accuracy. After the sample size exceeded 60, the accuracy of the algorithm fluctuated between 0.57 and 0.64, which was relatively stable. According to Fig. 6 (c), the average accuracy of Algorithm 3 was 0.46. When the sample size was less than 100, the accuracy fluctuated greatly, with poor stability. When the sample size reached 100, the accuracy gradually stabilized. The error rate and recall rate of the three algorithms are shown in Fig. 7.

As shown in Fig. 7 (a), the error rates of Algorithm 1, Algorithm 2, and Algorithm 3 stabilized at 0.05, 0.10, and 0.13, respectively. Algorithm 1 reached a maximum of 0.15 when the sample size was 40. When the sample size exceeded 40, the error rate gradually decreased and stabilized at 0.05. Algorithm 2 and Algorithm 3 were 0.21 and 0.22, respectively when the sample size was 60. When the sample size was greater than 80, the error rates of the two algorithms decreased and eventually stabilized at 0.15 and 0.13, respectively. As shown in Fig. 7 (b), the recall rates of Algorithm 1, Algorithm 2, and Algorithm 3 ultimately stabilized at 0.90, 0.79, and 0.68, respectively. The recall rate of Algorithm 1 gradually increased with the increase of sample size. The overall recall rate of Algorithm 2 increased. When the sample size was less than 100, the fluctuation range of error rate was large, with poor stability. The overall recall rate of Algorithm 3 also increased. When the sample size was less than 100, the recall rate was extremely unstable, with a large fluctuation range and frequency. After the sample size exceeded 100, the recall rate gradually stabilized. The signal detection transmission speed and minimum detectable signal are experimentally analyzed. The experimental results are shown in Fig. 8.

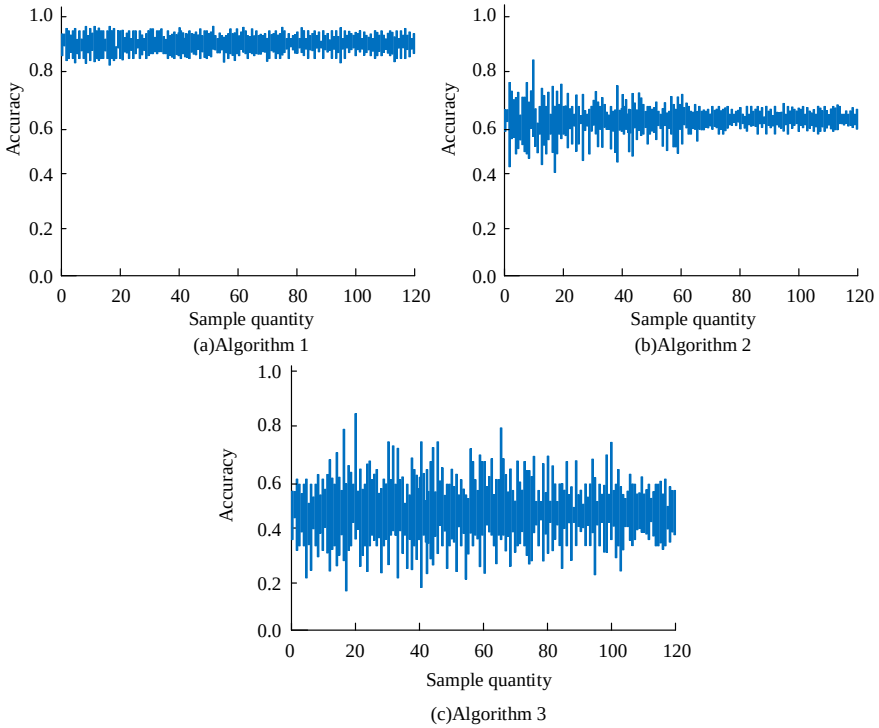


Fig. 6. Comparison of accuracy among three algorithms.

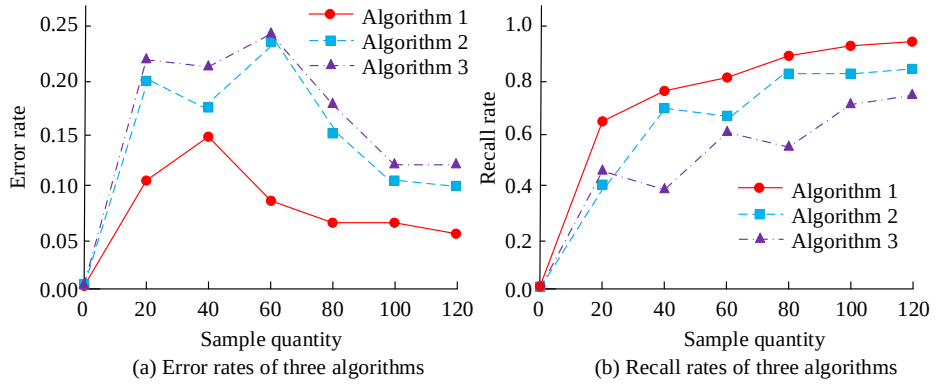


Fig. 7. Comparison of error rate and recall rate of three algorithms.

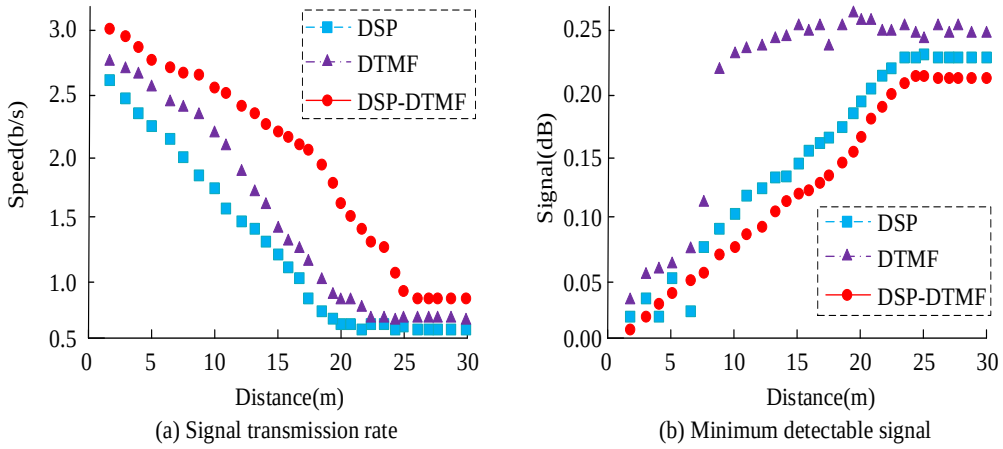


Fig. 8. Signal transmission rates and minimum detectable signals of three algorithms.

From Fig. 8 (a), the signal transmission rates of Algorithm 1, Algorithm 2, and Algorithm 3 all decreased with increasing distance. Among them, the signal transmission rate of Algorithm 1 slowed down and tended to increase when the distance reached 15m. When the distance reached 25m, the signal transmission rate reached its lowest value, at 0.8b/s, and then fluctuated around 0.8b/s. Algorithm 2 and Algorithm 3 reached the minimum transmission rate at 25m, which was 0.53b/s and 0.67b/s, respectively. As shown in Fig. 8 (b), as the distance increased, the minimum detectable signal values of all three algorithms increased. At 25m, Algorithm 1 and Algorithm 2 had minimum detectable signals of 0.19dB and 0.21dB, respectively. At 10m, the minimum detection signal of Algorithm 3 skyrocketed to 0.23. Subsequently, it fluctuated between 0.22 and 0.24. In summary, the DSP-DTMF algorithm proposed in the study has the best overall performance among the three algorithms. In order to more comprehensively verify the effectiveness and scalability of DSP-DTMF algorithm and its autonomous call model, it is tested with the model proposed in literature [13] and the model proposed in literature [14] in a variety of data sets. In addition to the radar radiation source identification signal dataset, several other key datasets were introduced to evaluate the algorithm's performance in different scenarios. The radar radiation source identification signal dataset comes from the measured data of AD9910 and USRP hardware, which contains 6 individuals, each individual has 6 modulation types, each modulation type contains 5000 pulses, a total of 180,000 samples; Fire simulation data set is a data set generated by simulating fire scenarios, which contains fire data under different temperature and smoke concentration conditions. The multi-source information fusion dataset synthesizes data from video, audio, sensor and other sources, and contains multi-modal data under various emergency scenarios. The performance of the three models on different data sets is shown in Table II.

TABLE II. THE PERFORMANCE OF THE THREE MODELS ON DIFFERENT DATA SETS

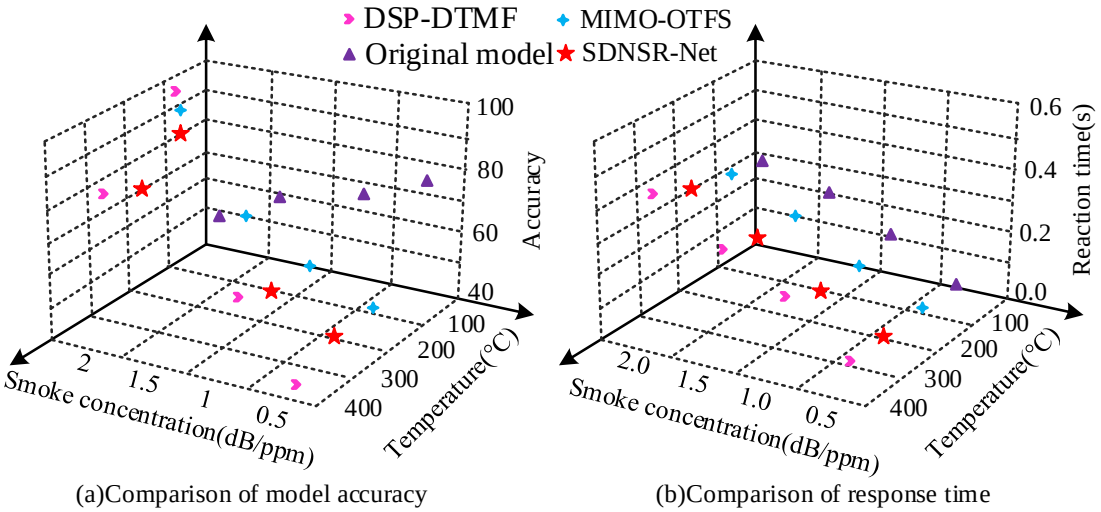
Model type	Data set	Signal detection accuracy	Error rate	Transmission speed (b/s)	Minimum detectable signal (dB)
DSP-DTMF	Radar radiation source identification signal dataset	0.91	0.05	0.81	0.19
	Fire simulation data set	0.93	0.04	0.77	0.18
	Multi-source information fusion dataset	0.89	0.06	0.66	0.22
Literature [13]	Radar radiation source identification signal dataset	0.72	0.10	0.53	0.21
	Fire simulation data set	0.71	0.12	0.48	0.23
	Multi-source information fusion dataset	0.69	0.13	0.45	0.26
Literature [14]	Radar radiation source identification signal dataset	0.77	0.13	0.67	0.23
	Fire simulation data set	0.68	0.15	0.61	0.25
	Multi-source information fusion dataset	0.73	0.17	0.56	0.24

As can be seen from Table II, the model proposed in this study has high signal inspection accuracy in the three data sets, which are 0.91, 0.93 and 0.89 respectively. Through testing on different data sets, the research proves that DSP-DTMF algorithm and its autonomous call model have good scalability. In addition, this result also proves that the proposed model has better performance than other studies. The above results show that the algorithm and model can adapt to various emergency scenarios and have wide application potential.

B. Performance Analysis of the Optimized Autonomous Call Rule Model

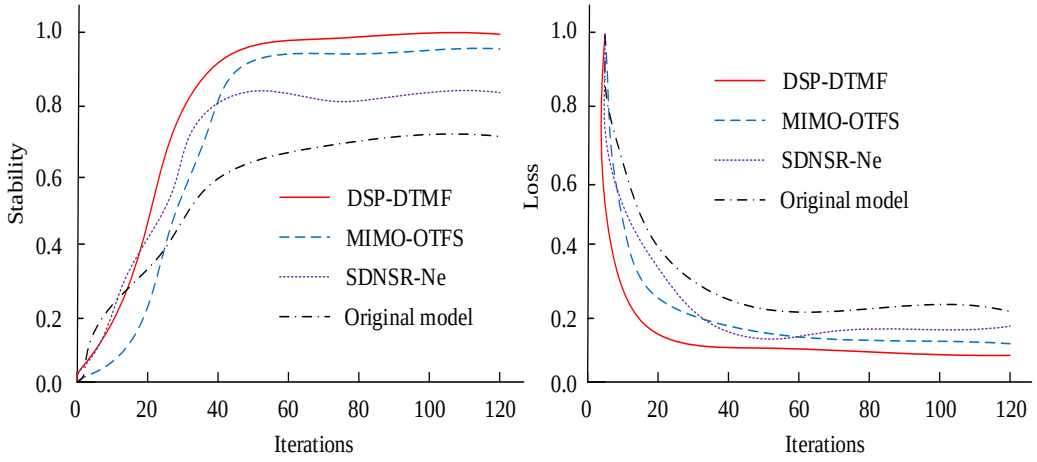
After verifying the superiority of the DSP-DTMF algorithm, in order to analyze the application effect of the proposed multi-scenario autonomous call rule model based on the DSP-DTMF algorithm, the DSP-DTMF algorithm, MIMO-OTFS algorithm, and SDNSR-Net algorithm are used in the autonomous call rule model for comparison. Taking the fire emergency scenario as an example, different models are used in the fire simulation environment to compare the comprehensive performance of the three models and the model without using the proposed algorithm. The accuracy and response comparison of the four models are shown in Fig. 9.

According to Fig. 9 (a), the average accuracy of the DSP-DTMF model, MIMO-OTFS model, SDNSR-Net model, and original model were 0.93, 0.79, 0.68, and 0.61, respectively. When the temperature was 200°C and the smoke concentration was 0.5dB/ppm, the accuracy of the DSP-DTMF, MIMO-OTFS, SDNSR-Net, and original model were 0.97, 0.90, 0.81, and 0.72, respectively. When the temperature and smoke concentration increased, the accuracy of all four models decreased. The DSP-DTMF model had an accuracy of 0.94 at 400°C and 2dB/ppm smoke concentration, while the accuracy of the MIMO-OTFS, SDNSR-Net, and original model under these conditions were 0.76, 0.65, and 0.58, respectively. According to Fig. 9 (b), the average response time of the four models was 0.08s, 0.12s, 0.23s, and 0.31s, respectively. The DSP-DTMF, MIMO-OTFS, SDNSR-Net, and original model had response time of 0.05s, 0.08s, 0.14s, and 0.20s, respectively at 200°C and 0.5dB/ppm. As the temperature and smoke concentration increased, the response time of the four models also gradually increased. Moreover, when the temperature was 400°C and the smoke concentration was 2dB/ppm, the response time of four models was 0.10s, 0.34s, 0.40s, and 0.42s, respectively. Fig. 10 shows the stability and loss function curves of four models.



(a)Comparison of model accuracy (b)Comparison of response time

Fig. 9. Comparison of accuracy and response time of four models.

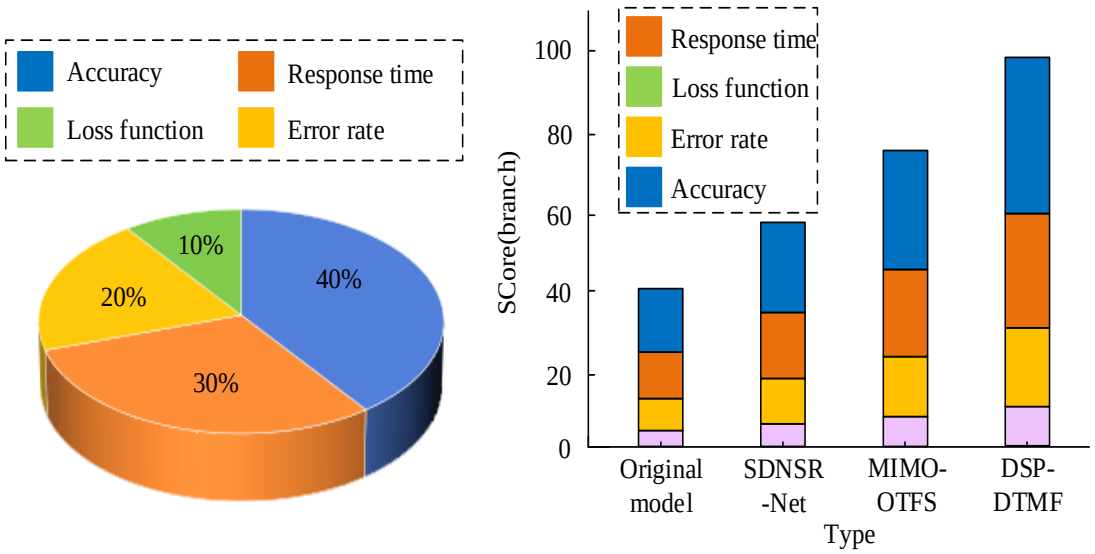


(a)Stability of Four Models (b)Loss function curves of four models

Fig. 10. Comparison of stability and loss function curves for four models.

From Fig. 10 (a), the stability increased with increasing iterations. The final stability of the DSP-DTMF model, MIMO-OTFS model, SDNSR-Net model, and original model was 0.93, 0.91, 0.87, and 0.63, respectively. The stability of all four models reached its maximum at an iteration of 40. From Fig. 10 (b), the loss values of all four models decreased with increasing iterations. Among them, the DSP-DTMF model, MIMO-OTFS model, SDNSR-Net model, and original model had final loss values of 0.07, 0.10, 0.16, and 0.21, respectively. Moreover, the loss values of all four models reached the lowest value at 20 iterations. Fig. 11 shows the comprehensive performance score of the four models.

From Fig. 11 (a), the comprehensive performance score is composed of the accuracy, response time, loss function, and error rate of the model. In the comprehensive performance score, the accuracy of the model accounts for the largest proportion of 40%, the loss function accounts for the smallest proportion, at 10%, and the proportion of response time and error rate is 30% and 20%, respectively. The comprehensive performance score of the model can be calculated from the proportion in Fig. 11 (a) to obtain Fig. 11 (b). The comprehensive score consisted of four parts, among which the DSP-DTMF model had the highest comprehensive score of 97 points, MIMO-OTFS model and SDNSR-Net model had a comprehensive score of 69 points and 60 points, respectively, and the original model had the lowest comprehensive score of 40 points. In summary, the comprehensive performance of the DSP-DTMF autonomous call model proposed in this study is the best. In order to more comprehensively verify the effect of the autonomous call rule model based on DSP-DTMF algorithm, this model and the novel multi-source information autonomous call rule model, swarm intelligence autonomous call rule model and adaptive autonomous call model are tested in five actual fire scenarios. This comparison is to further verify the performance of the optimization algorithm and ensure the effectiveness and reliability of the model in emergency scenarios. The test results are shown in Table III.



(a)The comprehensive performance score of the model and the proportion of each function

(b)Comprehensive performance evaluation of the model

Fig. 11. Comparison of comprehensive performance scores of four models.

TABLE III. PERFORMANCE OF DIFFERENT MODELS IN REAL FIRE SCENARIOS

Index	DSP-DTMF autonomous call rule model	Multi-source information autonomous call rule model	Swarm intelligence autonomous call rule model	Adaptive autonomous call model
Accuracy rate (%)	95.2%	90.1	89.6	88.1
Reaction time (s)	2.1	3.3	3.1	4.5
Number of scenarios	5	5	5	5
Total test duration (h)	100.1	98.8	98.9	95.6
Average frames per second (FPS)	30.1	24.2	25.1	20.6

From Table III, the autonomous call rule model based on DSP-DTMF algorithm performed well in the simulated fire scenario, with an accuracy of 95.2% and a reaction time of only 2.1s, which was far better than comparison models. It was tested in five different real scenarios, and the total test time reached 100.1h, showing the stability and durability of the model. In addition, from Table III, the average FPS of the autonomous call rule model based on DSP-DTMF algorithm was 30.1, which was better than 24.2, 25.1, and 20.6 of comparison models, indicating that the algorithm had higher processing speed. Finally, the paper also considers the scalability and robustness of the autonomous call rule model based on DSP-DTMF algorithm in many different scenarios. The robustness of the model is analyzed by comparing the accuracy and response time of the model in five specific scenarios. The specific results are shown in Table IV.

From Table IV, under different actual scenarios, DSP-DTMF algorithm maintained a high accuracy rate, and its mean value was 94.9%. In addition, the response time of the model fluctuated slightly in different environments, but the overall level was also kept low, with a mean response time of 2.2s. The above results show that the autonomous call rule model based on DSP-DTMF algorithm has scalability and robustness. The performance of the DSP-STMF autonomous call model is compared with the widely used disaster autonomous call model based on GA-SVM algorithm, as shown in Table V.

TABLE IV. ACCURACY AND RESPONSE TIME OF THE AUTONOMOUS CALL RULE MODEL BASED ON DSP-DTMF ALGORITHM IN DIFFERENT SCENARIOS

/	Accuracy rate	Response time (s)
Residential area fire	96.1%	1.9
Industrial area fire	94.5%	2.2
Forest fire	93.8%	2.4
Commercial fire	95.2%	2.0
Mountain fire	94.8%	2.3
Mean value	94.9%	2.2

TABLE V. MODEL PERFORMANCE COMPARISON

Model	DSP-DTMF model	GA-SVM model
Precision	95.2%	87.6%
Reaction time	2.1s	3.5s
Monitoring accuracy	97.5%	82.1%
False negative rate	1.1%	8.3%
False report rate	1.4%	9.6%
Fraction of coverage	98.9%	89.7%

According to Table V, the proposed DSP-STMF model outperformed the current GA-SVM autonomous call model in all aspects of performance. The monitoring accuracy, precision, and coverage of the DSP-DTMF autonomous call model were all above 95%, while the monitoring accuracy, precision, and

coverage of the GA-SVM model were all below 90%. The false negative rate, false alarm rate, and response time of the DSP-DTMF model were lower than those of the GA-SVM model. From the above results, it can be concluded that the DSP-DTMF autonomous call model proposed in the study outperforms current autonomous call models in all aspects of performance.

IV. Discussion

This study designed a comparative experimental analysis on the performance of the DSP-DTMF algorithm. Then, comparative experiments are conducted on autonomous call models based on DSP-DTMF algorithm, MIMO-OTFS algorithm, and SDNSR-Net algorithm. The results showed that the DSP-DTMF algorithm outperformed the other two algorithms in accuracy, stability, error rate, and signal transmission speed. In the accuracy waveform, the average accuracy of the DSP-DTMF algorithm was the highest at 0.9, with small fluctuation in accuracy and strong stability. This result was similar to the experimental results of Schaumont using the DSP-DTMF algorithm to process remote course selection for courses [21]. This result indicates that in practical applications, the DSP-DTMF algorithm can make more accurate judgments on emergency scenarios. In the recall and error rate curves of the algorithm, the DSP-DTMF algorithm had the lowest error rate of 0.05 and the highest recall rate of 0.9, further verifying the superiority of the algorithm. In terms of the signal transmission rate and minimum detectable signal, the DSP-DTMF algorithm outperformed the other two algorithms. Wibowo et al. also had similar conclusions [22]. This result indicates that the DSP-DTMF algorithm can more quickly and accurately determine emergency scenarios in practical applications.

Secondly, all three algorithms were applied to the autonomous call model. Through comparative experimental analysis between the three models and the original model, it was found that the autonomous call model based on the DSP-DTMF algorithm had strong advantages in accuracy, response time, stability, loss function, and comprehensive performance evaluation. In terms of accuracy and response speed in emergency scenario assessment, this model accurately assessed the emergency scenarios under high temperature and high concentration smoke, and made timely next steps. In the comparison of model stability and loss function curve, it was found that the DSP-DTMF model had the strongest stability of 0.93 and the smallest loss function. This result indicates that the model has strong anti-interference ability, which is not easily affected by other external factors in emergency scenarios. This model is relatively accurate in predicting emergency scenarios. Perng also conducted similar conclusions in the research on DSP digital filters [23]. In the comprehensive performance evaluation of the model, the DSP-DTMF model had the highest comprehensive score of 97 points, which was much higher than other models. This coincided with the conclusion on an automatic unlocking system based on improved DTMF proposed by Iwuji [24].

This result fully demonstrates that the autonomous call model based on DSP-DTMF algorithm can effectively predict emergency scenarios and respond quickly to them, meeting sudden emergency needs. The DSP-DTME autonomous call model is used to test the performance of the model in different fire actual scenario environments. The test results show that the DSP-DTMF autonomous call model maintains a high accuracy rate in different fire scenarios, and its average accuracy rate reaches 94.9%, and the average response time is only 2.2s. It can be concluded that the disaster autonomous call model based on DSP-STME algorithm can improve the speed of disaster emergency rescue, reduce the economic and property losses caused by disasters, and protect the safety of people's lives and properties.

V. Conclusion

Aiming at the problems of slow processing speed, weak anti-interference and low accuracy of autonomous call model in emergency command scenario, this paper innovatively integrates DSP and DTMF to construct an efficient hybrid algorithm, and designs a new autonomous call model based on this algorithm. Through a series of comparative experiments, the superiority and practical application value of DSP-DTMF algorithm and its autonomous call model are verified. The main contribution of the research is to propose a new DSP-DTMF hybrid algorithm, which is successfully applied to the autonomous call model in the emergency command scenario. The algorithm not only significantly improves the accuracy and efficiency of signal detection, but also enhances the anti-interference ability of the model, so that it can maintain stability and accuracy in complex and changeable emergency scenarios. In addition, the experimental results of the autonomous call model designed based on the algorithm in the fire simulation scene show that its comprehensive performance score is much higher than other comparison models, which proves the effectiveness and practicability of the model in emergency command.

In practical applications, the autonomous call model based on DSP-DTMF algorithm has significant advantages. First of all, the model can judge the situation in the emergency scenario more quickly and accurately, so as to start rescue operations in time and reduce disaster losses. Secondly, the model has strong anti-interference ability, can maintain stable operation in complex and changeable emergency environment, and improve the reliability and stability of emergency command. In addition, the model has good scalability and robustness, can adapt to the emergency needs in different scenarios, and provide comprehensive technical support for emergency command.

Although this study has achieved certain results, there are still some limitations. First of all, the current research mainly focuses on the optimization of autonomous call model in fire scenarios, and its applicability to other disaster types (such as earthquake, flood, etc.) needs to be further verified. Second, although the performance of the model has been improved, there may still be some limitations in extremely complex or specific emergency scenarios. In addition, due to the limitations of experimental conditions, some data in the study may have certain biases, which may affect the accuracy of the results.

In view of the above limitations, future research can be expanded and deepened from the following aspects. Firstly, the applicability of DSP-DTMF algorithm under different disaster types is further verified, and the model is optimized and

improved according to the characteristics of different disaster types. Second, explore fusing multimodal data into the model to achieve more comprehensive emergency scenario perception and more accurate call decisions. In addition, the real-time and robustness of the model can be enhanced to improve its response ability in complex emergency scenarios. Finally, the possibility of cross-domain application of the model in intelligent transportation, security monitoring and other fields can be explored to further expand its practical application value. Through these efforts, it is expected to further improve the efficiency and accuracy of emergency command, and provide more powerful technical support for disaster relief work.

ACKNOWLEDGEMENT

The research is supported by: Zhejiang Dayou Group Science and Technology project, "Research on the Application of Artificial Intelligence in Emergency Command Scenarios for Power Maintainment", (Project Number: DY2023-08).

REFERENCES

[1] Bandewad G, Datta K P, Gawali B W, Pawar, S. N. Review on Discrimination of Hazardous Gases by Smart Sensing Technology. Artificial Intelligence and Applications. 2023, 1(2): 86-97.

[2] Zardini G, Lanzetti N, Pavone M, Frazzoli E. Analysis and control of autonomous mobility-on-demand systems. Annual Review of Control, Robotics, and Autonomous Systems, 2022, 5(9): 633-658.

[3] Ghosh S, Zaboli A, Hong J, Kwon J. An integrated approach of threat analysis for autonomous vehicles perception system. IEEE Access, 2023, 11(6): 14752-14777.

[4] Shaheen M M Z, Amer H H, Ali N A. Robust air-to-air channel model for swarms of drones in search and rescue missions. IEEE Access, 2023, 11: 68890-68896.

[5] Chelbi N E, Gingras D, Sauvageau C. Worst-case scenarios identification approach for the evaluation of advanced driver assistance systems in intelligent/autonomous vehicles under multiple conditions. Journal of Intelligent Transportation Systems, 2022, 26(3): 284-310.

[6] Das D, Datta A, Kumar P, Kazancoglu Y, Ram M. Building supply chain resilience in the era of COVID-19: An AHP-DEMATEL approach. Operations Management Research, 2022, 15(1): 249-267.

[7] Liu X, Chen X, Yang Z, Xia H, Zhang C, Wei X. Surface acoustic wave based microfluidic devices for biological applications. Sensors & Diagnostics, 2023, 2(3): 507-528.

[8] Teixeira G M. An Overview of the mechanism-based thermomechanical fatigue method (DTMF) in the design of automotive components. Procedia Structural Integrity, 2024, 57(5): 670-691.

[9] Wang Y, Li C, Liu C,Liu, S. Advancing DSP into HPC, AI, and beyond: challenges, mechanisms, and future directions. CCF Transactions on High Performance Computing, 2021, 3(6): 114-125.

[10] Maity A, Prakasam P, Bhargava S. Machine learning based KNN classifier: towards robust, efficient DTMF tone detection for a Noisy environment. Multimedia Tools and Applications, 2021, 80(7): 29765-29784.

[11] Oluwole A S, Odekunle O P, Olubakinde E. Applications and recent development of DTMF based technology in home automation. European Journal of Electrical Engineering and Computer Science, 2021, 5(3): 60-67.

[12] Fan Q, Lu C, Lau A P T. Combined neural network and adaptive DSP training for long-haul optical communications. Journal of lightwave technology, 2021, 39(22): 7083-7091.

[13] Seshadri R, Ramakrishnan S, Kumar J S. Knowledge-based single-tone digital filter implementation for DSP systems. Personal and Ubiquitous Computing, 2022, 26(2): 319-328.

[14] Nisha B, Jose M V. DTMF: Decision-based Trimmed Multimode approach Filter for denoising MRI images. Soft Computing, 2023, 23(12): 1-16.

[15] Das N, Shakil M D T, Mehedy M, Rahman, M. Designing and Implementation of low-cost DTMF Based remotely controlled irrigation system with interactive voice response in Bangladesh. International Journal, 2022, 18(1): 63-72.

[16] Liu T, Banerjee S, Yan H, Mou, J. Dynamical analysis of the improper fractional-order 2D-SCLMM and its DSP implementation. The European Physical Journal Plus, 2021, 136(5): 506-510.

[17] Wang K, Cai W, Zhang Y, Hao, H. Numerical simulation of fire smoke control methods in subway stations and collaborative control system for emergency rescue. Process Safety and Environmental Protection, 2021, 147(12): 146-161.

[18] Qu H, Liu G, Imran M A, et al. Efficient channel equalization and symbol detection for MIMO OTFS systems. IEEE Transactions on Wireless Communications, 2022, 21(8): 6672-6686.

[19] Li L, Hu J. Fast-converging and low-complexity linear massive MIMO detection with L-BFGS method. IEEE Transactions on Vehicular Technology, 2022, 71(10): 10656-10665.

[20] Yang H, Zhang H, Wang H, Guo Z. A novel approach for unlabeled samples in radiation source identification. Journal of Systems Engineering and Electronics, 2022, 33(2): 354-359.

[21] Schaumont P. Socially-Distant Hands-On Labs for a Real-time Digital Signal Processing Course. Proceedings of the 2021 on Great Lakes Symposium on VLSI. 2021, 22(7): 425-430.

[22] Wibowo F W, Wihayati W. Goertzel Algorithm Design on Field Programmable Gate Arrays For Implementing Electric Power Measurement. International Conference on Computer Science, Information Technology and Engineering (ICCoSITE). IEEE, 2023, 23(9): 489-494.

[23] Perng J W, Hsieh T L, Guo C Y. A novel dentary bone conduction device equipped with laser communication in DSP. Sensors, 2021, 21(12): 4229-4233.

[24] Iwuji P C, Idajor J A. Automated GSM-based un-locker system. World Journal of Chemical and Pharmaceutical Sciences, 2020, 1(2): 44-51.

Enhancing User Comfort in Virtual Environments for Effective Stress Therapy: Design Considerations

Farhah Amaliya Zaharuddin[1], Nazrita Ibrahim[2], Azmi Mohd Yusof[3]
Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia[1]
College of Computing and Informatics, Universiti Tenaga Nasional (UNITEN), Putrajaya Campus,
Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia[2, 3]

***Abstract*—Mental stress has emerged as a widespread concern in modern society, impacting individuals from diverse demographic backgrounds. Therefore, exploring effective methods for therapy, such as virtual environments tailored for stress management, is vital for advancing mental health and improving coping strategies. Prioritising user comfort in the design of virtual environments is essential for enhancing their efficacy in alleviating stress. By considering four design aspects of virtual environments that influence user comfort: (i) visual clarity, (ii) safety features, (iii) cognitive preparedness, and (iv) social support, this study intends to (i) evaluate the effectiveness of these four user-centered design elements in facilitating stress reduction and (ii) explore the underlying rationale behind their stress-reducing properties. This study utilised a mixed-methods approach comprising (i) experiments, (ii) questionnaires, and (iii) interviews. Following evaluation with the Depression Anxiety Stress Scale (DASS), 40 participants (10 men and 30 females) were chosen from the 55 healthy adults aged 20 to 60 who volunteered for the study. The findings validated the efficacy of all four design aspects in enhancing users' comfort during therapeutic sessions in virtual environments. This study offers important insights not only into the importance of user-centered design in creating virtual environments for stress management, where comfort markedly improves therapy outcomes but also contributes valuable knowledge to the fields of mental health and human-computer interaction, paving the way for further exploration of innovative therapeutic solutions for mental stress.**

Keywords*—*Virtual environment design; virtual reality; stress therapy; user comfort

I. Introduction

Mental stress is now more prevalent than ever, affecting different demographic groups in several nations around the world [1], [2], [3], [4]. Afterwards, traditional therapeutic methods, for example, breathing exercises, progressive muscle relaxation, and guided imagery are often used to treat psychological pressure [5], [6]. Although these traditional therapeutic methods have been effective, they are also associated with several constraints such as low immersion, inadequate engagement, and high levels of contextual distractions. So, virtual reality (VR) technology; as one of the recent new devices has been applied in stress therapy sessions due to its advantages such as being more immersive, attractive, and less environmental distractions [7], [8]. Studies supporting the practical application of VR as therapeutic mechanisms have largely produced positive results, with users self-reporting decreased levels of stress following sessions [9], [10], [11], [12], [13].

Although these effects are positive, literature on the effectiveness of VR for stress therapy (e.g., [14], [15], [16]), focuses much more heavily on whether or not these systems can reduce stress, but leave out a discussion on design factors that are fundamental for the virtual context which enables this process of stress reduction to take place [17], [18]. Following that, [18] suggested eight design requirements for psychological therapy VR applications, whereby one of them is the importance of creating meaningful therapeutic spaces, underscoring the need to develop virtual environments that foster an appropriate therapeutic milieu and address users' emotional needs.

Furthermore, the studies in [19] and [20] in their scoping review of VR environments for stress reduction and management have similarly suggested that more attention should be given to the virtual environment design to promote stress reduction according to user-specific requirements. Their claim is consistent and relevant to the present study, which also prioritizes user comfort in the design of virtual environments for effective stress therapy. In the meantime, [17] proposed a framework that outlines factors and design elements pertinent to virtual environments used in stress therapy, particularly those affecting user comfort. Their framework addresses five factors and fourteen design elements: (i) visual clarity; (ii) safety features; (iii) mind preparedness; and (iv) social support. By referring to their framework, this study aims to (i) evaluate the effectiveness of these four user-centered design elements in facilitating stress reduction and (ii) explore the underlying rationale behind their stress-reducing properties.

Understanding and implementing these design elements is essential not only for the immediate benefits of stress reduction but also for the long-term development of more effective virtual therapeutic solutions. By tailoring virtual environments to meet the specific psychological and emotional needs of users as stated by [21], this study posits that researchers and developers can create more responsive, effective, and user-friendly virtual therapeutic environments that can adapt to diverse user populations.

Moreover, this study bridging theoretical design principles with practical applications fills an important gap in the fields of mental health and human-computer interaction and makes explicit contributions to both disciplines. Indeed, the findings of this study help lay the groundwork for future research on innovative therapeutic solutions for mental stress. Thus, this study not only contributes toward future research but also allows a way to pair interdisciplinary collaboration with state-of-the-art

technologies in treatment approaches, further improving the treatment landscape of mental health issues.

The remainder of this paper is structured as follows: Section II reviews existing studies on virtual reality-based stress therapy and highlights the identified gap. Section III presents a framework for designing virtual environments for stress therapy. Section IV outlines the methodology, including data collection and analysis methods. Section V presents the results, followed by a discussion in Section VI and a conclusion in Section VII.

II. Review of Existing Virtual Reality-Based Stress Therapy Studies

VR use in stress therapy to date has proved beneficial for its users as numerous scholars (e.g., [9], [10], [11], [12], [13]) have proved that VR-based stress therapy yields significant improvements after a session when compared to standard types of treatment by providing man-altering experiences. Recently, [22] undertook a systematic review and found that out of 50 studies conducted on VR and stress management, 48 studies confirmed the efficacy of VR for this purpose. Not only that, but a study by [14] whose systematic review identified that the 17 studies explore VR and immersive technology in workplace wellbeing also supports VR as a safe and effective tool for wellbeing stemming from stress. In agreement, other studies by [19], [23], [24], and [25] also demonstrated the feasibility of the VR stress therapy initiative.

Meanwhile, although many recent studies indicate VR could be a suitable tool for performing stress therapy, there remains a gap in the study of virtual environments designed for stress therapy [17], [18]. In other words, even though the majority of existing studies presented the results on the effectiveness of the system in reducing stress with design discussion focused on general virtual environment factors such as audio-visual elements, navigation, biofeedback integration, and interaction [19], [22], [26], [27], [28]; however, there is limited discussion on user comfort within the virtual environment.

Following that, this study believed that emphasizing the necessity to incorporate user comfort in the design of virtual environments for stress therapy is essential as it serves as a guideline for developing effective virtual environments that aid in the stress reduction process. This is because the implementation of user comfort in virtual environments will ensure the comfortability of the users while in the therapy session, thereby increasing the users' motivation to continue with the therapy session [29], [30]. Neglecting user comfort in the design of virtual environments for stress therapy may result in ineffective therapy sessions and a loss of interest in continuing the therapy.

III. Framework For Designing Virtual Environments For Stress Therapy

A. Overview of Framework for Designing Virtual Environments for Stress Therapy

Fig. 1 illustrates the framework for designing virtual environments for stress therapy. The development of the framework involved four studies; (i) user requirements, (ii) an existing system review, as presented in study [31], (iii) user evaluation, where the discussion was presented in study [32] and (iv) an expert review, where the details was presented in study [33]. This framework consists of five factors – environment setting, environment exploration, interaction, attention stimuli, and user comfort; and 14 design elements.

The framework was discussed and evaluated by [17], who found out that (i) the system developed based on the proposed framework was found to be effective, usable, useful, and easy to use and (ii) all 14 design elements proposed in the framework were deemed useful in reducing stress. These encouraging findings suggest that the design elements included in the framework significantly contribute to the stress reduction process and should be considered when designing virtual environments for stress therapy.

Fig. 1. Framework for designing virtual environments for stress therapy. Source: [34]

Among these five factors, three are general factors for virtual environments that have been shown to facilitate stress reduction, as evidenced by studies conducted by [18], [19], [27], [35], and [36]. These three general factors are:

- Environment setting, which includes audio and visual design elements that facilitate stress reduction.
- Environment exploration, which addresses navigation techniques to support stress reduction.
- Interaction, which underscores the design elements for suitable interaction types that aid the stress reduction process.

While the past studies (e.g., [18], [19], [27], [35], [36]) mostly focused on the general factor of the virtual environment while discussing audio-visual elements, navigation, and interaction factors, however, the factor of user comfort remains underexplored. In other words, past studies often overlook the importance of user comfort in enhancing the therapeutic outcomes of virtual environments. Therefore, this paper aims to fill in the gap by exploring effective therapeutic interventions, particularly user-comfort-focused virtual environments for effective stress therapy. This study intends to (i) evaluate the effectiveness of four user-centered design elements – visual

clarity, safety features, mind preparation, and social support – in facilitating stress reduction; and (ii) explore the underlying rationale behind their stress-reducing properties. By discussing the rationale behind the stress-reduction properties of these four design elements, this study seeks to provide an understanding of how the user comfort factor contributes to the effectiveness of virtual environments for stress therapy.

B. User Comfort Factor in Designing Virtual Environments for Stress Therapy

User comfort in the context of designing virtual environments for stress therapy refers to the degree to which users feel psychologically at ease while being in the virtual environment. Four design elements contribute to user comfort namely, (i) visual clarity, (ii) safety features, (iii) mind preparation, and (iv) social support. Table I provides descriptions of each design element.

TABLE I. DESCRIPTION OF DESIGN ELEMENTS THAT SUPPORT USER COMFORT IN VIRTUAL ENVIRONMENTS FOR STRESS THERAPY

Design Element	Description
Visual clarity	Objects included in the environment should be visually clear to ensure information is correctly delivered. The users should quickly gain the necessary information from the visual presented (easily recognize the objects at first glance), understand what is happening, and what they should do to react accordingly.
Safety features	The environment should provide features that make users feel safe and protected. The environment cannot be dangerous in any sense of appearance, either by using dangerous animals or any setting that can trigger the users to feel anxious.
Mind preparation	A short mind-framing session that involves the users prepares their minds to become aware of the environmental changes before the virtual therapy starts (from the real-world environment to the virtual world environment) and before the therapy session ends (from the virtual world environment to the real-world environment). The pre-therapy mind preparation scene is meant to help the users get ready and focus on the therapy scene. The post-therapy mind preparation scene is meant to set the user's mind to bring the relaxation feel from the therapy session to the real world.
Social support	The presence of avatars who resemble real people provides the opportunity to observe and interact.

Source: [34]

Following the description in Table I, Table II shows the recommendations for the design elements' application in virtual environments for stress therapy.

By referring to Tables I and II, the first design element identified to support user comfort is visual clarity. Visual clarity refers to the comprehensibility of the displayed information within the virtual environment. Information may be presented in various forms, such as images, 3D models, or text, with the primary objective being to ensure that users easily understand the displayed information. For example, a display of flying birds should be instantly recognizable, rather than appearing as jagged moving dots. Similarly, fish in a river should be identifiable at first glance. High visual clarity can be achieved through high-quality graphic elements and appropriate techniques, all while maintaining the system's performance [37], [38], [39].

TABLE II. RECOMMENDATIONS FOR DESIGN ELEMENTS' APPLICATION INTO VIRTUAL ENVIRONMENTS FOR STRESS THERAPY

Design Element	Application of Design Element
Visual clarity	It is recommended to provide a clear graphic display where users can directly get the necessary information presented by an object. A clear graphic display can be achieved by having high-quality graphics. Graphic quality is usually associated with technical graphic aspects such as the numbers of the polygon, resolution rate, texture quality, and anti-aliasing features.
Safety features	It is recommended to provide the users with space and situations that make them feel safe and protected. It is also recommended to consider safety elements before any objects or situations are included in the environment.
Mind preparation	It is recommended to include mind-preparation activities that guide the users to appropriate meditation practices, such as breathing exercises. In addition, it is also recommended to have two sessions of mind preparation: first, at the beginning (before the therapy environment starts), and second, at the end of the therapy session. As for before the session starts (the pre-therapy session), it is recommended to include a set of breathing exercises containing at least three rounds of breath-in, hold-out, and breath-out instructions. This breathing exercise will prepare the user to be in therapy mode. Whereas, the mind preparation scene before the session ends (post-therapy session) should include a script that reminds the users to maintain the positive feeling they experience all day long. At the end of the script, it is suggested to have a countdown for the user to prepare themselves to take off the head-mounted display (HMD). As the script is being read, no additional relaxing visuals are encouraged to be displayed. The application may continue displaying the therapy environment until the end of the countdown. The application should shut down automatically as the countdown reaches its end.
Social support	It is recommended to include the presence of virtual people who mingle around in the environment. The virtual person or avatar may be portrayed through cyclists cycling around a park, a family having a picnic under a tree, or men jogging around. In addition, it is also recommended to take the distance between the virtual people and the users into consideration. Consideration should be made to include some interpersonal distance between the virtual people and the users to provide comfort and privacy to the users. The distance that is too close may trigger anxiety in the users. Apart from that, the number of virtual people included should also be considered. The number should be in minimal quantity—neither too many that will crowd the environment nor a complete absence, which may cause the users to feel lonely instead. Besides, the physical appearances of the virtual people should also be taken into account. Virtual people should portray proper appearances that suit the user's culture.

Source: [34]

The second design element identified is safety features. Safety features pertain to conditions within virtual environments that provide users with a sense of safety and protection. This design element is essential for preventing adverse effects and ensuring a pleasant therapeutic experience [40]. Implementing safety features involves designing the environment to avoid any dangerous appearances, such as the inclusion of hazardous animals like snakes and scorpions or settings that could trigger

user anxiety. Incorporating soft representations of animals [41], such as fish, rabbits, and ducks to fulfil the safety feature design elements is advisable. Furthermore, it is prudent to evade perilous locations, such as cliffs or higher terrains. In environments that may provoke anxiety, appropriate safety precautions must be instituted. In a beach environment with tumultuous waves, the incorporation of wavebreakers could augment users' perception of safety. In the context of a lookout tower, the incorporation of gates or barriers would decrease the likelihood of anxiety-provoking scenarios. Thus, this study advocates for the evaluation of safety aspects before the integration of any objects or scenarios into virtual environments to establish secure and supportive therapeutic settings.

The third design element contributing to user comfort is mind preparation. Mind preparation denotes the individual's engagement in activities that augment their awareness of alterations in the surrounding environment. To incorporate the mental preparation design element in the virtual environment, it is advisable to integrate activities that direct users through suitable meditation practices, including breathing exercises [42], [43], [44], [45]. Mind preparation can be divided into two parts: (i) pre-therapy and (ii) post-therapy sessions. The two parts are designed to ensure a seamless visual transition for the user from the physical world to the virtual environment before therapy and from the virtual realm back to reality following the session's conclusion. The inclusion of this feature aims to enhance users' consciousness, therefore better equipping them for the next stage of therapy. As individuals enhance their preparedness, the probability of experiencing anxiety during the session diminishes.

The fourth design element identified for user comfort is social support. Social support is defined as the presence of avatars that resemble real people, which provides users with opportunities to observe and interact. The objective of implementing social support is to prevent users from feeling lonely and isolated while also helping to reduce the anxiety associated with being alone in an unfamiliar place [46]. Implementing social support requires the strategic placement of avatars within the virtual environment. The avatars may represent cyclists, families having picnics, or individuals jogging. However, it is crucial to consider the distance between the user and the avatars. The avatars should not be allowed to come too close to the user, as this may cause discomfort; nor should they be placed too far away, which could induce feelings of isolation [47], [48], [49]. The concept of interpersonal distance can be applied to determine the optimal positioning of avatars relative to the user, ensuring a balance that enhances comfort and reduces anxiety.

Having explained the descriptions and recommendations for the application of the four design elements, the next section presents the methodology for evaluating the effectiveness of these elements in reducing stress.

IV. METHODOLOGY

A. Participants

A total of 55 healthy adults aged 20 to 60 volunteered for this study. During the screening session, participants answered questions about their stress levels using the Depression Anxiety Stress Scale (DASS). Only those with a score above 14 or a stress intensity scale of four or higher were invited to experience the virtual environment and be interviewed. Those criteria indicated that participants were experiencing some level of stress. Ultimately, 40 participants met the selection criteria of this study. As illustrated in Table III, 10 are male, and 30 are female. Among these 40 participants, 13 participants are within the age range of 20 to 30 years old, 14 participants an age range between 31 and 40 years, nine participants with an age range between 41 and 50 years old, and four participants with an age range between 51 and 60 years old.

TABLE III. PARTICIPANTS' DEMOGRAPHIC PROFILES

Demographic Profiles		**Frequency**	**Percentage**
Gender	Male	10	25.0%
	Female	30	75.0%
Age	20 – 30 years old	13	32.5%
	31 – 40 years old	14	35.0%
	41 – 50 years old	9	22.5%
	51 – 60 years old	4	10.0%

B. Data Collection

Given that the four design elements to address user comfort were reported to be effective, usable, useful, and easy to use, this study therefore focused on evaluating the usefulness of these four design elements in reducing stress. Three methods were used to collect the required data: (i) experiment, (ii) questionnaire, and (iii) interview. In other words, all 40 participants went through all three data collection methods, beginning with the environmental stage, filling in the questionnaire, and ending with an interview session.

1) Experiment: All 40 participants were invited to this experimental stage, which allowed them to experience the virtual environments developed based on the proposed framework discussed in the earlier section and as an evaluation tool to assess the usefulness of the proposed design elements in a single session. Fig. 2 shows images of the virtual environments used in the experiment.

2) Questionnaire: The usefulness of the design elements was evaluated based on their helpfulness in assisting users in reducing stress. A 5-point Likert-type scale question was included in the questionnaire for participants to rate the helpfulness of each design element in reducing stress where 1 indicated 'not useful at all', 2 'not helping much', 3 'helpful', 4 'very helpful' and 5 'no opinion'.

3) Interview: Upon completion of the experimental session, face-to-face interviews were conducted with each participant. The objective of the interview was to gain a further explanation of the ratings received in the questionnaire. The questions asked during the interview sessions were based on feedback obtained from the questionnaire. Each interview session was audio-recorded for analysis purposes. The interview sessions provided insights into how each design element contributed to the stress reduction process and ensured accurate interpretation of the rating received.

Fig. 2. Images of virtual environments used for the experiment.

C. Data Analysis

Data analysis was conducted based on the type of data collected. Feedback received in numerical form was tabulated on a table and was then analyzed by using descriptive statistical analysis using Microsoft Excel 2021. Meanwhile, recorded audio was transcribed word-by-word before they were reviewed three times for accuracy and understanding and content analysis using ATLAS.ti version 9.

V. Results

The results of the analysis are presented in three parts; (i) helpfulness frequency, (ii) helpfulness mean values, and (iii) helpfulness understanding.

A. Usefulness of Design Elements to Support User Comfort Based on Helpfulness Frequency

The helpfulness of the design elements was assessed using a 5-point Likert-type rating grouped into four categories for analysis purposes as shown in Table IV.

Based on the analysis, this study found the majority of the participants agreed on the usefulness of all four design elements in assisting them to reduce stress. As detailed in Table VI, the agreement is reflected in the high numbers of 'helpful' frequency (3 and 4 helpfulness rating points) compared to 'not helping much', 'not helpful at all', and 'no opinion'.

B. Usefulness of Design Elements to Support User Comfort Based on Helpfulness Mean Values

As shown in Table V, this study also analyses the helpfulness of the design elements that facilitate user comfort through their helpfulness mean values. All four design elements were useful in reducing stress as their helpfulness mean values were above the mean scale value of 2.5.

C. Rationale Behind the Helpfulness of Design Elements in Facilitating Stress Reduction

The analysis of the interviews provided an understanding of how the design elements aid stress reduction as the participants were asked to explain their reasoning behind the helpfulness rating, they provided for the design elements during the interview sessions.

1) Visual clarity: For the first design element, visual clarity, the majority of the participants (39 out of 40 participants or 97.50%) rated it as 'helpful' in reducing their stress – as presented in Table VI earlier. Indeed, the interview participants revealed that this design element is helpful as the clarity of the visual displayed significantly contributed to their comfort. They elaborated that: "Nah… clear visual made it easy to capture the information (displayed)… I didn't have any problem to do that" (Participant 17) and "When it (visual display) clear… it helps us feel at ease and comfortable" (Participant 57).

Meanwhile, for the one who rated visual clarity as 'not helping much', the reason for that was the user claimed to experience an unclear visual display. The participant commented that it was difficult to recognize the information being conveyed, which made her feel uncomfortable. She commented: "*Unclear visual made me dizzy*" (Participant 50).

TABLE IV. Derivation of Helpfulness Categories based on the Helpfulness Rating Points

Helpfulness Rating Point	Helpfulness Category
4 - Very helpful	Helpful
3 - Helpful	
2 - Not helping much	Not helping much
1 - Not helpful at all	Not helpful at all
5 – No opinion	No opinion

TABLE V. Tabulation of Design Elements' Helpfulness Means Values

Design Element	Mean	SD
Visual Clarity	3.58	0.55
Safety Features	3.55	0.81
Mind Preparation	3.45	0.85
Social Support	3.05	0.99

TABLE VI. Tabulation of Design Elements Helpfulness Categories Frequency

	Helpful		Not Helping Much		Not Helpful At All		No opinion	
Design Elements	Frequency	Percent (%)	Frequency	Percent (%)	Frequency	Percent (%)	Frequency	Percent (%)
Visual Clarity	39	97.5	1	2.5	0	0.0	0	0.0
Safety Features	37	92.5	2	5.0	0	0.0	1	2.5
Mind Preparation	36	90.0	3	7.5	0	0.0	1	2.5
Social Support	33	82.5	3	7.5	3	7.5	1	2.5

2) Safety features: For the second design element, 37 participants, or 92.5% rated safety features as 'helpful'. Analysis from the interviews revealed that this design element aids in stress reduction by providing them with a feeling of being safe and protected within the virtual environment. The participants indicated that the implementation of safety features made them feel comfortable as the environments appeared to be not dangerous in any aspect, especially due to the absence of dangerous animals. Among comments received highlighting the importance of safety features included: "Felt safe as there were no harmful animals around… for example… in the garden… within that unfamiliar environment, we never know if monkeys might suddenly jumping out of nowhere… that would make me uncomfortable… so, safety elements help in that sense" (Participant 14) and "It helps… in the sense that if there were dangerous animals… there might be surprising elements… causing fear… so… may cause discomfort… unexpected elements don't contribute to relaxation" (Participant 15).

On the other hand, for the two participants who rated safety features as 'not helping much', the analysis revealed they preferred being in an adventurous setting to reduce their stress. These participants mentioned that they did not mind the inclusion of dangerous animals or extreme situations in virtual environments as that would help them better release stress.

3) Mind preparation: For the third design element, mind preparation, 37 participants, or 90% rated it as 'helpful' due to three reasons; (i) it helps to prepare for visual transitions, (ii) it aids first-time users to calm down and (iii) it helps to release mental burden. Besides, they also recommended this study to apply this design element in two parts; (i) pre-therapy scene and (ii) post-therapy scene. These recommendations received positive feedback from the participants believed that mind preparation had assisted them in preparing themselves for visual transitions. Visual transitions meant by the participants referred to the transition from the real world to the virtual environment and from the virtual environment back to the real world. They praised the implementation of this design element as it smoothens the process and reduces the possibility of dizziness. They narrated that: "It helps!!!... It's kind of providing preparation… mental readiness… before we went into it (therapy scene)" (Participant 8), "It is helpful… it provides readiness before we started (therapy scene)" (Participant 12), "Cause it sets our mind…it sets our readiness" (Participant 15) and "It's 100% helpful because it gets you in the state of mind before you go into the simulation (therapy scene)" (Participant 16).

Another rationale for the helpfulness of mind preparation in supporting user comfort within virtual environments for stress therapy was the design element's ability to help first-time users calm down. Some participants mentioned that the experiment sessions were their first experience with VR, making them nervous as they did not know what to expect from the virtual environments. According to them, this is the part where mind preparation helped the most, to ease their feelings and calm them down. Comments reflecting the helpfulness included: "For me personally, it helps… as I don't have any experience with VR beforehand… therefore, the breath in breath out activities… the preparation exercise… really helped me feel more at ease" (Participant 13), "This one (mind preparation) is actually really helpful…you know when someone has never had any experience with VR… they don't know what to do… starting the therapy session (scene) straight away… is not proper… this thing (mind preparation) is important before we start any therapy… we calm ourselves first" (Participant 14), and "Helpful to reduce my nervous feeling… I had no idea what was inside the environment (therapy scene) … hence… breathing in and out calmed me" (Participant 34).

Additionally, mind preparation was found to be helpful as the design element helps to release the participant's mental burdens. Such practices not only promote relaxation but also empower individuals to enter the virtual environment with a clearer, more focused mindset, ultimately enhancing their therapeutic experience. One participant commented: "Deep breathing in and out made me feel relieved… free of burdens" (Participant 35).

Despite the positive feedback, mind preparation also received three 'not helping much' ratings. Based on the analysis conducted, two reasons were identified. The first reason that caused mind preparation to be less helpful was the participant mentioned that he was more eager to know what was offered in the therapy scene. The eagerness made him restless and uninterested in the mind preparation scene. He commented: "Wanted to skip it (mind preparation scene) as I was more interested in what was offered in the environments (therapy scene)" (Participant 19). The second reason for the less helpfulness rating for mind preparation was some participants had the opinion that breathing exercises could be done anywhere, not necessarily before a therapy session. Hence, making it less interesting as a pre-therapy scene exercise.

4) Social support: For the fourth design element, 33 participants, or 82.5%, rated social support as 'helpful'. The interview sessions also revealed two reasons why the participants believed it was able to help reduce stress. The first rationale for social support being helpful is that its implementation helped users from feeling isolated and lonely. They also recommended this study to place the avatars around the environment as they commented that: "People (avatars) around are helpful… empty environment (without avatars) is not a good idea… may cause loneliness" (Participant 14), "Presence of people (avatars) around reduces loneliness" (Participant 52) and "I can see people around me… it helps… I am not alone" (Participant 55).

The second reason supporting the relevance of social support as a helpful design element in reducing stress is it enhances the realism effect. In the participants' context, realism referred to the imitation of real-world settings within the virtual environment. By incorporating social support, participants felt as if they were in the real world rather than a virtual one. In other words, social support may help to increase user immersion. Among comments received highlighting the benefit of social support in imitating the real-world behavior setting were: "Indeed really helpful… presence of people (avatars) made it feel like being in the real world" (Participant 18), "The presence of avatars… it makes I feel like… like… it is… realistic (real world)" (Participant 17) and "It is essential to have people

(avatars) around… if not… it feels weird to be the only one in the world (virtual environment)" (Participant 32).

For the three participants who rated social support as 'not helping much', the result of the analysis revealed that they preferred to be alone to relieve their stress. For those participants, having people around, even in the form of avatars, made them uncomfortable, as if someone was watching them. They narrated that: "It is better to be alone… don't need social support… I prefer to be on my own" (Participant 8), "I think avatars are not necessary… usually… we find peace with views and sounds…not social support" (Participant 10), and "It is not helpful… having people around feels like someone is watching us… I prefer to be alone to relax" (Participant 53).

Analyzing the feedback for 'not helpful at all', such a rating was given due to the appearance of the avatars used in the system for the experiment session which was found to be scary by three participants. Two participants commented: "It is not helpful at all because the people (avatars)… they look scary (Participant 51) and "Because of the people (avatars)… they are too big and scary-looking" (Participant 50). While the other one commented: "When coming close to the people (avatar)… we can see their scary face… ugly" (Participant 39).

VI. Discussion

The analysis of the data gathered from the experiment, questionnaires, and interview sessions demonstrated that the four design elements identified to facilitate user comfort in virtual environments for stress therapy were useful in supporting the stress therapy process. The encouraging findings were supported by the high number of 'helpful' categories frequency received for all four design elements. Additionally, the mean values for the helpfulness also indicate that the proposed design is useful in assisting stress reduction. The encouraging finding is evidenced by the mean value for each of the design elements surpassing the mean value of the scale used which was 2.5.

Furthermore, the explanation provided for the rationale behind how the design elements assist in stress reduction offers an understanding of the significance of implementing these design elements in virtual environments for stress therapy. This study concluded that visual clarity enhances user comfort by efficiently conveying information, rendering it instantly recognizable and identifiable. Further analysis of the 'not helping much' feedback for visual clarity revealed that the unclear display was not caused by the clarity of the visual environments used for the experiment, but rather a personal issue of the participant. For instance, one of the participants disclosed at the end of the interview that she was experiencing eye problems and was about to undergo corrective surgery. The clarification by the participant indicated that the discomfort experienced was not due to the clarity of the environment but the participant's health issue.

For safety features, this study confirmed that this element helps provide user comfort in virtual environments for stress therapy. The helpfulness is attributed to the sense of safety and protection, as safety features encourage developers to consider safety aspects before implementing any objects or settings into virtual environments. However, for the 'not helping much' feedback received, it was found that the ratings were based on participants' personal preferences rather than their experiences with the system during the experiment before the interview session. Consequently, it may be inferred that the ineffectiveness of the safety features in facilitating stress reduction was attributable to user preferences rather than the characteristics themselves.

Mind preparation was also deemed helpful by the majority of the participants. Based on the results presented, it can be seen that the majority of the participants agreed that mind preparation helped reduce stress. Mind preparation was acknowledged for its benefit in preparing users' minds for visual transitions, assisting first-time users in calming calm and easing their nervousness as well as relieving mental burdens. Regarding the 'not helping much' feedback, comments suggest that breathing exercises can be done anywhere and are not interesting to be implemented as pre-therapy scenes might be due to issues in the implementation strategies of the design element in the virtual environments used for the experiment sessions.

The issues mentioned could be from the design of the scene which might be less attractive or the implementation strategies of the breathing exercise guidelines which used expanding and contracting circles to indicate breath in and out instructions. Therefore, the implementation strategies should be improved to better present the breathing exercise. There is no issue with implementing breathing exercises in the pre-therapy scene as they have been evidenced to be effective in reducing stress as reported by past studies (e.g., [50], [51], [52], [53], [54]).

Similar to the other three design elements, social support also received a majority of 'helpful' feedback. The helpfulness of social support was contributed by its ability to prevent participants from being alone and lonely, as well as enhancing the realism effect of the virtual environment. Despite the positive results, social support was also minorly rated as 'not helping much' in reducing stress due to participants' preferences to be alone when relaxing. The result of the analysis revealed that there was no issue with the environments themselves, rather the preference of the participants did not match the setting of the environments. However, the 'not helpful at all' feedback received provided insight into issues regarding the appearance of the avatars. The problem was not the presence of the avatars but their appearance as the avatars used in the virtual environments for the experiment sessions may have had design issues. It was acknowledged that the appearance may be ugly as the avatars' 3D models incorporated in the virtual environments were not ones with high polygons. Therefore, improving the avatars' appearance is necessary to avoid user discomfort in virtual environments for stress therapy.

The findings regarding the efficacy of the four design aspects in enhancing user comfort in virtual environments for stress therapy were favourable. The majority of users concurred that the proposed design elements were 'beneficial' in alleviating stress. The interview findings elucidated the rationale for the helpfulness of the design features, thereby deepening our comprehension of their role in the stress reduction process. This comprehension is essential as it highlights the importance of integrating user comfort and its design components while creating virtual environments for stress therapy.

VII. CONCLUSION

Exploring effective therapeutic methods, especially user-comfort-oriented virtual environments for stress therapy, is crucial for tackling the widespread problem of mental stress in modern society and improving coping strategies. The findings of this study highlight the critical significance of user-centred design in creating virtual worlds specifically designed for stress therapy, whereby the four recognised design elements that enhance user comfort—visual clarity, safety features, mental preparedness, and social support—effectively mitigate stress. The interviews also clarify how these four design aspects help users in regulating their stress levels.

The findings underscore the imperative to prioritise user comfort in the design of virtual environments for stress therapy. Integrating user comfort into the design process is essential for creating effective virtual settings that promote stress relief. User-centred design advocates for the active participation of users in the design process, ensuring that their needs, preferences, and comfort levels are paramount. The study indicates that when design elements are meticulously incorporated, they not only foster a more immersive experience but also enhance users' emotional and psychological well-being, thereby mitigating mental stress and improving general well-being.

REFERENCES

[1] M. Moitra et al., "Global mental health: Where we are and where we are going," *Springer Science+Business Media*, vol. 25, no. 7, pp. 301-311, May 31, 2023. https://doi.org/10.1007/s11920-023-01426-8.

[2] S. Jamshaid et al., "Pre- and post-pandemic (COVID-19) mental health of international students: Data from a longitudinal study," *Dove Medical Press*, vol. 16, pp. 431-446, Feb. 1, 2023. https://doi.org/10.2147/prbm.s395035.

[3] I. Luberenga et al., "Mental health awareness programmes to promote mental well-being at the workplace among workforce in low-income and middle-income countries: A scoping review protocol," *BMJ*, vol. 13, no. 7, e073012-e073012, Jul. 1, 2023. https://doi.org/10.1136/bmjopen-2023-073012.

[4] World Health Organisation, "World Mental Health Report," Jun. 16, 2022. https://www.who.int/teams/mental-health-and-substance-use/world-mental-health-report.

[5] S. K. Norelli, A. Long, and J. M. Krepps, "Relaxation techniques," in StatPearls, Treasure Island, FL: StatPearls Publishing, 2024, [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK513238/ (accessed Feb. 18, 2024).

[6] L. Toussaint et al., "Effectiveness of progressive muscle relaxation, deep breathing, and guided imagery in promoting psychological and physiological states of relaxation," *Evidence-Based Complementary and Alternative Medicine: ECAM*, 2021, 5924040. https://doi.org/10.1155/2021/5924040.

[7] M. J. Park et al., "A literature overview of virtual reality (VR) in treatment of psychiatric disorders: recent advances and limitations," *Frontiers in Psychiatry*, vol. 10, Jul. 19, 2019. https://doi.org/10.3389/fpsyt.2019.00505.

[8] J. L. Maples-Keller, B. E. Bunnell, S. J. Kim, and B. O. Rothbaum, "The use of virtual reality technology in the treatment of anxiety and other psychiatric disorders," *Harvard Review of Psychiatry*, vol. 25, no. 3, pp. 103–113, 2017.

[9] A. Fang, H. Chhabria, A. Maram, and H. Zhu, "Practicing stress relief for the everyday: designing social simulation using VR, AR, and LLMs," *ArXiv*, 2024. https://arxiv.org/abs/2410.01672.

[10] M. May, "How virtual reality therapy is shaping mental health," *Nature Portfolio*, vol. 30, no. 7, pp. 1797-1799, Jun. 4, 2024. https://doi.org/10.1038/d41591-024-00032-2.

[11] S. Riches, "Virtual reality relaxation for stress in young adults," May 31, 2024. https://www.simonriches.com/blog/virtual-reality-relaxation-for-stress-in-young-adults.

[12] S. Saeed et al., "Review on the role of virtual reality in reducing mental health diseases specifically stress, anxiety, and depression," *Cornell University*, 2024. https://doi.org/10.48550/arxiv.2407.18918.

[13] M. Waqas et al., "using virtual reality for detection and intervention of depression — A systematic literature review," *Cornell University*, 2024. https://doi.org/10.48550/arxiv.2403.01882.

[14] S. Riches et al., "Virtual reality relaxation for stress in young adults: A remotely delivered pilot study in participants' homes," *Springer Science+Business Media*, Mar. 28, 2024. https://doi.org/10.1007/s41347-024-00394-x.

[15] D. Han, D. Kim, K. Kim, and I. Cho, "Exploring the effects of VR activities on stress relief: A comparison of sitting-in-silence, VR meditation, and VR smash room," *ArXiv*, vol. 26, pp. 875-884, Oct. 16, 2023. https://doi.org/10.1109/ismar59233.2023.00103.

[16] S. Riches et al., "Virtual reality relaxation for people with mental health conditions: A systematic review," *Springer Science+Business Media*, vol. 58, no. 7, pp. 989-1007, Jan. 20, 2023. https://doi.org/10.1007/s00127-022-02417-5.

[17] F. A. Zaharuddin, N. Ibrahim, and A. M. Yusof, "A conceptual framework for designing virtual environments for stress therapy," *Applied Sciences*, vol. 12, no. 19, p. 9973, 2022. https://doi.org/10.3390/app12199973.

[18] L. Tabbaa et al., "A reflection on virtual reality design for psychological, cognitive and behavioral interventions: Design needs, opportunities and challenges," *International Journal of Human-Computer Interaction*, vol. 37, no. 9, pp. 851–866, 2021. https://doi.org/10.1080/10447318.2020.1848161.

[19] I. Ladakis, D. Filos, and I. Chouvarda, "Virtual reality environments for stress reduction and management: a scoping review," *Virtual Reality*, vol. 28, no. 1, p. 50, 2024. https://doi.org/10.1007/s10055-024-00943-y.

[20] C. M. Fidopiastis, A. A. Rizzo, and J. P. Rolland, "User-centered virtual environment design for virtual rehabilitation," *Journal of NeuroEngineering and Rehabilitation*, vol. 7, Article 11, 2010. https://doi.org/10.1186/1743-0003-7-11.

[21] I. H. Bell et al., "Virtual reality as a clinical tool in mental health research and practice," *Dialogues in Clinical Neuroscience*, vol. 22, no. 2, pp. 169, 2020. https://doi.org/10.31887/DCNS.2020.22.2/lvalmaggia.

[22] S. Meshkat et al., "Virtual reality and stress management: A systematic review," *Cureus*, 2024. https://doi.org/10.7759/cureus.64573.

[23] N. D. Mohd Muhaiyuddin, A. Abdul Mutalib, S. N. Abdul Salam, and N. Alis, "Image-based virtual reality stress therapy application (VRT-stressNOmore): An alternative tool for self-therapy," *Journal of Information System and Technology Management*, vol. 7, no. 29, pp. 222–241, 2022. https://doi.org/10.35631/JISTM.729020.

[24] A. A. Mahmud et al., "Brief virtual reality exposure therapy and its effects on negative and positive emotions among healthy working adults: A feasibility study," *Alpha Psychiatry*, vol. 23, no. 5, pp. 223–229, 2022. https://doi.org/10.5152/alphapsychiatry.2022.21781.

[25] X. Lin et al., "Virtual reality-based musical therapy for mental health management," in *2020 10th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 948–952, 2020. https://doi.org/10.1109/CCWC47524.2020.9031157

[26] F. A. Zaharuddin, N. Ibrahim, and A. M. Yusof, "A conceptual framework for designing virtual environments for stress therapy," *Applied Sciences*, vol. 12, no. 19, p. 9973, 2022. https://doi.org/10.3390/app12199973.

[27] S. Riches et al., "Virtual reality relaxation for the general population: a systematic review," *Social Psychiatry and Psychiatric Epidemiology*, vol. 56, no. 10, pp. 1707–1727, 2021. https://doi.org/10.1007/s00127-021-02110-z.

[28] S. Riches et al., "Virtual reality and immersive technologies to promote workplace wellbeing: a systematic review," *Journal of Mental Health*, vol. 33, no. 2, pp. 253–273, 2024. https://doi.org/10.1080/09638237.2023.2182428.

[29] M. D. A. Rozmi et al., "Design considerations for a virtual reality-based nature therapy to release stress," in *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*, 2019, pp. 1–4. https://doi.org/10.1109/AECT47998.2020.9194175.

[30] G. Marques, R. Nóbrega, and R. N. Madeira, "Exploring virtual reality in exposure therapy for sensory food aversion," in *the 29th International ACM Conference on 3D Web Technology (WEB3D '24), Guimarães, Portugal*, Sep. 25–27, 2024. ACM. https://doi.org/10.1145/3665318.3677166.

[31] O. Cohavi and S. Levy-Tzedek, "Young and old users prefer immersive virtual reality over a social robot for short-term cognitive training," *International Journal of Human-Computer Studies*, vol. 161, p. 102775, 2022. https://doi.org/10.1016/j.ijhcs.2022.102775.

[32] F. A. Zaharuddin et al., "Virtual environment for VR-based stress therapy system design element: User perspective," *International Visual Informatics Conference*, pp. 25–35, 2019.

[33] F. A. Zaharuddin et al., "Virtual reality application for stress therapy: Issues and challenges," *International Journal of Engineering and Advanced Technology*, vol. 9, no. 1, pp. 2325–2329, 2019. https://doi.org/10.35940/ijeat.A2656.109119.

[34] F. A. Zaharuddin, N. Ibrahim, and A. M. Yusof, "Experts review on factors to consider when designing virtual environment for stress therapy," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 3, pp. 2114–2119, 2021. https://doi.org/10.17762/turcomat.v12i3.1153.

[35] F. A. Zaharuddin, "A design framework for designing a virtual environment for stress therapy: Malaysian context," PhD thesis, Universiti Tenaga Nasional, 2023. Unpublished.

[36] A. Gentile et al., "Nature through virtual reality as a stress-reduction tool: A systematic review," *International Journal of Stress Management*, vol. 30, no. 4, pp. 341–353, 2023. https://doi.org/10.1037/str0000300.

[37] N. A. Mohamad Yahaya et al., "Design of game-based virtual forests for psychological stress therapy," *Forests*, vol. 14, no. 2, 2023. https://doi.org/10.3390/f14020288.

[38] J. Li, "Quality optimization of virtual reality-based efficient net model in artworks," *CAD and Applications*, S28, pp. 211-223, 2024. https://doi.org/10.14733/cadaps.2024.S28.211-223.

[39] S. Baigabulov and M. Ipalakova, "Virtual reality enabled immersive data visualization for data analysis," in *Proceedings of the 8th International Conference on Digital Technologies in Education, Science and Industry, Almaty, Kazakhstan*, 2023. https://ceur-ws.org/Vol-3680/S1Paper1.pdf.

[40] S. M. LaValle, Virtual reality. University of Oulu, 2019. https://msl.cs.uiuc.edu/vr/vrbookbig.pdf.

[41] J. Cao et al., "Explorations in designing virtual environments for remote counselling," *ArXiv*, 2024. https://arxiv.org/abs/2409.07765.

[42] J. M. Armfield, "Understanding animal fears: A comparison of the cognitive vulnerability and harm-looming models," *BMC Psychiatry*, vol. 7, p. 68, 2007. https://doi.org/10.1186/1471-244X-7-68.

[43] Y. She et al., "An interaction design model for virtual reality mindfulness meditation using imagery-based transformation and positive feedback," *Computer Animation and Virtual Worlds*, vol. 34, no. 3-4, e2184, 2023. https://doi.org/10.1002/cav.2184.

[44] R. R. Feinberg, U. Lakshmi, M. J. Golino, and R. I. Arriaga, "ZenVR: Design evaluation of a virtual reality learning system for meditation," in *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '22)*, pp. 1-15. ACM, 2022. https://doi.org/10.1145/3491102.3502035.

[45] J. Xu, H. Jo, L. Noorbhai, A. Patel, and A. Li, "Virtual mindfulness interventions to promote well-being in adults: A mixed-methods systematic review," *Journal of Affective Disorders*, vol. 300, p. 571, 2022. https://doi.org/10.1016/j.jad.2022.01.027.

[46] E. Mazgelytė et al., "Effects of virtual reality-based relaxation techniques on psychological, physiological, and biochemical stress indicators," *Healthcare*, vol. 9, no. 12, p. 1729, 2021. https://doi.org/10.3390/healthcare9121729.

[47] K. Kenyon, V. Kinakh, and J. Harrison, "Social virtual reality helps to reduce feelings of loneliness and social anxiety during the Covid-19 pandemic," *Scientific Reports*, vol. 13, no. 1, p. 19282, 2023. https://doi.org/10.1038/s41598-023-46494-1.

[48] A. D. Fraser et al., "Do realistic avatars make virtual reality better? Examining human-like avatars for VR social interactions," *Computers in Human Behavior: Artificial Humans*, vol. 2, no. 2, p. 100082, 2024. https://doi.org/10.1016/j.chbah.2024.100082.

[49] M. T. Deighan, A. Ayobi, and A. A. O'Kane, "Social virtual reality as a mental health tool: How people use VRChat to support social connectedness and wellbeing," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, pp. 1-18. ACM, 2023. https://doi.org/10.1145/3544548.3581103.

[50] L. Leung, "Loneliness, social support, and preference for online social interaction: The mediating effects of identity experimentation online among children and adolescents," *Chinese Journal of Communication*, vol. 4, no. 4, pp. 381-399, 2011. https://doi.org/10.1080/17544750.2011.616285.

[51] S. P. Morgan, C. A. Lengacher, and Y. Seo, "A systematic review of breathing exercise interventions: An integrative complementary approach for anxiety and stress in adult populations," *Journal of Holistic Nursing*, 08980101241273860, 2024. https://doi.org/10.1177/08980101241273860.

[52] G. Birdee et al., "Slow breathing for reducing stress: The effect of extending exhale," *Complementary Therapies in Medicine*, vol. 73, 2023. https://doi.org/10.1016/j.ctim.2023.102937.

[53] M. Komariah et al., "Effect of mindfulness breathing meditation on depression, anxiety, and stress: A randomized controlled trial among university students," *Healthcare*, vol. 11, no. 1, 2023. https://doi.org/10.3390/healthcare11010026.

[54] H. Hamasaki, "Effects of diaphragmatic breathing on health: A narrative review," *Medicines*, vol. 7, no. 10, p. 65, 2020. https://doi.org/10.3390/medicines7100065.

A Machine Learning Model for Crowd Density Classification in Hajj Video Frames

Afnan A.Shah
Department of Computing-College of Engineering and Computing, Umm Al-Qura University, Makkah 21961, Saudi Arabia

***Abstract*—Managing the massive annual gatherings of Hajj and Umrah presents significant challenges, particularly as the Saudi government aims to increase the number of pilgrims. Currently, around two million pilgrims attend Hajj and 26 million attend Umrah making crowd control especially in critical areas like the Grand Mosque during Tawaf, a major concern. Additional risks arise in managing dense crowds at key sites such as Arafat where the potential for stampedes, fires and pandemics poses serious threats to public safety. This research proposes a machine learning model to classify crowd density into three levels: moderate crowd, overcrowded and very dense crowd in video frames recorded during Hajj, with a flashing red light to alert organizers in real-time when a very dense crowd is detected. While current research efforts in processing Hajj surveillance videos focus solely on using CNN to detect abnormal behaviors, this research focuses more on high-risk crowds that can lead to disasters. Hazardous crowd conditions require a robust method, as incorrect classification could trigger unnecessary alerts and government intervention, while failure to classify could result in disaster. The proposed model integrates Local Binary Pattern (LBP) texture analysis, which enhances feature extraction for differentiating crowd density levels, along with edge density and area-based features. The model was tested on the KAU-Smart-Crowd 'HAJJv2' dataset which contains 18 videos from various key locations during Hajj including 'Massaa', 'Jamarat', 'Arafat' and 'Tawaf'. The model achieved an accuracy rate of 87% with a 2.14% error percentage (misclassification rate), demonstrating its ability to detect and classify various crowd conditions effectively. That contributes to enhanced crowd management and safety during large-scale events like Hajj.**

***Keywords*—Hajj; moderate crowd; overcrowded; very dense crowd; machine learning**

I. Introduction

Mass surveillance video data recorded during Hajj require advanced processing techniques [1] to classify and detect crowd density levels in real-time ensuring timely responses to high-risk situations [2]. Manual crowd-monitoring systems are not only time-consuming and resource-intensive but also insufficient for managing the complexity of massive events like Hajj [3]. Machine learning methods have been implemented as a reliable solution for automating crowd analysis enabling the classification of crowd density levels directly from video frames [4]. Managing crowd density during Hajj also presents general challenges [5], such as noisy video frames [6], environmental factors like poor visibility and unrecognized objects [7], such as moving fans or vehicles near the Kaaba. These challenges can cause false alerts or missed detections [8], potentially leading to delays or dangerous situations. Furthermore, normal crowd behaviors during Hajj, such as group movements [9] and sacred congregations may appear high-risk in other contexts.

This study focuses on detecting and classifying crowd density into three levels: moderate crowd, overcrowded and very dense crowd. Unlike many current research efforts that predominantly rely on Convolutional Neural Networks (CNNs) for detecting abnormal behaviors [10], this research employs a Gradient Boosting Classifier (GBC) with a structured feature extraction model. Features such as Edge Density, Local Binary Pattern (LBP) texture and crowd area coverage provide a comprehensive representation of crowd characteristics, while a threshold-based method directly labels frames with Edge Density values above a critical level as very dense, ensuring the rapid identification of high-risk crowd regions.

The proposed model implements data augmentation techniques to improve generalization throughout various scenarios, uses class balancing to address the distribution of crowd density levels in the dataset and enhances the GBC through hyperparameter tuning with GridSearchCV to ensure robust performance. In addition, the model triggers visual alerts for very dense crowd scenes by overlaying a red flash indicator on detected frames, assisting organizers in real-time decision-making.

The contributions of this study are as follows: development of a Machine Learning model for classifying crowd density levels in video frames recorded during Hajj; integration of key features such as Edge Density, LBP texture and crowd area coverage to enhance classification accuracy; and use of performance metrics such as accuracy, precision, recall and F1-score to evaluate the model, with special emphasis on the 'very dense crowd' category.

The following sections are organized as, Section II provides the related works and Section III presents the proposed model. Section IV shows the result analysis and Section V concludes the paper.

II. Related Work

Felemban et al. [2] highlighted different levels of crowd density observed among pilgrims categorized as moderate crowd, overcrowded and very dense crowd. Such categories range from safe to potentially disastrous with very dense crowds posing significant risks such as stampedes. Although extensive efforts by Hajj authorities to manage crowds effectively, a method to classify hazardous crowd density conditions and provide timely notifications remains critical for enhancing crowd management. This paper builds on such insights by

proposing a data-driven machine-learning approach for accurate and timely crowd density classification.

Aldayri and Albattah [11] introduced a computer vision framework that uses a convolutional LSTM autoencoder to detect abnormal human behaviors in video sequences by combining convolutional neural network (CNN) to extract spatial features from video frames and LSTMs to analyze temporal patterns across consecutive frames. Their research, along with other similar studies [12] focuses on identifying abnormal behaviors whereas my model emphasizes crowd density classification based on structural features like Edge Density and LBP texture, addressing the specific challenge of real-time detection of very dense crowd levels.

In another research effort [13], a Crowd Anomaly Detection Framework (CADF) was developed which integrates multi-scale feature fusion and soft non-maximum suppression (soft-NMS) to detect anomalies in dense crowd scenarios. Bhuiyan [4] proposed the Crowd Anomaly Hajj Monitor to classify crowd anomalies by using optical flow and a fully convolutional neural network, which identifies specific behaviors but counts on predefined feature extraction techniques that may limit adaptability to varying crowd scenarios. On the other hand, my approach leverages Edge Density and other texture-based features for crowd density classification to ensure wider relevance in different crowd conditions. In addition, Alhothali et al. [14] introduced a deep CNN-based model to detect and localize anomalous events in dense crowd scenes classifying seven categories of abnormal events based on extracted spatial and temporal features. While that approach is effective for anomaly detection, such models often focus on individual behaviors rather than holistic crowd density levels. In a similar effort [15], other studies have utilized lightweight CNN and LSTM models to detect violent activities in surveillance footage, generating real-time alarms for law enforcement, which often enhance specific scenarios. Whereas, my study emphasizes a generalizable model for crowd density classification that focuses on critical safety categories like very dense crowds.

Previous works have widely applied CNNs and FCNNs to analyze surveillance videos for detecting anomalies [16]. Also, FCNN have been used for estimating crowd density from distant surveillance footage, showing significant improvements in crowd density classification [17]. However, many depend on fixed feature extraction techniques, such as optical flow or motion analysis [18] which can limit their adaptability to the unique dynamics of Hajj. Also, methods dependent on inflexible frameworks [19] may face challenges in handling the complicated temporal and spatial variations in crowd movement, especially during rituals and gatherings. The proposed model in this paper addresses these limitations by extracting meaningful features such as Edge Density, LBP texture and crowd area coverage which better capture the visual and structural characteristics of different crowd density levels. Furthermore, traditional approaches may classify normal crowd behaviors during Hajj, such as ritual-related movements or group activities as anomalies because of the unique nature of such event. This study avoids such pitfalls and provides a focused solution for monitoring critical crowd density levels in real-time by tailoring the model to classify crowd density rather than individual behaviors.

This study adds to the existing body of knowledge by presenting a machine-learning model designed to systematically classify crowd density levels. The model achieves enhanced accuracy and adaptability by employing a structured feature extraction process and fine-tuning classification with GBC, offering a practical solution for improving crowd management and safety during Hajj.

III. The Proposed Model

A. Data Preparation

I used the HAJJv2 dataset [20] which contains training and testing videos each set including nine videos capturing different holy sites during Hajj such as Massaa, Jamarat, Arafat and Tawaf. This dataset was originally created to focus on abnormal crowd behaviors during Hajj. Each video is divided into different frames and each frame's number is repeated multiple times in the label file with each instance assigned to a different type of abnormal behavior, thereby allowing for the capture of various behaviors occurring simultaneously or consecutively within the same frame.

For this study, I used the same segmented frames provided in the dataset with each video divided into a varying number of frames ranging between 466 and 501. However, I modified the label files to categorize the frames into three classifications; moderate crowd, overcrowded and very dense crowd eliminating the need to repeat the same frame's number to define multiple abnormal behaviors in a single frame. This was essential to reduce complexity and focus on overall density rather than specific behaviors.

Since the goal of this study is to categorize the overall crowd density in each frame rather than track specific abnormal behaviors or objects, I eliminated the need to repeat the same frame number to detect various abnormal behaviors in different areas within each frame. Instead, I simplified this by selecting one entry per frame to represent its density level. Each video's frames in the training and testing set are processed individually and assigned a single density label; moderate, crowded or very dense crowd based on the dominant or average density level observed in each frame.

To ensure accurate density classification, the dataset was preprocessed further. During this process, the total number of rows loaded from the dataset was 43,721 and frames with high EdgeDensity values were automatically updated with corresponding labels for very dense crowd conditions. The augmented dataset resulted in 131,163 entries which included features such as Edge Density, Area, Average Intensity and Local Binary Pattern (LBP) texture, significantly expanding the variety of training samples. In addition, an edge density threshold of 64.5 was set during the training process to facilitate the reliable identification of very dense crowds. That simplifies the dataset structure, making it more efficient for training a model aimed at detecting general crowd density levels without requiring additional details, such as bounding boxes or multiple labels per frame. By doing so, the training process is more streamlined, with improvements in efficiency and performance.

B. Model Design

The major goal of the proposed model is to detect and classify crowd density levels in video frames recorded during Hajj across various holy sites aiming to identify very dense crowd regions that may require immediate attention. In order to achieve this, the model uses a structured approach that starts with careful dataset preparation and well-defined labeling criteria as shown in Fig 1. The dataset comprises video frames categorized by crowd density levels: moderate crowd, overcrowded and very dense crowd. The model employs a threshold-based method to automatically label samples where frames with an EdgeDensity value above 64.5 are directly classified as very dense. This was selected through exploratory analysis of sample distributions and is intended to provide immediate labeling for the most critical category, ensuring timely detection of high-density scenes.

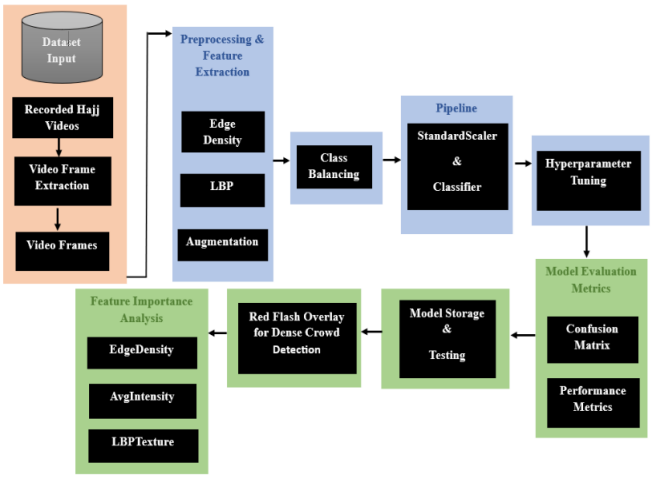


Fig. 1. Proposed model.

The model extracts Edge Density by calculating the density of edges detected in a designated area of each frame as this feature reflects changes in density. The Local Binary Pattern (LBP) texture feature was calculated by analyzing the local pixel patterns in grayscale crowd regions within the video frames, by which the LBP algorithm computes a binary code for each pixel by comparing it with its surrounding neighbors and generates a histogram capturing the frequency of specific binary patterns. That histogram reflects the texture characteristics of the crowd which is particularly useful in identifying varying density levels based on textural patterns that are prevalent in dense crowds. In this study, the histogram values were normalized and averaged to provide a single numerical representation of texture per region, which was included as the LBP Texture feature. The decision to use LBP stems from its ability to highlight subtle texture variations across crowd density levels aiding the model in making more accurate classifications and features like the area covered by the crowd region and average intensity contribute further contextual details. Also, data augmentation is used to enhance the training set's diversity including frame rotation and brightness adjustment. The reason for using augmentation is to provide the model with a wider array of samples thereby improving its generalization across new frames by allowing it to learn from a broader variety of crowd presentations.

Class balancing was essential due to some crowd levels having significantly fewer samples than others. The model addresses this by oversampling the minority classes, particularly the moderate crowd and very dense crowd categories. This ensures that each crowd level is equally represented during training thereby reducing the likelihood of bias and enhancing the classifier's performance across all classes.

The Gradient Boosting Classifier (GBC) was implemented as the primary classification algorithm because of its effectiveness in modeling complex non-linear relationships. That classifier works by sequentially building a series of decision trees where each tree attempts to correct the errors of its predecessor. That iterative boosting process allows the model to focus more on the misclassified samples from previous iterations, thereby improving its overall accuracy. For this study, GBC was integrated into a model that included feature scaling using StandardScaler to normalize the values of all features e.g., EdgeDensity, LBPTexture, Area, and AvgIntensity. The scaled features were then passed to the classifier. The model achieved optimal performance by incorporating GBC in a model and tuning its hyperparameters using GridSearchCV, as reflected in a cross-validation score of 0.97. That iterative approach also ensures the model's robustness in classifying different crowd densities despite the diversity in frame characteristics.

Hyperparameter tuning through GridSearchCV allows for the optimization of the GBC. A selection of key parameters such as the number of estimators, learning rate and maximum depth is tuned using cross-validation to ensure the model performs well without overfitting. Systematically, GridSearchCV tests various parameter options to find the set that yields the highest accuracy score during cross-validation.

Performance metrics such as accuracy, precision, recall and F1-score are used to evaluate the proposed model in this study as they provide an in-depth assessment of its ability to predict each crowd density level correctly with particular attention given to its precision and recall on the 'very dense crowd' classification due to the high importance of this category. In addition, a confusion matrix is employed to analyze further misclassifications offering insights into which crowd density levels may be confused by the model.

The final model, saved as a serialized file, is tested on new video frames to assess its practical performance in detecting very dense crowds, and importantly, it activates a red flash indicator as a visual alert for such frames. The design process is finalized with a feature importance analysis as each feature is examined to provide findings on which factors of crowd data affect the model's predictions validating the choice of features and pointing out the model's reliance on important crowd-related characteristics.

C. Training Process

The training dataset was balanced using oversampling techniques to address the class imbalance especially for the moderate crowd and very dense crowd categories. That ensured equal representation of all three crowd density levels. Features of each frame were extracted that included Edge Density calculated as the density of detected edges, Local Binary Pattern (LBP) texture, area covered by the crowd and average intensity.

During the training process, the LBP texture features played a significant role in distinguishing between moderately crowded and very dense crowd frames, since the histogram values captured by LBP provided insights into the structural patterns within the crowd, such as uniform regions indicating low density versus textured regions in high-density areas. That feature combined with EdgeDensity, area, and average intensity, allowed the model to detect high-density areas effectively. In addition, to enhance the diversity of the training set, data augmentation techniques such as rotation and brightness adjustments were employed.

D. Hyperparameter Optimization and Feature Importance

Hyperparameter tuning was performed using GridSearchCV, optimizing key parameters of GBC, such as the number of estimators 50, 100 and 200, learning rate 0.01, 0.05, 0.1, 0.3 and 0.5 and maximum depth 3, 4, 5, 6 and 7. The best parameters were a 0.5 learning rate, a maximum depth of 7, and 200 estimators, achieving a cross-validation score of 0.97 which provided the best cross-validation accuracy while balancing model complexity and performance.

An analysis of feature importance revealed that Edge Density was the most significant predictor for crowd density classification followed by LBP Texture and Average Intensity. The feature importance were: Area (0.69), Edge Density (0.24), Average Intensity (0.06) and LBP Texture (0.00). That confirms the relevance of these features in distinguishing between different crowd densities and supports their inclusion in the model.

IV. Results and Discussion

During the training stage of the model, I visualized the distribution of key features, such as area, position coordinates, edge density and average intensity across different crowd density levels including moderate, overcrowded and very dense crowd. These highlight how these features contribute to distinguishing between the crowd density categories which is essential for the model's accuracy. Fig. 2 shows that Area is a significant feature in distinguishing very dense crowd situations from moderate crowd or overcrowded as very dense crowds tend to cluster within a smaller area. In the HAJJv2 dataset, very dense crowds mostly appear in videos recorded during Tawaf around the Kaaba which is typically a quite small area. The yellow and green bars representing overcrowded and moderate crowds respectively, are distributed more widely across larger areas such as Sa'i, Mina, and Arafat compared to the very dense category. Such finding suggests that as crowd density decreases, the area covered by the crowd tends to increase due to people being more spread out.

Fig. 3 and Fig. 4 show the distribution of X and Y coordinates respectively, by crowd density category: moderate crowd, overcrowded and very dense crowd. Very dense crowd regions represented by the red bars are concentrated within specific ranges on both the X and Y axes. The clustering suggests that very dense crowds are likely to occur in particular locations within the frame reflecting areas where crowding frequently happens. Meanwhile, overcrowded and moderate crowds represented by the yellow bars and green bars respectively, are more widely distributed across the X and Y coordinates. In particular moderate crowds are spread across a larger range, indicating a more even distribution across the frame. Therefore, the specific positioning patterns of each crowd density level imply that spatial coordinates X and Y are useful features for distinguishing crowd densities. Thus, the model can use this information to recognize crowd density levels based on where they commonly appear within the frame.

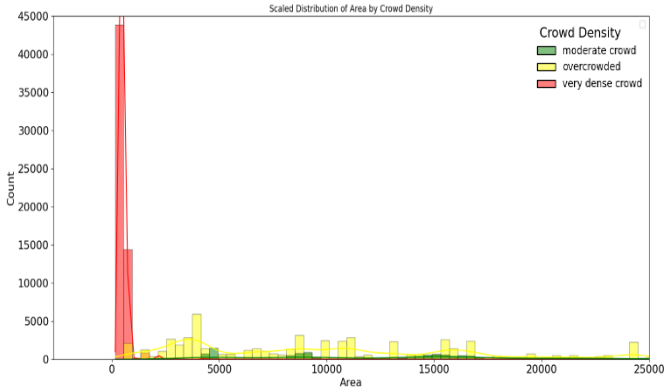


Fig. 2. Distribution of the "Area" feature by crowd density category (moderate crowd, overcrowded, and very dense crowd).

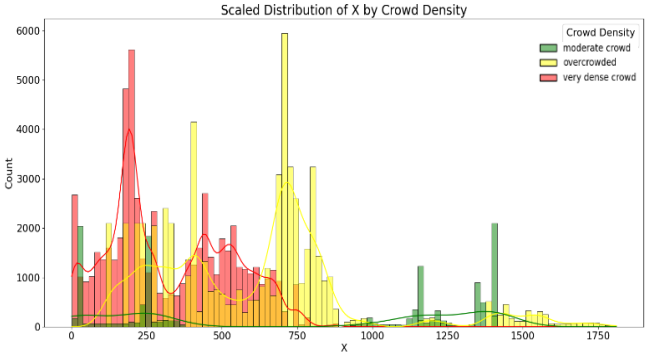


Fig. 3. The distribution of the " X " coordinate by crowd density category.

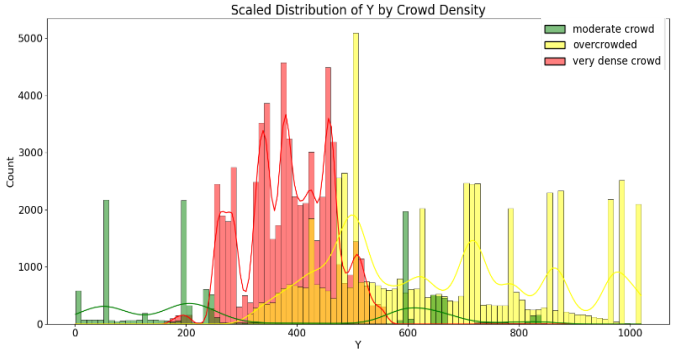


Fig. 4. The distribution of the "Y" coordinate by crowd density category.

For the distribution of the 'Edge Density' feature across the three crowd density categories, Fig. 5 shows that the very dense crowd represented by the red color, has the highest concentration of samples at larger edge density values, with a peak around 80–100.

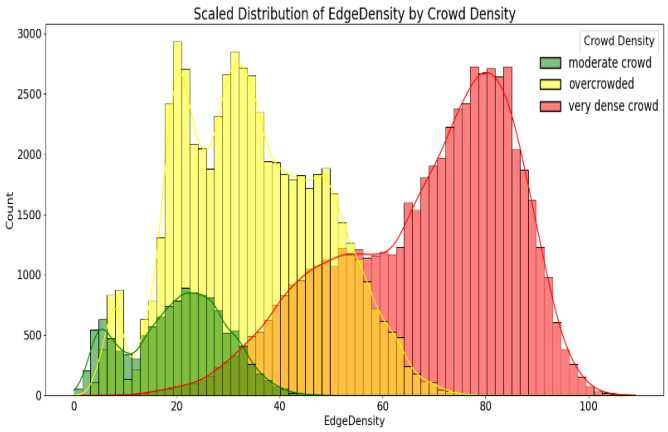


Fig. 5. Distribution of the "Edge Density" feature across crowd density categories.

Additionally, the figure shows that the overcrowded and moderate crowd categories, represented by yellow and green respectively, have a peak in the middle range of edge density values (around 40–60) and are concentrated in the lower range of edge density values below 20, respectively. This highlights the significance of edge density as a feature for classifying crowd density levels, as the distributions show clear separations, especially between the moderate and very dense crowd categories, with minimal overlap. Thus, the 'Edge Density' feature is a major factor contributing to the model's accuracy in distinguishing between these categories.

During the counting of the 'Average Intensity' feature, the visualized values show that, while there are slight shifts in the distributions as shown in Fig. 6, the very dense crowd peaks at slightly higher average intensity values around 125–150, the overcrowded category peaks in the mid-range of 100–125, and the moderate crowd is concentrated at lower intensity values below 100. All three crowd density levels overlap significantly in the range of 75 to 125 average intensity values. In addition, Fig. 5 demonstrates that average intensity captures variations in brightness across the frame that correlate with crowd density. While the overlap limits its standalone utility, it can enhance the model's accuracy when combined with other features like edge density and area.

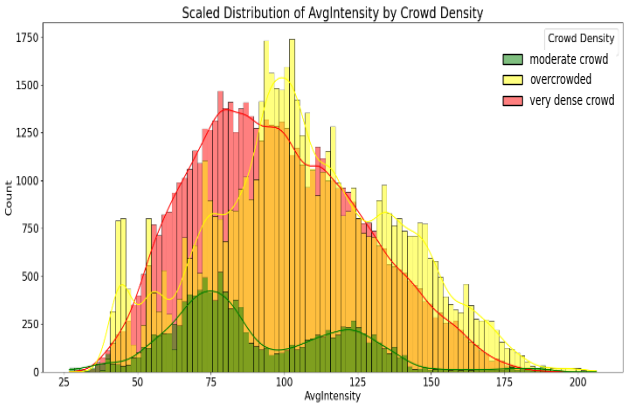


Fig. 6. Distribution of the "Average Intensity" (AvgIntensity) feature across crowd density categories.

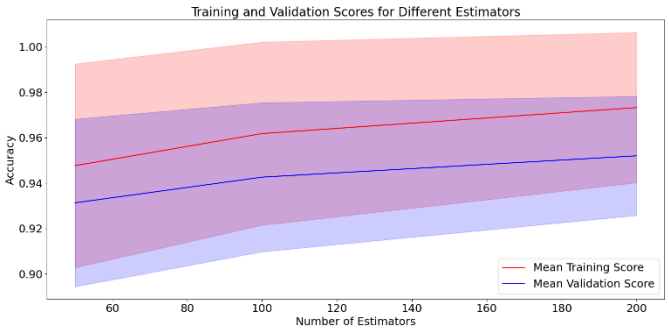


Fig. 7. Training and validation scores.

Finally, Fig. 7 shows the training and validation accuracy trends as the number of estimators increases, along with shaded regions representing the standard deviation for each. Both training and validation accuracy improve as the number of estimators increases and the model demonstrates stable performance without significant overfitting, as indicated by the relatively small gap between training and validation scores.

A. *Model Performance and Confusion Matrix Analysis*

Metrics including accuracy, precision, recall and F1-score were used to evaluate the proposed model. Fig. 8 shows the results of these metrics as 87% accuracy, 93% precision, 87% recall, and 84% F1-score, which therefore indicate the model's high reliability, especially in identifying frames with very dense crowd conditions, which is crucial for real-time monitoring systems.

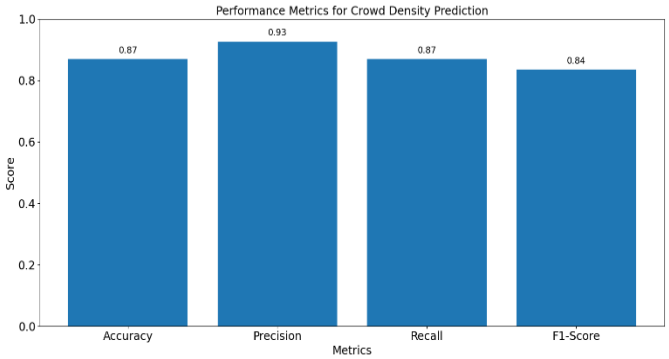


Fig. 8. Performance metrics.

In addition, a confusion matrix was constructed to analyze misclassifications. The matrix revealed that majority of misclassifications occurred between moderate crowd and overcrowded. While, frames classified as very dense crowd showed minimal false negatives thereby ensuring effective detection in critical situations. Fig. 9 presents the confusion matrix and provides a detailed breakdown of predictions versus true labels.

Moreover, an error percentage or misclassification rate based on the values of confusion matrix using the following equation:

$$Error\ Percentange = \frac{Total\ Misclassified\ Instances}{Total\ Instances} \times 100$$

$$Error\ Percentange = \frac{94}{4382} \times 100 \approx 2.14\%$$

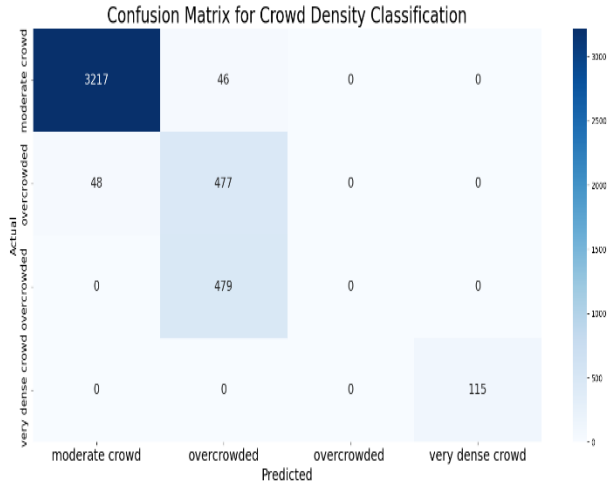


Fig. 9. Confusion matrix.

The 2.14% error percentage is ideal for high precision in a critical application such as classifying the surveillance videos of the holy sites during Hajj, especially in determining very dense crowd conditions, as false alerts or false estimations of the crowd situation could lead to either unnecessary triggering of emergency efforts or disasters. Furthermore, the model triggered visual alerts represented as red overlays for frames classified as 'very dense crowd'. This visual feedback could enhance the practical usability of the system in high-risk environments like Hajj.

The results demonstrate the success of the proposed approach in addressing crowd density classification challenges with high accuracy and low error percentage. The use of GridSearchCV for hyperparameter optimization and data augmentation for feature enhancement contributed significantly to the model's success. However, some limitations have been noted such as misclassification between moderate and overcrowded conditions, which could indicate the need for further enhancement to the methods used for extracting the features of the video frames. Also, real-time processing performance could benefit from hardware acceleration or further enhancement of the model.

Future work may look into contextual data integration such as temporal crowd flow patterns. In addition, extending the system to support multi-camera feeds could improve scalability for large-scale events like Hajj.

B. Comparsion with Previous Studies

The proposed model has been compared with other models in previous studies reviewed in Section II based on specific key evaluation criteria as shown in Table I. These criteria were collected from the previously reviewed studies. It was found that the proposed model in this study outperforms other models in its tailored feature extraction, real-time usability, focus on crowd density, and specific application to Hajj which make it uniquely suited for improving crowd management during Hajj. Although the proposed model achieves an accuracy of 87% which is slightly lower than the reported accuracy in some studies, that outperforms others in its targeted application of crowd density classification during Hajj. Unlike models that focus on detecting individual anomalies or specific behaviors, the proposed model is specifically designed to classify crowd density into actionable categories (moderate, overcrowded, very dense) and provide real-time visual alerts for very dense conditions. That is essential during Hajj to avoid crises such as the Mina stampede, which tragically led to the deaths of 1,000 pilgrims [21], and the collapse of a crawler crane at the Holy Mosque, resulting in 107 deaths and over 230 injuries [22] due to the dense crowds. That makes it more practical and reliable for real-world crowd management during Hajj.

Furthermore, with a low error rate of 2.14% and a balanced performance across metrics (precision: 93%, recall: 87%, F1-score: 84%), the proposed model ensures robust and consistent results tailored to Hajj-specific challenges.

TABLE I. COMPARISON OF PROPOSED MODEL WITH RELATED WORKS ON KEY EVALUATION CRITERIA (LEGEND √ MEANS INCLUDED/STRONGLY ADDRESSED, × MEANS NOT ADDRESSED AND ≈ MEANS PARTIALLY INCLUDED)

Criteria	Tailored Feature Extraction	Real-Time Visual Alerts	Low Misclassification Rate	Focus on Crowd Density	Application to Hajj-Specific Scenarios
Bhuiyan [4]	×	×	×	≈	√
Aldayri and Albattah [11]	×	×	×	×	×
Bhuiyan et al.[12]	×	×	×	×	≈
Nasir et al.[13]	×	×	×	×	≈
Alhothali et al.[14]	≈	×	×	×	≈
Habib et al.[15]	×	√	×	×	×
Alafif et al. [16]	×	×	×	×	≈
Miao et al.[19]	√	×	×	×	√
Alafif et al.[20]	√	×	×	×	√
Proposed model	√	√	√	√	√

C. Discussion

Even though this study uses the HAJJv2 dataset, which represents different crowd scenarios from various holy sites, such as Tawaf, Sa'i, Mina and Arafat, thereby covering different crowd density levels, biases may exist due to the under representation of certain density levels or specific scenarios, such as moderately crowded conditions during less busy times. To address this, data augmentation techniques were applied to balance the dataset and reduce classification bias. The model's performance on unseen data indicates its generalizability, though further validation with external datasets could provide additional insights into fairness.

The observed misclassifications between moderate and overcrowded categories highlight areas for improvement. Introducing additional contextual features, such as temporal crowd patterns or environmental conditions may enhance fairness and reduce boundary-level classification errors.

The proposed model in this study demonstrates trustworthiness by providing interpretable outputs based on clear and relevant features like area, edge density and spatial coordinates. These features align with real-world crowd dynamics, as evidenced by the strong correlations observed in very dense crowd situations, and the real-time visual alerts represented as red overlays for very dense crowd frames, further enhancing trust by offering actionable and transparent feedback to decision-makers.

The proposed model's low error rate of 2.14% combined with high precision 93% and recall 87% metrics ensures its reliability in critical scenarios, especially in avoiding false negatives for very dense crowd conditions. That makes the system suitable for real-time applications during high-stakes events like Hajj.

Studies like Bhuiyan et al. [12] and Alhothali et al. [14] focused on detecting individual anomalies, which differ from the comprehensive crowd density classification targeted by this study which uniquely combines tailored feature extraction with advanced machine learning techniques. Also, the proposed model balances computational efficiency with an accuracy of 87% offering a practical solution for detecting hazardous crowds during Hajj. While hybrid approaches as shown by Miao et al. [19] and Alafif et al. [20] which integrate deep learning techniques (e.g., CNNs, ResNet-50, YOLOv2) with traditional methods (e.g., Random Forests, Kalman filters) have demonstrated high accuracy in abnormal behavior detection, their reliance on computationally intensive models, advanced infrastructure (e.g., UAVs, edge-cloud models) and heavily annotated datasets often limits their scalability and real-time applicability during Hajj.

Finally, to further improve fairness and trustworthiness, future work may incorporate temporal crowd flow data, environmental factors and multi-camera feeds to enhance model robustness and scalability. Validation with external datasets can help address potential biases and improve generalizability, ensuring broader applicability to diverse crowd management scenarios.

V. Conclusion

This study proposed a model to classify crowd density in video surveillance recorded at the holy sites during Hajj into three classifications: moderate crowd, overcrowded and very dense crowd. The videos were segmented into frames, and features such as Edge Density, LBP texture, and area were extracted, which were identified as critical predictors for crowd density classification. Data augmentation and class balancing were employed to enhance the training dataset. A GBC was chosen for its robustness in handling imbalanced datasets and its ability to capture complex relationships between features. GridSearchCV was utilized to optimize the classification process, tuning key parameters such as the number of estimators, learning rate and maximum depth through cross-validation to achieve optimal performance.

The findings illustrate the capability of the proposed model to accurately classify crowd density levels. This proves its usefulness for real-time monitoring and managing crowds in critical locations.

References

[1] L. Al-Salhie, M. Al-Zuhair, and A. Al-Wabil, "Multimedia Surveillance in Event Detection: Crowd Analytics in Hajj," in Marcus, A. (eds) Design, User Experience, and Usability. User Experience Design for Diverse Interaction Platforms and Environments. DUXU 2014. Lecture Notes in Computer Science, vol 8518. Springer, Cham, 2014, pp. 383–392. doi: https://doi.org/10.1007/978-3-319-07626-3_35.

[2] E. A. Felemban et al., "Digital Revolution for Hajj Crowd Management: A Technology Survey," IEEE Access, vol. 8, pp. 208583–208609, 2020, doi: 10.1109/ACCESS.2020.3037396.

[3] F. Gazzawe and M. Albahar, "Reducing traffic congestion in makkah during Hajj through the use of AI technology," Heliyon, vol. 10, no. 1, Jan. 2024, doi: 10.1016/j.heliyon.2023.e23304.

[4] M. R. Bhuiyan, "Video analytics using deep learning for hajj pilgrimage crowd monitoring," MMU Institutional Repository, 2022.

[5] A. M. Al-Shaery et al., "Open Dataset for Predicting Pilgrim Activities for Crowd Management During Hajj Using Wearable Sensors," IEEE Access, vol. 12, pp. 72828–72846, 2024, doi: 10.1109/ACCESS.2024.3402230.

[6] W. Albattah, M. H. K. Khel, S. Habib, M. Islam, S. Khan, and K. A. Kadir, "Hajj crowd management using CNN-based approach," Computers, Materials and Continua, vol. 66, no. 2, pp. 2183–2197, 2020, doi: 10.32604/cmc.2020.014227.

[7] W. Halboob, H. Altaheri, A. Derhab, and J. Almuhtadi, "Crowd Management Intelligence Framework: Umrah Use Case," IEEE Access, vol. 12, pp. 6752–6767, 2024, doi: 10.1109/ACCESS.2024.3350188.

[8] Y. Salih and M. Simsim, "Visual Surveillance for Hajj and Umrah: A Review," IAES International Journal of Artificial Intelligence (IJ-AI, vol. 3, no. 2, pp. 90–104, 2014.

[9] A. Jabbari, "Tracking and Analysis of Pilgrims' Movement Throughout Umrah and Hajj Applying Artificial Intelligence and Machine Learning," 2023 7th International Conference On Computing, Communication, Control And Automation, ICCUBEA 2023, p. 2024, 2023, doi: 10.1109/ICCUBEA58933.2023.10392217.

[10] A. J. Showail, "Solving Hajj and Umrah Challenges Using Information and Communication Technology: A Survey," IEEE Access, vol. 10, pp. 75404–75427, 2022, doi: 10.1109/ACCESS.2022.3190853.

[11] A. Aldayri and W. Albattah, "A deep learning approach for anomaly detection in large-scale Hajj crowds," Vis Comput, vol. 40, pp. 5589–5603, 2024.

[12] M. R. Bhuiyan, J. Abdullah, N. Hashim, F. Al Farid, and J. Uddin, "Hajj pilgrimage abnormal crowd movement monitoring using optical flow and FCNN," J Big Data, vol. 10, no. 1, Dec. 2023, doi: 10.1186/s40537-023-00779-4.

[13] R. Nasir, Z. Jalil, M. Nasir, U. Noor, M. Ashraf, and S. Saleem, "An Enhanced Framework for Real-Time Dense Crowd Abnormal Behavior Detection Using YOLOv8," 2024, doi: 10.22541/au.172542291.10740660/v1.

[14] A. Alhothali, A. Balabid, R. Alharthi, B. Alzahrani, R. Alotaibi, and A. Barnawi, "Anomalous event detection and localization in dense crowd scenes," Multimed Tools Appl, vol. 82, pp. 15673–15694, 2023.

[15] S. Habib et al., "Abnormal activity recognition from surveillance videos using convolutional neural network," Sensors, vol. 21, no. 24, Dec. 2021, doi: 10.3390/s21248291.

[16] A. A. Shah, "Enhancing Hajj and Umrah Rituals and Crowd Management through AI Technologies: A Comprehensive Survey of Applications and Future Directions," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3487923.

[17] M. R. Bhuiyan et al., "A deep crowd density classification model for Hajj pilgrimage using fully convolutional neural network," PeerJ Comput Sci, vol. 8, 2022, doi: 10.7717/peerj-cs.895.

[18] T. Alafif, B. Alzahrani, Y. Cao, R. Alotaibi, A. Barnawi, and M. Chen, "Generative adversarial network based abnormal behavior detection in massive crowd videos_ a Hajj case study," J Ambient Intell Humaniz Comput, vol. 13, pp. 4077–4088, 2022.

[19] Y. Miao et al., "Abnormal Behavior Learning Based on Edge Computing toward a Crowd Monitoring System," IEEE Netw, vol. 36, no. 3, pp. 90–96, 2022, doi: 10.1109/MNET.014.2000523.

[20] T. Alafif et al., "Hybrid Classifiers for Spatio-Temporal Abnormal Behavior Detection, Tracking, and Recognition in Massive Hajj Crowds," Electronics (Switzerland), vol. 12, no. 5, Mar. 2023, doi: 10.3390/electronics12051165.

[21] Y. A Alaska, A. D. Aldawas, N. A. Aljerian, Z. A. Memish, and S. Suner, "The impact of crowd control measures on the occurrence of stampedes during Mass Gatherings," The National Center for Biotechnology Information , 2016.

[22] W. Alazmy, O. Samarkandi, and B. Williams, "The history of emergency medical services response to mass casualty incidents in disasters, Saudi Arabia," Journal of Emergency Medicine, Trauma and Acute Care, vol. 2020, no. 1, Jul. 2020, doi: 10.5339/JEMTAC.2020.3.

Towards an Ontology to Represent Domain Knowledge of Attention Deficit Hyperactivity Disorder (ADHD): A Conceptual Model

Shahad Mansour Alsaedi, Aishah Alsobhi, Hind Bitar
Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Abstract*—Attention deficit/hyperactivity disorder (ADHD) represents a highly heterogeneous and complex medical domain with numerous multidisciplinary research areas. Despite the rising number of research on the pathophysiology of ADHD, the available information in the ADHD domain is still scattered and disconnected. This research study mainly aims to develop a conceptual model of ADHD by applying knowledge engineering processes to structure the domain knowledge, elucidating key concepts and their interrelationships. The methodology for developing the conceptual model is derived from established practices in ontology construction. It adopts a hybrid approach, integrating principles from prominent methodologies such as Ontology Development 101, the Uschold and King methodology, and METHONTOLOGY. The proposed ADHD conceptual model links various aspects of ADHD including subtypes, symptoms, behaviors, diagnostic criteria, treatment, risk factors, comorbidities, and patient profile. Comprising eight top-level classes and highlighting 13 key relationships, it establishes connections between symptoms and recommended treatments, as well as symptoms and their diverse manifestations, risk factors, ADHD subtypes, and potential comorbidities. While the model captures a broad range of ADHD-related concepts, it has certain limitations. It does not extensively address genetic or neurobiological mechanisms, nor does it capture cultural and contextual variations in ADHD manifestations. These limitations highlight opportunities for future expansion, such as incorporating real-world data and diverse demographic contexts. Nevertheless, the model developed in this study is well-suited to serve as a cornerstone for constructing a comprehensive ADHD domain knowledge ontology. Ontologies play a crucial role as a layer for transferring knowledge and serve as a foundation for developing advanced systems, such as decision-support tools and expert systems, to enhance ADHD research and clinical practice.

Keywords*—Conceptual model; ontology; ADHD; knowledge engineering

I. Introduction

Attention-deficit/hyperactivity disorder (ADHD) is the most prevalent behavioral disorder and the second most frequent chronic condition among children [1]. ADHD is a neurodevelopmental disorder characterized by a recurring pattern of inattention, hyperactivity, and impulsivity [2], [3]. The symptoms of inattention, hyperactivity and impulsivity include a wide range of behaviors such as difficulty completing tasks, difficulty following instructions, making careless mistakes, and being overly impulsive [3]. These symptoms of ADHD often have a direct negative impact on the individual's academic achievement, social relationships, self-esteem, and emotional functioning [2], [4]. ADHD was already a common condition in the past. However, it is becoming more common nowadays. Over the past two decades, there has been such a significant increase in rates of ADHD diagnoses that researchers describe it as a "dramatic change" that occurred between 1997 and 2016 [5], [6]. The global prevalence of ADHD is estimated at 2% to 7%, with an average of around 5% in school-aged children [7] and 2.5% to 5% in adults [8], [9].

In the last two decades, the field of ADHD has witnessed remarkable progress in research. Different studies often explore different aspects of ADHD. It incorporates multiple disciplines such as psychology (e.g. [10], [11]), neuroscience (e.g. [12], [13]), genetics (e.g. [9], [14]), epidemiology (e.g. [15], [16]), pharmacology (e.g. [17], [18]), and psychosocial and educational interventions (e.g. [19], [20]). Thus, ADHD research is quite diverse and multi-faceted. Bridging the gap between such different disciplines is particularly difficult due to this diversity.

Despite the rising number of studies on the pathophysiology of ADHD, the disorder remains a complex psychological condition that is challenging to characterize [21]. Many previous studies have repeatedly emphasized the complex and heterogeneous nature of ADHD [22]. This heterogeneity is reflected in its wide range of psychiatric comorbidities, diverse clinical profiles, patterns of neuropsychological impairment and developmental trajectories, and a broad spectrum of structural and functional brain abnormalities [22]. In addition to its complex nature, one of the significant challenges in the ADHD research field is the scattered and unstructured nature of the available information. ADHD knowledge is dispersed across various types of studies, making it difficult for researchers, clinicians, and caregivers to access and synthesize relevant insights effectively [23]. Thus, this research focuses on addressing two core problems: the complexity and heterogeneity of ADHD and the scattered, unstructured information in the domain.

As the research environment grows increasingly complex, it is crucial to establish methods that make the diverse output of ADHD studies more accessible and integrative [24]. Recently, there has been a striking rise in the use of biomedical ontologies as the preferred method for bridging different disciplinary gaps. They ensure wide access and exchange of information among researchers and professionals from different backgrounds and a wide range of disciplines [25], [26]. The remarkable success of

biomedical ontologies is due to their power to offer a comprehensive framework within which researchers can publish their new findings and physicians can conveniently access the studies that serve as the foundation for their diagnosis and treatment approaches.

Gruber [27] defined the computer science ontology in 1993 as " explicit specification of a conceptualization". This definition placed an emphasis on the conceptualization, which serves as the foundational phase influencing subsequent processes in ontology development. Conceptualization consumes the majority of the time dedicated to ontology construction. It commences with the knowledge acquisition phase, wherein a description of the domain ontology is formulated. Subsequently, the acquired knowledge is arranged and structured within a conceptual model [28]. Kabilan [29] defined the conceptual model as a theoretical representation of a segment of reality, outlining fundamental concepts and their interrelations. It captures all relevant knowledge in a given field, organizing it into a structured and unified framework. By doing so, a conceptual model serves as the foundation for integrating and harmonizing diverse and scattered information. This leads us to the central research question of this study: How to develop a conceptual model that offers a unified framework to integrate the heterogeneous and scattered knowledge in the ADHD domain?"

As we mentioned above, the ADHD domain, with its multifaceted nature and interdisciplinary scope, presents unique challenges in structuring and consolidating knowledge. Addressing the complexity and scattered nature of ADHD-related information requires a systematic approach to conceptualization, one that captures the domain's key components and their intricate interrelations. Recognizing this need, in this paper, we mainly aim to develop a conceptual model that describes the reality of the ADHD domain with its key concepts and relationships. The resulting conceptual model aims primarily to be a basis for building an ADHD domain knowledge ontology in order to provide a step towards reducing the knowledge gap in the field of ADHD.

The rest of this paper is organized as follows: Section II demonstrates the Related work, and the research methodology is detailed in section III. The findings of this research are explained in Section IV. The discussion is presented in Section V before concluding the work.

II. Related Work

Conceptual models have been widely used in the study of neurodevelopmental disorders to provide structured frameworks for understanding the relationships between symptoms, risk factors, functional impairments, and outcomes. These models are essential for integrating knowledge across disciplines, identifying intervention targets, and advancing clinical and research practices. They play a crucial role in organizing and integrating knowledge about complex disorders such as ADHD. However, within the field of ADHD, there is a notable paucity of comprehensive conceptual models that unify its various dimensions. Existing efforts often focus on isolated aspects, leaving gaps in understanding the disorder holistically. The following section highlights existing conceptual models in ADHD research and underscores the gaps this study aims to address.

Rapport et al. [30] introduced a conceptual model that highlights the underlying assumptions about what causes ADHD and how it is treated through behavioral and pharmacological approaches. The model suggests that biological factors, such as genetics or prenatal complications, lead to differences in how the brain's systems function. For example, changes in neurotransmitter systems like dopamine and norepinephrine are thought to be directly responsible for the main psychological features of ADHD, including cognitive and behavioral symptoms. The model also explains that other characteristics of ADHD, often called "associated features" in the DSM-IV (e.g., academic struggles, poor social skills, or family conflicts), are secondary effects caused by these core symptoms. For instance, a child's academic difficulties might result from their inability to pay attention or persist in tasks over time. Eventually, this model categorizes therapeutic interventions into pharmacological treatment, cognitive behavioral therapy (CBT), and skills training, correlating each type to be specifically directed to treat one aspect of the disorder: pharmacological treatment for biological factors, CBT for cognitive and behavioral symptoms (subtypes), and skills training for associated features.

Brod et al. [31] in their study focuses on understanding how ADHD symptoms and impairments impact the quality of life (QoL) in adults. The authors proposed a conceptual model based on data from experts, patients, and literature. The model identifies five areas of impact: work, daily activities, relationships, psychological well-being, and physical well-being, which are grouped into three key QoL domains—productivity, relationships, and health. The model highlights how ADHD symptoms interact synergistically, creating cascading impairment pathways that affect daily functioning. To operationalize the model, the Adult ADHD QoL Measure was developed, enabling clinicians to assess ADHD's impact on QoL and design more comprehensive, individualized treatment plans.

In the same context of ADHD, Dosreis & Myers [32] propose a conceptual model explaining how parents of children with ADHD recognize problems, seek medical evaluations, and decide to pursue treatment. Based on interviews with 48 parents, the model identifies three phases: (1) problem recognition, (2) motivation for evaluation, and (3) therapeutic intervention. Parental decisions are shaped by life circumstances (e.g., family changes, financial shifts), experiences at home, school, and within cultural/religious settings. The model distinguishes between family subgroups based on their help-seeking behaviors, offering insights for tailored, patient-centered interventions to enhance treatment delivery and family engagement.

In neurodevelopmental disorders fields such as autism spectrum disorder (ASD), dyslexia, and child psychopathology, numerous conceptual models have been developed to address the complexity and heterogeneity of these conditions. Alsobhi et al. [33] propose a conceptual model to support personalized learning experiences for students with dyslexia. The conceptual model links dyslexia types with assistive technologies and learning styles to create tailored educational materials. The

model includes key components such as Dyslexia Type, Assistive Technologies, and Learning Style, ensuring flexibility and adaptability. By aligning learning resources with student needs, the model provides a structured framework to enhance e-learning outcomes for dyslexic learners.

These studies illustrate the growing recognition of the importance of conceptual models in the field of neurodevelopmental disorders. They highlight the complexity of these conditions and the need for frameworks that integrate various factors. However, there remains a significant gap in developing comprehensive, unified models, which points to the need for continued research and further work to build more inclusive and practical models for understanding and addressing neurodevelopmental disorders such as ADHD.

III. METHODOLOGY

This paper embarks on a journey into the methodology employed for crafting not only a robust ontology but also the consequential development of an associated conceptual model. Recognizing the symbiotic relationship between these two endeavors, we acknowledge that the path toward a comprehensive ontology necessitates a deliberate and strategic approach in conceptualization. The conceptualization process produces the conceptual model which contributes to the ontology development process by enhancing clarity through its graphical representation, making it easier to understand and utilize. Additionally, the conceptual model promotes reusability, as it can be adapted into various ontology representation languages [28]. Consequently, our methodology for conceptual model development is basically derived from and informed by the proven techniques and practices established in the realm of ontology construction.

To develop the conceptual model proposed by the current study, a hybrid methodology of the most prominent ontology construction methodologies was adopted, drawing upon the foundational principles recommended by the Ontology Development 101 method [34], the Uschold and King methodology [35] and the METHONTOLOGY [36]. Specifically, the foundational framework presented by Uschold and King will be integrated through the iterative phases of the Ontology Development 101 method, complemented by the crucial conceptualization step outlined in METHONTOLOGY. The selection of this hybrid approach is grounded in its capacity to leverage the respective merits of each methodology. The Uschold and King methodology affords a clear and systematic framework for ontology development, facilitating the establishment of an initial ontology structure. Complementarily, the Ontology Development 101 method empowers us to incorporate finer-grained details and engage in iterative stages, thereby fostering the creation of a more thorough and comprehensive ontology. Furthermore, the integration of METHONTOLOGY augments our approach by harnessing the distinct benefits offered by its conceptualization step. This encourages a deep understanding of the domain and helps ensure that the conceptual model captures the semantics accurately, thereby enriching the overall development process of the ontology. Using these methodologies, the conceptual model is constructed following a set of prescribed guidelines and procedural steps: (1) Identify the domain, scope, and purpose, (2) Capture the knowledge and conceptualization by a set of iterative steps (Enumerate important terms in the domain, Define the classes, and define the properties (object and data properties), (3) Build the conceptual model). Fig. 1 illustrates the workflow of the methodology.

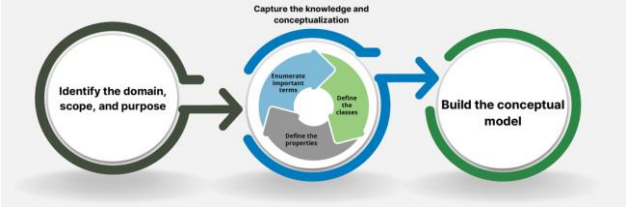


Fig. 1. The workflow of the conceptual model building methodology.

A. Identify the Domain, Scope, and Purpose

The foundational phase in constructing a robust conceptual model, and equally essential in ontology development, involves a meticulous exploration to identify the fundamental aspects that shape its essence. At the forefront of this endeavor is the imperative task of pinpointing the domain, determining the scope, and elucidating the purpose that the conceptual model aims to serve [34], [35]. This initial step serves as the compass, guiding the subsequent development process. Through a comprehensive examination of the domain's intricacies, a clear delineation of the model's boundaries, and a lucid definition of its overarching objectives, this step lays the groundwork for a conceptual model poised to effectively encapsulate and represent the targeted knowledge domain. Recognizing that this step is equally pivotal in constructing ontologies underscores its significance in fostering a seamless transition from conceptualization to formal knowledge representation. Table I illustrates the domain scope of the proposed conceptual model.

B. Capture the Knowledge and Conceptualization

1) Enumerate important terms in the domain: The starting point in Capturing the knowledge in a field is to write down a list of all relevant terms in the field. As outlined in the 101 method [34], the focus is on obtaining a comprehensive list of terms, regardless of any overlap between the concepts they represent, relations among the terms, or any properties that the concepts may possess. In the context of ADHD, essential terms include but are not limited to Symptoms, Inattention, Hyperactivity, Impulsivity, Subtypes (such as predominantly inattentive, predominantly hyperactive-impulsive, combined), Diagnostic criteria [43] (e.g., DSM-5 criteria), Treatment options (e.g., medication, behavioral therapy), Medication (e.g., stimulants, non-stimulants), Behavioral therapy, Risk factors, Genetics, Environmental factors, Comorbidities (e.g., anxiety disorders, depression), Executive dysfunction, Neurodevelopmental disorder, Neurobiology, Cognitive impairments, Academic performance ,Impairment in daily functioning. Subsequent steps in the knowledge capture process build upon this foundational list, often executed iteratively (as depicted in Fig. 1). From this list, the most significant terms are selected to represent the classes forming the conceptual model, with the remaining terms likely serving as subclasses or properties that provide detailed descriptions of these classes.

TABLE I. SCOPE, DOMAIN AND KNOWLEDGE SOURCE OF THE ADHD CONCEPTUAL MODEL

Domain	The domain of interest of this work is the ADHD domain
Date	2023 - 2024
Built By	Research student in Information Systems department at the Faculty of Computing and Information Technology – King Abdulaziz University
Purpose	• Knowledge Representation Blueprint: Acting as a blueprint, the conceptual model outlines key concepts, relationships, and entities crucial for accurate and comprehensive conceptualization and representation of the available knowledge in various aspects of ADHD. • Foundation for Ontology Construction: The conceptual model serves as the foundational framework upon which the subsequent ontology is constructed. • Defining Ontological Boundaries: By identifying the domain, scope, and purpose, the conceptual model aids in clearly defining the boundaries and parameters that subsequently shape the ontology. • Enhancing Ontological Consistency: It contributes to ensuring consistency and coherence in the ontology by presenting a conceptual framework that aligns with the nuances and intricacies of the targeted domain. • Facilitating Seamless Transition: The conceptual model facilitates a seamless transition from conceptualization to formal ontology construction, streamlining the overall knowledge modeling process.
Scope	The scope of the conceptual model is a manifestation of the domain knowledge encapsulated within the semantic model. Specifically, the ADHD domain conceptual model is finely tuned to address the intricacies of the ADHD domain. It covers aspects which include: ADHD types, symptoms, Diagnosis criteria and treatments, Comorbidities …
Source of Knowledge	• The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, (DSM-5) [3] • The World Health Organization (WHO) Eleventh Revision of the International Classification of Diseases (ICD-11) [2] • The National Institutes of Health (NIH) [37] • The American Academy of Pediatrics (AAP) [38] • American Psychological Association (APA) [39] • Children and Adults with Attention-Deficit/Hyperactivity Disorder (CHADD) organization [40] • Centers for Disease Control and Prevention (CDC) [41] • Clinical Expert interview • The first researcher also reviewed various materials, including websites and academic papers, to gain comprehensive insights into the specified domain.

2) Define the classes: As we delve into constructing the conceptual model for the ADHD domain, a crucial initial step involves defining classes. The top-level classes of the ADHD conceptual model (illustrated in Table II) act as fundamental building blocks, representing key concepts and terms integral to the realm of ADHD. The process of defining classes not only brings clarity to our conceptual model but also lays the groundwork for developing an organized and thorough representation of ADHD, forming the basis for future ontology construction.

3) Define the properties: As outlined in study [34], properties serve as a form of association between concepts within the domain, describing relationships among individuals belonging to different classes. Typically, ontologies include ordered binary relations where one argument signifies the domain and the other represents the range. These relations are called Object properties which establish connections between individuals of the domain class and those of the range class. Object properties may possess corresponding inverse properties. For example, the properties treatedBy (s, t) and isTreatmentFor (t, s) that relate a symptom with a treatment are inverse properties. Additionally, properties may include cardinality constraints, specifying the number of relationships an individual can engage in for a given property. Table III provides an overview of the properties within the ADHD conceptual model, detailing their domain, range, inverse property (if applicable), and cardinality.

TABLE II. DESCRIPTION OF THE ADHD CONCEPTUAL MODEL MAIN CLASSES

Class	Description
Subtypes	Describes the three different subtypes of ADHD
Symptoms	A compilation of signs and symptoms endorsed by DSM-5, indicative of ADHD
Behaviors	A compilation of diverse expressions and behaviors exhibited by individuals, each representing varied manifestations falling under each specific symptom
Diagnostic Criteria	A compilation of diagnostic criteria outlined in DSM-5, officially endorsed for the identification and diagnosis of ADHD
Risk Factors	A list of genetic, environmental, and psychosocial factors that could potentially contribute to the manifestation of ADHD symptoms
comorbidities	Different common coexisting conditions or comorbidities frequently observed in individuals diagnosed with ADHD
Treatment	Different safe and effective treatments and management interventions suitable for individuals with ADHD throughout their lifespan
Patient profile	Personal medical history profile and information

TABLE III. ADHD CONCEPTUAL MODEL PROPERTIES

Property Name	Domain	Range	Cardinality	Inverse property
hasSymptoms	Subtypes	Symptoms	Multiple: a Subtype may have more than one Symptoms	hasSubtype
hasDiagnosticCriteria	Subtypes	Diagnostic Criteria	Multiple: a Subtype may have more than one Symptoms	-
hasBehviors	Symptoms	Behaviors	Multiple: a Symptom may have more than one Behaviors	hasRelatedSymptoms
hasDSM5Criteria	Symptoms	Diagnostic Criteria	Single: a Symptom relates to one DSM 5 Criteria	-
treatedBy	Symptoms	Treatment	Multiple: a Symptom may have more than one Treatments	isTreatmentFor
hasInfluenceOn	Risk Factors	Subtypes	Single: a risk factor may have an influence on the occurrence of one Subtype	influencedBy
hasInfluenceOn	Risk Factors	comorbidities	Multiple: a risk factor may have an influence on the occurrence of more than one comorbidity	-
hasSymptoms	Patient profile	Symptoms	Multiple: a patient may manifest more than one Symptoms	-
hasRiskFactors	Patient profile	Risk Factors	Multiple: a patient may have more than one Risk Factors	-

IV. ADHD CONCEPTUAL MODEL: UNVEILING THE INTERCONNECTED LANDSCAPE OF ADHD DOMAIN KNOWLEDGE

Fig. 2 illustrates the ADHD conceptual model that is meticulously crafted in an attempt to contribute to elucidate its multifaceted nature. This conceptual model serves as a guiding framework, delineating top-level classes alongside their corresponding properties and interrelationships. Each of these classes embodies a pivotal concept within the ADHD domain, encompassing a spectrum of subtypes, symptoms, behaviors, risk factors, comorbidities, treatment, and diagnostic criteria.

The relationships between these classes are not merely arbitrary connections but rather solid interconnections firmly grounded in evidence derived from reputable scientific research. Through a meticulous synthesis of findings from reliable sources, the conceptual model unveils the connections that exist between these fundamental concepts, serving as the initial blueprint for the ADHD domain knowledge ontology. As we navigate through this conceptual model, we embark on a journey toward a deeper understanding of ADHD, guided by the structured interplay of its constituent elements.

The 'Symptom' class is one of the most central concepts within the domain of ADHD, as underscored by its pervasive association with numerous concepts throughout the conceptual model. The relationship 'has_DSM5_criteria' between the symptom class and the diagnostic criteria class, elucidating the direct correlation between identified symptoms and the diagnostic standards delineated in DSM-5 [3]. This alignment with DSM-5 criteria, acknowledged as the authoritative reference for diagnostic guidelines worldwide, substantiates the reliability and validity of symptoms enumerated within the symptom class.

The 'Symptom' class further establishes a relationship with the 'Behavior' class through the designated relationship 'has_related_behaviors' and its corresponding inverse relationship 'has_related_symptoms'. While DSM-5 delineates diagnostic symptoms characteristic of ADHD, empirical evidence from numerous studies suggests that these symptoms may manifest diversely as behavioral patterns across different developmental stages—children, adolescents, and adults alike [35], [36], [37]. This interrelation serves to bridge each symptom outlined in DSM-5 with its varied behavioral manifestations across distinct age cohorts, thereby enhancing the diagnostic precision tailored to the nuanced needs of different age groups. Consequently, this alignment holds promise for refining the diagnostic process, rendering it more comprehensive and tailored to the developmental context of individuals presenting with ADHD symptoms.

Symptoms wield significant influence in subtype delineation within the disorder, thereby informing potential treatment modalities. Consequently, it is associated with the 'Subtypes' class through the relationship (has_subtype) and its inverse relationship (has_symptoms), as well as with the 'Treatments' class via another designated relationship (treated_by) and its corresponding inverse relationship (is a treatment for). This interconnection underscores the pivotal role of symptomatology not only in subtype classification but also in guiding treatment and interventions approaches. Such relational frameworks within the ontology facilitate a comprehensive understanding of the intricate interplay between symptoms, subtypes, and treatment strategies, thereby enriching the diagnostic and therapeutic landscape of ADHD management.

The relationship between the 'Risk Factors', 'Subtypes', and 'Comorbidities' classes within the ADHD ontology is illustrated in the conceptual model through relationships (has_influence_on) (a_risk_factor_for). Research, including the study by Freittag et al. [38], as well as numerous corroborating studies, have underscored the influential role of various risk factors in shaping both ADHD subtypes manifestations and comorbidities. By elucidating these relationships, the ontology contributes to the understanding of how risk factors contribute to the heterogeneity of ADHD presentation and the onset of comorbidities discussion.

In this paper, we propose a cohesive conceptual model that integrates various concepts of ADHD, including symptoms, subtypes, diagnostic criteria, treatment, risk factors, comorbidities, and patient profile. These concepts are related to each other through many relationships between them. This conceptual model aims to be the cornerstone for building an ontology of existing knowledge within the domain of ADHD.

A. Conceptual Models in the ADHD Field

The In the realm of ADHD, few works have delved into conceptualizing knowledge within a conceptual framework, while existing studies have made important contributions toward understanding various aspects of ADHD, the development of comprehensive conceptual models remains limited.

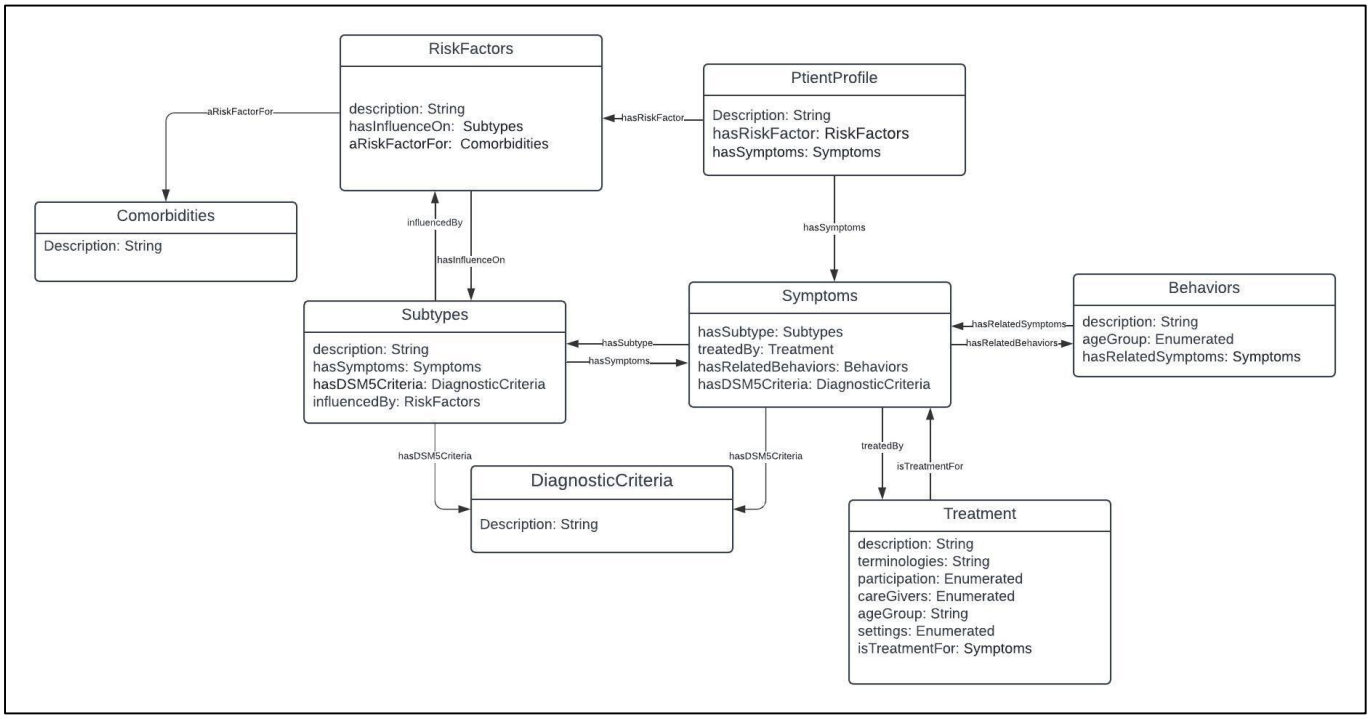


Fig. 2. ADHD conceptual model.

Such as symptoms, risk factors, or treatment pathways without integrating these elements into a unified framework. This highlights a clear need for a holistic approach that bridges these gaps, providing a more structured and interconnected representation of ADHD domain knowledge. The proposed work aims to address this gap by presenting a conceptual model that unifies these diverse components.

Unlike previous works in the field of ADHD, which focus solely on developing conceptual models, this research uniquely aims to transform a conceptual model into an ontology. Ontologies play a crucial role as a knowledge transfer layer, enabling the structured representation and integration of ADHD domain knowledge. By formalizing key concepts and relationships, an ontology serves as a foundation for building advanced systems such as decision-making systems and expert systems. These systems can enhance diagnosis, treatment planning, and research, ultimately improving outcomes for clinicians, researchers, and individuals with ADHD.

Building on the studies discussed in the "Related Work" section, the following section examines the similarities and differences between these studies and the current research, highlighting the unique contributions of this study. Rapport et al. [30] introduced a conceptual model focusing on the classification of therapeutic interventions for children with ADHD. However, given the disorder's inherent complexity and heterogeneity, such arbitrary classifications may not always prove effective. In the context of ADHD, several critical clinical considerations influence the selection of treatment strategies. These include factors such as the patient's age, the severity of ADHD symptoms [42], and the presence of comorbidities [46]. For instance, the (AAP) strongly advocates for behavioral therapy as the initial treatment approach for children aged 4-5 years, emphasizing the importance of evaluating its efficacy before considering pharmacological interventions [47]. In contrast to Rapport's model [30], the proposed ADHD conceptual model associates treatments primarily with symptoms. This association arises from the acknowledgment that there is no definitive cure for ADHD; rather, treatments focus on managing and alleviating the symptoms exhibited by individual patients [22], [48]. Additionally, it considers patient age through the "age group" property within the treatment class and addresses ADHD severity based on the number and type of symptoms exhibited by the patient.

Additionally, Brod et al. [31] introduced a conceptual model focused on assessing ADHD's impact on the quality of life among adults with the disorder. Central to their model is the interplay between ADHD symptoms and associated behaviors, including executive dysfunction. This reciprocal relationship underscores the intricate connection between symptoms and their outward manifestations, a principle mirrored in the findings of our proposed conceptual model (the relationship between Symptoms and behaviors classes). Both models highlight the complex dynamics between ADHD symptoms and their associted behaviors, emphasizing the multifaceted nature of the disorder's impact on individuals' lives. By recognizing this intricate relationship, our model aligns with Brod et al.'s [31] conceptual framework, contributing to a more comprehensive understanding of ADHD and its implications for quality of life.

B. Theoretical Implications on the Research Landscape in the ADHD Field

In advancing the theoretical conceptualization of ADHD, researchers have made significant strides through diverse studies. For instance, the review by Nigg et al. [49] underscores the importance of adopting a broader conceptualization of ADHD beyond the specific inattention and hyperactivity

symptoms. Instead, it raises the question of the possibility of inclusion of a wide spectrum of phenotypes that should be encompassed within the diagnostic framework of ADHD, such as emotion dysregulation and executive dysfunction. This perspective has been a focal point of numerous studies within the field, emphasizing the need for a comprehensive understanding of ADHD manifestations (e.g. [44]). The conceptual model presented in this paper directly addresses this imperative by integrating diverse viewpoints and establishing connections between symptoms delineated in the DSM-5 (Symptoms class) and the multifaceted array of phenotypes (Behaviors class), spanning functional and emotional dimensions.

Also, Faraone et al. [50] offers a comprehensive examination of ADHD, encompassing many of the core concepts of the disorder like diagnosis, treatment, and associated risk factors and comorbidities. Similarly, Drechsler et al. [51] delve into the current concepts, comorbidities and treatment in children and adolescents with ADHD. Additionally, Zayats & Neale [52] investigates how the genetic and neurobiological foundations [45] underpinnings on our conceptualization of ADHD, focusing specifically on two key aspects within the field: diagnosis and treatment. The present study, along with existing literature, reveals commonalities regarding fundamental concepts of ADHD, emphasizing the importance of understanding ADHD symptoms, subtypes, and their implications for diagnosis and treatment. Additionally, risk factors and comorbidities are recognized as significant contributors to the heterogeneity and complexity of ADHD presentations. Although these shared concepts provide a foundational conceptualizing of ADHD, they underscore the need for comprehensive models that can account for its multifaceted nature. While each of these works offers valuable contributions to the field, they often focus on specific aspects of ADHD in isolation. They frequently do not establish connections between different domain concepts, leading to a fragmented understanding of the disorder. The present research endeavors to bridge these conceptual gaps by proposing a comprehensive conceptual model that integrates multiple facets of ADHD. By elucidating the interconnectedness of these concepts, this research aims to provide a more holistic and cohesive framework for understanding ADHD. This fundamental difference in the theoretical implications of the present research and previous research in the field of ADHD emphasizes the importance of adopting an integrated approach to conceptualizing the disorder.

C. Practical Applications in the ADHD Field

In terms of contributing to practical applications in the field of ADHD, there are many knowledge-based systems that mainly rely on "background knowledge" as their input. For example, Göker & Tekedere [53] implemented an expert system that predicts ADHD during childhood, Silva et al. [54], Delavarian et al. [55] and Chu et al. [56] developed decision support systems with various goals serving the field. When data is collected specifically for each system individually the knowledge collected to develop these systems is exclusive to them and was not transferable to other systems, leaving future projects with no way of benefitting from what they had achieved. This means that any time a new system is created, it has to reinvent the wheel and go through the same process of collecting knowledge from scratch. Considering the complexity of ADHD domain knowledge and related sources that are of many sorts and sources, there is an increasing need for a unified knowledge base that provides easy access to information and the ability to share, transfer and reuse the knowledge. As previously mentioned, the primary objective of constructing this conceptual model is to serve as the foundational framework for the development of a comprehensive domain knowledge ontology pertaining to ADHD. It also plays a major role as a layer for transferring and reusing domain knowledge [57]. In general, the utilization of ontologies for knowledge structuring enables the creation of reusable and shareable knowledge that, once developed, can be used entirely or partially by anybody [58]. Therefore, building a conceptual model is frequently considered as the first step in formalizing and standardizing domain knowledge while building various systems.

V. CONCLUSION

In conclusion, this research has developed a comprehensive conceptual model that encapsulates the domain knowledge of ADHD, shedding light on its multifaceted nature and interrelationships between key concepts. By adopting a hybrid approach derived from established practices in ontology construction, the model effectively integrates principles from prominent methodologies. The resulting ADHD conceptual model delineates connections between various aspects of the disorder, including subtypes, symptoms, behaviors, diagnostic criteria, treatment options, risk factors, comorbidities, and patient profiles. Through its 8 top-level classes and 13 relationships, the model elucidates critical links between symptoms and treatments, symptom manifestations, risk factors and ADHD subtypes, and potential comorbidities. By bridging these domain concepts, the model contributes to a more holistic understanding of ADHD and serves as a foundational framework for constructing an ADHD domain knowledge ontology. This research thus plays a pivotal role in advancing knowledge within the field of ADHD and offers valuable insights for researchers, clinicians, and policymakers navigating the multidisciplinary landscape of ADHD research and practice.

Despite its contributions, this study has limitations that present opportunities for future research. While the model captures a broad range of ADHD-related concepts, it does not delve deeply into the genetic underpinnings or specific neurobiological mechanisms of the disorder, which are increasingly recognized as critical to understanding ADHD's etiology. Similarly, cultural and contextual variations in the manifestation, diagnosis, and treatment of ADHD remain underexplored, leaving gaps in how the disorder is understood and addressed across diverse populations. The model also relies predominantly on existing literature and expert consensus, potentially overlooking emerging research and patient perspectives, particularly those from underrepresented demographic groups.

Future work should address these limitations by incorporating real-world data, such as longitudinal studies and patient-reported outcomes, to validate and expand the model's applicability. Efforts to integrate genetic and neurobiological insights could enrich the conceptual framework, providing a

more comprehensive understanding of ADHD's mechanisms. Additionally, capturing cultural and contextual variations could make the model more globally relevant and inclusive. Expanding the scope to include emerging research trends and interdisciplinary perspectives will further enhance the model's utility in research and clinical practice.

The most critical next step is to use the resulting conceptual model as the foundation for constructing a comprehensive ontology for ADHD. This ontology would formalize the domain knowledge, enabling standardized representation, enhanced interoperability across systems, and the development of tools for diagnosis, treatment planning, and research. By advancing from conceptual modeling to ontology development, this work can catalyze significant progress in understanding and addressing ADHD, ultimately improving outcomes for individuals affected by this complex disorder.

ACKNOWLEDGMENT

The The first author conducted a number of interviews with doctors: Dr. Marwa Al-Aoufi, Dr. Abdul-Halim, and the psychological specialist Al-Anoud Al-Shehri, in the Department of Growth and Behavior, King Salman bin Abdulaziz Medical City, Medina, Kingdom of Saudi Arabia.

REFERENCES

[1] M. L. Wolraich et al., "ADHD Diagnosis and Treatment Guidelines: A Historical Perspective," Pediatrics, vol. 144, no. 4, p. e20191682, Oct. 2019, doi: 10.1542/peds.2019-1682.

[2] "ICD-11 for Mortality and Morbidity Statistics." Accessed: Jan. 07, 2023. [Online]. Available: https://icd.who.int/browse11/l-m/en#/http%3a%2f%2fid.who.int%2ficd%2fentity%2f821852937

[3] "DSM-5.pdf(Shared) - Adobe cloud storage." Accessed: Jan. 05, 2023. [Online]. Available: https://acrobat.adobe.com/link/track?uri=urn%3Aaaid%3Ascds%3AUS%3A907fa51f-b6cb-494c-95b1-5cacf626fc55&viewer%21megaVerb=group-discover

[4] S. Cortese and D. Coghill, "Twenty years of research on attention-deficit/hyperactivity disorder (ADHD): looking back, looking forward," Evidence-Based Mental Health, vol. 21, no. 4, pp. 173–176, Nov. 2018, doi: 10.1136/ebmental-2018-300050.

[5] "Is There an Increase in ADHD?," CHADD. Accessed: Dec. 20, 2022. [Online]. Available: https://chadd.org/adhd-weekly/is-there-an-increase-in-adhd/

[6] "General Prevalence of ADHD," CHADD. Accessed: Jan. 16, 2023. [Online]. Available: https://chadd.org/about-adhd/general-prevalence/

[7] K. Sayal, V. Prasad, D. Daley, T. Ford, and D. Coghill, "ADHD in children and young people: prevalence, care pathways, and service provision," The Lancet Psychiatry, vol. 5, no. 2, pp. 175–186, Feb. 2018, doi: 10.1016/S2215-0366(17)30167-0.

[8] P. Song, M. Zha, Q. Yang, Y. Zhang, X. Li, and I. Rudan, "The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis," J Glob Health, vol. 11, p. 04009, doi: 10.7189/jogh.11.04009.

[9] G. C. Akutagava-Martins, L. A. Rohde, and M. H. Hutz, "Genetics of attention-deficit/hyperactivity disorder: an update," Expert Review of Neurotherapeutics, vol. 16, no. 2, pp. 145–156, Feb. 2016, doi: 10.1586/14737175.2016.1130626.

[10] M. W. Handler and G. J. DuPaul, "Assessment of ADHD: Differences Across Psychology Specialty Areas," J Atten Disord, vol. 9, no. 2, pp. 402–412, Nov. 2005, doi: 10.1177/1087054705278762.

[11] M. J. Manos, "Nuances of assessment and treatment of ADHD in adults: A guide for psychologists," Professional Psychology: Research and Practice, vol. 41, pp. 511–517, 2010, doi: 10.1037/a0021476.

[12] K. Rubia, "Cognitive Neuroscience of Attention Deficit Hyperactivity Disorder (ADHD) and Its Clinical Translation," Frontiers in Human Neuroscience, vol. 12, 2018, Accessed: Feb. 17, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2018.00100

[13] M. Milham, D. Fair, M. Mennes, and S. Mostofsky, "The adhd-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience," Frontiers in Systems Neuroscience, vol. 6, 2012, Accessed: Feb. 17, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnsys.2012.00062

[14] D. Demontis et al., "Discovery of the first genome-wide significant risk loci for attention-deficit/hyperactivity disorder," Nat Genet, vol. 51, no. 1, pp. 63–75, Jan. 2019, doi: 10.1038/s41588-018-0269-7.

[15] A. S. Rowland, C. A. Lesesne, and A. J. Abramowitz, "The epidemiology of attention-deficit/hyperactivity disorder (ADHD): A public health view," Mental Retardation and Developmental Disabilities Research Reviews, vol. 8, no. 3, pp. 162–170, 2002, doi: 10.1002/mrdd.10036.

[16] F. E. de la Barra, B. Vicente, S. Saldivia, and R. Melipillan, "Epidemiology of ADHD in Chilean children and adolescents," ADHD Atten Def Hyp Disord, vol. 5, no. 1, pp. 1–8, Mar. 2013, doi: 10.1007/s12402-012-0090-6.

[17] F. D. Mowlem, M. A. Rosenqvist, J. Martin, P. Lichtenstein, P. Asherson, and H. Larsson, "Sex differences in predicting ADHD clinical diagnosis and pharmacological treatment," Eur Child Adolesc Psychiatry, vol. 28, no. 4, pp. 481–489, Apr. 2019, doi: 10.1007/s00787-018-1211-3.

[18] F. D. Crescenzo, S. Cortese, N. Adamo, and L. Janiri, "Pharmacological and non-pharmacological treatment of adults with ADHD: a meta-review," BMJ Ment Health, vol. 20, no. 1, pp. 4–11, Feb. 2017, doi: 10.1136/eb-2016-102415.

[19] N. Zendarski et al., "Examining the Educational Gap for Children with ADHD and Subthreshold ADHD," J Atten Disord, vol. 26, no. 2, pp. 282–295, Jan. 2022, doi: 10.1177/1087054720972790.

[20] G. J. DuPaul, S. W. Evans, J. A. Mautone, J. S. Owens, and T. J. Power, "Future Directions for Psychosocial Interventions for Children and Adolescents with ADHD," Journal of Clinical Child & Adolescent Psychology, vol. 49, no. 1, pp. 134–145, Jan. 2020, doi: 10.1080/15374416.2019.1689825.

[21] C.-H. Lin, T.-W. Chien, and Y.-H. Yan, "Predicting the number of article citations in the field of attention-deficit/hyperactivity disorder (ADHD) with the 100 top-cited articles since 2014: a bibliometric analysis," Ann Gen Psychiatry, vol. 20, no. 1, p. 6, Jan. 2021, doi: 10.1186/s12991-021-00329-3.

[22] Y. Luo, D. Weibman, J. M. Halperin, and X. Li, "A Review of Heterogeneity in Attention Deficit/Hyperactivity Disorder (ADHD)," Frontiers in Human Neuroscience, vol. 13, 2019, Accessed: Feb. 02, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00042

[23] F. M. Abomelha, H. AlDhalaan, M. Ghaziuddin, N. A. Al-Tassan, and B. R. Al-Mubarak, "Autism and ADHD in the Era of Big Data; An Overview of Digital Resources for Patient, Genetic and Clinical Trials Information," Genes, vol. 13, no. 9, Art. no. 9, Sep. 2022, doi: 10.3390/genes13091551.

[24] O. Bodenreider, "Biomedical Ontologies in Action: Role in Knowledge Management, Data Integration and Decision Support," Yearb Med Inform, vol. 17, no. 1, pp. 67–79, 2008, doi: 10.1055/s-0038-1638585.

[25] M. Gong et al., "Toward early diagnosis decision support for breast cancer: Ontology-based semantic interoperability.," JCO, vol. 37, no. 15_suppl, pp. e18072–e18072, May 2019, doi: 10.1200/JCO.2019.37.15_suppl.e18072.

[26] F. Ali et al., "A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion," Information Fusion, vol. 63, pp. 208–222, Nov. 2020, doi: 10.1016/j.inffus.2020.06.008.

[27] T. R. Gruber, "A translation approach to portable ontology specifications," Knowledge Acquisition, vol. 5, no. 2, pp. 199–220, Jun. 1993, doi: 10.1006/knac.1993.1008.

[28] "An Ontological Approach to Model Software Quality Assurance Knowledge Domain.pdf."

[29] V. Kabilan, Ontology for information systems (04IS) design methodology: conceptualizing, designing and representing domain

ontologies. Kista: Data- och systemvetenskap, Kungliga Tekniska högskolan, 2007.

[30] M. D. Rapport, K. M. Chung, G. Shore, and P. Isaacs, "A conceptual model of child psychopathology: implications for understanding attention deficit hyperactivity disorder and treatment efficacy," J Clin Child Psychol, vol. 30, no. 1, pp. 48–58, Mar. 2001, doi: 10.1207/S15374424JCCP3001_6.

[31] M. Brod, A. Perwien, L. Adler, T. Spencer, and J. Johnston, "Conceptualization and Assessment of Quality of Life for Adults with Attention-Deficit/Hyperactivity Disorder," Primary Psychiatry, vol. 12, no. 6, pp. 58–64, 2005.

[32] S. dosReis and M. A. Myers, "Parental attitudes and involvement in psychopharmacological treatment for ADHD: A conceptual model," International Review of Psychiatry, vol. 20, no. 2, pp. 135–141, Jan. 2008, doi: 10.1080/09540260801933084.

[33] A. Y. Alsobhi, N. Khan, and H. Rahanu, "Personalised Learning Materials Based on Dyslexia Types: Ontological Approach," Procedia Computer Science, vol. 60, pp. 113–121, Jan. 2015, doi: 10.1016/j.procs.2015.08.110.

[34] H. P. P. Filho, "Ontology Development 101: AGuide to Creating Your First Ontology".

[35] M. Uschold and M. King, "Towards a Methodology for Building Ontologies".

[36] M. Fernández-López, A. Gomez-Perez, and N. Juristo, "METHONTOLOGY: from ontological art towards ontological engineering," Engineering Workshop on Ontological Engineering (AAAI97), Mar. 1997.

[37] "National Institutes of Health (NIH)," National Institutes of Health (NIH). Accessed: Jul. 25, 2024. [Online]. Available: https://www.nih.gov/

[38] "Home | AAP." Accessed: Jul. 25, 2024. [Online]. Available: https://www.aap.org/?srsltid=AfmBOopHZ4_xCWrHPZoq1chDM-vBrMG17At2pvj4ClVR92_g3MomhwAi

[39] "American Psychological Association (APA)," https://www.apa.org. Accessed: Jul. 25, 2024. [Online]. Available: https://www.apa.org

[40] "CHADD - Improving the lives of people affected by ADHD," CHADD. Accessed: Jul. 25, 2024. [Online]. Available: https://chadd.org/

[41] CDC, "Centers for Disease Control and Prevention." Accessed: Jul. 25, 2024. [Online]. Available: https://www.cdc.gov/index.html

[42] "The Relationship Between Executive Function Deficits and DSM-5-Defined ADHD Symptoms - Michael J. Silverstein, Stephen V. Faraone, Terry L. Leon, Joseph Biederman, Thomas J. Spencer, Lenard A. Adler, 2020." Accessed: Jul. 25, 2024. [Online]. Available: https://journals.sagepub.com/doi/abs/10.1177/1087054718804347?journalCode=jada

[43] "ADHD in Adults: New Symptom Tests, Diagnostic Criteria Needed." Accessed: Jul. 25, 2024. [Online]. Available: https://www.additudemag.com/adhd-in-adults-new-diagnostic-criteria/

[44] L. A. Adler, S. V. Faraone, T. J. Spencer, P. Berglund, S. Alperin, and R. C. Kessler, "The structure of adult ADHD," Int J Methods Psychiatr Res, vol. 26, no. 1, p. e1555, Feb. 2017, doi: 10.1002/mpr.1555.

[45] C. M. Freitag et al., "Biological and psychosocial environmental risk factors influence symptom severity and psychiatric comorbidity in children with ADHD," J Neural Transm, vol. 119, no. 1, pp. 81–94, Jan. 2012, doi: 10.1007/s00702-011-0659-9.

[46] A. Caye, J. M. Swanson, D. Coghill, and L. A. Rohde, "Treatment strategies for ADHD: an evidence-based guide to select optimal treatment," Mol Psychiatry, vol. 24, no. 3, Art. no. 3, Mar. 2019, doi: 10.1038/s41380-018-0116-3.

[47] S. C. on Q. I. and M. Subcommittee on Attention-Deficit/Hyperactivity Disorder, "ADHD: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents," Pediatrics, vol. 128, no. 5, pp. 1007–1022, Nov. 2011, doi: 10.1542/peds.2011-2654.

[48] "Attention deficit hyperactivity disorder (ADHD) - Treatment," nhs.uk. Accessed: Jan. 07, 2023. [Online]. Available: https://www.nhs.uk/conditions/attention-deficit-hyperactivity-disorder-adhd/treatment/

[49] J. T. Nigg, M. H. Sibley, A. Thapar, and S. L. Karalunas, "Development of ADHD: Etiology, Heterogeneity, and Early Life Course".

[50] S. V. Faraone et al., "The World Federation of ADHD International Consensus Statement: 208 Evidence-based conclusions about the disorder," Neuroscience & Biobehavioral Reviews, vol. 128, pp. 789–818, Sep. 2021, doi: 10.1016/j.neubiorev.2021.01.022.

[51] R. Drechsler, S. Brem, D. Brandeis, E. Grünblatt, G. Berger, and S. Walitza, "ADHD: Current Concepts and Treatments in Children and Adolescents," Neuropediatrics, vol. 51, pp. 315–335, Jun. 2020, doi: 10.1055/s-0040-1701658.

[52] T. Zayats and B. M. Neale, "Recent advances in understanding of attention deficit hyperactivity disorder (ADHD): how genetics are shaping our conceptualization of this disorder," F1000Res, vol. 8, p. F1000 Faculty Rev-2060, Feb. 2020, doi: 10.12688/f1000research.18959.2.

[53] H. Göker and H. Tekedere, "Dynamic Expert System Design for the Prediction of Attention Deficit and Hyperactivity Disorder in Childhood," Bilişim Teknolojileri Dergisi, vol. 12, no. 1, Art. no. 1, Jan. 2019, doi: 10.17671/gazibtd.458102.

[54] S. D. Silva, S. Dayarathna, G. Ariyarathne, D. Meedeniya, S. Jayarathna, and A. M. P. Michalek, " Computational Decision Support System for ADHD Identification ," IJAC, vol. 18, no. 2, pp. 233–255, Dec. 2020, doi: 10.1007/s11633-020-1252-1.

[55] M. Delavarian, F. Towhidkhah, P. Dibajnia, and S. Gharibzadeh, "Designing a Decision Support System for Distinguishing ADHD from Similar Children Behavioral Disorders," J Med Syst, vol. 36, no. 3, pp. 1335–1343, Jun. 2012, doi: 10.1007/s10916-010-9594-9.

[56] K.-C. Chu, Y.-S. Huang, C.-F. Tseng, H.-J. Huang, C.-H. Wang, and H.-Y. Tai, "Reliability and validity of DS-ADHD: A decision support system on attention deficit hyperactivity disorders," Computer Methods and Programs in Biomedicine, vol. 140, pp. 241–248, Mar. 2017, doi: 10.1016/j.cmpb.2016.12.003.

[57] X. Xing, B. Zhong, H. Luo, H. Li, and H. Wu, "Ontology for safety risk identification in metro construction," Computers in Industry, vol. 109, pp. 14–30, Aug. 2019, doi: 10.1016/j.compind.2019.04.001.

[58] E. Strakhovich, Ontological Engineering in Education: Tools for Knowledge Transfer and Knowledge Assessment. 2014, p. 715. doi: 10.1109/ICALT.2014.207.

Leiden Coloring Algorithm for Influencer Detection

Handrizal[1]*, Poltak Sihombing[2], Erna Budhiarti Nababan[3], Mohammad Andri Budiman[4]
Student Doctoral Program in Computer Science[1]
Department of Computer Science-Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia[1, 2, 4]
Department of Information Technology-Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia[3]

***Abstract*—In today's digital age, the role of influencers, especially on social media platforms, has grown significantly. A commonly used feature by business professionals today is follower grouping. However, this feature is limited to identifying influencers based solely on mutual followership, highlighting the need for a more sophisticated approach to influencer detection. This study proposes a novel method for influencer detection that integrates the Leiden coloring algorithm and Degree centrality. This approach leverages network analysis to identify patterns and relationships within large-scale datasets. Initially, the Leiden coloring algorithm is employed to partition the network into various communities, considered potential influencer hubs. Subsequently, Degree centrality is utilized to identify nodes with high connectivity, indicating influential individuals. The proposed method was validated using data crawled from Twitter (X) social media, employing the keyword "GarudaIndonesia." The data was collected using Tweet-Harvest between January 1, 2020, and October 16, 2024, resulting in a dataset of 22,623 rows. The dataset was subjected to two experimental scenarios: 1,000 and 5,000 rows. Compared to the Louvain coloring method, the proposed approach demonstrated an increase in the modularity value of the Leiden coloring algorithm by 0.0306, a reduction in time processing by 14.4848 seconds, and a decrease in the number of communities by 1,290.**

***Keywords*—Influencer; Louvain coloring; Leiden; Leiden coloring**

I. Introduction

The role of influencers, especially on social media platforms, has grown significantly, necessitating the development of more sophisticated influencer detection methods [1]. Influencer detection is a part of community detection. Businesses must accurately identify customers and respond to their needs to remain competitive [2]. As the business landscape evolves, businesses are increasingly adopting digital marketing strategies to keep pace with the competition [3].

Digital marketing involves promoting and disseminating information, as well as searching for markets, through digital media by utilizing various means such as social media [4]. Digital marketing simplifies the process for businesses to connect with and satisfy the desires of potential consumers [5]. From the perspective of potential consumers, digital marketing provides easy access to product information through cyberspace, making it faster and more convenient to search for information [6]. To increase their customer base, businesses need to identify individuals or groups who can help market their products or services. These individuals or groups, with the potential to attract more consumers, are known as influencers [7].

Contemporary digital marketing applications, particularly on Twitter (X) [8], provide features such as follower grouping to assist businesses in promoting their products.

However, the follower grouping feature is limited to identifying influencers based solely on mutual followership, without providing information about the topics discussed by the group or the size and influential members of the group. This can be challenging for business owners who are new to using social media as a promotional tool. Therefore, a method is needed to detect influencers based on specific topics or keywords using Social Network Analysis (SNA) on platforms like Twitter (X), represented graphically. One widely used SNA method is community detection.

There are several community detection algorithms, such as the Louvain algorithm [9] and the Leiden algorithm [10]. The latter was created to identify communities in large networks with complex modularity patterns. The Leiden algorithm is one of the approaches used in social network analysis [11], aiming to identify subgraph groups with strong internal connections. This can help understand the social structure and relationships within a social network, but the algorithm has limitations, such as difficulty in interpreting community visualizations due to the lack of color coding. Graph coloring can accelerate the process by assigning unique colors to nodes, similar to indexing in a relational database. Based on this, this study aims to influencer detection using the Leiden algorithm combined with graph coloring. By incorporating color coding into community detection, we can improve the interpretability and effectiveness of the results.

II. Materials and Methods

A. Leiden Algorithm

The Leiden algorithm is an algorithm used for community detection in networks or graphs. Designed to work efficiently, the Leiden algorithm detects communities in networks or graphs based on modularity optimization [13]. Modularity is a measure that describes the extent to which the density of connections within a community exceeds that of a random network. The modularity value is a scale value ranging from [-1,1]. The modularity of a network is calculated using the following formula:

*Corresponding Author. Email ID: handrizal@usu.ac.id

$$Q = \frac{1}{2m}\sum A_{i,j} - \frac{k_i k_j}{2m}\delta(c_i, c_j)$$

where:

- $A_{i,j}$ represents the edge weights between nodes i and j
- k_i and k_j are the sum of the weights of the edges attached to nodes i and j
- m is the sum of all edge weights in the graph
- c_i and c_j are the node communities
- δ is the Kronecker delta function
- $(\delta(c_i, c_j) = 1$ if $c_i = c_j$, otherwise 0

The Leiden algorithm comprises three primary stages:

1) Moving nodes: Iteratively moving nodes between communities to optimize the overall modularity of the network.

2) Refinement: Dividing the network into distinct, connected components by separating unconnected communities

3) Aggregation: Form a hierarchical network by iteratively aggregating nodes into communities and then treating each community as a single node in the next level of the hierarchy

Steps in the Leiden Algorithm:

1) Initialization: Create a representation of the graphunder analysis. This can be achieved using either an adjacency list or an adjacency matrix. Subsequently, each node within the graph is initialized as an independent community

2) Iteration: Repeat the following steps iteratively until convergence is attained:

a) Local Move Step

- Assess the feasibility of relocating each node within the graph to a different community.
- Calculate the change in modularity resulting from the movement of the node
- If a community exists that provides a greater increase in modularity, relocate the node accordingly.

b) Community Aggregation

- Once all nodes have been evaluated, nodes within the same community are merged into a single new node
- Construct a new graph in which each node corresponds to a community generated in the preceding step

c) Weight Update

- Recompute the edge weights in the new graph according to the number of nodes represented in each community.

3) Termination: The iteration process continues until a stable state is reached, where neither community assignments nor modularity values change significantly.

B. Graph Coloring

Graph coloring, a technique introduced in 1979 [14], involves assigning colors to vertices in a graph such that no adjacent vertices share the same color. This study explores the application of graph coloring [12] within the context of network analysis. Consider a graph G(V, E), where V represents the set of vertices and E represents the set of edges in the graph. We aim to assign a color, w(i) ∈ {1, 2, ..., k}, to each vertex i ∈ V, where k is the number of colors used, such that:

$$\forall (i,j) \in G, w(i) \neq w(j)$$

C. Leiden Coloring Algorithm

In 2024, the study in [15] stated that one of the shortcomings of the Leiden algorithm is its inability to efficiently process large networks, resulting in relatively long processing times.To address this issue, the Leiden algorithm was modified using graph coloring, hereinafter referred to as Leiden Coloring.

The steps of the Leiden coloring algorithm are as follows:

- Phase 1: Community Detection

Community detection within the graph is performed using the Leiden algorithm. The implementation strictly adheres to the algorithmic steps and mathematical formulations definedby the Leiden algorithm.

- Phase 2: Community Coloring

After community identification, the graph coloring principle is applied to the resulting community structure. Each community is assigned a unique color, ensuring that no more colors are used than the number of identified communities.

This process effectively integrates the Leiden algorithm with graph coloring, resulting in each node being labeled with both its community membership and a unique color.

The stages of influencer detection research using Leiden coloring are as follows:

- Business understanding

This stage involves the collection of information about influencers, encompassing indicators that can identify potential influencers, relevant phenomena, and pertinent facts.

- Twitter data collection

Data collection for this study was conducted on Twitter (X) using the Tweet-Harvest service. Tweet-Harvest was employed to retrieve tweets related to the keyword "GarudaIndonesia" from January 1, 2020, to October 16, 2024. This process resulted in a dataset of 22,623 tweets.

- Network construction

This stage of dataset processing encompasses data selection based on research objectives, subsequent data cleaning to eliminate irrelevant entries, and finally, data transformation into a graph format for further analysis.

- Community detection

After the preceding stage, the dataset will undergo further analysis for influencer detection utilizing Social Network

Analysis methodologies. Both the Leiden and Leiden Coloring algorithms will be employed in this phase.

- Analysis results

This stage involves analyzing, concluding, and evaluating the results of influencer detection from the previous step. The information generated may include graph visualizations, the number of communities formed, and the identification of influencers within each community.

- Evaluation

At this stage, the algorithm's performance is evaluated using three matrices: modularity, time processing, and the number of communities to determine the algorithm's overall effectiveness.

III. Results and Discussion

This study employed the Leiden coloring algorithm for influencer detection, conducting tests with two datasets: a smaller dataset of 1,000 rows and a larger dataset of 5,000 rows. The objective was to analyze how the identification of influential individuals within a social network might be affected by the size and scale of the available data.

A. Detection of Influencers using the Leiden Coloring Algorithm

This section presents the results obtained from the proposed Leiden coloring algorithm. These results encompass the influencer detection matrix, modularity values, processing times, and the number of identified communities. Two dataset testing scenarios were conducted in this study. The first scenario utilized a dataset comprising 1,000 rows, while the second scenario employed a dataset consisting of 5,000 rows.

1) Results of influencer detection with dataset 1000 rows: This section presents the results obtained from the proposed Leiden coloring algorithm. These results are presented as an influencer detection matrix using a dataset of 1,000 rows.

TABLE I. Results of Influencer Detection with Dataset 1000 Rows

No.	Leiden Coloring Algorithm
1	IndonesiaGaruda
2	GarudaCares
3	wandiseptian11
4	PinterPoin
5	idbcpr

Table I shows that in the dataset containing 1000 rows, the username 'IndonesiaGaruda' holds the top influencer position, while 'idbcpr' ranks fifth.

2) Result of influencer detection with dataset 5000 rows: This section presents the results obtained from the proposed Leiden coloring algorithm. These results are presented as an influencer detection matrix using a dataset of 5,000 rows.

TABLE II. Result of Influencer Detection with Dataset 5000 Rows

No.	Leiden Coloring Algorithm
1	IndonesiaGaruda
2	disemuacom
3	GarudaCares
4	astuceclover
5	TiketPesawatPro

According to Table II, in the 5000-row dataset, the username 'IndonesiaGaruda' is identified as the top influencer, while 'TiketPesawatPro' is ranked fifth in terms of influencer.

B. Result of Matrix Modularity

This section presents the results obtained from the proposed Leiden coloring algorithm. The results are shown as a modularity matrix using datasets containing 1,000 rows and 5,000 rows.

TABLE III. Results of Matrix Modularity

No.	Dataset	Modularity of Leiden Coloring
1	1000 rows	0.9396
2	5000 rows	0.9381
Average		0.9388

Table III shows the modularity values for the 1000-row dataset (0.9396) and the 5000-row dataset (0.9381). Additionally, the matrix above is presented graphically in Fig. 1.

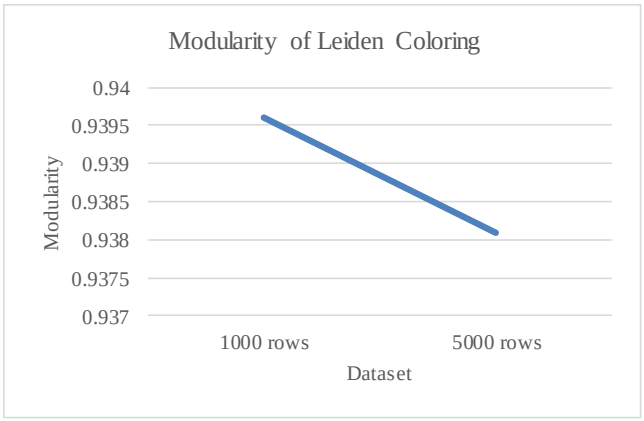


Fig. 1. Results of matrix modularity.

C. Result of Matrix Time Processing

This section presents the results obtained from the proposed Leiden coloring algorithm. The results are shown as a time-processing matrix using datasets containing 1,000 rows and 5,000 rows.

TABLE IV. Result of Matrix Time Processing

No.	Dataset	Time Processing Leiden Coloring (second)
1	1000 rows	29.5493
2	5000 rows	434.1838
Average		231,8666

In Table IV, it can be seen that for a dataset with 1000 rows, the processing time is 29.5491 seconds, and for a dataset with 5000 rows, it is 434.1838 seconds. The matrix above is also presented in graphical form, as shown in Fig. 2.

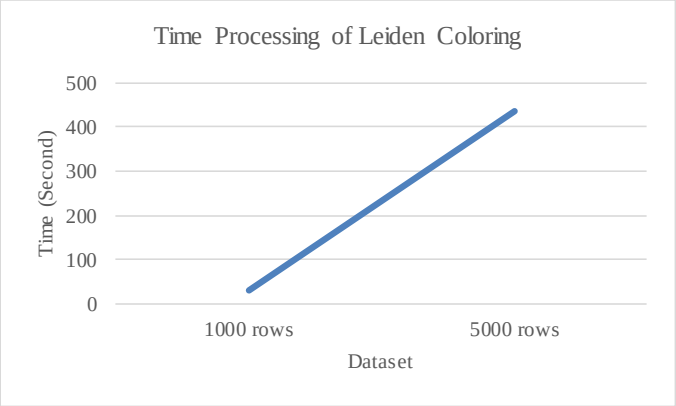


Fig. 2. Results of matrix time processing.

D. Results of Matrix Number Communities

This section presents the results obtained from the proposed Leiden coloring algorithm. The results are shown as a number of community matrix using datasets containing 1,000 rows and 5,000 rows.

TABLE V. RESULT OF MATRIX NUMBER OF COMMUNITIES

No.	Dataset	Number Communities Leiden Coloring
1	1000 rows	505
2	5000 rows	1969
Average		1237

As shown in Table V, the number of communities for the 1000-row dataset is 505, whereas for the 5000-row dataset, it is 1969. The corresponding matrix is graphically depicted in Fig. 3.

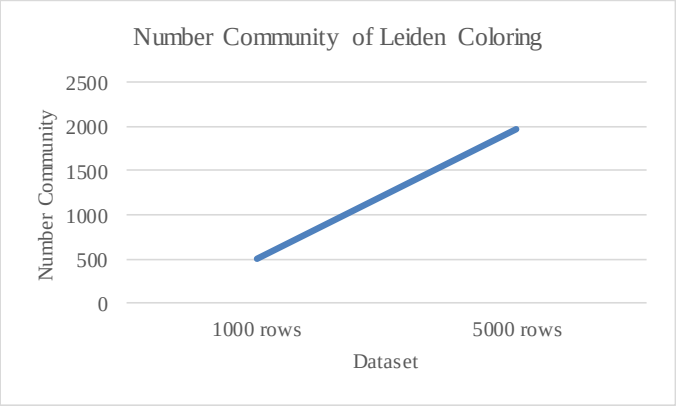


Fig. 3. Results of matrix number communities.

In the graph in Fig. 3 above, it can be seen that the number of communities for the 1,000-row dataset is 505, and it increases to 1,969 for the 5,000-row dataset.

E. Comparison of Louvain Coloring and Leiden Coloring Algorithm

This section presents the results of a comparison between the proposed Leiden coloring algorithm and the Louvain coloring algorithm, focusing on influencer detection, modularity value, time processing, and the number of identified communities. Two testing scenarios were conducted with varying dataset sizes: the first using a dataset with 1,000 rows, and the second using a dataset with 5,000 rows.

1) Comparison of influencer detection with dataset 1000 rows: This section presents a comparison of the results obtained using the proposed Leiden coloring algorithm with the Louvain coloring algorithm based on the influencer detection metric, using datasets containing 1,000 rows.

TABLE VI. COMPARISON OF INFLUENCER DETECTION WITH DATASET 1000 ROWS

Louvain Coloring Algorithm	Leiden Coloring Algorithm
IndonesiaGaruda	IndonesiaGaruda
GarudaCares	GarudaCares
astuceclover	wandiseptian11
TiketPesawatPro	PinterPoin
disemuacom	idbcpr

Table VI shows that the Louvain and Leiden coloring algorithms produce identical influencer detections for the first and second ranks in the 1000-row dataset: 'IndonesiaGaruda' and 'GarudaCares', respectively. However, the rankings diverge starting from the third rank.

2) Comparison of influencer detection with dataset 5000 rows: This section presents a comparison of the results obtained using the proposed Leiden coloring algorithm with the Louvain coloring algorithm based on the influencer detection metric, using datasets containing 5,000 rows.

TABLE VII. COMPARISON OF INFLUENCER DETECTION WITH DATASET 5000 ROWS

Louvain Coloring Algorithm	Leiden Coloring Algorithm
IndonesiaGaruda	IndonesiaGaruda
GarudaCares	disemuacom
TiketPesawatPro	GarudaCares
disemuacom	astuceclover
astuceclover	TiketPesawatPro

In Table VII, it can be seen that the Louvain coloring and Leiden coloring algorithms on the 5000-row dataset produce the same influencer detection for the first rank, namely the username 'IndonesiaGaruda.' Meanwhile, the second to fifth ranks produce different influencer detections.

Analysis of Tables VI and VII reveals that both the Louvain and Leiden coloring algorithms identify 'IndonesiaGaruda' as the top-ranked influencer. However, discrepancies in influencer rankings emerge from the second to the fifth positions.

F. Comparison of Matrix Modularity

This section presents a comparison of the results obtained using the proposed Leiden coloring algorithm with the Louvain coloring algorithm based on the modularity metric, using datasets containing 1,000 rows and 5,000 rows (Table VIII).

TABLE VIII. COMPARISON OF MATRIX MODULARITY

Dataset	Modularity	
	Louvain Coloring	Leiden Coloring
1000 rows	0.9114	0.9396
5000 rows	0.9050	0.9381
Average	0.9082	0.9388

The Leiden coloring algorithm consistently demonstrated higher modularity values compared to the Louvain coloring algorithm across all two test scenarios. The modularity values for the Leiden coloring algorithm ranged from a minimum of 0.9367 to a maximum of 0.9396, with an average of 0.9388. In contrast, the Louvain coloring algorithm exhibited a lower range, with a minimum of 0.9050 a maximum of 0.9252, and an average of 0.9082. This translates to an average increase in modularity of 0.0306 for the Leiden coloring algorithm. A comparative graph illustrating this modularity matrix is presented in Fig. 4.

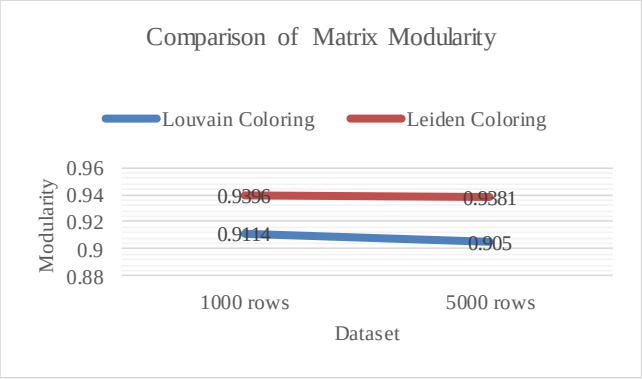


Fig. 4. Comparison of matrix modularity.

G. Comparison of Matrix Time Processing

This section presents a comparison of the results obtained using the proposed Leiden coloring algorithm with the Louvain coloring algorithm based on the time processing metric, using datasets containing 1,000 rows and 5,000 rows (Table IX).

TABLE IX. COMPARISON OF MATRIX TIME PROCESSING

Dataset	Time Processing (second)	
	Louvain Coloring	Leiden Coloring
1000 rows	41.85	29.5493
5000 rows	450.86	434.1838
Average	246,355	231,8666

The time processing of the Leiden coloring algorithm is better than that of the Louvain coloring algorithm. In the two test scenarios conducted, the Leiden coloring algorithm outperforms the Louvain coloring algorithm in all scenarios. The processing time for the Leiden coloring algorithm ranges from a minimum of 29.5493 seconds to a maximum of 434.1838 seconds, with an average of 246.355 seconds. Meanwhile, the Louvain coloring algorithm has a minimum value of 41.85 seconds, a maximum of 450.86 seconds, and an average of 231.8666 seconds. Thus, the Leiden coloring algorithm shows a reduction in processing time of 14.4848 seconds. The comparison of these processing times is also illustrated in the graph in Fig. 5.

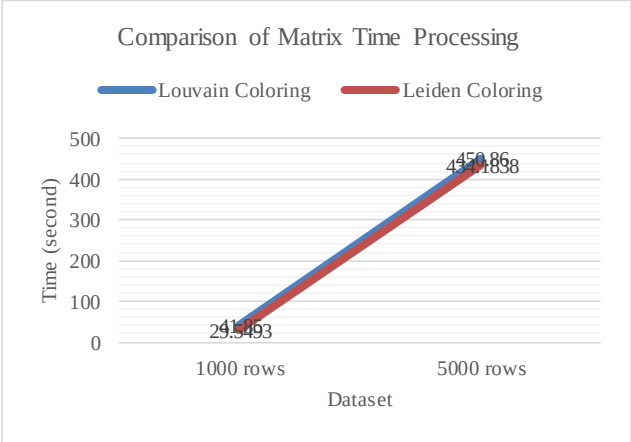


Fig. 5. Comparison of matrix time processing.

H. Comparison of Matrix Number Communities

This section presents a comparison of the results obtained using the proposed Leiden coloring algorithm with the Louvain coloring algorithm based on the number of communities metric, using datasets containing 1,000 rows and 5,000 rows (Table X).

TABLE X. COMPARISON OF MATRIX NUMBER COMMUNITIES

Dataset	Number Communities	
	Louvain Coloring	Leiden Coloring
1000 rows	936	505
5000 rows	4119	1969
Average	2527	1237

The Leiden coloring algorithm produced a significantly lower number of communities compared to the Louvain coloring algorithm across the two test scenarios. The number of communities detected by the Leiden algorithm ranged from 505 to 1969 with an average of 1237, while the Louvain coloring algorithm produced a range of 936 to 4119 with an average of 2527. This resulted in a reduction of 1290 communities on average. A comparison graph of the community numbers for both algorithms is presented in Fig. 6.

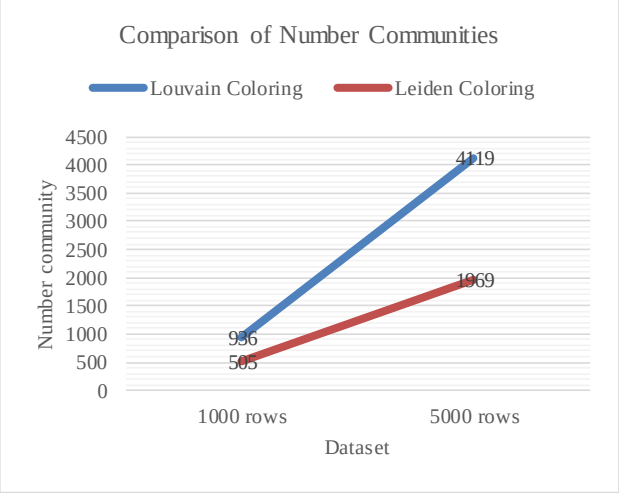


Fig. 6. Comparison of matrix number communities.

IV. Conclusion

This study successfully applied the Leiden coloring algorithm to identify influencers on Twitter (X). Influencer detection was conducted using the Leiden coloring algorithm on a dataset of Twitter (X) interactions related to "GarudaIndonesia." Data was collected using Tweet-Harvest between January 1, 2020, and October 16, 2024, resulting in a dataset of 22,623 tweets.

The dataset was evaluated using two experimental scenarios: one with 1,000 tweets and another with 5,000 tweets. The analysis identified five influential usernames in each scenario. For the 1,000-tweet dataset, the identified influencers were IndonesiaGaruda, GarudaCares, Wandiseptian11, PinterPoin, and idbcpr. For the 5,000-tweet dataset, the identified influencers were IndonesiaGaruda, disemuacom, GarudaCares, astuceclover, and TiketPesawatPro.

Compared to the Louvain coloring algorithm, the Leiden coloring algorithm demonstrated improved performance. The Leiden coloring algorithm exhibited a 0.0306 increase in modularity value, a 14.4848-second reduction in processing time, and a 1290-community reduction.

The primary contributions of this study include improved modularity, reduced processing time, and a more concise community structure when using the Leiden coloring algorithm for influencer detection.

Several limitations and suggestions have emerged during this study, such as the rapid development of Twitter (X) data due to bold processing methods, highlighting the need for new approaches to detect influencers. Therefore, future research could focus on applying the Leiden coloring algorithm to real-time Twitter (X) data.

References

[1] A. Rababah, L. Al-Haddad, M. S. Sial, Z. Chunmei, and J. Cherian, "Analyzing the effects of COVID-19 pandemic on the financial performance of Chinese listed companies," J. Public Aff., vol. 20, no. 4, 2020, doi: 10.1002/pa.2440.

[2] D. Ushakov, E. Dudukalov, E. Kozlova, and K. Shatila, "The Internet of Things impact on smart public transportation," Transp. Res. Procedia, vol. 63, pp. 2392–2400, 2022, doi: 10.1016/j.trpro.2022.06.275.

[3] Y. J. Purnomo, "Digital Marketing Strategy to Increase Sales Conversion on E-commerce Platforms," J. Contemp. Adm. Manag., vol. 1, no. 2, pp. 54–62, Aug. 2023, doi: 10.61100/ADMAN.V1I2.23.

[4] M. T. Khanom, "Using social media marketing in the digital era: A necessity or a choice," Int. J. Res. Bus. Soc. Sci. (2147- 4478), vol. 12, no. 3, pp. 88–98, 2023, doi: 10.20525/ijrbs.v12i3.2507.

[5] M. K. Peter and M. Dalla Vecchia, The Digital Marketing Toolkit: A Literature Review for the Identification of Digital Marketing Channels and Platforms, vol. 294, no. March. Springer International Publishing, 2021. doi: 10.1007/978-3-030-48332-6_17.

[6] D. R. Piranda, D. Z. Sinaga, and E. E. Putri, "ONLINE MARKETING STRATEGY IN FACEBOOK MARKETPLACE AS A DIGITAL MARKETING TOOL," J. Humanit. Soc. Sci. Bus., vol. 1, no. 3, pp. 1–8, Mar. 2022, doi: 10.55047/JHSSB.V1I2.123.

[7] D. Vrontis, A. Makrides, M. Christofi, and A. Thrassou, "Social media influencer marketing: A systematic review, integrative framework and future research agenda," Int. J. Consum. Stud., vol. 45, no. 4, pp. 617–644, Jul. 2021, doi: 10.1111/IJCS.12647.

[8] S. S. Veleva and A. I. Tsvetanova, "Characteristics of the digital marketing advantages and disadvantages," IOP Conf. Ser. Mater. Sci. Eng., vol. 940, no. 1, 2020, doi: 10.1088/1757-899X/940/1/012065.

[9] V. D. Blondel, J. L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech. Theory Exp., vol. 2008, no. 10, Mar. 2008, doi: 10.1088/1742-5468/2008/10/p10008.

[10] V. Traag, L. Waltman, and N. J. van Eck, "From Louvain to Leiden: guaranteeing well-connected communities," Sci. Rep., vol. 9, no. 1, Oct. 2018, doi: 10.1038/s41598-019-41695-z.

[11] F. Nguyen, "Leiden-Based Parallel Community Detection," no. September, 2021, [Online]. Available: www.kit.edu

[12] Ü. V. Çatalyürek, J. Feo, A. H. Gebremedhin, M. Halappanavar, and A. Pothen, "Graph coloring algorithms for multi-core and massively multithreaded architectures," Parallel Comput., vol. 38, no. 10–11, pp. 576–594, Oct. 2012, doi: 10.1016/J.PARCO.2012.07.001.

[13] S. H. H. Anuar et al., "Comparison between Louvain and Leiden Algorithm for Network Structure: A Review," J. Phys. Conf. Ser., vol. 2129, no. 1, p. 012028, Dec. 2021, doi: 10.1088/1742-6596/2129/1/012028.

[14] D. Brélaz, "New methods to color the vertices of a graph," Commun. ACM, vol. 22, no. 4, pp. 251–256, Apr. 1979, doi: 10.1145/359094.359101.

[15] S. Sahu, K. Kothapalli, and D. S. Banerjee, "Fast Leiden Algorithm for Community Detection in Shared Memory Setting," ACM Int. Conf. Proceeding Ser., pp. 11–20, 2024, doi: 10.1145/3673038.3673146.

Construction and Optimal Control Method of Enterprise Information Flaw Risk Contagion Model Based on the Improved LDA Model

Jun Wang[1]*, Zhanhong Zhou[2]
Yiwu Industrial & Commercial College, Yiwu, Zhejiang, 322000, China[1]
School of International Economics and Trade, Lanzhou University of Finance and Economics, Lanzhou Gansu 730010, China[2]

***Abstract*—In this study, we construct a risk contagion model for corporate information disclosure using complex network methods and incorporate the manipulative perspective of management tone into it. We employ an enhanced LDA model to analyze and refine the relevant data and models presented in this paper. The results of quantitative analysis show that the improved LDA algorithm optimizes the classification decision boundary, making similar samples closer and different samples more dispersed, thus improving the classification accuracy. Additionally, we combine multi-objective evolutionary optimization techniques with an improved particle swarm optimization algorithm to solve the proposed model while incorporating enhancements through the use of weighted Smote algorithm. The quantization results show that using the weighted Smote algorithm to deal with the imbalance in the dataset significantly improves the classification performance. Furthermore, we compare our proposed method with classical algorithms on four real enterprise information disclosure datasets and observe that our approach exhibits higher efficiency and accuracy compared to traditional optimal control methods. Accounting information disclosure reveals moral hazard and adverse selection, alleviating information asymmetry. Transparent information improves the availability of financing, preventing liquidity risk. High-quality information disclosure reduces financing costs, alleviates confidence crises, ensures capital adequacy, and avoids capital outflows. Research constructs a corporate information disclosure risk contagion model, using an improved LDA model and multi-objective evolutionary optimization methods for analysis, showing high efficiency and good accuracy, effectively controlling environmental and related effects.**

***Keywords*—Management tone manipulation; enterprise information disclosure; risk contagion; optimal control**

I. Introduction

A. Research Background

The multi-objective optimization algorithm is designed to address the challenges of optimizing multiple objective functions simultaneously and identifying a set of solutions that are optimal in terms of these objectives [1]. In this context, enhancing the performance of one target may result in a decline in the performance of one or more other targets. Particle swarm optimization (PSO) algorithm is an optimization technique inspired by swarm intelligence, specifically bird flock predation behavior [2]. It simulates individual collaboration and information sharing mechanisms observed in bird flocks to seek the optimal solution for a given problem [3]. Each solution within the particle swarm optimization algorithm represents a "particle" within the search space, possessing unique velocity and position properties [4]. By continuously updating their speed and position, particles navigate through the search space to find an optimal solution. SMOTE (Synthetic Minority Over-sampling Technique) algorithm is an oversampling technique utilized to address data category imbalance issues [5]. Its fundamental concept involves analyzing limited samples and augmenting new samples into the dataset [6] [7].

The enterprise is different from the market resource allocation mechanism [1], the essence of the enterprise is the contract collection of stakeholders. The importance of corporate information disclosure is self-evident, and it is an important basis [8] for listed companies to sign contracts and execute contracts with stakeholders in the market. To apply enterprise risk management to address the company's vulnerability to cyber risks, thereby achieving control over risks for the company [9].Corporate information disclosure can enable external investors to assess the real or potential value [10] [11] [12] Companies that have experienced data breaches have become aware of their vulnerabilities and have implemented oversight at the board level, leading them to have the audit committee increase supervision [13]. The main form of modern accounting information disclosure is financial report, which is a useful mechanism [14]. The disclosure of enterprise information has gone through three stages: voluntary disclosure, mandatory disclosure, and the combination of voluntary disclosure and mandatory disclosure.

Risk information disclosure encompasses both informational and risk attributes. From an informational perspective, risk information and financial digital information are interconnected and mutually reinforcing, playing a crucial role in enhancing the comprehensiveness of financial information. Following the disruption of "rigid exchange" in the bond market, changes in the risk environment can lead to a decline in investor sentiment, further reducing market liquidity and exacerbating internal and external information asymmetry. During such times, bond issuing enterprises may provide more detailed risk information to mitigate this asymmetry. From a risk standpoint, disclosing risk information unveils heterogeneous risks within an enterprise while conveying

*Corresponding Author.

negative news that can have adverse effects on the company. Simultaneously, risk information possesses characteristics such as high discretion, low verifiability, and low violation costs; thus making it susceptible to being utilized as a tool for impression management by enterprises. When disruptions like "gang" breakouts induce market aversion towards risks, companies may engage in impression management by reducing their disclosure of risk-related details to diminish participants' perception of heterogeneity risks associated with the firm.

B. Literature Review

Past scholars studied the impact [12] of disclosure motivation, corporate governance structure and corporate scale on corporate information disclosure. They believe that because of corporate strategy, corporate governance and other reasons, they have the motivation to actively disclose a certain degree of corporate information. Some scholars believe that the level of voluntary disclosure is significantly related [16] to the size of the company, and another scholar finds that the degree of ownership concentration is negatively related [17-19] to the degree of voluntary disclosure. The academic circle has not yet formed a unified view on the evaluation of the quality of accounting information disclosure of enterprises. Traditional studies on the quality of accounting information mainly focus on three aspects [20-22]: the characteristics of information quality, the measurement of accounting information quality and economic consequences, and pay more attention to the relationship between accounting information quality and earnings management, financing cost and performance management. It mainly focuses on the theoretical development, influencing factors and statistical evaluation of the quality of accounting information [23].

Some scholars believe that earnings management will distort stakeholders' interpretation of corporate performance, causing them to misunderstand [11]. The reliability of accounting information can be achieved through three indicators: the degree [24] of earnings management, the measurement error of earnings and the audit opinion. The relevance of accounting information can be realized through the value relevance [11] of accounting information and the continuity [11] of earnings. Comparability is mainly measured by the difference between expected earnings and actual earnings. Timeliness is measured [11] by the time lag in earnings releases. As for the study on the influencing factors of management, it is proposed that when the equity is highly concentrated in the management level, the voluntary disclosure will decrease. Some scholars point out that the information communication process of financial reports will be hindered under the following three conditions: (1) Managers have more information about the company's business strategy and operation than investors; (2) the pursuit of managers is not completely consistent with the interests of all shareholders; And (3) imperfect accounting standards and auditing. Senior executives can influence or even decide accounting behaviors and policies such as corporate information disclosure and earnings management, thus affecting the quality [28] of accounting information.

In the existing literature, the research on the risk contagion model of enterprise information disclosure has made some progress, especially in the application of complex network analysis and optimization algorithm. However, there is still relatively little research on managing how tone manipulation affects the risk contagion of corporate disclosure [15], and how this risk can be more effectively controlled through improved models and algorithms. This study aims to fill this gap by combining the perspective of management tone manipulation, using the improved LDA model and multi-objective evolutionary optimization technology, to construct a more accurate enterprise information disclosure risk contagion model. Firstly, it utilizes text data from Chinese listed companies, including MDA texts, annual reports of performance presentations, and other relevant materials, to investigate the risk associated with corporate information disclosure [29]. Secondly, this study explores how management tone influences corporate information disclosure and risk issues. Previous studies primarily focused on disclosure motivation, governance structure, financing cost, etc., whereas this paper expands the research perspective by analyzing management tone data. Thirdly, an information risk contagion model between enterprises is constructed from a dynamic correlation perspective while incorporating accounting information based on emotions to establish a dynamic model. This effectively controls for environmental effects and correlation effects. In addition, this study also discusses the effectiveness of the weighted Smote algorithm in dealing with unbalanced data sets, which provides a new direction for future research. Lastly, particle swarm optimization algorithm is employed for data verification and algorithm comparison to study optimal control methods. The purpose of this article is to enrich the influence of management tone on enterprise information disclosure by improving the accuracy and stability of the risk infection model of enterprise information disclosure.

This article utilizes text data from listed companies, explores innovative ways to enrich the impact through examining the influence of managerial tone, constructing models, and employing algorithms for validation and comparison. It also studies and constructs a risk contagion model for corporate information disclosure, applying an improved model and optimization methods for analysis and solution. The improvements to the algorithms are empirically proven to enhance the accuracy and stability of the models, providing strong support for corporate information disclosure risk control. Additionally, it discusses the connection between accounting information disclosure and corporate agency, information asymmetry, and the benefits of high-quality disclosure to enterprises.

II. Model Construction

A. Preparation of Knowledge

In the realm of real-world enterprises, organizations establish networks via social connections generated by interactive behaviors. This research develops a risk contagion model for corporate disclosure based on the social network framework. In recent times, given the rapid progression of information, the scale of social networks has been expanding exponentially. Usually, the inter-enterprise network is abstracted as a figure $G = (V, E, W)$, where $V = \{v_1, v_2, \ldots, v_n\}$ is the set of all nodes in the graph, that is, the collection of all enterprises in the network, and n represents

the number of enterprise nodes in the network; The edge set formed by the interaction relationship between enterprise nodes in the graph. $E = \{e_1, e_2, \dots, e_m\}$ is the set formed by the interaction relationship between enterprises in the network, and m is the number of the network edge between enterprises; $W = \{w_1, w_2, \dots, w_m\}$ represents the set of weights for all inter-enterprise edges in the graph. Among them, the inter-firm network has the characteristics of small world, scale-free and power-law distribution, etc. In this study, the network structure is assumed to compound the above characteristics. In inter-firm networks, the distance between firms is usually relatively short. And in complex networks, the nodes k with degrees of are power-law distribution$p(k) = ck^{-\gamma}$. Numerous studies have discovered that within the inter-enterprise relationship network, the connection pattern of nodes adheres to the scale-free attribute, with most businesses possessing only a limited number of connections, whereas a select few users maintain a substantial number of connections. The notion of clustering coefficient, existing within the network, is employed to assess the intimacy between associates. In a scale-free network, nodes exhibiting larger degree values generally possess lower clustering coefficients, conversely, nodes with smaller degree values demonstrate higher clustering coefficients. Therefore, the clustering coefficient within the enterprise network will exhibit a power-law distribution pattern. In the inter-firm information disclosure risk contagion network, the influence of a company varies based on its position and other variables. As the network evolves, nodes will continuously accumulate and exert their risk influence on other entities, leading to alterations in the risks faced by surrounding individuals.

The influence maximization problem requires the calculation of information propagation in a network, considering the risk associated with information disclosure. In this study, we employed the Linear Threshold model (LT), which is a widely used model for simulating influence propagation and risk interaction among users in an inter-enterprise information disclosure network [30]. Each node in the network was defined to have two states during information transmission: active and inactive. At any given moment, each node can only exist in one state, with active nodes having the ability to activate their inactive neighboring nodes while inactive nodes cannot activate other neighbors. The probability of successful activation increases as a node has more active neighbors. When an active node attempts to activate its inactive neighbor, and if successful, that neighbor will continue attempting to activate its own inactive neighbors until no further activations occur within the network, marking the end of the propagation process. All transitions between states are limited to going from inactive to active or vice versa. A linear threshold model is a simple and efficient graph model that assumes that the relationships between nodes are boolean-valued, i.e. either exist or do not exist. In this case, the linear threshold model can predict the state of the node by performing threshold processing on the relationship between the nodes. In contrast, recently popular models related to graph networks, such as convolutional neural networks (CNNS) and recurrent neural networks (RNNS), are more complex and flexible. These models can deal with continuous relationships between nodes and can learn complex interaction patterns between nodes. However, these models are also more complex and computationally intensive, requiring more computational resources and data. Therefore, when building graph models based on social networks, the linear threshold model is chosen because it is simple, efficient, and can be trained and predicted on small data sets. In addition, linear threshold models can also be used in combination with other models to improve prediction accuracy and generalization ability.

The independent cascade model is a widely utilized framework for modeling influence propagation. Once a node v has been successfully influenced, it is deemed capable of activating inactive nodes w in its neighboring network and will attempt to do so. The probability of being affected by success is P_{vw}, and this attempt is made only once, and these attempts are independent of each other, that is, v activation of w is not affected by other nodes. The specific propagation process is as follows: First, given the initial set of active nodes S, when the node v is activated at the moment t, it has a chance to influence its neighbor nodes w that are not activated. The probability of successful activation is P_{vw}, this value is a random system parameter, which has nothing to do with nodes in the network. The greater the value, the more likely the node w is to be affected. If multiple nodes are newly activated and these nodes are neighbors of the node w, then these nodes will try to activate the node w in any order. If the node w is successfully activated by the node v, the node w will become active at the moment $t + 1$. At moment $t + 1$, the node w will have an influence on other neighbors that are not activated. The process is repeated until there are no active nodes with influence in the network. The propagation process ends.

Multi-objective optimization function is realized based on the multi-objective optimization algorithm. These algorithms trade off and compromise between multiple objective functions to find a set of optimal solutions that are not inferior to any other on all objective functions, i. e., there is no solution that is better on all objective functions. It can be summarized as the following steps: randomly generate a set of initial solutions as the initial population. The initial population was evaluated to calculate the value of each solution on the individual objective function. According to a certain selection mechanism, a part of the solution is selected from the current population as the parent of the next generation population. The selected parent solutions are subjected to evolutionary operations such as crossover and variation to generate the next generation population. Repeat the above steps until the termination conditions are met. Multi-objective optimization can be applied to the recommendation system, logistics distribution, product design and so on.

B. Management Intonation Embedding

This study uses natural language processing to process text data embedded by management. For the probability of a text sequence $S = w_1, w_2, \dots, w_T$, it can be expressed as shown in Eq. (1). That is, the researcher can convert the text probability into the product of the conditional probabilities of the antecedent.

$$P(S) = P(w_1, w_2, \dots, w_T) = \prod_{t=1}^{T} p\,(w_t \mid w_1, w_2, \dots, w_{t-1}) \quad (1)$$

The CBOW model serves as the initial implementation of the Word2Vec algorithm, encompassing both the Embedding matrix from the bag of words model and continuous Embedding

vectors [31]. By predicting target words based on contextual information, the CBOW model enables analysis of word vector expressions. In fact, it can also acquire word vectors by sequentially predicting text content associated with target words. This variant is referred to as the Skip-gram model, as depicted in Eq. (2).

$$p(w_o \mid w_i) = \frac{e^{U_o \cdot V_i}}{\sum j e^{U_j \cdot V_i}} \quad (2)$$

Among them, V_i is the column vector in the embedding layer matrix, also known as the input vector of w_i. U_j is the row vector in the Softmax layer matrix, and it is also the output vector of w_i. The essence of Skip-gram model is to calculate the cosine similarity between the input vector of the input word and the output vector of the target word, and carry out Softmax normalization. Skip-gram model is a natural language processing model. In the disclosure risk task, input words and target words can represent different concepts. Input words can represent the information disclosed by the enterprise, such as financial reports [26] [27], announcements of major events, etc. Target words can represent concepts related to disclosure risks, such as risk factors, uncertainties, laws and regulations. By using the Skip-gram model, companies can predict the target words associated with disclosure risks and take timely steps to mitigate them. Complex normalized probability is a way to describe the probability distribution of a quantum state. In quantum mechanics, a quantum state can be represented by a wave function, and the modular square of the wave function gives the probability of getting a certain result when measuring a physical quantity in that quantum state. The complex normalized probability is introduced to facilitate the description of the properties of quantum states, because it normalizes the probability distribution to 1, so that the probability distribution between different quantum states can be directly compared. In this study, the complex normalized probability is decomposed into a series of conditional probability products, where v is expressed as the possible values of the physical quantity, bi(v) is expressed as the wave function of the quantum state, and p(v) is expressed as the probability of v obtained by measuring the physical quantity in this quantum state as shown in Eq. (3).

$$p(v) = \prod_{i=1}^{m} p(b_i(v) \mid b_1(v), \ldots, b_{i-1}(v), \text{context}) \quad (3)$$

In addition, in order to understand this process more intuitively, a classification binary tree can be constructed. First, the original dictionary D is divided into two subsets D_1, D_2, while assuming that given Context, the probability that the target word belongs to the subset D_1 is described in Eq. (4), U_{Droot} and Vw_t both sums in the formula are parameters of the model.

$$p(w_t \in D_1 \mid context) = \frac{1}{1+e^{-U_{D_{\text{root}}} * v_{w_t}}} \quad (4)$$

Second, the researcher can further divide the subsets $D_1 D_2$, and after going through this process, the original dictionary D of size V is converted into a binary tree of depth $\log V$. There is only one path from the root to any leaf node, and along this path, the corresponding leaf node's class is encoded. From the root to the leaf, it's actually a random walk. According to this principle, researchers can calculate the likelihood probability of leaf nodes based on the binary tree. For example, assuming a target word w_t in the training sample, the corresponding binary tree code is $\{1,0,1,\ldots,1\}$, then the likelihood function can be constructed as shown in Eq. (5).

$$p(w_t \mid \text{context}) = p(D_1 = 1 \mid \text{context}) p(D_2 = 0 \mid D_1 = 1) \ldots p(w_t \mid D_k = 1) \quad (5)$$

The term product in the given formula represents a logistic regression function, and solving for the maximum likelihood function provides us with the probability of computing a binary tree parameter (a vector on a non-leaf node) to a specific child node. By constructing a binary tree, we reduce the complexity of the target probability calculation, i. e., from the original V to the $\log V$. The original likelihood function of Skip-gram model corresponds to the distribution of multinomial. When solving this likelihood function using the maximum likelihood method, the loss function of cross-entropy is obtained as shown in Eq. (6).

$$J(\theta) = -\sum_{t=1}^{T} \frac{\log p(w_{t+j} \mid w_t)}{T} \quad (6)$$

Where, $p(w_{t+j} \mid w_t)$ is the probability after normalization over the entire dictionary,and define the logistic regression function as shown in Eq. (7).

$$p(w, \text{context}) = \sigma\left(\log p(w \mid \text{context}) - \log k p_n(w)\right) \quad (7)$$

Where, the target word w represents the text-driven probability; k is a prior parameter, which represents the sampling frequency of noise; $p(w \mid \text{context})$ represents the non-normalized probability distribution, using the molecular part of the Softmax normalized function; $p_n(w)$ word distribution representing background noise, often using the word Unigram distribution. New data sets can be obtained after the k sampling of the noise distribution. In this study, the dependence of the NCE likelihood function on the noise distribution is removed, and the logistic regression function is defined directly with the molecules in the original softmax function [32]. The corresponding objective function of the model can be defined as shown in Eq. (8).

$$J(\theta) = \log \sigma(U_o \cdot V_i) + \sum_{j=1}^{\kappa} E_{w_j \sim p_n(w)} \left[\log \sigma\left(-U_j \cdot V_i\right)\right] \quad (8)$$

Therefore, the aim is to capture statistical laws and patterns in language data, and then be used for a variety of natural processing tasks. Improve performance by providing principled and statistics-based approaches to language patterns and regularities. Based on the aforementioned model, an analysis is conducted on the MD&A text, performance presentation, annual report, and social responsibility report. Management intonation manipulation refers to the strategic adjustment of language, wording, and emotional tone in information disclosure by management to selectively convey the company's performance, strategy, prospects, and other content with the aim of influencing the company's image, investors' decisions, and market sentiment. Building upon existing research findings, abnormal optimistic tone exhibited by management is utilized as a measure for assessing the extent of management intonation manipulation. Initially utilizing LM's emotional vocabulary [33] as a basis for analysis yielded the management intonation result Tone = (positive words - negative words) / (positive words + negative words). The affective analysis within management

intonation can be estimated through evaluating the total number of positive emotional messages associated with time-specific information. In this study, separate constructs were developed for emotionality assessment including polarity and negative word identification. Furthermore, the improved LDA algorithm outlined in Algorithm 1 was employed to adjust social media emotion data pertaining to aforementioned listed companies. The sentiment thesaurus is based on the sentiment thesaurus published by CNKI.com, Detailed Dictionary of Commonly used commendatory and derogatory Words, Student Commendatory and Derogatory Words, Commendatory Word Dictionary and derogatory Word Dictionary. It excludes emotion words with low frequency while incorporating network and oral emotion words, resulting in a total of 4637 positive words and 5139 negative words. The constructed emotional thesaurus is categorized into five levels (KW1-KW5) based on word usage frequency, ranging from the simplest version to the complete version. Additionally, this study identifies polarity lexemes within the emotional lexicon that exhibit strong polarity for certain emotion words, particularly derogatory ones. When identifying opinion sentences containing these aforementioned words, their polarity determines that of the entire sentence (with opposite polarity for negative sentence patterns). To differentiate them from other lexemes within emotions lexicon, these specific lexemes are referred to as polar lexemes which consist of 158 positive words and 281 negative words.

Algorithm 1: Improve the LDA algorithm

Input: MD&A, performance presentation, annual report, social responsibility report and other corpus documents
Output: document theme emotion score
01: **for on topic** and emotionzm
02: Extract multinomial distribution$\emptyset_{z,m}$
03: **for** document, **emotion**dj
04: Extract multinomial distributionφ_d, θ_{dj}
05: **for** each sentence in the sentence, $s$words$w_{d,n}$
06: Extract binomial distribution m_s, π_{n-1}
07: **if** $x_n = 0$
08: Extract multiple variable topics $Z_n$and wordsw_n
09: **else if** $x_n = 1$
10: Extract the topic $\delta$and words from the LDAZ_{n-1} distribution of the parameterw_{n-1}
11: **end if**
12**:** **end for**
13: **end for**
14: **end for**
15: return: document topic emotion score

The LDA (Linear Discriminant Analysis) here is a classical linear classification method that is also widely used in dimensionality reduction tasks. The main idea of LDA is to project high-dimensional data into a low-dimensional space while maintaining category information so that the sample projection points between the same species are as close as possible. The Xgboost algorithm is employed in this study to enhance the performance of the LDA algorithm. The core concept involves constructing CART regression trees by continuously splitting continuous features [34]. Each generated CART tree represents a new function that fits the negative gradient (approximate residual) of the previous predicted outcome. Upon completion of training, n CART tree is obtained. To predict the score of a new sample, it only requires traversing each tree based on the characteristics of this sample and obtaining its final position in each tree's leaf node along with its corresponding score. Finally, the scores from each tree are aggregated to obtain the final predicted score for the sample, The LDA algorithm is improved to improve the classification accuracy, and the improved LDA algorithm optimizes the classification decision boundary to make the similar samples closer and the different class samples more dispersed, thus improving the classification accuracy. Improve the computational efficiency, optimize the calculation process of the algorithm, reduce the computational complexity and improve the calculation efficiency. In addition, the robustness is enhanced by introducing data preprocessing steps to reduce the impact of data noise and outliers on the algorithm, and enhance the robustness of the algorithm.as expressed in Eq. (9).

$$\hat{y}_i = \hat{y}_i^{(N)} = \sum_{n=1}^{N} f_n(x_i), f_n \in \mathrm{F} \tag{9}$$

Where, N represents the number of CART decision trees,$f_n(x_i)$ represents the n th tree, the final prediction result of the model is equal to fitting a new tree $f_n(x_i)$ based on the prediction value of $n-1$, $\hat{y}_i^{(n-1)}$, so as to reduce the target function as much as possible. The space of CART tree can be represented as shown in Eq. (10).

$$\mathrm{F} = \{f(x) = \omega_{q(x)}\}, q: R^m \to \text{or} \to T, \omega \in R^T \tag{10}$$

Among them, each decision tree consists of two parts: structure q and leaf node weight ω, q represents the structure of each decision tree, and the sample is divided into corresponding leaf nodes through traversal features, T represents the number of leaf nodes; ω represents the set of leaf node scores of each tree, so the predicted value of the model is equal to the sum of the scores of the corresponding leaf nodes of the sample in each tree, and its objective function is shown in Eq. (11).

$$L(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k) \tag{11}$$

In addition, in order to prevent overfitting, In the process of establishment, if the target function is less than γ, the feature is not selected for division, the more serious the model, the greater the value of γ. Therefore, the objective function can be defined as shown in Eq. (12).

$$L^m = \sum_{j=1}^{T} \left[\sum_{i \in I_j} g_i \omega_j + \frac{1}{2} \left(\sum_{i \in \{i | q(x_i) = j\}} h_i + \lambda \right) \omega_j^2 \right] + \gamma T + \text{constant} \tag{12}$$

Where, I_j represents the sample set belonging to the first leaf node j in the training process. Take the derivative of the objective function to get the parameter expression of the smallest objective function as shown in Eq. (13).

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \tag{13}$$

Due to the vast number of features in the training sample, there may exist multiple structural patterns in constructing a decision tree. Enumerating all possible decision tree structures to calculate weight is obviously undesirable. Therefore, the

Xgboost model employs a greedy algorithm that iteratively splits nodes from the starting tree, generating two nodes for each split. The sample data will be divided into the left subtree and the right subtree respectively according to the node rules, and set as I_l and I_r. Then the loss reduction after the node splitting can be defined as Eq. (14).

$$L_s = \frac{1}{2}\left[\frac{G_l^2}{H_l+\lambda} + \frac{G_r^2}{H_r+\lambda} - \frac{(G_l+G_r)^2}{H_l+H_r+\lambda}\right] - \gamma \quad (14)$$

This study uses the probability distribution of $C = 0$ on the learned t trading day after q first represents all the information of the enterprise on the time series. The Beta distribution is used to describe the probability q, which is defined as a continuous distribution over an interval $[0,1]$.Its two shape parameters, α and β are used as exponents of the random variables, controlling the shape of the distribution. The Beta distribution is characterized by two shape parameters and is used to model phenomena with constraints between 0 and 1, and its probability density function can be defined as shown in Eq. (15).

$$P(q;\alpha,\beta) = \frac{1}{\int_0^1 q^{\alpha-1}(1-q)^{\beta-1}dq} q^{\alpha-1}(1-q)^{\beta-1} \quad (15)$$

In this study, either $C = 0$ or $C = 1$ was considered as a single Bernoulli experiment. In this study, N the likelihood function in the Bernoulli experiment is represented by Eq. (16).

$$L(v, N-v \mid q) = \binom{N}{v} q^{N-v}(1-q)^{v} \quad (16)$$

Combining the above methods can help enterprises better understand and manage the risk of information disclosure. For example, enterprises can use natural language processing technology to identify potential risk factors in information disclosure text, use CBOW and Skip-gram models to predict possible risk factors, and finally use LDA model to identify topics and topics in information disclosure text, so that enterprises can better understand the content of disclosure text and identify possible risk factors.

III. Optimal Control Algorithm

A. Multi-objective Optimization

Multi-objective optimization is a crucial domain within the realm of multi-criteria decision making, encompassing mathematical problems that aim to optimize multiple objective functions simultaneously. This necessitates making optimal decisions in situations where tradeoffs between two or more conflicting objectives arise. The multi-objective optimization problem refers to the simultaneous optimization of multiple objectives, often leading to conflicts among them. In other words, in multi-objective optimization, enhancing the performance of one objective may result in a decline in the performance of one or more other objectives. Many real-life problems can be transformed into multi-objective optimization problems. For maximization problems, multi-objective optimization can be represented as shown in Eq. (16).

$$\text{maximize}[F(x)] = \left(f_1(x), f_2(x), \dots, f_{no}(x)\right) \text{ s.t. } x = (x_1, x_2, \dots, x_k) \in \Omega \quad (16)$$

Where, Ω is the decision space, no is the number of objective functions, and $x = (x_1, x_2, \dots, x_k)$ is the candidate solution with one variable. As for the evaluation index of multi-objective optimization problem, it mainly analyzes the convergence and diversity of the solution set. The two widely used performance measures are the rewind distance, as shown in Eq. (17).

$$IGD = \frac{\sum_{i=1}^{n} d_i}{n} \quad (17)$$

Where, n denotes the number of solutions in the real Pareto front, d_i is the Euclidean distance between the i th solution uniformly sampled from the real Pareto front and the solution generated by the algorithm, where the smaller the IGD value, the better the convergence and diversity of the algorithm. In addition, the hypothesis $z^r = (z_1^r, z_2^r, \dots, z_m^r)$ is a reference point in the target space, which is dominated by all Pareto optimal solutions. S is a solution set generated by the algorithm, HV represents the size of the target space dominated by the solution in S and the boundary z^r, m is the target number, which can be expressed in detail as shown in Eq. (18).

$$HV(S) = VOL \bigcup_{x \in S}[f_1(x), z_1^r] * [f_m(x), z_m^r] \quad (18)$$

Where, the function $\mathrm{VOL}(*)$ represents the Lebesgue measure, and the larger the value HV, the higher the quality of the solution set S. The study also refers to the multi-objective evolutionary optimization method, which initially employs Pareto non-dominated sorting to categorize the solutions in a combined population into distinct frontiers. Subsequently, it selects solutions with higher frontier ordering until reaching the desired population size. In case additional solutions are required within the same frontier, the crowding distance operator is utilized to arrange and select some solutions with superior ranking in descending order. Decomposition-based algorithms partition a multi-objective problem into numerous single-objective subproblems and concurrently optimize them. Among this category of algorithms, the decomposition-based multi-objective evolutionary algorithm stands out as widely adopted.

B. Improve the Particle Swarm Optimization Algorithm

The particle swarm optimization algorithm is designed based on the principle of avian foraging behavior. During the process of food search, birds are initially fed randomly. If one bird discovers food along its path, other birds adjust their speed and position according to this discovery. Hence, the algorithm is named as particle swarm optimization algorithm. First, the particle population N is randomly initialized within the search space, where the dimension of each particle is D.Then, the maximum number of iterations and initial speeds $v_{i,d}^0$ are set for each individual particle. Secondly, after defining the adaptive function, the best value found by this particle is the local optimal solution *(*pbest*)*, and the best value selected by all particles in the optimal solution is the global optimal solution *(*gbest*)*. Finally, in the iteration process, the iteration to the kth i particle is $x_i = \left(x_{i,1}, x_{i,2}, \dots, x_{i,D}\right)$, through the historical global optimal solution and the local optimal solution of the particle at this time, according to the velocity update Eq. (19) to find the particle's location.

$$x_{i,d}^{k+1} = x_{i,d}^{k} + wv_{i,d}^{k} + c_1 r_1\left(\text{pbest}_{i,d}^{k} - x_{i,d}^{k}\right) + c_2 r_2\left(\text{gbest}_{d}^{k} - x_{i,d}^{k}\right) \quad (19)$$

Where, $x_{i,d}^{k}$ is the position of the i th particle with the d-dimensional vector in the k th iteration; $v_{i,d}^{k}$ is the velocity of the ith particle with the d-dimensional vector in the k th iteration. w is the initial weight value, and the acceleration constant c_1, c_2 to control the relationship between *pbest* and *gbest*, and r_1, r_2 is the random number between 0 and 1. However, in the actual problem, the problem that needs to be solved is not a single goal, but a number of goals need to be considered. Like this, the problem that requires to solve multiple goals but cannot make multiple goals achieve the best at the same time is called the multi-goal problem. Therefore, the convolutional neural network is combined to improve the particle swarm optimization algorithm.

The core components of convolutional neural networks primarily consist of diverse neural network layers. Generally, these networks encompass an Input Layer, Convolutional Layer, Pooling Layer, Activation Layer, Dropout Layer, Fully Connected Layer and more. By arranging and combining these layers based on their respective functions, convolutional neural networks can form distinct models with varying capabilities. Subsequently, the subsequent sections will elaborate on the individual functionalities of each layer.

The input layer functions to feed the preprocessed data into the convolutional neural network, thus typically positioned at the inception of the network. The primary purpose of the convolutional layer is to extract data features by means of convolving the input data within this layer. In general, the number of convolutional cores is equal to the number of channels, and a set of filters is composed of multiple convolutional cores, and the number of filters is equal to the number of output channels of the convolutional layer. Therefore, if convolution is thought of as a function, the first thing to consider is the size of the filter. The features x_i are obtained in the upper layer, weighted by the convolutional kernel i, j, and then solved. As shown in Eq. (20).

$$x_j^l = f\left(\sum_{i=M_i} x_i^{l-1} * \text{kernel}_{ij}^l + B^l\right) \quad (20)$$

After the convolution, the inclusion of complex parameters may lead to overfitting or extensive training time consumption during the iterative process. Therefore, it is necessary to pass through a pooling layer for down sampling after the convolutional layer. This approach reduces neural network parameters and facilitates further feature extraction to mitigate overfitting. Following the convolutional layer, all operations remain linear, which limits model training effectiveness and hinders continuous learning adaptation to real-world scenarios. Moreover, in reality, most classifications are non-linear; hence we require a mechanism for nonlinear transformations that can enhance semantic information extraction capabilities. In this study, we employ the Sigmoid activation function as depicted in Eq. (21).

$$\sigma(x) = \frac{1}{(1+e^{-x})} \quad (21)$$

The application of particle swarm optimization algorithm to the improvement of neural network is mainly reflected in parameter optimization, network structure optimization and training process optimization. In terms of parameter optimization, PSO algorithm can be used to optimize the parameters of neural network and improve the performance of the network by searching for the optimal parameter combination. In terms of network structure optimization, the PSO algorithm can be used to optimize the structure of neural networks and improve the performance of the network by searching for the optimal network structure. In terms of the training process optimization, the PSO algorithm can be used to optimize the training process of the neural network to improve the performance of the network by improving the training speed and avoiding falling into the local optimal solution.

C. Improvement of Unbalanced Data Processing

In the scenario of information risk control, the occurrence of defaulting users is typically relatively low, resulting in an imbalanced data set issue. Taking binary classification samples as an example, if the majority of training samples are positive and only a small portion are negative, without any balancing treatment during training, the model may yield positive predictions for all samples. Although achieving a 90% accuracy rate, this renders the model ineffective as it fails to identify any negative instances. To address this class imbalance problem in our study, we primarily employ a weighted Smote algorithm that combines up-sampling and down-sampling techniques.

Smote algorithm is one of the methods of up-sampling. It uses K-NN algorithm and linear interpolation method to artificially generate new minority class samples randomly. Specifically for the minority class sample set S in each sample x_i, calculate the Euclidean distance to all other minority samples, This get the k minority samples closest to x_i (KNN algorithm). And randomly select one sample x_j in k samples around x_i, link x_i and x_j, and randomly generate a new sample in the middle of the two by linear interpolation method, as shown in Eq. (22).

$$x_{ij} = x_i + \text{rand}(0,1)\left(x_j - x_i\right) \quad (22)$$

According to the proportion of positive and negative samples in the original data set, an upsampling ratio column ais manually set, and the above step is repeated a times for each minority class sample. While Smote effectively addresses the issue of sample overfitting, it also introduces a challenge regarding sample quality [25], specifically the likelihood of generating artificial noise within the majority class. Therefore, this study employs a weighted Smote algorithm primarily aimed at denoising certain class samples initially and subsequently determining their positions based on Euclidean distance. Different positions correspond to different weights, with higher weights resulting in more newly generated samples. Using the SMOTE algorithm improves the classification performance, and he can increase the number of minority class samples, allowing the classifier to better learn the characteristics of the minority classes during the training process. This helps to reduce the bias of the classifier against majority classes and improve the ability of the classifier to identify minority classes, thus improving the classification performance. In addition, the SMOTE algorithm can synthesize

new minority class samples, rather than simply copy the existing samples, which helps to alleviate the overfitting problem.

For each sample x_i in the minority class sample set S, calculate the Euclidean distance from all other samples (including the majority class samples) to obtain the k samples closest to the x_i. For sample x_i, if the nearest k samples belong to the majority class samples, it is judged to be noisy data. Calculate each minority samples x_i to all other minority samples the sum of D_j, A smaller D_j value indicates that the closer the sample point is to the center of a few class of samples, otherwise the closer to the boundary of positive and negative samples, the two types of points contains the few class more information, more representative, so need to give more weight. The D_j of all few samples is calculated, the mean $\widehat{D}$ is obtained, and the center point and boundary point are selected according to the absolute value difference between D_j and $\widehat{D}$. Finally, the weight ω_j is determined according to the proportion of d_j in $\sum_j d_j$. For the remaining minority samples, the algorithm KNN is performed again (including only a minority sample), and the algorithm steps are repeated. The only difference is that $a * n * \omega$ new samples are generated in each minority sample based on the weight, which is defined in this study as shown in Algorithm 2. In this study, the above methods are combined to solve the multi-objective optimization problem. The correlation between graph model building and managing integration embeddings is very tight. A graph model is a data structure used to represent and process complex relationships, while management integration embedding is a way to integrate different data sources into a unified data model. In a graph model, nodes represent entities and edges represent relationships between entities. Through the use of graph models, enterprises can better understand and manage their business processes, customer relationships, supply chains and other complex relationships. Management integration embeddings can help enterprises integrate different data sources into a unified data model to better manage and analyze data. In practical applications, graph models and management integration embeddings are often used together. For example, an enterprise can use a graph model to represent its business processes and then use management integration embeddings to integrate different data sources into this graph model. In this way, enterprises can better understand and manage their business processes and better analyze and utilize data. The proposed method in this study is defined as Optimal Control of Enterprise Information Disclosure Risk from the Perspective of management tone manipulation Management Tone Manipulation, ECTR-MTM).

Algorithm 2: Improve the weighted Smote algorithm

Enter:S, x_{ij}
Output:ω_j
01: **for** samples in a small Ssample **set**x_i
02: Calculate the Euclidean distance from the sample to all other samples (including most samples)x_i
03: Determine the nearest sample$x_i k$
04: **for** samplex_i
05: **if** the nearest sample belongs to the $k$majority sample
06: Then it is judged as noise data
07: Calculate the sum of Euclidean distances from each minority class sample to all other minority class samples$x_j D_j$
08: Calculate the mean of all minority samples$D_j \widehat{D} = \frac{\sum_j^n D_j}{n}$
09: Select the center point and boundary point according to the absolute value of the difference $D_j \widehat{D}$as shown in Eq. (22)
10: According to the proportion to determine the weight$d_j \sum_j d_j \omega_j = \frac{d_j}{\sum_j^n d_j}$
11: **end if**
12: **return** ω_j

$$d_j = \left| D_j - \frac{\sum_j^n D_j}{n} \right| \tag{23}$$

IV. Numerical Example

A. Experimental Design

The experiment in this study is conducted using the NS3 simulation platform (Network Simulator 3), which utilizes the API interface to simulate a real network environment and establish a topology simulation. The BBR module within the NS3 simulation platform is developed by Google. Since its implementation, researchers have made improvements on this framework. In this paper, the enhanced algorithm is implemented based on the BBR module framework. The experiment described in this paper is based on the BBR module released by Google on the NS3 simulation platform, which already incorporates the congestion control algorithm of standard BBR. The algorithm proposed in this paper builds upon and improves upon this existing framework [29]. Typically, steps involved in simulating with the NS3 platform are as follows: firstly, determining and constructing an appropriate network topology according to specific requirements; then setting relevant parameters for modules within that network topology, such as channel bandwidth and packet loss rate; finally, configuring additional properties for conducting simulations while collecting and analyzing data. For implementing the algorithms proposed in this paper as well as related comparison algorithms, Tensorflow 1.5.1 was utilized by our research team. The data samples used in this study consist of text data from MD&A reports, performance presentations, annual reports, social responsibility reports, Directors' CVs and supervisors' CVs of listed companies from both China and United States. The specific features of the data set are presented in Table I. The data set mainly consists of four real corporate information disclosure network datasets, including: corporate network of listed companies established in China, chain director network of listed companies in China, corporate network established by listed companies in the United States, and chain director network of listed companies in the United States. These data sets contain the enterprise's social network information, such as the number of nodes, the number of node boundaries, the average degree, the average path of nodes, and the clustering coefficient. Each dataset has its own unique characteristics, such as the number of nodes and the number of edges, which reflect the complex network of relationships between different enterprises. When working with these datasets, we found that there was a class imbalance, where

some classes had far more samples than others, which could affect the model's ability to generalize. To solve this problem, we use the weighted Smote algorithm to balance the dataset by generating new minority class samples, thereby improving the performance and generalization ability of the model. In this way, we ensure that the model can more accurately capture the propagation characteristics of enterprise information disclosure risk. The experiment was conducted in groups, where the data were divided into 10 groups and cross-validation method was employed [35]. This involved splitting the data set into 10 equal parts, with one part being selected as the test set each time while the remaining parts served as training sets. Finally, an average value was obtained. All data were stored in a MySQL database in CSV format for further processing. For the network data set, we utilized Rapidminer, a data mining tool, to randomly select 10% of each user's rating data as the test set and used the remaining 90% of user data as training sets. To address performance analysis and scheduling issues, we propose a cost minimization algorithm framework by examining problem parameters and comparing different algorithm components. All algorithms were implemented and tested using Matlab R2018a on an i5-3470 CPU@3.20GHz processor with 8GB RAM.

TABLE I. ENTERPRISE DISCLOSURE RISK DATA SET

Network serial number	Social network name	Type	Number of nodes	Number of node boundaries	Average degree	Average path of nodes	Clustering coefficient
1	Enterprise network of listed companies in China	Directed	542522	24139204	34.350	2.245	0.154
2	Chain director network of listed companies in China	Directed	6331023	455238591	41.245	2.519	0.148
3	Corporate Network of American public Companies	Directed	38120	585291	10.501	2.490	0.130
4	Network of interlocking directors of U.S. public companies	Directed	452112	3522222	14.246	2.529	0.137

In this study, the actual implementation involved several key steps to ensure the effectiveness and reliability of the research method. First, we collected text data such as M&A reports, performance reports, annual reports, social responsibility reports, directors' resumes and supervisors' resumes from listed companies in China and the United States, which constituted the data set for this study. These data sets undergo rigorous preprocessing, including text cleaning, feature extraction, and vectorization, to facilitate model input and processing. In the model construction stage, we adopted the improved LDA model to analyze the topic and emotion of the document, and combined the multi-objective evolutionary optimization technology to solve the optimization problem of the risk contagion model. In addition, to deal with the imbalance in the data set, we apply the weighted Smote algorithm to improve the model's ability to recognize a few classes. In the experimental phase, we used the NS3 simulation platform for network simulation and cross-validated the model to evaluate the performance of the model. Finally, we compare the performance of the proposed model with other classical algorithms, and the results show that our model has significant advantages in terms of efficiency and accuracy. The whole research process strictly follows the methodology of scientific research to ensure the reliability and validity of the research results.

The identification of information used in this article involves the recognition of direct identifiers and quasi-identifiers. The purpose is to more accurately describe, distinguish, or interpret concepts. By quoting these two symbols, we can better elaborate our own views, provide richer information, and enhance the persuasion and readability of the text. In this study, the methods for identifying direct identifiers are primarily divided into two categories. Firstly, regular expressions are utilized as search patterns defined by character sequences to identify data that strictly adheres to specific composition patterns, such as ID cards, bank cards, mobile phone numbers, email addresses, etc. Secondly, a deep learning-based named entity recognition method is employed for recognizing and extracting names from text sequences [36-38]. Depending on the data composition type within a field (e.g., pure number value composition, mixed number and string text composition, pure string text composition), different recognition methods and combination strategies are adopted to identify direct identifiers accordingly. Additionally, in order to enhance efficiency during recognition process, identification is performed within the unique value set of each field while establishing a mapping dictionary that links back the recognized results with their original field sequence elements [39]. The identification method for quasi-identifiers mainly relies on metadata recognition using keyword thesauri. Due to the complexity of quasi-identifier types often lacking standard composition patterns; structured datasets inherently convey certain information through their structure itself which can be leveraged by searching information types within the collection of field names using keyword thesauri in order to identify corresponding quasi-identifiers present in the dataset. Following completion of machine-based recognition phase; manual sampling is conducted to verify machine-generated results [40-41].

The calculation of management intonation is based on the literature practice, employing simple word frequency statistics for computation. The process encompasses the following steps: Firstly, a computer program is utilized to position the text from MD&A, performance presentation, annual report, and social responsibility report. Subsequently, the text information pertaining to future outlook is selected. A manual review of numerous financial annual reports is conducted with a focus on identifying common characteristics in the initial and final stages of text information disclosure. Start keywords and end keywords are chosen respectively as indicators for the beginning position and ending position of the text information. The required text information is then selected as the standard. Due to changes in annual report disclosure criteria, adjustments have been made multiple times to these positioning keywords;

for instance, starting stage keywords include "outlook" and "future development," while end stage keywords encompass phrases like "investment situation." After filtering out basic text information, any unqualified content undergoes manual adjustment to enhance screening accuracy. Manual screening criteria involve examining excessively long or short texts based on their size and adjusting texts according to iconic words that represent specific content. Secondly, a dictionary is constructed to translate positive and negative English words into Chinese based on the Chinese context. The dictionary is created by referring to commonly used Chinese emotion dictionaries, deleting positive and negative words that are not relevant to the text [42-43], and retaining commonly used Chinese words found in frequently used textual information. The Jieba word segmentation program in Li Python is utilized for segmenting the text information, selecting positive and negative emotion words based on the segmentation results, constructing the benchmark emotion word database for this study, and calculating management's intonation variable data.

The present study proposes an optimization control method for enterprise information disclosure risk (ECR-MTM) from the perspective of management intonation manipulation. It compares with SMOTE method (SMOTE), Gaussian Mixture Clustering (GMC), Weighting SMOTE (WSMOTE), Ant Colony Optimization (ACO), Swarm Optimization (SWO), K-Shell Centrality (KSC), and Weighted K-Shell Degree Neighborhood (WKS-DN) [44-48]. In addition to these algorithms, there are also decision tree method (DT), artificial neural network method (ANN), random forest method (RF) and so on. These methods are not designed to solve a single problem, so direct comparisons require data preprocessing, including data augmentation for unbalanced data, cross-validation methods to divide data sets into training and test sets to assess the performance of the model, and optimize each parameter to ensure they compare at the same level.

Risk control problems in complex networks are commonly regarded as binary classification tasks. In the evaluation confusion matrix of binary classification tasks with two classes, True Positive (TP) represents the number of accurately predicted links, while True Negative (TN) represents the number of correctly predicted non-links. False Positive (FP) indicates the count of incorrectly predicted links, and False Negative (FN) signifies the count of inaccurately predicted non-links [49-50]. Based on this framework, the evaluation metrics employed in this study include accuracy, accuracy rate, recall rate, and F-measure expressed by Eq. (24)-(27), respectively. Additionally, to complement existing literature findings, two precision functions are utilized: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The specific calculation methods for these functions are presented in Eq. (28) and Eq. (29), respectively.

$$Precision = \frac{TP}{TP+FP} \tag{24}$$

$$Accuracy = \frac{TP+TN}{P+N} \tag{25}$$

$$Recall = \frac{TP}{TP+FN} \tag{26}$$

$$F-measure = \frac{2*precision*recall}{precision+recall} \tag{27}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|f_i - y_i| \tag{28}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(observed_t - predicted_i)^2} \tag{29}$$

B. Experimental Results

In Table II, two evaluation indexes $Accuracy@k$ and MAP are used to compare the performance of the enterprise information disclosure risk optimization control method proposed in this paper with other information disclosure risk control algorithms in four real enterprise information disclosure network datasets from the perspective of management tone manipulation by ECTR MTM. The values in each cell in Table II represent the likelihood of a result resulting from completing a specific task under specific conditions. Specifically, each cell represents a combination of event, condition, task, and result. The value of each cell is usually a percentage, which represents the possibility of the result generated by the completion of a specific task under certain conditions, so as to better help the team understand the business process and system, identify potential risks and problems, and formulate corresponding solutions. By comparing the values of different cells, you can identify possible bottlenecks and opportunities for optimization in the process. The experimental results of this study are shown in Table II. It can be seen that the results of the algorithm proposed in this paper are superior to other algorithms, indicating that the algorithm proposed in this paper has higher accuracy than other algorithms.

The results of the area under the curve for the proposed ECTR-MTM risk optimization control method in the real social network dataset, from the perspective of management tone manipulation, are presented in Table III. The area under the curve values measure the average performance of the algorithm on all possible samples. The area under the curve is between 0 and 1, and a larger value indicates better performance. As can be seen from the table, the area under the curve value of the proposed ECR-MTM algorithm is greater than that of other algorithms, indicating that the proposed algorithm has better performance than other algorithms. This study demonstrates that the proposed risk optimization control method for corporate information disclosure, considering management intonation manipulation, yields superior experimental outcomes when applied to real corporate information disclosure network datasets.

The collaborative recommendation optimization method of CAHT-ROM, based on an enhanced ant colony algorithm and hypernetwork technology as presented in Table IV, exhibits a general superiority over alternative approaches. This can be attributed to its ability to swiftly respond, dynamically adjust in real-time, and optimize social networks effectively. As can be seen from the table, the running time of the proposed ECTR-MTM algorithm is significantly lower than other algorithms, indicating that the ECTR-MTM algorithm can complete the task in a shorter time than other algorithms, and has faster response speed and higher efficiency when processing data. Furthermore, it significantly reduces the overall loss within the social network.

TABLE II. COMPARISON OF *Accuracy@k* AND MAP INDEXES IN DIFFERENT DATA SETS

Name of algorithm		SMOTE	GMC	WSMOTE	ACO	SWO	KSC	WKS-DN	ECTR-MTM
Enterprise dataset of listed companies in China	Acc@1	0.022	0.031	0.033	0.052	0.073	0.083	0.098	**0.124**
	Acc@5	0.025	0.036	0.037	0.057	0.077	0.092	0.129	**0.173**
	Acc@10	0.037	0.048	0.042	0.068	0.104	0.128	0.135	**0.245**
	MAP	0.042	0.051	0.054	0.073	0.125	0.152	0.153	**0.319**
Data set of chain directors of listed companies in China	Acc@1	0.012	0.028	0.038	0.052	0.063	0.129	0.130	**0.238**
	Acc@5	0.024	0.031	0.042	0.058	0.082	0.142	0.142	**0.263**
	Acc@10	0.037	0.048	0.051	0.063	0.121	0.163	0.253	**0.301**
	MAP	0.045	0.062	0.065	0.072	0.135	0.172	0.279	**0.322**
Corporate data set of US public companies	Acc@1	0.014	0.022	0.035	0.048	0.082	0.120	0.128	**0.218**
	Acc@5	0.023	0.025	0.039	0.052	0.085	0.143	0.142	**0.237**
	Acc@10	0.031	0.041	0.052	0.071	0.102	0.175	0.150	**0.373**
	MAP	0.054	0.065	0.068	0.083	0.189	0.188	0.235	**0.468**
Data set of chain directors of US public companies	Acc@1	0.012	0.024	0.038	0.049	0.073	0.108	0.117	**0.225**
	Acc@5	0.018	0.027	0.039	0.051	0.082	0.110	0.119	**0.271**
	Acc@10	0.023	0.024	0.041	0.053	0.064	0.122	0.142	**0.391**
	MAP	0.028	0.035	0.048	0.061	0.072	0.153	0.175	**0.428**

Note: Values in bold all indicate that the algorithm they correspond to performs well.

TABLE III. AREA UNDER THE CURVE VALUES OF EACH DATA SET IN DIFFERENT METHODS

Level of cross validation	Data set name	Optimization algorithm			
		SMOTE	GMC	WSMOTE	ACO
2-fold	Enterprise data set of listed companies in China	0.113	0.219	0.229	0.354
	Data set of chain directors of listed companies in China	0.136	0.230	0.201	0.300
	Corporate data set of US public companies	0.139	0.241	0.285	0.385
	Data set of chain directors of US public companies	0.142	0.235	0.318	0.335
4-fold	Enterprise dataset of listed companies in China	0.143	0.294	0.257	0.339
	Data set of chain directors of listed companies in China	0.134	0.285	0.239	0.318
	Corporate data set of US public companies	0.125	0.202	0.248	0.309
	Data set of chain directors of US public companies	0.113	0.238	0.316	0.326
10-fold	Enterprise data set of China's listed companies	0.150	0.248	0.285	0.329
	Data set of chain directors of listed companies in China	0.168	0.285	0.344	0.318
	Corporate data set of US public companies	0.248	0.319	0.353	0.420
	Data set of chain directors of US public companies	0.214	0.341	0.423	0.412
Cross verification rating	Data set name	Coordinated recommendation algorithm			
		SWO	KSC	WKS-DN	ECTR-MTM
2-fold	Enterprise data set of listed companies in China	0.384	0.438	0.542	**0.773**
	Data set of chain directors of listed companies in China	0.342	0.381	0.470	**0.721**
	Us public Company enterprise dataset	0.394	0.402	0.451	**0.702**
	Data set of chain directors of US public companies	0.365	0.436	0.466	**0.683**
4-fold	Enterprise dataset of listed companies in China	0.355	0.443	0.555	**0.790**
	Data set of chain directors of listed companies in China	0.302	0.430	0.409	**0.729**
	Us public company enterprise dataset	0.359	0.429	0.420	**0.809**
	Data set of chain directors of US public companies	0.335	0.443	0.535	**0.753**
10-fold	Enterprise data set of China's listed companies	0.378	0.470	0.491	**0.785**
	Data set of chain directors of listed companies in China	0.311	0.438	0.409	**0.722**
	Us public Company enterprise dataset	0.330	0.402	0.593	**0.736**
	Data set of chain directors of US public companies	0.352	0.462	0.538	**0.749**

Note: Values shown in bold all indicate that their corresponding algorithms perform well.

TABLE IV. RESULTS OF ALGORITHM RUNNING TIME COMPARISON IN DIFFERENT DATA SETS

Data set Name	SMOTE	GMC	WSMOTE	ACO
Corporate data set of listed companies in China	242.351	819.849	492.503	849.232
Data set of chain directors of listed companies in China	324.831	779.304	879.394	809.753
Us Public Company Enterprise dataset	741.416	693.351	792.429	532.230
Data set of chain directors of American public companies	683.315	684.348	602.691	548.249
Data set name	SWO	KSC	WKS-DN	ECTR-MTM
Enterprise dataset of listed companies in China	323.624	242.402	435.317	**59.529**
Data set of chain directors of listed companies in China	792.938	310.985	782.295	**40.495**
Us public Company Enterprise dataset	630.402	248.312	591.940	**32.390**
Data set of chain directors of US public companies	534.102	320.382	483.293	**41.204**

Note: The values shown in bold all indicate that the algorithm they correspond to performs well.

V. CONCLUSION

Accounting information disclosure involves the agency problem of enterprises and reflects the moral hazard and adverse selection tendencies of enterprise managers and accounting practitioners through the influence of information asymmetry. In this process, enterprises can not only convey positive signals to banks through timely and transparent accounting information disclosure but also facilitate sufficient bank lending and financing during financial difficulties by enhancing financing availability, thereby preventing the formation and contagion of liquidity risks. Moreover, high-quality information disclosure enables positive interaction with investors. Additionally, from the perspectives of financing cost and robustness, enterprises can reduce their financing costs while mitigating the risk of confidence crisis occurrence, ensuring short-term capital adequacy, avoiding sudden capital outflows, as well as reducing the probability of liquidity risk occurrence and diffusion. The present study constructs an enterprise information disclosure risk contagion model from the perspective of management intonation manipulation, which is subsequently analyzed and solved using the improved LDA model and multi-objective evolutionary optimization method. The results demonstrate that the proposed approach exhibits higher efficiency and accuracy, enabling better control over environmental and association effects. Furthermore, this paper enhances the weighted Smodule algorithm. Experimental findings indicate that this method effectively improves the accuracy and stability of the enterprise information disclosure risk infection model. Consequently, this model can provide robust support for enterprise information disclosure risk control.

With technological advancements and the increasing complexity of financial markets, accounting information disclosure faces new challenges and opportunities. Big data and artificial intelligence technologies will enhance the efficiency of financial data processing, provide more accurate and timely information, and reduce information asymmetry. Blockchain technology will make financial data immutable and transparent, enhancing investor trust, and smart contracts will automate financial reporting and auditing processes. Global financial integration requires companies to strengthen internal control and compliance management, and regulatory bodies need to establish unified international standards. The importance of ESG factors in investment decisions is increasing, prompting companies to pay more attention to non-financial information disclosure. Technological innovation will have a significant impact on accounting information disclosure, and companies, investors, and regulatory agencies need to work together to ensure that information is transparent, accurate, and timely, in order to stabilize and promote the development of capital markets.

ACKNOWLEDGMENT

The research was supported by Research on the Effect Measurement and Path of Management Tone on the Quality of Accounting Information in Enterprises (Grant No. 2022JYTYB09).

REFERENCES

[1] Drissi A, Khemiri A, Sassi S, et al. LDA plus: An Extended LDA Model for Topic Hierarchy and Discovery. 14th Asian Conference on Intelligent Information and Database Systems (ACIIDS), 2022, 1716:14-26.

[2] A. Madani, B. Ombuki-Berman and A. Engelbrecht. Decision Space Scalability Analysis of Multi-Objective Particle Swarm Optimization Algorithms. IEEE Congress on Evolutionary Computation (CEC), 2021:2179-2186.

[3] S. Chansamorn and W. Somgiat. Improved Particle Swarm Optimization using Evolutionary Algorithm. 19th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2022:1-5.

[4] Sedighizadeh D, Masehian, E, et al. GEPSO: A new generalized particle swarm optimization algorithm. MATHEMATICS AND COMPUTERS IN SIMULATION, 2021, 179(194-212).

[5] Nedjar I, Mahmoudi S, Chikh M A. A topological approach for mammographic density classification using a modified synthetic minority over-sampling technique algorithm. International Journal of Biomedical Engineering and Technology, 2022, 38(2):193-214.

[6] Anusha Y, Visalakshi R, Srinivas K. Imbalanced data classification using improved synthetic minority over-sampling technique. Multiagent and Grid Systems, 2023,19(2):117-131.

[7] R. H. Coase. The Nature of the Firm. Economica, 1937, 4(16):386~405.

[8] Watts R. L., Zimmerman J. L. A Positive Accounting Theory. The Accounting Review, 1986, 65(5):455-468.

[9] Jung K, Kim C, Yun J. The effect of corporate risk management on cyber risk mitigation: Evidence from the insurance industry[J]. The Geneva Papers on Risk and Insurance-Issues and Practice, 2024: 1-43.

[10] Miller G S, Skinner D J. The evolving disclosure landscape: How changes in technology, the media, and capital markets are affecting disclosure. Journal of Accounting Research, 2015, 53(2): 221-239.

[11] Blankespoor E, Miller G S, White H D. The role of dissemination in

market liquidity: Evidence from firms' use of Twitter ™. The Accounting Review, 2013, 89(1): 79-112.

[12] Chau G K, Gray S J. Ownership structure and corporate voluntary disclosure in Hong Kong and Singapore. The International Journal of Accounting, 2002, 37(2): 247-265.

[13] Lankton N, Price J B, Karim M. Cybersecurity breaches and the role of information technology governance in audit committee charters[J]. Journal of Information Systems, 2021, 35(1): 101-119.

[14] Beaver, W.H. What should be the FASB's objectives?. The Journal of Accountancy,1973(August): 49-56.

[15] Paul M. Healy and Krishna G. Palepu. Information Asymmetry Corporate Disclosure and the Capital Markets: A Review of the Empirical Disclosure Literature. Journal of Accounting and Economics, 2001 (31):405-440

[16] Chow C W, Wong-Boren A. Voluntary Financial Disclosure by Mexican Corporations. The Accounting Review, 1987, on conversion (3): 533-541.

[17] Raffournier B., 1995, "The Determinants of Voluntary Financial Disclosure by Swiss Listed Companies", The European Accounting Review, July, PP261-280.

[18] Mahmaud, Hossain, Andrew, K., Prevost and Ramesh, P., Rao, 2001, "Corporate Governance in New Zealand: the Effect of the 1993 Companies Act on the Relation between Board Composition and Firm Performance ", Pacific-Basin Finance Journal, Vol.9.

[19] Mitchell, Jason. D, Chia, Chris. W. L, Loh, Andrew. S., 1995, "Voluntary Disclosure of Segment Information", Accounting and Finance, Vol.35, PP1-16.

[20] Morck, R., Shleifer, A. and Vishny, R., 1988, "Management Ownership and Market Valuation: an Empirical Analysis. ", Journal of Financial Economics, Vol. 20, PP293-315.

[21] Paul M. Healy and Krishna G. Palepu. The Effect of Firms Financial Disclosure Strategies on Stock Prices. Accounting Horizons, 1993(3): 1-11.

[22] A.C. Littleton, Structure of Accounting Theory, American Accounting Association, 1953.

[23] G. Biddle, G. Hilary, Accounting Quality and Firm-level Capital Investment, The Accounting Review, Vol. 81, No. 5,2006, pp. 963-982.

[24] Dechow P M, Sloan R G, Hutton A p. Dtecting earnings management. Accounting Review,1995,70(2):193-225.

[25] Dechow P, Ge W, Schrand C. Understanding earnings quality: A review of the proxies, their determinants and their consequences. Journal of accounting and economics.2010,50(2-3):344-401.

[26] De Franco G, Kothari SP, Verdi RS. The benefits of financial statement comparability. Journal of Accounting Research. 2011,49 (4): 895-931.

[27] Healy PM, Hutton AP, Palepu KG. Stock performance and intermediation changes surrounding sustained increases in Disclosure. Contemporary accounting research. 1999 (3): 485-520.

[28] Porter, M. and Van, D. L. Toward a New Conception of the Environment Competitiveness Relationship. Journal of Economic Perspective, 1995, 9(4): 97-118.

[29] Prajogo, D. and Sohal, A. The Integration of TQM and Technology/R&D Management in Determining Quality and Innovation Performance. Omega, 2016, 34(3): 296-312.

[30] Tran C, Zheleva E. Heterogeneous peer effects in the linear threshold model. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(4): 4175-4183.

[31] Agrawal A, Shanly K S A, Vaishnaw K, et al. Reverse dictionary using an improved CBOW model. Proceedings of the 3rd ACM India Joint International Conference on Data Science & Management of Data (8th ACM IKDD CODS & 26th COMAD), 2021: 420-420.

[32] Matsuda T, Uehara M, Hyvarinen A. Information criteria for non-normalized models. Journal of Machine Learning Research, 2021, 22(158): 1-33.

[33] Guo H, Wang Y, Wang B, et al. Does prospectus AE affect IPO underpricing? A content analysis of the Chinese stock market. International Review of Economics & Finance, 2022, 82: 1-12.

[34] Abedinia A, Seydi V. Building semi-supervised decision trees with semi-cart algorithm. International Journal of Machine Learning and Cybernetics, 2024: 1-18.

[35] Rajan, R. G. Insiders and Outsiders: The Choice between Informed and Arm's-Length Debt. The Journal of Finance, 1992, 47(4): 1367-1400.

[36] Rioja, F. and Valev, N. Does One Size Fit All?: A Reexamination of the Finance and Growth Relationship. Journal of Development Economics, 2004, 74(2): 429-447.

[37] Romer, P. M. Increasing Returns and Long-Run Growth. Journal of Political Economy, 1986, 94(5): 1002-1037.

[38] Rong, Z., Wu, X. and Boeing, P. The Effect of Institutional Ownership on Firm Innovation: Evidence from Chinese listed Firms. Research Policy, 2017, 46(9): 1533-1551.

[39] Ross, S. The Determination of Financial Structure: The Incentive Signaling Approach. Bell Journal of Economics, 1977, 8(1): 23-40.

[40] Rothwell, A., Herbert, I. and Rothwell, F. Self-perceived Employability: Construction and Initial Validation of A Scale for University Students. Journal of Vocational Behavior, 2008, 73(1): 1-12.

[41] Sasidharan, S., Lukose, P. J. J. and Komera, S. Financing Constraints and Investments in R&D: Evidence from Indian Manufacturing Firms. The Quarterly Review of Economics and Finance, 2015, 55(1): 28-39.

[42] Satrovic, E., Muslija, A., J. and Abul, S., et al. Interdependence between Gross Capital Formation, Public Expenditure on R&D and Innovation in Turkey. Journal of Balkan and Near Eastern Studies, 2021, 23(1): 163-179.

[43] Stads, G. J. and Beinitema, N. Agricultural R&D Expenditure in Africa: An Analysis of Growth and Volatility. The European Journal of Development Research, 2015, 27(3): 391-406.

[44] Stiglitz, J. and Weiss, A. Credit Rationing in Markets with Imperfect Information. American Economic Review, 1981, 71(3): 393-410.

[45] Sueyoshi, T. and Goto, M. Can R&D Expenditure Avoid Corporate Bankruptcy? Comparison between Japanese Machinery and Electric Equipment Industries using DEA–discriminant Analysis. European Journal of Operational Research, 2009, 196(1): 289-311.

[46] Thorwarth, S., Kraft, K. and Czarnitzki, D. The Knowledge Production of R&D. Economics Letters, 2009, 105(1): 141-143.

[47] Tomlinson, P. R. Co-operative Ties and Innovation: Some New Evidence for UK Manufacturing. Research Policy, 2010, 39(6): 762-775.

[48] Tong L, Wang H, Xia J. Stakeholder Preservation or Appropriation? The Influence of Target CSR on Market Reactions to Acquisition Announcements. Academy of Management Journal, 2020, 63(5): 1535-1560.

[49] Tong, X. and Frame, J. D. Measuring National Technological Performance with Patent Claims Data. Research Policy, 1994, 23(2): 133-141.

[50] Tsai, K. H. and Wang, J. C. External Technology Acquisition and Firm Performance: A Longitudinal Study. Journal of Business Venturing, 2008, 23(1): 91-112.

A Machine Learning-Based Intelligent Employment Management System by Extracting Relevant Features

Yiming Wang, Chi Che*
School of Economics and Management, Jilin Engineering Normal University, Changchun, 130000, China

***Abstract*—In recent years, there has been a significant increase in the number of students trying to broaden the work opportunities available to college graduates. This study presents an intelligent employment management system that may be used in educational institutions for students to gain a better understanding of their occupations and analyzing the sectors in which they will work. In this article, the fundamental concepts of information recommendation are discussed, as well as a customized recommendation system for entrepreneurship that is provided. The fundamental information and personal interest points of college students are represented by feature vectors. These feature vectors provide positive theoretical support for the career planning and employment and entrepreneurship information suggestions of college students. In conclusion, an analysis of the performance of the proposed model is performed to provide college students with a system that is both convenient and quick in terms of information recommendation. This will result in an indirect improvement in the employment rate of graduates and will provide solutions that correspond to the problem of difficult employment.**

Keywords*—Employment management system; recommendation system; feature index; accuracy and employment intention index

I. INTRODUCTION

The college job information that is available now has the characteristics of being diverse, very efficient, and quite extensive. As a consequence of the continuous advancement of information technology, this is one of the characteristics that has come into existence within the field [1]. There has been a significant amount of time that has passed since the conventional data statistics system was able to meet the actual requirements of data statistics [2]. The statistics of employment information in colleges and universities need to design a new model with intelligent technology to carry out technological innovation. This is in addition to the fact that they are responsible for conducting scientific data screening and doing a thorough job of data analysis. When there are a large number of collections, some of which include real data, there are also a significant number of collections that contain misleading information. Because of this, statisticians are necessary to design a data screening technique that is both scientific and efficient to carry out the real statistical process successfully. Two components make up this system: artificial screening and intelligent screening In the field of statistics, it is essential to make use of computers to provide a smooth operation process. This is important to guarantee a high level of efficiency in the collection of statistical data and in designing creative data statistics [3, 4].

It is essential to perform successfully while simultaneously developing new methods of data collection [5-8]. The establishment of a system for the collecting of data over an extended period is necessary in order to improve the rate at which high-quality data mining is performed. In and of itself, data mining is a kind of data deep processing that is focused on achieving certain goals. To accomplish this objective, active data users will be required to do statistical data deep processing for extended periods in real-world contexts [9, 10]. These are the most common types of operations that fall under the category of practical application. Effective management of cumulative data statistics and data analysis with a specific goal are examples of these types of operations.

Users of an intelligent data mining system are required to provide statistical data that is relevant to the information needs of the issue, taking into consideration the many real scenarios. This is necessary for the system to work properly. The source of the data and the method of computation must be able to satisfy the format criteria of the other units for the data statistics to be accurate when they are recorded and utilized in various other units. The capabilities of intelligent data mining technology have increased, and it is now able to better satisfy the requirements of complex data management [11].

Education that focuses on career planning helps to promote entrepreneurial education in higher education by increasing its effectiveness, affinity, and appeal. At the same time, it provides students with a more complete education. In conclusion, it discusses how this education may be used as a defence mechanism against the professional goals of college students [12, 13]. The demands of contemporary data processing technologies are too much for old recommendation systems to handle, especially considering the growing number of consumers. One application of data processing and collection in the context of remote learning is the utilization of virtual reality gesture recognition to provide recommendations for educational resources that may be used in the classroom environment [14]. The influence of user comments on their surroundings, the social platform itself, and content active filtering and recommendation are some of the fundamental user behaviours and social software services that are available on the platform [15]. Conceptual ideas, which are the user's perceptions of the items that are inferred from the usage of label data for free categorization, are utilized to achieve the selection of recommended things. This is performed via the utilization of conceptual concepts. Table I presents the comparison of proposed method with literature studies [16-19].

TABLE I. COMPARISON OF PROPOSED METHOD WITH LITERATURE STUDIES

Work focus	Key Findings	Relevant Literature
Machine Learning in HR	- ML algorithms effectively screen resumes, predict candidate suitability, and reduce bias in hiring. - Predictive models forecast employee performance, identify high-potential individuals, and provide targeted development plans. - ML algorithms predict attrition risks, enabling proactive retention strategies. - Personalized learning paths and skill development recommendations enhance employee growth.	Hong Zhu (2021) [16]
Feature Engineering for HR Analytics	- Feature selection and dimensionality reduction improve model performance and interpretability. - Domain expertise from HR professionals is crucial for effective feature engineering.	Ali Raza, et al., (2022) [17]
Ethical Considerations in HR Tech	- ML models can perpetuate existing biases, leading to unfair outcomes. - Addressing bias and ensuring fairness and transparency is crucial. - Data privacy and security measures are essential to protect employee data.	Andrieux, P., et al. (2024) [18]
Gaps in Existing Employment Management Systems	- Lack of personalization in recommendations. - Data silos and incompatibility across HR systems. - Limited predictive capabilities for future workforce needs and risks. - Potential for bias and unfair outcomes. - Complex and user-unfriendly interfaces.	Negt, P., et al., (2024) [19]
Proposed Solution: ML-Based Intelligent Employment Management System	- Personalized recommendations for career development and skill enhancement. - Data integration and analysis for a holistic view of the workforce. - Predictive analytics for proactive HR decision-making. - Mitigating bias and ensuring fair and equitable outcomes. - User-friendly interfaces for improved accessibility and engagement.	Proposed framework for an employment system

Additionally, the recommendation algorithm of the system creates a collection of synonymous labels in order to gather labels that are similar to one another to use them in definition classification, which is the foundation of label labelling [20]. The user-available resources were subjected to the cluster seepage methodology, which ultimately led to the creation of this selection [21]. There are a significant number of unfinished areas in the process of transferring learning that will be used in information recommendation systems [22, 23]. Before beginning the process of searching for information, it is required to first do an analysis of the information access history records of users, then extract their locations, then label them in accordance with semantic approaches, and finally search for individuals [24, 25].

A. Contribution

The study proposes a novel ML-based employment management system architecture or approach. A unique mix of algorithms, data sources, or feature extraction methods may be used. A suggested system may have benefits over current systems [16-19]. Relevant characteristics may improve employee recruiting, performance assessment, and retention forecasts and classifications. The system may handle massive datasets quicker due to computational efficiency. HR specialists may find the system easier to utilize. The study identifies and extracts key employee data elements for certain activities. This may reveal employee performance, satisfaction, and retention variables. The report might boost industrial use of ML technologies by showing how they solve real-world HR problems.

II. PROPOSED RECOMMENDATION SYSTEM: METHODOLOGY

The proposed model collects and filters the information that the user provides, originating from a variety of sources [26], and then converts the information content into text information. This is the process that is referred to as the content suggestion process. Fig. 1 presents the block diagram of the proposed recommendation model.

Step 1: Remodel the content. The assumption is made that the text information after the content conversion is transformed into vector points of various dimensions of the space vector. Each feature point (i) of the content is therefore turned into vector points (x_i) is content feature point, and y_i is weight. The expression for information content $I(i)$ is

$$I(i) = \{x_1 : y_1; x_2 : y_2; \dots; x_i : y_i\}$$

The feature points of the information are represented by vectors:

$$I(i) = \{y_1, y_2, y_3, \dots y_i\}$$

Step 2. An analysis is performed to determine the feature index (F_i) as well as the similarity connection between the content feature points. The many users each have their own unique preference models, and the statement may be stated as,

$$S(i) = \frac{\sum_i F_i I(i)}{F_i}$$

Step 3: Determine the degree of resemblance between each of the content's three points of similarity is written as,

$$\sin(S, I) = \cos\left(\vec{S}, \vec{I}\right) = \frac{\vec{S}, \vec{I}}{\left|\vec{S}\right| \times \left|\vec{I}\right|}$$

The three points of similarity is often computed for features like, Job Satisfaction, Job Security (Both the Employment Intention Index (EII) and the proposed Employment Management System emphasize the importance of job satisfaction in predicting employee retention), Career Advancement Opportunities (Both EII and proposed system recognize the significance of career advancement opportunities in influencing employee retention and job satisfaction) and Work-Life Balance (Both EII and proposed system)

acknowledge the impact of work-life balance on employee retention, job satisfaction, and overall well-being.

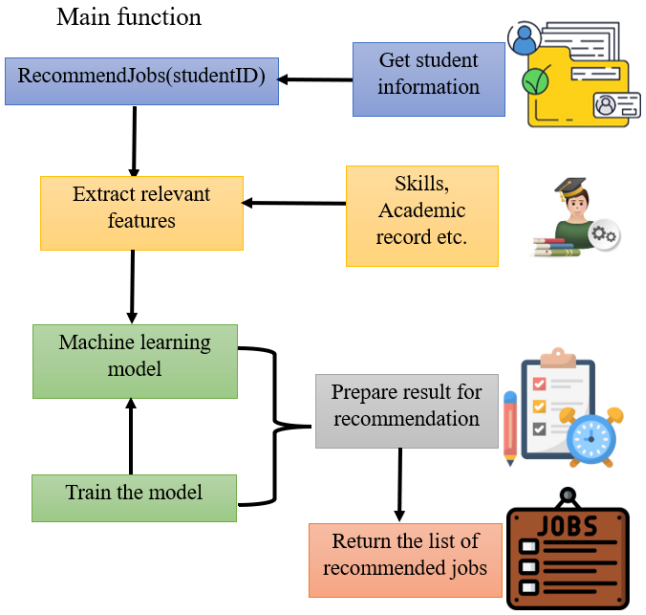


Fig. 1. Block diagram of the proposed recommendation model.

Step 4: The outcomes of the information suggestion should be generated. Users are provided with recommendations based on the information that is integrated and has the greatest similarity. The construction of the information distribution model is done in order to accomplish the mission of providing college students with information recommendations for career planning, employment, and entrepreneurial endeavors. Personal information should be collected from educational institutions, and data sequences of information characteristics should be produced such that

$$x = \{x_1, x_2, x_3, \dots, x_i\}$$

Step 5: Integrate the characteristic points of the personal information of the students, which is to say, the expression of the characteristic quantity of the learning model is written as

$$C(i) = \frac{x_i}{\sum_i F_i y_i}$$

Step 6: Compute the sample model (M) of the characteristic points of the personal information of college students such that

$$M = \frac{kC(i)}{q\sum_i x_i y_i}$$

The index of employment intention is denoted by q in the information feature point sampling model, while the quantity of employment information is denoted by k.

Step 7: Calculate the final feature point value Students in higher education develop their own career goals based on their majors and the interests they have outside of school and then combine these aspects. In the context of college students' career planning and job information referral, the learning model (C) describes $C = Min\{\max(C_i)\}$.

The following is an example of the self-adaptive sampling model that divides the group size of college students into distinct professional sectors from one another such that

$$M = C[1 - C(i)]^{k-1}$$

To search the employment requirements and interest feature points of colleges, the deep learning approach is used, and the final feature point value $(\mathcal{M})$ is written as

$$\mathcal{M} = \sum_0^{\infty}\left[\frac{1 - C(i)}{M}\right]^k$$

B. Pseudocode for the Proposed System

Function: RecommendJobs(studentId)

// This function recommends jobs to a student based on their profile and machine learning model

Function RecommendJobs(studentId)

// Input: studentId *(unique identifier for a student)*

// Output: List of recommended jobs *(job IDs or descriptions)*

1. Get student data

studentData = GetStudentData(studentId)

// Call to StudentData class *(refer to previous explanation)*

2. Extract relevant features (e.g., skills, academic record)

skills = ExtractSkills(studentData)

experience = ExtractExperience(studentData)

majors = ExtractMajors(studentData)

3. Load pre-trained machine learning model

model = LoadModel("JobRecommendationModel.pkl")

// Replace with actual model loading method

4. Prepare data for prediction

preparedData = PrepareDataForModel(skills, experience, majors)

// Feature engineering

5. Make prediction using the model

predictedJobIDs = model.predict(preparedData)

6. Retrieve details of recommended jobs

recommendedJobs = GetJobDetails(predictedJobIDs)

// Call to EmploymentData class

7. Return the list of recommended jobs

return recommendedJobs

End Function

The above pseudocode is an example of a recommendation system that has been streamlined according to following steps:

1) The student ID serves as the foundation for the collection of information about the students.

2) The data on the students are taken into consideration, and from the information that is accessible, significant aspects such as the student's skills, experiences, and majors are extracted.

3) It loads a machine learning model that has previously been trained and is used for job recommendation.

4) The gathered attributes are then formatted for the model.

5) The student profile serves as the foundation upon which the model bases its ability to generate predictions for relevant job IDs.

6) This step involves retrieving information about the task from the database by using the required IDs.

7) As a last step, the function gives the learner a list of potential careers that are recommended for them to pursue.

C. Data Preparation

1) Collect pertinent information from a variety of human resource management (HRM) systems, such as learning management, performance management, and employee recruiting.

2) Inconsistencies, outliers (which may be eliminated or adjusted), and missing numbers (imputation) are all things that need to be addressed during the data cleaning process.

3) Scale numerical attributes to a common range, such as between 0 and 1, or with zero mean and unit variance, as part of the process of normalizing and standardizing the data. This is an example of a data transformation.

4) The dataset should be divided into three parts: one for training, one for validation, and one for training and testing. Listed below is the distribution of the available resources: 70% will be allocated for teaching, 15% will be used for validation, and 15% will be used for testing.

D. Feature Engineering

We performed a correlation research is to identify the characteristics that have a strong relationship with the variable that is being studied (for example, employee performance or turnover).

E. Training the Model

1) Make use of the chosen training data in order to train the model that was made selection.

2) Monitor how far along the training process you are: There are a number of measures that have to be monitored, including the accuracy, precision, F1-score, and feature index value.

F. Parameter Tuning

During the process of optimization, the learning rate is responsible for regulating the step size. Parameter tuning is the process of modifying the parameter. When it comes to machine learning networks, the design complexity of the neural network is determined by the number of hidden layers and neurons that constitute the network.

III. The Database Acquisition System: Capturing Student ID

The design of this system's network topology is comprised of three distinct services: the database service, the cluster management service, and the business management service. These services are not connected. The architecture must be broken down into these three different elements to function properly. Among the many examples of online firm management services, one example is website platforms that provide customers with a data collection system. This service may be used by commercial enterprises. The acquisition system can't perform its functions correctly if it does not gather the cluster management service, which is an essential component of the big data service. Assisting the data carrier that is included in the acquisition system is one of the functions that the database service is responsible for [18–20].

Since the platform provides services which are delivered via the external network, the platform needs to have a firewall in order to guarantee that users may access the system without risk. A multitude of subsystems that are quite similar to one another are what make up the business management service. Some examples of these subsystems are the portal, data gathering, task scheduling, visualization, and many more. To guarantee that data transfer between web servers is continuous, the creation of routing between subsystems is essential. For this reason, it is of the utmost importance to maintain continual connections between subsystems. Data collection from this vast data cluster and the maintenance of correct access among the machines that comprise the cluster are both obligations that fall within the purview of the acquisition system. It is the purpose of the acquisition system to ensure that this occurs. At this point, it is of the utmost importance that you provide the Web service access to the cluster for it to be able to get data that is stored inside the cluster.

Following the hierarchical architecture, the overall design of the system is composed of three levels: one level is dedicated to visual configuration, another level is dedicated to aggregation processing, and the third level is dedicated to storage. The acquisition and processing layer is the most important component of the system since it is the one that is responsible for actually gathering data in line with the acquisition technique that has been defined. Together, the term "storage" refers to both the end destination of the data as well as the initial method that was used to obtain it. In addition to that, it offers the capability of caching services. The visually configured layer, which is directly user-oriented, is responsible for carrying out all of the interactions that take place between the user and the acquisition system. On the browser side of the interface, the user can configure the components of the collection that are responsible for processing, terminal, and source. Additionally, the user can construct their collecting technique, which is an additional benefit.

The functions of managing processes such as data collection and the actual transactions that make use of it are the responsibility of the layer that is responsible for acquisition and processing. The collection may be carried out by the information on the workflow setup after the procedure of acquiring the data has been completed once and for all.

IV. Analysis of Results

The data set was selected from the time series categorization collection at UC Riverside [27]. Following the tagging of the time series included within each dataset, the datasets are then divided into training and test sets, each of which is characterized by a distinct scale. Table II summarizes the dataset description. A summary of the properties of the dataset may be seen in the table that I have provided below. This system intends to offer intelligent management by using the same workflow architecture that is used in automated offices. The use of data-gathering techniques contributes to the development of the employer's assessment system as well as the enhancement of the school's administrative capabilities. Therefore, students have the opportunity to make use of the findings of the assessment to get a more thorough knowledge of the company, which provides them with other references to choose the firm.

TABLE II. Dataset Description

Title	Description
Number of records	10,000
Number of attributes	15
Data Sources	Time Series Categorization Collection, UC Riverside [27]
Data Type	Time Series, Categorical, Numerical
Time Period	2010-2020
Sampling Frequency	Monthly
Missing Values	2%
Imbalanced Data	Yes (70% majority class, 30% minority class)
Relevant Features	Job Title, Industry, Location, Experience, Salary

For this investigation, we have used scientific approaches to classify, analyze, and evaluate the data that was gathered from a particular socioeconomic event. The investigation of the dependability and quality of this data is one of the methods that is considered to be of crucial importance. An evaluation of the dependability of the information must be carried out first.

A. Correlation Analysis

The correlation analysis method is used to investigate the structural validity of the scale. This is done to ensure that the scale is accurate. This indicates that the validity of the scale is evaluated based on the correlation that exists between the several components that comprise the scale, as well as the correlation that exists between each factor and the total score of the scale. The statistical indicators of the scale correspond with these results, which are congruent with them.

In the correlation results shown in Fig. 2-4, the relationship between employee intention index (q) and feature index (F_i) depends on the specific context of the feature being measured.

1) Positive correlation: If the feature index measures aspects that contribute to employee satisfaction, like work-life balance, compensation, or career development opportunities, then a higher FI might correlate with a higher Emplyment Intention Index (EII) (positive intention to stay). If the feature index measures the alignment between employee skills and job requirements, then a higher FI could indicate a better fit and potentially a higher EII (intention to stay and contribute).

2) Negative correlation: If the feature index measures factors that contribute to employee stress, like long working hours, heavy workload, or lack of resources, then a higher FI might correlate with a lower EII (intention to leave).

3) Neutral relationship: If the feature index measures aspects unrelated to employee retention, like office layout or color scheme, then there might not be a significant correlation with EII.

The specific interpretation of the relationship depends heavily on what the feature index represents. Statistical analysis of employee data with both EII and FI measurements would be necessary to determine the exact nature of the relationship (positive, negative, or neutral). Fig. 2-4 represents the positive, neutral and negative correlation graphs respectively.

Following the completion of the quality standards, we are now in a position to investigate how college students engage in entrepreneurship education. There is a link between the accuracy and feature point value in the subject matter of prediction of the recommendation, as seen in Fig. 5. When the facts are taken into consideration, it would seem that the current notion of entrepreneurship education has to be broadened significantly. A little less than ten per cent of the student population has a genuine interest in gaining knowledge about how to manage a business. In a different way of putting it, a sizeable proportion of students are enthralled by the idea of entrepreneurship education; they are just interested in it. There is a favorable correlation between the length of time that students spend receiving instruction on the subject of entrepreneurship and the level of excitement that they have for learning about entrepreneurship as shown in Fig. 6. At the same time, students who have a greater interest in completing degrees have been receiving more consistent instruction in the field of entrepreneurship. There is evidence to imply that educational institutions such as universities and colleges have not been successful in sufficiently promoting the concept of entrepreneurial education.

Table III presents accuracy results. The Employment Intention Index (EII) is strongly correlated with work happiness, employment stability, and career progression, according to correlation study. This reveals that contented, secure, and advancement-oriented individuals are more inclined to remain with their present employment. EII may be linked to work-life balance since balanced workers are more likely to have long-term career objectives. The feature correlation results are presented in Table IV-VI. The two variables are linked, although not as much as the other categories. This study might assist create an intelligent human resource management system that emphasizes employee retention-related traits.

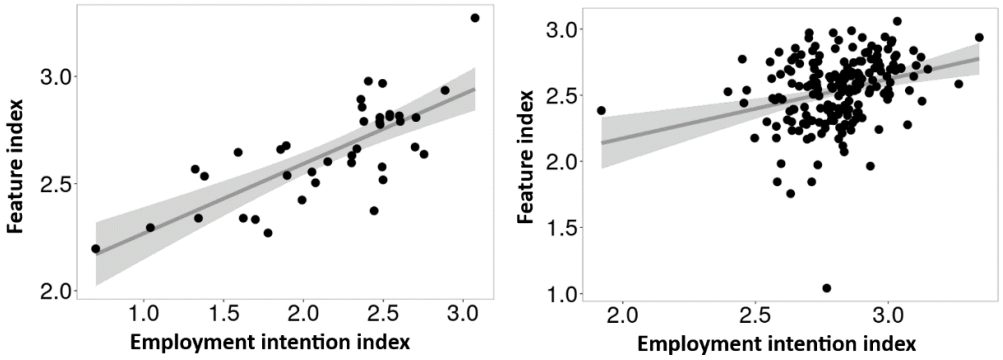


Fig. 2. Analysis with positive correlation: (a) Sample size 150. (b) Sample size 50.

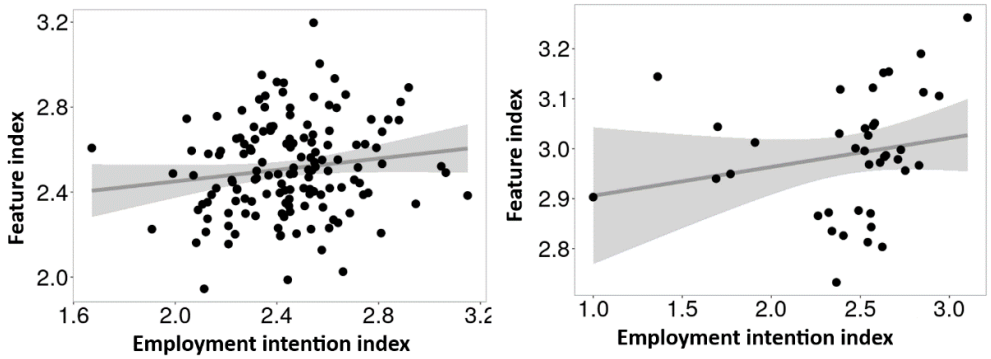


Fig. 3. Analysis with neutral correlation: (a) Sample size 150. (b) Sample size 50.

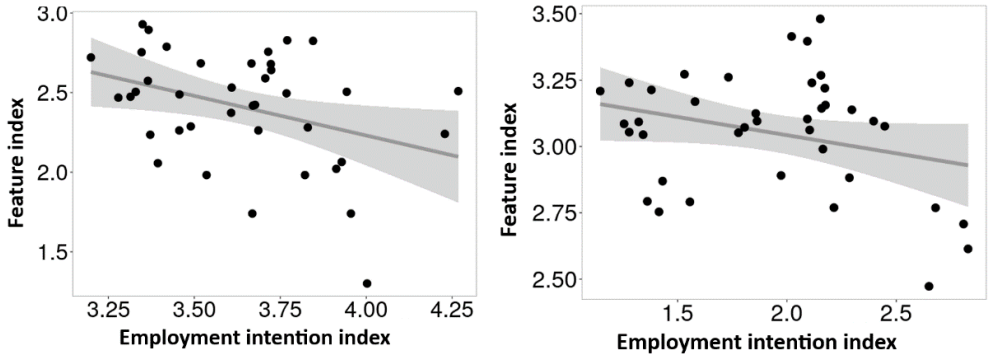


Fig. 4. Analysis with negative correlation: (a) Sample size 150. (b) Sample size 50.

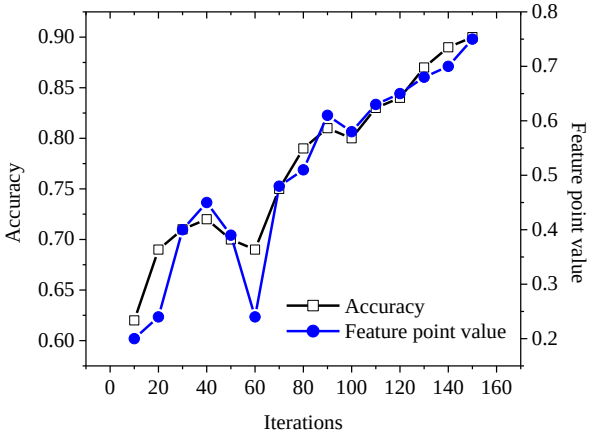


Fig. 5. Plot between accuracy (%) and feature point value for multiple iterations using the recommendation algorithm.

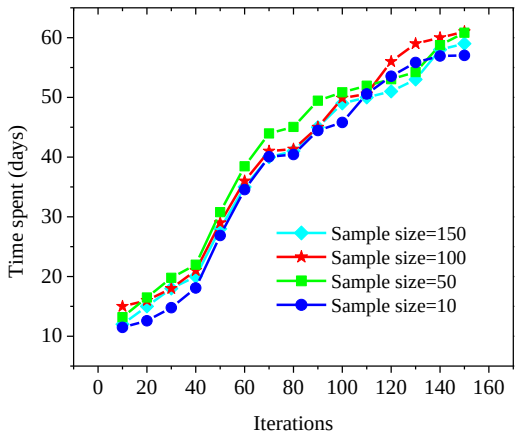


Fig. 6. Time spent for different sample sizes and iterations.

TABLE III. ACCURACY RESULTS

Feature	Accuracy	Precision	Recall	F1-Score	Feature index
Job Satisfaction	0.85	0.83	0.87	0.85	0.23
Job Security	0.82	0.81	0.83	0.82	0.21
Career Advancement	0.88	0.86	0.90	0.88	0.18
Work-Life Balance	0.85	0.84	0.86	0.85	0.15

TABLE IV. FEATURE CORRELATION RESULTS

Correlation results	Job Satisfaction	Job Security	Career Advancement	Work-Life Balance
Job Satisfaction	1.00	0.85	0.78	0.69
Job Security	0.85	1.00	0.83	0.73
Career Advancement	0.78	0.83	1.00	0.81
Work-Life Balance	0.69	0.73	0.81	1.00

TABLE V. CORRELATION ANALYSIS RESULTS

Correlation results	Correlation coefficient	Standard error	t-value	p-value	Confidence Intervals
Job Satisfaction	0.43	0.08	5.38	<0.001	(0.28, 0.58)
Job Security	0.31	0.09	3.44	0.001	(0.14, 0.48)
Career Advancement	0.23	0.23	2.30	0.021	(0.04, 0.42)
Work-Life Balance	0.29	0.20	3.32	<0.001	(0.16, 0.51)

TABLE VI. FEATURE BASED CORRELATION ANALYSIS RESULTS WITH CONFIDENCE INTERVALS

Correlation results	Job Satisfaction	Job Security	Career Advancement	Work-Life Balance
Job Satisfaction	1	0.85(0.79,0.90)	0.78 (0.72, 0.84)	0.69 (0.63, 0.75)
Job Security	0.85 (0.79, 0.90)	1	0.83 (0.77, 0.88)	0.73 (0.67, 0.79)
Career Advancement	0.78 (0.72, 0.84)	0.83 (0.77, 0.88)	1	0.81 (0.75, 0.86)
Work-Life Balance	0.69 (0.63, 0.75)	0.73 (0.67, 0.79)	0.81 (0.75, 0.86)	1

B. Computational Complexity and Time Complexity

By considering the algorithmic factors and implementing appropriate strategies the computational complexity of the proposed ML-based employment management systems is presented in Table VII.

Table VIII presents the time complexity results. Fig. 6 illustrates the amount of time that was allotted for the various sample sizes and iterations. The temporal complexity of the method for machine learning is linear with respect to the number of iterations and the size of the sample. Based on this, it seems that the approach can be expanded to accommodate bigger datasets; however, it may need a substantial amount of processing resources when applied to really large datasets.

The lack of excitement that college students often exhibit when it comes to beginning their enterprises is only one of the numerous factors that contribute to this unfortunate circumstance.

TABLE VII. COMPUTATIONAL COMPLEXITY OF PROPOSED ML-BASED EMPLOYMENT MANAGEMENT SYSTEM

S. no	Computation	Complexity	Details
1.	Data Cleaning	$O(n)$	n is the number of documents (job descriptions)
2.	Data Transformation	$O(n)$	n is the number of documents (job descriptions)
3.	Feature Engineering	$O(n \times d)$	n is the number of documents (job descriptions) and d is the average document length.
4.	Feature Selection	$O(n \times d^2)$	We use, correlation analysis where, n is the number of documents (job descriptions) and d is the average document length.
5.	Model Training	$O(n \times d^2)$ *or* $O(d^3)$	This is for training complexity.
6.	Parameter tuning	O(n)	Most common metrics (accuracy, precision, recall, F1-score) have linear complexity

TABLE VIII. TIME COMPLEXITY RESULTS

Iterations	Time Spent	Training and testing time	Sample size complexity	Iteration complexity
50	0.05	The training time increases linearly with the sample size and number of iterations. The testing time is relatively constant, with a slight increase for larger sample sizes.	$O(n)$ The sample size complexity is O(n), where n is the sample size.	$O(n)$ The iteration complexity is O(n), where n is the number of iterations.
60	0.25			
70	0.50			
80	0.25			
90	1.25			
100	2.50			
110	0.50			
120	2.50			
130	5.00			
140	2.50			
150	12.50			
160	25.00			

In conclusion, the majority of college students do not need an education in entrepreneurship because they do not have a strong sense of purpose, they do not have a strong sense of self-awareness, and they do not believe that having a thorough understanding of one's work is sufficient to manage the problems that they would experience while doing it. For the reasons that were discussed before, they are excluded from the requirement that they get an education in entrepreneurship.

V. DISCUSSION

There is a lack of clarity on the ultimate purpose of incorporating entrepreneurship courses in colleges, which ultimately results in a curriculum that is not relevant to the lives of the students. To fulfil the ever-evolving requirements of students, college entrepreneurship programs had to modify their entrepreneurial curriculum. In addition, these programs must cultivate an entrepreneurial attitude, excitement, and the ability to solve challenges that are encountered in the real world. Because of this, students will be able to develop into well-rounded businesspeople who possess strong character attributes. Additionally, the education curriculum of college and university entrepreneurship programs has failed to address the present mental health of entrepreneurs and the entrepreneurial spirit among today's students. This is a problem since these programs are designed to teach students about entrepreneurship. According to the data, college students who are not yet prepared to deal with the challenges of entrepreneurship are afraid of failing terribly in entrepreneurial endeavors. Many people give up on their dreams of being entrepreneurs because they just do not have the financial means to deal with the challenges and failures that they are certain to encounter. This psychological condition affects the perspectives and attitudes that college students have about business, namely entrepreneurship.

The current state of the labor market is so terrible that more than thirty-one per cent of students who are contemplating starting their own company are doing it as a last choice. Students who are motivated in this manner often strive to establish their firms as rapidly as they can in the expectation of achieving substantial financial gains. In the course of their entrepreneurial endeavors, people may find themselves in a state of confusion when they face obstacles and adverse conditions. Furthermore, according to the findings of a separate poll, more than eighty per cent of students are dissatisfied with the way lessons on entrepreneurship are currently being taught in schools. According to these students, the existing curriculum for entrepreneurship teaches students a great deal of theory but does not evaluate any actual business problems. In contrast, research conducted on student expectations about the growth of entrepreneurship education revealed that more than eighty percent of students had the desire that their school would provide a basis for entrepreneurial activity and provide them with a multitude of possibilities to put what they learn into practice. According to the findings, there is an immediate need to improve the practical teaching system of entrepreneurship education for students who are enrolled in college. In order to offer college students with advice and assistance throughout the process of establishing their own firms, the content of education that is included into entrepreneurship education should be correctly represented in real-world teaching situations [28].

The expansion of entrepreneurial education seems to have any impact on their ambitions for their future careers, according to their subjective perspective. This concept is the outcome of a mixture of many psychological factors that have come together. The education programs generally do not provide a sufficient amount of content and information about business. It is essential to do an analysis of this likelihood. Additionally, the two most common types of education when it comes to entrepreneurship are theoretical education and practical education. In order for students to have a significant influence on their future efforts in entrepreneurship, there must be a sufficient quantity of hands-on experiences [29-31]. This is necessary for entrepreneurship education to have a meaningful impact. The majority of students believe that education in entrepreneurship does not provide much in the way of development for their own personal growth.

Due to the fact that the education that is provided by educational institutions has not been able to successfully integrate theory and practice, and has not fully utilized the complementary role that entrepreneurship education plays, the majority of students have a negative view of entrepreneurship

education. The development of these traits may be accomplished via the careful application of creative skill and imaginative thought. Two of these qualities include thinking that is inclusive and thinking with a focus on the broader picture. Within the realm of entrepreneurship education, there is no doubt that this is the path that will be taken in the future. To summarize, the realm of education should not be the only purview of classroom teaching; rather, theory and practice should collaborate in order to guarantee that this area is able to play the most important role possible.

VI. Conclusion

This paper aims to highlight need for entrepreneurial education to be integrated with education curricula. Up until now, very few people have looked at how entrepreneurial education has developed over time. When thinking about entrepreneurial education, it's important to take into account both internal and external factors. Afterwards, three nearby universities were surveyed for quality assurance purposes, and the findings showed that college students do not fully understand the entrepreneurial classes they take. Incorporating entrepreneurial education into students' thoughts requires drawing on the inclusive and thought-directing qualities of entrepreneurial education. But that isn't all. Some gaps remain. Inadequate theoretical comprehension raises the possibility that views may become somewhat biased. Consequently, it's likely that the study's results aren't comprehensive enough to gain broader recognition of the integration of entrepreneurial education for college students. What's more, there's a need for more research into potential methods to improve both ideological and political education and entrepreneurship education. This leads to an inadequate understanding of the current state of entrepreneurship education in higher education and an inadequate level of analysis. Eventually, more college students will embrace entrepreneurship education as a result of improvements in both content and delivery. The goal of this essay is to look at the practical significance of improving the entrepreneurial education system for college students. This will help produce more high-quality college graduates and increase their chances of finding work after graduation.

HR data may expose biases in hiring, promotions, and other HR decisions. If machine learning models are trained on biased data, they will undoubtedly perpetuate these prejudices. If previous data shows that males are promoted more than females, the model may unfairly favor men in future promotions. The figures may not represent the whole workforce, leading to biased conclusions. If the data is derived from high-performing people, the system may mispredict lower-performing personnel. Methodological flaws in data collection may cause bias. For example, reviewers' biases might influence subjective performance assessments.

Model training and maintenance may need a significant amount of processing resources for big datasets and considerable feature engineering. Deep neural networks may be difficult to grasp. This lack of transparency makes it difficult to explain model decisions and identify biases.

Real-time data streams may be integrated by combining learning platforms, communication tools (email and chat), and staff performance tracking systems. This helps to deliver up-to-date information and personalized guidance. Streaming algorithms examine real-time data and provide timely alerts or actions in order to build real-time prediction models. These models may identify employees who need assistance or who are at risk of leaving. Machine learning for evaluating candidate data, such as social media profiles and applications, has the potential to enhance recruitment technology by enhancing hiring, job suitability, and culture fit. Within the firm, this strategy may help you find mentors, partners, and knowledge networks. Internal employee contacts may be the focus of a future social network research.

Acknowledgments

1) Funding project: 2024 Scientific Research Planning Project of the Education Department of Jilin Province (JJKH20240212JY).

Project name: Research on the obstacles, systems and paths for college students in Jilin Province to stay for employment.

2) Funding project: Research Plan on Huang Yanpei's Vocational Education Thoughts (ZJS2024YB041).

Project name: Strategy Study on Enhancing College Students' Vocational Literacy from the Perspective of Huang Yanpei's Educating People Concept.

3) Funding project: 2024 Higher Education Research Project of Jilin Province (JGJX24C127).

Project name: Research on the integration of mental health education in vocational colleges (middle school, high school, and undergraduate) in Jilin Province.

References

[1] F.Y. Zhang. Thinking on The Information Construction and Standardized Management of University Construction Engineering Archives. Acta Informatica Malaysia. 2023; 7(2): 105-107.

[2] G.M. MD. N. Rabbani, A. Nath, M. H. Emon. Challenges and Opportunities in The Implementation of Big Data Analytics in Management Information Systems in Bangladesh. Acta Informatica Malaysia. 2023; 7(2): 122-130.

[3] F. Sheng. Design and implementation of internship management system based on Web. Journal of Hunan City University (NATURAL SCIENCE EDITION, 2016 (05).

[4] Yirui Song. Optimization of Quantitative Research Methods in Social Sciences in the Era of Big Data. Acta Informatica Malaysia. 2023; 7(2): 92-96. http//:doi.org/ 10.26480/aim.02.2023.92.96

[5] Hu Wenhong, sun Xinxin. Application of data mining technology based on time series in urban waterlogging disaster. Science and Technology Bulletin, 2016,32 (06): 229-231.

[6] Wang Shuxia, Xiong Zenggang. Research on dangerous web data mining technology under massive data interference. Microelectronics and computer, 2016 (02): 87-91.

[7] Xu Kun. Application and development of information management system. Digital communication world, 2018 (05).

[8] Li Yong, Zhang Min, Liu Hao, et al. Research on the current situation and trend of data mining and applied statistics -- academic summary of the 8th annual meeting of the International Association of Data Mining and Applied Statistics. Friends of accounting, 2016 (22): 24-26.

[9] Li Yong, Zhu Jianping, Yang Yunfeng. Theory and Application Research of excellent data sharing statistics: summary of the 8th International Conference on data mining and applied statistics. China Statistics, 2016 (08): 22-25.

[10] Tang Yongfeng. Research on the development of smart grid big data technology. Computer knowledge and technology: academic exchange, 2017, 013 (031): 242-243.

[11] Deng Song, Yue Dong, Zhu Lipeng, et al. Technical framework for intelligent and efficient analysis and mining of power big data. Journal of electronic measurement and instrument, 2016,30 (11): 1679-1686.

[12] Yang Lian Bao, Li Ping, Xu GUI Hong, et al. Research on railway safety management based on big data technology. Railway computer application, 2016 (9): 83-87.

[13] Hu Lijuan, Diao Yinglong, Liu Keji, et al. Reliability analysis of distribution network operation based on big data technology. Power grid technology, 2017 (01): 265-271.

[14] Ren Kai, Deng Wu, Yu Yan. Research on Weblog analysis system based on big data technology. Modern electronic technology, 2016, 39 (02): 39-41.

[15] Zhao Rui. Research on intelligent supply chain model of business big data technology in industry 4.0 era. Business Economics Research, 2018.

[16] Hong Zhu, Research on Human Resource Recommendation Algorithm Based on Machine Learning, Scientific Programming, 2021, https://doi.org/10.1155/2021/8387277.

[17] Ali Raza, et al., Predicting Employee Attrition Using Machine Learning Approaches, Appl. Sci. 2022, 12(13), 6424; https://doi.org/10.3390/app12136424.

[18] Andrieux, P., Johnson, R. D., Sarabadani, J., & Van Slyke, C. (2024). Ethical considerations of generative AI-enabled human resource management. Organizational Dynamics, 53(1), 101032.

[19] Negt, P., Haunschild, A. Exploring the gap between research and practice in human resource management (HRM): a scoping review and agenda for future research. Manag Rev Q (2024). https://doi.org/10.1007/s11301-023-00397-7

[20] Gao Xia, Wu Tao, Gao yueren. Filling and classification algorithm of e-government big data system in cloud computing environment. Electronic Design Engineering, 2020, V.28; no.445 (23): 79-85.

[21] Xiong Wen. Research on Key T echnologies of benchmark test and performance optimization of big data system. 2017.

[22] Xiong Jun, Xia yuan, Yang Yong, et al. Laser active avoidance flight object system based on CCD image intelligent analysis. Infrared and laser engineering, 2016, 45 (0z1): 65-70.

[23] Zeng Yitang, Zhang Yufeng. Research on the application platform of logistics information intelligent analysis based on cloud mining. China circulation economy, 2016, v.30; no.256 (01): 31-36.

[24] Zhang Yufeng, Zeng Yitang. Research on intelligent analysis method of logistics information based on cloud clustering mining. Information work, 2016, No.208 (01): 44-49.

[25] Ju Gengen, Sun Jieping, Chen Li, et al. Design framework of intelligent analysis platform for computer network virtual experiment under big data. Laboratory research and exploration, 2017, v.36; no.262 (12): 113-115.

[26] L. Liu, "Learning information recommendation based on text vector model and support vector machine," Journal of Intelligent and Fuzzy Systems, vol. 40, no. 2, pp. 2445–2455, 2021.

[27] https://www.cs.ucr.edu/~eamonn/time_series_data/

[28] P. Xu, "A probe into the construction of online and offline mixed courses of college students' career planning under the background of first-class curriculum construction," International Journal of Social Science and Education Research, vol. 4, no. 5, pp. 233–239, 2021.

[29] M. Lee and S. Oh, "An information recommendation technique based on influence and activeness of users in social networks," Applied Sciences, vol. 11, no. 6, p. 2530, 2021.

[30] M. Bartkowiak and A. Rutkowska, "Vague expert information/ recommendation in portfolio optimization-an empirical study," Axioms, vol. 9, no. 2, p. 38, 2020.

[31] C. Zhou, J. Shen, Y. Wang, and X. Guo, "Battlefield situation information recommendation based on recall-ranking," Intelligent Automation & Soft Computing, vol. 26, no. 4, pp. 1429–1440, 2020.

Optimizing the Fault Localization Path of Distribution Network UAVs Based on a Cloud-Pipe-Side-End Architecture

Lan Liu*, Ping Qin, Xinqiao Wu, Chenrui Zhang
Digital Power Transmission Grid Department, China Southern Power Grid Digital Grid Technology (Guangdong) Co., Ltd., Guangzhou 510000, China

***Abstract*—The currently proposed optimization algorithm for cooperative fault inspection of distribution network UAVs struggles to accurately detect fault points quickly, leading to low inspection efficiency. To address these issues, we investigate a new fault localization path optimization algorithm for distribution network UAVs based on a cloud-pipe-edge-end architecture. This architecture employs multiple drones for coordinated control, allowing for the simultaneous detection of suspected fault areas. Communication links facilitate interaction at both the drone and system levels, enabling the transmission of fault diagnosis information. Fault defects are identified, and the information is analyzed within an edge computing framework to achieve precise fault localization. Experimental results demonstrate that the proposed algorithm significantly enhances detection speed and accuracy, providing robust technical support for UAV operations.**

***Keywords*—Cloud-pipe-edge-end architecture; distribution network UAV; cloud-edge collaboration; edge computing**

I. Introduction

With the continuous development of high and new technology, UAVs are more and more widely used in various industries. With the support of IoT technology, UAV, as a new type of electric power inspection tool, has the advantages of wide inspection range, high efficiency, low cost, etc., and has been widely used in power systems. Therefore, it is necessary to continuously improve the distribution network UAV inspection and fault location system, optimize the flight path planning and positioning algorithms to adapt to the development needs of the power system, and improve the fault location accuracy and inspection efficiency. At present, the path planning of distribution network UAVs mainly adopts the rule-based method, which is simple and easy to implement but cannot adapt to the complex environment and task requirements. The study in [1] uses YOLOv5 machine vision technology to optimize the planning of unmanned aerial vehicle inspection paths, which requires control and coordination of multiple unmanned aerial vehicles. However, this method does not have a supporting multi-objective unmanned aerial vehicle cluster control technology, making it difficult to achieve multi-path collaborative planning. The study in [2] uses an adaptive multi heuristic ant colony algorithm to improve the problem of missing local path planning for unmanned aerial vehicles. It sets threshold restrictions on pheromones and uses adaptive heuristic function factors to transition the states of all nodes in the inspection path, achieving local path planning. However, when facing complex inspection environments, the overall path planning of this method is easily affected by local factors, resulting in path redundancy and increased time consumption. The study in [3] adopts the fast search random tree RRT algorithm to randomly generate connections through task nodes, analyze the distance between the path and obstacles, and plan the path of multi UAV collaborative tasks, which has certain feasibility in shortening the flight path distance. But the search range of this algorithm is limited, which will have an impact on the running time of the drone. In response to the above shortcomings, this article proposes a drone cluster control technology based on cloud management edge architecture to optimize the planning of fault location paths for distribution network drones. The establishment of cloud management side end architecture for joint control of multiple UAVs, the determination of suspected fault areas through synchronous detection, and the analysis of fault information in the edge computing logic framework can achieve cluster control and optimal path planning for distribution network UAVs, accurate fault diagnosis and location, effectively remedy the shortcomings of multi UAV patrol in complex environments, such as the difficulty of coordination and long control time, and provide new ideas and methods for the application of UAVs in power patrol and other fields.

II. Research Method

A. Distribution Network UAV Inspection Control based on Cloud-Pipe-Side-End Architecture

1) Cloud pipe edge technology architecture: Cloud-managed edge-end architecture is an architecture that combines cloud computing, edge computing, and end devices to realize efficient management and control of IoT devices [4]. In cloud-pipe-edge-end architecture, cloud computing is responsible for storing and processing large amounts of data and providing various applications and services; edge computing is responsible for processing data close to the data source in order to reduce the latency and cost of data transmission; and end devices are responsible for collecting and transmitting data.

Cloud-pipe-side-end architecture is an emerging architecture that provides new ideas and methods for the development of the Internet of Things, and the application of

this technology to the distribution network UAV inspection system has the following advantages:

a) Improve efficiency: By distributing computing and storage resources on different levels, the efficiency and response speed of the system can be improved.

b) Reducing cost: By processing data on edge computing nodes, the delay and cost of data transmission can be reduced.

c) Improve reliability: By distributing computing and storage resources on different levels, the reliability and fault tolerance of the system can be improved [5].

d) Expanding application scope: By combining cloud computing, edge computing, and terminal devices, the application scope of IoT can be expanded.

2) Cloud edge cooperative UAV inspection control system: UAVs play an essential role in power systems in distribution network fault inspection and line monitoring. However, UAV inspection and control systems usually face challenges such as large data volume, high transmission latency, and limited computational resources. In order to solve these problems, the cloud-pipe-side-end architecture is applied to the UAV inspection and control system. The structural framework of the cloud-edge cooperative UAV inspection control system based on the cloud-pipe-edge-end architecture is shown in Fig. 1.

a) Cloud platform: The cloud platform is responsible for storing and processing a large amount of data generated during UAV inspection [6]. It provides robust computation and storage capabilities that enable in-depth analysis and processing of data, as well as a variety of applications and services, such as data visualization, mission planning, and flight path optimization.

b) Edge nodes; Edge nodes are located at the edge of the network, close to UAVs and sensors. They are responsible for processing and analyzing the data collected by UAVs in real time to reduce the delay of data transmission; edge nodes can also perform some critical control tasks, such as flight control, mission assignment, etc.

c) UAV terminal: The UAV terminal includes the UAV itself and various sensors, which are responsible for collecting data and transmitting them to the edge node or cloud platform [7]. The UAV terminal can also receive control commands and perform corresponding tasks.

Through the cloud-pipe-edge architecture, the UAV inspection and control system can realize efficient data processing and transmission. The cloud platform provides powerful computing and storage capabilities, the edge node reduces the delay of data transmission, and the UAV terminal ensures real-time data collection and task execution [8]. It brings higher efficiency, reliability, and scalability for UAV distribution network inspection and fault localization and provides better support for the application of UAVs in various fields. Fig. 2 shows overall structure of the UAV inspection control system.

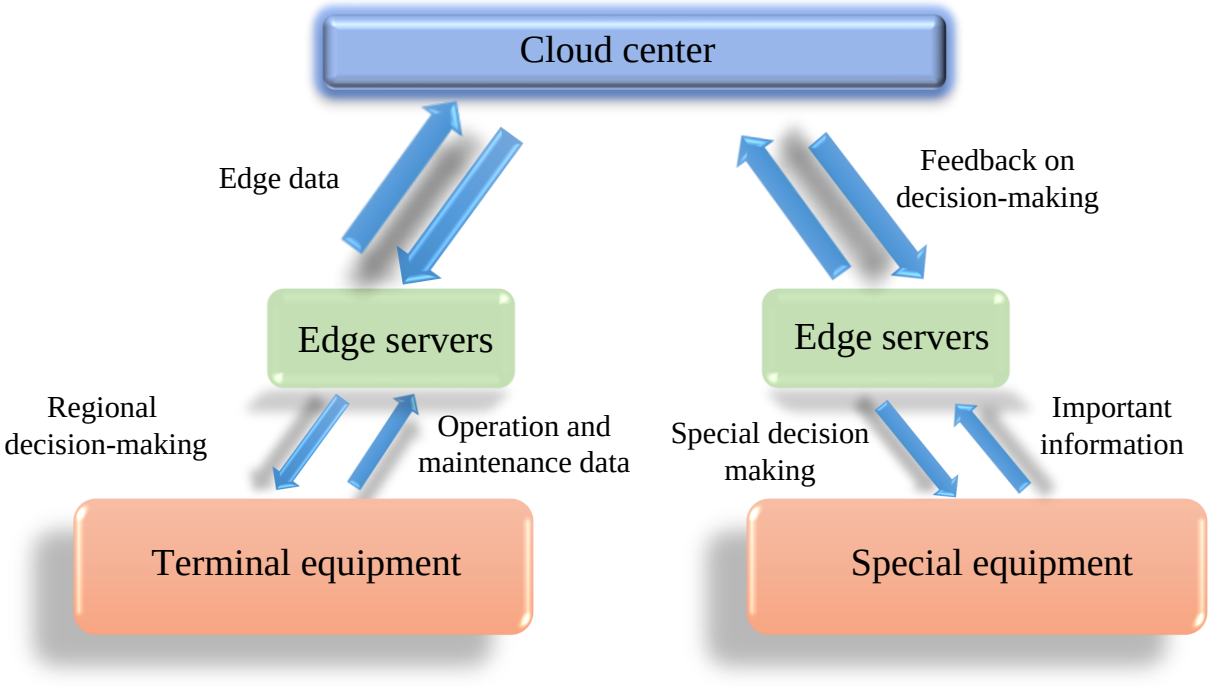


Fig. 1. Technical architecture of cloud-pipe-edge-end architecture.

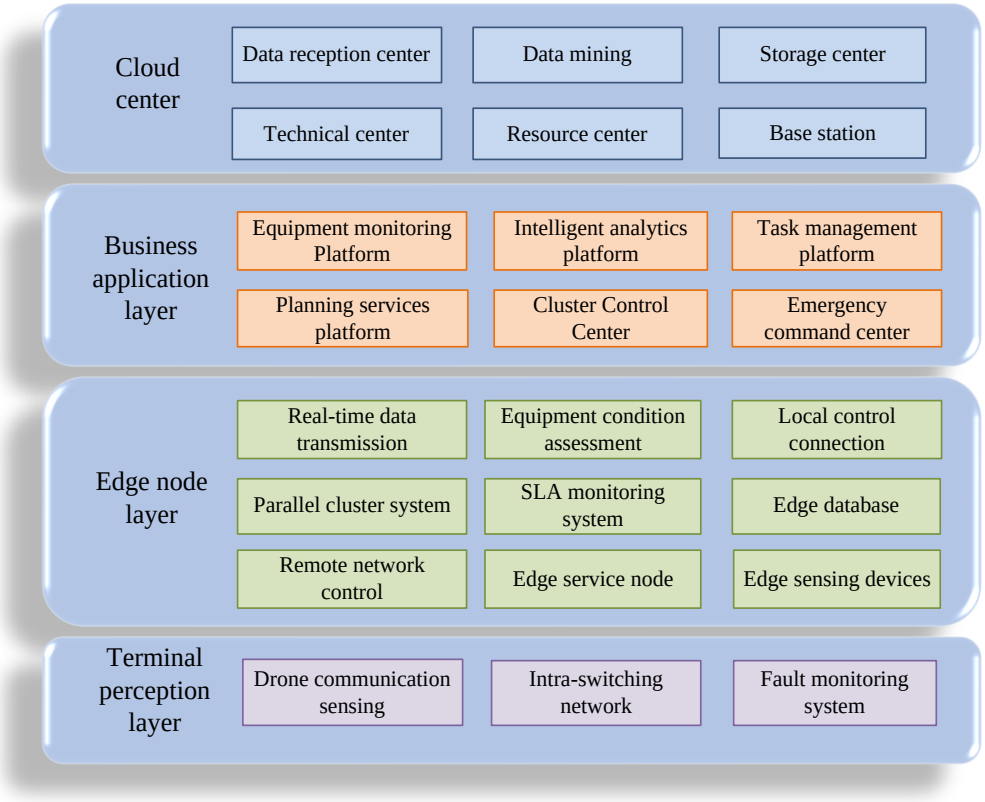


Fig. 2. Overall structure of the UAV inspection control system at the edge of the cloud pipe.

III. Distribution Network Fault Localization UAV Inspection Path Planning

A. Fault Localization based on Edge Computing

In the process of distribution network inspection, the UAV collects distribution network-related localization data, including location information such as lines, equipment, receiving base station, relative distance, etc., by means of the onboard LiDAR sensor. The collected data is transmitted in real time to the grid system server for edge computing [9]. The received edge server data is processed in real time to extract the positional feature information of the edge deployment, and the UAV cognitive edge computing network model is constructed (see Fig. 3).

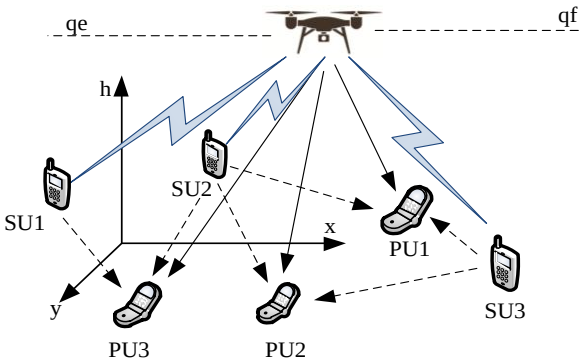


Fig. 3. UAV cognitive edge computing network model.

All servers, distribution network nodes and user terminal location information is imported into the network model, and the UAV inspection channel model is constructed using a three-dimensional Cartesian coordinate system [10]. Assuming that the distribution network node coordinates are represented as (x, y, z), and setting the fixed elevation of the UAV inspection flight as h, the node coordinate channel gain coordinate transformation under the time slot sequence is described as:

$$G_t(x, y, z) = \frac{\Delta g}{h_t^2 + q[y_t] - x_t^2} \tag{1}$$

In the formula, $G_t(x, y, z)$ represents the distribution network node coordinate time slot channel gain result, and Δg is the channel gain amplitude per unit distance. The network model is used to monitor the dynamic information of the UAV inspection and to monitor the faults of the distribution network equipment and lines, and when an abnormal signal location is identified, the fault point is quickly located and marked by comparing the data in the normal state with the currently collected data, and the signal is transmitted to the edge server at the nearest distance [11]. Carrier linear fuzzy adjustment based on real-time UAV position and edge server distance:

$$\begin{bmatrix} N_{0,0} \\ \vdots \\ N_{n-1,n} \end{bmatrix} = \begin{bmatrix} \Delta X_{0,0} & \Delta Y_{0,n} \\ \vdots & \vdots \\ \Delta X_{n-1,n} & \Delta Y_{n-1,n} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \tag{2}$$

In the formula, N is the horizontal distance carrier linear correction parameter of the time series data, and ΔX , ΔY are the original coordinate signals of the fault localization point. The distance between the corrected coordinate parameters and the edge server is calculated using the least squares method L_n:

$$\left(x_n - x_{n-1}\right)^2 + \left(y_n - y_{n-1}\right)^2 + \left(z_n - z_{n-1}\right)^2 = L_n^2 \tag{3}$$

The fault localization results are fed back to the UAV inspection control system through edge computing, and the system carries out optimal path planning according to the fault localization information and manipulates the UAV to carry out collaborative inspection and maintenance operations [12].

B. UAV Inspection Path Planning Algorithm

The inspection path power model is constructed according to the UAV flight law, and the pitch angle, yaw angle, and roll angle are calculated during the UAV flight:

$$\begin{cases} \theta = \left[\left(V_y - V_z\right)\phi\varphi + v_x F\right] / V_x \\ \varphi = \left[\left(V_z - V_x\right)\theta\phi + v_y F\right] / V_y \\ \phi = \left[\left(V_x - V_y\right)\theta\varphi + v_z F\right] / V_z \end{cases} \tag{4}$$

In the formula, θ , φ , ϕ represent the pitch, roll and yaw angles of the UAV during flight, V_x , V_y , V_z are the flight acceleration in the horizontal and vertical axes, respectively, v_x , v_y , v_z are the rotor angular velocities corresponding to the sailing directions, and F is the coefficient of the UAV flight moment [13]. Through the Laplace transform, the UAV pitch angle, roll angle, and yaw angle (see Fig. 4) attitude mechanics function matrices are obtained:

$$\begin{cases} \theta(t) / F(t) = \bar{h} / V_y t^2 \\ \varphi(t) / F(t) = \bar{h} / V_z t^2 \\ \phi(t) / F(t) = \bar{h} / V_x t^2 \end{cases} \tag{5}$$

In the formula, t is the UAV timing node parameter, $\bar{h}$ is the ideal altitude for UAV hovering.

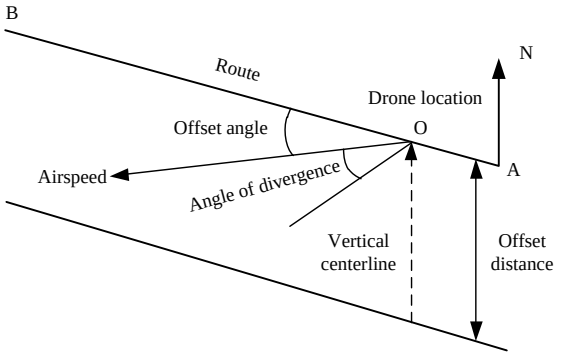


Fig. 4. Schematic of yaw angle attitude mechanics.

The unit angle error correction factor e_{ref} is introduced to correct the error for flight altitude at different attitude angles:

$$\begin{cases} \Delta h_\theta = h\left(\theta_{ref} - \theta\right) \cdot e_{ref} \\ \Delta h_\varphi = h\left(\varphi_{ref} - \varphi\right) \cdot e_{ref} \\ \Delta h_\phi = h\left(\phi_{ref} - \phi\right) \cdot e_{ref} \end{cases} \tag{6}$$

Based on the corrected UAV flight dynamics parameters, a PID tracking controller is utilized for path planning of the inspection nodes:

$$Line_n = \sqrt{\left(free + \alpha L_{on_1}\right)^2 + \left(l_{safe} + L_{on_n}\right)^2} \tag{7}$$

In the formula, $free$ indicates the maximum degree of freedom of the UAV flight path, α is the inspection path safety coefficient, l_{safe} is the inspection safety distance for fault localization, and L_{on_1} , L_{on_n} are the distances from the initial position to the inspection node, respectively. According to the formula, the UAV inspection path can be planned to carry out precise fault point search inspection for distribution network tripping, disasters and other suspected fault segments, with high real-time, high accuracy and fast response, which can improve the operation and maintenance efficiency and reliability of the distribution network [14].

IV. UAV Path Optimization Algorithm based on Cloud Pipe Edge Ends

A. UAV Cluster Control Objective Function

In order to carry out distribution network fault detection more efficiently, multi-UAV cluster control will be used for

collaborative inspection, so multi-objective optimization is needed on the basis of the original path planning. Firstly, the shortest distance function between multi-path nodes is obtained by edge computing as:

$$\min L_n = \sum_{n}^{i=1}\left(ql_n + ql_n'\right) + f\left(e_n\right) \tag{8}$$

In the formula, where l_n, l_n' are the straight line length and the actual effective length of the n node distance, q is the distance minimization weighting factor, and $f\left(e_n\right)$ is the error penalty function [15]. The nearest path matching objective function that removes the overlapping area of the node space is obtained by constraining the node in and out of the path edge points:

$$\sum_{n}^{i=0} L_n^i = 0\,|\,1, \forall i = 1, 2\cdots, N, \forall n = 1, 2\cdots, n \tag{9}$$

The directed band-weighted relationship between the number of UAV cluster controls and the total distance of the inspection path can be described as:

$$\min L_{all} = \sum_{N}^{n}\sum_{n}^{i=0}\sum_{n}^{j=0} l_{on}^i h_{nt} \tag{10}$$

The shortest distance between the monitoring points is calculated to match the neighboring nodes in the inspection path, and the directed weighted path (Fig. 5) is obtained schematically as follows:

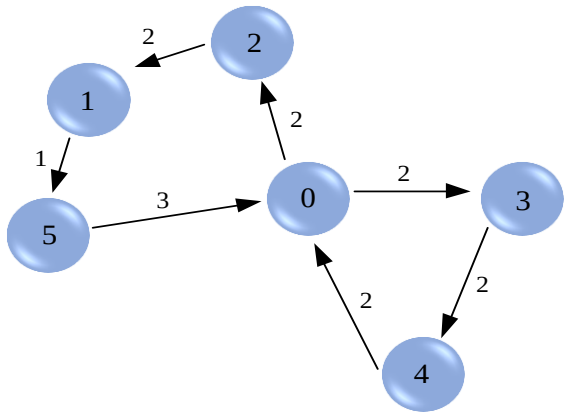


Fig. 5. Schematic diagram of directed weighted paths.

Based on the UAV speed and path distance parameters, the total UAV cluster inspection route time can be predicted:

$$\min T_{all} = \max\left\{\sum_{n\le N}^{i=1}\max\left\{t_i^n \times \left(t_i - e_i\right), 0\right\}\right\} \tag{11}$$

In the formula, where t_i^n is the UAV inspection time per unit distance for line n, and e_i is the time error parameter of the path [16]. The server monitors the flight paths and times of multiple UAVs, and collaboratively monitors the entire UAV cluster through centralized control to achieve efficient inspection tasks.

B. Multi-objective Function Constraints

For the UAV range, energy consumption, and load-carrying performance, further optimal efficiency constraints need to be imposed on the multi-objective UAV path[17]. Assuming that all UAVs take the distribution network base control center as the departure starting point, the maximum number of cooperative control UAVs is:

$$\max N = \sum_{n}^{i=1} l_i^n, i = 1, 2\cdots, n \tag{12}$$

Distribution network UAV inspection needs to be conFig.d with relevant sensing and communication devices to collect information on distribution machinery lines and other equipment, as well as monitor and feedback fault problems [18]. The UAV load will affect its flight speed and elevation threshold; in order to achieve optimal path planning, the total amount of UAV load needs to be constrained:

$$\max M = \sum_{n}^{i=0} v_i^n \times m_i, i \in 1, 2\cdots, n \tag{13}$$

In the formula, where v_i^n is the expected optimal speed of path inspection, m_i is the UAV unit distance supportable carrying capacity threshold. According to the UAV endurance performance and inspection path length, the maximum working time of a single UAV is constrained and controlled by considering the UAV startup operation and waiting service time t_i:

$$\max T = \sum_{n}^{i=0} l_i^n v_i^{-n} + \sum_{n}^{i=1} l_i^n t_i \tag{14}$$

For the fault localization process, the UAV path has a dynamic nature, and the particle velocity and position update formulation of the adjacency matrix of the inspection node[19]. Setting the dimension particle of node position as W_n, the extended edge set of particles in the adjacency matrix is $n_k^l = \left[n_k^{l_1}, \cdots, n_k^{l_n}\right]$:

$$n_k^{l_i} = \left[\left\langle n_i, W_n\right\rangle, \left\langle W_n, n_j\right\rangle\right] \otimes k_W \tag{15}$$

The extended edge set data of each node particle can form a Hamiltonian circle in the actual position, and in the Hamiltonian circle constraints, the UAV path can be dimensionally adjusted, and the probabilistic constraints can be imposed on the edge flight speeds of the path nodes in the case where the velocity component v_k^l is.

$$P\left(v_k^l\right) = \left\{\frac{n_i, n_j}{P\left(n_i, n_j\right)}\right\} \tag{16}$$

After the above multi-objective function constraints, the multi-UAV cluster control capability can be strengthened to promote the distribution network fault inspection and localization efficiency [20], reduce the work intensity and error rate, and realize high-precision and high-efficiency fault inspection so as to guarantee safe and stable operation of the power system.

V. Discussion

In order to test the effectiveness of the research method, a comparative experiment is designed to simulate the UAV distribution network fault inspection process and compare and analyze the experimental result data. The distribution network of a power system is taken as the inspection object, and the inspection nodes are set according to the location of the distribution network; the server model of the control center is PC i5-8250U CPU, and the inspection node data are numbered by using VRPTW. Select and adjust the UAV cluster parameters as shown in Table I.

TABLE I. UAV Cluster Test Parameter Settings

Targets	Parameters
UAV SYSID	2、3、5
Number of nodes	20
Path type	zigzag path
UAV2 point set coordinates	HOME (121.441 121 6,31.028 408 30) TARGET1(121.441 269 2,31.028 381 84,10) TARGET2(121.441 308 2,31.028 426 12,10)
UAV3 point set coordinates	HOME (121.441 167 1,31.028 364 30,10) TARGET1(121.441 197 3,31.028 611 22,17) TARGET2(121.441 009 2,31.028 666 81,17) TARGET3(121.441 988 3,31.028 800 62,17)

The operating drone control system automatically formulates the drone inspection plan according to the calculation results of the optimization algorithm, planning the number, starting point, flight path, and other contents of the cluster inspection drone. Adopting vehicle navigation, planning the path navigation of the vehicle to the take-off point of the faulty section, guiding the team members to the survey location to carry out the machine patrol operation, navigating and guiding the team to the landing point of the drone to recover the aircraft after the drone takes off, and real-time according to the faulty search and patrol and localization process of the collection of data situation, and further make the drone response strategy.

A. UAV Fault Location Inspection Time Analysis

This experiment is based on edge computing algorithm support, combined with multi-objective function constraints for path planning and task control of UAV clusters. All drones take the point HOME (121.441 121 6,31.028 408 30) as the starting point for inspection, and the lightweight defect recognition algorithm is deployed at the edge end of the drones to identify the defective faults of the distribution network in real-time and troubleshoot them on site. Through the real-time linkage function, the UAV is maneuvered to return to the flight, upload the fault information, and complete this fault-finding task. The average inspection time of each fault node of the UAV under the support of different algorithms is monitored and counted, and the time parameters of one group, three groups, and five groups of UAV cooperative control are recorded, respectively.

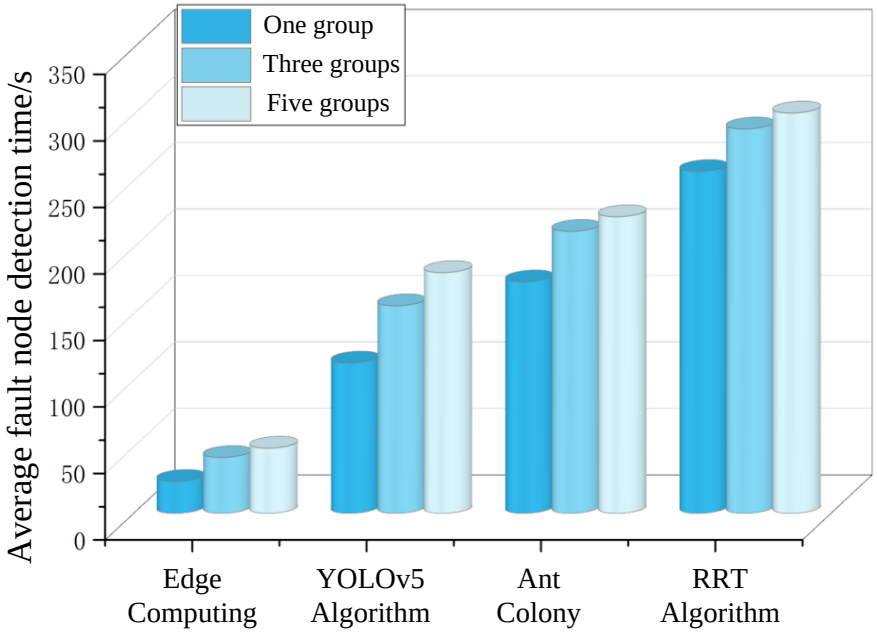


Fig. 6. UAV fault localization inspection time.

From the data in Fig. 6, it can be clearly seen that when controlling a group of drones, the cloud edge collaboration method has a shorter working time for the drones, with an average inspection time of only 24 seconds for a single node. The response time for single drone inspections using the other three traditional methods, such as YOLOv5, exceeds 100 seconds. At the same time, as the number of drones in collaborative control increases, the average inspection time per unit drone has also increased, and the fault detection time is directly proportional to the number of drones. When the number of UAV test groups is 5, the average detection time based on edge computing control is 49 seconds, while the RRT forest algorithm is limited by the search range, and the response speed is slow, and the unit time is up to 301 seconds. This shows that the edge computing method used in this paper for cloud edge collaborative control of UAVs is faster, and the average time for distribution network fault patrol is far lower than the traditional method.

B. UAV Flight Attitude Stability Analysis

According to the UAV flight path dynamics model, the pitch angle, roll angle, and yaw angle of UAV flight will affect the course direction and cause the optimal path offset to a certain extent. Therefore, the UAV attitude offset angle is monitored during the experiment, and the jitter angle range of different flight attitudes of each group of UAVs is counted to analyze the UAV flight stability.

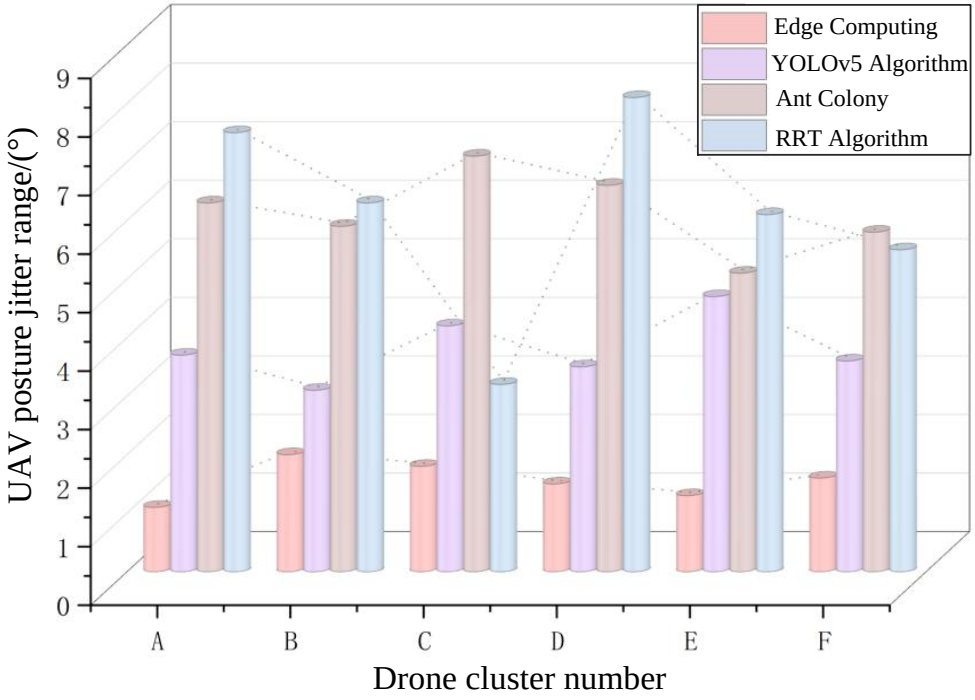
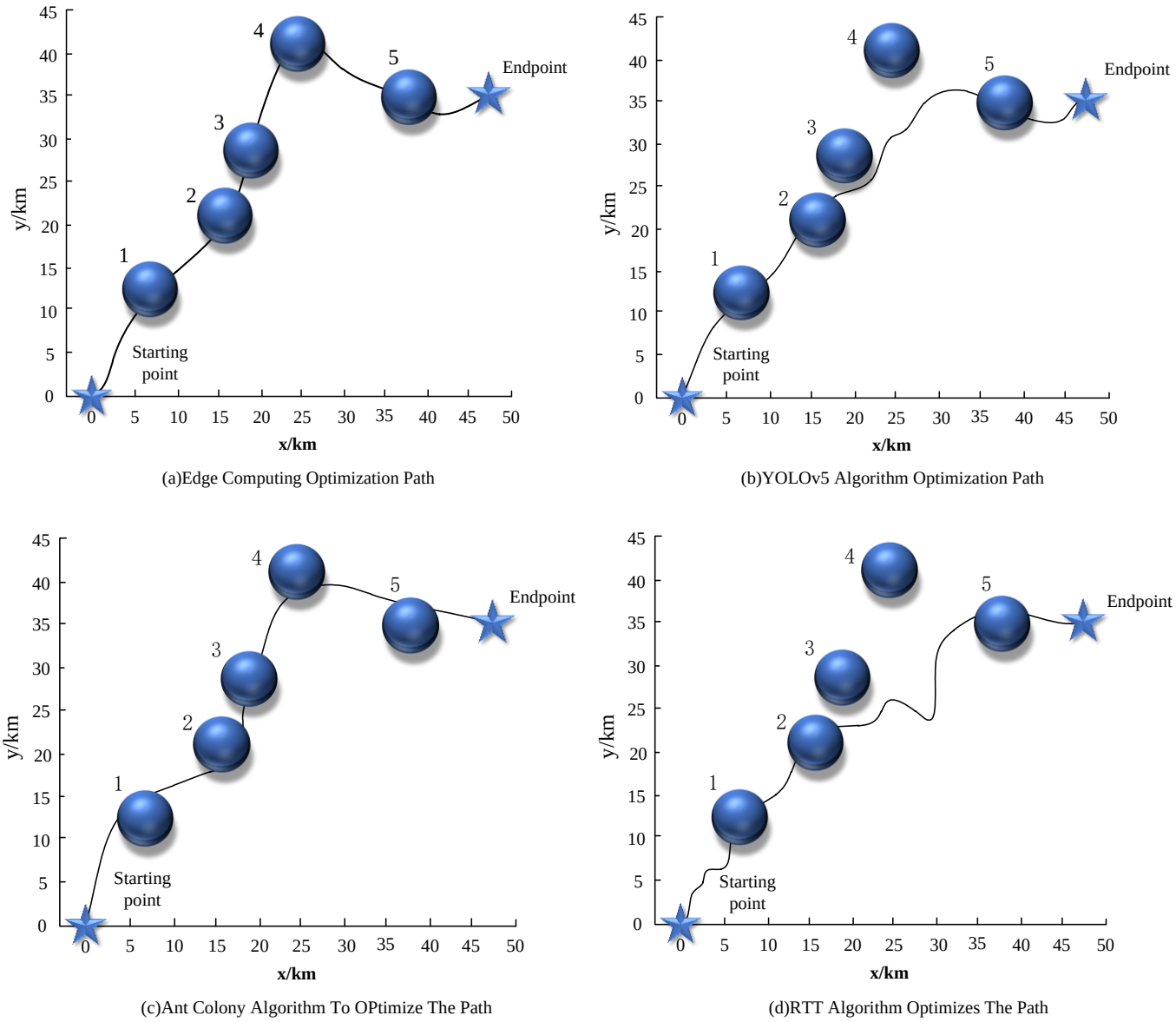


Fig. 7. UAV flight attitude jitter range.

As shown in the Fig. 7, this paper is based on edge computing to construct the power model of the UAV inspection path. The UAV flight angle attitude control is more accurate, and the jitter range of each group's UAV posture is no more than 3°, which has good stability. The control system based on the adaptive ant colony algorithm is less stable. The jitter range of all six groups of UAVs is more than 5°, while the jitter situation of UAVs using RRT forest algorithm path planning has great uncertainty. The jitter offset angle reaches a maximum of 8.1°, and the minimum angle deviation is only 3.2°. In the comprehensive analysis, using the edge computing method for path planning, fully considering the UAV yaw and jitter angle problems, and using power law for path optimization effectively improves the UAV inspection attitude stability.

C. Drone Inspection Path Planning Analysis

Using MATLAB software to simulate the fault location inspection path of the distribution network UAV, a set of path node position coordinates are extracted and imported into the system program, and four groups of UAV flight real-time path coordinate parameters are substituted to obtain the training path visualization image as follows.



(a)Edge Computing Optimization Path

(b)YOLOv5 Algorithm Optimization Path

(c)Ant Colony Algorithm To OPtimize The Path

(d)RTT Algorithm Optimizes The Path

Fig. 8. Comparison of UAV inspection paths.

As shown in Fig. 8, five fault detection points are set in the selected path. The optimized patrol path based on edge computing completely covers five fault points, almost passing through the joint point of the fault center, and there is no redundant path between fault nodes. The YOLOv5 algorithm lacks supporting multi-objective unmanned aerial vehicle cluster control technology, which makes it impossible to achieve collaborative control of multiple unmanned aerial vehicle targets, resulting in path planning missing fault detection points 3 and 4, which does not meet the expected planning and inspection requirements; The path optimization effect of ant colony algorithm is relatively good, achieving fault localization for 5 monitoring points. However, the detection synchronization is poor, and there is a certain lag in information, resulting in a deviation in the inspection path of fault points 1 and 2, only reaching the edge range of the fault point, and there is a certain error in the positioning data; RRT path planning also missed the distant detection points 3 and 4, and did not reach the center range of other fault monitoring points, resulting in low accuracy of fault localization and detection information.

In summary, the fault localization path optimization algorithm for distribution network UAVs with cloud-pipe-edge-end architecture studied in this paper has good application performance, the coordinate information of fault localization is more accurate, and the power model is used to improve the stability of the UAV flight attitude. With the support of cloud-edge cooperative technology, we can better realize multi-objective UAV cluster control, edge computing optimization path coverage is more comprehensive, the unit interval distance is shorter, and it is more advantageous than

the traditional method in the distribution network fault inspection and positioning work.

VI. CONCLUSION

Aiming at the deficiencies in UAV cluster control and path planning, this paper proposes a fault location path optimization algorithm for distribution network UAVs based on cloud-pipe-edge-end architecture. It obtains the following conclusions through theoretical and experimental research:

Based on the cloud-pipe-edge-end technology architecture, the UAV is controlled by cloud-edge cooperative control and interacts with edge nodes to improve the data processing and response rate.

Based on edge computing to process UAV inspection and fault localization data, carrier linear fuzzy adjustment is carried out through the relative distance of edge servers to improve the accuracy of coordinate localization parameters.

Construct a UAV flight dynamics model, analyze UAV posture angle and flight law, and improve the real-time control performance of the path planning system.

Carry out multi-objective path optimization for UAV clusters and constrain the load time and expansion dynamics of the number of UAVs to reduce the error rate of inspection path deviation and obtain the optimal path planning for inspection with high precision and high efficiency.

Through experiments, it is proved that the method studied in this paper has good stability and application efficiency, and more secure and efficient algorithms and technologies will be further explored in future research to improve the accuracy and efficiency of fault location path planning for distribution network UAVs, so as to make a more significant contribution to the intelligent development of the work of UAV inspection and fault detection of distribution network.

REFERENCES

[1] Wang Yunbing; Fu Xiaogang; Niu Yuan. Route optimization and fault detection based on UAV photovoltaic inspection. Journal of Shanghai Dianji University, 2023(005):275-280.

[2] Yin Yanan; Zhen Ran; Wu Xiaojing; Zhang Chunyue; Wu Xueli. Research on UAV route planning based on adaptive multi heuristic ant colony algorithm. Journal of Hebei University of Science and Technology, 2021(001):038-047.

[3] Chen Jin-Tao; LI Hong-Yi; Ren Hong-Ru; Lu Ren-Quan. Cooperative Indoor Path Planning of Multi-UAVs for High-rise Fire Fighting Based on RRT-forest Algorithm[J]. Acta Automatica Sinica, 2023, 49(12):2615-2626.

[4] Tian Xiao-Zhuang; Li Song; Fu Guo-ping; Tan Qi-yun; Shan De-shuai; Wang Wei-guang; WANG Zhu. Time-optimal Obstacle Avoidance Path Planning for UAV Inspection. Computer and Modernization, 2023(3):43-47.

[5] Kong Weili; Wang Feng; Zhou Pinghua; Wang Hefeng. Three-Dimensional Path Planning of UAVs Based on Improved Ant Colony Algorithm. Electronics Optics & Control, 2023, 30(3):63-69.

[6] Han Ren, Gao Yujie Zhang Sheng. Research on Path Planning in Unmanned Aerial Vehicle-Assisted Edge Computing Environments. Modeling and Simulation, 2023, 12(6):5949-5958.

[7] Huang Hui; Nian Fugeng. Research on Multi-UAV Collaborative Reconnaissance Path Planning. Ship Electronic Engineering, 2021,(9):58-61.

[8] Liu T, Sun Y, Wang C, et al. Unmanned aerial vehicle and artificial intelligence revolutionizing efficient and precision sustainable forest management. Journal of Cleaner Production, 2021, 311:127-146.

[9] Li Q, Xu Y. Intelligent Early Warning Method Based on Drone Inspection. Journal of Uncertain Systems, 2022, (10):1142-1149.

[10] Jeong E, Seo J, Wacker J. Grayscale Drone Inspection Image Enhancement Framework for Advanced Bridge Defect Measurement. Transportation Research Record, 2021, 2675(8):603-612.

[11] Aliyari M, Ashrafi B, Ayele Y Z. Drone-Based Bridge Inspection in Harsh Operating Environment: Risks and Safeguards. International Journal of Transport Development and Integration, 2021(2):118-135.

[12] Zhang N, Zhang M, Low K H. 3D path planning and real-time collision resolution of multirotor drone operations in complex urban low-altitude airspace. Transportation Research Part C Emerging Technologies, 2021, 129:103-123.

[13] Ganesan R G, Kappagoda S, Loianno G, et al. Comparative Analysis of Agent-Oriented Task Assignment and Path Planning Algorithms Applied to Drone Swarms. 2021,101:051-061.

[14] Wu Y, Low K H, Pang B, et al. Swarm-Based 4D Path Planning For Drone Operations in Urban Environments. IEEE Transactions on Vehicular Technology, 2021(8):309-331.

[15] Hong D, Lee S, Cho Y H, et al. Energy-Efficient Online Path Planning of Multiple Drones Using Reinforcement Learning. IEEE Transactions on Vehicular Technology, 2021(10):9725-9740.

[16] Luo Y, Ding W, Zhang B. Optimization of Task Scheduling and Dynamic Service Strategy for Multi-UAV-Enabled Mobile-Edge Computing System. IEEE Transactions on Cognitive Communications and Networking, 2021,9 (03):970-984.

[17] Wang J, Ke H, Liu X, et al. Optimization for computational offloading in multi-access edge computing: A deep reinforcement learning scheme. Computer Networks, 2022, 204(26):1-17.

[18] Luo G C Z. Adaptive entropy theory polymerization method for path optimization in edge computing educational systems. Journal of intelligent & fuzzy systems: Applications in Engineering and Technology, 2021, 40(2):2941-2951.

[19] Zuo B, Xu Y, Yang D, et al. Joint resource optimization and trajectory design for energy minimization in UAV-assisted mobile-edge computing systems. Computer communications, 2023,203:312-323.

[20] Cui E, Yang D, Zhang H, et al. Improving Power Stability of Energy Harvesting Devices with Edge Computing-Assisted Time Fair Energy Allocation IEEE Transactions on Green Communications and Networking, 2021,1(5):540-551.

Predicting the Number of Video Game Players on the Steam Platform Using Machine Learning and Time Lagged Features

Gregorius Henry Wirawan, Gede Putra Kusuma
Computer Science Department-BINUS Graduate Program-Master of Computer Science,
Bina Nusantara University, Jakarta 11480, Indonesia

***Abstract*—Predicting player count can provide game developers with valuable insights into players' behavior and trends on the game population, helping with strategic decision-making. Therefore, it is important for the prediction to be as accurate as possible. Using the game's metadata can help with predicting accuracy, but they stay the same most of the time and do not have enough temporal context. This study explores the use of machine learning with lagged features on top of using metadata and aims to improve accuracy in predicting daily player count, using data from top 100 games from Steam, one of the biggest game distribution platforms. Several combinations of feature selection methods and machine learning models were tested to find which one has the best performance. Experiments on a dataset from multiple games show that Random Forest model combined with Pearson's Correlation Feature Selection gives the best result, with R^2 score of 0.9943, average R^2 score above 0.9 across all combinations.**

Keywords*—Video games; regression method; feature selection; time series forecasting; machine learning

I. INTRODUCTION

The video game industry has seen a massive growth over the past few years, especially during the COVID-19 pandemic, when people were encouraged to stay at home, increasing gaming activity. The increased gaming activity was due to either stress relief [1], seeking social interactions [2], or having no other activity to do at home [3]. Steam is one of the rapidly growing game distribution platforms and makes a major contribution in the growth of the industry, along with the transition from physical distribution of games to digital distribution in the form of licenses, and along with online social networking services for gamers [4].

With more than 50,000 video games available from various developers and publishers, Steam gives gamers the liberty to buy and play their favorite video games and share their captured moments with their friends on the platform. Gamers can also leave their impressions and share their opinions about the game they play through a review on the game's store page [5]. All the things happening on the platform will certainly generate some data. Fortunately, Steam allows access to said data by providing various Application Programming Interfaces (APIs), allowing publishers and researchers alike to generate their own sets of data and gain meaningful insights [6]. As a result, a lot of studies on the platform have appeared over the past few years from various disciplines, including consumer behavior [7], human-computer interaction [8], economics [9], education [10], health [11], social sciences [12], law [13], business [14], and gaming specific engagement [15].

While publishers have access to proprietary APIs that provide sensitive, non-public user data, general APIs still allow access to publicly available data on platforms, including game prices, genres, reviews (game-specific data), as well as player activity, game ownership, and achievement statistics (public user data) [16]. This can be useful for business applications through data-driven analysis and can also be used in machine learning algorithms to predict various variables, such as game prices, discount trends, player count, game ratings, and many others.

Predicting daily player count is essential for game developers and publishers, as it provides insights into player activity, engagement, and behavior, enabling informed decision-making. Therefore, it is important for the prediction model to be as accurate as possible to avoid any potential misinformation. The features used play an important role in the model's predictive performance and must provide meaningful information for the model. Existing studies on predicting daily player count used various metadata features in addition to historical data, such as game's genres, supported languages, number of achievements, etc. However, those features stay the same most of the time, and don't give enough temporal context for the model, as historical data like daily player count can have patterns, such as trends and seasonality depending on the time of the observation. But the historical data can be transformed into lagged features using feature engineering to capture such historical patterns and enhance the prediction performance of the model. However, there is a limited number of studies in exploring the use of time-lagged features for predicting player count.

In this paper, we proposed a new method utilizing lagged features on top of existing metadata features to accurately predict daily player count on Steam. Historical player count data were transformed into new lagged features using the sliding window technique, providing more temporal context than using meta-data only. This will be explained further in Section III. Several combinations of feature selection methods and machine learning models were tested and compared on their predictive performance to find out which one gives the most accurate prediction. The models are Random Forest, Support Vector Regression (SVR), and XGBoost, with feature

selection methods such as Pearson's Correlation feature selection, Recursive Feature Elimination (RFE), and the models' embedded feature selection method. The structure of the paper is as follows: Section I presents the background of this study, Section II discusses the works related to this study, Section III explains the methodology, Section IV the results and discussion, and Section V contains the conclusion of this study and things to be addressed for future works.

II. Related Works

As stated in Section I, there have been a lot of studies on the Steam platform in the past few years. A study by Prathama et al. [6] created a system to provide data analysis on current game trends and predict game trends for the next two weeks using game and user data obtained with Steam API. The trend prediction is through predicting future game rating and future player count, using Multiple Linear Regression (MLR) method. H. Zhang [16] used the same MLR method to predict game sales, and also investigated various factors and how they are related with game sales. Zendle et al. [7] analyzed trends on ingame microtransactions using historical data from 463 most-played games in the Steam platform.

Wannigamage et al. [17] analyzed the changes in player population and weekly player count patterns during the COVID-19 pandemic, and also analyzed the changes in game sales. They also tried to identify which games that became popular during the pandemic by comparing player population from before and during the pandemic. Vuorre et al. [2] also analyzed the changes in players behavior during COVID-19 pandemic using various data from popular games, like play time and player count from both before the pandemic and during the pandemic. Wu et al. [18] conducted an analysis on the impact of the COVID-19 pandemic on the video game industry overall, by analyzing and comparing the number of games released and also player count from before the pandemic and during the pandemic, and also predicted the demand for online games with machine learning, using historical player data combined with COVID-19 features and human mobility features to predict daily player count. Several machine learning models were used, including SVR, Random Forest, and Ridge Regression.

Varghese et al. [19] discuss an online game's success upon release using the game's historical player data on the Steam platform with models including SVR, Random Forest, and Bayesian Regression. Teja et al. [20] compared various machine learning algorithms and predicted the rating from Metacritic for games on Steam by comparing variables that are related to the score, like genres and player count. Abdul-Rahman et al. [21] developed a model for churn prediction using Vector Autoregression enhanced with sentiment analysis on user reviews from various games on the Steam platform.

III. Methodology

This section explains the data and methods used in the study. The workflow is depicted in Fig. 1. The data was gathered from various public sources. The gathered data was then pre-processed to get it ready for machine learning models. New lagged features were created using the sliding window technique, and several feature selection methods were selected to select the most relevant features: Pearson's Correlation feature selection, RFE, and embedded method. The selected features from each method were used in three machine learning models: Random Forest, SVR, and XGBoost. Grid Search hyperparameter tuning is used to find the best parameters for each model. The results from each combination were then compared to find out which one is the best in predictive performance.

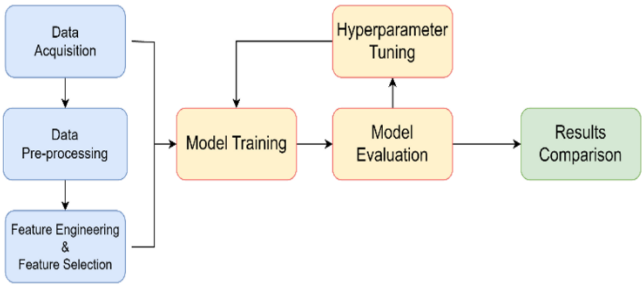


Fig. 1. Workflow of the study.

A. Data Acquisition

This study uses publicly available data from various sources. The historical dataset was collected through manual download from SteamDB, a third-party database providing information on games from the Steam platform. The historical dataset consisted of player count, number of positive and negative reviews from 100 most-played games on Steam, sorted by peak number of players during the date of collection, 8 August 2024. All observations were in UTC time zone. Then a code was made for scraping metadata from each game. The code was written in python programming language. It works by sending an HTTP request using the Steam 'getdetails' API to get each game's details, utilizing the unique ID number from each game called appid. The details to be extracted were pre-determined to avoid getting unnecessary information. This process was repeated for all games. Then all the data from the games was parsed and then compiled into a single CSV file. The metadata included genres, release date, supported languages, and many more. Over 20 columns of raw data were collected using the API.

B. Data Pre-processing

Since the raw data were collected from different sources, it needed to be pre-processed to be ready for model training. Missing values on the metadata were removed, and several features were transformed into new ones with feature engineering. Table I shows an example of new features created from the raw metadata. For historical data, three new features were created. First was the game's age since the release date, then Cumulative Moving Average (CMA) for both positive and negative reviews. All the new features were calculated for each observation. The result was a total of six features for the historical data excluding the time variable. Table II shows an example of the resulting data from the game Counter-Strike 2, one of the games on the dataset.

Additional features were created for the historical data using the sliding window technique. The sliding window technique transforms the time-series into a supervised learning problem by shifting the data, taking prior observations as

lagged features, depending on the window size. For example, the current observation is labeled as *x* and the sliding windows technique was applied to capture observations for the past 30 days, the data would be shifted 30 days, and the window would contain observations from *x*-29 to *x*, with *x* being the most recent. In this research, a window size of 7 was randomly chosen, meaning that observations from previous seven days were used as lagged features. The technique was applied to all features on the historical data, resulting in a total of 42 features. Then the data was shifted once more to obtain the target feature *x*+1, which was the observation on the next day. Then the processed historical data was merged with the metadata, and then split into training/validation/testing sets, with a ratio of 60/20/20. The split was done to ensure the model robustness against unseen data. All features were normalized after split using Min-Max scaler, including the target feature.

TABLE I. EXAMPLE OF NEW FEATURES CREATED FROM METADATA

name	app_id	required_age	is_free	dlc	achievements	full_controller_support
Grand Theft Auto V	271590	17	0	1	77	1
No Man's Sky	275850	0	0	0	27	1
BeamNG.drive	284160	0	0	0	4	0
Sid Meier's Civilization® VI	289070	0	0	10	320	0

TABLE II. EXAMPLE OF HISTORICAL DATA FROM COUNTER-STRIKE 2

DateTime	Players	Positive reviews	Negative reviews	Positive_CMA	Negative_CMA	days_since_release
2024-03-14	1447897	2414	-862	2025.5721809169765	-304.2506195786865	4223
2024-03-15	1474990	0	0	2024.9448745741715	-304.1563951687829	4224
2024-03-16	1490175	2468	-858	2025.0820433436531	-304.32786377708976	4225
2024-03-17	1425033	0	0	2024.4552770040236	-304.23367378520584	4226

C. Feature Selection

This study uses three feature selection methods, which are Pearson's Correlation Feature Selection for filter method, Recursive Feature Elimination (RFE) for wrapper method, and embedded feature selection method from the model itself.

1) Pearson's correlation feature selection: This feature selection method uses Pearson's Correlation Coefficient (PCC), which measures linear relationship between two or more variables. The correlation value *r* between variables *X* and *y* can be obtained through Eq. (1).

$$r = \frac{\Sigma_{i=1}^{n}(X_i-\bar{X})(y_i-\bar{y})}{\sqrt{\Sigma_{i=1}^{n}(X_i-\bar{X})^2}\cdot\sqrt{\Sigma_{i=1}^{n}(y_i-\bar{y})^2}} \quad (1)$$

where $\bar{X}$ and $\bar{y}$ are mean values of variables *X* and *y*. The correlation value ranges from -1 to +1, where values closer to -1 or +1 indicate stronger correlation, and values closer to 0 indicate weaker correlation. This method selects features with correlation value higher than a certain threshold.

2) Recursive Feature Elimination (RFE): This feature selection method selects the most important features in the dataset by recursively removing the least important features until the number of features to select is reached.

3) Embedded method: Embedded method refers to feature selection method that is built-in to the model itself. The model performs feature selection during training. Models based on decision tree like Random Forest and XGBoost use feature importance to select the most relevant features.

D. Machine Learning Models

This study used three machine learning models: Random Forest (RF), Support Vector Regression (SVR), and XGBoost.

1) Random forest: Random Forest is an ensemble model of decision trees that use random sub-samples from a dataset for prediction, and can be used for both classification and regression tasks [19]. For regression, the model takes predictions from all decision trees then averages them for the final result.

2) Support Vector Regression (SVR): Support Vector Regression (SVR) is a type of Support Vector Machine (SVM) that is used for regression tasks, and is a commonly used method for time-series forecasting [22]. It tries to fit an optimal hyperplane for predicting continues values.

3) XGBoost: XGBoost stands for Extreme Gradient Boosting. It can also be used in both classification and regression tasks. XGBoost is based on Gradient Boosting algorithm, where the decision trees are added to the model sequentially.

E. Model Evaluation

This study used and compared the combinations of feature selection methods machine learning models. Three machine learning models were used and tested: Random Forest (RF), Support Vector Regression (SVR), and XGBoost. The models were first trained using default parameters, then tuned using Grid Search hyperparameter tuning to find the best settings for each model. We used four evaluation metrics to evaluate the model's performance: Coefficient of Determination (R^2), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE),

and Mean Absolute Percentage Error (MAPE), all of which can be defined in Eq. (2), (3), (4), and (5) respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \quad (3)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (4)$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \quad (5)$$

where, y_i is the actual value, $\hat{y}_i$ is the predicted value, and $\bar{y}_i$ is the mean value.

IV. Results and Discussion

A. Training and Validation Results

The results for model training with default parameters along with the validation can be seen on Table III, Table IV, and Table V. The results were grouped based on the feature selection method for readability purposes. The best results from each method are highlighted in bold.

TABLE III. Training and Validation – Pearson's Feature Selection

Model	Training				Validation			
	R²	***RMSE***	***MAE***	***MAPE***	***R²***	***RMSE***	***MAE***	***MAPE***
RF	**0.9971**	**22568.1233**	**10925.2073**	**5.81%**	0.9932	25517.8834	13777.1053	6.70%
SVR	0.9913	13470.8283	5308.5905	6.05%	0.8868	6192.4275	3751.0109	7.01%
XGBoost	0.9959	26876.6614	13251.0304	8.11%	**0.9943**	**23282.8510**	**12679.5643**	**6.69%**

TABLE IV. Training and validation – RFE

Model	Training				Validation			
	R²	***RMSE***	***MAE***	***MAPE***	***R²***	***RMSE***	***MAE***	***MAPE***
RF	**0.9983**	**17410.7001**	**9081.5890**	**4.89%**	**0.9938**	**24321.2331**	**13347.6318**	19.29%
SVR	0.9941	32088.4842	19245.7056	31.47%	0.9930	25748.6940	17236.5587	22.32%
XGBoost	0.9969	23422.6130	13053.8578	8.72%	0.9936	24719.3096	14251.4393	**10.08%**

TABLE V. Training and validation – Embedded Feature Selection

Model	Training				Validation			
	R²	***RMSE***	***MAE***	***MAPE***	***R²***	***RMSE***	***MAE***	***MAPE***
RF	**0.9962**	**25747.2753**	12930.4714	6.93%	0.9913	28877.6052	15217.6164	6.98%
XGBoost	0.9962	26001.9723	**11681.7900**	**6.20%**	**0.9945**	**22880.2620**	**11970.0420**	**5.91%**

B. Testing Results

The models were then put to test to see how well the model generalizes with truly unseen data using the testing set. The results of the experiments can be seen on Table VI, Table VII, and Table VIII. The results were grouped based on the feature selection method for readability purposes. Models with all the features were also tested for comparison and can be seen on Table IX. The best results from each method are highlighted in bold.

TABLE VI. Experiment Results – Pearson's Feature Selection

Model	R^2	RMSE	MAE	MAPE
RF	**0.9943**	**30926.7590**	**16087.1698**	**5.49%**
SVR	0.8280	60648.5777	21614.1249	10.21%
XGBoost	0.9925	35243.8582	18644.4787	6.72%

TABLE VII. Experiment Results – RFE

Model	R^2	RMSE	MAE	MAPE
RF	0.9933	33282.8687	18839.3967	19.90%
SVR	**0.9937**	**32290.6700**	**20560.2261**	14.30%
XGBoost	0.9845	50718.4633	25457.2603	**11.26%**

TABLE VIII. Experiment Results – Embedded Feature Selection

Model	R^2	RMSE	MAE	MAPE
RF	0.9900	40899.9580	22687.4373	8.09%
XGBoost	**0.9926**	**35068.2627**	**17777.9998**	**5.90%**

TABLE IX. Experiment Results – No Feature Selection

Model	R^2	RMSE	MAE	MAPE
RF	**0.9934**	**33140.1820**	**18877.3783**	**17.72%**
SVR	0.9926	34990.9246	22427.8018	23.58%
XGBoost	0.9766	62394.0293	32761.8084	23.78%

Random Forest combined with Pearson's Correlation for feature selection gives the best results overall, with an R^2 score of 0.9943, a slight improvement from the model using all features with an R^2 score of 0.9934. Fig. 2 shows the prediction error of the model. This indicates that even without feature selection, the model was still able to explain more than 95% of the variance. The MAPE also dropped significantly from 17.72% to 5.49%. XGBoost seemed to benefit from feature selection the most, based on the improved results compared to no feature selection.

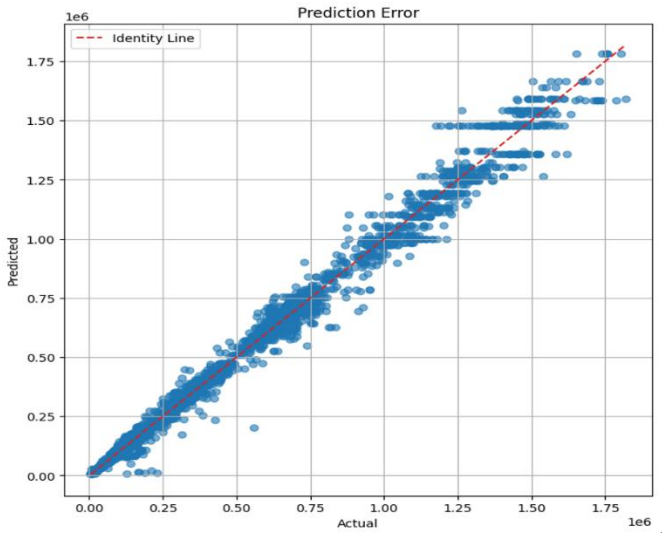


Fig. 2. Prediction error for Random Forest – Pearsons' Feature Selection

C. Results Comparison

A previous study using a linear model [16] had achieved an R^2 score of 0.6756, indicating that the model managed to capture 67.56% of the variation. The results indicated that linear models might not be suitable for predicting the number of players. Another study [18] tested various models, including SVR, RF, and Ridge Regression, with Pearson's Correlation feature selection method. The method reduced the number of features to be used in the models from 889 to 163. The result achieved with the best model was an R^2 score of 0.805, indicating that the model managed to capture 80.5% of the variation. The best model in this study achieved an R^2 score of 0.9943, and an average R^2 score above 0.9 across all models, a better result compared to the previous studies mentioned above.

V. Conclusion and Future Work

In this paper, we presented a method using lagged features in predicting daily player count, using historical data from video games on the Steam platform. The lagged features were created using sliding window method, using data from the last 7 days. Several feature selection methods and machine learning models were tested. Random Forest and Pearson's Correlation Feature Selection shows the best predictive performance amongst all combinations. Despite that, all the other combinations have an average R^2 score above 0.9, showing the effectiveness of our method.

However, there are several limitations in this study. The data used was from video games with the most player populations at the time of the study, so it might be different in the future. This study didn't consider the price of each video game, but only whether they were free or not. Future works may consider adding individual game prices, and the price during sales for more detailed analysis. Game ratings obtained from sentiment analysis of user reviews may also be used for more accurate results.

References

[1] Y. S. Balhara, D. Kattula, S. Singh, S. Chukkali, and R. Bhargava, "Impact of lockdown following COVID-19 on the gaming behavior of college students," Indian J Public Health, vol. 64, no. 6, p. 172, 2020, doi: 10.4103/ijph.IJPH_465_20.

[2] M. Vuorre, D. Zendle, E. Petrovskaya, N. Ballou, and A. K. Przybylski, "A Large-Scale Study of Changes to the Quantity, Quality, and Distribution of Video Game Play During a Global Health Pandemic," Technology, Mind, and Behavior, vol. 2, no. 4, 2021, doi: 10.1037/tmb0000048.

[3] E. Haug et al., "Increased Gaming During COVID-19 Predicts Physical Inactivity Among Youth in Norway—A Two-Wave Longitudinal Cohort Study," Front Public Health, vol. 10, Feb. 2022, doi: 10.3389/fpubh.2022.812932.

[4] M. N. Rizani, M. N. A. Khalid, and H. Iida, "Application of Meta-Gaming Concept to the Publishing Platform: Analysis of the Steam Games Platform," Information, vol. 14, no. 2, p. 110, Feb. 2023, doi: 10.3390/info14020110.

[5] T. Guzsvinecz and J. Szűcs, "Length and sentiment analysis of reviews about top-level video game genres on the steam platform," Comput Human Behav, vol. 149, p. 107955, Dec. 2023, doi: 10.1016/J.CHB.2023.107955.

[6] N. Y. Prathama, R. Asmara, and A. R. Barakbah, "Game Data Analytics using Descriptive and Predictive Mining," in 2020 International Electronics Symposium (IES), IEEE, Sep. 2020, pp. 398–405. doi: 10.1109/IES50839.2020.9231949.

[7] D. Zendle, R. Meyer, and N. Ballou, "The changing face of desktop video game monetisation: An exploration of exposure to loot boxes, pay to win, and cosmetic microtransactions in the most-played Steam games of 2010-2019," PLoS One, vol. 15, no. 5, 2020, doi: 10.1371/journal.pone.0232780.

[8] C. Phillips, M. Klarkowski, J. Frommel, C. Gutwin, and R. L. Mandryk, "Identifying Commercial Games with Therapeutic Potential through a Content Analysis of Steam Reviews," Proc ACM Hum Comput Interact, vol. 5, no. CHI PLAY, pp. 1–21, Oct. 2021, doi: 10.1145/3474682.

[9] A. M. Thorhauge and R. K. L. Nielsen, "Epic, Steam, and the role of skin-betting in game (platform) economies," Journal of Consumer Culture, vol. 21, no. 1, pp. 52–67, Feb. 2021, doi: 10.1177/1469540521993929.

[10] C. Moro, C. Phelps, and J. Birt, "Improving serious games by crowdsourcing feedback from the STEAM online gaming community," Internet High Educ, vol. 55, p. 100874, Oct. 2022, doi: 10.1016/J.IHEDUC.2022.100874.

[11] A. O. Thunström, I. Sarajlic Vukovic, L. Ali, T. Larson, and S. Steingrimsson, "Prevalence of virtual reality (VR) games found through mental health categories on STEAM: a first look at VR on commercial platforms as tools for therapy," Nord J Psychiatry, vol. 76, no. 6, pp. 474–485, Aug. 2022, doi: 10.1080/08039488.2021.2003859.

[12] J. Kohlburn, H. Cho, and H. Moore, "Players' perceptions of sexuality and gender-inclusive video games a pragmatic content analysis of steam reviews," Convergence: The International Journal of Research into New Media Technologies, vol. 29, no. 2, pp. 379–399, Apr. 2023, doi: 10.1177/13548565221137481.

[13] L. Y. Xiao and L. L. Henderson, "Illegal video game loot boxes with transferable content on steam: a longitudinal study on their presence and non-compliance with and non-enforcement of gambling law," Int Gambl Stud, pp. 1–27, Aug. 2024, doi: 10.1080/14459795.2024.2390827.

[14] A. M. Thorhauge, "The steam platform economy: From retail to player-driven economies," New Media Soc, vol. 26, no. 4, pp. 1963–1983, Apr. 2024, doi: 10.1177/14614448221081401.

[15] K. Stecuła, "Analysis of asymmetric VR games – Steam platform case study," Technol Soc, vol. 78, p. 102673, Sep. 2024, doi: 10.1016/J.TECHSOC.2024.102673.

[16] H. Zhang, "The Establishment of Multi-variable Linear Regression in Steam Sales," in Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022), 2022, pp. 853–856. doi: 10.2991/aebmr.k.220307.137.

[17] D. Wannigamage, M. Barlow, E. Lakshika, and K. Kasmarik, "Analysis and Prediction of Player Population Changes in Digital Games During the COVID-19 Pandemic," 2020, pp. 458–469. doi: 10.1007/978-3-030-64984-5_36.

[18] S. Wu, H. Hu, Y. Zheng, Q. Zhen, S. Zhang, and C. Zhan, "The Impact of COVID-19 on Online Games: Machine Learning and Difference-in-Difference," Communications in Computer and Information Science, vol. 1492 CCIS, pp. 458–470, 2022, doi: 10.1007/978-981-19-4549-6_35.

[19] R. R. Varghese, D. R. Aiswarya, A. Roy, V. Muraly, and S. Renjith, A Novel Approach to Predict Success of Online Games Using Random Forest Regressor for Time Series Data, vol. 881. 2022. doi: 10.1007/978-981-19-1111-8_3.

[20] A. S. Teja, M. L. I. Hanafi, and N. N. Qomariyah, "Predicting Steam Games Rating with Regression," E3S Web of Conferences, vol. 388, p. 02001, May 2023, doi: 10.1051/e3sconf/202338802001.

[21] S. Abdul-Rahman, M. F. A. M. Ali, A. A. Bakar, and S. Mutalib, "Enhancing churn forecasting with sentiment analysis of steam reviews," Soc Netw Anal Min, vol. 14, no. 1, 2024, doi: 10.1007/s13278-024-01337-3.

[22] J. M. Valente and S. Maldonado, "SVR-FFS: A novel forward feature selection approach for high-frequency time series forecasting using support vector regression," Expert Syst Appl, vol. 160, p. 113729, Dec. 2020, doi: 10.1016/j.eswa.2020.113729.

Cross-Entropy-Driven Optimization of Triangular Fuzzy Neutrosophic MADM for Urban Park Environmental Design Quality Evaluation

Xing She[1], Xi Xie[2]*, Peng Xie[3]
School of Arts and Design, Anhui University of Technology, Maanshan, Anhui, China[1, 2]
Faculty of Humanities and Arts, Macao University of Science and Technology, Macao, China[3]

Abstract—The evaluation of urban park environmental design quality focuses on functionality, aesthetics, ecology, and user experience. Functionality ensures practical facilities, clear zoning, and accessibility. Aesthetics emphasizes visual harmony, cultural integration, and artistic appeal. Ecological quality assesses vegetation, biodiversity, and sustainability, promoting environmental protection. User experience evaluates comfort, safety, inclusivity, and the ability to meet diverse needs. A well-designed park balances these elements, fostering harmony between humans and nature while enhancing public well-being, environmental awareness, and the overall urban living experience. The quality evaluation of urban park environmental design is multi-attribute decision-making (MADM). In this study, triangular fuzzy neutrosophic number cross-entropy (TFNN-CE) approach is executed under triangular fuzzy neutrosophic sets (TFNSs). Furthermore, Then, entropy is employed to execute the weight and TFNN-CE approach is executed for MADM under TFNSs. Finally, numerical example for quality evaluation of urban park environmental design is executed the advantages of TFNN-CE approach through different comparisons. The major contributions of this study could be executed: (1) entropy is employed to execute the weight under TFNSs; (2) TFNN-CE approach is executed under TFNSs; (3) TFNN-CE approach is put forward for MADM under TFNSs; (4) numerical example for quality evaluation of urban park environmental design is executed the advantages of TFNN-CE approach through different comparisons.

Keywords—Multiple-Attribute Decision-Making (MADM) problems; Triangular Fuzzy Neutrosophic Sets (TFNSs); cross-entropy approach; TFNN-CE approach; urban park environmental design

I. INTRODUCTION

Nowadays, the construction of urban parks has attracted the attention of the whole society. Integrating the design concepts and approaches of the original ecological environment landscape into the construction of urban parks is an important trend in the development of urban parks [1-3]. Integrating the design concept of original ecological environment into the construction of urban parks is an important core of urban park construction. It not only fully utilizes the existing ecological resources of the city itself, but also maintains the ecological environment of the city itself and protects the original biological communities, enabling them to continue to survive [4-6]. Therefore, it is necessary to fully investigate and analyze the original ecological environment of human cities, adapt to the original ecological conditions, incorporate artificial design and transformation, combine human activities with nature, protect the original ecological environment, maintain its richness, ensure the scientific rationality of urban park planning and construction, and create a livable and harmonious living environment for urban residents [7-9]. In the natural world, biological landscapes are full of vitality, so when designing urban parks, designers can use this as a starting point to integrate rich plants into landscape units, and use species diversity to maintain the ecological balance of the city [10-13]. Firstly, vegetation selection should be based on the geographical location and urban climate of each city, ensuring not only the reasonable allocation of trees, but also a high survival rate of the selected tree species. Secondly, it is necessary to fully consider the habits of animals [14-16]. When designing landscapes, unique vegetation is utilized to attract animal habitats, thereby establishing a stable and balanced ecosystem. Finally, based on the ecological environment of the city itself, design and treat its lakes and corridors, such as adding landscape patches, transforming natural rivers and lakes, and improving their water quality. Add different types of plant areas (such as functional, productive, ornamental, etc.), grassland resources, wetland resources, etc. according to different landscape functions, in order to enrich the biological diversity and stabilize the ecological balance in urban parks [17-19]. Protecting the ecological environment and conserving natural resources are the most important aspects of landscape design in the original ecological environment [1, 20]. Therefore, in the planning process of urban parks, it is necessary to combine them with the actual ecological environment and minimize the waste of natural resources as much as possible. Firstly, when performing artificial scenery, environmentally friendly or energy-saving materials can be selected; secondly, when transforming natural landscapes, ecological engineering should be used to avoid polluting the natural environment [21-23]. Once again, when laying roads, natural soil can be chosen, which not only causes minimal damage to the environment, but also maintains the natural appearance; Finally, it is important to pay attention to the water quality of urban parks, focus on their recycling, and use specially designed pipelines to circulate water quality in a circular manner, thereby playing a role in irrigating green spaces and cleaning parks [24-26]. Natural ecology can effectively remove impurities from air and water sources and make them cleaner, greatly improving air quality and water quality [27, 28]. However, once a certain ecological chain of the original environment is destroyed, it is very easy to damage the overall ecological environment. Therefore, when designing the original

ecological landscape, the principle of non-interference or low intervention should be followed to reduce damage to the ecological rules of the natural environment, and to build scientific urban parks [29-31]. During construction, it is not allowed to change the biological elements in the original ecosystem, adhere to the natural law of survival of the fittest, reduce or even eliminate artificial traces, and follow the development rules of natural ecology. The design principle of low intervention can ensure that the ecological environment maintains its complete function and the balance of the original ecology, optimize the artistic expression of artificial design, and reduce the cost of park construction [32, 33]. In terms of artistic creation, the original ecological environment landscape design should be combined with the different cultural characteristics of each city, highlighting the cultural value of urban park ecological environment [34-36]. Based on the cultural characteristics of each region, inject cultural emotions into the cultural landscape, arouse people's resonance, and integrate the urban parks executed into modern civilization on the basis of meeting the requirements of the original ecosystem. Designers should fully and reasonably consider the degree of cultural integration, and maximize the use of natural elements without excessive human processing, integrating cultural and economic benefits into urban parks [37-39].

Due to the complexity, diversity, and different preferences of decision-makers in the decision-making environment, MADM problems have a certain degree of fuzziness and uncertainty [40-43]. For this reason, domestic and foreign scholars have focused on complex decision-making problems, involving fields such as construction, industry and agriculture [44-50], but in terms of initial information expression, they are based on the fuzzy sets proposed by Zadeh [51], which does not fully characterize decision-makers' hesitation, fuzziness, and different biases. The assignment of indicators is influenced by the experience of decision-makers and the results of fuzzy comprehensive evaluation are influenced by the principle of maximum probability, which is not universal [52-55]. In addition, due to the increasing amount of decision information, the complexity and uncertainty of decision problems become higher. The triangular fuzzy neutrosophic sets (TFNSs) proposed by Biswas [56] overcomes the limitations of Zadeh's fuzzy set theory, allows DMs to represent the membership, indeterminacy-membership and falsity-membership which is depicted through triangular fuzzy numbers (TFNs) and meticulously expresses the decision information of uncertainty and preferences in the MADM process, thereby improving the scientific and rational nature of decision results. The quality evaluation of urban park environmental design is MADM. TFNSs [56] is efficient tool for managing fuzzy information during quality evaluation of urban park environmental design. The CE approach [57] was put forward the MADM. Furthermore, many approaches utilize CE approach [58-64] and entropy approach [57, 65-67] to administrate the MADM. Until now, no or few approaches have been administrated on CE approach for MADM based on entropy approach along with TFNSs. Thus, in this study, TFNN-CE approach is executed under TFNSs. Then, the entropy is employed to execute the weight under TFNSs and TFNN-CE approach is executed for MADM under TFNSs. Finally, numerical example for quality evaluation of urban park environmental design is executed the advantages of TFNN-CE approach through different comparisons. The major contributions of this study could be executed: (1) entropy is employed to execute the weight under TFNSs; (2) TFNN-CE approach is executed under TFNSs; (3) TFNN-CE approach is put forward for MADM under TFNSs; (4) numerical example for quality evaluation of urban park environmental design is executed the advantages of TFNN-CE approach through different comparisons.

The research structure is executed. In Section II, the TFNNSs is executed. The TFNN-CE is produced in Section III. In Section IV, TFNN-CE is fully put forward MADM under TFNSs. Section V executed numerical example for quality evaluation of urban park environmental design and comparative analysis. Final remark is executed in Section VI.

II. PRELIMINARIES

Biswas [56] built the TFNSs.

Definition 1[56]. The TFNSs is depicted:

$$ST=\left\{\left(\theta,SA(\theta),SB(\theta),SC(\theta)\right)\middle|\,\theta\in\Theta\right\} \quad (1)$$

where $SA(\theta),SB(\theta),SC(\theta)\in[0,1]$ execute the membership, indeterminacy-membership and falsity-membership which is executed through TFNs.

$$SA(\theta)=\left(SA^{L}(\theta),SA^{M}(\theta),SA^{U}(\theta)\right),0\le SA^{L}(\theta)\le SA^{M}(\theta)\le SA^{U}(\theta)\le 1 \quad (2)$$

$$SB(\theta)=\left(SB^{L}(\theta),SB^{M}(\theta),SB^{U}(\theta)\right),0\le SB^{L}(\theta)\le SB^{M}(\theta)\le SB^{U}(\theta)\le 1 \quad (3)$$

$$SC(\theta)=\left(SC^{L}(\theta),SC^{M}(\theta),SC^{U}(\theta)\right),0\le SC^{L}(\theta)\le SC^{M}(\theta)\le SC^{U}(\theta)\le 1 \quad (4)$$

$ST=\left\{\begin{matrix}\left(SA^{L},SA^{M},SA^{U}\right),\\ \left(SB^{L},SB^{M},SB^{U}\right),\left(SC^{L},SC^{M},SC^{U}\right)\end{matrix}\right\}$ is called a TFNN, $0\le SA^{U}+SB^{U}+SC^{U}\le 3$.

Definition 2[56]. Let

$ST_1=\left\{\begin{matrix}\left(SA_1^{L},SA_1^{M},SA_1^{U}\right),\\ \left(SB_1^{L},SB_1^{M},SB_1^{U}\right),\left(SC_1^{L},SC_1^{M},SC_1^{U}\right)\end{matrix}\right\}$,

$ST_2=\left\{\begin{matrix}\left(SA_2^{L},SA_2^{M},SA_2^{U}\right),\\ \left(SB_2^{L},SB_2^{M},SB_2^{U}\right),\left(SC_2^{L},SC_2^{M},SC_2^{U}\right)\end{matrix}\right\}$ and

$$ST=\left\{\begin{array}{l}\left(SA^{L},SA^{M},SA^{U}\right),\\ \left(SB^{L},SB^{M},SB^{U}\right),\left(SC^{L},SC^{M},SC^{U}\right)\end{array}\right\},\ \chi>0,$$

the operation laws are executed:

$$(1)\ ST_1\oplus ST_2=\left\{\begin{array}{l}\left(\begin{array}{l}SA_1^L+SA_2^L\text{-}SA_1^LSA_2^L,\\ SA_1^M+SA_2^M\text{-}SA_1^MSA_2^M,SA_1^U+SA_2^U\text{-}SA_1^USA_2^U\end{array}\right),\\ \left(SB_1^LSB_2^L,SB_1^MSB_2^M,SB_1^USB_2^U\right),\\ \left(SC_1^LSC_2^L,SC_1^MSC_2^M,SC_1^USC_2^U\right)\end{array}\right\};$$

$$(2)\ ST_1\otimes ST_2=\left\{\begin{array}{l}\left(SA_1^LSA_2^L,SA_1^MSA_2^M,SA_1^USA_2^U\right),\\ \left(\begin{array}{l}SB_1^L+SB_2^L\text{-}SB_1^LSB_2^L,\\ SB_1^M+SB_2^M\text{-}SB_1^MSB_2^M,SB_1^U+SB_2^U\text{-}SB_1^USB_2^U\end{array}\right),\\ \left(\begin{array}{l}SC_1^L+SC_2^L\text{-}SC_1^LSC_2^L,\\ SC_1^M+SC_2^M\text{-}SC_1^MSC_2^M,SC_1^U+SC_2^U\text{-}SC_1^USC_2^U\end{array}\right)\end{array}\right\};$$

$$(3)\ \chi ST=\left\{\begin{array}{l}\left(1-\left(1-SA^L\right)^{\chi},1-\left(1-SA^M\right)^{\chi},1-\left(1-SA^U\right)^{\rho\chi}\right),\\ \left(\left(SB^L\right)^{\chi},\left(SB^M\right)^{\chi},\left(SB^U\right)^{\chi}\right),\\ \left(\left(SC^L\right)^{\chi},\left(SC^M\right)^{\chi},\left(SC^U\right)^{\chi}\right)\end{array}\right\};$$

$$(4)\ ST^{\chi}=\left\{\begin{array}{l}\left(\left(SA^L\right)^{\chi},\left(SA^M\right)^{\chi},\left(SA^U\right)^{\chi}\right),\\ \left(1-\left(1-SB^L\right)^{\chi},1-\left(1-SB^M\right)^{\chi},1-\left(1-SB^U\right)^{\chi}\right),\\ \left(1-\left(1-SC^L\right)^{\chi},1-\left(1-SC^M\right)^{\chi},1-\left(1-SC^U\right)^{\chi}\right)\end{array}\right\}.$$

From Definition 2, the operation laws have different executed properties.

$$(1)\ ST_1\oplus ST_2=ST_2\oplus ST_1; \tag{5}$$

$$(2)ST_1\otimes ST_2=ST_2\otimes ST_1,\left(\left(ST_1\right)^{\chi_1}\right)^{\chi_2}=\left(ST_1\right)^{\chi_1\chi_2}; \tag{6}$$

$$(3)\ \chi\left(ST_1\oplus ST_2\right)=\chi ST_1\oplus\chi ST_2,\left(ST_1\otimes ST_2\right)^{\chi}=\left(ST\right)^{\chi}\otimes\left(ST_2\right)^{\chi}; \tag{7}$$

$$(4)\chi_1ST_1\oplus\chi_2ST_1=\left(\chi_1+\chi_2\right)ST_1,\left(ST_1\right)^{\chi_1}\otimes\left(ST_1\right)^{\chi_2}=\left(ST_1\right)^{\left(\chi_1+\chi_2\right)}. \tag{8}$$

Definition 3 [56]. Let $ST=\left\{\begin{array}{l}\left(SA^L,SA^M,SA^U\right),\\ \left(SB^L,SB^M,SB^U\right),\left(SC^L,SC^M,SC^U\right)\end{array}\right\}$, the score and accuracy functions of ST is:

$$SF\left(ST\right)=\frac{1}{12}\left[\begin{array}{l}8+\left(SA^L+2SA^M+SA^U\right)\\ -\left(SB^L+2SB^M+SB^U\right)\\ -\left(SC^L+2SC^M+SC^U\right)\end{array}\right],$$
$$SF\left(ST\right)\in\left[0,1\right] \tag{9}$$

$$AF\left(ST\right)=\frac{1}{4}\left[\begin{array}{l}\left(SA^L+2SA^M+SA^U\right)\\ -\left(SB^L+2SB^M+SB^U\right)\end{array}\right],$$
$$AF\left(WW\right)\in\left[-1,1\right] \tag{10}$$

For ST_1 and ST_2, then

(1) *if* $SF\left(ST_1\right)\prec SF\left(ST_2\right),\ ST_1\prec ST_2$;
(2) *if* $SF\left(ST_1\right)=SF\left(ST_2\right),AF\left(ST_1\right)\prec AF\left(ST_2\right),\ ST_1\prec ST_2$;
(3) *if* $SF\left(ST_1\right)=SF\left(ST_2\right),AF\left(ST_1\right)=AF\left(ST_2\right),ST_1=ST_2$.

III. Cross-entropy with TFNSs

Bhandari and Pal [67] created the cross entropy.

Definition 4[67]. Let $s\alpha=\left(s\alpha\left(s_1\right),s\alpha\left(s_2\right),\cdots,s\alpha\left(s_n\right)\right)$ and $s\beta=\left(s\beta\left(s_1\right),s\beta\left(s_2\right),\cdots,s\beta\left(s_n\right)\right)$. The cross-entropy of $s\alpha$ from $s\beta$ is executed:

$$CE\left(s\alpha,s\beta\right)=\sum_{j=1}^{n}\left(\begin{array}{l}s\alpha\left(s_j\right)\ln\frac{s\alpha\left(s_j\right)}{s\beta\left(s_j\right)}\\ +\left(1-s\alpha\left(s_j\right)\right)\ln\frac{1-s\alpha\left(s_j\right)}{1-s\beta\left(s_j\right)}\end{array}\right) \tag{11}$$

which is the discrimination degree of $w\alpha$ from $w\beta$.

Shang and Jiang [57] created the modified cross-entropy.

Definition 5 [57]. Let $s\alpha=\left(s\alpha\left(s_1\right),s\alpha\left(s_2\right),\cdots,s\alpha\left(s_n\right)\right)$ and $s\beta=\left(s\beta\left(s_1\right),s\beta\left(s_2\right),\cdots,s\beta\left(s_n\right)\right)$. The cross-entropy of $s\alpha$ from $s\beta$ is executed:

$$CE\left(s\alpha, s\beta\right) = \sum_{j=1}^{n}\left(\begin{array}{l} s\alpha\left(s_j\right)\ln\dfrac{s\alpha\left(s_j\right)}{\frac{1}{2}\left(s\alpha\left(s_j\right)+s\beta\left(s_j\right)\right)} + \\ \left(1-s\alpha\left(s_j\right)\right)\ln\dfrac{1-s\alpha\left(s_j\right)}{1-\frac{1}{2}\left(s\alpha\left(s_j\right)+s\beta\left(s_j\right)\right)} \end{array}\right) \quad (12)$$

which is discrimination degree of $s\alpha$ from $s\beta$.

Then, TFNN cross-entropy (TFNN-CE) is executed in light with cross-entropy [57] and TFNNs [56].

Definition 6. Let $ST_1 = \left\{\begin{array}{l}\left(SA_1^L, SA_1^M, SA_1^U\right), \\ \left(SB_1^L, SB_1^M, SB_1^U\right), \left(SC_1^L, SC_1^M, SC_1^U\right)\end{array}\right\}$, $ST_2 = \left\{\begin{array}{l}\left(SA_2^L, SA_2^M, SA_2^U\right), \\ \left(SB_2^L, SB_2^M, SB_2^U\right), \left(SC_2^L, SC_2^M, SC_2^U\right)\end{array}\right\}$. The TFNN-CE is produced between ST_1 and ST_2:

$$\begin{aligned} &\text{TFNN-CE}\left(ST_1, ST_2\right) \\ &= \left(\begin{array}{l} \left(\dfrac{SA_1^L+SA_1^M+SA_1^U}{3}\right)\ln\dfrac{\frac{SA_1^L+SA_1^M+SA_1^U}{3}}{\rho\left(\frac{SA_1^L+SA_1^M+SA_1^U}{3}+\frac{SA_2^L+SA_2^M+SA_2^U}{3}\right)} \\ +\left(1-\dfrac{SA_1^L+SA_1^M+SA_1^U}{3}\right)\ln\dfrac{1-\frac{SA_1^L+SA_1^M+SA_1^U}{3}}{1-\rho\left(\frac{SA_1^L+SA_1^M+SA_1^U}{3}+\frac{SA_2^L+SA_2^M+SA_2^U}{3}\right)} \end{array}\right) \\ &+ \left(\begin{array}{l} \left(\dfrac{SB_1^L+SB_1^M+SB_1^U}{3}\right)\ln\dfrac{\frac{SB_1^L+SB_1^M+SB_1^U}{3}}{\rho\left(\frac{SB_1^L+SB_1^M+SB_1^U}{3}+\frac{SB_2^L+SB_2^M+SB_2^U}{3}\right)} \\ +\left(1-\dfrac{SB_1^L+SB_1^M+SB_1^U}{3}\right)\ln\dfrac{1-\frac{SB_1^L+SB_1^M+SB_1^U}{3}}{1-\rho\left(\frac{SB_1^L+SB_1^M+SB_1^U}{3}+\frac{SB_2^L+SB_2^M+SB_2^U}{3}\right)} \end{array}\right) \\ &+ \left(\begin{array}{l} \left(\dfrac{SC_1^L+SC_1^M+SC_1^U}{3}\right)\ln\dfrac{\frac{SC_1^L+SC_1^M+SC_1^U}{3}}{\rho\left(\frac{SC_1^L+SC_1^M+SC_1^U}{3}+\frac{SC_2^L+SC_2^M+SC_2^U}{3}\right)} \\ +\left(1-\dfrac{SC_1^L+SC_1^M+SC_1^U}{3}\right)\ln\dfrac{1-\frac{SC_1^L+SC_1^M+SC_1^U}{3}}{1-\rho\left(\frac{SC_1^L+SC_1^M+SC_1^U}{3}+\frac{SC_2^L+SC_2^M+SC_2^U}{3}\right)} \end{array}\right) \end{aligned} \quad (13)$$

which is discrimination degree of ST_1 and ST_2 and $\rho \in [0,1]$.

In light with Shannon's inequality [66], it's easily verify that $\text{TFNN-CE}\left(ST_1, ST_2\right) \geq 0$, and $\text{TFNN-CE}\left(ST_1, ST_2\right) = 0$ if and only if $\left(SA_1^L, SA_1^M, SA_1^U\right) = \left(SA_2^L, SA_2^M, SA_2^U\right)$, $\left(SB_1^L, SB_1^M, SB_1^U\right) = \left(SB_2^L, SB_2^M, SB_2^U\right)$, $\left(SC_1^L, SC_1^M, SC_1^U\right) = \left(SC_2^L, SC_2^M, SC_2^U\right)$.

IV. CROSS-ENTROPY TECHNIQUE FOR MADM WITH TFNNS

The TFNN-CE technique is executed for TFNN-MADM. Suppose that m alternatives $\{SX_1, SX_2, \ldots, SX_m\}$, n attributes $\{SG_1, SG_2, \ldots, SG_n\}$ with weight $sw = (sw_1, sw_2, \cdots, sw_n)$. TFNN-CE technique is executed for MADM with TFNNs.

Step 1. Execute the TFNN-matrix $STFNN = \left[STFNN_{ij}\right]_{m\times n}$:

$$STFNN = \left[STFNN_{ij}\right]_{m\times n} = \begin{array}{c} \\ SX_1 \\ SX_2 \\ \vdots \\ SX_m \end{array} \begin{array}{c} \begin{array}{cccc} SG_1 & SG_2 & \ldots & SG_n \end{array} \\ \begin{bmatrix} STFNN_{11} & STFNN_{12} & \ldots & STFNN_{1n} \\ STFNN_{21} & STFNN_{22} & \ldots & STFNN_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ STFNN_{m1} & STFNN_{m2} & \ldots & STFNN_{mn} \end{bmatrix} \end{array} \quad (14)$$

$$STFNN_{ij} = \begin{Bmatrix} \left(\left(SA_{ij}^L\right), \left(SA_{ij}^M\right), \left(SA_{ij}^U\right)\right), \\ \left(\left(SB_{ij}^L\right), \left(SB_{ij}^M\right), \left(SB_{ij}^U\right)\right), \\ \left(\left(SC_{ij}^L\right), \left(SC_{ij}^M\right), \left(SC_{ij}^U\right)\right) \end{Bmatrix}.$$

ωηερε

Step 2. Normalize the $STFNN = \left[STFNN_{ij}\right]_{m\times n}$ to $NSTFNN = \left[STFNN_{ij}^N\right]_{m\times n}$.

For benefit attributes:

$$STFNN_{ij}^N = \begin{Bmatrix} \left(\left(NSA_{ij}^L\right), \left(NSA_{ij}^M\right), \left(NSA_{ij}^U\right)\right), \\ \left(\left(NSB_{ij}^L\right), \left(NSB_{ij}^M\right), \left(NSB_{ij}^U\right)\right), \\ \left(\left(NSC_{ij}^L\right), \left(NSC_{ij}^M\right), \left(NSC_{ij}^U\right)\right) \end{Bmatrix}$$

$$= \begin{Bmatrix} \left(\left(SA_{ij}^L\right), \left(SA_{ij}^M\right), \left(SA_{ij}^U\right)\right), \\ \left(\left(SB_{ij}^L\right), \left(SB_{ij}^M\right), \left(SB_{ij}^U\right)\right), \\ \left(\left(SC_{ij}^L\right), \left(SC_{ij}^M\right), \left(SC_{ij}^U\right)\right) \end{Bmatrix} \quad (15)$$

For cost attributes:

$$STFNN_{ij}^N = \begin{Bmatrix} \left(\left(NSA_{ij}^L\right), \left(NSA_{ij}^M\right), \left(NSA_{ij}^U\right)\right), \\ \left(\left(NSB_{ij}^L\right), \left(NSB_{ij}^M\right), \left(NSB_{ij}^U\right)\right), \\ \left(\left(NSC_{ij}^L\right), \left(NSC_{ij}^M\right), \left(NSC_{ij}^U\right)\right) \end{Bmatrix}$$

$$= \begin{Bmatrix} \left(\left(SC_{ij}^L\right), \left(SC_{ij}^M\right), \left(SC_{ij}^U\right)\right), \\ \left(\left(SB_{ij}^L\right), \left(SB_{ij}^M\right), \left(SB_{ij}^U\right)\right), \\ \left(\left(SA_{ij}^L\right), \left(SA_{ij}^M\right), \left(SA_{ij}^U\right)\right) \end{Bmatrix} \quad (16)$$

Step 3. Produce the TFNN positive ideal solution (TFNNPIS) and $STFNN_i$;

$$TFNNPIS_j = \begin{Bmatrix} \left(\left(NSA_j^L\right)^+, \left(NSA_j^M\right)^+, \left(NSA_j^U\right)^+\right), \\ \left(\left(NSB_j^L\right)^+, \left(NSB_j^M\right)^+, \left(NSB_j^U\right)^+\right), \\ \left(\left(NSC_j^L\right)^+, \left(NSC_j^M\right)^+, \left(NSC_j^U\right)^+\right) \end{Bmatrix} \quad (17)$$

$$SF \begin{Bmatrix} \left(\left(NSA_j^L\right)^+, \left(NSA_j^M\right)^+, \left(NSA_j^U\right)^+\right), \\ \left(\left(NSB_j^L\right)^+, \left(NSB_j^M\right)^+, \left(NSB_j^U\right)^+\right), \\ \left(\left(NSC_j^L\right)^+, \left(NSC_j^M\right)^+, \left(NSC_j^U\right)^+\right) \end{Bmatrix}$$

$$= \max_i SF \begin{Bmatrix} \left(\left(NSA_{ij}^L\right), \left(NSA_{ij}^M\right), \left(NSA_{ij}^U\right)\right), \\ \left(\left(NSB_{ij}^L\right), \left(NSB_{ij}^M\right), \left(NSB_{ij}^U\right)\right), \\ \left(\left(NSC_{ij}^L\right), \left(NSC_{ij}^M\right), \left(NSC_{ij}^U\right)\right) \end{Bmatrix} \quad (18)$$

$$NSTFNN_i = \begin{Bmatrix} \left(\left(NSA_{ij}^L\right), \left(NSA_{ij}^M\right), \left(NSA_{ij}^U\right)\right), \\ \left(\left(NSB_{ij}^L\right), \left(NSB_{ij}^M\right), \left(NSB_{ij}^U\right)\right), \\ \left(\left(NSC_{ij}^L\right), \left(NSC_{ij}^M\right), \left(NSC_{ij}^U\right)\right) \end{Bmatrix} \quad (19)$$

Step 4. The weight numbers are important for MADM [68-72]. Entropy technique [65] is put forward weight numbers. The TFNN decision matrix (TFNNDM) is executed:

$$TFNNDM_{ij}=\frac{1}{2}\left(\frac{SF\left\{\begin{matrix}\left(\left(NSA_{ij}^{L}\right),\left(NSA_{ij}^{M}\right),\left(NSA_{ij}^{U}\right)\right),\\ \left(\left(NSB_{ij}^{L}\right),\left(NSB_{ij}^{M}\right),\left(NSB_{ij}^{U}\right)\right),\\ \left(\left(NSC_{ij}^{L}\right),\left(NSC_{ij}^{M}\right),\left(NSC_{ij}^{U}\right)\right)\end{matrix}\right\}+2}{\sum_{i=1}^{m}\left(SF\left\{\begin{matrix}\left(\left(NSA_{ij}^{L}\right),\left(NSA_{ij}^{M}\right),\left(NSA_{ij}^{U}\right)\right),\\ \left(\left(NSB_{ij}^{L}\right),\left(NSB_{ij}^{M}\right),\left(NSB_{ij}^{U}\right)\right),\\ \left(\left(NSC_{ij}^{L}\right),\left(NSC_{ij}^{M}\right),\left(NSC_{ij}^{U}\right)\right)\end{matrix}\right\}+2\right)}\right), \quad (20)$$

Then, TFNN Shannon entropy (TFNNSE) is produced:

$$TFNNSE_{j}=-\frac{1}{\ln m}\sum_{i=1}^{m}TFNNDM_{ij}\ln TFNNDM_{ij} \quad (21)$$

and $TFNNDM_{ij}\ln TFNNDM_{ij}=0$ if $TFNNDM_{ij}=0$.

Then, the weight numbers are executed:

$$sw_{j}=\frac{1-TFNNSE_{j}}{\sum_{j=1}^{n}\left(1-TFNNSE_{j}\right)} \quad (22)$$

Step 5. Execute the TFNN-CE model between TFNNPIS and $STFNN_{i}$ and $\rho\in[0,1]$:

$$TFNNCE\left(NSTFNN_{i},TFNNPIS\right)$$

$$=\sum_{j=1}^{n}sw_{j}\left(\begin{matrix}\left(\frac{NSA_{ij}^{L}+NSA_{ij}^{M}+NSA_{ij}^{U}}{3}\right)\ln\frac{\frac{NSA_{ij}^{L}+NSA_{ij}^{M}+NSA_{ij}^{U}}{3}}{\rho\left(\begin{matrix}\frac{NSA_{ij}^{L}+NSA_{ij}^{M}+NSA_{ij}^{U}}{3}\\ +\frac{\left(NSA_{j}^{L}\right)^{+}+\left(NSA_{j}^{M}\right)^{+}+\left(NSA_{j}^{U}\right)^{+}}{3}\end{matrix}\right)}\\ +\left(1-\frac{NSA_{ij}^{L}+NSA_{ij}^{M}+NSA_{ij}^{U}}{3}\right)\ln\frac{1-\frac{NSA_{ij}^{L}+NSA_{ij}^{M}+NSA_{ij}^{U}}{3}}{1-\rho\left(\begin{matrix}\frac{NSA_{ij}^{L}+NSA_{ij}^{M}+NSA_{ij}^{U}}{3}\\ +\frac{\left(NSA_{j}^{L}\right)^{+}+\left(NSA_{j}^{M}\right)^{+}+\left(NSA_{j}^{U}\right)^{+}}{3}\end{matrix}\right)}\end{matrix}\right)$$

$$+\sum_{j=1}^{n}sw_{j}\left(\begin{matrix}\left(\frac{NSB_{ij}^{L}+NSB_{ij}^{M}+NSB_{ij}^{U}}{3}\right)\ln\frac{\frac{NSB_{ij}^{L}+NSB_{ij}^{M}+NSB_{ij}^{U}}{3}}{\rho\left(\begin{matrix}\frac{NSB_{ij}^{L}+NSB_{ij}^{M}+NSB_{ij}^{U}}{3}\\ +\frac{\left(NSB_{j}^{L}\right)^{+}+\left(NSB_{j}^{M}\right)^{+}+\left(NSB_{j}^{U}\right)^{+}}{3}\end{matrix}\right)}\\ +\left(1-\frac{NSB_{ij}^{L}+NSB_{ij}^{M}+NSB_{ij}^{U}}{3}\right)\ln\frac{1-\frac{NSB_{ij}^{L}+NSB_{ij}^{M}+NSB_{ij}^{U}}{3}}{1-\rho\left(\begin{matrix}\frac{NSB_{ij}^{L}+NSB_{ij}^{M}+NSB_{ij}^{U}}{3}\\ +\frac{\left(NSB_{j}^{L}\right)^{+}+\left(NSB_{j}^{M}\right)^{+}+\left(NSB_{j}^{U}\right)^{+}}{3}\end{matrix}\right)}\end{matrix}\right)$$

$$+\sum_{j=1}^{n} s\omega_j \left(\left(\frac{NSC_{ij}^{L}+NSC_{ij}^{M}+NSC_{ij}^{U}}{3} \right) \ln \frac{\frac{NSC_{ij}^{L}+NSC_{ij}^{M}+NSC_{ij}^{U}}{3}}{\rho \left(\frac{NSC_{ij}^{L}+NSC_{ij}^{M}+NSC_{ij}^{U}}{3} + \frac{\left(NSC_{j}^{L}\right)^{+}+\left(NSC_{j}^{M}\right)^{+}+\left(NSC_{j}^{U}\right)^{+}}{3} \right)} + \left(1-\frac{NSC_{ij}^{L}+NSC_{ij}^{M}+NSC_{ij}^{U}}{3} \right) \ln \frac{1-\frac{NSC_{ij}^{L}+NSC_{ij}^{M}+NSC_{ij}^{U}}{3}}{1-\rho \left(\frac{NSC_{ij}^{L}+NSC_{ij}^{M}+NSC_{ij}^{U}}{3} + \frac{\left(NSC_{j}^{L}\right)^{+}+\left(NSC_{j}^{M}\right)^{+}+\left(NSC_{j}^{U}\right)^{+}}{3} \right)} \right) \tag{23}$$

Step 6. In light with $TFNNCE\left(NSTFNN_i, TFNNPIS\right)$, the smaller $TFNNCE\left(NSTFNN_i, TFNNPIS\right)$, the better alternative is.

V. Numerical Example and Comparative Analysis

A. Numerical Example

The current state of urban park construction in China is far from ideal, reflecting a relatively outdated urban planning concept compared to that of developed countries. Traditionally, urban parks in China have been designed with a focus on improving urban ecology, providing recreational spaces for citizens, beautifying the city, and showcasing cultural heritage. However, these designs often fail to prioritize fostering environmental awareness among the public. Influenced by commercialization and an anthropocentric mindset, some designers overly emphasize the entertainment and leisure functions of parks. This approach inadvertently reinforces self-centered tendencies, creating false needs that lead to a distorted relationship with nature. For example, excessive consumption, thrill-seeking, and superficial curiosity reduce parks to tools for personal gratification rather than spaces for genuine connection with the natural world. This perception of nature as a mere resource to be exploited is fundamentally flawed and causes significant harm to both the environment and humanity. This erroneous perspective exacerbates critical issues such as environmental pollution, climate change, and ecological crises. On a societal level, it deepens anthropocentric attitudes, making it harder for individuals to break free from self-centered worldviews. Without a true connection to nature, people struggle to form meaningful relationships with the natural world, fail to develop genuine respect for life, and lose the capacity to appreciate and protect the environment. A lack of harmony with nature ultimately undermines one's ability to connect authentically with other humans and society as a whole. From this analysis, it is evident that achieving harmonious coexistence between humans and nature, rooted in a genuine awareness of environmental protection, is a fundamental human necessity. Moving forward, whether renovating existing urban parks or constructing new ones, it is imperative to adopt advanced planning concepts such as "harmonious coexistence between humans and nature," "sustainable development," and "ecological cities." Urban parks should be designed to guide people out of self-centeredness, encouraging them to deeply connect with nature on an emotional and spiritual level. Parks must inspire individuals to embrace, integrate with, understand, appreciate, and love nature. Only through such a transformation can urban parks fulfill their role in fostering a profound respect for life and a commitment to protecting the natural world. The quality evaluation of urban park environmental design is MADM. In this section, numerical example for quality evaluation of urban park environmental design is executed through TFNN-CE approach. Five urban park environmental design schemes $SX_i\left(i=1,2,3,4,5\right)$ are assessed with different attributes:

1) SG_1 is functional design which forms the foundation of the evaluation, focusing on the rationality and practicality of park facilities. This includes whether the functional zones are clearly defined, whether facilities such as pathways, seating, drinking fountains, and restrooms are complete, and whether accessibility features for people with disabilities are adequately implemented. Additionally, the connectivity of the park, such as the placement of entrances and parking areas, plays a crucial role in its functionality.

2) SG_2 is aesthetic and artistic appeal which determines the park's visual attractiveness and uniqueness. Excellent landscape design should achieve harmony between elements such as vegetation, architecture, and water features while incorporating local cultural or historical symbols to showcase a distinctive artistic style. The color scheme should be natural and harmonious, with diverse seasonal vegetation changes. Furthermore, artistic installations (e.g., sculptures, fountains) and lighting design can enhance the overall beauty of the park.

3) SG_3 is Ecological quality which reflects the park's environmental sustainability. Key indicators include vegetation

coverage and biodiversity, as well as the cleanliness of water bodies, air quality, and noise pollution levels. Whether the park incorporates sustainable design concepts, such as rainwater collection systems or energy-saving facilities, is also an important aspect of modern urban parks.

4) SG_4 is user experience and comfort which focus on the actual usability of the park. The spatial layout should avoid overcrowding or overly desolate areas, and facilities should meet the needs of diverse groups (e.g., playgrounds for children, fitness equipment for seniors). Additionally, safety and privacy measures, such as surveillance systems and proper lighting, greatly influence user satisfaction.

The TFNN-CE approach is executed for TFNN-MADM to select the best urban park environmental design schemes.

Step 1. Execute the $STFNN = \left[STFNN_{ij} \right]_{5\times4}$ (Table I).

TABLE I. THE TFNN INFORMATION

	SG_1
SX_1	((0.12,0.45,0.78), (0.23,0.47,0.81), (0.14,0.36,0.69))
SX_2	((0.14,0.48,0.80), (0.22,0.46,0.79), (0.12,0.37,0.65))
SX_3	((0.16,0.44,0.79), (0.28,0.50,0.81), (0.15,0.34,0.68))
SX_4	((0.13,0.47,0.77), (0.23,0.51,0.82), (0.14,0.37,0.67))
SX_5	((0.15,0.46,0.81), (0.27,0.54,0.90), (0.12,0.35,0.64))
	SG_2
SX_1	((0.31,0.53,0.85), (0.24,0.43,0.76), (0.11,0.30,0.59))
SX_2	((0.29,0.52,0.83), (0.25,0.40,0.74), (0.10,0.31,0.57))
SX_3	((0.32,0.55,0.86), (0.26,0.43,0.78), (0.13,0.39,0.62))
SX_4	((0.30,0.49,0.88), (0.20,0.42,0.75), (0.12,0.34,0.61))
SX_5	((0.33,0.56,0.89), (0.21,0.44,0.73), (0.11,0.30,0.58))
	SG_3
SX_1	((0.19,0.48,0.74), (0.21,0.44,0.77), (0.15,0.39,0.63))
SX_2	((0.18,0.47,0.75), (0.20,0.41,0.72), (0.16,0.38,0.64))
SX_3	((0.17,0.46,0.73), (0.23,0.49,0.80), (0.11,0.36,0.60))
SX_4	((0.18,0.45,0.72), (0.24,0.48,0.79), (0.15,0.38,0.63))
SX_5	((0.19,0.43,0.76), (0.25,0.47,0.78), (0.14,0.39,0.62))
	SG_4
SX_1	((0.27,0.50,0.82), (0.18,0.42,0.66), (0.13,0.35,0.58))
SX_2	((0.26,0.54,0.87), (0.19,0.45,0.69), (0.13,0.33,0.56))
SX_3	((0.25,0.53,0.84), (0.21,0.44,0.71), (0.12,0.32,0.52))
SX_4	((0.28,0.52,0.85), (0.22,0.46,0.70), (0.10,0.31,0.55))
SX_5	((0.29,0.51,0.83), (0.20,0.40,0.68), (0.13,0.32,0.53))

Step 2. Normalize the $STFNN = \left[TFNN_{ij} \right]_{5\times4}$ to $NSTFNN = \left[NSTFNN_{ij} \right]_{5\times4}$ (Table II).

TABLE II. THE NORMALIZED TFNNS

	SG_1
SX_1	((0.12,0.45,0.78), (0.23,0.47,0.81), (0.14,0.36,0.69))
SX_2	((0.14,0.48,0.80), (0.22,0.46,0.79), (0.12,0.37,0.65))
SX_3	((0.16,0.44,0.79), (0.28,0.50,0.81), (0.15,0.34,0.68))
SX_4	((0.13,0.47,0.77), (0.23,0.51,0.82), (0.14,0.37,0.67))
SX_5	((0.15,0.46,0.81), (0.27,0.54,0.90), (0.12,0.35,0.64))
	SG_2
SX_1	((0.31,0.53,0.85), (0.24,0.43,0.76), (0.11,0.30,0.59))
SX_2	((0.29,0.52,0.83), (0.25,0.40,0.74), (0.10,0.31,0.57))
SX_3	((0.32,0.55,0.86), (0.26,0.43,0.78), (0.13,0.39,0.62))
SX_4	((0.30,0.49,0.88), (0.20,0.42,0.75), (0.12,0.34,0.61))
SX_5	((0.33,0.56,0.89), (0.21,0.44,0.73), (0.11,0.30,0.58))
	SG_3
SX_1	((0.19,0.48,0.74), (0.21,0.44,0.77), (0.15,0.39,0.63))
SX_2	((0.18,0.47,0.75), (0.20,0.41,0.72), (0.16,0.38,0.64))
SX_3	((0.17,0.46,0.73), (0.23,0.49,0.80), (0.11,0.36,0.60))
SX_4	((0.18,0.45,0.72), (0.24,0.48,0.79), (0.15,0.38,0.63))
SX_5	((0.19,0.43,0.76), (0.25,0.47,0.78), (0.14,0.39,0.62))
	SG_4
SX_1	((0.27,0.50,0.82), (0.18,0.42,0.66), (0.13,0.35,0.58))
SX_2	((0.26,0.54,0.87), (0.19,0.45,0.69), (0.13,0.33,0.56))
SX_3	((0.25,0.53,0.84), (0.21,0.44,0.71), (0.12,0.32,0.52))
SX_4	((0.28,0.52,0.85), (0.22,0.46,0.70), (0.10,0.31,0.55))
SX_5	((0.29,0.51,0.83), (0.20,0.40,0.68), (0.13,0.32,0.53))

Step 3. Execute the TFNNPIS (Table III).

TABLE III. TFNNPIS

	TFNNPIS
SG_1	((0.16,0.44,0.79), (0.28,0.50,0.81), (0.15,0.34,0.68))
SG_2	((0.33,0.56,0.89), (0.21,0.44,0.73), (0.11,0.30,0.58))
SG_3	((0.19,0.43,0.76), (0.25,0.47,0.78), (0.14,0.39,0.62))
SG_4	((0.29,0.51,0.83), (0.20,0.40,0.68), (0.13,0.32,0.53))

Step 4. Execute the $TFNN\text{-}CE\left(NSTFNN_i, TFNNPIS \right)$ (Table IV)and $\rho = 0.6$.

TABLE IV. THE $TFNN\text{-}CE\left(NSTFNN_i, TFNNPIS \right)$

Alternatives	$TFNN\text{-}CE\left(NSTFNN_i, TFNNPIS \right)$
TFNN-CE$\left(NSTFNN_1, TFNNPIS \right)$	0.5840
TFNN-CE$\left(NSTFNN_2, TFNNPIS \right)$	0.6421
TFNN-CE$\left(NSTFNN_3, TFNNPIS \right)$	0.4276
TFNN-CE$\left(NSTFNN_4, TFNNPIS \right)$	0.4933

TFNN-CE$\left(NSTFNN_5, TFNNPIS\right)$	0.6907

Step 5. In light with $TFNN\text{-}CE\left(NTFNN_i, TFNNPIS\right)$, the order is: $SX_3 > SX_4 > SX_1 > SX_2 > SX_5$ and SX_3 is optimal urban park environmental design scheme.

B. Compare with Some Existing Approaches

Then, the TFNN-CE approach is compared with TFNNWA approach [56], TFNNWG approach [56], TFNN-VIKOR approach [73], TFNN-MABAC approach [74], TFNN-EDAS approach [75], TFNN-GRA approach [76]. The order of different approaches is executed in Table V.

TABLE V. ORDER OF DIFFERENT APPROACHES

	Order
TFNNWA approach[56]	$SX_3 > SX_4 > SX_1 > SX_2 > SX_5$
TFNNWG approach[56]	$SX_3 > SX_4 > SX_2 > SX_1 > SX_5$
TFNN-VIKOR approach [73]	$SX_3 > SX_4 > SX_1 > SX_2 > SX_5$
TFNN-MABAC approach [74]	$SX_3 > SX_4 > SX_1 > SX_2 > SX_5$
TFNN-EDAS approach [75]	$SX_3 > SX_4 > SX_1 > SX_2 > SX_5$
TFNN-GRA approach [76]	$SX_3 > SX_4 > SX_1 > SX_2 > SX_5$
TFNN-CE approach	$SX_3 > SX_4 > SX_1 > SX_2 > SX_5$

In accordance with WS coefficients [77, 78], the WS coefficient between TFNNWA approach [56], TFNNWG approach [56], TFNN-VIKOR approach [73], TFNN-MABAC approach [74], TFNN-EDAS approach [75], TFNN-GRA approach [76] and the proposed TFNN-CE approach is 1.0000, 0.8437, 1.0000, 1.0000, 1.0000, 1.0000. The WS coefficient shows the order of TFNN-CE approach are same to order of TFNNWA approach [56], TFNN-VIKOR approach [73], TFNN-MABAC approach [74], TFNN-EDAS approach [75], TFNN-GRA approach [76]; the WS coefficient shows order of TFNN-CE approach are slightly different to order of TFNNWG approach [56]. Furthermore, the major advantages of TFNN-CE approach are executed: TFNN-CE approach could reflect the uncertainty and has strong differentiation ability and could overcome the design defects of existing CE approach. The major limits of TFNN-CE approach didn't mention the psychological behavior of DMs.

VI. CONCLUSION

The evaluation of urban park environmental design quality focuses on four key aspects: functionality, aesthetics, ecology, and user experience. Functionality assesses whether the park's facilities are practical, accessible, and cater to diverse needs, such as clear zoning for recreation, rest, and activities, as well as the inclusion of essential amenities and barrier-free designs. Aesthetics examines the visual appeal, harmony of landscape elements, and integration of cultural or artistic features, ensuring the park is both attractive and reflective of local identity. Ecological quality evaluates sustainability, including vegetation coverage, biodiversity, water and air quality, and the use of eco-friendly materials or systems. Finally, user experience measures comfort, safety, and inclusivity, considering whether the park meets the needs of various groups, maintains a balance between crowdedness and tranquility, and provides spaces for relaxation and interaction. A high-quality urban park balances these elements, fostering harmony between humans and nature while enhancing public well-being and environmental awareness. The quality evaluation of urban park environmental design could be attributed to MADM problem. In this study, TFNN-CE approach is executed under TFNSs. Then, the entropy is employed to execute the weight under TFNSs and TFNN-CE approach is executed for MADM under TFNSs. Finally, numerical example for quality evaluation of urban park environmental design is executed the advantages of TFNN-CE approach through different comparisons. The major contributions of this study could be executed: (1) entropy is employed to execute the weight under TFNSs; (2) TFNN-CE approach is executed under TFNSs; (3) TFNN-CE approach is put forward for MADM under TFNSs; (4) numerical example for quality evaluation of urban park environmental design is executed the advantages of TFNN-CE approach through different comparisons.

There may be possible study limitations, which could be executed in future research: (1) TFNN-CE approach doesn't take into account the irrational state of DMs, and it may be worthwhile research point to execute prospect theory [79-82] for quality evaluation of urban park environmental design under TFNSs; (2) It may be meaningful to avoid regretful MADM, and it could also worthwhile research point to execute regret theory [83-86] for quality evaluation of urban park environmental design under TFNSs.

ACKNOWLEDGMENT

The work was supported by the 2023 Anhui Philosophy and Social Science Planning Project "Research on the Digital Protection and Application Path of Industrial Heritage in Anhui River Basin" under Grant No. AHSKYY2023D040.

REFERENCES

[1] L.J. Xing, Y.F. Liu, X.J. Liu, X.J. Wei, Y. Mao, Spatio-temporal disparity between demand and supply of park green space service in urban area of wuhan from 2000 to 2014, Habitat International, 71 (2018) 49-59.

[2] S.H. Guo, G.G. Yang, T. Pei, T. Ma, C. Song, H. Shu, Y.Y. Du, C.H. Zhou, Analysis of factors affecting urban park service area in beijing: Perspectives from multi-source geographic data, Landscape and Urban Planning, 181 (2019) 103-117.

[3] Y.Q. Jiang, G.L. Huang, B. Fisher, Air quality, human behavior and urban park visit: A case study in beijing, Journal of Cleaner Production, 240 (2019) 7.

[4] M. Blaszczy, M. Suchocka, M. Wojnowska-Heciak, M. Muszynska, Quality of urban parks in the perception of city residents with mobility difficulties, Peerj, 8 (2020) 25.

[5] C.X. Chen, W.J. Luo, H.W. Li, D.T. Zhang, N. Kang, X.H. Yang, Y. Xia, Impact of perception of green space for health promotion on willingness to use parks and actual use among young urban residents, International Journal of Environmental Research and Public Health, 17 (2020) 21.

[6] S.L. Chen, O. Sleipness, Y.N. Xu, K. Park, K. Christensen, A systematic review of alternative protocols for evaluating non-spatial dimensions of urban parks, Urban Forestry & Urban Greening, 53 (2020) 15.

[7] S. Fares, A. Conte, A. Alivernini, F. Chianucci, M. Grotti, I. Zappitelli, F. Petrella, P. Corona, Testing removal of carbon dioxide, ozone, and

atmospheric particles by urban parks in italy, Environmental Science & Technology, 54 (2020) 14910-14922.

[8] K. Herman, M. Rodgers, From tactical urbanism action to institutionalised urban planning and educational tool: The evolution of park(ing) day, Land, 9 (2020) 19.

[9] A. Jahani, M. Saffariha, Aesthetic preference and mental restoration prediction in urban parks: An application of environmental modeling approach, Urban Forestry & Urban Greening, 54 (2020) 14.

[10] D. Ariyani, B. Sulistyantara, T. Budiarti, Formulation of design concept of urban park using butterflies as a good urban environment bio-indicator, in: 3rd International Symposium for Sustainable Landscape Development (ISSLD), Iop Publishing Ltd, Bogor, INDONESIA, 2017.

[11] P. Carinanos, M. Casares-Porcel, C.D. de la Guardia, M.J. Aira, J. Belmonte, M. Boi, B. Elvira-Rendueles, C. De Linares, S. Fernandez-Rodriguez, J.M. Maya-Manzano, R. Perez-Badia, D. Rodriguez-de la Cruz, F.J. Rodriguez-Rajo, J. Rojo-Ubeda, C. Romero-Zarco, E. Sanchez-Reyes, J. Sanchez-Sanchez, J. Rojo-Ubeda, C. Romero-Zarco, E. Sanchez-Reyes, J. Sanchez-Sanchez, R. Tormo-Molina, A.M.V. Maray, Assessing allergenicity in urban parks: A nature-based solution to reduce the impact on public health, Environmental Research, 155 (2017) 219-227.

[12] N. Torabi, J. Lindsay, J. Smith, L.A. Khor, O. Sainsbury, Widening the lens: Understanding urban parks as a network, Cities, 98 (2020) 12.

[13] X.R. Yang, X.W. Tan, C.W. Chen, Y.P. Wang, The influence of urban park characteristics on bird diversity in nanjing, china, Avian Research, 11 (2020) 9.

[14] E. van Vliet, G. Dane, M. Weijs-Perree, E. van Leeuwen, M. van Dinter, P. van den Berg, A. Borgers, K. Chamilothori, The influence of urban park attributes on user preferences: Evaluation of virtual parks in an online stated-choice experiment, International Journal of Environmental Research and Public Health, 18 (2021) 20.

[15] G.Q. Di, J.L. Xiang, Y. Yao, C. Chen, Q.H. Lin, Develop a public response model of soundscape for urban landscape garden parks, Urban Ecosystems, 25 (2022) 453-463.

[16] Z.Q. Li, Q. Liu, Y.X. Zhang, K. Yan, Y.Y. Yan, P. Xu, Characteristics of urban parks in chengdu and their relation to public behaviour and preferences, Sustainability, 14 (2022) 16.

[17] X.Y. Ren, C.H. Guan, Evaluating geographic and social inequity of urban parks in shanghai through mobile phone-derived human activities, Urban Forestry & Urban Greening, 76 (2022) 10.

[18] B. Srdjevic, Z. Srdjevic, K.M. Reynolds, M. Lakicevic, S. Zdero, Using analytic hierarchy process and best-worst method in group evaluation of urban park quality, Forests, 13 (2022) 17.

[19] R. Sabouri, J. Breuste, A. Rahimi, A planted forest in the mountain steppe of tabriz, iran: Visitor's perceptions of eynali urban woodland park, Frontiers in Forests and Global Change, 6 (2023) 15.

[20] B.B. Lin, R.A. Fuller, R. Bush, K.J. Gaston, D.F. Shanahan, Opportunity or orientation? Who uses urban parks and why, Plos One, 9 (2014) 7.

[21] F.Z. Li, N. Yao, D.M. Liu, W.P. Liu, Y.H. Sun, W.W. Cheng, X. Li, X.L. Wang, Y.N. Zhao, Explore the recreational service of large urban parks and its influential factors in city clusters - experiments from 11 cities in the beijing-tianjin-hebei region, Journal of Cleaner Production, 314 (2021) 16.

[22] F. Liu, J.B. Lu, Ecological engineering approaches to restoring the aquatic biological community of an urban pond ecosystem and its effects on water quality - a case study of the urban xixi national wetland park in china, Knowledge and Management of Aquatic Ecosystems, (2021) 12.

[23] Y.S. Wu, Y. Li, X. Gao, J. Pan, N. Wang, Y.W. Cheng, C.G. Yang, Y.C. Yang, Sewage treatment system planning for dianchi urban wetland park in kunming, in: 5th International Conference on Advances in Energy, Environment and Chemical Science (AEECS), E D P Sciences, Electr Network, 2021.

[24] H. Yin, W.Y. Liang, X. Cao, Self-purification mode of still-water ponds in urban parks based on in situ ecological remediation design, Land, 11 (2022) 25.

[25] M. Bottero, M. Bravi, C. Caprioli, F. Dell'Anna, Combining revealed and stated preferences to design a new urban park in a metropolitan area of north-western italy, Ecological Modelling, 483 (2023) 11.

[26] W.J. Mitsch, L. Zhang, L.N. Griffiths, J. Bays, Contrasting two urban wetland parks created for improving habitat and downstream water quality, Ecological Engineering, 192 (2023) 18.

[27] Q.Y. Liu, Z.P. Zhu, Z.X. Zhuo, S.P. Huang, C.Y. Zhang, X.B. Shen, C.C.K. van den Bosch, Q.T. Huang, S.R. Lan, Relationships between residents' ratings of place attachment and the restorative potential of natural and urban park settings, Urban Forestry & Urban Greening, 62 (2021) 10.

[28] L.J. Xing, Y. Mao, Y.F. Liu, X.J. Wei, Using multiscenario assessment framework to measure access to urban parks for refuge in reference to survival justice, Journal of Urban Planning and Development, 147 (2021) 12.

[29] D.L. Zhang, S.F. Ma, J.H. Fan, D.X. Xie, H.Y. Jiang, G.W. Wang, Assessing spatial equity in urban park accessibility: An improve two-step catchment area method from the perspective of 15-mintue city concept, Sustainable Cities and Society, 98 (2023) 15.

[30] Y. Zhang, F.J. Rao, J. Xue, D.Y. Lai, Dependence of urban park visits on thermal environment and air quality, Urban Forestry & Urban Greening, 79 (2023) 11.

[31] Y. Zheng, S. Wang, J.L. Zhu, S. Huang, L.L. Cheng, J.W. Dong, Y.X. Sun, A comprehensive evaluation of supply and demand in urban parks along "luck greenway" in fuzhou, Sustainability, 15 (2023) 19.

[32] S.F. Liu, S.H. Tan, Building a new framework for urban parking facilities research with quality improvement: The case of chongqing, china, International Journal of Environmental Research and Public Health, 20 (2023) 24.

[33] N.X. Mou, J.H. Wang, Y.H. Zheng, L.X. Zhang, T. Makkonen, T.F. Yang, J.Q. Niu, Flowers as attractions in urban parks: Evidence from social media data, Urban Forestry & Urban Greening, 82 (2023) 12.

[34] L.L. Tian, D. Winterbottom, J.J. Liu, Soundscape optimization strategies based on landscape elements in urban parks: A case study of greenlake park in kunming, Sustainability, 15 (2023) 14.

[35] X. Xiao, J. Gao, J.Y. Lu, P.Z. Li, Y.L. Zhang, Social carrying capacity and emotion dynamics in urban national parks during the covid-19 pandemic, Journal of Outdoor Recreation and Tourism-Research Planning and Management, 41 (2023) 10.

[36] M.W. Yang, R.W. Wu, Z.Y. Bao, H. Yan, X.E. Nan, Y.X. Luo, T.F. Dai, Effects of urban park environmental factors on landscape preference based on spatiotemporal distribution characteristics of visitors, Forests, 14 (2023) 18.

[37] H.I. Jo, J.Y. Jeon, Overall environmental assessment in urban parks: Modelling audio-visual interaction with a structural equation model based on soundscape and landscape indices, Building and Environment, 204 (2021) 21.

[38] Y.L. Guo, X.M. Jiang, L.F. Zhang, H. Zhang, Z.Q. Jiang, Effects of sound source landscape in urban forest park on alleviating mental stress of visitors: Evidence from huolu mountain forest park, guangzhou, Sustainability, 14 (2022) 22.

[39] Z. Yin, Y.X. Zhang, K.M. Ma, Evaluation of pm2.5 retention capacity and structural optimization of urban park green spaces in beijing, Forests, 13 (2022) 13.

[40] T. Mahmood, Z. Ali, M. Naeem, Aggregation operators and critic-vikor method for confidence complex q-rung orthopair normal fuzzy information and their applications, Caai Transactions on Intelligence Technology, 8 (2023) 40-63.

[41] T. Mahmood, Z. Ali, D. Pamucar, Applications to biogas-plant implementation problem based on type-2 picture fuzzy matrix game under new minkowski type measures, Journal of Intelligent & Fuzzy Systems, 44 (2023) 6545-6571.

[42] T. Mahmood, U.U. Rehman, Z. Ali, I. Haleemzai, Analysis of topsis techniques based on bipolar complex fuzzy n-soft setting and their applications in decision-making problems, Caai Transactions on Intelligence Technology, (2023) 22.

[43] T. Mahmood, U. ur Rehman, Z. Ali, Analysis and application of aczel-alsina aggregation operators based on bipolar complex fuzzy information in multiple attribute decision making, Information Sciences, 619 (2023) 817-833.

[44] K. Zhang, Y.J. Xie, S.A. Noorkhah, M. Imeni, S.K. Das, Neutrosophic management evaluation of insurance companies by a hybrid todim-bsc

method: A case study in private insurance companies, Management Decision, 61 (2023) 363-381.

[45] R. Verma, E. Alvarez-Miranda, Group decision-making method based on advanced aggregation operators with entropy and divergence measures under 2-tuple linguistic pythagorean fuzzy environment, Expert Systems with Applications, 231 (2023) 32.

[46] P.H. Tsai, C.J. Chen, W.H. Hsiao, C.T. Lin, Factors influencing the consumers? Behavioural intention to use online food delivery service: Empirical evidence from taiwan, Journal of Retailing and Consumer Services, 73 (2023) 17.

[47] N. Thakkar, P. Paliwal, Multi-criteria valuation for sustainable autonomous microgrid planning: A comparative analysis of technology mix with different madm techniques, Iranian Journal of Science and Technology-Transactions of Electrical Engineering, (2023) 24.

[48] T. Senapati, An aczel-alsina aggregation-based outranking method for multiple attribute decision-making using single-valued neutrosophic numbers, Complex & Intelligent Systems, (2023) 15.

[49] F. Lei, Q. Cai, G.W. Wei, Z.W. Mo, Y.F. Guo, Probabilistic double hierarchy linguistic madm for location selection of new energy electric vehicle charging stations based on the msm operators, Journal of Intelligent & Fuzzy Systems, 44 (2023) 5195-5216.

[50] F. Lei, Q. Cai, N.N. Liao, G.W. Wei, Y. He, J. Wu, C. Wei, Todim-vikor method based on hybrid weighted distance under probabilistic uncertain linguistic information and its application in medical logistics center site selection, Soft Computing, 27 (2023) 8541-8559.

[51] L.A. Zadeh, Fuzzy sets, Information and Control, 8 (1965) 338-353.

[52] D. Filev, R.R. Yager, Analytic properties of maximum-entropy owa operators, Information Sciences, 85 (1995) 11-27.

[53] A. Smith, B. Page, K. Duffy, R. Slotow, Using maximum entropy modeling to predict the potential distributions of large trees for conservation planning, Ecosphere, 3 (2012).

[54] X.L. Zhang, H.Z. Huang, Z.L. Wang, N.C. Xiao, Y.F. Li, Uncertainty analysis method based on a combination of the maximum entropy principle and the point estimation method, Eksploatacja I Niezawodnosc-Maintenance and Reliability, 14 (2012) 114-119.

[55] R.R. Yager, On the maximum entropy negation of a probability distribution, Ieee Transactions on Fuzzy Systems, 23 (2015) 1899-1902.

[56] P. Biswas, S. Pramanik, B.C. Giri, Aggregation of triangular fuzzy neutrosophic set information and its application to multi-attribute decision making, Neutrosophic Sets and Systems, 12 (2016) 20-40.

[57] X. Shang, W. Jiang, A note on fuzzy information measures, Pattern Recognition Letters, 18 (1997) 425-432.

[58] M. Foumani, A. Moeini, M. Haythorpe, K. Smith-Miles, A cross-entropy method for optimising robotic automated storage and retrieval systems, International Journal of Production Research, 56 (2018) 6450-6472.

[59] Y.M. Wang, H. Yang, K.Y. Qin, The consistency between cross-entropy and distance measures in fuzzy sets, Symmetry-Basel, 11 (2019) 11.

[60] P.D. Liu, Z. Ali, T. Mahmood, The distance measures and cross-entropy based on complex fuzzy sets and their application in decision making, Journal of Intelligent & Fuzzy Systems, 39 (2020) 3351-3374.

[61] P.D. Liu, T. Mahmood, Z. Ali, The cross-entropy and improved distance measures for complex q-rung orthopair hesitant fuzzy sets and their applications in multi-criteria decision-making, Complex & Intelligent Systems, 8 (2022) 1167-1186.

[62] W. Yang, Y.F. Pang, T-spherical fuzzy oreste method based on cross-entropy measures and its application in multiple attribute decision-making, Soft Computing, 26 (2022) 10371–10387.

[63] L. Li, Cross-entropy method for efficiency evaluation of integrated development of agriculture and tourism to promote rural revitalization under the triangular fuzzy neutrosophic sets, Journal of Intelligent & Fuzzy Systems, 44 (2023) 6151-6161.

[64] X.R. Zhu, Modified cross-entropy method for evaluating the higher vocational education management quality under the fuzzy number intuitionistic fuzzy sets, Journal of Intelligent & Fuzzy Systems, 45 (2023) 3461-3471.

[65] C.E. Shannon, A mathematical theory of communication, Bell System Technical Journal, 27 (1948) 379-423.

[66] J. Lin, Divergence measures based on the shannon entropy, IEEE Transactions on Information theory, 37 (1991) 145-151.

[67] D. Bhandari, N.R. Pal, Some new information measures for fuzzy sets, Information Sciences, 67 (1993) 209-228.

[68] Tehreem, A. Hussain, A. Alsanad, M.A.A. Mosleh, Spherical cubic fuzzy extended topsis method and its application in multicriteria decision-making, Mathematical Problems in Engineering, 2021 (2021) 14.

[69] R.P. Tan, W.D. Zhang, S.Q. Chen, Decision-making method based on grey relation analysis and trapezoidal fuzzy neutrosophic numbers under double incomplete information and its application in typhoon disaster assessment, Ieee Access, 8 (2020) 3606-3628.

[70] J.H. Kim, B.S. Ahn, The hierarchical vikor method with incomplete information: Supplier selection problem, Sustainability, 12 (2020) 15.

[71] M.S.A. Khan, F. Khan, J. Lemley, S. Abdullah, F. Hussain, Extended topsis method based on pythagorean cubic fuzzy multi-criteria decision making with incomplete weight information, Journal of Intelligent & Fuzzy Systems, 38 (2020) 2285-2296.

[72] P.D. Liu, W.Q. Liu, Multiple-attribute group decision-making method of linguistic q-rung orthopair fuzzy power muirhead mean operators based on entropy weight, International Journal of Intelligent Systems, 34 (2019) 1755-1794.

[73] J. Wang, G.W. Wei, M. Lu, An extended vikor method for multiple criteria group decision making with triangular fuzzy neutrosophic numbers, Symmetry-Basel, 10 (2018) 15.

[74] I. Irvanizam, N.N. Zi, R. Zuhra, A. Amrusi, H. Sofyan, An extended mabac method based on triangular fuzzy neutrosophic numbers for multiple-criteria group decision making problems, Axioms, 9 (2020) 18.

[75] J.P. Fan, X.F. Jia, M.Q. Wu, A new multi-criteria group decision model based on single-valued triangular neutrosophic sets and edas method, Journal of Intelligent & Fuzzy Systems, 38 (2020) 2089-2102.

[76] B. Xie, Modified gra methodology for madm under triangular fuzzy neutrosophic sets and applications to blended teaching effect evaluation of college english courses, Soft Computing, (2023) https://doi.org/10.1007/s00500-00023-08891-00506.

[77] W. Salabun, K. Urbaniak, A new coefficient of rankings similarity in decision-making problems, in: 20th Annual International Conference on Computational Science (ICCS), Springer International Publishing Ag, Amsterdam, NETHERLANDS, 2020, pp. 632-645.

[78] W. Sałabun, J. Wątróbski, A. Shekhovtsov, Are mcda methods benchmarkable? A comparative study of topsis, vikor, copras, and promethee ii methods, Symmetry, 12 (2020) 1549.

[79] Tversky, K. Amos, Prospect theory: An analysis of decision under risk, Econometrica, 47 (1979) 263-291.

[80] A. Tversky, D. Kahneman, Prospect theory: An analysis of decision under risk, Econometrica, 47 (1979) 263-291.

[81] T. Kahneman, Advances in prospect theory: Cumulative representation of uncertainty. Journal of risk and uncertainty, 5 (1992) 297–323.

[82] L. Gomes, L.A.D. Rangel, An application of the todim method to the multicriteria rental evaluation of residential properties, European Journal of Operational Research, 193 (2009) 204-211.

[83] H. Bleichrodt, A. Cillo, E. Diecidue, A quantitative measurement of regret theory, Management Science, 56 (2010) 161-175.

[84] Y. Lin, Y.M. Wang, S.Q. Chen, Hesitant fuzzy multiattribute matching decision making based on regret theory with uncertain weights, International Journal of Fuzzy Systems, 19 (2017) 955-966.

[85] H. Zhou, J.Q. Wang, H.Y. Zhang, Grey stochastic multi-criteria decision-making based on regret theory and topsis, International Journal of Machine Learning and Cybernetics, 8 (2017) 651-664.

[86] Y. Wang, J.Q. Wang, T.L. Wang, Fuzzy stochastic multi-criteria decision-making methods with interval neutrosophic probability based on regret theory, Journal of Intelligent & Fuzzy Systems, 35 (2018) 2309-2322.

Improved YOLOv11pose for Posture Estimation of Xinjiang Bactrian Camels

Lei Liu, Alifu Kurban,Yi Liu
Department School of Software, Xinjiang University, Urumqi, China

***Abstract*—Automatic pose estimation of camels is crucial for long-term health monitoring in animal husbandry. There is currently less research on camels, and our study has certain practical application value in actual camel farms. Due to the high similarity of camels, this has brought us a huge challenge in pose estimation. This study proposes YOLOv11pose-Camel, a pose estimation algorithm tailored for Bactrian camels. The algorithm enhances feature extraction with a lightweight channel attention mechanism (ECA) and improves detection accuracy through an efficient multi-scale pooling structure (SimSPPF). Additionally, C3k2 modules in the neck are replaced with dynamic convolution blocks (DECA-blocks) to strengthen global feature extraction. We collected a diverse dataset of Bactrian camel images with farm staff assistance and applied data augmentation. The optimized YOLOv11pose model achieved 94.5% accuracy and 94.1% mAP@0.5 on the Xinjiang Bactrian camel dataset, outperforming the baseline by 2.1% and 2.2%, respectively. The model also maintains a good balance between detection speed and efficiency, demonstrating its potential for practical applications in animal husbandry.**

***Keywords*—YOLOv11pose; efficient channel attention; multi-scale pooling structure; DECA-block; Bactrian camel posture estimation； SimSPPF; ECA**

I. INTRODUCTION

With the continuous large-scale development of the modern camel farming industry, their farming methods have gradually changed from traditional free-range farming to cluster farming. The cluster farming method makes the breeders pay more and more attention to the healthy development of camels. Breeders increasingly recognize the critical role of health management in selecting and cultivating high-quality camel breeds. As an essential metric of health management, posture estimation is gaining growing attention. It provides scientific data for complex behavior analysis and health monitoring tasks, extending beyond the scope of traditional detection methods and better adapting to diverse scenarios. By estimating the key points of camels, we can further analyze their behaviors. The external behaviors of camels often reflect their health conditions. For instance, daily postures such as standing or walking can indicate their activity level, as camels tend to reduce activity when ill [1]. However, long-term manual observation and recording of individual information are time-consuming, costly, and subjective. Currently, many researchers use various sensors to monitor animal behavior [2-5]. Nevertheless, wearable sensors have limitations, as they may stress the animals and impact their natural growth and development. In contrast, machine vision offers long-term, non-contact continuous monitoring and has gradually been applied in precision livestock farming to monitor livestock activities and health conditions. With the rapid advancements in image processing [6], pattern recognition [7], and artificial intelligence [8], machine learning applications in China's animal husbandry are also increasing. Machine vision-based camel posture detection facilitates comprehensive analysis and rapid assessment of camel growth and development [9].

The rapid development and widespread application of computer vision technology provide robust technical support for the intelligent, precise, and scientific modernization of livestock farms. The mature implementation of these technologies has laid a solid foundation for the advancement of modern animal husbandry [10]. Through object detection techniques and keypoint detection algorithms, it is possible to extract the key features of animal postures, which can then be connected to form skeletal structures, enabling accurate posture estimation [11]. In the task of pose estimation, our focus is on ensuring the accuracy of keypoint detection. During the detection process, we often encounter various complex problems that can affect the accuracy of keypoint detection. For example, camels have relatively large bodies, and diverse poses, and tend to move in close groups. These problems can cause key points on certain parts of camels to be obscured by other camels or objects. In addition, the deformation of features on camels in different poses can cause key points to blur, thereby affecting the model's detection of skeletal key points. These problems not only affect the accuracy of keypoint localization but also the precision of detection boxes. If the skeletal features extracted from the image are incomplete, accurate pose estimation cannot be performed.

To address the aforementioned challenges, we have proposed a Bihumped Camel pose estimation framework, named YOLOv11pose-Camel, which is designed to effectively process camel information in various forms and environments. Specifically, we first utilize channel attention and multi-scale pooling to enhance sensitivity to detailed features. Secondly, the error under complex postures is further reduced by incorporating our DECA-Block module. The main contributions of this paper are as follows:

1) We used data augmentation techniques to cover a variety of environments with camels in the dataset and applied the improved YOLOv11-pose algorithm to our self-built dataset.

2) We introduced the ECA lightweight channel attention mechanism into the model to enhance the algorithm's multi-scale feature extraction ability. The model effectively reduced

the extraction of non-camel features and focused more on the precise identification of camel skeleton points [12].

3) To further enhance the algorithm's ability to identify skeleton key points, we combined the algorithm with the efficient multi-scale pooling structure SimSPPF. This can effectively reduce the precision error caused by occlusion.

4) We introduced the improved dynamic convolution module DECA-Block into the neck module of the YOLOv11-pose model. This module combines dynamic convolution and an efficient attention mechanism. This design not only reduces the model's weight but also maintains a high accuracy. In particular, the module shows strong adaptability to feature deformation problems.

The algorithm adopted in this study can meet the actual needs for detecting camel skeleton key points. Meanwhile, the method can infer the information of occluded skeleton points. Section II of the paper reviews the current research status on Bactrian camel posture estimation both domestically and internationally, as well as its application in animal husbandry. Section III describes the overall network structure and focuses on the proposed YOLOv11pose-Camel method, which integrates the ECA channel attention mechanism, the multi-scale feature module SimSPPF, and the dynamic convolution module DECA-Block. Section IV introduces the experimental dataset, evaluation metrics, and recognition performance, followed by comparative analyses and experimental results on the custom dataset. Finally, Section V summarizes the main contributions and outlines potential future research directions.

II. Related Work

In recent years, pose estimation based on keypoint detection has significantly advanced image recognition and analysis technologies, achieving notable progress in animal posture detection. In early studies, Fang et al. (2017) introduced a novel multi-person pose estimation framework called RMPE, aimed at enhancing the adaptability of deep convolutional networks to inaccuracies in bounding box localization. This innovation not only improved feature extraction efficiency but also laid the foundation for subsequent developments in multi-person pose estimation. However, since RMPE was primarily designed for humans and camels possess unique body shapes and complex keypoint distributions, it cannot effectively estimate camel poses [13]. In 2018, Zheng et al. incorporated Faster R-CNN into a deep learning framework for pig posture estimation, enabling the detection of five posture changes (standing, sitting, sternal recumbency, ventral recumbency, and lateral recumbency) to reflect the health status of sows [14]. However, the more drastic posture variations in camels make it challenging for a single Faster R-CNN framework to capture their dynamics. Similarly, Song et al. (2018) proposed a model for skeletal extraction based on keypoint prediction in walking states, which performed well in detecting cow walking postures [15]. Comparatively, research focused directly on camel posture estimation remains limited. In 2020, Chen et al. utilized the real-time instance segmentation framework YOLACT for pig body part tracking [16]. However, camels' unique physiological structures, especially the presence of humps, add significant challenges in segmentation and localization. In 2023, Zheng et al. developed the ViT-BERT pose estimation model for rapid posture changes in small animals within dynamic scenes. While this model demonstrated good accuracy and speed in bird testing, it is less suitable for large animals like camels [17]. Also in 2023, Natesan et al. proposed a YOLOv5-based algorithm incorporating adaptive attention mechanisms to estimate occluded key points, improving measurement accuracy and achieving precise livestock behavior recognition [18]. In 2023, Agullo and colleagues combined MobileNetV3 with Vision Transformer (ViT) for animal pose detection, aiming to perform real-time pose estimation with lower computational resources. This approach aligns well with the current needs of automated livestock farms [19]. However, for large animals with complex postures, higher resolution and deeper feature extraction are required, and resource constraints may limit accuracy. In 2022, Wang et al. introduced a dynamic convolution and bidirectional LSTM time series ensemble, which considered long-term posture tracking and showed good performance in complex groups [20]. However, this method has not been fully explored in applications where posture changes frequently or under occlusion conditions. In 2023, Zhao and his colleagues enhanced their model by combining ResNet with LSTM for behavior recognition and pose detection, allowing for the identification of abnormal situations while recognizing poses. This algorithm primarily applies to animals with uniform body types, and the complexity and specificity of camel bodies pose significant challenges for applying this technology to camel pose detection [21]. In 2023, Barney and his research team used deep learning methods for cattle detection and pose estimation. They modified Mask-RCNN to estimate poses in video sequences for each cow and applied the CatBoost gradient boosting algorithm to combine all features, using triple cross-validation to determine accuracy. This method proved highly effective for detecting hoof issues in cattle, which is critical for animal health. However, the same measurement point localization is not suitable for camels, as their body structure is more complex [22]. Additionally, in 2024, Dhivya Mohanavel and others proposed an animal detection early warning system based on deep learning and computer vision technologies for the livestock environment, further highlighting the growing importance of these technologies in the agricultural industry [23]. This research offers valuable insights into the practical application of animal management in real-world agricultural settings; however, it has not been thoroughly explored in the context of precise analysis of bipedal pose characteristics.

The aforementioned research demonstrates the effectiveness of non-contact monitoring methods and highlights significant progress in this field. In the future, non-contact computer vision methods are likely to become mainstream in research, not only effectively reducing labor costs but also promoting animal welfare. Existing technologies have shown certain success in animals such as humans, cattle, and pigs, but camel pose estimation still faces numerous challenges. Due to camels' unique body structure, gregarious nature, and variability, pose estimation becomes particularly difficult. With the development of computer vision technologies [24], there is potential for further research on optimization models focused on camels, which could better

support their morphological evaluation and health monitoring. This would not only enhance the practical value of the model but also provide valuable insights for the application of deep learning and computer vision in agriculture.

III. Research Methodology

The YOLOv11-pose model is a deep convolutional network based on YOLOv11, designed with an end-to-end unified structure rather than the traditional two-branch composition commonly used in pose estimation. This single-stage network performs both keypoint localization and partial affinity field prediction simultaneously, enabling pose estimation through joint training of object detection and keypoint detection. Initially developed for human pose estimation, the model has demonstrated promising results in animal pose estimation as well. The YOLOv11pose-Camel model consists of four main components: input, backbone, neck, and head. The lower-level features provide detailed information about object positions, while higher-level features capture stronger semantic information. In the input stage, multiple convolutional layers (Conv) preprocess the input image to the dimensions required for model training and extract preliminary features. The backbone includes several C3K2 modules, which merge features from different layers and pass the refined features to the SimSPPF module for further multi-scale processing. Subsequently, the ECAttention module enhances the feature extraction process by emphasizing attention mechanisms, capturing the relational information between different spatial regions through feature fusion. This allows the model to focus on critical feature regions, improving detection performance and convergence speed. In the neck, the PAN structure is employed [25], and the DECA-Block module is integrated to enhance the network's capability in multi-scale feature integration. The head network adopts depthwise separable convolution to process the network's outputs and convert them into the required format for the output layer. This design reduces computational overhead without compromising keypoint accuracy, ensuring model stability. The proposed model architecture is illustrated in Fig. 1.

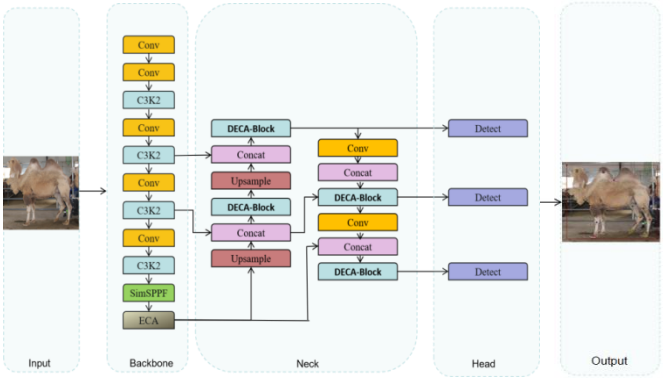


Fig. 1. Architecture of the YOLOv11pose-camel model.

A. ECAttention Mechanism

In deep convolutional neural networks, the primary function of channel attention mechanisms is to enhance the model's focus on significant features, enabling more precise feature extraction and improved performance. As input images pass through convolutional networks, spatial information is gradually embedded into channels. However, repeated spatial compression or channel expansion may lead to the loss of some semantic information. By dynamically assigning higher weights to important features, the model can concentrate more effectively on task-relevant features, thereby enhancing accuracy [26]. Thus, channel attention mechanisms are critical in deep convolutional neural networks. To achieve a balance between detection speed and efficiency, we chose to employ the lightweight Efficient Channel Attention (ECA) mechanism [27]. ECA improves upon the SENet architecture by removing fully connected layers and replacing them with a 1*1 convolution kernel. This introduces a local cross-channel interaction strategy with reduced dimensionality, effectively meeting the requirements for balancing detection accuracy and speed. Specifically, the ECA module adaptively determines the kernel size k for one-dimensional convolution based on the number of input channels. The calculation formula is as follows:

$$F = CNN(I)[k = |\frac{log_2(C)}{\gamma} + \frac{b}{\gamma}|_{odd}] \quad (1)$$

Where C represents the number of channels for the input feature, γ and b is the hyperparameter. The core idea of the ECA module is to introduce a one-dimensional convolution operation into the relationship between input feature channels after determining kernel size k. The ECA attention mechanism first performs global average pooling operations on the input feature maps to generate global information. Then, the module performs one-dimensional convolution processing with kernel size k on the globally pooled information. The activation function maps the weights nonlinearly to obtain the weights for each channel. Finally, these weights are multiplied with the original input feature graph channels to produce output features that incorporate the attention weights. The detailed flow of the attention mechanism is shown in Fig. 2.

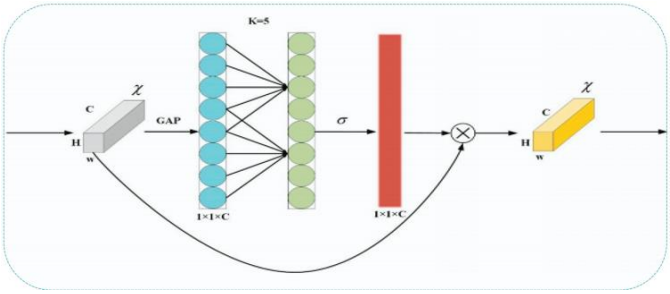


Fig. 2. ECAttention workflow diagram.

B. SimSPPF

SimSPPF (Simplified Space Pyramid Pooling Fast) is an optimized version of the fast version of the space pyramid pooling (SPPF) based on the classic space pyramid pooling (SPP). SPP is an improved version of the classic space pyramid pooling. The feature map is pooled at multiple scales to improve the receptive field in SPP. Although this method has significant effects, the computational cost is high. SPPF reduces redundant content, not only improving the efficiency of feature extraction but also reducing memory usage. The main innovation of SimSPPF is the hierarchical structure, which divides nodes into different layers according to their

size, and the number of nodes in each layer is twice that of the previous layer. SimSPPF can reuse allocated nodes, which reduces the computational cost. In addition, this improvement reduces the dimensionality of the feature network. Specifically, SimSPPF replaces the activation function in the original SPPF with a new activation function (ReLU). This combination of multi-scale pooling methods not only captures multi-level feature information but also ensures a certain detection ability in different scenarios. Fig. 3 shows the structure design of SPP, and Fig. 4 shows SPPF and SimSPPF.

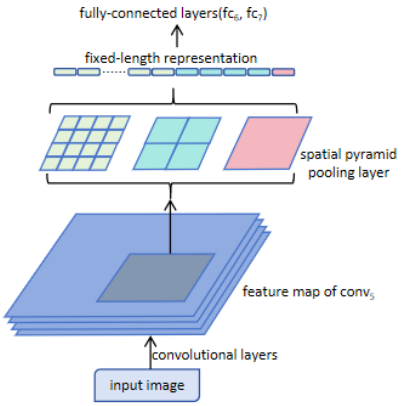


Fig. 3. Structure of SPP.

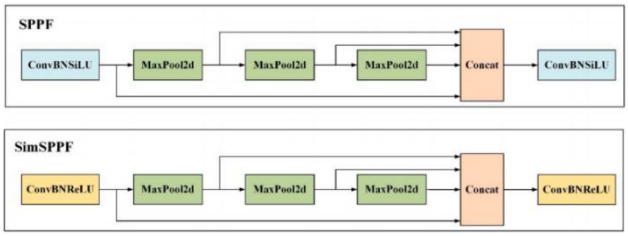


Fig. 4. Structures of SPPF and SimSPPF.

C. DECA-Block

The neck module primarily optimizes and fuses the features from the backbone network, improving the performance of key point detection tasks. When handling features of varying scales and backgrounds, the neck module addresses issues such as feature variation and keypoint occlusion, particularly in dealing with multi-scale features [29].

To enhance the model's understanding of feature details and structures, inspired by feature extraction components, we propose a dynamic convolution module called DECA-Block, designed for keypoint detection of various parts of Xinjiang Bactrian camels. This method helps mitigate keypoint localization errors and feature deformation among camels, adapting effectively to complex environments to balance model accuracy and efficiency. Specifically, the module incorporates the efficient feature extraction capability of the EMAttention mechanism and the deformable convolution network (DCNv3) into the bottleneck structure. EMAttention focuses on critical features and guides their transmission to DCNv3, which adapts to feature variations spatially. This enables the model to better capture key points across different scales and semantic levels, thereby improving detection accuracy. Such direct interaction enhances the attention mechanism's performance on deformed features, providing distinct advantages for complex shapes and occluded key points. Additionally, it reduces redundant computation and improves overall model performance and efficiency. The specific structure of the DECA-Block module is shown in Fig. 5.

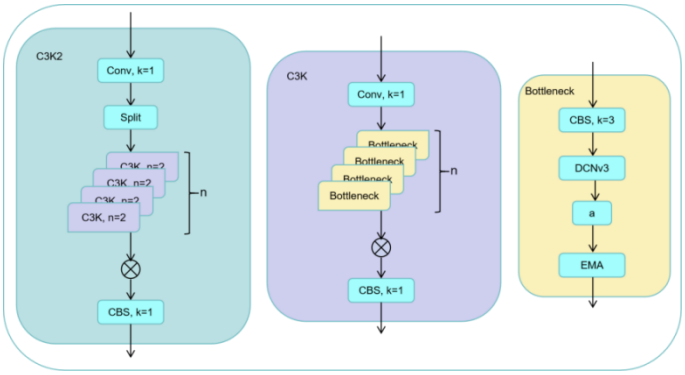


Fig. 5. D3E-C3K2 module.

D. Bactrian Camel Pose Estimation Method

The algorithm analyzes the three-dimensional joint position information contained in the two-dimensional image to achieve the analysis of the target pose. Based on this, the algorithm fully utilizes contextual information to solve the problem of joint overlap. Contextual information includes precise detection of joint positions, modeling of camel pose through structured constraints, and estimation of joint relative position relationships through analysis. Each object possesses its unique anatomical structure, and these structural constraints provide critical prior knowledge to the model. In cases where certain joints are occluded, this prior knowledge effectively aids the model in achieving more accurate pose estimation [30].

$$J = \{j_1, j_2, \dots, j_n\} \quad (2)$$

Where j_n is the position vector of the n-th joint. The structure of the camel pose can be represented by a set of distance constraints between joints.

$$D_{ij} = \| j_i - j_i \|, \forall i, j \in \{1,2, \dots, n\} \quad (3)$$

These distance constraints D_i can be learned from the training data. During the pose estimation process, they are widely used to help the model infer the positions of joints that are not directly observable due to occlusion. YOLOv11-pose extracts feature maps from input images through convolutional neural networks (CNN). Assuming the input image is represented as i, the feature maps extracted by the CNN are denoted as F . These feature maps provide multi-scale structured information, laying the foundation for subsequent pose estimation tasks.

$$F = \mathrm{CNN}(I) \quad (4)$$

The feature maps F contain multi-scale features and rich contextual information of the image, which play a vital role in predicting the positions of skeletal joints. During the prediction of joint locations, the model relies not only on local features but also heavily on global features derived from the overall structure. The rich contextual information further provides comprehensive support. Additionally, the model leverages

multi-level information to accurately analyze pose characteristics. For instance, for certain occluded skeletal key points j_k, the model can utilize the information from other detected key points in the feature map and infer the position by combining global contextual features. Annotated data sets are fed into the model for training, preparing it for subsequent pose estimation tasks. The annotations for Bactrian camels are illustrated in Fig. 6.

$$j_k = f(F, j_1, j_2, \dots, j_{k-1}, j_{k+1}, \dots, j_n) \quad (5)$$

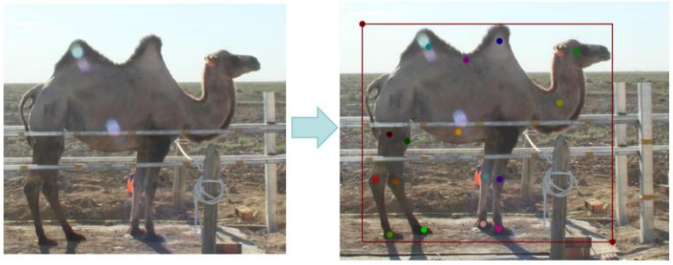


Fig. 6. Annotation of Bactrian camels.

IV. Preparation and Analysis

A. Dataset Preparation

In this study, a Canon 60D camera equipped with an 18-135mm lens was used for data collection. The initial image capture took place in December 2020 in Ziniquanzi Town, Fukang City, Xinjiang Uygur Autonomous Region, where images of Xinjiang Bactrian camels were recorded. Later, in March and April 2021 and July 2021, further collection was carried out, which included various lighting conditions, different angles and distances, and diverse obstruction scenarios, resulting in the acquisition of 90 high-quality images. Considering the limited dataset size, additional data collection was conducted in September 2023 in Keeping County, located in the western Aksu Prefecture of Xinjiang, a region renowned as the "Camel City of Xinjiang." This supplementary collection carried out in collaboration with the local Xintuomilk Group, added approximately 252 effective images to the dataset.

1) Images selection: We constructed a dataset containing 1,084 images for the study of Xinjiang Bactrian camel pose estimation. The data collection was conducted across multiple scenarios and seasons, resulting in images captured under varying lighting conditions influenced by changes in sunlight.

2) Key points selection: We utilized the open-source image annotation tool Labelme [31] for manual annotation of the images. During the annotation process, we focused on creating bounding boxes around Bactrian camels to minimize the area of detection boxes and reduce the influence of unrelated factors. Due to the unique structure of Bactrian camels, we annotated 1 rectangular box ("camel") and 16 key points (A, B, C...P), representing the head, neck, first hump, back, second hump, tail, right hind knee, right hind ankle, right hind hoof, abdomen, right foreleg knee, right foreleg hoof, left foreleg knee, left foreleg hoof, left hind knee, left hind ankle, and left hind hoof. Invisible skeletal key points were labeled as "1." An example of the key point annotations is shown in Fig. 1. These annotations were saved as text files, corresponding to the image filenames, providing a solid foundation for subsequent training.

3) Data augmentation: We introduced various data augmentation techniques [32] in the training set to enhance the model's robustness in different scenarios and its adaptability to complex environments. These techniques included adding random noise, simulating varying levels of occlusion, and performing operations such as mirroring, flipping, rotating, cropping, and stitching at random locations. These methods effectively simulated real-world scenarios the model might encounter. The model not only provides a reliable scientific basis and data support for Bactrian camel morphological evaluation based on pose estimation but also ensures accuracy and comprehensiveness in the evaluation process. Data augmentation examples are shown in Fig. 7.

4) Dataset formatting: To ensure the dataset's representativeness and the reliability of model training, we randomly sampled images collected at different times. These samples were divided into three parts: the training set, validation set, and test set. The training and validation sets comprised 80% of the total dataset, with a 9:1 split between them. The remaining 20% was allocated to the test set for final model evaluation. Special attention was given to ensuring that the collection times of images in the test set differed from those in the training set, effectively reducing potential interference caused by temporal sequence effects on model performance evaluation.

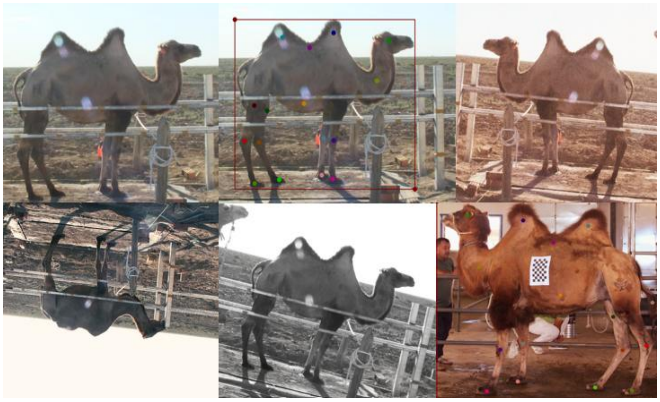


Fig. 7. Results of data augmentation.

B. Main Hardware Equipment

The data collection equipment includes a Canon 60D camera, an 18-135mm lens, and a stereo camera with frame synchronization. For model training, the computational hardware utilized an NVIDIA GeForce RTX 4090 GPU, with PyTorch version 2.1.0 and Python version 3.10. The GPU computation platform was based on CUDA version 12.1. The model's input size was set to 640×640.

C. Evaluation Metrics

The evaluation metrics for Camel-YOLOv11pose include Mean Average Precision (mAP50), Mean Average Precision at thresholds ranging from 0.50 to 0.95 (mAP50-95), Precision

(P), Recall (R), and Parameters (Params). The mAP50-95 metric evaluates the average precision at different IoU thresholds from 0.5 to 0.95, providing a more comprehensive assessment of the model's performance [33].

D. Experimental Results and Analysis

1) Comparative experiments: We conducted comparative experiments to verify the performance of the proposed model in the pose estimation task and compared it with several classic pose estimation algorithms across multiple metrics. To ensure the scientific rigor of the experiment, we adopted the same strategy as in the ablation experiments. All experiments used the same evaluation metrics and were tested on the same dataset.

TABLE I. MODEL COMPARISON RESULTS

Camel	P	R	mAP50	mAP50-95	Parameters	FPS
YOLOv7n-pose	90.2	89.6	87.7	68.5	34930376	140.8
YOLOv8n-pose	91	94.3	90.5	77.9	2971999	163.9
YOLOv8m-pose	90	94.2	89.8	75.9	23820511	181.8
YOLOv8s-pose	91.4	95.1	91.7	78.3	10324383	144.9
YOLOv11n-pose	92.4	90.7	91.9	79.6	2869783	166.6
YOLOv11pose-Camel	94.5	94.3	94.1	82.8	2678642	147.6

As shown in Table I, our YOLOv11pose-Came model performs better in the camel pose estimation task for the Bactrian camels of Xinjiang. Although both YOLOv7-pose and YOLOv8-pose used pre-trained weights as initial settings, their keypoint prediction [34] accuracy did not reach the baseline model's level. While the FPS slightly decreased compared to the original YOLOv11-pose baseline model, other metrics were improved. Compared to YOLOv7n-pose, YOLOv8n-pose, YOLOv8m-pose, YOLOv8s-pose, and YOLOv11n-pose, the improved model's mAP50 increased by 6.4, 3.6, 4.3, 2.4, and 2.2 percentage points, respectively. Although YOLOv8m-pose has a higher FPS than the improved model, its larger parameter size leads to higher memory consumption. Therefore, this study effectively improves the accuracy of camel pose estimation for Xinjiang Bactrian camels while ensuring detection speed, validating the effectiveness of the proposed method. The model comparison results are shown in Table I [28].

2) Ablation experiments: Based on the current YOLOv11pose-Camel model, we designed ablation experiments to evaluate the enhancement effects of the proposed modules and their contributions to algorithm performance. These experiments aim to analyze the impact of individual modules in depth. First, we introduced the ECAttention mechanism into the backbone feature extraction network to enhance the network's focus on key points. Next, we replaced the original SPPF module in the existing feature extraction network with the SimSPPF module. Finally, we integrated the C3K2 module with the DECA-Block module, which combines the EMAttention mechanism and DCNv3 dynamic convolution. We conducted experimental evaluations of each improvement strategy on the same dataset. To ensure the accuracy of the results, we used the same evaluation metrics to measure the contribution of each improvement to the network. From these experiments, it can be observed that each module contributes to varying degrees of improvement in the overall performance of the model. Through the pose estimation algorithm, we successfully obtained accurate skeletal key points. The relevant experimental details and test results are shown in Table II.

TABLE II. ABLATION EXPERIMENT RESULTS

	P	R	mAP50	mAP50-95
YOLOv11-pose	92.4	90.7	91.9	79.6
ECAttention	93.3	91.9	92.5	81.7
SimSPPF	92.9	90.8	93.4	79.7
SimSPPF-ECA	93.8	92.2	93.3	82.5
YOLOv11-pose-Camel	94.5	94.3	94.1	82.8

3) Bactrian camel pose estimation: We determined the position information of the key points by using the pose estimation algorithm, and then inferred the occluded key points using Eq. (5). These key points allow for a more accurate description of the camel's skeletal structure and dynamic changes, providing scientific data for subsequent studies on the morphology of Bactrian camels [35]. Some training results of YOLOv11pose-Camel are shown in Fig. 8.

Fig. 8. Partial training results.

V. SUMMARY AND OUTLOOK

A. Summary

In this study, we explored the pose estimation method for Bactrian camels and conducted detailed technical analysis and experimental validation. By introducing the YOLOv11pose-Camel optimization design, we successfully achieved an efficient fusion of object detection and pose estimation, improving recognition accuracy and real-time performance under complex backgrounds. The model significantly enhanced the ability to recognize skeletal key points of different parts of Bactrian camels. YOLOv11pose-Camel maintained good.

Performance under various experimental conditions, achieving evaluation metrics (precision, recall, mean average precision) of 94.5%, 94.3%, and 94.1%, respectively. These performances surpassed existing models such as YOLOv8pose, YOLOv7pose, YOLOv5pose, and the base model YOLOv11pose used in this study. The improved model reduced parameters by 0.191 million, and the FPS reached 158.73, reflecting a significant improvement in detection speed

and real-time performance, making it highly suitable for rapid deployment and cross-device portability.

In the context of the current development of intelligent farming, pose estimation and real-time monitoring of Bactrian camels are becoming increasingly important. Pose estimation lays the technical foundation for subsequent morphological evaluation (such as hoof health and rumination), which can ensure the healthy development of camels at an early stage, providing valuable insights into enhancing economic value. Therefore, the results of this study hold significant appeal for camel breeders [36], [29].

B. Outlook for the Future

Although this study has made certain progress in some areas, there are still some limitations.

1) This research focuses on pose estimation for Bactrian camels in Xinjiang, and we plan to extend the study to include more breeds in the future to address this issue. For example, we aim to study camels from different regions, environments, and breeds, thereby improving the practical applicability of the model developed in this research [37].

2) While we have achieved the expected results in addressing feature deformation issues through algorithm optimization, there is still a gap to higher performance metrics. Therefore, future work should explore methods to further improve the model's accuracy and performance.

Looking ahead, integrating more advanced computer vision techniques and incorporating multimodal frameworks could better facilitate information fusion [38], enhancing the model's ability to extract features. This would significantly improve the accuracy and practicality of pose estimation, especially for large-scale precision breeding farms with abundant resources. These improvements could make our framework more practical and promote the development of livestock farming in Xinjiang.

ACKNOWLEDGMENT

This research is funded by the Natural Science Foundation of Xinjiang Uygur Autonomous Region (No. 2021D01C082). We would also like to express our sincere gratitude to Xinjiang Xintuo Dairy Group for their strong support.

REFERENCES

[1] Wu Saisai, Wu Jianzhai, Cheng Guodong, et al. Research progress on animal behavior recognition based on pose estimation in agricultural universities [J]. Journal of Agricultural University, 2023, 28(6): 22-35.

[2] Yang Liang, Wang Hui, Chen Ruipeng, et al. Research progress on pig-specific sensors [J]. Journal of Intelligent Agricultural Machinery (Chinese and English), 2023, 4(2): 22.

[3] He Dongjian, Liu Dong, Zhao Kaixuan. Research progress on intelligent perception and behavior detection of animals in precision livestock farming [J]. Journal of Agricultural Machinery, 2016, 47(5): 231-244.

[4] Yongfeng, Wang Wensheng, Guo Leifeng, et al. Research progress on daily behavior recognition of cattle based on wearable sensors [J]. Journal of Livestock Ecology, 2022, 43(10): 1-9.

[5] Guo Dongdong. Research on the recognition of goat behavior features based on triaxial acceleration sensor [D]. Taiyuan: Taiyuan University of Technology, 2015.

[6] Liu Haiying. Image processing technology based on computer vision algorithms [J]. Computer and Digital Engineering, 2019, 47(3): 672-677.

[7] Wang Hao. Application of computer vision in smart medical care [J]. Electronic Communication and Computer Science, 2024, 6(1): 86-88.

[8] Bai Yang, Bai Yun, Zhan Xini, et al. Technological characteristics and morphological evolution of artificial intelligence generated content (AIGC) [J]. Library and Information Science Knowledge, 2023, 40(1): 66-74.

[9] Yang Jiejing, Liu Zhigang. Exploration of the path of high-quality development of livestock farming under the model of **-style modernization [J]. Feed Research, 2023, 46(11).

[10] Yan Pengfei, Xiao Sha, Zhang Zhiyong, et al. Application and Research Progress of Computer Vision Technology in Modern Animal Husbandry[J]. **Pig Industry, 2024, 19(5): 83-89.

[11] Qi Yu, Su Han, Hou Rong, et al. A Method for Estimating Giant Panda Pose Based on High-Resolution Network[J]. Journal of Zoology in China, 2022, 42(4): 451.

[12] Chen Y, Fang R, Liang T, et al. Stock Price Forecast Based on CNN‐BiLSTM‐ECA Model[J]. Scientific Programming, 2021, 2021(1): 2446543.

[13] Fang H S, Xie S, Tai Y W, et al. Rmpe: Regional multi-person pose estimation[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2334-2343.

[14] Zheng C, Zhu X, Yang X, et al. Automatic recognition of lactating sow postures from depth images by deep learning detector[J]. Computers and electronics in agriculture, 2018, 147: 51-63.

[15] Song H, Jiang B, Wu Q, et al. Detection of dairy cow lameness based on fitting line slope feature of head and neck outline[J]. Transactions of the CSAE, 2018, 34(15): 190-199.

[16] Chen F, Liang X, Chen L, et al. Novel method for real-time detection and tracking of pig body and its different parts[J]. International Journal of Agricultural and Biological Engineering, 2020, 13(6): 144-149.

[17] Zheng W, Yan L, Chen L, et al. Knowledge-Embedded Mutual Guidance for Visual Reasoning[J]. IEEE Transactions on Cybernetics, 2023.

[18] Natesan B, Liu C M, Ta V D, et al. Advanced robotic system with keypoint extraction and YOLOv5 object detection algorithm for precise livestock monitoring[J]. Fishes, 2023, 8(10): 524.

[19] Rodriguez-Juan J, Berenguer-Agullo A, Benavent-Lledo M, et al. Bird Action Recognition in Wetlands using Deep Learning[C]//Proceedings of the 2024 International Conference on Information Technology for Social Good. 2024: 350-357.

[20] Research Progress and Technology Trend of Intelligent Monitoring of Dairy Cow Exercise Behavior[J]. Smart Agriculture, 2022, 4(2): 36.

[21] Wang J, Zhang X, Gao G, et al. Open Pose Mask R-CNN Network for Individual Cattle Recognition[J]. IEEE Access, 2023.

[22] Barney S, Dlay S, Crowe A, et al. Deep learning pose estimation for multi-cattle lameness detection[J]. Scientific Reports, 2023, 13(1): 4499.

[23] Mohanavel D, Ishwarya A M. Deep Learning and Computer Vision Based Warning System for Animal Disruption in Farming Environments[C]//2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT). IEEE, 2024: 1-6.

[24] Application of Deep Learning in Wildlife Conservation[J]. Acta Zoologica Sinica, 2023, 43(6): 734.

[25] Yuan, Zijian, et al. "YOLOv8-ACU: improved YOLOv8-pose for facial acupoint detection." Frontiers in Neurorobotics 18 (2024): 1355857.

[26] Deng, Ye, et al. "Context Adaptive Network for Image Inpainting." IEEE Transactions on Image Processing (2023).

[27] Wang, Shuqi, et al. "A Human Posture Estimation Method for Image Interaction System Based on ECA." International Forum on Digital TV and Wireless Multimedia Communications. Singapore: Springer Nature Singapore, 2023.

[28] Ren, Fei, et al. "Steel Surface Defect Detection Using Improved Deep Learning Algorithm: ECA-SimSPPF-SIoU-Yolov5." IEEE Access (2024).

[29] Peng, Yue, Alifu Kurban, and Mengmei Sang. "An Improved YOLOv8 Method for Measuring the Body Size of Xinjiang Bactrian

Camels." International Journal of Advanced Computer Science & Applications 15.8 (2024).

[30] Qu, X.; Long, Y.; Wang, X.; Hu, G.; Tao, X. Research on the Cable-to-Terminal Connection Recognition Based on the YOLOv8-Pose Estimation Model. Appl. Sci. 2024, 14, 8595.

[31] Ali Alameer. "Facial Emotion Recognition Datasets for YOLOv8Annotation." University of Salford, 2023.

[32] Hu Aijun, Tang Guiji, An Lianlian. Noise Reduction Method of Vibration Signal of Rotating Machinery Based on Mathematical Morphology[J]. Journal of Mechanical Engineering, 2006, 42(4): 127-130.

[33] Jiang T, Li Y, Feng H, et al. Research on a Trellis Grape Stem Recognition Method Based on YOLOv8n-GP[J]. Agriculture, 2024, 14(9): 1449.

[34] D. I. Krasnov, S. N. Yarishev, V. A. Ryzhova, and T. S. Djamiykov, "Improved YOLOv8 Network for Small Objects Detection," 2024 XXXIII International Scientific Conference Electronics (ET), 2024, doi: 10.1109/ET63133.2024.10721517.

[35] LI Menghe, XU Hongji, SHI Leixin, et al. Multi-person Behavior Recognition Based on Bone Key Point Detection[J]. Computer Science, 2021, 48(4): 138-143.

[36] Gong C, Zhang Y, Wei Y, et al. Multicow pose estimation based on keypoint extraction[J]. PloS one, 2022, 17(6): e0269259.

[37] Shao D, He Z, Fan H, et al. Detection of cattle key parts based on the improved Yolov5 algorithm[J]. Agriculture, 2023, 13(6): 1110.

[38] Yang Q, Liu Y, Lu L, et al. GFIDF: Gradual Fusion Intent Detection Framework[J]. 2024.

A Hybrid Machine Learning Approach for Continuous Risk Management in Business Process Reengineering Projects

RAFFAK Hicham[1], LAKHOUILI Abdallah[2] , MANSOURI Moahmed[3]
Faculty of Sciences and Techniques University, Hassan 1 st Settat, Morocco[1]
Faculty of Sciences and Techniques University, Hassan 1 st Settat, Morocco[2]
National School of Applied Sciences University, Hassan 1 Berrechid, Morocco[3]

***Abstract*—This study proposes a hybrid machine learning approach for continuous risk management in Business Process Reengineering (BPR) projects. This approach combines supervised and unsupervised learning techniques, integrating feature selection and preprocessing through Principal Component Analysis (PCA), clustering with K-means, and visualization with t-SNE. The labeled data are then used as input for predictive modeling with XGBoost, optimized using Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Grid Search algorithms.PCA reduces data dimensionality, simplifying analysis and improving model performance. K-means and t-SNE are employed for data clustering and visualization, enabling the identification of risk segments and uncovering hidden patterns. XGBoost, a powerful boosting algorithm, is utilized for predictive modeling due to its efficiency, accuracy, and ability to handle missing values. Optimization techniques further enhance XGBoost's performance by fine-tuning its hyperparameters. The approach was applied to a risk database from the automotive sector, demonstrating its practical applicability. Results show that PSO achieves the lowest mean squared error (MSE) and root mean squared error (RMSE), followed by GWO and Grid Search. Mahalanobis distance yields more accurate clustering results compared to Euclidean, Manhattan, and Cosine distances. This hybrid machine learning approach significantly enhances risk detection, evaluation, and mitigation in BPR projects, offering a robust framework for proactive decision-making.**

***Keywords*—BPR; Risk management; PCA; K-means; XGBoost; PSO; GWO**

I. Introduction

In today's fast-paced and competitive business environment, organizations, companies, and enterprises consist of a series of organized and interconnected business processes and activities arranged sequentially, requiring effective and efficient management to achieve strategic objectives [1]. Business Process Management (BPM) provides a systematic approachto managing work and achieving goals [2]. Furthermore, due to the dynamic nature of business, organizations often evolve through growth, transformation, or expansion into new markets [3]. This evolution impacts business processes, which must be adjustedto align with the company's needs. Since the First Industrial Revolution, when Henry Ford introduced the assembly line, business processes have played a crucial role in managing and improving productivity [4]. Consequently, the science of processes has emerged, introducing numerous tools and techniques, such as Business Process Reengineering (BPR), as powerful methods to improve process efficiency and productivity [5]. Additionally, as dynamic components, business processes are influenced by external events and other internal processes within the same organization [6]. Thus, Business Process Management has evolved from the initial concept of Business Process Reengineering to awell-established management approach [7]. These strategies have improved the monitoring and control of efficiency, productivity, profitability, service levels, and other business objectives [8]. As companies grow, transform, or expand, the efficiency of business processes can be affected, sometimes requiring a redesign of processesto adapt to business changes [9].

With process automation and digital transformation, many manual tasks have been converted into digital platforms, suchas workflow management systems, thereby increasing productivity, efficiency, and effectiveness [10]. Automationhas provided organizations with an abundance of data and detailed records [11]. Rapid advancements in information technology, automation, and digital transformation have elevated expectations regarding the purpose of processes, even before considering their improvement or reengineering [12].

The integration of hybrid methods combining supervised and unsupervised learning offers promising prospects for continuous risk management in BPR projects. Supervised learning, which relies on labeled data, enables the creation of accurate predictive models for identifying and quantifying risks [13]. On the other hand, unsupervised learning, which does not require labeled data, excels at uncovering hidden structures and unknown patterns within the data, offering a deep understanding of potential risks. The combination of these two approaches leverages the complementary advantages of each method.

To strengthen this approach, a hybridization of unsupervised algorithms such as K-means and t-SNE with supervised algorithms like XGBoost, optimized by methods such as PSO, GWO, and Grid Search, can be used for risk management in BPR projects. K-means and t-SNE are particularly effective for clustering and visualizing data, enabling the identification of emerging risk segments and anomalies in operational data in real-time [14]. XGBoost is known for its performance in terms of precision and speed in classification and regressiontasks [15].

Optimizing XGBoost with techniques such as Particle Swarm Optimization (PSO), Grey Wolf Optimizer (GWO), and Grid Search further enhances the accuracy and robustness of predictive models [16]. PSO and GWO are particularly effective in searching for optimal solutions within complex parameter spaces, while Grid Search provides an exhaustive method for exploring possible parameter combinations.

In Section II, a literature review of the various concepts addressed in this study will be presented. Subsequently, in Section III, we will introduce the proposed approach along with the computational conditions of our model. Section IV will focus on the results and their discussion before concluding the study in Section V.

II. Literature Review

A. Business Process Re-engineering

Since its emergence in the early 1990s, Business Process Reengineering (BPR) has attracted considerable interest. Both scholars and industry professionals have extensively discussed its significance, methodologies, impacts, and success factors [9]. BPR emerged as a groundbreaking strategy aimed at fundamentally rethinking and restructuring business operations to achieve substantial improvements in metrics such as cost, quality, service, and speed [10]. Reengineering involves a radical redesign of business processes, characterized by an extensive overhaul of the organization's processes, technologies, management systems, organizational structures, and core values. The goal is to realize significant performance enhancements throughout the organization. For BPR to succeed, it must be integrated with other organizational components, leverage advanced technology, and employ various methodologies. BPR cannot thrive in isolation. Information Technology (IT) plays a critical role in BPR by providing the tools needed for exceptional organizational achievements, though its role is often misunderstood [12].

For any implementation team, the ultimate objective is to achieve a high success rate in their projects. However, the outcomes of business process reengineering initiatives have been mixed, often due to the adoption of best practices or industry benchmarks from various sectors without fully understanding the specific needs of the target industry. Notably, approximately 70% of such projects fail, largely due to the absence of an appropriate framework or methodology [8]. Nonetheless, numerous factors influence a project's outcome. These factors serve as critical indicators in predicting the project's trajectory or assessing its potential for success. BPR inherently involves risks, and its successful implementation relies on several critical success factors [17].

The successful implementation of BPR relies on several key factors, offering valuable practical insights [18]. Change management and organizational culture play a central role, emphasizing effective communication, robust motivation and reward systems, employee empowerment, continuous training and development, and a collaborative work environment. Similarly, managerial competence and support are essential, requiring strong leadership, expertise in risk management, active engagement and support from senior management, as well as an appropriate organizational structure. The BPR process itself must align seamlessly with organizational goals through strategic planning, effective project management, proper methodological application, productive consultation, and a clear BPR vision. Finally, IT capabilities are indispensable, encompassing a robust IT infrastructure, enhanced IT functionality, and the alignment of IT systems with BPR strategies to ensure successful implementation.

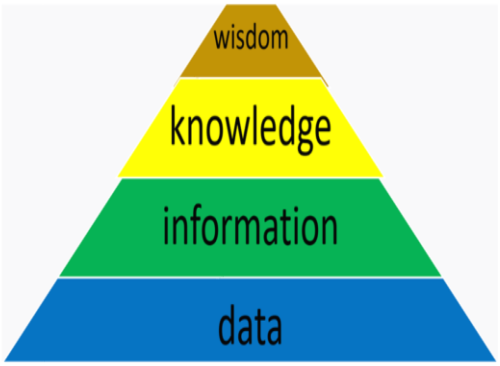


Fig. 1. Data transformation pyramid.

Agile development principles are progressively replacing traditional tools, leading to a significant transformation in engineering and management methods [19]. This evolution creates new requirements for the design and management of knowledge bases.

Data, as illustrated in Fig. 1, even in their raw state, form the foundation of layered processing systems. Their quality directly impacts their ability to generate added value, particularly in optimizing processes and improving product development [20]. An effective database should, therefore, act as a key resource to support management tools and performance indicators essential for structured and informed development [21].

However, the lack of sufficiently complete data and usable knowledge poses a major challenge. This gap hinders strategic decision-making, team coordination, and the overall optimization of product lifecycle management [22].

Data exploitation in risk management, while advantageous, faces significant challenges. The application of risk management tools such as PFMEA generates volumes of data that are often unmanageable for engineers, with heterogeneous sources and varied formats. This lack of standardization, combined with poorly structured data entry, results in duplicates and irrelevant data. Historically, risk evaluation relied on human expertise, but this approach is biased by personal experiences, leading to inconsistent and ambiguous risk identification and management across projects. This complicates the reuse of historical data and highlights the importance of data quality and standardization for effective processing [23].

B. Principal Component Analysis

Reducing the number of dimensions in large, high-dimensional datasets is crucial for effective analysis. This process can either serve as the primary objective for visualizing complex data or act as a preliminary step before further analysis, such as clustering. Principal Component Analysis (PCA) is one

of the earliest and most renowned techniques for dimensionality reduction. Initially introduced by Pearson in 1901 and later independently refined by Hotelling in 1933, where the concept of "principal components" was formally established, PCA is also known by several other names, including the Karhunen-Loeve method, eigenvector analysis, and empirical orthogonal functions. PCA remains one of the most widely used methods for creating low-dimensional representations of multivariate data (Fig. 2) [24].

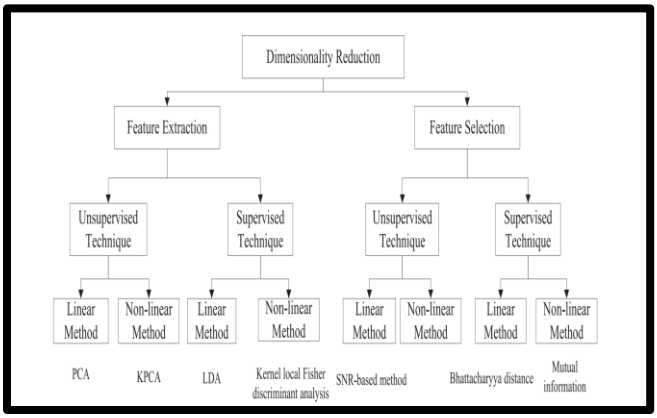


Fig. 2. Typical nomenclature of dimensionality reduction techniques.

PCA is a linear technique because it constructs components as linear combinations of the original variables (features). Despite its linearity, PCA can preserve the non-linearity of the data, making it effective for visualization purposes. The process involves iteratively calculating the direction of maximum variance and then projecting the data onto a perpendicular hyperplane. This method quickly identifies a few orthogonal directions that capture most of the data's variability, resulting in a low-dimensional representation. When all principal components are considered, the process can be visualized as a rotation in the space of the original variables. For a thorough exploration and historical context of principal component analysis, refer to [25].

C. Clustering

Clustering divides a group of individuals into several categories based on their similarities, where the differences among individuals within the same category should be as small as possible [26]. The most representative clustering methods are based on geometric distance measurement. Clustering techniques can analyze complex input data patterns and suggest solutions that might not be evident otherwise. They reveal customer typologies, enabling highly effective marketing strategies [27].

The goal of classification is to build a function or model based on the characteristics of the entire dataset and then categorize each object into a known object class. Classification has a wide range of applications, such as medical diagnosis, credit scoring, image pattern recognition, target market positioning, defect detection, efficiency analysis, graphic processing, insurance fraud analysis, [26].

Prediction involves using knowledge generated from historical and current data to deduce future data trends. While classification is used to predict classes, analysts often want to predict certain values of missing or unknown data. In other words, the desired prediction outcome corresponds to numerical data [26].

a) Positioning Euclidean distance, Manhattan distance, Mahalanobis distance and Cosine similarity: The Euclidean distance and Manhattan distance are both specific cases of the Minkowski distance. Let X and Y be two data samples, each consisting of T, elements, defined as follows:

$$X = [X_1, X_2, \cdots, X_T] \quad ; Y = [Y_1, Y_2, \cdots, Y_T] \tag{1}$$

The Minkowski distance of order p (where p is an integer) between two samples X and Y is defined by the following equation:

$$d(X,Y) = \left(\sum_{i=1}^{n} |X_i - Y_i|^p\right)^{\frac{1}{p}} \tag{2}$$

This distance metric evaluates the difference between two data samples as vectors in a multi-dimensional space, with the order p determining the emphasis on individual component differences. As p increases, larger differences in components have a more pronounced effect on the overall distance. Specifically, when p=1 and p=2, the Minkowski distance simplifies to the Manhattan distance and the Euclidean distance, respectively. As p approaches infinity, it converges to the Chebyshev distance (Fig. 3) [28].

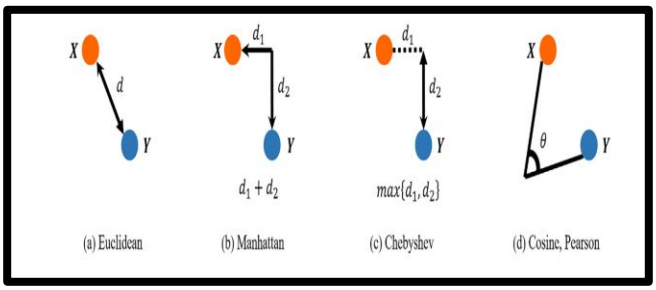


Fig. 3. Diagram of several distance measures.

The Euclidean distance is a measure of the straight-line distance between two points in Euclidean space, forming the basis of classical geometry. It is widely used as a similarity measure, particularly suitable for cases where there is no inherent correlation between different features [27]. y default, unless specified otherwise, the Euclidean distance is often employed. However, it is sensitive to the scale of the features, which can skew the results if the units are not consistent. Hence, it is generally necessary to normalize or standardize the data before applying this distance measure [28].

The Manhattan distance, also known as the city block distance or taxicab distance, calculates the distance between two points as the sum of the absolute differences of their Cartesian coordinates. The term "Manhattan distance" arises from the grid-like street layout of Manhattan, where the shortest path between two points involves a series of right-angle turns [29]. For example, in a study on personalized visual comfort control in buildings, individual user preferences and energy consumption profiles were analyzed. Collaborative user preferences were computed based on the Manhattan distance between the target occupant and others, leading to recommended adjustments in light intensity.

The Mahalanobis distance, as defined in Eq. (2), measures the distance between two vectors X and Y from the same distribution, using the covariance matrix S [30].

$$d(X,Y) = \sqrt{(X-Y)^{\top} S^{-1} (X-Y)} \quad (3)$$

where S is the covariance matrix. This distance metric extends the Euclidean distance by incorporating correlations between data points through the covariance matrix S. It is particularly effective for datasets with reduced features, although the covariance matrix can introduce unwanted redundancies. The Mahalanobis distance remains stable against projections or scaling of the data, making it useful for identifying outliers. For instance, Westermann et al. utilizedthe Mahalanobis distance to filter outliers from building energy data, classifying points farthest from the center of a multivariate Gaussian distribution as outliers [31].

Cosine similarity measures the similarity between vectors by focusing on their direction and angle rather than magnitude. The cosine similarity between two vectors X and Y is defined as:

$$cos(\theta) = \frac{\sum_{i=1}^{T} X_i Y_i}{\sqrt{\sum_{i=1}^{T} (X_i)^2} \times \sqrt{\sum_{i=1}^{T} (Y_i)^2}} \quad (4)$$

where θ is the angle between X and Y. A smaller angle indicates a higher similarity between the two vectors. The value of this equation ranges between -1 and 1. Based on this equation, the cosine distance can be defined, ranging from 0 to 2.

A study proposed a modified cosine similarity measure for initializing input weights in a building energy consumption prediction model, defined as follows [32]:

$$cos'(\theta) = \frac{\sum_{i=1}^{T} (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{T} (x_i - \bar{X})^2} \times \sqrt{\sum_{i=1}^{T} (y_i - \bar{Y})^2}} \quad (5)$$

where $\bar{X}$ and $\bar{Y}$ are the mean values of X and Y, respectively.

Instead of using Euclidean distance, which is highly sensitive to magnitudes, the modified cosine similarity coefficient is used to initialize the weights connecting the input neurons and the hidden neurons in the extreme learning machine, thereby improving its generalization ability [33].

b) K-means algorithm: K-means is one of the clustering algorithms [26]. It takes the number of clusters as a parameter and partitions the data into the specified number of clusters so that the similarity within each cluster is high [34]. K-means is an iterative approach that calculates the centroid values before each iteration. It requires precise numbers of clusters k, as the initial cluster center can change, which may lead to unstable data grouping [35]. Data points are moved between different clusters based on the centroids calculated in each iteration [36]. The process is repeated until the sum of distances cannot be decreased further. The advantages of K-means include its speed and scalability: it is one of the fastest clustering models and can efficiently handle large datasets with many records and numerous input clustering fields [26]. The K-means algorithm is presented in Algorithm 1.

Algorithm 1: K-means Algorithm.

1. Initially, based on the value of k, k random points are chosen as initial centroids.
2. The distances from each data point to the previously chosen centroids are calculated.
3. The distance values are compared, and each data point is assigned to the centroid with the shortest Euclidean distance.
4. The previous steps are repeated. The process stops if the clusters obtained are the same as those in the previous iteration.

c) Evaluation criteria: Unsupervised evaluation criteria are based on internal clustering information, such as the distance between objects within a cluster and the centroid of that cluster [37]. These criteria often rely on the simplest clustering definition, which states that objects within the same cluster should be as close as possible, and objects fromdifferent clusters should be as far apart as possible. To determine if a clustering respects this intuitive definition, distance measures are calculated between cluster representatives and residual objects. These unsupervised measures evaluate both the compactness and separability of clusters. Since the definitionof cluster quality is not formally defined, numerous criteria evaluate results differently. Some criteria are used directly as an objective function and optimized by a clustering algorithm. Others are too costly to evaluate during algorithm execution and are calculated after its application.

- Silhouette Coefficient (CS):

The silhouette coefficient evaluates the compactness of clusters and their separability [35]. It can be calculated for each object, each cluster, and the entire clustering. For an object x,it is defined as:

$$CS(x) = \frac{b_x - a_x}{max(a_x, b_x)} \quad (6)$$

where a_x is the average distance between object x and all other objects in the same cluster, and b_x is the average distance between xx and all objects not in the same cluster. The coefficient $C(x)$ ranges between -1 and 1. A positive value (a_x<b_x) indicates that objects in the same cluster as x are closer to xx than objects in other groups. For a cluster, the silhouette coefficient is the average of the coefficients of objects in that cluster:

$$CS(C_i) = \frac{1}{|C_i|} \sum_{x \in C_i} CS(x) \quad (7)$$

- Clustering Evaluation

The silhouette coefficient for clustering is equal to the average of the silhouette coefficients of its clusters:

$$CS(C) = \frac{1}{K} \sum_{i=1}^{K} CS(C_i) \quad (8)$$

The silhouette coefficient ranges between -1 and 1, with a positive value indicating that the clusters are very compact and well-separated. It should be noted that calculating this index is relatively time-consuming because many distance calculations are required for its evaluation.

D. The T-Distributed Stochastic Neighbor Embedding

Nonlinear techniques offer major advantages in processing nonlinear and complex datasets. The t-distributed stochastic neighbor embedding (t-SNE), among these techniques, has become a benchmark method for dimensionality reduction and data visualization across various fields [38]. Its applications encompass a wide range of domains, including microbiome data, single-cell RNA sequencing, bird song analysis, computational fluid dynamics, genomic data, and remote sensing images, among others. The t-SNE algorithm projects complex datasets onto a 2D or 3D space while preserving the local structure of the original high-dimensional space. However, while t-SNE excels in data visualization, it lacks an intrinsic mechanism to map new data points onto the low-dimensional representation, limiting its use in classification and regression tasks [39, 40].

E. The Extreme Gradient Boosting

It is built on gradient boosting trees, Extreme Gradient Boosting (XGBoost) is an algorithm delivering significant performance improvements over traditional gradient boosting methods. Based on the Classification and Regression Tree (CART) theory, XGBoost stands out as an effective tool for addressing regression and classification problems [41]. Furthermore, XGBoost is a flexible computing library that incorporates innovative algorithms with Gradient Boosting Decision Trees (GBDT) methods [42].

The objective function of XGBoost, post-optimization, comprises two components: one for model deviation and another regularization term to mitigate overfitting. Consider D = {(x_i,y_i)} as a dataset containing nn samples and m features, where the predictive model is an additive ensemble of k base models. The prediction for a sample is expressed as:

$$\hat{y}_i = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \varphi, \qquad \varphi = \{ f(x) = w_s(x) \mid (s: \mathbb{R}^m \to T, w_s \in \mathbb{R}^T) \} \tag{10}$$

where $\hat{y}_i$ is the predicted label for the i-th sample, x_i is the i-th sample $f_k(x_i)$ is the predicted score, and φ represents the set of regression trees. which is a tree structure parameter of s, f(x) and w representing the weight of leaves and the number of leaves. The objective function in XGBoost is a combination of the traditional loss function and a term for model complexity, described by:

$$\text{Obj} = \sum_{i=1}^{m} l(\hat{y}_i, y_i^{(t-1)}) + f_i(x_i)) + \Omega(f_k) \qquad \Omega(f_k) = \gamma T + \frac{1}{2}\lambda w^2. \tag{11}$$

In this formula, the first term, $l(\hat{y}_i, y_i)$ is the traditional loss function, and the second term, $\Omega(f_k)$, accounts for the model's complexity. Here, γ and λ are parameters used to tune the tree's complexity, helping smooth the final learning weights and preventing overfitting.

F. Optimization Techniques

a) GWO algorithm: The Grey Wolf Optimizer (GWO) is an optimization method inspired by the social hierarchy and hunting strategies of grey wolves. This algorithm mimics the leadership and cooperative behavior observed in wolf packs. The wolf pack is classified into four types of wolves [43, 44]:

Algorithm 2: GWO Algorithm

(a) Alphas: The leaders of the pack, responsible for decision-making and guiding the group.
(b) Betas: These wolves act as deputies to the alphas, assisting in decision-making and other critical tasks.
(c) Deltas: Subordinate to both alphas and betas, these wolves still hold authority over omegas and include roles such as scouts, sentinels, elders, hunters, and caretakers.
(d) Omegas: The lowest ranking members of the pack, often serving as scapegoats, and must submit to all other wolves in the hierarchy.

b) PSO algorithm: Particle Swarm Optimization (PSO), developed by Kennedy and Eberhart, is inspired by the study of bird flocking behavior during their search for food [45]. The algorithm updates each particle's velocity and position by considering the best position found by any particle in the swarm and each particle's personal best position within the search space. The PSO procedure encompasses five key steps:

Algorithm 3: PSO Algorithm

(1) initialization,
(2) evaluation,
(3) updating the particle's best position (Pbest),
(4) updating the global best position (Gbest),
(5) updating the particles' velocity and position. Particles adjust their trajectories based on Pbest and Gbest, progressively converging towards the optimal solution.

Furthermore, XGBoost is a flexible computing library that incorporates innovative algorithms with Gradient Boosting Decision Trees (GBDT) methods [42].

III. Proposed Approach

Fig. 4 illustrates the structure of the proposed framework for a hybrid machine learning (ML) approach aimed at risk assessment and classification in the context of an operational process reengineering project. This framework integrates feature selection and preprocessing, unsupervised learning (UL), and supervised learning (SL) paradigms as its core components.

In the preprocessing phase, the primary objectives are attribute transformation, composite attribute splitting, dimensionality reduction, and attribute rank analysis. Principal component analysis (PCA) is the pivotal tool employed for feature selection. The UL tools employed include the k-means algorithm and t-SNE for clustering and assigning target labels to the dataset. K-means identifies the number of clusters that correspond to different risk impact levels within the dataset. t-SNE provides detailed visualizations, facilitating the understanding of clusters, patterns, and relationships between risk input parameters, a key objective in data mining. Additionally, t-SNE maps each data point to a specific cluster (target class), thereby generating a target vector for the dataset. The main purpose of the UL stage is to transform the previously unlabeled risk dataset into a labeled dataset suitable for SL.

In the SL phase, XGBoost is employed to perform regression, classification, and risk forecasting using the labeled dataset obtained from the UL stage. To further enhance the model's accuracy, grid search, particle swarm optimization, and the grey wolf optimizer algorithms are used to optimize the parameters of the XGBoost model, with the results from each

method compared. The implementation of the design is carried out in three stages: data collection, description and preprocessing, rank analysis, clustering visualization, and evaluation of the overall approach.

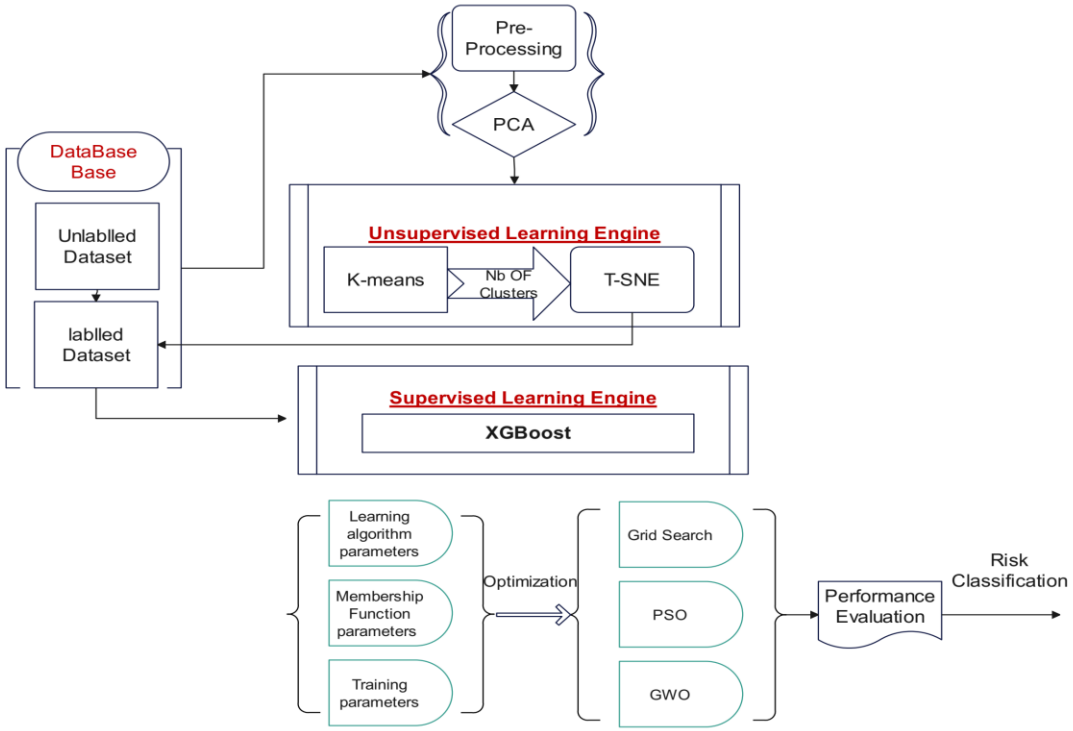


Fig. 4. Proposed approach.

IV. Results and Discussion

We applied the program to a risk database based on PFMEA register of a company operating in the automotive sector. The database consists of 11 attributes, 4 of which are numerical: the severity, occurrence, and detectability factors, and the fourth is the risk impact index, which is the RPN (Risk Priority Number), used to categorize and predict the risk.

The other categorical attributes are the risk title, owner, description, concerned part, detection tools, action to be applied, and description.

PCA is used in this code to reduce the dimensionality of the input data. The principal components capture the majority of the variance in the data while reducing the number of dimensions, which simplifies the analysis and can improve the performance of machine learning models.

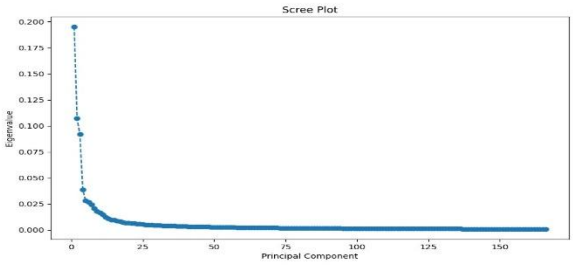


Fig. 5. Variance explained by principal components in PCA.

The Fig. 5 illustrates the eigenvalues associated with each principal component, showing the amount of variance captured by each component from the data. The graph reveals a sharp decline in eigenvalues from the first principal component to the second and third, indicating that the initial components capture most of the data's variance. After the initial drop, the eigenvalues level off, suggesting that the subsequent components contribute minimally to the total variance. This pattern generally indicates that only the first principal components are significant for explaining the dataset's variability, with the others being less important.

TABLE I. K-Means vs. Agglomerative Clustering Silhouette Scores

Algorithm	Silhouette Score
K-means	0.17712200253115506
Agglomerative Clustering	0.15189647725476427

The silhouette score is considered as a criterion for evaluating the distances between the data points and the clustering. Based on this score and by comparing the four distances: Euclidean, Manhattan, Mahalanobis, and Cosine, the Manhattan and Mahalanobis distances were shown to be more effective. However, given the advantages of the Mahalanobis distance, the authors opted for Mahalanobis distance due to its ability to consider variable correlations, which can be crucial depending on the context and nature of the data.

As shown in the Table I, when comparing the two algorithms, K-means and Agglomerative Clustering, K-means proves to be more effective with a silhouette score of 0.17712200253115506.

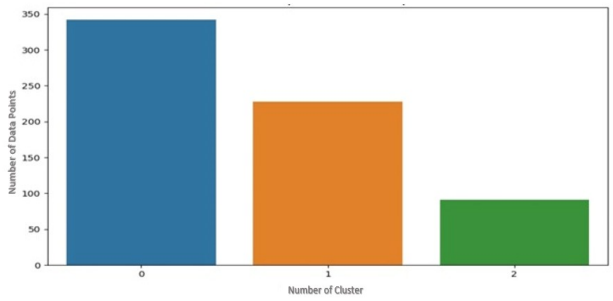


Fig. 6. Distribution of data points by cluster.

The clustering performed by the k-means algorithm identified three clusters (Fig. 6), corresponding to three levels of risk. However, based on human evaluation and as outlined in the PFMEA grid, four levels of risk are distinguished. Cluster 0 has the highest number of data points, with slightly over 300 points, indicating it is the most populated cluster. Cluster 1 follows with around 225 data points, and Cluster 2 has the fewest data points, with slightly less than 150 points. This distribution suggests that the data points are not evenly distributed among the clusters, with Cluster 0 containing the majority of the data points, Cluster 1 having a moderate amount, and Cluster 2 containing the least.

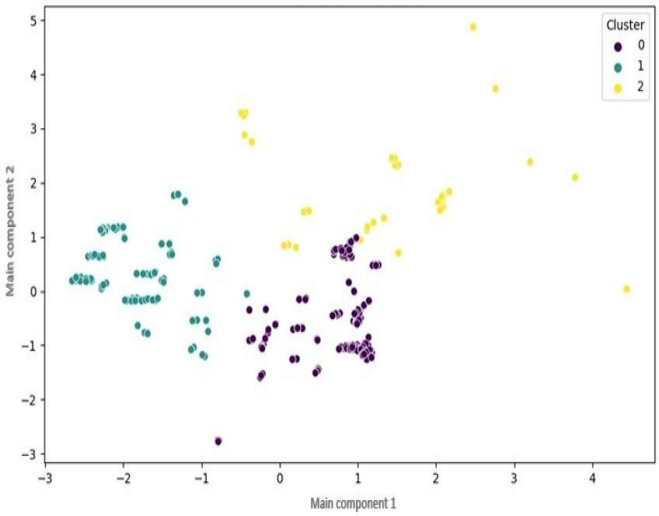


Fig. 7. Cluster visualization using ACP.

To better understand the relationship between the data, two dimensionality reduction and visualization methods were employed: PCA, a linear technique, and t-SNE, a non-linear method. For the first method (Fig. 7), although the clusters are relatively well-separated, some proximity between clusters 0 and 1 is observed.

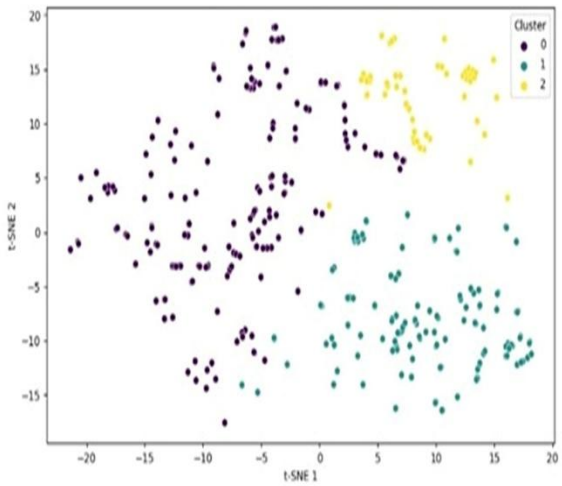


Fig. 8. Cluster visualization using t-SNE.

On the other hand, t-SNE, which leverages non-linear relationships to optimize the local representation of the data, allows for a clearer separation, particularly between clusters 0 and 2, but at the cost of less intuitive axis interpretation (Fig.8). In conclusion, PCA is suitable for data with a linear structure, while t-SNE is more effective in identifying complex relationships in non-linear data. For our dataset we adopt the t-SNE algorithm.

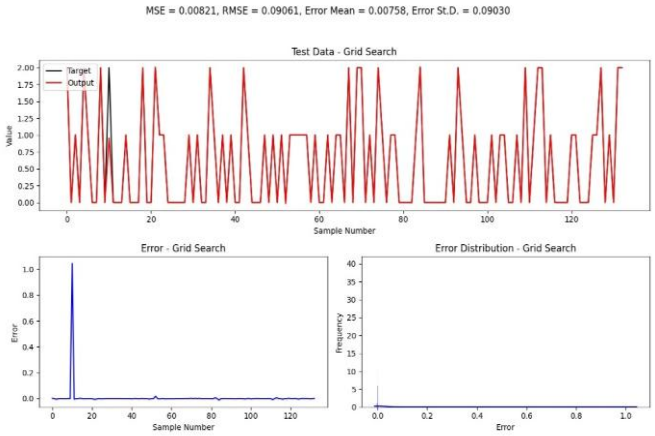


Fig. 9. Grid search optimization results: test data, error, and error distribution.

XGBoost is particularly appreciated for its performance and efficiency in machine learning competitions, as well as its ability to handle data with missing values and reduce overfitting through regularization. XGBoost is therefore a relevant choice for this type of analysis due to its robustness and ability to provide accurate results on complex datasets. In our study, we optimized XGBoost using the three techniques: GWO, PSO, and Grid Search, and evaluated the performance of our model based on two error metrics: MSE and RMSE.

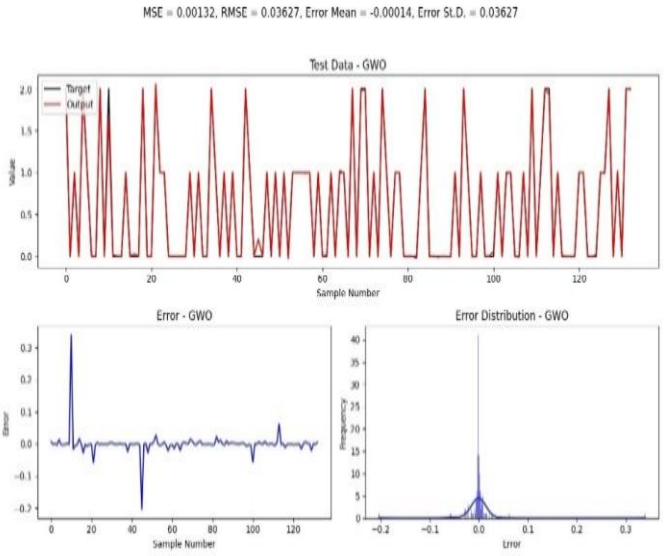


Fig. 10. GWO optimization results: test data, error, and error distribution.

The analysis of the three optimization methods—PSO (Fig. 11), GWO (Fig. 10), and Grid Search (Fig. 9)—reveals significant differences in their performance. The PSO model shows the best results with an MSE of 0.00119 and RMSE of 0.03446, indicating high accuracy as the predicted values closely follow the target values, and the error distribution is tightly centered around zero.

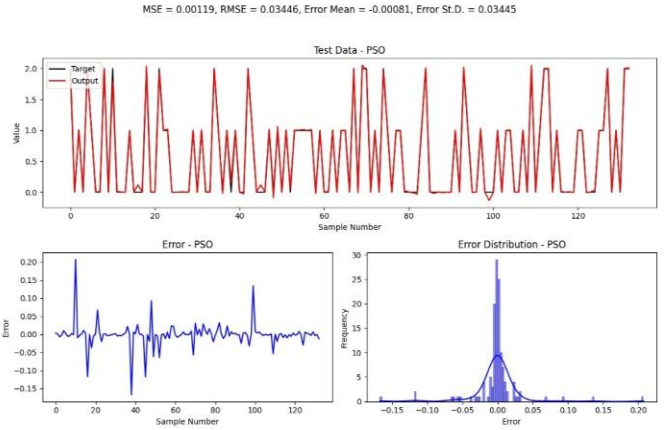


Fig. 11. PSO optimization results: test data, error, and error distribution.

The GWO model follows, with slightly higher MSE of 0.00132 and RMSE of 0.03627, showing more variation and a wider error spread compared to PSO, but still demonstrating good performance. The Grid Search model exhibits the highest error metrics MSE of 0.00821 and RMSE of 0.09061, with significant deviations and a broad error distribution, indicating poor alignment between predicted and target values and the least consistent performance among the three. Therefore, PSO is the most effective optimization method, followed by GWO, while Grid Search is the least effective.

V. Conclusion

The approach presented in this study, designed for continuous risk management in BPR projects, offers a comprehensive method that combines supervised and unsupervised machine learning techniques for effective risk management. The study present the dynamic nature of business processes and the importance of adopting strategies tailored to organizational changes. This hybrid approach leverages the strengths of unsupervised learning to uncover hidden relationships within the data and supervised learning for predictive modeling.

Algorithms such as K-means and t-SNE for clustering and visualization, and XGBoost for classification and regression, in this approach, aims to provide a robust framework for risk assessment. The optimization of XGBoost using advanced techniques like PSO, GWO and Grid Search further enhances the model's accuracy and reliability.

The article underscores the importance of using Mahalanobis distance due to its ability to consider variable correlations, which is crucial for accurate risk assessment. The authors applied their methodology to a real-world risk database in the automotive sector, demonstrating the practical applicability and effectiveness of their approach.

The results indicate that PSO outperforms other optimization methods in terms of accuracy, followed by GWO and Grid Search. The study concludes that adopting a hybrid machine learning approach can significantly improve the detection, evaluation, and mitigation of risks in BPR projects, ultimately contributing to the success of such initiatives.

In summary, the article emphasizes the value of combining supervised and unsupervised learning techniques, along with advanced optimization methods, to manage risks in BPR projects effectively. This integrated approach not only enhances predictive accuracy but also provides valuable insights into the underlying risk patterns, facilitating proactive and informed decision-making.

References

[1] G Harmon, P. Business Process Change: A Business Process Management Guide for Managers and Process Professionals; Morgan Kaufmann, Publishers: Burlington, MA, USA, 2019.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] P., Ehrlich, H.-C., Steinke, T.: ZIB structure prediction pipeline: composing a com- plex biological workflow through web services. In: Nagel, W.E., Walter, W.V., Lehner, W. (eds.) Euro-Par 2006. LNCS, vol. 4128, pp. 1148_1158. Springer, Heidelberg (2006). doi:10.1007/11823285_121.

[3] Nisar, Q.A., Ahmad, S. and Ahmad, U. (2014) 'Exploring factors that contribute to success of business process reengineering and impact of business process reengineering on organizational performance: a qualitative descriptive study on banking sector at Pakistan', Asian Journal of Multidisciplinary Studies, Vol. 2, No. 6, pp.219–224 http://ajms.co.in/sites/ajms/index.php/ajms/article/viewFile/405/365.

[4] Tsakalidis, G.; Vergidis, K. Towards a Comprehensive Business Process Optimization Framework. In Proceedings of the 2017 IEEE 19th Conference on Business Informatics (CBI), Thessaloniki, Greece, 24–27 July 2017; IEEE: Piscataway, NJ, USA, 2017; Volume 1, pp. 129–134. doi: 10.1109/CBI.2017.39Y.

[5] Ghanadbashi, S. and Ramsin, R. (2016) 'Towards a method engineering approach for business process reengineering', IET Software, Vol. 10, No. 2, pp.27–44, doi: 10.1049/ietsen.2014.0223.

[6] Erim, A. and Vayvay, O. (2010) 'Is the business process reengineering (BPR) proved itself to be a trustable change management approach for multinational corporations' case studies from the literature', Journal of Aeronautics and Space Technologies, Vol. 4, No. 4, pp.23–30.

[7] Bhaskar, H.L. (2016) 'A critical analysis of information technology and business process reengineering', Int. J. Productivity and Quality Management, Vol. 19, No. 1, pp.98–115. doi:10.1504/IJPQM.2016.078018.

[8] Alghamdi, H.A., Alfarhan, M.A. and Abdullah, A.L. (2014) 'BPR: evaluation of existing methodologies and limitations', International Journal of Computer Trends & Technology, Vol. 7, No. 4, pp.224–227 [online]http://www.ijcttjournal.org/Volume7/number-4/IJCTTV7P154.pdf. doi: 10.14445/22312803/IJCTT-V7P154.

[9] Bhaskar, H.L. (2018) 'Business process reengineering framework and methodology: a critical study', Int. J. Services and Operations Management, Vol. 29, No. 4, pp.527–556. doi:10.1504/IJSOM.2018.090456.

[10] Yin, G. (2010) 'BPR application', Modern Applied Science, Vol. 4, No. 4, pp.96–101. doi: http://dx.doi.org/10.5539/mas.v4n4p96.

[11] Hammer, M. and Champy, J. (1993) 'Reengineering the corporation: a manifesto for business revolution', Business Horizons, Vol. 36, No. 5, pp.90–91, ISBN: 9781857880977.

[12] Eke, G.J. and Achilike, A.N. (2014) 'Business process reengineering in organizational performance in Nigerian banking sector', Academic Journal of Interdisciplinary Studies, Vol. 3, No. 5, pp.113–124, doi: http://dx.doi.org/10.5901/ajis.2014.v3n5p113.

[13] Mlay, S.V., Zlotnikova, I. and Watundu, S. (2013) 'A quantitative analysis of business process reengineering and organizational resistance: the case of Uganda', The African Journal of Information Systems, Vol. 5, No. 1, pp.1–26.

[14] Maaten, L. v. d., & Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research, 9(Nov), 2579-2605.

[15] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794). ACM.

[16] Holland, J. H. (1992). Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT Press.

[17] Jamali, G., Abbaszadeh, M.A., Ebrahimi, M. and Maleki, T. (2011) 'Business process reengineering implementation: developing a causal model of critical success factors', International Journal of e-Education, e-Business, e-Management and e Learning, Vol. 1, No. 5, pp.354–358. doi:10.7763/IJEEEE.2011.V1.58.

[18] Hicham, R., Abdallah, L., Mohamed, M. (2024). Agile Framework that Integrates Continuous Risk Management for the Implementation of BPR. In: Ezziyyani, M., Kacprzyk, J., Balas, V.E. (eds) International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD'2023). AI2SD 2023. Lecture Notes in Networks and Systems, vol 930. Springer, Cham. https://doi.org/10.1007/978-3-031-54318-0_24A. Cheriyadat, and L. M. Bruce, "Why principal Component analysis is not an Appropriate feature extractionmethod for hyperspectral data," in Proceedings of the IEEE Geosci. and remote Sens. Symp., 2003, pp. 3420–3422.

[19] Baumeister, J., Seipel, D., & Puppe, F. (2009). Agile development of rule systems. Dans Handbook of Research on Emerging Rule-Based Languages and Technologies: Open Solutions and Approaches (1nd Edi., Vol. 1, pp. 253-272). United States of America: IGI Global.

[20] Jifa, G., & Lingling, Z. (2014). Data, DIKW, Big Data and Data Science. Procedia Computer Science, 31, 814-821. doi:10.1016/j.procs.2014.05.332.

[21] Choi, T.-M., Chan, H. K., & Yue, X. (2017). Recent Development in Big Data Analytics for Business Operations and Risk Management. IEEE Transactions on Cybernetics, 47(1), 81-92. doi: 10.1109/tcyb.2015.2507599.

[22] Zhang, Y., Ren, S., Liu, Y., & Si, S. (2017). A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. Journal of Cleaner Production, 142, 626-644. doi: 10.1016/j.jclepro.2016.07.123.

[23] Hicham, R.; Abdallah, L.;Mohamed, M. Risk Management and Assessment Hybrid Framework for Business Process Reengineering Projects: Application in Automotive Sector. Eng 2024, 5, 1360–1381.https://doi.org/10.3390/eng5030071.

[24] X. Jia, and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remotesensing image display and classification," IEEE Trans. Geosci. Remote Sens., Vol. 37, no. 1, pp. 538–542, 1999.

[25] Cheng, Ying, Ken Chen, Hemeng Sun, Yongping Zhang, and Fei Tao. "Data and knowledge mining with big data towards smart production." Journal of Industrial Information Integration 9 (2018) : 1-13.

[26] Tsiptsis, Konstantinos K., and Antonios Chorianopoulos. "Data mining techniques in CRM : inside customer segmentation." (2011).

[27] F. Iglesias, W. Kastner, Analysis of similarity measures in times series clustering for the discovery of building energy patterns, Energies 6 (2) (2013) 579–597.

[28] Fan, C., Yan, D., Xiao, F., Li, A., An, J., & Kang, X. (2020, October). Advanced data analytics for enhancing building performances: From data-driven to big datadriven approaches. In Building Simulation (pp. 1-22). Tsinghua University Press.

[29] P. Kar, A. Shareef, A. Kumar, K.T. Harn, B. Kalluri, S.K. Panda, ReViCEE: A recommendation based approach for personalized control, visual comfort & energy efficiency in buildings, Build. Environ. 152 (2019) 135–144.

[30] R. De Maesschalck, D. Jouan-Rimbaud, D.L. Massart, The mahalanobis distance, Chemometrics and intelligent laboratory systems 50 (1) (2000) 1–18.

[31] R. Ruiz de la Hermosa González-Carrato, Wind farm monitoring using Mahalanobis distance and fuzzy clustering, Renewable Energy 123 (2018) 526–540.

[32] P. Westermann, C. Deb, A. Schlueter, R. Evins, Unsupervised learning of energy signatures to identify the heating system and building type using smart meter data, Appl. Energy 264 (2020) 114715, https://doi.org/10.1016/j.apenergy.2020.114715.

[33] Y. Xu, M. Zhang, L. Ye, Q. Zhu, Z. Geng, Y.L. He, Y. Han, A novel prediction intervals method integrating an error & self-feedback extreme learning machine with particle swarm optimization for energy consumption robust prediction, Energy 164 (2018) 137–146.

[34] Ao Li, Cheng Fan, Fu Xiao, Zhijie Chen, Distance measures in building informatics: An in-depth assessment through typical tasks in building energy management, Energy and Buildings, Volume 258, 2022, 111817, ISSN 0378-7788, https://doi.org/10.1016/j.enbuild.2021.111817.

[35] Christy, A. Joy, A. Umamakeswari, L. Priyatharsini, and A. Neyaa. "RFM ranking–An effective approach to customer segmentation." Journal of King Saud University-Computer and Information Sciences 33, no. 10 (2021) : 1251-1257.

[36] Syakur, M. A., B. K. Khotimah, E. M. S. Rochman, and Budi Dwi Satoto. ""Integrationk-means clustering method and elbow method for identification of the best customer profile cluster."" In IOP conference series : materials science and engineering, vol. 336, no. 1, p. 012017. IOP Publishing, 2018.

[37] M. K. Pakhira, S. Bandyopadhyay, et U. Maulik. Validity index for crisp and fuzzy clusters. Pattern Recognition, 37(3) :487 — 501, 2004.

[38] L. Kaufman et P. Rousseeuw. Finding Groups in Data An Introduction to Cluster Analysis. Wiley Interscience, New York, 1990.

[39] Maaten, L. v. d., & Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research, 9(Nov), 2579-2605.

[40] Linderman, G. C., Rachh, M., Hoskins, J. G., Steinerberger, S., & Kluger, Y. (2019). Fast interpolation-based t-SNE for improved visualization of single-cell RNA-seq data. Nature Methods, 16(3), 243-245.

[41] Wu, J., Xia, H., Cheng, Q., & Li, L. (2017). Application of Machine Learning Algorithms to Predict Central Neuropathic Pain in People with Spinal Cord Injury. Journal of Pain Research, 10, 1627-1634.

[42] Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. the 22nd ACM SIGKDD International Conference. 2016: 785-794.

[43] S. Shoghian, M. Kouzehgar, A comparison among wolf pack search and four other optimization algorithms, Int. J. Comput. Inf. Eng. 6 (2012) 1619–1624.

[44] S. Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw.69(2014)46–61, http://dx.doi.org/10.1016/J.ADVENGSOFT.2013.12.007.

[45] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of the International Conference on Neural Networks, 1995, pp. 1942–1948.

Enhancing CURE Algorithm with Stochastic Neighbor Embedding (CURE-SNE) for Improved Clustering and Outlier Detection

Dewi Sartika Br Ginting*, Syahril Efendi, Amalia, Poltak Sihombing
Department of Computer Science and Technology Information, Universitas Sumatera Utara, Medan, Indonesia

***Abstract*—This study focuses on analyzing stunting data using the CURE and CURE-SNE algorithms for clustering and outlier detection. The primary challenge is identifying patterns in stunting data, which includes variables such as age, gender, height, weight, and nutritional status. Both algorithms were employed to group the data and detect outliers that may affect the results of the analysis. The evaluation methods included determining the optimal number of clusters using the silhouette score and assessing cluster quality using the Davies-Bouldin Index (DBI). The results showed that both algorithms formed four clusters, with CURE-SNE detecting 6,050 outliers, while CURE detected 5,047 outliers. Silhouette score analysis revealed that both algorithms formed four optimal clusters. However, when validated using DBI, CURE achieved a score of 0.523, while CURE-SNE produced a lower score of 0.388, indicating that CURE-SNE outperformed CURE in terms of cluster quality. This suggests that CURE-SNE not only detects more outliers but also produces clusters with better separation and compactness. The findings highlight that both algorithms are effective for clustering stunting data, but CURE-SNE excels in terms of outlier detection and overall cluster quality. Thus, CURE-SNE is more suitable for handling complex datasets with potential outliers, providing more accurate insights into the structure of the data. In conclusion, CURE-SNE demonstrates superior performance compared to CURE, offering a more reliable and detailed clustering solution for stunting data analysis.**

***Keywords*—Stunting; clustering algorithm; CURE; CURE-SNE; outliers**

I. INTRODUCTION

Dataset readiness is a crucial step in the clustering process, because clean and structured data will produce more accurate and reliable clusters. Optimal dataset readiness is also an important foundation for producing accurate and representative clusters. In the clustering process, detecting and handling outliers is a critical step to prevent biased and unreliable results. In its application, outlier detection is very important, because outliers often cause distortion in cluster formation, especially when the distribution of the dataset is not uniform, resulting in the formation of less accurate clusters and even potentially influencing business decisions or misguided analysis. Outliers can come from recording errors, irrelevant data, or extreme variations that do not reflect the general trend of the dataset. Without proper detection and handling, outliers can attract cluster centers or blur the boundaries between clusters, thus making clustering results less than optimal and even misleading. A recent study published in IEEE Transactions on Knowledge and Data Engineering (2021) [1] emphasizes that the presence of outliers in a retail company's customer dataset interferes with the interpretation of customer segments and leads to less representative clustering results. Another in the Journal of Cleaner Production (2021) [2] shows that clustering algorithms density-based ones, such as DBSCAN, have better performance in automatically detecting outliers than conventional algorithms, thereby increasing the accuracy of segmentation results in user behavior analysis.

The CURE (Clustering Using Representatives) algorithm is a clustering approach that is superior in handling large datasets and has the ability to detect non-spherical cluster shapes, and is more resistant to outliers than classical algorithms such as K-Means. CURE works by selecting multiple point representations from each cluster, then compressing (shrinking) these points to the center of the cluster to increase robustness against outliers. However, despite these advantages, CURE still has limitations in handling large datasets optimally, especially if there are many outliers. The use of sampling in CURE to reduce the scale of large datasets can result in important information being missed or even failing to identify relevant outliers. Additionally, point compression methods sometimes sacrifice certain details that are actually important in complex datasets, so clustering results may not always be optimal.

The development of the CURE algorithm to overcome the challenges of large datasets and outliers is increasingly becoming the main focus in various research due to the need for more accurate and efficient clustering. One of the newest approaches that is being developed is a combination of CURE with machine learning models, such as using autoencoders to identify complex features and reduce data dimensions before clustering. In this way, the algorithm can remove noise and outliers more effectively, while preserving important information in large datasets. Additionally, research in ACM Transactions on Knowledge Discovery from Data (2022) [3] shows that integration between CURE and density-based models, such as DBSCAN, produces better clustering results on high-density data, where outliers can cause significant distortion. This approach helps CURE group data more precisely in dense areas, while isolating outliers. Furthermore, experiments on large-scale datasets show that implementing CURE in distributed computing environments, such as Apache Hadoop and Spark, allows these algorithms to handle very large datasets more quickly without sacrificing accuracy. The use of a platform like this also allows the development of more adaptive CURE algorithms, for example by automating the selection

parameters of point representations and compression measures, which are instrumental in avoiding bias from outliers.

Several recent studies have made significant contributions to the development of the CURE algorithm to improve its performance in handling large datasets and outliers problems. Research in IEEE Transactions on Big Data (2021) [26] [4] developed CURE by leveraging Apache Spark, which speeds up processing of large datasets and reduces the impact of outliers on clustering results. On the other hand, a study in ACM Transactions on Knowledge Discovery from Data (2022) [3] combined CURE with density-based DBSCAN to pre-separate outliers, which proved effective for anomaly detection in network data. Additionally, research in the Journal of Cleaner Production (2021) [2] introduces a hybrid CURE and Isolation Forest approach for handling high-dimensional data, thereby increasing precision in customer segmentation and fraud detection. Dimensionality reduction techniques using PCA before applying CURE, as proposed in Data Mining and Knowledge Discovery (2021) [5], also help simplify the data and reduce the effects of outliers, increasing clustering efficiency. Meanwhile, research in Information Sciences (2019) [6] developed an adaptive version of CURE for distributed computing environments, such as Hadoop and Spark, with optimized parameters to be more robust to outliers in large datasets. These studies emphasize the importance of developing CURE to be more efficient and accurate in real applications on large and complex data.

To address these limitations, this research introduces CURE-SNE, an enhanced version of CURE that integrates Stochastic Neighbor Embedding (SNE) for dimensionality reduction. SNE optimizes the mapping of high-dimensional data into a lower-dimensional space, effectively preserving local and global structures. This integration allows CURE-SNE to detect complex patterns and improve the identification of outliers, resulting in more accurate clustering results. By leveraging SNE's ability to emphasize neighborhood relationships, CURE-SNE enhances CURE's robustness and accuracy in clustering and outlier detection.

This paper is organized as follows:

- Section II discusses related work and advancements in CURE-based clustering methods.
- Section III describes the proposed CURE-SNE methodology and its implementation.
- Section IV presents experimental results, evaluated using *Silhouette Score* and *Davies-Bouldin Index* to assess clustering quality.
- Section V provides a discussion on the comparative analysis between CURE and CURE-SNE.
- Section VI concludes the study, emphasizing CURE-SNE's contributions to clustering accuracy and outlier detection.

The integration of SNE into the CURE algorithm represents a significant step toward improving clustering performance, particularly for datasets with complex structures and a high number of outliers.

II. Related Work

In recent years, clustering research has made significant strides in addressing critical challenges, particularly in dealing with the increasing complexity of modern datasets. Key issues such as handling large-scale datasets, managing uneven data distributions, and mitigating the disruptive impact of outliers have been at the forefront of this research. Among the most robust and widely recognized clustering algorithms designed to tackle these challenges is the CURE algorithm, short for Clustering Using Representatives. CURE stands out for its ability to effectively handle non-spherical cluster shapes and exhibit strong resistance to outliers. However, despite its robustness, CURE is not without its limitations. One of its primary challenges lies in detecting subtle, complex, and high-dimensional outlier patterns, which can significantly distort clustering outcomes if not adequately addressed.

To overcome these limitations, researchers have proposed and implemented numerous enhancements and hybrid adaptations of the CURE algorithm. For instance, the integration of CURE with Gaussian Mixture Models (GMM) has been shown to greatly improve the detection and representation of outlier patterns in high-dimensional datasets [7]. This hybrid approach has proven particularly effective in domains where data complexity is a significant factor. Similarly, the combination of CURE with Support Vector Data Description (SVDD) has resulted in a highly effective anomaly detection framework specifically tailored for network data analysis [8]. This framework capitalizes on the strengths of both density-based clustering and boundary-based detection techniques to improve the reliability of clustering results in such specialized domains.

Another major challenge associated with clustering large datasets is their high dimensionality, which can complicate data representation and analysis. To address this issue, dimensionality reduction techniques have been successfully employed. For example, research has demonstrated that combining CURE with Principal Component Analysis (PCA) not only reduces the dimensionality of the data but also preserves critical information essential for accurate clustering [9]. This combination enhances the clustering algorithm's performance and ensures better scalability for handling extensive datasets. Additionally, recent advancements have showcased the implementation of CURE on the Apache Flink framework, which is a powerful distributed computing system. This adaptation has resulted in a 45% improvement in processing speed compared to traditional methods, while maintaining sensitivity to outliers, making it an ideal solution for real-time and large-scale data processing needs [10].

The integration of deep learning with clustering algorithms has also opened up new opportunities for innovation. Hybrid models that combine CURE with deep learning techniques, such as autoencoders, have significantly enhanced the algorithm's ability to detect complex, nonlinear relationships in high-dimensional data. For example, in the context of e-commerce applications, this combination has led to remarkable improvements in clustering precision and the detection of subtle outlier patterns [13]. Similarly, hybrid approaches like CURE-DBSCAN [11] and CURE-SNE [12] leverage density-based

clustering and dimensionality reduction techniques, respectively, to handle datasets with intricate structures. These advancements underscore the adaptability, precision, and efficiency of modern clustering algorithms in addressing the challenges posed by diverse and complex datasets.

These developments are not only theoretical but also have practical implications across various fields, including health data analysis and decision-making systems. For instance, Ginting et al. (2023) [28] explored the application of fuzzy logic methods to predict neurotic disorder types. This study emphasized the critical role of precision in computational methods when dealing with sensitive medical data. Similarly, Ginting et al. (2024) [29] developed a perceptron neural network model for predicting postpartum depression. This research demonstrated the significant potential of hybrid and advanced computational techniques in addressing public health challenges and improving the accuracy of predictive analytics.

Moreover, in the domain of decision-making systems, Ginting et al. (2021) [30] introduced an innovative integration of the AHP and TOPSIS methods. This approach optimized the performance of decision support systems for identifying recipients of the Family Hope Program. This combination of multi-criteria decision-making techniques aligns closely with the broader theme of leveraging diverse computational methods to achieve optimal outcomes in complex and multi-dimensional datasets. Such methodologies not only enhance the accuracy of decision-making systems but also ensure scalability and reliability across various application areas.

In summary, the continuous evolution of the CURE algorithm and its hybrid adaptations reflects the growing demand for advanced clustering techniques capable of addressing the ever-increasing complexity of real-world datasets. These innovations have proven to be invaluable tools across multiple domains, providing practical solutions for challenges ranging from network security and anomaly detection to large-scale public health analysis and market segmentation.

III. METHODOLOGY

CURE (Clustering Using Representatives) algorithm is a clustering algorithm designed to handle large datasets and is able to work with data that has a non-spherical cluster shape, while reducing the impact of outliers. [14] One of the main advantages of CURE is its unique approach of representing each cluster with several representative points carefully selected from the data, rather than just one central point as in K-Means. [15] CURE has the advantage of generating clusters of various shapes and sizes, which makes it very effective in the analysis of complex data, such as spatial data or data that is not symmetrically distributed. By using these representative points, CURE maintains flexibility in grouping data that does not conform to simple distribution assumptions, thereby providing more robust clustering results [16].

The main steps in the CURE algorithm consist of several key stages, namely sampling, initial cluster formation, selection of representative points, and compression of representative points towards the cluster center. First, the algorithm samples the data to reduce the number of points that need to be processed, thereby speeding up computing. After that, the sampled data is grouped into initial clusters using a hierarchical clustering approach. [25] Next, for each cluster formed, the algorithm selects a number of representative points (usually several points in the area around the cluster) which will be used to describe the characteristics of the cluster. These representative points were chosen to define the shape and boundaries of the clusters more clearly, including clusters that have asymmetrical shapes [17].

To be more resistant to outliers, CURE applies a compression technique (shrinkage) to each representative point, namely shifting these points towards the cluster center by a certain factor [19]. Suppose C is a cluster with a center of mass μ and a representative point Ri (with i referring to the ith representative point in cluster C), then each point Ri is compressed towards the center μ using a shrinkage factor α which satisfies $0< \alpha <1$. The formula for moving the representative point Ri to Ri' is as given in Eq. (1):

$$Ri' = \mu + \alpha (Ri - \mu) \quad (1)$$

Ri' : New representative point after compression.

μ : The cluster center point (centroid) of cluster C.

Ri : The initial representative point selected for the cluster C.

α : Compression factor, where $0<\alpha<1$.

The parameter α controls how far the representative points will be compressed towards the cluster center. For example, a value of $\alpha = 0.5$ means that each representative point is shifted towards the center by 50% of its distance to the center. Here, the value of α plays an important role in determining how far the representative point moves towards the center. The larger the α value, the greater the influence of the cluster center on the representative point, which reduces the effect of outliers on the cluster. By performing shrinkage, the CURE algorithm reduces the influence of outliers located far from the cluster center, thereby increasing cluster stability and producing more accurate results [27].

Distance Measurement between Clusters (Hierarchical Clustering): In the CURE algorithm, clusters are initially generated through a hierarchical clustering approach. To combine two clusters, CURE measures the distance between two clusters Ci and Cj based on the closest representative point in each cluster. For example, if the representative points of cluster Ci are $\{ri1, ri2, \ldots, rim\}$ and of cluster Cj are $\{rj1, rj2, \ldots, rjn\}$, then the distance between clusters is calculated as in Eq. (2):

$$d(Ci, Cj) = \min_{p \in Ci, q \in Cj} \|rp - rq\| \quad (2)$$

$d(Ci, Cj)$: Distance between clusters Ci and Cj.

rp : Representative point in cluster Ci.

rq : Representative point in cluster Cj.

$\|rp - rq\|$: Euclidean distance between rp and rq.

This distance measure determines how close two clusters are to each other, so CURE can decide whether two clusters should be combined.

Centroid or cluster center for representative point compression, CURE calculates cluster centers using centroids. If cluster C has data points {x1,x2,…,xn}, then the cluster center μ can be calculated using the Eq. (3):

$$\mu = \frac{1}{n}\sum_{k=1}^{n} Xk \quad (3)$$

μ : Centroid or cluster center *C.*

Xk : *K* data point in the cluster *C.*

N : Number of data points in the cluster *C.*

These cluster centers are used to determine the direction and compression level of representative points.

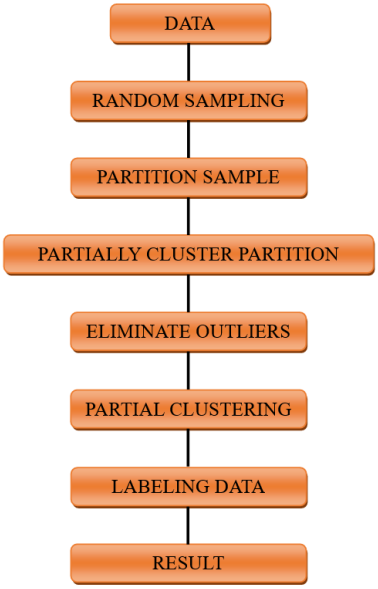


Fig. 1. CURE algorithm.

Fig.1 is the pseudocode for the CURE algorithm.

TABLE I. CURE ALGORITHM FOR CLUSTERING

Step	Description
1	Choose representative points for each cluster
2	For each point, calculate the distance to the representative points.
3	Cluster points based on the minimum distance to the representatives.
4	Repeat steps 2 and 3 until convergence or a stopping criterion is met.
5	Remove outliers: points that are far from any cluster representatives.
6	Output the final cluster with their representative points.

In Table I, summarizes the key steps involved in the CURE algorithm for clustering, focusing on the selection of representative points and the process of refining the clusters while detecting outliers.

A. Cure-SNE

CURE-SNE (Clustering Using Representative Objects with Stochastic Neighbor Embedding) is an advanced clustering algorithm that combines the strengths of CURE with the dimensionality reduction technique of Stochastic Neighbor Embedding (SNE) [18]. While CURE focuses on selecting representative points to form clusters, CURE-SNE enhances this process by first mapping the data into a lower-dimensional space using SNE. This transformation helps reveal complex patterns and relationships that may not be apparent in higher-dimensional spaces [20]. In CURE-SNE, the clustering is performed by calculating the distance between data points and the representative points in this reduced space, making the algorithm more sensitive to underlying structures. One of the key advantages of CURE-SNE is its ability to detect and handle outliers more effectively. By identifying points that are far from any cluster representatives in the low-dimensional space, CURE-SNE ensures that these outliers are excluded from the final clusters, resulting in more accurate and refined clustering outcomes. This hybrid approach makes CURE-SNE particularly useful for datasets with complex structures or a high number of outliers.

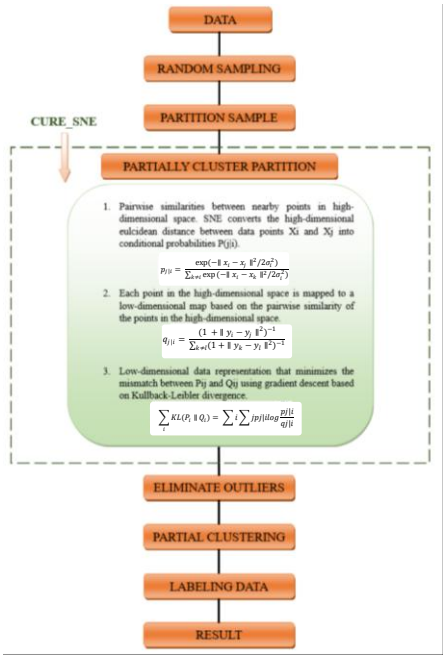


Fig. 2. CURE-SNE algorithm.

In Fig. 2, outlines the key steps involved in the CURE-SNE algorithm, emphasizing the integration of SNE to map the data to a lower-dimensional space and refining the clustering process by detecting outliers effectively.

B. Clustering Evaluation

The clustering evaluation methodology in this research uses two main metrics: Silhouette Score and Davies-Bouldin Index (DBI). The Silhouette Score is used to measure the extent to which each data point is separated from other clusters, with values ranging from -1 to 1. A positive value close to 1 indicates that the data point is well located in the right cluster, while a value close to -1 indicates that the closer to the wrong cluster. This process is carried out by calculating the average distance between each data point to other points in the same cluster, as well as the average distance to the closest point in another cluster. Meanwhile, DBI evaluates the quality of clustering by comparing the distance between clusters with the size of the cluster itself. Lower DBI indicates better separation between

clusters and higher compactness. These two metrics provide a comprehensive picture of the effectiveness of the applied clustering algorithm, thereby allowing the selection of the optimal clustering model based on the structure of the analyzed data.

1) Silhouette score: Silhouette Score is used to measure the extent to which each data point is separated from other clusters. [21] The formula for calculating the Silhouette Score $S(i)$ for data point i is as given in Eq. (4):

$$S(i) = \frac{b(i)-a(i)}{max(a(i),b(i))} \quad (4)$$

where, $a(i)$ is the average distance between point i and all other points in the same cluster, and $b(i)$ is the average distance between point i and the nearest point in another cluster. The Silhouette Score value ranges from -1 to 1; a positive value close to 1 indicates that the data point is well located in the correct cluster, while a value close to -1 indicates that the point is closer to the incorrect cluster. [22]

2) Davies-Bouldin Index (DBI): The Davies-Bouldin Index (DBI) evaluates the quality of clustering by comparing the distance between clusters with the size of the cluster itself. [23] The DBI formula for C cluster is given in Eq. (5):

$$DBI = \frac{1}{C}\sum_{i-1}^{C} \max(\frac{Si+Sj)}{Dij}) \quad (5)$$

where, Si is the size (in terms of distance) of cluster i, Sj is the size of cluster j, and Dij is the distance between the center of cluster i and the center of cluster j [24]. Lower DBI indicates better separation between clusters and higher compactness. These two metrics provide a comprehensive picture of the effectiveness of the applied clustering algorithm, thereby allowing the selection of the optimal clustering model based on the structure of the analyzed data.

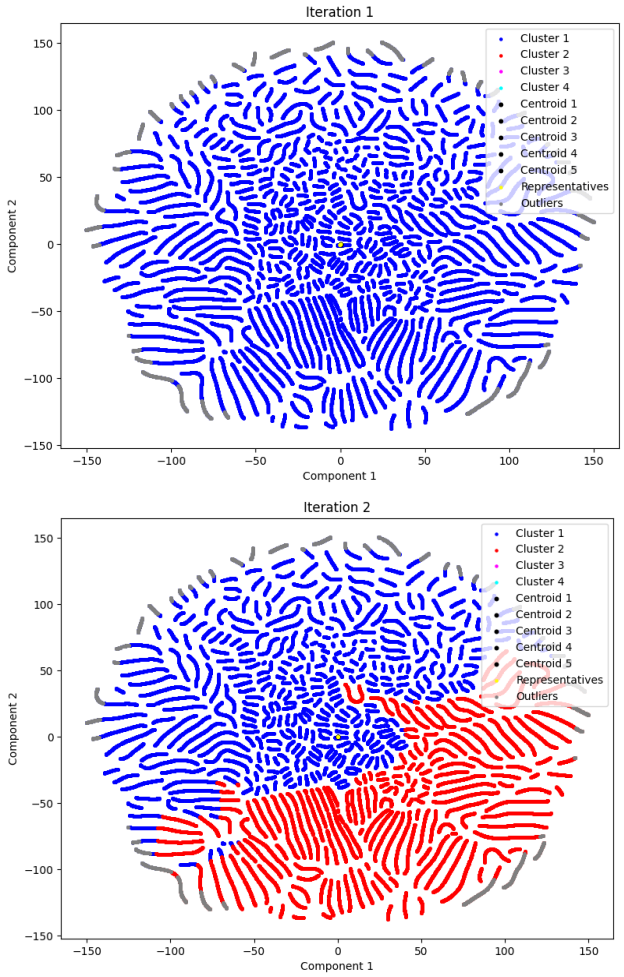
IV. Experimental Results and Analysis

The dataset under analysis contains data related to stunting cases, comprising 121,000 rows, with several key variables essential for health analysis, including age (in months or years), gender (male or female), height (in centimeters), weight (in kilograms), and nutritional status (categories indicating nutritional conditions such as well-nourished, undernourished, or stunted). The aim of analyzing this dataset is to understand the patterns and factors associated with the occurrence of stunting, which can assist in formulating more effective interventions or policies to address this public health issue. To ensure accurate and reliable analysis, outlier detection and handling will be performed on the data. Outliers, which may appear as extreme or anomalous values in certain variables, can significantly impact the results of the analysis. By identifying and addressing outliers, we can ensure that the dataset maintains high quality, allowing for more valid insights into the patterns and factors influencing stunting.

In Table II, a visualization of the clustering results using the CURE (Clustering Using Representatives) model. This model group's data based on representative points that capture the characteristics of each cluster, employing an approach that identifies patterns in the data and handles variations in distribution. The image shows several iterations, resulting in clusters with different characteristics, with each cluster represented by a different color, indicating groups of data with similarities in the analyzed variables. This visualization helps in understanding the distribution and patterns within the cluster space generated by the CURE algorithm.

TABLE II. Table Datasets

No	Age (month)	Gender	Height (cm)	Weight (kg)	Nutritional Status
1	18	Boy	80.5	10.1	Stunted
2	23	Girl	101.2	16	Over
3	18	Boy	74.1	7.2	Severely Stunted
4	30	Boy	102	16.4	Over
5	8	Boy	76.1	8.0	Normal
6	2	Boy	52.0	3.3	Normal
7	32	Boy	101.9	17.6	Over
8	24	Girl	100.0	15.1	Over
9	50	Boy	112.2	21.0	Over
10	18	Boy	75.7	8.2	Severely Stunted
...	...	...	...	...	...
120.999	5	Boy	50.0	4.1	Severely Stunted



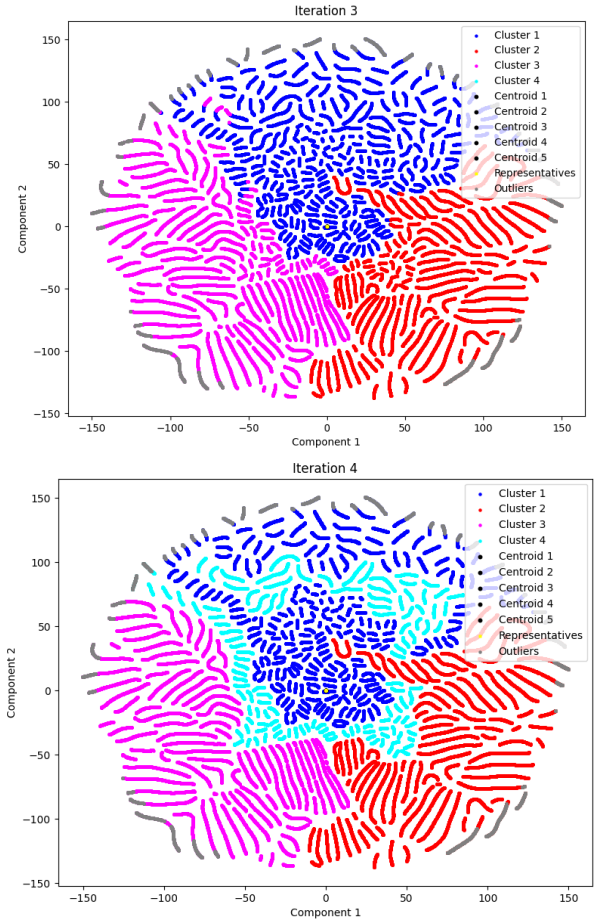


Fig. 3. Visualization of clustering iteration with CURE.

Fig. 3 is a visualization of the clustering results using the CURE-SNE (Clustering Using Representatives and Stochastic Neighbor Embedding) model. This model groups data based on the proximity between points, employing an approach that effectively handles outliers. The image shows four iterations, resulting in four distinct clusters, with each cluster represented by a different color, indicating groups of data with similar characteristics based on the analyzed variables. This visualization aids in understanding the distribution and patterns within the cluster space generated by the algorithm.

In Fig. 4:

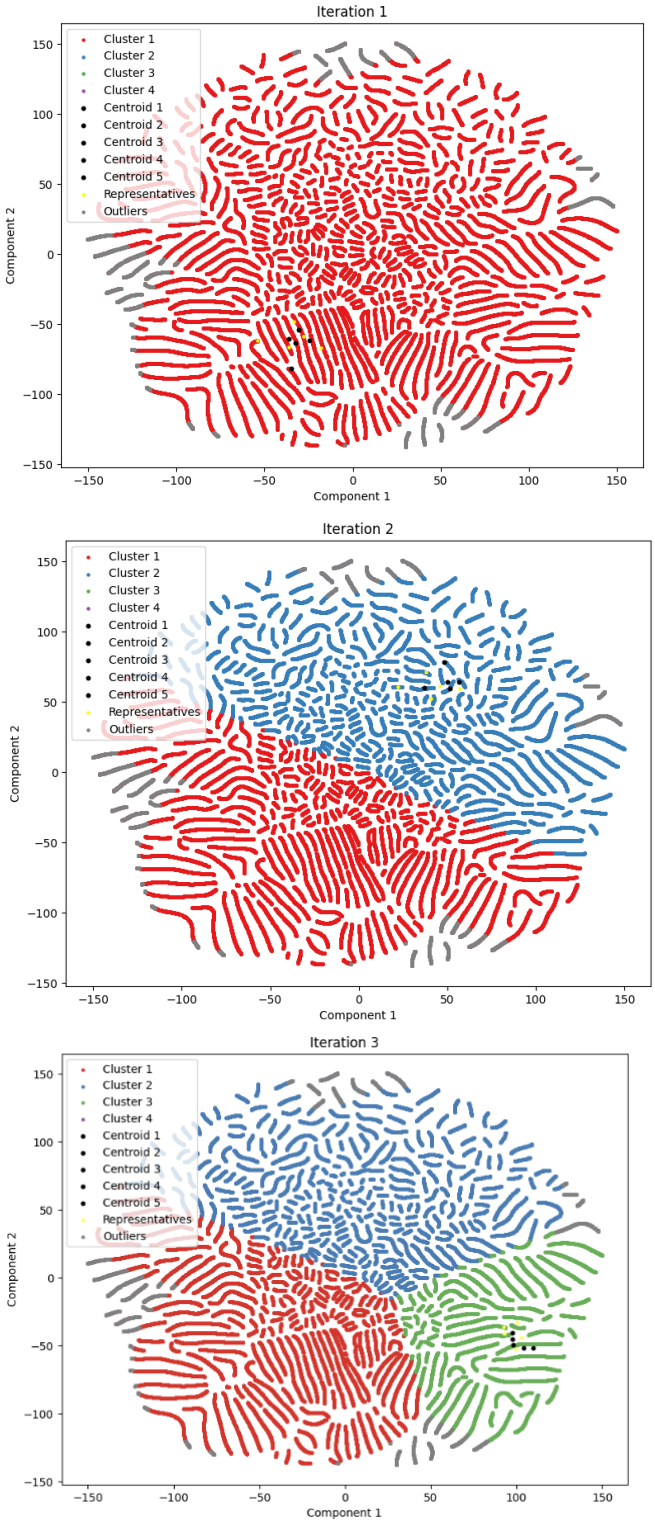
Iteration 1: In the first iteration, the CURE-SNE algorithm initializes 121K clusters based on the data, just as in the original dataset. Each data point represents a separate cluster.

Iteration 2: In the second iteration, the CURE-SNE algorithm begins merging nearby clusters. Clusters that are close to each other in the SNE visualization tend to share similar characteristics in the original data, leading them to merge into larger clusters.

Iteration 3: The merging of clusters continues in the next iteration. Clusters that exhibit similar patterns in the SNE space, as indicated by their proximity in the visualization, continue to merge. The size of the dominant clusters increases, while smaller clusters or outliers remain separate.

Iteration 4: In the final iteration, the CURE-SNE algorithm reaches the desired number of clusters, which is 4 clusters.

The following images show the clustering results from both CURE and CURE-SNE, highlighting the differences in the distribution of points for each cluster.



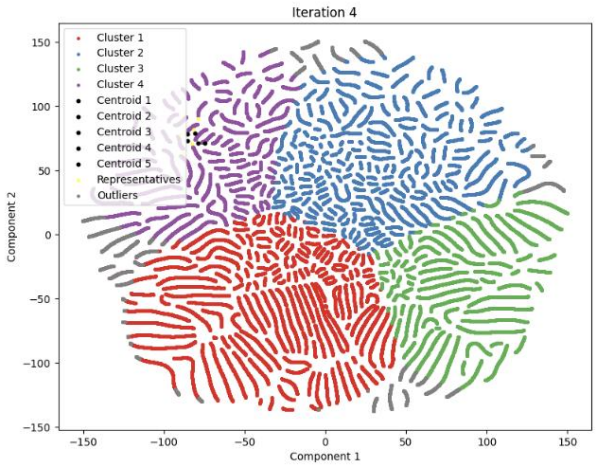


Fig. 4. Visualization of clustering iteration with CURE-SNE.

From Fig. 5 and Fig. 6, these four clusters represent groups of toddlers with distinct characteristics, as described below:

Cluster 1: Male toddlers with a very stunted nutritional status.

Cluster 2: Female toddlers with a normal nutritional status.

Cluster 3: Male toddlers with a normal nutritional status.

Cluster 4: Male toddlers with a high nutritional status.

Table III presents the composition of the clustering results for nutritional status using two different approaches: CURE and CURE-SNE. In the CURE method, the data is grouped based on representative points to capture the cluster structure, while CURE-SNE combines the outlier-handling capabilities of CURE with dimensionality reduction through Stochastic Neighbor Embedding (SNE). Each table displays the number of individuals in each cluster, along with their distribution across nutritional status categories such as well-nourished, undernourished, and stunted, providing valuable insights into the patterns and characteristics within the data.

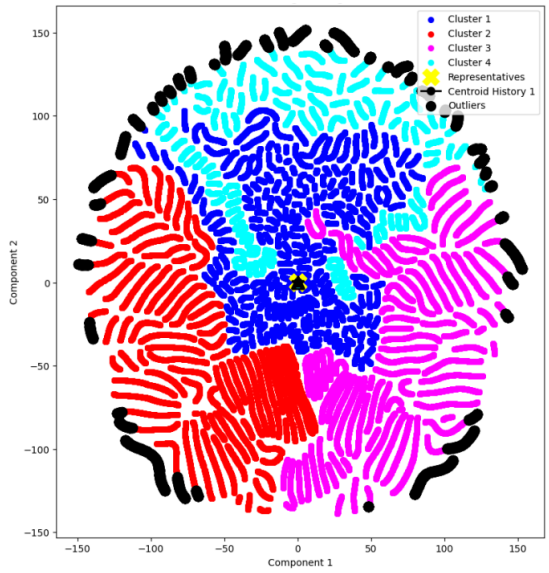


Fig. 5. Clustering results using the CURE.

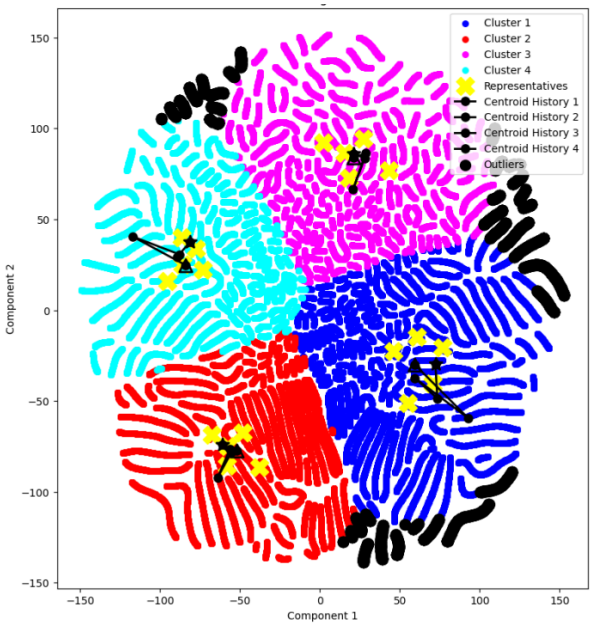


Fig. 6. Clustering results using the CURE-SNE.

TABLE III. CLUSTERING COMPOSITION OF NUTRITIONAL STATUS WITH CURE (%)

Nutritional Status/Cluster	Normal	Severely Stunted	Stunted	Over
1	72.855	1.112	13.015	13.018
2	22.683	32.860	17.248	27.209
3	43.932	29.154	9.078	17.934
4	100.000000	0.000000	0.000000	0.000000

TABLE IV. CLUSTERING COMPOSITION OF NUTRITIONAL STATUS WITH CURE-SNE (%)

Nutritional Status/Cluster	Normal	Severely Stunted	Stunted	Over
1	46.811	3.436	35.665	14.018
2	37.017	24.788	23.076	15.119
3	38.390	29.199	10.078	22.766
4	100.000000	0.000000	0.000000	0.000000

From Table IV, the clustering results using the CURE and CURE-SNE algorithms demonstrate the capabilities of both methods in grouping data while detecting outliers, though with differences in the number of outliers identified. The CURE algorithm effectively clusters data based on representative points within each cluster and detects outliers using a distance-based approach. On the other hand, the CURE-SNE algorithm, which integrates Stochastic Neighbor Embedding (SNE) to optimize the mapping of data in a lower-dimensional space, is able to detect a greater number of outliers compared to CURE. This indicates that CURE-SNE is more sensitive in identifying data points that do not conform to general patterns, resulting in a more detailed clustering outcome.

Fig. 7 and Fig. 8 shows the outlier detection process by CURE, where a total of 5,047 outlier data points were identified. In comparison, CURE-SNE detected 6,050 outliers.

Outliers in data_CURE:

	Age (month)	Gender	Height (cm)	Nutritional Status
12024	6	male	60.6	severely stunted
12028	6	male	59.1	severely stunted
12031	6	male	60.4	severely stunted
12038	6	male	60.7	severely stunted
12041	6	male	59.4	severely stunted
...	...	...	...	...
86829	43	female	111.5	normal
86836	43	female	111.4	normal
86889	43	female	111.4	normal
86929	43	female	111.6	normal
86948	43	female	111.2	normal

[5047 rows x 5 columns]

Number of outliers in original data: 5047

Fig. 7. Outliers detection by CURE.

Outliers in data_CURE-SNE:

	TSNE1	TSNE2
12016	10.998404	134.536285
12022	8.577623	138.188461
12024	7.951420	146.315186
12028	7.508828	142.431824
12031	7.873251	145.992142
...	...	...
119898	-5.304644	-136.068207
119909	-5.924666	-135.762711
119931	-5.790865	-135.835434
119953	-5.790865	-135.835434
119965	-6.047623	-135.675613

[6050 rows x 2 columns]

Number of outliers in data: 6050

Fig. 8. Outliers detection by CURE-SNE.

Fig. 9 presents a comparison of the clustering results using the CURE and CURE-SNE algorithms across several iterations. Both graphs illustrate the changes in cluster sizes over iterations and show how the algorithms detect and handle outliers. CURE uses a distance-based approach to group data and detect outliers, while CURE-SNE combines this approach with dimensionality reduction mapping, resulting in more detailed clustering outcomes.

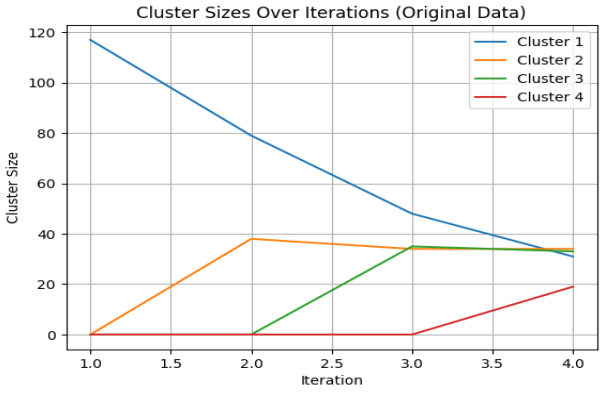


Fig. 9. Graph of CURE.

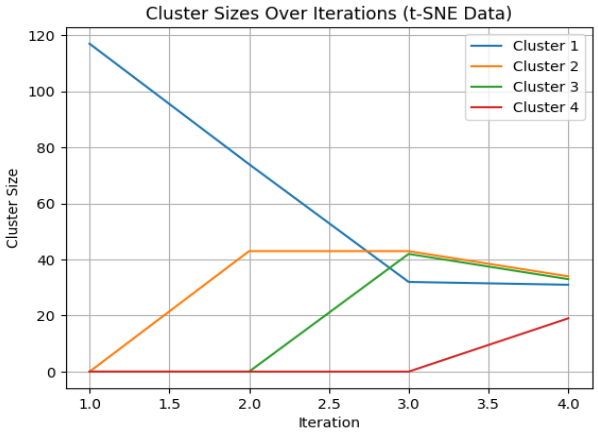


Fig. 10. Graph of CURE-SNE.

Fig. 10 also shows the clustering results using the CURE algorithm on the original data. The initial clusters also decrease in size during the iterations. However, the changes in cluster size are less dramatic, indicating that fewer outliers were detected compared to the CURE-SNE approach.

In the first graph, it can be observed that the CURE-SNE algorithm detects significant changes in cluster sizes over iterations. The initially large clusters gradually break down into smaller clusters, with a clearer data distribution in the final iterations. This indicates CURE-SNE's sensitivity in handling outliers, reflected in the reduction of main cluster sizes and the formation of additional clusters.

The comparison of these two graphs shows that CURE-SNE is generally more effective in detecting outliers and producing clusters with a more segmented data distribution.

It is essential to evaluate the performance of the clustering results generated by both the CURE and CURE-SNE methods to ensure the validity of the cluster structures. In this study, two evaluation metrics are used: the silhouette score, which helps determine the optimal number of clusters by measuring cluster cohesion and separation, and the Davies-Bouldin Index (DBI), which assesses the quality of the clusters by considering their compactness and separation. These evaluations provide valuable insights into the effectiveness of the clustering methods in capturing meaningful patterns in the data.

1) Sillhouette score evaluation: Fig. 11 illustrates the silhouette score analysis for determining the optimal number of clusters generated by both the CURE and CURE-SNE algorithms. The silhouette score, which measures the quality of clustering by assessing the separation and cohesion of clusters, indicates that both methods achieve the highest clustering performance at 4 clusters. As shown, the silhouette score initially increases and peaks at 4 clusters before dropping significantly as the number of clusters increases. This suggests that 4 clusters provide the best balance between intra-cluster cohesion and inter-cluster separation for the given dataset. The similarity in results highlights the effectiveness of both CURE and CURE-SNE in identifying the optimal cluster structure.

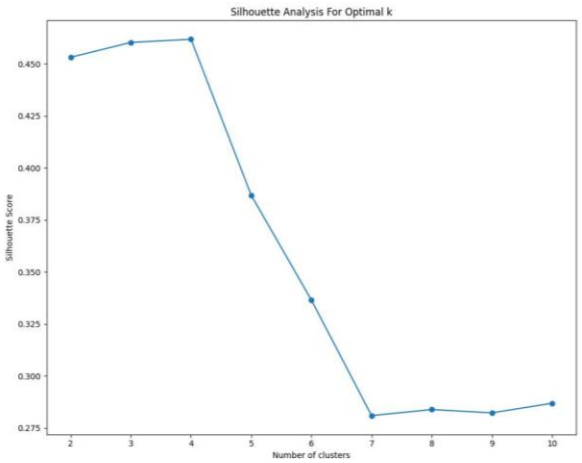


Fig. 11. Graph of sillhouette score evaluation.

2) Davies-Bouldin Index (DBI) Evaluation: From Table V, clustering evaluation can be performed using the Davies-Bouldin Index (DBI), which measures the quality of clusters based on the separation between clusters and the compactness within clusters. The smaller the DBI value, the better the clustering quality, as it indicates well-separated and tightly-knit clusters. In this study, the CURE-SNE method yielded a DBI value of 0.388, which is smaller than the DBI value of 0.523 obtained by the CURE method. This demonstrates that the CURE-SNE method performs better in generating clusters with higher quality, featuring clearer separation between clusters and greater compactness within them.

Evaluation of CURE Clustering:

$R_{12} = 0.314$, $R_{13} = 0.611$, $R_{14} = 0.366$, $R_{23} = 0.431$, $R_{24} = 0.363$, $R_{34} = 0.438$

$D_1 = \max(R_{12}, R_{13}, R_{14}) = \max(0.314, 0.611, 0.366) = 0.611$

$D_2 = \max(R_{21}, R_{23}, R_{24}) = \max(0.314, 0.431, 0.363) = 0.431$

$D_3 = \max(R_{31}, R_{32}, R_{34}) = \max(0.611, 0.431, 0.438) = 0.611$

$D_4 = \max(R_{41}, R_{42}, R_{43}) = \max(0.366, 0.363, 0.438) = 0.438$

$DBI = ¼ (0.611+0.431+0.611+0.438) = 0.523$

Evaluation of CURE-SNE Clustering:

$R_{12} = 0.4$, $R_{13} = 0.203$, $R_{14} = 0.164$, $R_{23} = 0.422$, $R_{24} = 0.272$, $R_{34} = 0.309$

$D_1 = \max(R_{12}, R_{13}, R_{14}) = \max(0.4, 0.203, 0.164) = 0.4$

$D_2 = \max(R_{21}, R_{23}, R_{24}) = \max(0.4, 0.422, 0.272) = 0.422$

$D_3 = \max(R_{31}, R_{32}, R_{34}) = \max(0.203, 0.422, 0.309) = 0.422$

$D_4 = \max(R_{41}, R_{42}, R_{43}) = \max(0.164, 0.272, 0.309) = 0.309$

$DBI = ¼ (0.4+0.422+0.422+0.309) = 0.388$

TABLE V. COMPARISON OF CLUSTER EVALUATION

Algorithm	Sillhouette Score	Davies-Bouldin Index
CURE	4	0.523
CURE-SNE	4	0.388

V. DISCUSSION

The clustering results using the CURE and CURE-SNE algorithms show significant differences in how the two approaches process data and detect outliers. In the CURE algorithm, the clustering process is based on data representation through representative points that reflect the characteristics of each cluster. The results show that this algorithm is effective in grouping data but less sensitive to detecting outliers, with 5047 outliers detected. This results in clusters with more stable data distributions and less drastic changes in size at each iteration.

In contrast, the CURE-SNE algorithm, which combines the CURE representation approach with dimensionality reduction through Stochastic Neighbor Embedding (SNE), exhibits a higher ability to detect outliers. This is evident from the dramatic reduction in the size of large clusters in the early iterations and the formation of smaller, more separated clusters in subsequent iterations. CURE-SNE detected 6050 outliers, which is more than the CURE algorithm. The sensitivity of CURE-SNE in handling outliers makes it more effective at identifying deviating data points, resulting in more segmented clusters.

Overall, both algorithms have their respective advantages. CURE is better suited for clustering data with a more regular distribution and is less influenced by outliers, while CURE-SNE excels in detecting complex patterns and handling data with many outliers. Therefore, the choice of algorithm should be tailored to the characteristics of the dataset and the analysis goals. In this case, the results from CURE-SNE provide deeper insights into the data structure, particularly in the context of identifying significant outliers for further analysis.

VI. CONCLUSION

The clustering analysis results using the CURE and CURE-SNE algorithms provide valuable insights into the capabilities of both methods in grouping data and detecting outliers. Both algorithms resulted in four clusters, but with differences in outlier detection, where CURE-SNE was able to detect 6050 outliers, while CURE detected 5047 outliers. Cluster validation using silhouette score showed that both CURE and CURE-SNE formed four optimal clusters. However, when validated with the Davies-Bouldin Index (DBI), CURE achieved a value of 0.523, while CURE-SNE achieved a value of 0.388, indicating that CURE-SNE outperformed CURE in terms of the quality of the clusters formed. The CURE algorithm demonstrated strong performance in generating stable clusters with well-organized data distributions but had limitations in sensitively detecting outliers. On the other hand, CURE-SNE, with the integration of the Stochastic Neighbor Embedding (SNE) technique, was able to detect more outliers and generate more segmented clusters, reflecting complex patterns within the data. This difference indicates that CURE-SNE is more effective for datasets with irregular distributions or many outliers, while CURE is better suited for data with a more homogeneous structure. Therefore, the choice of algorithm should consider the characteristics of the dataset and the analysis objectives. These findings can serve as a reference for selecting the appropriate clustering method for analyzing complex data, such as public health cases, including identifying factors contributing to stunting.

ACKNOWLEDGMENT

We would like to express our deepest gratitude to the Directorate of Research, Technology, and Community Service (DRTPM) for funding this research under the Doctoral Dissertation Research (PDD) scheme. We also extend our sincere thanks to Universitas Sumatera Utara for providing invaluable support and resources that greatly contributed to the success of this study.

REFERENCES

[1] Anonymous "2021 Index IEEE Transactions on Knowledge and Data Engineering Vol. 33," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, *(1)*, pp. 1-37, 2022.

[2] P. V. Balachandran, "Data-driven design of B20 alloys with targeted magnetic properties guided by machine learning and density functional theory," *Journal of Materials Research*, vol. 35, *(8)*, pp. 890-897, 2020.

[3] C. C. Aggarwal, "Communication from the Editor-in-Chief: State of the ACM Transactions on Knowledge Discovery from Data," *ACM Transactions on Knowledge Discovery from Data*, vol. 16, *(2)*, pp. 1-2, 2022.

[4] Anonymous "IEEE Transactions on Big Data," *IEEE Software*, vol. 38, *(5)*, pp. 22-22, 2021.

[5] J. Plasse, H. Hoeltgebaum and N. M. Adams, "Correction to: Streaming changepoint detection for transition matrices (Data Mining and Knowledge Discovery, (2021), 10.1007/s10618-021-00747-7)," *Data Mining and Knowledge Discovery*, 2021.

[6] Anonymous "Corrigendum to Spam Profiles Detection on Social Networks Using Computational Intelligence Methods: The Effect of The Lingual Context (Journal of Information Science, (2019), 10.1177/0165551519861599)," *Journal of Information Science*, 2019.

[7] A. N. M. B. Rashid *et al*, "Correction to: Cooperative co-evolution for feature selection in Big Data with random feature grouping (Journal of Big Data, (2020), 7, 1, (107), 10.1186/s40537-020-00381-y)," *Journal of Big Data*, vol. 7, *(1)*, 2020.

[8] D. Zhang, Z. Ye, G. Feng and H. Li, "Intelligent Event-Based Fuzzy Dynamic Positioning Control of Nonlinear Unmanned Marine Vehicles Under DoS Attack," in *IEEE Transactions on Cybernetics*, vol. 52, no. 12, pp. 13486-13499, Dec. 2022, doi: 10.1109/TCYB.2021.3128170.

[9] C. Garibotto, A. Sciarrone, F. Lavagetto, L. Pronzati, A. Baljak and G. Tagliabue, "Performance Analysis of an IoT-Based Personal Vocal Assistant for Cruise Ships Over Satellite Networks," in *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 14857-14866, 15 Aug.15, 2022, doi: 10.1109/JIOT.2021.3116435.

[10] Y. Zhang *et al*, "Research on resource allocation technology in highly trusted environment of edge computing," *Journal of Parallel and Distributed Computing*, vol. 178, pp. 29-42, 2023.

[11] C. Boya-Lara, O. Rivera-Caballero and J. Alfredo Ardila-Rey, "Clustering by communication with local agents for noise and multiple partial Discharges discrimination," *Expert Systems with Applications*, vol. 225, pp. 120067, 2023.

[12] H. Xu, P. Tong and Y. Li, "Correction to: Different treatments of pixels in unlabeled images for semi-supervised sonar image segmentation (International Journal of Machine Learning and Cybernetics, (2024), 15, 2, (637-646), 10.1007/s13042-023-01930-6)," *International Journal of Machine Learning and Cybernetics*, 2024.

[13] Jamei, M., Alkouh, A.B., Karbasi, M. *et al.* Thermo-physical properties estimation of an oil-based hybrid nanofluid: application of a new hybrid neurocomputing approach. *J Therm Anal Calorim* (2024).

[14] V. Bhadra Pratap Singh *et al*, "Hierarchical cluster analysis implementation using the algorithm of clustering using representatives," in 2022, . DOI: 10.1109/GlobConET53749.2022.9872414.

[15] J. Park and M. Choi, "A K-Means Clustering Algorithm to Determine Representative Operational Profiles of a Ship Using AIS Data," *Journal of Marine Science and Engineering*, vol. 10, *(9)*, pp. 1245, 2022.

[16] A. Ghimire, M. Alkurdi and F. Amsaad, "Enhancing hardware trojan security through reference-free clustering using representatives," in 2024, . DOI: 10.1109/VLSID60093.2024.00084.

[17] T. Gu *et al*, "A robust reconstruction method based on local Bayesian estimation combined with CURE clustering," *Information Sciences*, vol. 680, pp. 121132, 2024.

[18] T. He *et al*, "Rolling Bearing Fault Diagnosis Using a Deep Convolutional Autoencoding Network and Improved Gustafson–Kessel Clustering," *Shock and Vibration*, vol. 2020, *(2020)*, pp. 1-17, 2020.

[19] M. Cebecauer *et al*, "Revealing representative day-types in transport networks using traffic data clustering," *Journal of Intelligent Transportation Systems*, vol. 28, *(5)*, pp. 695-718, 2024.

[20] C. Manyfield-Donald, T. A. Kwembe and J. C. Cheng, "A modified clustering using representatives to enhance and optimize tracking and monitoring of maritime traffic in real-time using automatic identification system data," in 2021, . DOI: 10.1109/CSCI54926.2021.00119.

[21] D. Joshi *et al*, "Prediction of sonic log and correlation of lithology by comparing geophysical well log data using machine learning principles," *Geojournal*, vol. 88, *(Suppl 1)*, pp. 47-68, 2023.

[22] B. Rim *et al*, "Semantic cardiac segmentation in chest CT images using K-means clustering and the mathematical morphology method," *Sensors (Basel, Switzerland)*, vol. 21, *(8)*, pp. 2675, 2021.

[23] F. Ros, R. Riad and S. Guillaume, "PDBI: A partitioning Davies-Bouldin index for clustering evaluation," *Neurocomputing (Amsterdam)*, vol. 528, pp. 178-199, 2023.

[24] A. Idrus *et al*, "Distance Analysis Measuring for Clustering using K-Means and Davies Bouldin Index Algorithm," *TEM Journal*, vol. 11, *(4)*, pp. 1871-1876, 2022.

[25] L. Dahlström *et al*, "Identification of representative building archetypes: A novel approach using multi-parameter cluster analysis applied to the Swedish residential building stock," *Energy and Buildings*, vol. 303, pp. 113823, 2024.

[26] D. Kumar *et al*, "A Hybrid Approach to Clustering in Big Data," *IEEE Transactions on Cybernetics*, vol. 46, *(10)*, pp. 2372-2385, 2016.

[27] L. Ma and S. Fan, "CURE-SMOTE algorithm and hybrid algorithm for feature selection and parameter optimization based on random forests," *BMC Bioinformatics*, vol. 18, *(1)*, pp. 169-169, 2017.

[28] D. S. Br Ginting, F. Y. Manik, R. Arrahmi, M. A. A. Saragih, M. D. A. A. Dalimunthe, and M. I. Aldeena, "Performance of Fuzzy Tsukamoto and Fuzzy Sugeno Methods in Predicting Types of Neurotic Disorder," 2023, pp. 194–199.

[29] D. S. Br Ginting, F. Y. Manik, F. N. Nasution, and M. I. Aldeena, "Perceptron neural network model on predicting postpartum depression in the puerperium," in AIP Conf. Proc., vol. 2987, 2024, p. 020038.

[30] D. S. Br Ginting, R. L. Sipahutar, F. Natalida, C. N. Kudadiri, and D. E. R. Purba, "Combination AHP and TOPSIS methods optimizes performance of decision support system for the recipients family hope program in Huta Limbong Padang Sidempuan," in 2021 International Conference on Data Science, Artificial Intelligence, and Business Analytics (DATABIA), 2021, doi: 10.1109/DATABIA53375.2021.9650342.

Distributed Networks for Brain Tumor Classification Through Temporal Learning and Hybrid Attention Segmentation

Sayeedakhanum Pathan[1]*, Savadam Balaji[2]
Research Scholar, Department of CSE, Koneru Lakshmaiah Education Foundation,
Aziznagar, Hyderabad, Telangana, 500075, India[1]
Assistant Professor, Department of CSE – AIML & IoT, VNRVJIET, Hyderabad, Telangana, 500090, India[1]
Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Aziznagar, Hyderabad, Telangana, 500075, India[2]

***Abstract*—Brain Tumor (BT), which is the progress of abnormal cells in brain surface is categorized into different types based on the symptoms and the affected parts in brain. Classification of BT using Magnetic Resonance Imaging (MRI) is an important and challenging task for BT diagnosis. Various approaches are designed to solve the issues and there are so many inconsistencies in detecting the tumor at early stage. The changes in variability and the complexity of size, shape, location and texture of lesions, automatic detection of BT still results a challenging task in the medical research community. Hence, a proposed Hybrid Attention Temporal Difference Learning with Distributed Convolutional Neural Network-Bidirectional Long Short-Term Memory (HATDL-DCNN-BiLSTM) is developed in this research to detect and classify the BT at beginning stage that enables to improve the survival rate of humans. The proposed model uses Gaussian filter for input image enhancement, Hybrid Attention-VNet segmentation to generate region of interest and solves the computational issues through the attention modules by minimizing the dimensions. The proposed model consumed less memory utilization and increase the training speed globally using the distributed learning mechanism. The features extracted using Hybrid Attention based Efficient Statistical Triangular ResNet (HA-ESTER) supports the classification model to increase the training efficiency more accurately. The proposed HATDL-DCNN-BiLSTM attains higher efficiency by the metrics of accuracy, recall, F1-score, and precision of 98.93%, 99.21%, 97.67%, and 96.17% with training data, and accuracy, recall, F1-score, and precision of 96.34%, 96.51%, 96.33%, and 96.15% with k-fold using BraTS 2019 dataset.**

***Keywords*—Brain tumor; magnetic resonance imaging Gaussian filter; hybrid attention-VNet; distributed convolution neural network**

I. INTRODUCTION

Human brain is a vital organ in the physical body because it is responsible for the various governing processes of humans like feeling, memory, responses, vision, motor skills, and breathing [1] [2]. These regulatory functions are significantly disrupted when BT begin to form inside the brain, which arise due to the unfamiliar development of cells in certain brain tissues. BT is primarily categorized as benign or malignant, the benign tumor can't able to diffuse to diverse parts of the brain, so it is considered non-cancerous. However, the malignant tumor is cancerous because it grows uncontrollably and diffuses to various parts of the brain [3]. There are about 200 various types of BT that can arise in diverse areas of the brain. These types of tumors cause more life-varying impact on affected individual's lives [4]. The symptoms of BT appear when the illness is in the stage of advanced and the early phase of BT doesn't reveal any symptoms to the affected person. This phenomenon is due to the position and small size of the tumor in the early phases [5] [6]. Remembrance issues, deviations in the power of eyesight, unfamiliar actions, misperception, seizures, and stability issues are the indications of BT and sometimes it varies depend on the type and location of tumors [7]. After the surgery, the survival rate of BT patients is 14% but, if it is detected in the early stage, the survival rate increases to 70% [6].

Various imaging procedures, like positron emission tomography (PET), MRI and computed tomography (CT), are employed to scan the complete structure of the brain [8]. Compared to PET and CT, MRI is regarded as a better imaging modality and it is broadly used to recognize and categorize the BT because of its better resolution. Moreover, the MRI is highly useful and important in the domain of radiology, because it offers various alterations between a variety of body's soft tissues [9]. Recently, Machine Learning (ML) algorithms have been combined with the automatic BT detection system in various studies to detect BT from the brain MRI [10]. Many efforts have been made to create very effective and trustworthy methods for automatically classifying BT. Handmade features are utilized in the traditional ML techniques, which limited the robustness of the solution and increase the cost. Nevertheless, occasionally supervised learning methods can exceed the unsupervised learning methods, thus resultant in an overfitted approach, that is not fit for additional large repository [11]. Additionally, conventional ML methods also rely on handmade features, that impose drawbacks on the durability and effectiveness of the solution [12]. Though there are various systems available for recognizing irregularities in brain MRI, yet there is a possibility for improving the performance and making the categorization within an appropriate amount of time. The effort in examining and identifying BT using conventional methods become more difficult owing to the expanding size of medical information [13].

Globally, there is no effective method has been found for segmenting and identifying the BT in recent studies irrespective

of its position, structure, and intensity [14] [15]. Many researchers utilized numerous traditional feature extraction approaches, such as Histogram of Oriented Gradients (HOG), bag of word (BoW), local binary patterns (LBP), gray level co-occurrence matrix (GLCM), and density histogram to extract the relevant features. Though, these methods unsourced to extract the exact features, that are needed for accurate BT detection [16] [17]. The approach in study [18] applies k-means clustering approach, whereas feature reduction and extraction processes depend on the Principal Component Analysis (PCA) approach and Discrete Wavelet Transform (DWT) approach respectively. Lastly, SVM is utilized for classifying the BT. However, these approaches consumed more time to complete the procedure of accurate BT detection as well as classification [18]. Some methods apply manually defined tumor regions for detection of BT, which forbids them from being entirely computerized. Deep Learning (DL) is gaining more popular, because of the facility to extract features automatically. Still, DL consumes lot of processing volume and memory [17]. The experimental results of various BT detection frameworks are yet in the initial phase since several characteristics influence the detection method, like poor localization of tumor, deficiency of training volume, poor quality such as the deficiency of training data, poor tumor localization, low-quality images and features [19]. However the combination of deep learning and transfer learning aims to enhance the accuracy and efficiency of brain tumor diagnosis by leveraging the power of pre-trained models to improve classification and segmentation tasks [20].

The main motivation of the research is to design and develop a model for BT classification using proposed HATDL-DCNN-BiLSTM. The Gaussian filter is subjected to MRI for generating the pre-processing image result and with the pre-processed result, segmentation is done by HA-based VNet model to enhance the detection rate, and furthermore, the important and the suitable features are being extracted by applying HA-ESTR model. The features extracted through individual feature extraction mechanism are concatenated to generate a feature map that helps to obtain better performance in image classification. Fine tuning the model provides better generalizability by minimizing the error value through categorical cross entropy measure. The research contribution is briefly discussed as follows:

1) Hybrid Attention-VNet (HA-VNet) based segmentation: The segmentation model is designed through the incorporation of attention models, like HA with the VNet model to extract the color, contrast, texture, and boundary details of image modality, as it helps to increase the detection rate. It highly focuses in extracting key features in both the split and double attention to increase the classification performance.

2) Hybrid Attention-Efficient Statistical Triangular ResNet (HA-ESTR) based feature: The feature extraction process performed using HA-ESTR is to extract useful features to minimize the classification error. This feature extraction model is designed through the integration of different feature extraction models that helps to derive the tumor related features including statistical information. Accordingly, this information is more helpful in finding cancerous and non-cancerous category.

3) Proposed Hybrid Attention Temporal Difference Learning Distributed CNN-BiLSTM (HATDL-DCNN-BiLSTM): The proposed framework is designed through the incorporation of attention models, and temporal difference learning with distributed deep learning framework. Due to the higher learning capability and facility of proposed model, it shows optimal performance in BT classification using imaging modality. The trained images are coordinated with test images to minimize the loss function such that minimal error value returns best performance.

The subsequent sections of this manuscript are delineated as follows: Section II deals with the traditional methodologies employed in the processes of feature extraction, segmentation, optimization, and classification of Brain Tumor. Section III offers an in-depth explanation of the proposed framework. Section IV and V addresses the results and discussion part respectively. Finally, Section VI encapsulates the entirety of the research, highlighting key findings, implications and future work.

II. Literature Survey

Few traditional approaches used for the classification of BT are reviewed in this section. Rahman, T. and Islam, M.S [1] designed a parallel deep convolutional Neural Network (PDCNN) mechanism for classification of BT. Here local as well as global features were extracted in a parallel way to solve the over fitting issues by applying dropout regularizer with batch normalization. The performance was evaluated with three different BT datasets, and reported efficient and accurate performance by extracting low-level as well as high-level features. However, it failed to use the model with 3D structure for identifying the tumors. Ullah, Net al. [17] designed a unified end-to-end model based on deep learning approach (TumorDetNet) for detecting and classifying BT. The distinctive features were effectively learned and minimized the over fitting issues. The SoftMax layer was used to detect the tumor and their grades. It showed higher accuracy measure, but failed to detect complex tumor types. Z. Atha and J. Chaki [3] designed a semi supervised deep learning model for detecting BT. This model was derived through the integration of unsupervised autoencoder mechanism with supervised classification model. It trains the learning parameter of descriptors for better classification. Accordingly, the instances were created through fuzzy logic using augmented data. This model showed higher accuracy, but failed to detect the tumor with location.

Anantharajan, S.et al. [2] modelled a deep learning framework to detect the tumor at early stage. Here, the images were captured and pre-processed by applying adaptive contract enhancement algorithm (ACEA) with median filter. The segmentation was done using fuzzy c-means model and the features were extracted effectively. It resulted higher accuracy rate, but faced time complexity issues. Saha, P. et al. [3] designed a deep learning network with ensemble model for detecting the tumor into different classes. The deep features were captured by CNN and these features are used to classify the BT types. It showed better performance, but failed to evaluate the analysis with large sized dataset. Mathivanan, S.K.

et al. [21] focused a deep and transfer learning approach for detection of BT accurately using MRI. The model was trained with benchmark datasets and increase the performance by applying image enhancement methods. It showed higher accuracy rate and classifies the tumor into different grades.

Amin, J. et al. [9] designed a random forest classifier to classify BT into three different regions, namely non-enhancing, enhancing and complete tumor. The cross-validation schemes were applied in the model to reduce over fitting issues. This model was feasible in generating segmented results without manual interactions. Asiri, A.A.et al. [4] introduced a dual module mechanism to improve the accuracy and speed of BT detection. The first module was said as image enhancement approach that used a machine learning strategy for normalizing the images and solve the issues, like low contrast and noise. Accordingly, second module applies support vector machine (SVM) to perform segmentation as well as classification of BT. It increased the robustness and generalizability of model.

A. Challenges

The challenges discovered from the existing BT detection approaches are labelled below.

- In study [6], the suggested approach offered faster processing time and improved accuracy due to the incorporation of an efficient algorithm. However, the suggested method was inefficient in dealing with imbalanced and large-scale datasets.
- The suggested HBTC framework [22] reduced the complexity of inherent. However, it failed to generalize over different datasets and revealed low exactness and toughness.
- The suggested method [23] was not examined with openly available imaging datasets having irregularity features and on the other hand, the suggested method also had the issue of scalability.
- In study [12], the suggested method reduced feature irrelevance, duplication, and dimensionality to identify the important features, however, the feature selection and extraction process consumed more time because of the high computational complexity in managing huge amounts of data.
- The suggested method [24] effectively classifies the brain tumors, however, it failed to do the cropping and rotating process to the data before it was subjected to the classification process.

B. Problem Statement

In the recent years, BT has grown uncontrollably in different locations across human body, but it mainly causes in brain. When the tumor grows, it starts to increase the pressure inside the brain, which typically affects the brain and cause brain damage and also it threatens human life. Various BT detection methods introduced in the medical research helps the physicians to detect the cancer using MRI, but it undergoes several challenges. The medical professionals and radiologists examine number of MRI slices that results labor intensive and time-consuming process. However, this manual scheme causes human error and potential leads to delay or misdiagnosis. Hence, it is an urgent requirement in the medical research to design an accurate BT detection method for accurate diagnosis of BT at earlier stage. To accomplish this task, an input MRI is fetched and the image is captured from BraTS dataset. Assume the dataset as T with s number of MRI brain images as,

$$T = \{C_1, C_2, ..., C_i, ..., C_s\} \quad (1)$$

The input image is represented as C, which is subjected to the pre-processing stage to increase the quality of image by eliminating the noise. Accordingly, pre-processed result is represented as C, which is fetched by applying Gaussian filter to the input image. All the input images do have same dimension and so it is required to pre-process the image to provide uniformity in the training process. The image C is fed to the segmentation phase, which simplify and change the image representation into meaningful factor, such that segmented image result is specified as, H

$$H = \{C\} \quad (2)$$

Segmented image result is fed to feature extraction phase to extract useful and essential features and based on the features, a feature vector is generated and it is noted as, A, respectively.

Accordingly, the generated feature map is applied to the model, where the tumor grades are classified into various stages, like no tumor, enhancing tumor, non-enhancing tumor, and peritumoral edema and this is accomplished through reduction of error value by applying categorical cross entropy loss function that is specified as,

$$L = -\sum M \log(M) \quad (3)$$

Here, M denotes predicted value, and M is the actual ground truth value.

$$M = t = \begin{cases} 0; \text{ if "C " is normal} \\ 1; \text{ else if "C "is non enhancing tumor} \\ 2; \text{ else if "C "is peritumoral edema} \\ 3; \text{ otherwise "C "is enhancing tumor} \end{cases} \quad (4)$$

Hence, the proposed BT detection model generate accurate results by minimizing the error rate through the loss function.

III. METHODOLOGY

The HATDL-DCNN-BiLSTM is developed in this research for classification of BT with MRI modality. Input image is acquired from the BraTS dataset and subject to pre-processing phase, in which Gaussian filter is applied for enhancing image quality. The pre-processed image is subjected to segmentation phase, where HA based VNet segmentation framework is applied to generate region of interest. Based on the segmented image result, the features connected with the modality is extracted using HA-ESTR that includes different feature extraction models. After extracting the features, BT classification is performed using proposed HATDL-DCNN-BiLSTM. Fig. 1 represents the schematic view of proposed BT classification framework.

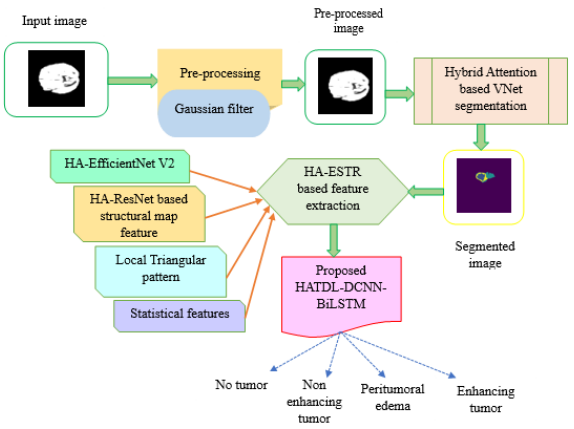


Fig. 1. Schematic representation of proposed HATDL with DCNN-BiLSTM for BT classification.

A. *Gaussian Filter-Based Image Pre-processing*

MRI plays an active role in human brain scanning, as MRI provides detailed information about brain soft tissue structure. Due to the better efficiency in detecting the soft tissues of brain, MRI contributes more significant in BT classification. Consider the dataset as T composed with S number of MRI, and each image used for the processing is specified as,

$$T = \{C\}; k \in \{1, 2, ..., s\} \tag{5}$$

Here, T is the dataset, C_k denotes k^{th} input image with the dimension of $[224 \times 224]$, and s implies total number of images. The brain images are sensitive towards unwanted noise and distortions and hence, pre-processing is needed to remove the noise for increasing greatest quality of image. Gaussian filter [25] is applied to pre-process the input image C for smoothen image by reducing the noise. Gaussian filter is an essential tool in computer vision and image processing system, and it is considered as an optimal filter to solve imaging issues. It is referred as the linear smoothing filter that selects the weights with respect to shape of Gaussian function. Gaussian filter is a category of low pass filter specifically designed to eliminate the noise subject to normal distribution, and hence it is commonly used in the image processing system. By applying Gaussian filter, the noises are suppressed and smoothed out that further increases the image quality. The Gaussian filter for the image C with pixel values is expressed as,

$$F(u, v) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\left(x + y\right) / 2\sigma\right) \tag{6}$$

$$C = F(C) \tag{7}$$

Here, σ denotes the variance of Gaussian filter, x and y are the filter kernel, and F denotes Gaussian filter. The resultant pre-processed image after removing the noise is expressed as, C with the dimension of $[224 \times 224]$, respectively.

B. *Hybrid Attention Based VNet Segmentation*

Segmentation process is applied to the image processing task for changing and simplifying image representation into meaningful form by selecting similar pixels. The pre-processed result C is applied to segmentation phase, where HA-VNet model is utilized to generate segmentation result. The HA-VNet is designed through the integration of VNet with HA, in which HA is modelled by integrating split attention and double attention model. The major advantage of using HA-VNet is that it allows seamless image segmentation with higher performance and accuracy. The benefit of using attention model in the segmentation is to minimize the spatial and channel dimension of VNet model. The split attention is the computational unit composed of split attention operations and feature map group. On the other hand, the double attention computes pooled features and capture the complex appearances and more efficient in correlating features with each specific location. Accordingly, these benefits are incorporated with the VNet model to provide accurate segmentation results that helps to increase the training process. The structure of VNet [26] [27] contains two different paths, namely compression path (left side) and decompression path (right side). The convolution with suitable padding is performed for exploiting the features from input and reduce the resolution with suitable stride at the end of every stage. The compression path is divided into different phases that operates at varying resolutions such that each stage contains one to three convolutional layers. The VNet takes C as input, and the input taken by each phase is processed by a residual function used in convolutional layers, and finally output is re through final convolution layer. The HA mechanism is added at the convolutional layer-2 to increase the training efficiency. The HA layer into the VNet takes previous convolution layer output as input of size $[N \times 6 \times 6 \times 16]$ and resulted output is passed as input to the next convolution layer. Accordingly, the benefit of using VNet architecture is that it offers better convergence than other non-residual learning network, and also it offers $[5 \times 5]$ size of convolutional layer at each stage. Accordingly, the convolution layer is represented as,

$$X = \sum\left\{\sum\sum C \,.E(\alpha, \beta)\right\} + P \tag{8}$$

where, X denotes the output of convolution operation, P denotes bias factor, C is the input, $E(\alpha, \beta)$ is the weight among the locations (α, β). The convolutional operations are used to double the feature maps due to the involvement of residual framework and number of increasing feature channels considered in compression path. The convolutional layer aims to minimize the memory required for training procedure and process the input at higher resolution. The increase in the segmentation rate is achieved by reducing the error rate of model, in which the error value is computed through dice loss function that is expressed as,

$$L = \frac{1}{R}\sum\left[1 - \frac{2\sum M \,.M}{\sum M + \sum M}\right] \tag{9}$$

Here, R denotes number of classes, r specifies number of class samples, M implies predicted value, and M shows actual

ground truth value. The activation function contains various functions, like rectified function, tanh, and so on.

$$X = f\left(\sum X.E + P\right) \tag{10}$$

Here, X is the input of activation function, E and P shows weight and bias vector. Finally, the rectified linear unit (ReLU) is expressed as,

$$H = \begin{cases} X & ; X \geq 0 \\ \frac{X}{X}; & X \leq 0 \end{cases} \tag{11}$$

The segmented image result obtained through HA-VNet is represented as H with the size of $[N \times 224 \times 224]$.

- Architecture of Hybrid Attention

The HA [28] is designed through the integration of double attention and split attention mechanism that enable the model to increase the training speed by consuming less memory during training process. Both the attention models are operated by taking the input with dimension of $[N \times 6 \times 6 \times 16]$. Fig. 2 shows the structure of Hybrid Attention model. For any input value, the HA compute the attention map separately through split and double attention and the dimension is adjusted using reshape layer by applying sigmoid function. Assume the input taken by the HA is D with size $[N \times 6 \times 6 \times 16]$ and it is divided into two parts, but the channel number of these parts must be equal. Accordingly, these two parts extracts the boundary information through split attention and double attention. Structure of split attention contains different layers, namely reshape, pooling, dense, and softmax. With split attention, the attention process is enabled across the feature maps. Moreover, the original feature dimension is reduced with reshape layer and the result generated with the respace layer have the size of $[N \times 6 \times 6 \times 16]$. On the other hand, the double attention model contains different layers, like convolution, reshape, softmax, and matrix multiplication. Here, the convolution layer performs the convolution operation, whereas reshape layer enable to reduce the dimension of feature maps into $[N \times 6 \times 6 \times 12]$. The output of two different layers is fused by the matrix multiplication layer and finally reshape layer is used to generate the output. Finally, the output of split attention d and the output of double attention d_2 are fused together and the resultant is specified as,

$$d = \{d + d\}; d \in [N \times 6 \times 6 \times 16] \tag{12}$$

Here, the output of HA mechanism is specified as, d having the dimension of $[N \times 6 \times 6 \times 16]$.

C. Hybrid Attention Based Efficient Statistical Triangular ResNet Feature Extraction

In this work HA-ESTR feature extraction method is proposed which is formed through the integration of different feature extraction models, like Hybrid Attention based EfficientNet V2 feature (HA based EfficientNet V2), Hybrid Attention based ResNet structural map (HA based ResNet structural map), Local Triangular pattern (LTrP), and statistical features. Accordingly, the Hybrid Attention based EfficientNet V2 is modeled through the integration of Hybrid attention with standard EfficientNet V2 feature, and also Hybrid Attention based ResNet structural map is formed through the integration of Hybrid attention and structural map with ResNet feature. The HA-ESTR feature extraction model takes the segmentation output (H) as input and extract useful features that helps to boost the classification accuracy. The detailed explanation of each feature extraction models is clearly explained as follows.

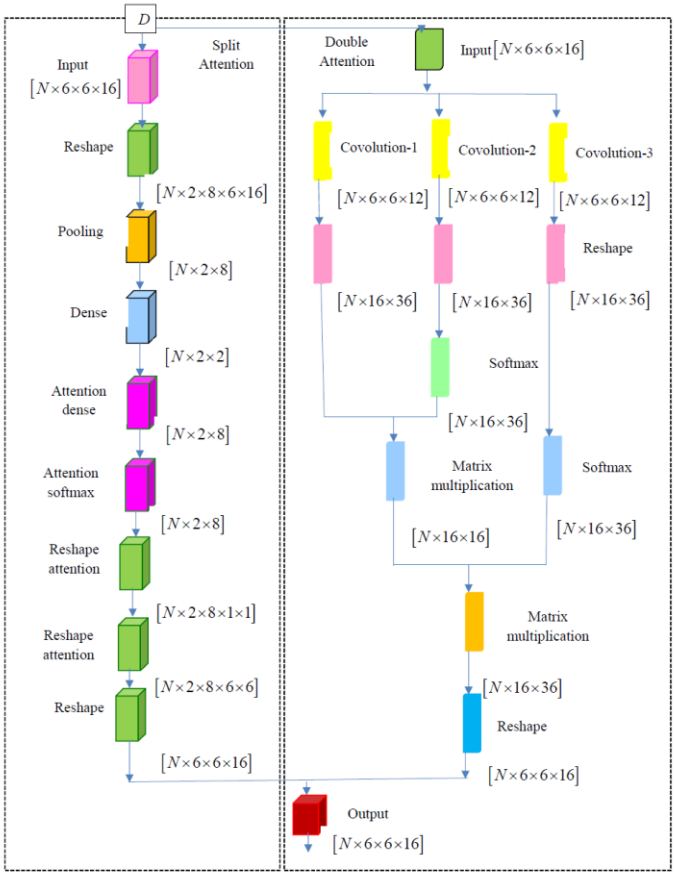


Fig. 2. Architecture of hybrid attention model.

1) Hybrid attention based EfficientNet V2 feature: In this feature is formed by incorporating the attention model features, like HA into the standard EfficientNetV2 feature. The benefit of using this feature model is that it effectively alters the regularization value with respect to image size due to the progressive learning mechanism. It effectively balances the network width, depth, and resolution and helps to increase the performance. The EfficientNetV2 model provides notable enhancement in the training efficiency and offers faster convergence [29] [30]. The EfficientNet V2 is composed with different layers, such as convolution, MBConv, HA, pooling, and fully connected layer. Initially, the input (H) with dimension $[N \times 224 \times 224]$ is passed to the convolution layer with stride $[3 \times 3]$, and the output is fed to MBConv layer,

which contains an inverted residual connection followed with the separable convolution. To reduce the overfitting issues, the dropout is used in MBConv block [30]. The HA layer is integrated into the model and the output of HA mechanism is passed to the convolution with stride $[1\times1]$ layer and finally, the pooling and the fully connected layer is utilized to generate the output feature as A with the size of $[1\times112\times112\times32]$ such that its dimension is resized into $[1\times32\times32]$ for further processing.

2) Hybrid attention based ResNet structural map feature: The HA based ResNet structural map feature is designed by integrating the attention mechanism, such as HA into ResNet model with structural map features. The structure of HA model contains different layers, convolution, batch normalization, Maxpooling, flatten, fully connected layer, and LTP. The input H with size $[N\times224\times224]$ is passed to the convolution layer, where the HA layer is applied, and followed by flatten and fully connected layer are utilized to reduce the dimension of original features. Each layer contains same number of filters and if the size of feature map halved, then filters used in the model get doubles [31]. The LTP is applied to the texture analysis of the grey scale images and this feature is more superior in computational efficiency and description performance. LTP refers to a descriptor, which describes relationship among a selected pixels in the image with the neighborhood pixels [32]. This feature operates by setting a gray value of certain pixel as G and the gray value of pixels in the $[3\times3]$ neighborhood as $G\left(j=0,1,...,7\right)$. It analyzes the relationship between G and G, if $G<G$, the value is set to 0, otherwise the value is set to 1. By doing this process, an 8-bit numbers with 0's and 1's are generated such that the decimal value corresponds to the binary number is referred as the TLP feature value and it is expressed as,

$$A=\sum\chi(G-G)2\ ;\ \chi(a)=\begin{cases}1 & ;a\geq0\\ 0 & ;Otherwise\end{cases} \tag{13}$$

Here, A denotes the structural map features. In general, the HA layer is integrated into the ResNet-50 model, and the output is fed to the LTP, which generates the structural map features as, A with the dimension of $[1\times112\times112\times32]$ and for easier processing, the original dimension is resized into $[1\times32\times32]$, which is used for further processing. The purpose of extracting this feature is to enhance classification accuracy by modifying the residual block with Rectified Linear Unit (ReLU) activation function. Due to the great representation ability, this feature extraction model is applied in BT classification.

3) Local triangular pattern feature: The LTrP is designed using 8-bit binary code and the steps involved in the extraction of LTrP feature is briefly explained as follows. At first, a $[3\times3]$ image matrix is generated for each image and then three neighbour pixels are selected with 600 triangle formation, and these triangles are formed in four different directions [33]. At each direction, the center pixel is referred as the threshold for the remaining neighboring pixels. A binary value of '1' is generated when the threshold value is higher than any of two neighboring pixels, and otherwise the binary value is set to '0'. Thus, a binary value is created for all the three pixels using the below equation as,

$$\begin{aligned}Nw&=BY\left(\max[w-N],[w-N],[w-N]\right)\\ Nw&=BY\left(\max[w-N],[w-N],[w-N]\right)\\ Nw&=BY\left(\max[w-N],[w-N],[w-N]\right)\\ Nw&=BY\left(\max[w-N],[w-N],[w-N]\right)\end{aligned} \tag{14}$$

After finding the values at each direction, the triangle is flipped to 1800and repeat the same process to find the binary code value. Here, each triangle contains four values that includes $w,N,w,$ and w. Accordingly, w is the center pixel, and hence set it as the threshold for $w_6,N,w_2,$ .If w is higher than any other two neighboring pixel values, that is w_6 or N or w_2, then binary value is set to '1', otherwise set to '0'. The process involved in generating the above values are expressed as,

$$\begin{aligned}Nw&=BY\left(\max[w-w],[N-w],[w-w]\right)\\ Nw&=BY\left(\max[w-w],[N-w],[w-w]\right)\\ Nw&=BY\left(\max[w-w],[N-w],[w-w]\right)\\ Nw&=BY\left(\max[w-w],[N-w],[w-w]\right)\end{aligned} \tag{15}$$

Finally, the eight coded patterns are generated through the above process using the below equation as,

$$Nw_h=\sum_{h=0}^{t}BY\left(w_h-N\right)\geq\left(t+1\right)/2 \tag{16}$$

$$A=LTP\left(\alpha,\beta\right)=\sum\gamma\left(Nw-N\right)2 \tag{17}$$

where,

$$\gamma\left(\mathrm{X}\right)=\begin{cases}1,\ \mathrm{X}\geq0\\ 0,Otherwise\end{cases} \tag{18}$$

Here, BY shows binary value, LTP represents the LTrP features, which is represented as A with the dimension of $[1\times32\times32]$.

4) Statistical features: Some of the statistical features extracted from the segmented image results are mean, standard deviation, variance, median, harmonic mean, geometric mean, and entropy [34] [35].

Mean: homogeneity of brightness of MRI, which is represented as

$$S=\frac{1}{Z}\sum H\left(z\right) \tag{19}$$

Here, Z is total number of pixels, S refers to mean feature with the dimension of $[1\times1]$, and $H\left(z\right)$ is the segmented image.

Variance: heterogeneity that strongly connected with standard deviations and it is expressed as,

$$S = \sum\sum (m - S)\, H(m,n) \tag{20}$$

where, S refers to the variance feature with the size of $[1\times1]$.

Standard deviation: shows the difference exists in each observation from the mean value and it is represented as,

$$S = \sqrt{\frac{\sum (H(z) - S)}{Z}} \tag{21}$$

Here, S implies standard deviation feature with size of $[1\times1]$.

Median: It shows the central tendency of the features and it is generally represented as, S with the size of $[1\times1]$.

Harmonic mean: It extracts the average sample size from the number of groups and it is expressed as,

$$S = \frac{z}{\frac{1}{\lambda} + \ldots \frac{1}{\lambda}} \tag{22}$$

Here, z implies total number of pixels, λ and λ_z are the pixel values, and S is the harmonic mean features having the size of $[1\times1]$.

Geometric mean: refers to the appropriate measure, where the value changes exponentially. It is a category of mean with central tendency and it is computed as,

$$S = \sqrt{\lambda, \ldots, \lambda} \tag{23}$$

Here, Z refers to the count of pixel values, and S is the geometric mean feature having the size of $[1\times1]$.

Entropy:It is used to calculate dissimilarity in MRI and its value will be very high for better performance. It is computed as,

$$S = \sum\sum H(m,n) \log H(m,n) \tag{24}$$

Accordingly, the entropy feature is denoted as, S with the size of $[1\times1]$. Finally, the statistical features generated from the segmented image results are expressed as,

$$A = \{S, \ldots S\} \tag{25}$$

where, A shows the statistical features with the size of $[1\times7]$ and this dimension is resized into $[1\times32\times32]$ for smoothing the training process. The feature map generated using HA-ESTR are represented as,

$$A = \{A_1 \| A_2 \| A_3 \| A_4 \|\} \tag{26}$$

Here, A is the feature map of HA-ESTR features with the size of $[1\times32\times32\times4]$, respectively.

D. HATDL-DCNN-BiLSTM Classification

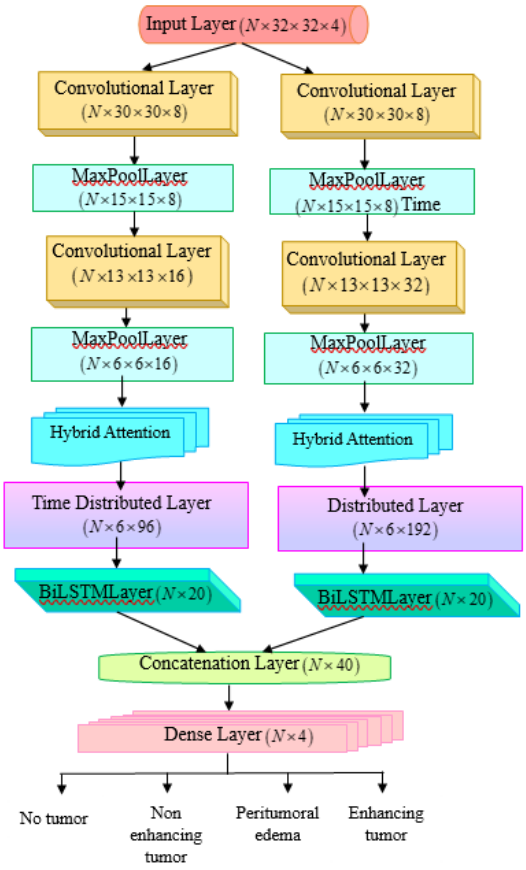


Fig. 3. Architecture of proposed HATDL-DCNN-BiLSTM.

The above Fig. 3 depicts the proposed framework, is modelled by applying the attention models, like split attention and double attention with Temporal Difference Learning (TDL) and CNN-BiLSTM using Distributed training [36] [37]. The benefit of applying this model for classification is due to its efficiency in higher training process. The integration of TDL helps to detect the output more accurately based on the working principle of unsupervised learning model. The distributed model provides completely effective solution with limited number of epochs by allocating the resources in a distributed manner. Hence, it reduces communication overhead and solve the communication delay caused due to the improper allocation of resources. The distributed learning framework partition the dataset and distribute it among the machines, and each data uses same weight factor but trains under various batches and finally, the results are averaged to get global gradient and it also enable to achieve faster training. The proposed HATDL with DCNN-BiLSTM model is made up of various layers, like convolution,

maxpooling, time distributed, Bidirectional, and dense layer [37]. The input layer gets the input by the dimension of $[N \times 32 \times 32 \times 4]$ and fed to the convolution layer. The model considers two convolution layers that contains the filter size of 64 and 128, and followed by the convolution layer, maxpooling layers are utilized to minimize the dimension size. The aim of convolution and the pooling layer is to filter the incoming information for extracting important features, in which the convolution layer performs convolution operation between input features and the smaller matrices referred as filters or kernels.

Let us assume the input matrix as x, where $x \to (A)$, y refers a kernel matrix, and O be the result matrix with rows and columns as p and q.

$$O[p,q] = x.y[p,q] = \sum\sum y[a,c].x[p-a,q-c] \tag{27}$$

Convolution layer uses ReLU activation function, which is commonly used activation function in the CNN model. The major benefit of using this activation function is that it does not require to activate each neuron at same time, as it converts all the negative value into '0', and due to this reason, ReLU offers more computational efficiency. The ReLU is expressed as,

$$F(J) = J = \max(0, J) \tag{28}$$

The pooling is the sub-sampling model aims to minimize the size of convolution matrix and enable to enhance the robustness of framework. After the maxpooling layer performs the operations, HA layer is applied and it generates the features with the dimension of $[N \times 6 \times 6 \times 16]$. Accordingly, the flatten or time distributed layer is utilized to normalize the features from the dimension $[N \times 6 \times 6 \times 16]$ into $[N \times 9 \times 96]$, and also the BiLSTM layer reduce the feature size into $[N \times 20]$. Each unit in the BiLSTM network contains a memory cell as well as three gates, namely forget, input and output gate and they are used to maintain the flow of information. The input passed to the input gate is expressed as,

$$K = \sigma(I.[g, Y] + Q) \tag{29}$$

Here, σ denotes activation function, Y shows the current input that is obtained from the output of flatten layer, I shows weight matrix, Q represents bias vector, and g represents previous hidden state. The sigmoid output is represented as,

$$w = \sigma(I.[g, Y] + Q) \tag{30}$$

where, w indicates sigmoid output.

$$\tilde{W} = \tanh(I.[g, Y] + Q) \tag{31}$$

Here, $\tilde{W}$ shows sigmoid output. The output of new cells state is specified as,

$$W = K.W + w.\tilde{W} \tag{32}$$

where, W shows new cells state and the output obtained through the output gate is expressed as,

$$N = \sigma(I[g, Y] + Q) \tag{33}$$

$$g = N.\tanh(W) \tag{34}$$

Here, σ shows sigmoid function, Y is the input at time c, g indicates hidden state at time c, tanh refers hyperbolic tangent function, I and Q are the weight and bias vector. The output receives from two BiLSTM layers are fused together by concatenating the features with the size of $[N \times 40]$, which undergoes to dense layer that creates classification result having the dimension of $[N \times 4]$ through sigmoid activation function. Accordingly, the accurate classification is accomplished by minimizing the error rate through loss function. Here, loss function considered to reduce the error value is accomplished using categorical cross entropy measure that is expressed in Eq. (3). Hence, the model generates the output of four different grades, as normal case, non-enhancing tumor, peritumoral edema, and indicates enhancing tumor.

IV. RESULTS

This section illustrates results of proposed HATDL-DCNN-BiLSTM by varying two different datasets using the performance metrics.

A. Experimental Setup

The proposed HATDL with DCNN-BiLSTM model is simulated in the PYTHON tool with windows 10 OS, intel core processor, and 16 GB RAM.

B. Dataset Description

This experimentation is done using BraTS 2018 dataset [38], and BraTS 2019 dataset [39]. BraTS 2018 dataset [38] contains multimodal scans available at four different modalities, namely T1, post contrast T1-weighted (T1Gd), T2-weighted (T2), and T2 fluid Attenuated Inversion Recovery (T2-FLAIR). The BraTS 2019 dataset [39] contains multimodal scans and they are available at four different modalities as similar that of above dataset. Also, both the dataset contains four different labels, namely grade-0 as no tumor, grade-1 as non-enhancing tumor, grade-2 as peritumoral edema, and grade-3 as enhancing tumor.

C. Evaluation Metrics

The effectiveness of the proposed framework is measured with metrics, namely accuracy, F1-score, precision, and recall.

Accuracy: It shows the percentage of accurately predicted samples to the total number of samples predicted.

$$A = \frac{G + G}{G + G + H + H} \tag{35}$$

Here, A is the accuracy, G and G are the true positive and the true negative, whereas H and H are the false positive and the false negative.

Precision: It shows the number of true positives towards total number of true and false positives.

$$P = \frac{G}{G + H} \tag{36}$$

Recall: It refers to the percentage of total number of relevant instances found in the data sample.

$$R = \frac{G}{G + H} \tag{37}$$

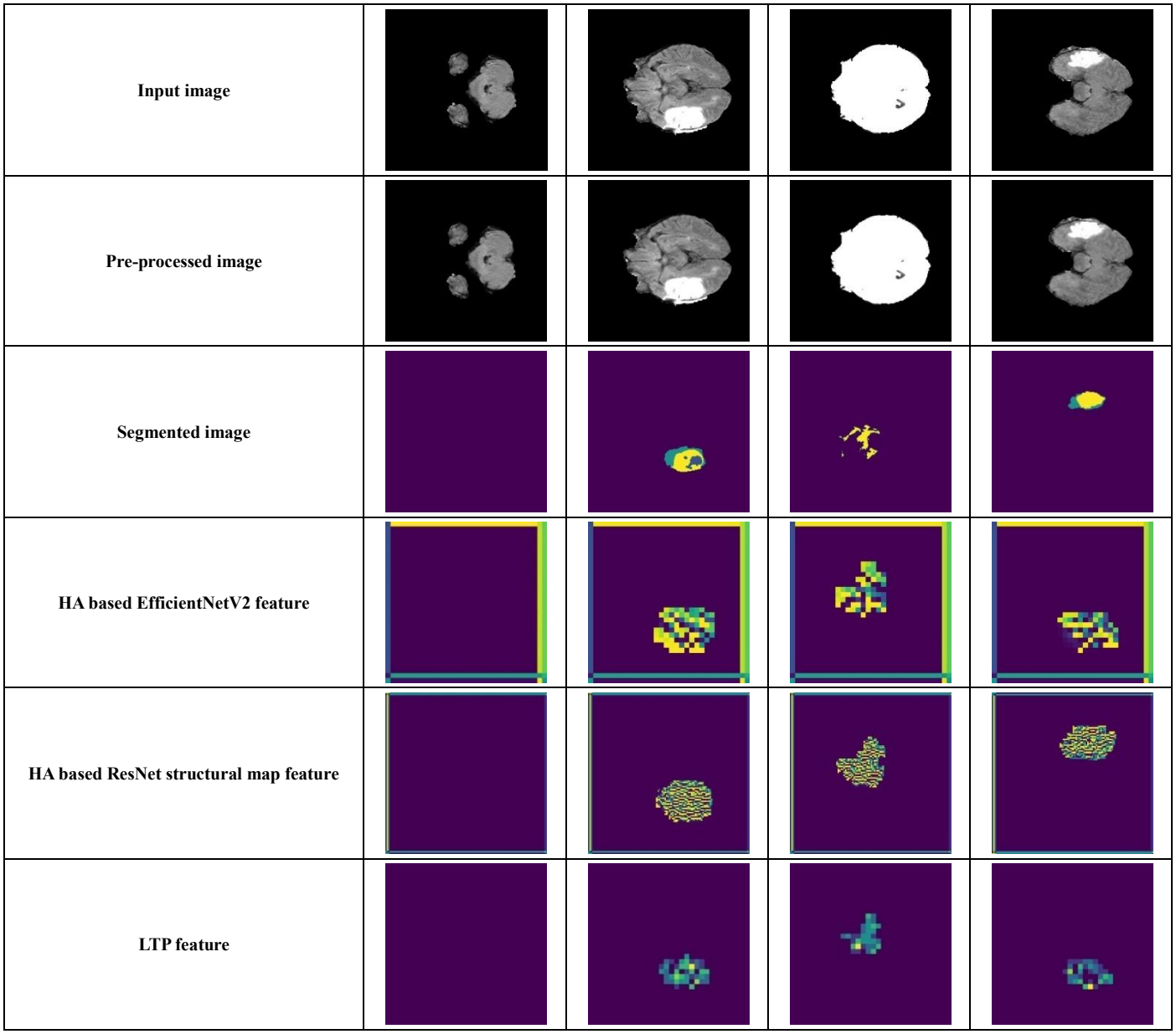
F1-score: It is the weighted average value of precision and recall measure.

$$F = 2 * \frac{P \times R}{P + R} \tag{38}$$

Here, P is the precision, R shows recall, and F is the F1-score.

D. Sample Image Results

This section presents the image results collected in BT classification with two different datasets. Fig. 4 shows the sample image results collected using BraTS 2018 dataset. The sample images of the proposed methods that undergoes each phase are captured and are briefly explained in the below figure. The imaging modality belongs to each class, like normal, enhancing tumor, non-enhancing tumor and peritumoral edema are captured and shown below.



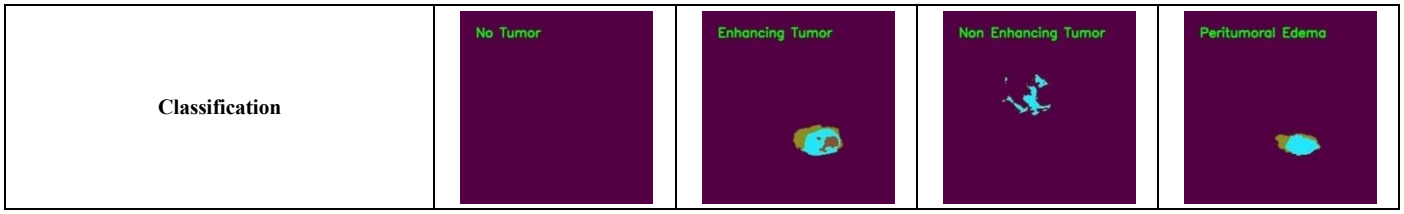


Fig. 4. Sample image results observed by BraTs 2018 dataset.

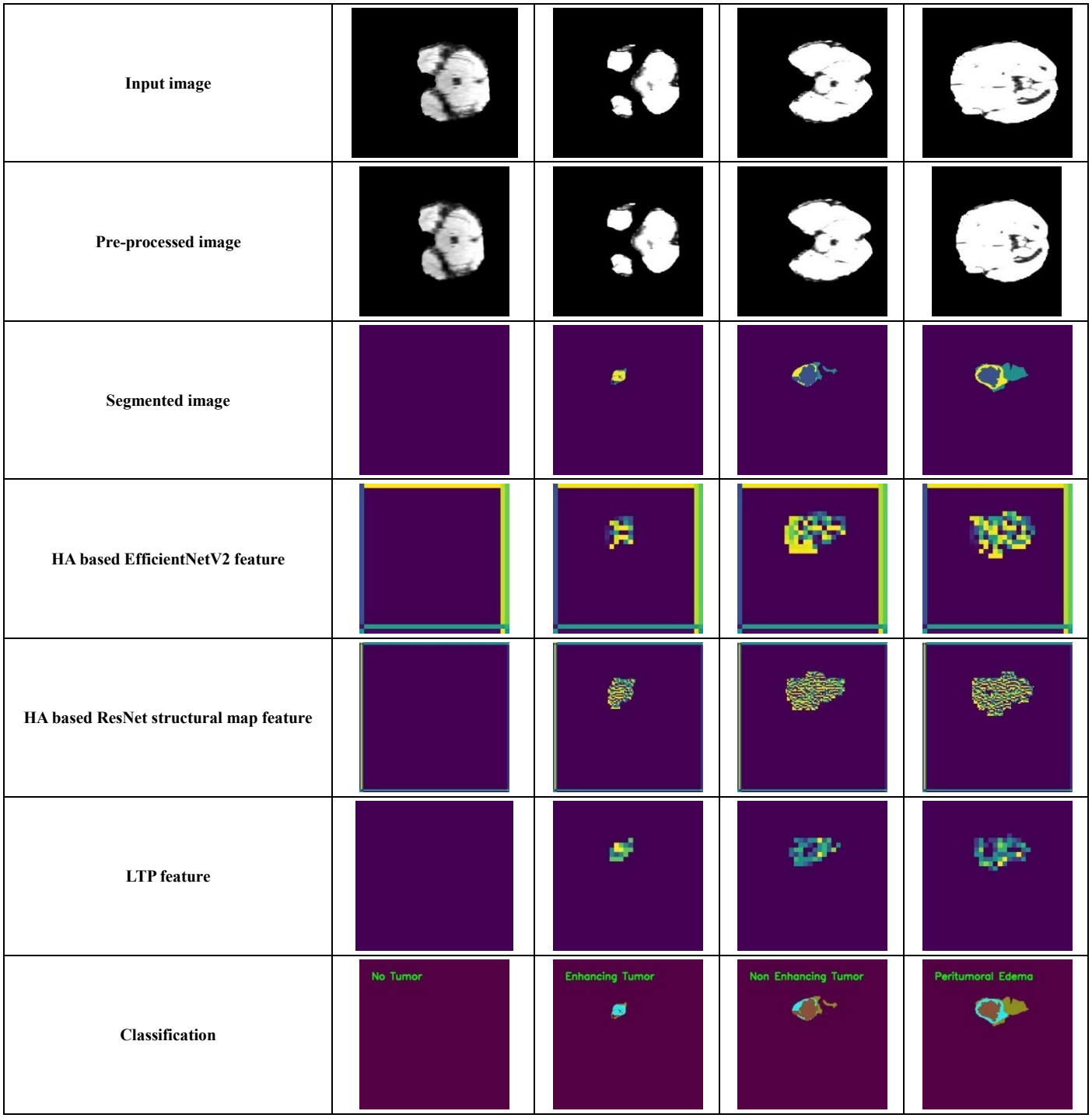


Fig. 5. Sample image results obtained by BraTs 2019 dataset.

The Fig. 5 depicts the sample image results captured by BraTS 2019 dataset. For each grade of BT classification, the images are collected using proposed HATDL with DCNN-BiLSTM, and the collected samples images are clearly represented in the below figure with each classification stage.

E. Performance Analysis

This section explains the performance analysis of HATDL-DCNN-BiLSTM by varying the training percentage of k-fold with two different datasets.

1) Analysis based on BraTS 2018 dataset: Fig. 6 depicts the analysis done with BraTS 2018 dataset with training data. Fig. 6 (a) shows accuracy measure. With 80% of training data, the accuracy observed by the HATDL-DCNN-BiLSTM by varying the epoch from 20, 40, 60, 80, and 100 is 87.12%, 88.72%, 91.36%, 4.75%, 97.55%. The analysis made by F1-score is depicted in Fig. 6 (b). For 80% training data, the F1-score computed by HATDL-DCNN-BiLSTM for epoch 20 is 87.11%, epoch 40 is 88.71%, epoch 60 is 91.35%, epoch 80 is 94.73%, and epoch 100 is 97.54%. The performance analysis observed using precision is illustrated in Fig. 6 (c). The precision measured by HATDL-DCNN-BiLSTM at 80% training data for epoch 20 is 87.01%, epoch 40 is 89.11%, epoch 60 is 91.21%, epoch 80 is 95.77%, and epoch 100 is 96.70%. Fig. 6 (d) depicts analysis of recall measure. The recall observed by the proposed HATDL-DCNN-BiLSTM for 80% training data by varying epoch from 20 to 100 is 87.21%, 88.32%, 91.49%, 93.72%, and 98.38%, respectively. The epoch 20 refers that 20 times, the model is trained with 20 iterations, and epoch 40 means that 40 times the model is trained with 40 iterations, and is similar for remaining epochs. Hence, increasing the value of epoch, the performance of HATDL-DCNN-BiLSTM increases effectively due to the increasing volume of trained data. If training data is 60% used, the remaining data are used for testing process and so the performance degrades. The training data of 80% refers that 80% data are allowed in the training procedure and only remaining 20% is considered for testing, and this could allow the model to improve the performance. Hence, increasing the training percentage increases the model performance.

Fig. 7 depicts the analysis done with BraTS 2018 dataset by varying k-fold. Fig. 7 (a) illustrates accuracy measure. Accuracy obtained at k-fold of 10 by the proposed HATDL-DCNN-BiLSTM with epoch 20 is 88.96%, epoch 40 is 91.18%, epoch 60 is 94.18%, epoch 80 is 95.74%, and epoch 100 is 96.62%. The analysis described by F1-score is shown in Fig. 7 (b). The F1-score computed by HATDL-DCNN-BiLSTM at 10 k-fold value by varying the epoch from 20 to 100 is 88.95%, 91.62%, 93.92%, 95.72%, and 96.75%. The performance analysis made using precision is illustrated in Fig. 7 (c). The precision obtained by HATDL-DCNN-BiLSTM with k-fold 10 for epoch 20, 40, 60, 80, and 100 is 88.81%, 90.47%, 93.47%, 95.47%, and 95.91%. Fig. 7 (d) illustrates analysis of recall measure. The recall observed by HATDL-DCNN-BiLSTM with k-fold of 10 for epoch 20 is 89.09%, 40 is 92.79%, 60 is 94.36%, 80 is 95.98%, and 100 is 97.61%, respectively. For K-fold of 8, the entire dataset is partitioned into eight sets and the model is trained for eight times to analyze the performance. The dataset is divided into different sets with respect to k-fold and model is trained based on the value of k-fold such that it simulates the performance to increase the efficiency. As, increasing the k-fold systematically increases the number of times that the model could be trained, which boost the performance of proposed HATDL-DCNN-BiLSTM.

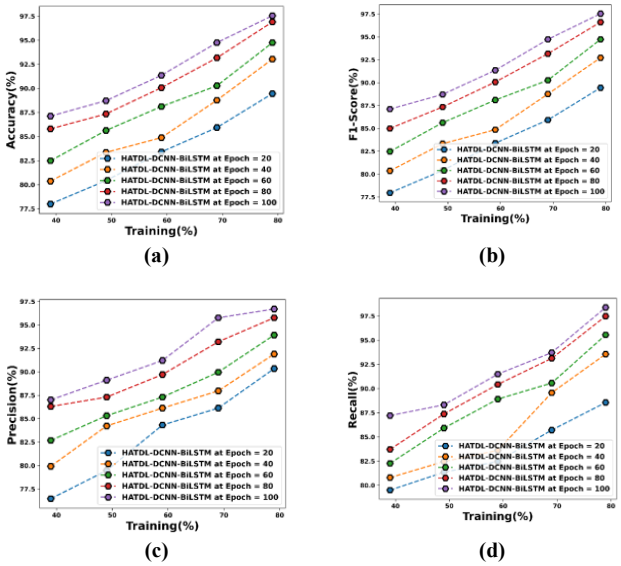


Fig. 6. Performance analysis-BraTS 2018 dataset by varying training percentage, a) accuracy, b) F1-score, c) precision, d) recall.

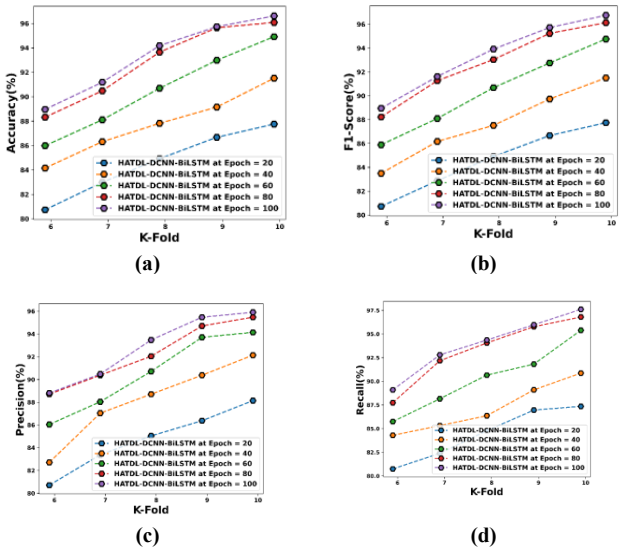


Fig. 7. Performance analysis with BraTS 2018 dataset by varying k-fold, a) accuracy, b) F1-score, c) precision, d) recall.

2) Analysis based on BraTS 2019 dataset: Fig. 8 illustrates the performance analysis carried out with BraTS 2019 dataset varying by training data. Fig. 8 (a) illustrates accuracy measure. Accuracy obtained by HATDL-DCNN-BiLSTM for 80% training data by varying the epoch 20 to 100 is 86.38%, 89.44%, 91.88%, 94.14%, and 98.93%. Fig. 8 (b) depicts F1-score

metric. The F1-score obtained by proposed HATDL-DCNN-BiLSTM with 80% training data for epoch 20 is 86.36%, epoch 40 is 89.43%, epoch 60 is 91.87%, epoch 80 is 94.13%, and epoch 100 is 97.67%. The precision used to analyze the performance is illustrated in Fig. 8 (c). With 80% training data, the precision of HATDL-DCNN-BiLSTM by varying the epoch from 20 to 100 is 85.51%, 89.04%, 91.89%, 93.53%, and 96.17%. Fig. 8 (d) illustrates the recall measure. The recall of proposed HATDL-DCNN-BiLSTM at 80% training data for epoch 20, 40, 60, 80, and 100 is 87.24%, 89.82%, 91.84%, 94.74%, and 99.21%.

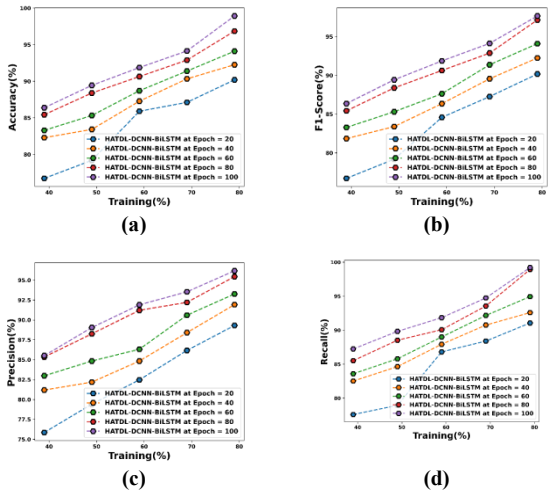


Fig. 8. Performance analysis with BraTS 2019 dataset by varying training data, a) accuracy, b) F1-score, c) precision, d) recall.

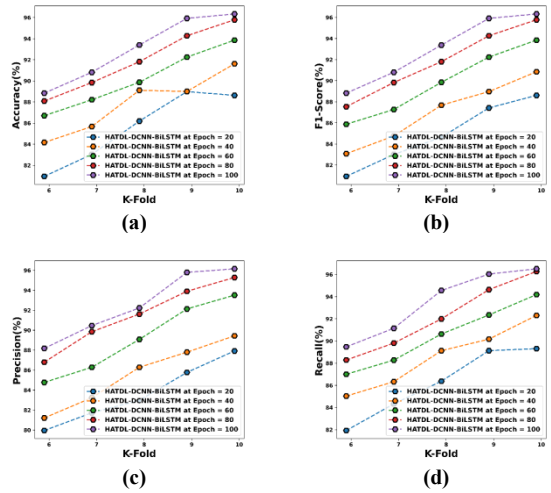


Fig. 9. Performance analysis with BraTS 2019 dataset by varying k-fold, a) accuracy, b) F1-score, c) precision, d) recall.

Fig. 9 shows the performance analysis done withBraTS 2019 dataset with k-fold. Fig. 9 (a) depicts accuracy measure. At k-fold 10, the accuracy observed by HATDL-DCNN-BiLSTM for epoch 20 is 88.84%, 90.81%, 93.40%, 95.93%, and 96.34%. Analysis done by varying F1-score is depicted in Fig. 9 (b). For k-fold 10, the F1-score measured by HATDL-DCNN-BiLSTM by varying the epoch from 20 to 100 is 88.82%, 90.80%, 93.38%, 95.92%, and 96.33%. The precision used to analyze the performance is shown in Fig. 9 (c). The precision obtained by proposed HATDL-DCNN-BiLSTM for k-fold 10 with epoch 20 is 88.18%, epoch 40 is 90.46%, epoch 60 is 92.23%, epoch 80 is 95.80%, and epoch 100 is 96.15%. Fig. 9 (d) illustrates the recall measure. With 10 k-fold value, the recall computed by HATDL-DCNN-BiLSTM by varying the epoch from 20 to 100 is 89.47%, 91.15%, 94.55%, 96.03%, and 96.51%.

F. Comparative Methods

To evaluate the performance, the proposed HATDL-DCNN-BiLSTM is compared with the existing techniques, like parallel deep convolutional neural network (PDCNN) [11], unified deep learning model (TumorDetNet) [17], Semi-Supervised Brain Tumor Classification Network (SSBTCNet) [10], Ensemble Deep Neural Support Vector Machine (EDN-SVM), and DCNN-BiLSTM [36] [37].

V. Discussion

This section elaborates the comparative discussion of proposed HATDL-DCNN-BiLSTM model with two different datasets.

1) Analysis with BraTS 2018 dataset: Fig. 10 illustrates the comparative analysis of proposed HATDL-DCNN-BiLSTM model with existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM methods using BraTS 2018 dataset in terms of training percentage. For training percentage of 80%, the proposed HATDL-DCNN-BiLSTM model obtained the accuracy of 97.55%, which is superior than existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM methods by 12.80%, 18.59%, 8.38%, 16.06% and 6.45% respectively. The F1-Score of the proposed HATDL-DCNN-BiLSTM model achieves the result of 12.73%, 18.61%, 9.34%, 16.05%, and 6.44% higher than the existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM approaches for 80% of training percentage. The precision of the HATDL-DCNN-BiLSTM model attains the result of 96.70%, which is improved over the existing PDCNN, by 11.41%, EDN-SVM by 16.33%, TumorDetNet by 5.74%, SSBTCNet by 14.62%, and DCNN-BiLSTM by 4.70% for 80% training data. Finally, the recall of the HATDL-DCNN-BiLSTM model is 98.38%, which is improved to 14.05% with PDCNN, 20.80% with EDN-SVM, 12.74% with TumorDetNet, 17.46% with SSBTCNet, and 8.16% with DCNN-BiLSTM with 80% training data.

Fig. 11 show the comparative results of the proposed HATDL-DCNN-BiLSTM model with existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM methods using BraTS 2018 dataset in terms of k-fold. The HATDL-DCNN-BiLSTM model achieves the accuracy of 96.62%, which shows that the accuracy is superior over the existing PDCNN by 5.46%, EDN-SVM by 12.71%, TumorDetNet by 3.75%, SSBTCNet by 7.01%, and DCNN-BiLSTM by 2.09% for k-fold of 10 respectively. The HATDL-DCNN-BiLSTM model achieves the F1-score of 96.75%, which shows that the F1-score of the HATDL-DCNN-BiLSTM model is 6.07%, 12.84%, 3.90%, 7.32%, and 2.24% advanced

than the existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM approaches respectively. The precision of the HATDL-DCNN-BiLSTM model for k-fold of 10 is 95.91%, which increased 4.62% with PDCNN, 11.91% ith EDN-SVM, 2.53% with TumorDetNet, 5.66% with SSBTCNet, and 1.14% with DCNN-BiLSTM. Moreover, for k-fold of 10, recall of the HATDL-DCNN-BiLSTM model is 97.61%, which is higher than the existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM methods by 7.50%, 13.77%, 5.26%, 8.95%, and 3.34% respectively.

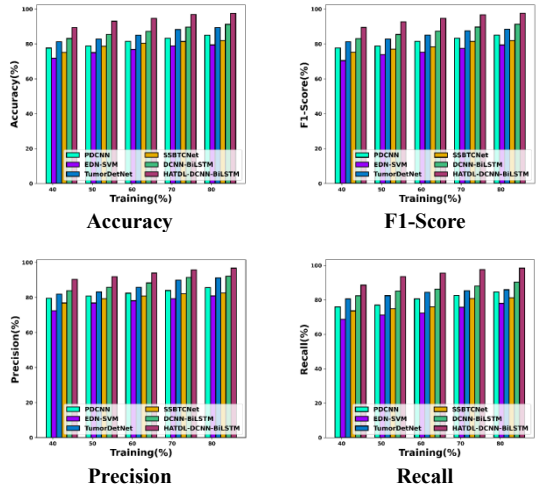


Fig. 10. Comparative analysis of HATDL-DCNN-BiLSTM with training percentage using BraTS 2018 dataset.

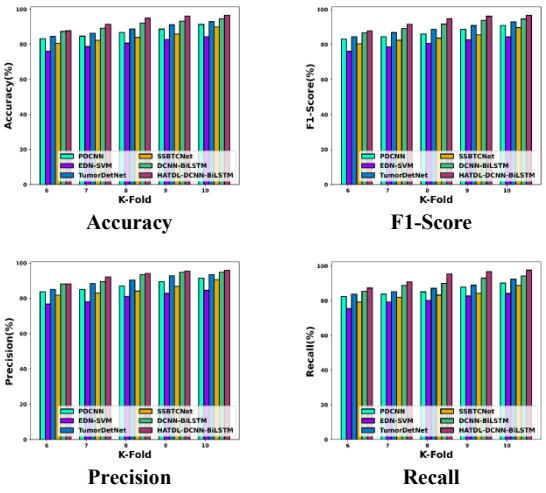


Fig. 11. Comparative analysis of HATDL-DCNN-BiLSTM with k-fold using BraTS 2018 dataset.

2) Analysis with BraTS2019 dataset: Comparative analysis of proposed HATDL-DCNN-BiLSTM model with existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM methods using BraTS 2019 dataset in terms of training percentage is shown in Fig. 12. The accuracy of the proposed HATDL-DCNN-BiLSTM model for 80% of training percentage is 98.93 which is compared with existing schemes, that reveals the improvement of 12.63% with PDCNN, 17.78% with EDN-SVM, 11.31% with TumorDetNet, 13.82% with SSBTCNet, and 6.39% with DCNN-BiLSTM respectively. The F1-Score of the HATDL-DCNN-BiLSTM model at 80% training percentage is 97.67%, which is higher than the existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM techniques by 11.52%, 16.75%, 10.18%, 12.76%, and 6.25% respectively. The HATDL-DCNN-BiLSTM model attains the precision of 96.17%, which depicts that it is higher than the conventional PDCNN by 9.48%, EDN-SVM by 14.19%, TumorDetNet by 8.44%, SSBTCNet by 10.52%, and DCNN-BiLSTM by 3.53% for training percentage of 80%. Finally, the recall of the proposed HATDL-DCNN-BiLSTM model is 13.53%, 19.24%, 11.91%, 14.96% and 8.91% superior than the existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM approaches respectively.

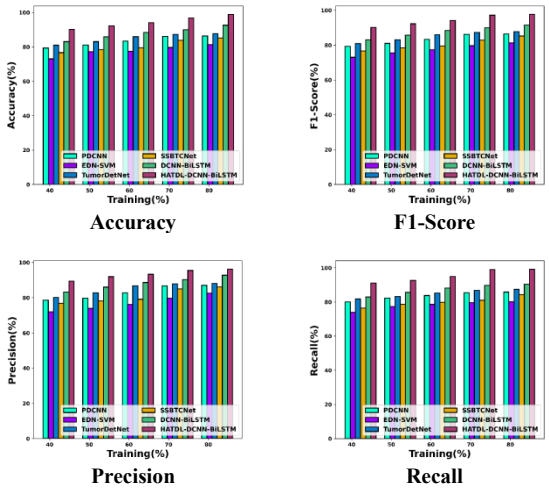


Fig. 12. Comparative analysis of HATDL-DCNN-BiLSTM with training percentage using BraTS 2019 dataset.

Fig. 13 illustrates the comparative outcomes of proposed HATDL-DCNN-BiLSTM model with existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM methods using BraTS 2019 dataset in terms of k-fold. For k-fold 10, accuracy of HATDL-DCNN-BiLSTM model is 96.34%, which is higher than the existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM methods by 7.31%, 14.71%, 3.93%, 10.20%, and 1.69% respectively. The HATDL-DCNN-BiLSTM model attains the F1-Score of 96.33%, which shows it is superior than the existing PDCNN by 7.31%, EDN-SVM by 14.72%, TumorDetNet by 3.93%, SSBTCNet by 10.21%, and DCNN-BiLSTM by 1.69% for k-fold 10 respectively. The precision of proposed HATDL-DCNN-BiLSTM is 7.27%, 14.15%, 3.56%, 10.71% and 1.19% advanced than the existing PDCNN, EDN-SVM, TumorDetNet, SSBTCNet, and DCNN-BiLSTM approaches respectively. The recall of the HATDL-DCNN-BiLSTM model for k-fold of 10 is 96.51%, which increased 7.36% with PDCNN, 15.28% with EDN-SVM, 4.30% with TumorDetNet, 9.70% with SSBTCNet, and 2.19% with DCNN-BiLSTM respectively.

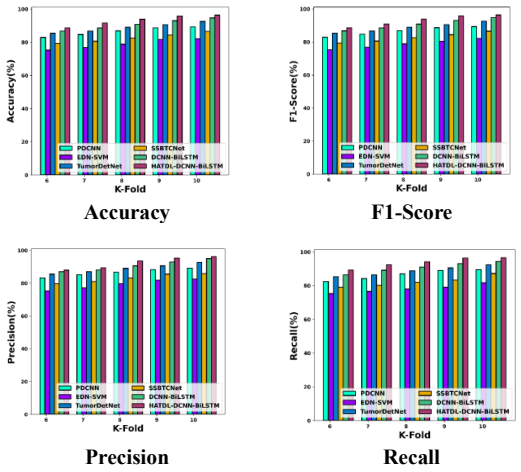


Fig. 13. Comparative analysis of HATDL-DCNN-BiLSTM with k-fold using BraTS 2019 dataset.

3) Comparative discussion: Table I illustrates the comparative discussion of HATDL-DCNN-BiLSTM with other existing methods using BraTS 2018 dataset. The existing methods evaluated for BT classification obtains inaccurate results with less accuracy. The PDCNN method [11] failed to use the model with 3D structure for identifying the tumors. Also, the TumorDetNet designed in [17] higher accuracy measure, but failed to identify the tumor grades. The SSBTCNet for BT classification failed to detect the tumor types exactly [10]. The EDN-SVM developed for the tumor classification provide higher performance but results computational complexity issues. The above stated issues are resolved through the proposed HATDL-DCNN-BiLSTM using the distributed learning mechanism of deep neural network model. The model effectively reduces the overfitting issues and offers higher performance in terms of the evaluation measures. The proposed HATDL-DCNN-BiLSTM framework minimizes the computational issues globally through the extraction of optimal features. The features extracted through different models have varying dimensions and these varying dimensional features are resized to have same dimension for all the features such that this process minimizes the computational complexity issues effectively. The proposed HATDL-DCNN-BiLSTM value is measured by altering the training percentage and k-fold value. With training percentage, the proposed obtained higher value of 97.55%, 97.54%, 96.70%, and 98.38% for accuracy, F1-score, precision, and recall. The proposed HATDL-DCNN-BiLSTM showed 96.62%, 96.75%, 95.91%, and 97.61% for accuracy, F1-score, precision, and recall and these values are measured byBraTS 2018 dataset.

Table II illustrates the comparative discussion of HATDL-DCNN-BiLSTM with other existing methods using BraTS 2019 dataset. With this dataset, the proposed HATDL-DCNN-BiLSTM obtained the accuracy, F1-score, precision, and recall as 98.93%, 97.67%, 96.17%, and 99.21 by changing the training percentage. On the other hand, proposed HATDL-DCNN-BiLSTM obtained accuracy, F1-score, precision, and recall as 96.34%, 96.33%, 96.15%, and 96.51% with k-fold value.

TABLE I. COMPARATIVE DISCUSSION OF HATDL-DCNN-BILSTM WITH BRATS 2018 DATASET

Methods		PDCNN	EDN-SVM	TumorDetNet	SSBTCNet	DCNN-BiLSTM	Proposed HATDL-DCNN-BiLSTM
TP=80%	Accuracy (%)	85.07	79.42	89.38	81.89	91.26	**97.55**
	F1-Score (%)	85.11	79.38	88.42	81.87	91.24	**97.54**
	Precision (%)	85.67	80.91	91.15	82.56	92.15	**96.70**
	Recall (%)	84.56	77.92	85.85	81.20	90.36	**98.38**
K-fold=10	Accuracy (%)	91.34	84.33	92.99	89.84	94.59	**96.62**
	F1-Score (%)	90.88	84.32	92.98	89.67	94.58	**96.75**
	Precision (%)	91.48	84.48	93.48	90.48	94.81	**95.91**
	Recall (%)	90.29	84.17	92.48	88.87	94.35	**97.61**

TABLE II. COMPARATIVE DISCUSSION OF HATDL-DCNN-BILSTM USING BRATS 2019 DATASET

Methods		PDCNN	EDN-SVM	TumorDetNet	SSBTCNet	DCNN-BiLSTM	Proposed HATDL-DCNN-BiLSTM
TP=80%	Accuracy (%)	86.43	81.33	87.73	85.25	92.60	**98.93**
	F1-Score (%)	86.42	81.31	87.72	85.21	91.56	**97.67**
	Precision (%)	87.05	82.52	88.05	86.05	92.78	**96.17**
	Recall (%)	85.79	80.13	87.40	84.38	90.38	**99.21**
K-fold=10	Accuracy (%)	89.29	82.16	92.55	86.51	94.71	**96.34**
	F1-Score (%)	89.28	82.15	92.54	86.50	94.70	**96.33**
	Precision (%)	89.16	82.55	92.73	85.86	95.01	**96.15**
	Recall (%)	89.41	81.76	92.35	87.15	94.39	**96.51**

VI. CONCLUSION

This research designs a proposed framework named HATDL-DCNN-BiLSTM for detection and classification of BT with MRI. This method uniquely addresses the issues in BT classification through the extraction of important and essential features by applying the attention mechanisms. It shows considerable advancements in efficiency and contrast for differentiating the tumor grades. The extracted features enable the model to increase the training speed that in turn enable to generate more reliable classification results. The transfer learning model introduced in the mechanism provides more beneficial activities, like faster training, less data requirements, higher learning rate, less training time and enhanced generalization. Due to the pre-trained models located in the transfer learning, the tasks to be executed are quickly learned and prevents the overfitting issues. It showed outstanding performance using the metrics, like accuracy, recall, F1-score, and precision of 98.93%, 99.21%, 97.67%, and 96.17% with training data. Also, the proposed scheme measured higher performance in terms of accuracy, recall, F1-score, and precision of 96.34%, 96.51%, 96.33%, and 96.15% using k-fold by considering the BraTS 2019 dataset. However, the proposed method has some limitations. Model performance is dependent on data quality and size, so further improvements can be brought about by using larger diverse datasets for better generalization.

The future direction of research would be the consideration of a hybrid optimization algorithm to improve efficiency in the network so as to give better classification results during BT classification. For example, combining genetic algorithm and particle swarm optimization can enhance the feature selection and the optimization of parameters for achieving an even more accurate and stable classification result.

REFERENCES

[1] E. Schulz, and S.J. Gershman, "The algorithmic architecture of exploration in the human brain", *Current opinion in neurobiology,* vol.55, pp.7-14, 2019.

[2] S. Anantharajan, S. Gunasekaran, T. Subramanian, and R. Venkatesh, "MRI brain tumor detection using deep learning and machine learning approaches", *Measurement: Sensors,* vol.31, pp.101026, 2024.

[3] P. Saha, R. Das, and S.K. Das, "BCM-VEMT: Classification of brain cancer from MRI images using deep learning and ensemble of machine learning techniques", *Multimedia Tools and Applications*, vol.82, no.28, pp.44479-44506, 2023.

[4] A.A. Asiri, T.A. Soomro, A.A. Shah, G. Pogrebna, M. Irfan, and S. Alqahtani, "Optimized Brain Tumor Detection: A Dual-Module Approach for MRI Image Enhancement and Tumor Classification", *IEEE Access,* vol.12, pp.42868-42887, 2024.

[5] A.N. Mavrakis, E.F. Halpern, F.G. Barker, R.G. Gonzalez, and J.W. Henson, "Diagnostic evaluation of patients with a brain mass as the presenting manifestation of cancer" *Neurology,* vol.65, no.6, pp.908-911, 2005.

[6] S.M. Malakouti, M.B. Menhaj, and A.A. Suratgar, "Machine learning and transfer learning techniques for accurate brain tumor classification", *Clinical eHealth,* vol.7, pp.106-119, 2024.

[7] S. Deepa, J. Janet, S. Sumathi, and J.P. Ananth, "Hybrid optimization algorithm enabled deep learning approach brain tumor segmentation and classification using MRI", *Journal of Digital Imaging*, vol.36, no.3, pp.847-868, 2023.

[8] M. Yasmin, S. Mohsin, M. Sharif, M. Raza, and S. Masood, "Brain image analysis: a survey", *World Applied Sciences Journal*, vol.19, no.10, pp.1484-1494, 2012.

[9] J. Amin, M. Sharif, M. Raza, and M. Yasmin, "Detection of brain tumor based on features fusion and machine learning", *Journal of Ambient Intelligence and Humanized Computing,* pp.1-17, 2024.

[10] Z. Atha and J. Chaki, "SSBTCNet: Semi-Supervised Brain Tumor Classification Network", *IEEE Access,* vol. 11, pp. 141485-141499, 2023.

[11] T. Rahman, and M.S. Islam, "MRI brain tumor detection and classification using parallel deep convolutional neural networks", *Measurement: Sensors,* vol.26, pp.100694, 2023.

[12] M. Wageh, K. Amin, A. D. Algarni, A. M. Hamad and M. Ibrahim, "Brain Tumor Detection Based on Deep Features Concatenation and Machine Learning Classifiers With Genetic Selection," *IEEE Access,* vol. 12, pp. 114923-114939, 2024.

[13] M.A. Talukder, M.M. Islam, M.A. Uddin, A. Akhter, M.A.J. Pramanik, S. Aryal, M.A.A. Almoyad, Hasan, and K.F. Moni, "An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning", *Expert systems with applications*, vol. 230, pp.120534, 2023.

[14] R.L. Siegel, K.D. Miller, and A. Jemal, "Cancer statistics, 2019", *CA: a cancer journal for clinicians, vol*.69, no.1, pp.7-34, 2019.

[15] N. Abiwinanda, M. Hanif, S.T. Hesaputra, A. Handayani, and T.R. Mengko, "Brain tumor classification using convolutional neural network", *In World Congress on Medical Physics and Biomedical Engineering 2018: Prague, Czech Republic,* vol.1, pp.183-189, Springer Singapore, 2019.

[16] E.S.A. El-Dahshan, H.M. Mohsen, K. Revett, and A.B.M. Salem, "Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm", *Expert systems with Applications,* vol.41, no.11, pp.5526-5545, 2014.

[17] N. Ullah, A. Javed, A. Alhazmi, S.M. Hasnain, A. Tahir, and R. Ashraf, "TumorDetNet: A unified deep learning model for brain tumor detection and classification.," *Plos one,* vol.18, no.9, pp.0291200, 2023.

[18] F.P. Polly, S.K. Shil, M.A. Hossain, A. Ayman, and Y.M. Jang, "Detection and classification of HGG and LGG brain tumor using machine learning", *International Conference on Information Networking (ICOIN),* pp. 813-817, 2018.

[19] M. Aamir, Z. Rahman, Z.A. Dayo, W.A. Abro, M.I. Uddin, I. Khan, A.S. Imran, Z. Ali, M. Ishfaq, Y. Guan, and Z. Hu, "A deep learning approach for brain tumor classification using MRI images", *Computers and Electrical Engineering,* vol.101, pp.108105, 2022.

[20] S.K. Mathivanan, S. Sonaimuthu, S. Murugesan, H. Rajadurai, B.D. Shivahare, and M.A. Shah, "Employing deep learning and transfer learning for accurate brain tumor detection", *Scientific Reports,* vol.14, no.1, pp.7232, 2024.

[21] S.A. Nawaz, D.M. Khan, and S. Qadri, "Brain tumor classification based on hybrid optimized multi-features analysis using magnetic resonance imaging dataset", *Applied Artificial Intelligence,* vol.36, no.1, pp.2031824, 2022.

[22] A. Vidyarthi, R. Agarwal, D. Gupta, R. Sharma, D. Draheim, and P. Tiwari, "Machine learning assisted methodology for multiclass classification of malignant brain tumors", *IEEE Access,* vol.10, pp.50624-50640, 2022.

[23] M.Z. Khaliki, and M.S. Başarslan, "Brain tumor detection from images and comparison with transfer learning methods and 3-layer CNN", *Scientific Reports,* vol.14, no.1, pp.2664, 2024.

[24] G. Deng, and L.W. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection", *IEEE conference record nuclear science symposium and medical imaging conference,* pp.1615-1619, October 1993.

[25] A. Abdollahi, B. Pradhan, and A. Alamri, "VNet: An end-to-end fully convolutional neural network for road extraction from high-resolution remote sensing data", *IEEE Access,* vol.8, pp.179424-179436, 2020.

[26] X. Liu, R. Yin, and J. Yin, "Attention V-Net: a modified V-Net architecture for left atrial segmentation", *Applied Sciences*, vol.12, no.8, pp.3764, 2022.

[27] H. Huang, Z. Zuo, B. Sun, P. Wu, and J. Zhang, "DSA-SOLO: Double split attention SOLO for side-scan sonar target segmentation", *Applied Sciences,* vol.12, no.18, pp.9365, 2022.

[28] R.S. Devi, V.R. Kumar, and P. Sivakumar, "EfficientNetV2 Model for Plant Disease Classification and Pest Recognition", *Computer Systems Science & Engineering,* vol.45, no.2, 2023.

[29] M.M. Hasan, N. Alfaz, M.A.M. Alam, A. Rahman, H.M. Shakhawat, and S. Rahman, "Detection of parkinson's disease from t2-weighted magnetic resonance imaging scans using efficientnet-v2", *International Conference on Computer and Information Technology (ICCIT), IEEE,* pp.1-6, December 2023.

[30] E. Rezende, G. Ruppert, T. Carvalho, F. Ramos, and De P. Geus, "Malicious software classification using transfer learning of resnet-50 deep neural network", *IEEE international conference on machine learning and applications (ICMLA),* pp.1011-1014, December 2017.

[31] H. Zhou, and G. Yu, "Research on pedestrian detection technology based on the SVM classifier trained by HOG and LTP features", *Future Generation Computer Systems,* vol.125, pp.604-615, 2021.

[32] R. Arya, and E.R. Vimina, "Local triangular coded pattern: A texture descriptor for image classification", *IETE Journal of Research,* vol.69, no.6, pp.3267-3278, 2023.

[33] K.D. Kharat, V.J. Pawar, and S.R. Pardeshi, "Feature extraction and selection from MRI images for the brain tumor classification", *International Conference on Communication and Electronics Systems (ICCES) IEEE,* pp.1-5, October 2016.

[34] A. Chaddad, and R.R. Colen, "Statistical feature selection for enhanced detection of brain tumor", *In Applications of Digital Image Processing XXXVII, SPIE,* Vol.9217, pp.260-267, September 2014.

[35] M. Aspri, G. Tsagkatakis, and P. Tsakalides, "Distributed training and inference of deep learning models for multi-modal land cover classification", *Remote Sensing*, vol.12, no.17, p.2670, 2020.

[36] M. Marjani, M. Mahdianpari, and F. Mohammadimanesh, "CNN-BiLSTM: A Novel Deep Learning Model for Near-Real-Time Daily Wildfire Spread Prediction", *Remote Sensing,* vol.16, no.8, pp.1467, 2024.

[37] M. Méndez, M.G. Merayo, and M. Núñez, "Long-term traffic flow forecasting using a hybrid CNN-BiLSTM model", *Engineering Applications of Artificial Intelligence,* vol.121, pp.106041, 2023.

[38] BraTS 2018 dataset, "https://www.med.upenn.edu/sbia/brats2018/data.html", accessed on October 2024.

[39] BraTS 2019 dataset, "https://www.med.upenn.edu/cbica/brats2019/data.html", accessed on October 2024

A Distributed Framework for Indoor Product Design Using VR and Intelligent Algorithms

Yaoben Gong*, Zhenyu Gao
College of Art and Media Design, Nanchang Institute of Technology, Nanchang 330108, Jiangxi, China

***Abstract*—This paper presents an innovative approach to the digital design of indoor home products by integrating virtual reality (VR) technology with intelligent algorithms to enhance design accuracy and efficiency. A model combining the Red deer Optimization Algorithm with a Simple Recurrent Unit (SRU) network is proposed to evaluate and optimize the design process. The study develops a digital design framework that incorporates key evaluation factors, optimizing the SRU network through the Red deer Optimization Algorithm to achieve higher precision in design applications. The model's performance is validated through extensive experiments using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). Results show that the RDA-SRU model outperforms other methods, with the smallest MAE of 0.133, RMSE of 0.02, and MAPE of 0.015. Additionally, the model achieved an R^2 value of 0.968 and the shortest evaluation time of 0.028 seconds, demonstrating its superior performance in predicting and evaluating digital design applications for home products. These findings indicate that the integration of VR with intelligent algorithms significantly improves user experience, customizability, and the overall accuracy of digital design processes. This approach offers a robust solution for designers to create more efficient and user-centric home product designs, meeting growing customer demands for immersive and interactive design experiences.**

Keywords*—*Interior home products; virtual reality technology; digital design algorithms; improved simple cyclic units; intelligent algorithms for design application evaluation

I. INTRODUCTION

With the development and popularization of Internet technology and artificial intelligence (AI) algorithms, the design field is diversifying, especially in the direction of digital applications, which has become the trend of the times and has unlimited development potential [1]. Virtual reality (VR) technology as the development of active new technology, constantly integrated into people's lives, especially in the field of scene visualization is widely used [2]. The combination of VR technology and realistic real interior home product design, constructing an interior home design scene close to the real world, promoting the design of interior home products, improving the degree of tacit understanding between the product and the user to a certain extent, and improving the user's experience of the product [3]. With the arrival of the product experience-oriented consumer design era, indoor home product design needs to adapt to the needs of humanization, combined with VR technology has become one of the ways to solve the problem of meeting the needs of users in home product design [3]. Based on VR technology indoor home products digital design method from the visual can improve the sense of interesting online home operation experience, while improving the efficiency of product design [4]. The accurate evaluation of digital design technology as a combination of VR technology in the home product design process feedback is an important link to help improve the accuracy of customer-oriented demand indoor home product design, but also improve the integrity of the home product design process [5]. The research direction of the digital design of indoor home products under the VR scene mainly focuses on the construction of scenes in virtual home product design, product virtualization design, product interaction design, digital application test analysis and other aspects of the research [6]. Scholars have made many contributions. Basha et al. [7] optimized the method and scene of secondary modeling and proposes a rapid modeling system based on VR technology; Sobhy and Abouelnaga [8] studied the value of the integration of VR technology and indoor home product design and introduces the stage-by-stage method of integration design; Wei et al. [9] analyzed the development trend of VR technology and combines with the characteristics of the experiential display design, and analyzes the characteristics of the experiential display design from a variety of perspectives VR technology in indoor home design application characteristics; Ye and Li [10] puts forward a virtual scene optimization layout strategy that can adapt to the real-time interactive display of different cell phone devices, and realizes the virtual scene of indoor home product design; Hewitt et al. [11] researched the design and implementation of a three-dimensional product interactive display website using computer language; Rangaswamy et al. [12] built a digital design based on VR technology process, and extracted relevant digital design evaluation indexes, and used hierarchical analysis to solve the problem of digital design evaluation of indoor furniture products.

With the development of AI algorithms, digital design application methods based on machine learning algorithms and deep learning algorithms enter the field experts and scholars [13]. Currently, design application evaluation methods based on intelligent algorithms include k-means clustering algorithm, support vector machine, decision tree, neural network, deep learning and so on [14]. At present, the research on the method of digital design of indoor home products based on intelligent algorithms combined with VR technology is less, in its infancy, and the effect and accuracy of the application formed cannot meet the needs of customers [15, 16]. For this reason, this paper proposes a digital design method for indoor home products based on the Red deer Optimization Algorithm to improve the simple recurrent unit neural network [17, 18]. Focusing on the digitalization of indoor home products, the digital design framework of indoor home products is constructed by analyzing

*Corresponding Author.

the digital design process, and the key technologies of digital design are illustrated and introduced; for the evaluation of the digital design application of indoor home products in combination with VR technology, the simple recurrent unit network is trained and optimized using Red deer Optimization Algorithm, and the digital design evaluation model based on RDA-SRU network is constructed; the digital design application evaluation model is verified through experiments. The digital design application evaluation model based on RDA-SRU network is constructed by using the Red deer optimization algorithm; the application feasibility and computational efficiency of the method proposed in this paper are proved through experimental verification.

II. Digital Design Framework for Interior Home Products

A. Virtual Reality

VR technology is through the comprehensive use of computer graphics science on the seeming, multimedia teaching science on the seeming, human-computer interaction science and technology, Internet information, sensors and check measurement technology, three-dimensional performance technology and simulation technology and other types of disciplines and technologies, to create a real audio-visual, touch one of the artificial virtual environment [19], specifically as shown in Fig. 1. In the virtual environment, the user with the help of professional technical equipment, the most natural way to interact with the objects in life, so as to achieve a stronger sense of immersion and real feelings.

The key components of VR technology include headset devices, intrusion devices, 3D audio and sound technologies, and simulated environment modeling and rendering [19] (Fig. 2), which are mainly used in the fields of gaming, education, healthcare, industrial design, entertainment, and military [20] (Fig. 3).

B. Digital Design Process

1) Definitions and key techniques, tools: Digital design of interior home products is the process of designing, simulating, optimizing and manufacturing home products using advanced computer technology, information technology and digital tools. This design methodology improves design efficiency, reduces errors, lowers costs, and allows for better customization and service. Digital design can also help companies achieve scale customization, flexible production, and networked services, thus improving competitiveness and market responsiveness [21].

The key technologies and tools used for the digital design of interior home products include CAD/CAE/CAPP/CAM/PDM integration software, parametric modeling technology, VR (VR)/augmented reality (AR) technology, product lifecycle management (PLM), and enterprise resource planning (ERP) systems, etc., as shown in Fig. 4.

The process of digital design for interior home products is shown in Fig. 5, with the following steps: a) Demand analysis: collecting information on customers' individual needs and market trends; b) Conceptual design: based on the results of the demand analysis, preliminary design ideas; c) Detailed design: using digital tools to carry out accurate product design, including dimensions, materials, colours, etc.; d) Simulation and testing: by computer simulation to verify the feasibility and performance of the design; e) Prototyping: making samples based on the design drawings for actual testing and evaluation; f) Production preparation: after completing the design, preparing the documents and processes required for production; g) Production and quality control: monitoring the product quality during the production process to ensure compliance with the design; and h) Feedback services: providing installation guidance, maintenance services and customer support [22].

Technology can help designers and clients preview and experience design solutions in three-dimensional space, improving the interactivity and realism of design, and is mainly used in the c) detailed design phase.

Fig. 1. Virtual reality technology.

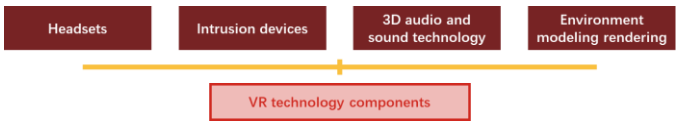


Fig. 2. VR technology components.

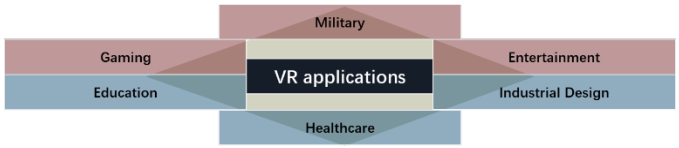


Fig. 3. Application areas of VR technology.

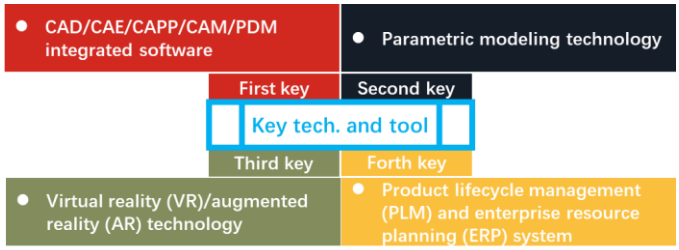


Fig. 4. Key technologies and tools for digital design of interior home products.

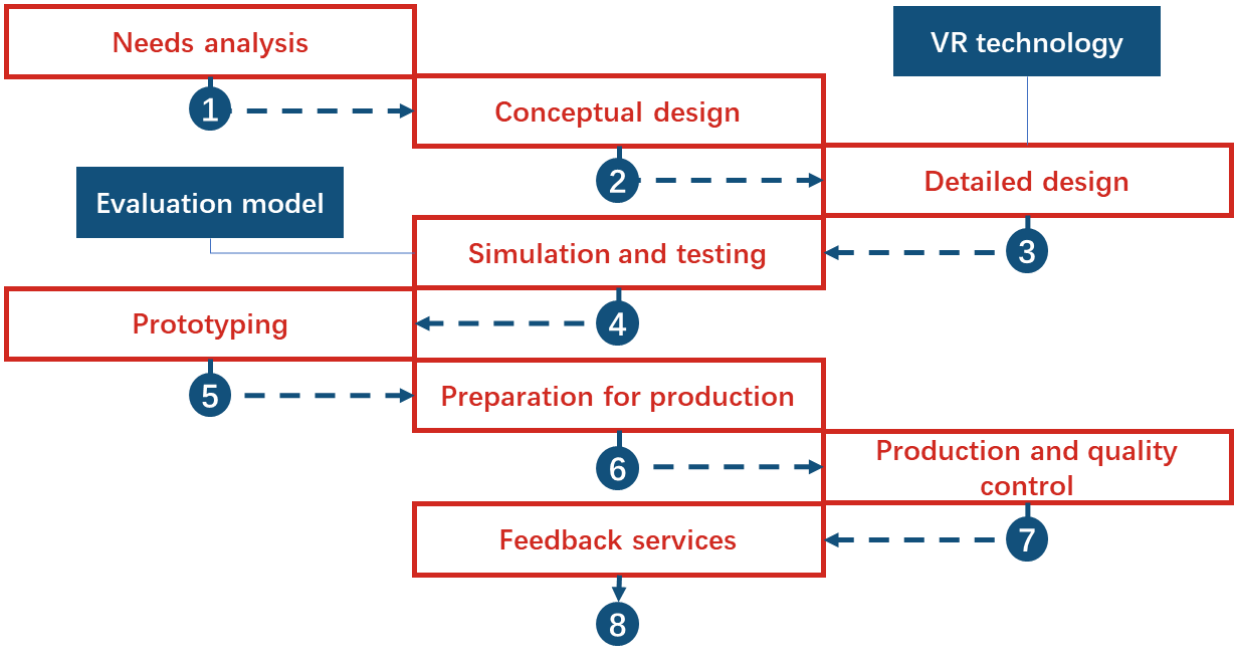


Fig. 5. Process of digital design for interior home products.

Digital design application evaluation can help designers quantify the suitability of interior home products and accelerate the improvement of the efficiency of digital design of interior home products, which is mainly used in the d) simulation and testing phases.

C. Digital Design Framework and Key Technology Analysis

1) Digital design framework: In order to improve the efficiency of digital design of indoor home products based on VR technology and enhance the evaluation accuracy of digital design application of indoor home products, this paper combines the intelligent optimization algorithm (Red deer Optimization Algorithm) and the deep learning algorithm (Simple Recurrent Unit Network) to propose a digital design method of indoor home products based on the RDA-SRU model, and the specific framework is shown in Fig. 6.

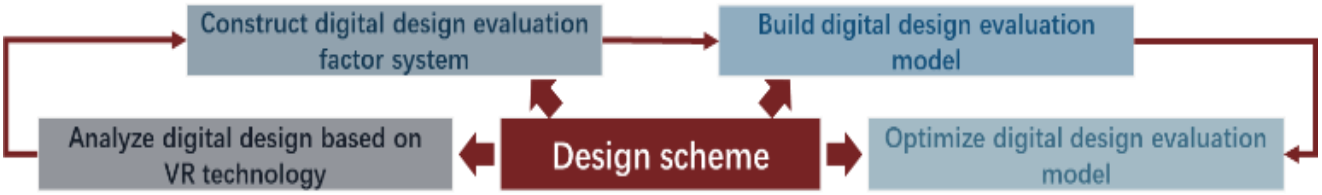


Fig. 6. Digital design framework for interior home products.

According to the analysis of the digital design framework, the digital design method of indoor home products based on RDA-SRU model by combining VR technology mainly consists of subsystems such as evaluation factor set construction, data preprocessing, digital design evaluation model construction optimization, evaluation performance analysis, etc., which are shown in Fig. 7.

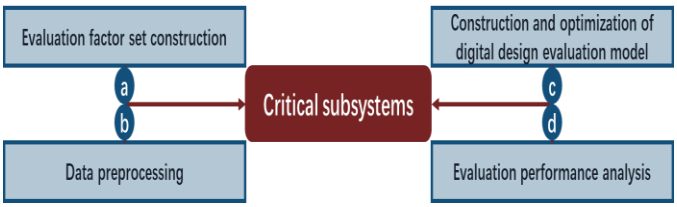


Fig. 7. Key subsystems for digital design of indoor home products.

2) Analysis of key technologies

a) Evaluation factor set construction: In the evaluation factor set construction subsystem, the digital design process of indoor home products combined with VR technology is analyzed to extract the digital design evaluation feature indicators and construct the evaluation factor set, and the specific subsystem inputs, outputs, and structures are shown in Fig. 8.

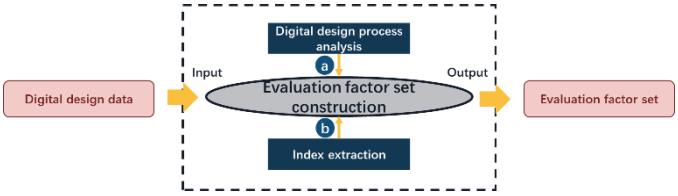


Fig. 8. Subsystem for constructing the factor set for digital design evaluation.

According to the principles of factor selection, such as demand orientation, objective scientificity, and quantitative comparability, the evaluation factors of digital design of indoor home products combined with VR technology are extracted from the phases of environment design F, 3D modeling S, and human-computer interaction T, and the evaluation factor set shown in Fig. 9 is constructed.

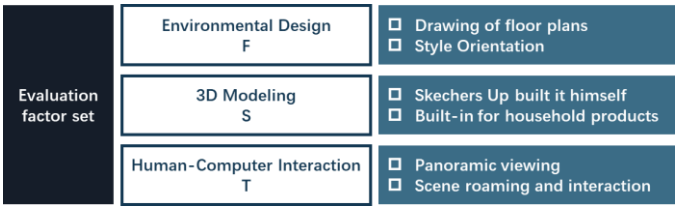


Fig. 9. Evaluation factor set construction.

b) Data pre-processing: In the data preprocessing subsystem of evaluation indexes for the digital design process of indoor home products, the data set used for training the evaluation model is obtained in a regular way through the outlier elimination technique, missing value regression filling technique, normalization processing technique, and dimensionality reduction technique [23], and the specific subsystem inputs, outputs, methods, and structures are shown in Fig. 10.

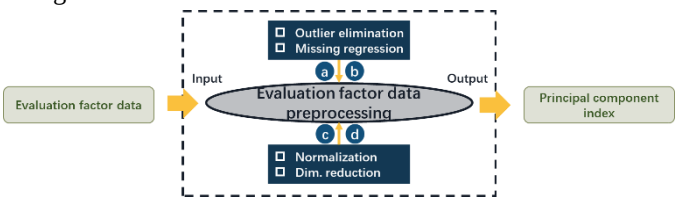


Fig. 10. Schematic diagram of the digital design evaluation data preprocessing subsystem.

c) Optimization of digital design evaluation model construction: In the optimization subsystem of evaluation model construction, the Red deer optimization algorithm is used to learn and train the weights and biases of the simple cyclic unit network, and to construct the digital design evaluation model based on the Red deer optimization algorithm-simple cyclic unit network, and the specific subsystem inputs, outputs, methods, and structures are shown in Fig. 11.

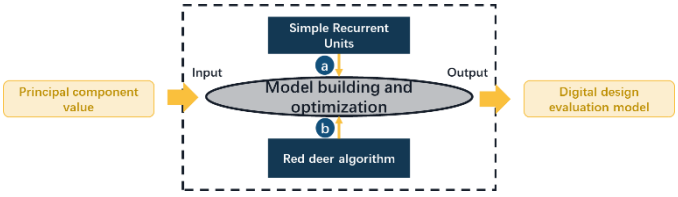


Fig. 11. Schematic diagram of the optimization subsystem for digital design evaluation model construction.

d) Evaluation performance analysis: In the digital design evaluation model performance analysis subsystem, MAE, RMSE, MAPE, R2, and evaluation time [24] are used as the performance evaluation indexes of the digital design evaluation model, as shown in Fig. 12.

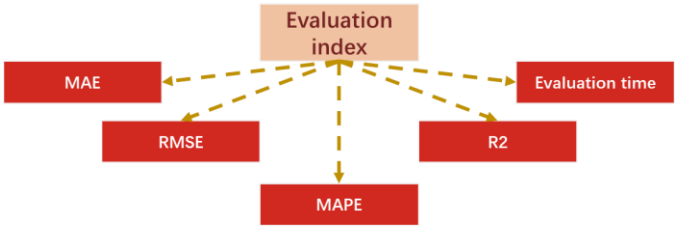


Fig. 12. Performance evaluation indicators.

III. Red Deer Optimization Algorithm for Interior Home Products

A. Simple Cyclic Unit Networks

1) Fundamentals: SRU (Simple Recurrent Units) [25] is a simplified recurrent neural network (RNN) architecture designed to increase the training speed of RNNs and allow for more efficient parallel computation. SRUs enable the network to be trained faster on modern GPUs by reducing the dependencies between time steps. SRUs are designed with the concept of optimizing the main portion of the computation is optimized to be a complete computation that does not depend on previous time steps, thus enabling an easily parallelizable network structure, the structure of which is shown in Fig. 13.

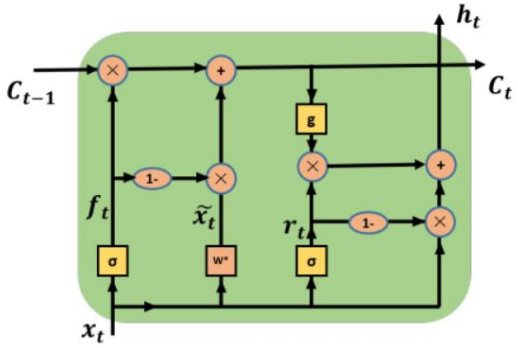


Fig. 13. Structure of SRU network model.

$$\tilde{x}_t = W x_t \tag{1}$$

$$f_t = \sigma\left(W_f x_t + b_f\right) \tag{2}$$

$$r_t = \sigma\left(W_r x_t + b_r\right) \tag{3}$$

$$c_t = f_t \odot c_{t-1} + \left(1 - f_t\right) \odot \tilde{x}_t \tag{4}$$

$$h_t = r_t \odot g\left(c_t\right) + \left(1 - r_t\right) \odot x_t \tag{5}$$

Where f_t is the forget gate and r_t is the reset gate.

2) Characteristics of SRU: The SRU network is characterized by the following features: 1) parallelization capability, 2) reduced computational complexity, and 3) gradient propagation improvement [26], as shown in Fig. 14.

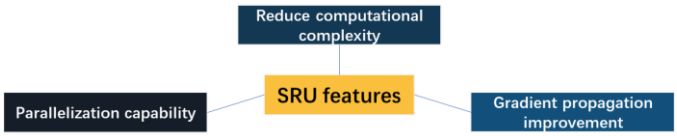


Fig. 14. SRU network characteristics.

3) SRU applications: SRU networks are suitable for a variety of deep learning tasks that require the processing of sequential data, including but not limited to the fields of natural language processing, speech recognition, and time series analysis, due to their improved training speeds and parallel computing capabilities [26] (Fig. 15).

Fig. 15. SRU web application.

B. The Redouble Optimization Algorithm

1) Algorithmic principles: The Red deer optimization algorithm (RDA) [27] was inspired by observing the mating activity of the horse deer species living in the UK during the breeding season. The population of horse deer is divided into four main species: male deer (Male), female deer (Hind), commander deer (Com), and stag (Stag). The RDA algorithm simulates the behaviors of roaring, fighting, and mating in a group of horse deer to solve the optimization problem, and demonstrates a better performance in complex problems.

a) Generation of initialized horse deer populations: In the RDA algorithm, the optimization object is a horse deer and each horse deer is represented as follows:

$$X_i = \left[x_1, x_2, x_3, \cdots, x_n\right] \tag{6}$$

Where, X_i is the i^{th} individual horse deer and n is the dimension of the RDA problem to be optimized.

In order to evaluate the merit of individual horse deer, it is necessary to introduce a fitness function:

$$f\left(X_i\right) = f\left(x_1, x_2, x_3, \cdots, x_n\right) \tag{7}$$

where $f\left(X_i\right)$ is the value of X_i fitness function.

Based on the size of the fitness value, the better N_{male} stags were selected as males and the remaining stags as females ($N_{hind} = N_{pop} - N_{male}$), the size of N_{male} denotes the algorithmic elite criterion, which maintains the augmentation properties of the algorithm, and N_{hind} takes into account the diversification phase of the algorithm.

b) Roaring stage between male deer: The roaring phase (Fig. 16) is searching the neighborhood space, and if a better Red deer is found in the neighborhood, the previous Red deer is replaced by the better one, as modeled below:

$$New_{male} = \begin{cases} Old_{male} + a_1\left(\left(UB - LB\right) \times a_2 + LB\right) & a_3 \geq 0.5 \\ Old_{male} - a_1\left(\left(UB - LB\right) \times a_2 + LB\right) & a_3 < 0.5 \end{cases} \tag{8}$$

Among them, Old_{male} is the current position of male deer, New_{male} is the updated position of male deer, UB and LB are the upper and lower bounds of the search space, respectively, and a_1 , a_2 and a_3 are randomly generated numbers.

a) Commander deer selection: The division of males into Commander Deer (Com) and Stag (Stag), primarily to differentiate between stags that are more powerful in combat and more successful in expanding their territories, is calculated as follows:

$$N_{Com} = round\left\{\gamma \cdot N_{male}\right\} \tag{9}$$

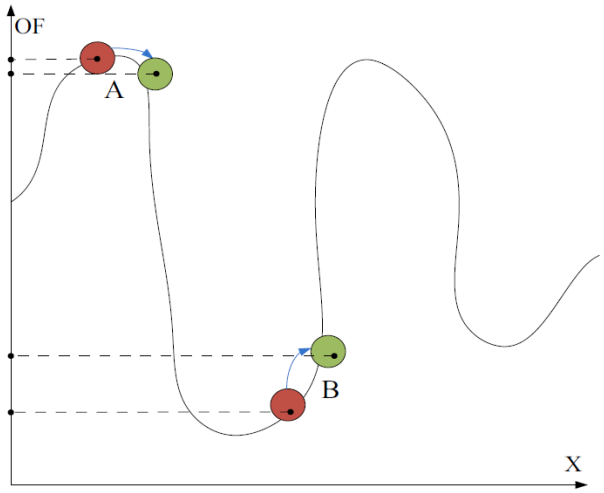


Fig. 16. The process of yelling.

Where, N_{Com} is the number of commander deer, γ is the initial value of the algorithm model, the value range is 0 to 1, the number of stags is $N_{Stag} = N_{male} - N_{Com}$, the specific distribution of the horse deer population is shown in Fig. 17.

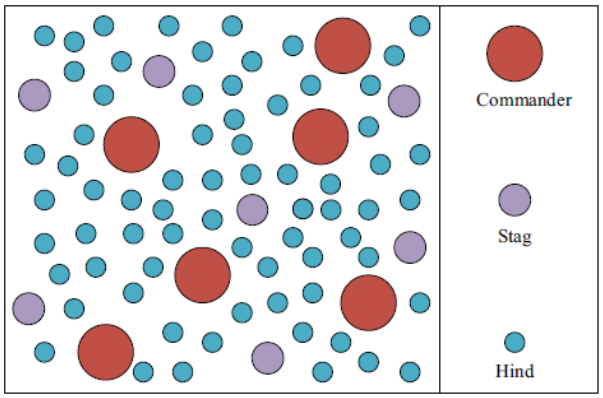


Fig. 17. Distribution of red deer populations.

b) Combat phase: From Fig. 18, Each commander deer will randomly fight a stag by approaching the other, causing a change in position, comparing fitness values, and replacing the commander deer with the superior stag individual, as calculated below:

$$New1 = \frac{(Com + Stag)}{2} + b_1 \times \left((UB - LB) \times b_2 + LB\right) \quad (10)$$

$$New2 = \frac{(Com + Stag)}{2} - b_1 \times \left((UB - LB) \times b_2 + LB\right) \quad (11)$$

Where $New1$ and $New2$ are two solutions generated by the combat process, Com and $Stag$ are commander deer and stag respectively, and b_1 and b_2 are uniformly distributed random numbers.

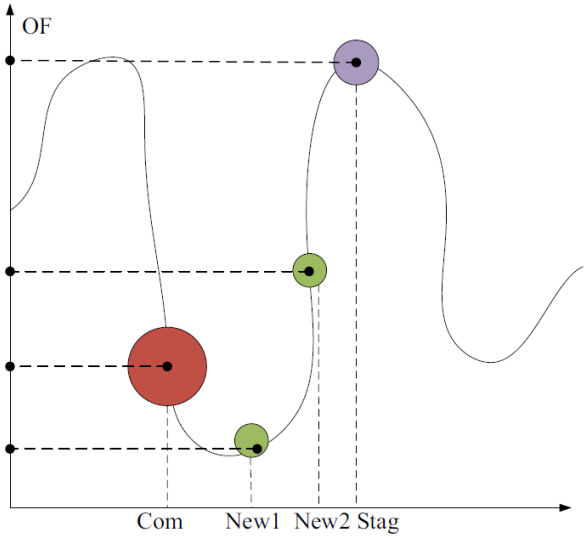


Fig. 18. Fighting process.

c) Formation of polygamy: The formation of polygamy can be referred to as the formation of a harem (Fig. 19), which is a group of female deer captured by a single commander deer. The number of female deer in the harem depends on the commander deer combat effectiveness. Forming a harem requires defining the commander deer standardized cost, which is modeled as follows:

$$V_n = v_n - \max_i \{v_i\} \quad (12)$$

$$P_n = \left| \frac{V_n}{\sum_{i=1}^{N_{Com}} V_i} \right| \quad (13)$$

$$N.harem_n = round\{P_n.N_{hind}\} \quad (14)$$

Where v_n is the fighting strength of the nth commander deer, $\max_i \{v_i\}$ denotes the maximum value of the fitness function for all stags, V_n is the standardized fighting strength, P_n is the standardized fighting strength, and N_{Com} denotes the commander deer. The standardized fighting power of commander deer includes a part of female deer, then the number of female deer in the harem is $N.harem_n$.

a) Mating stage: Commander deer mated with α% females, calculated as follows:

$$N.harem_n^{mate} = round\{\alpha.N.harem_n\} \quad (15)$$

Where, $N.harem_n^{mate}$ is the number of females in the nth harem that mated with the commander deer, and α is the value of the initial parameters of the RDA model. The locations of the next generation of mating-produced Red deer were as follows:

$$offs = \frac{(Com + Hind)}{2} + (UB - LB) \times c \quad (16)$$

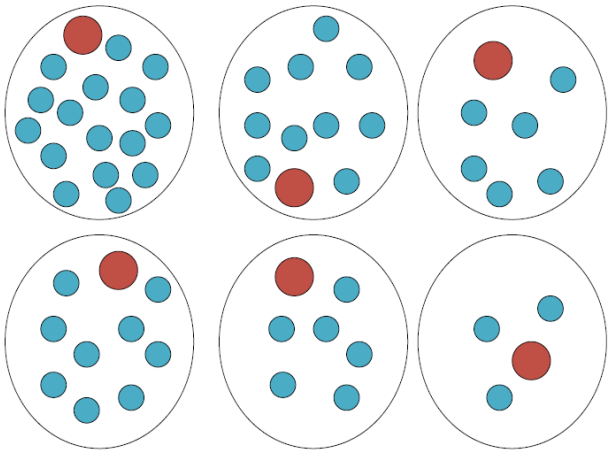


Fig. 19. Polygamy.

where $offs$ denotes a new problem-solving individual horse deer and c denotes a random number.

In the mating of a commander deer with a β% female, a harem is randomly selected and the commander road is allowed to mate with its β% female, mainly to exaggerate territory, calculated as follows:

$$N.harem_k^{mate} = round\{\beta.N.harem_k\} \quad (17)$$

where $N.harem_k^{mate}$ is the number of female deer in the k^{th} harem and β is the initial parameter that produces new individuals.

Male deer mated with the most recent female deer, calculated as follows:

$$d_i = \sqrt{\sum_{j \in J} \left(stag_j - hind_j^i\right)^2} \quad (18)$$

$$offs = \frac{(Stag + Hind)}{2} + (UB - LB) \times c \tag{19}$$

where d_i is the distance between the i^{th} female deer and the stag.

b) Choosing the next generation: In order to speed up convergence and avoid falling into local optimality, the RDA algorithm chooses two different strategies. The first strategy is to retain all male deer; the second strategy is to select among the fitness values and the offspring produced by the mating process using a roulette mechanism.

2) Algorithm flowchart: According to the optimization strategy and principle of Red deer's optimization algorithm, the flow of RDA algorithm is shown in Fig. 20.

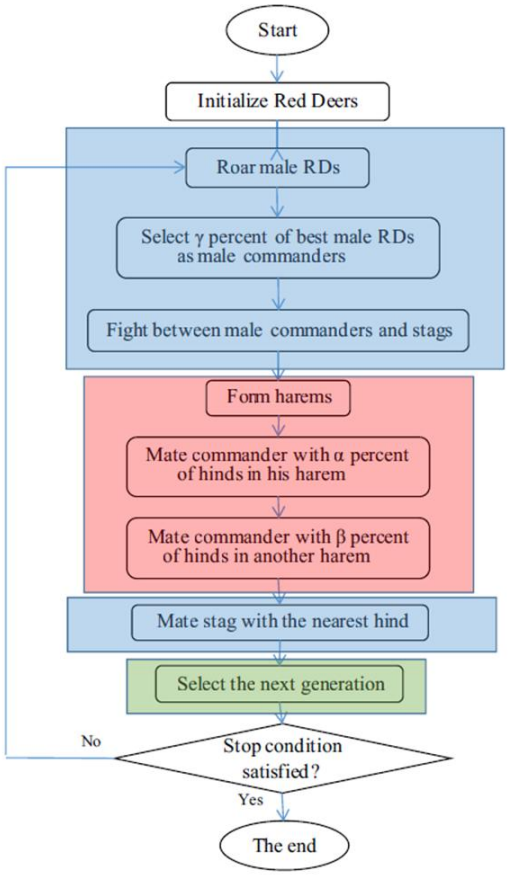


Fig. 20. Flowchart of the redouble optimization algorithm.

C. Digital Design Application Method Based on RDA-SRU Network

The application of RDA-SRU network for digital design of indoor home products is analyzed in three main aspects, i.e., decision variables, fitness function, and optimization steps.

In terms of decision variables, the optimized variables are W ,W_f ,W_r ,b_f ,b_r of the SRU network; in terms of the fitness function, the chosen function is the MAE value; the steps of the application of the RDA-SRU network model in the problem are shown in Fig. 21.

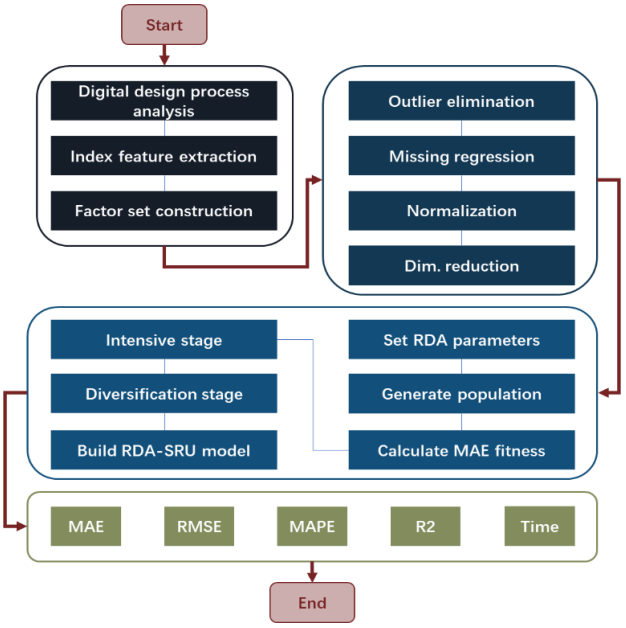


Fig. 21. Flowchart of RDA-SRU model application.

IV. Experimental Analysis

A. Experimental Setup

Digital design software includes CAD, Photoshop, Sketch up and Unity 3D engine; the computer processor is Intel7 Core Quad Core i7-7700HQ 2.8GHz with NVIDIA GeForce GTX10708GB discrete video memory, and the VR headset device is Xiaomi headset.

Using the MIVR headset device in conjunction with the NOLO interactive positioning device, the built-in sensors go out to receive signals and transmit them to Unity3D for processing.

B. Analysis of Experimental Results

With the help of drawing tool AutoCAD, the indoor scene plan is drawn as shown in Fig. 22. After being processed by Photoshop software, it is imported into Sketch up software to get the 3D model, as shown in Fig. 23. Through the VR software, build the virtual furniture scene, as shown in Fig. 24. The virtual scene is refined to get the VR realization effect as shown in Fig. 25.

In order to evaluate the effect of digital design of indoor home products based on VR technology, RNN, LSTM, GRU, SRU, RDA-SRU and other evaluation models are used to compare and analyze the design methods, and the specific results are shown in Fig. 26, 27 and 28.

The MAE, RMSE, MAPE, R2, and evaluation time performance comparison of different algorithms is given in Fig. 26. In terms of MAE, the RDA-SRU model has the smallest MAE value of 0.133, followed by SRU, GRU, LSTM, and RNN; in terms of RMSE, the RDA-SRU model has the smallest value of 0.02; in terms of MAPE, the RDA-SRU model has the smallest value of 0.015, followed by GRU, LSTM, SRU, and RNN; in terms of R2 , the RDA-SRU model is the largest with 0.968, followed by GRU, SRU, LSTM, and RNN; in terms of evaluation time, RDA-SRU model takes the shortest time with 0.028.

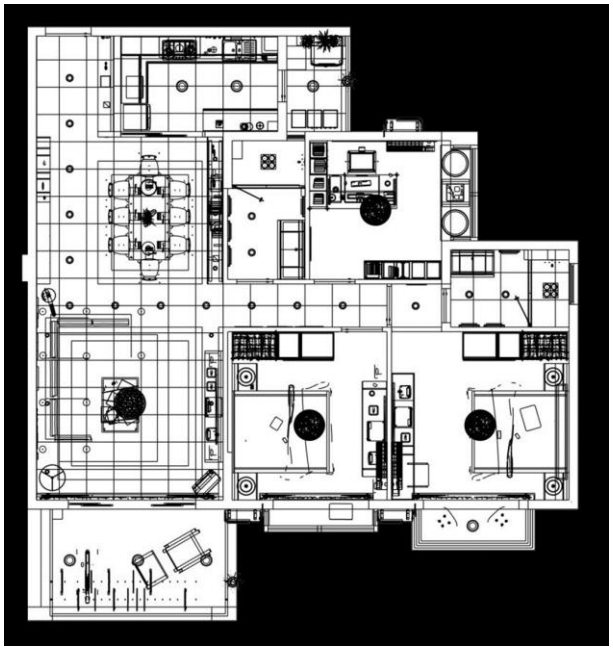


Fig. 22. Indoor scene floor plan.

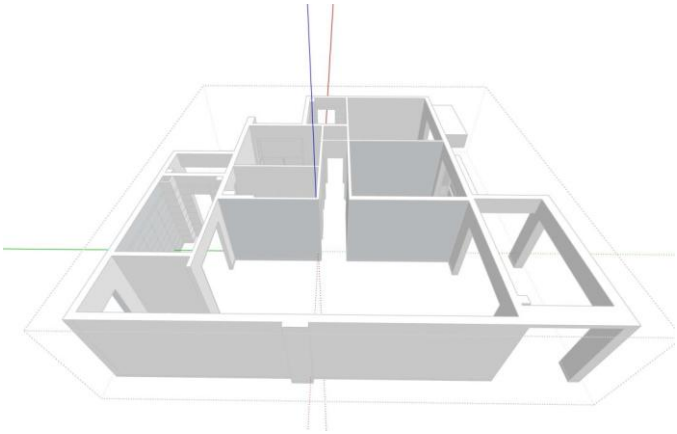


Fig. 23. Schematic diagram of 3D model.

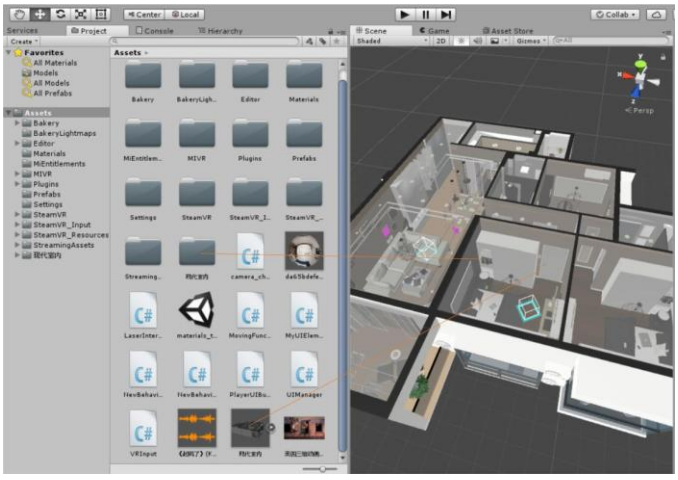


Fig. 24. Virtual engine scene.

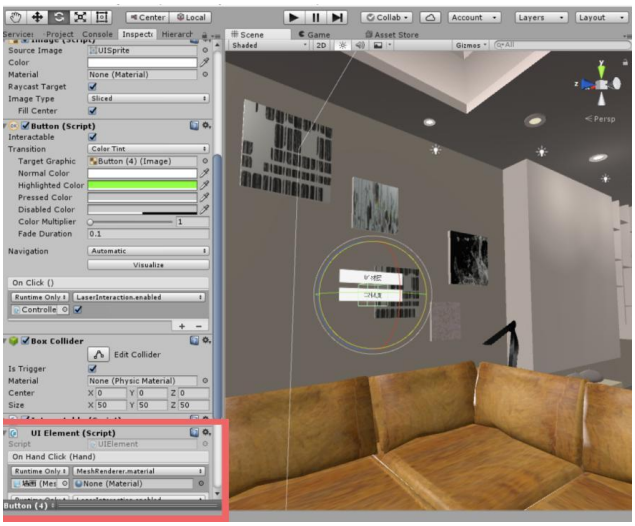


Fig. 25. Virtual reality interactive interface.

No.	Evaluation models	MAE	RMSE	MAPE	R2	Time/s
1	RNN	0.554	0.066	0.070	0.799	0.122
2	LSTM	0.480	0.045	0.042	0.846	0.089
3	GRU	0.443	0.058	0.017	0.956	0.085
4	SRU	0.257	0.044	0.066	0.881	0.036
5	RDA-SRU	0.133	0.02	0.015	0.968	0.028

Fig. 26. Comparison of performance metrics of the contrasting algorithms.

Fig. 27 gives the evaluation error of digital design application effectiveness of different comparison algorithms. From Fig. 27, it can be seen that the RDA-SRU model has the smallest RMSE value of 0.02 for evaluating the effectiveness of digital design application, followed by SRU, LSTM, GRU, and RNN.

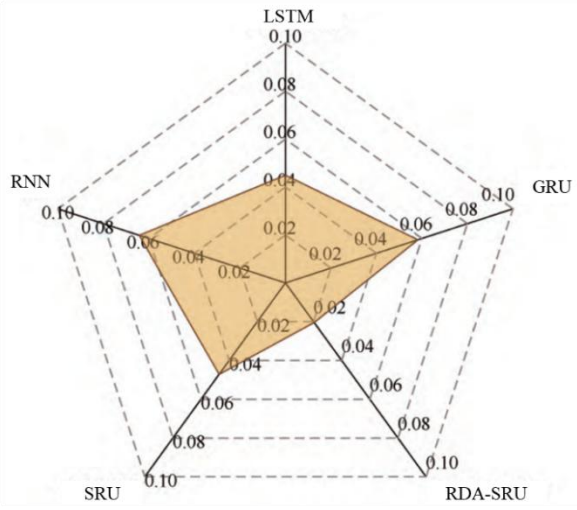


Fig. 27. Comparison of root mean square error of evaluation for contrasting algorithms.

Fig. 28 gives the convergence curve of the digital design evaluation of interior home products based on the RDA-SRU model. In Fig. 28, as the number of iterations increases, the value of the fitness function decreases until the RDA fitness value converges to about 0.184 at 50 iterations.

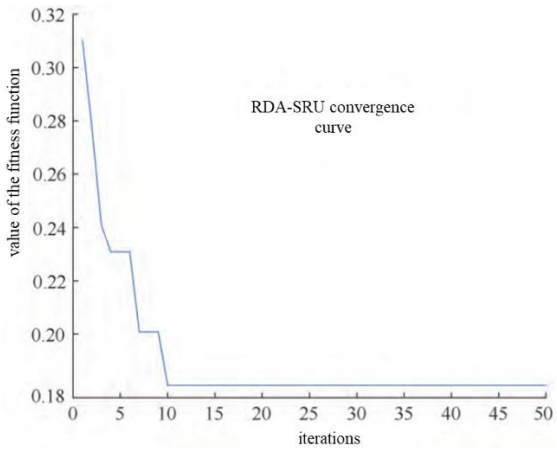


Fig. 28. Convergence curves based on RDA-SRU models.

V. CONCLUSION

In this paper, for predicting and evaluating the effect of digital design application combined with VR technology, firstly, according to the digitization process, we designed the evaluation framework of digital application effect, extracted the digital design evaluation factor set, and optimized the W , W_f , W_r , b_f , b_r of the Red Deer Optimization Algorithm degree SRU model to improve the performance of the digital design evaluation model; and then, through the experiment of digital design process of indoor home products, we chose to use the design process evaluation factor data set for training through pre-processing input into the RDA-SRU model. Then through the indoor home products digital design process experiment, the design evaluation factor dataset is selected to be input into the RDA-SRU model for training through preprocessing, and compared with the other four algorithms, and the results show that the RDA-SRU model proposed in this paper possesses a better accuracy, and good evaluation prediction results have been achieved.

ACKNOWLEDGMENT

This work is supported by 1) The 2022 Campus level Education Reform Project of Nanchang University of Technology: Research and Practice of Integrated Teaching Reform Based on the "SPOC+OBE" Concept - Taking the 3DsMAX Software Course as an Example. Grant No.: NGJG-2022-56; 2) Nanchang City Social Science Project: Research on Planning and Design of Industrial Heritage Protection and Adaptive Reuse in Nanchang City. Grant No.: FX202301; 3) Nanchang Institute of Technology 2022 School-level Humanities and Social Sciences Project: Research on the Sustainable Development of "Ancient Town" from the Perspective of Rural Revitalization-- A case study of Yangkou Town, a small town without heritage. Grant No.: NGRW-22-07.

REFERENCES

[1] Braa J, Sahay S, Monteiro E. Design Theory for Societal Digital Transformation: the Case of Digital Global Health. *Information Systems*, 2023, 24(6):1645-1669.

[2] He X, Zhang W, Feng P, Mai Z, Gong X, Zhang G. Role of Surface Coverage of Sessile Probiotics in Their Interplay with Pathogen Bacteria Investigated by Digital Holographic Microscopy. *Langmuir: The ACS Journal of Surfaces and Colloids*, 2023(48):39.

[3] Ujihashi Y, Watanabe S. Changes in Media Use in School Education and Home Learning Driven by the Progress of "GIGA School Concept". *The NHK Monthly Report on Broadcast Research*, 2022, 72(6):52-86.

[4] Baum M, Sui S, Malhotra S. A vicarious learning perspective on the relationship between home-peer performance and export intensity among SMEs. *International Marketing Review*, 2023, 40(2):197-223.

[5] Royer M. What Participation Creates in Experimental Design Practices. The Case of a Mobile Third Place Built in a Retirement Home. 2022, 2.

[6] Haque M, Indah D. Design of Digital Library Prototype Using The Design Thinking Method. *Jurnal Riset Informatika*, 2022.

[7] Basha-Jakupi A, Kuleta G, Navakazi V. The effect of architecture design as the non-digital component to the digital industry development -case study 'Brick Factory' in Prishtina, Kosovo. *City, Territory and Architecture*, 2022, 9(1).

[8] Sobhy M H M S H, Abouelnaga H M M. The role of digital graphic design in providing social communication solutions for the deaf and dumb. *International Journal of Design and Fashion Studies*, 2022.

[9] Wei D, Fan W, Liao Z. The future prospect of dynamic poster design in the context of new media. *Symposium on Creative Technology and Digital Media*, 2022.

[10] Ye W, Li Y. Digital Media Art Display Design and Research under the Research of 3D Point Cloud Data Acquisition Technology Based on Sequence Images. *Mobile Information Systems*, 2022, 2022(Pt.3):7106900.1-7106900.12.

[11] Hewitt J, Cluckie G, Smart A, Harris R, Reid J, Chandler C. Young Adults Rehabilitation Needs and Experiences following Stroke (YARNS): a review of digital accounts to inform the development of age-appropriate support and rehabilitation. *Journal of Advanced Nursing*, 2022, 78(3): 869-882.

[12] Rangaswamy E, Nawaz N, Changzhuang Z. The impact of digital technology on changing consumer behaviors with special reference to the home furnishing sector in Singapore. *Palgrave Communications*, 2022, 9.

[13] Morrison K, Hughes T, Doi L. Understanding the use of telehealth in the context of the Family Nurse Partnership and other early years home visiting programs: a rapid review. *Digital Health*, 2022, 8.

[14] Diefenbach S. Social norms in digital spaces: conflict reports and implications for technology design in the teleworking context. *Zeitschrift fur Arbeitswissenschaft*, 2022:1-22.

[15] Padmapriya V, Srivenkatesh M. Digital Twins for Smart Home Gadget Threat Prediction using Deep Convolution Neural Network. *International Journal of Advanced Computer Science and Applications*, 2023.

[16] Vaccaro R, Aglieri V, Rolandi E, Rossi M, Pettinato L, Ceretti A. The Remote Testing in Abbiategrasso (RTA) Study Protocol: a Counter-Balanced Crossover Trial to Assess the Feasibility of Direct-to-Home-Neuropsychology with Older People. *Health (English)*, 2022(005):014.

[17] Fathollahi-Fard A M, Hajiaghaei-Keshteli M, Tavakkoli-Moghaddam R. Red deer algorithm (RDA): a new nature-inspired meta-heuristic. *Soft Computing*, 2020(2): 1-29.

[18] Zhao J.N., She Q.S., Meng M., Chen Y. Skeleton action recognition based on multi-stream spatial attention graph convolutional SRU network. *Journal of Electronics*, 2022, 50(7):1579-1585.

[19] Yu X, Wang W. Personnel evacuation decision of underground complex based on VR test. *Safety and Environmental Engineering*, 2024, 31(04):125-132.

[20] Wang D. From immersion to multidimensional - A study of multimodal digital pavilion integrating VR, AR and AI technologies. *Art Education Research*, 2024, (13):83-85.

[21] Wu W, Ye Z. Digital transformation of home furnishing enterprises and the development trend of product digital design. *Wood Science and Technology*, 2023, 37(03):1-11.

[22] Sun D, Chen Y, Lu Z. Automation optimization technology in the process of transferring digital design results. *Power and Energy*, 2021, 42(04):488-491.

[23] Xie K, Yang Z. Ship power data preprocessing technology based on Spark framework. *Ship Electricity Technology*, 2024, 44(07):69-72.

[24] Zhang R, Huang W, Ma T. Construction of evaluation indexes of civil airport intelligence degree. *Journal of Southeast University (Natural Science Edition)*, 2024, 54(03):524-530.

[25] Hu F, Wu Y.R., Dong F.M., Zou Y.B., Sun S. High-speed simple cyclic unit networks. *Control and Decision Making*, 2022, 37(2):6.

[26] Yue Z.B., Lu J.B., Wan L. SRU model radar HRRP target recognition based on attention mechanism. *Ship Electronic Engineering*, 2023, 43(4):44-48.

[27] Gulec O, Sahin E. Red Deer Algorithm based nano-sensor node clustering for IoNT. *Journal of Network and Computer Applications*, 2023.

Convolutional Layer-Based Feature Extraction in an Ensemble Machine Learning Model for Breast Cancer Classification

Shofwatul 'Uyun[1], Lina Choridah[2], Slamet Riyadi[3], Ade Umar Ramadhan[4]
Department of Information-Faculty of Science and Technology, UIN Sunan Kalijaga, Yogyakarta, Indonesia[1, 4]
Department of Radiology-Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Indonesia[2]
Department of Information Technology-Faculty of Engineering, Universitas Muhammadiyah Yogyakarta, Indonesia[3]

***Abstract*—Mammography and ultrasound are the main medical imaging modalities for identifying breast lesions. Computer-assisted diagnosis (CAD) is an important tool for radiologists, helping them differentiate benign and malignant lesions more quickly and objectively. The use of appropriate features in mammography and ultrasound is one of the key factors determining the success of computer-assisted diagnosis (CAD) results for breast cancer systems. The diversity of feature forms and extraction techniques is a challenge. Additionally, the use of a single classification algorithm often causes noise, bias, and is not robust. We propose a convolutional layer-based feature extraction technique in the ensemble learning model for the classification of breast cancer. This study uses 439 mammography images (203 benign, 236 malignant) and 421 ultrasound images (244 benign, 177 malignant). This research consists of several stages, including data pre-processing, feature extraction, classification, and performance evaluation. We used four convolution layer-based feature extraction techniques: simple convolution (SC), feature fusion convolution (FFC), feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC). The model uses five machine learning algorithms (support vector machine, random forest, k nearest neighbours, decision tree, and logistic regression) that are part of ensemble learning. The experimental results show that the use of the FFC convolution layer in ensemble learning has the best performance for both datasets. In the ultrasound data set, the FFC achieved a value of 0.90 in each of the accuracy, precision, recall, specificity, and F1 score metrics. In the mammography data set, the FFC achieved a value of 0.98 on each of the same metrics. These results show the effectiveness of feature fusion in improving classification performance in the soft voting classifier for ensemble learning.**

***Keywords*—Ensemble learning; feature extraction; convolutional layer; breast cancer**

I. INTRODUCTION

Breast cancer is the main cause of cancer-related mortality in women. Early detection of cancer, especially breast cancer, will contribute to the treatment process [1]. Currently, computerised tomography (CT), magnetic resonance imaging (MRI), mammography, thermal imaging, and ultrasound are common screening for breast cancer. These methods have their own unique approaches and tools, and the expected results of these methods depend on different factors, so it is recommended to validate the results using multiple methods. Although mammography is considered by many physicians and specialists the gold standard method for the detection of breast cancer, the demand for more reliable methods is increasing. Mammography and ultrasound are critical tools in breast cancer screening, but they have different applications, effectiveness, and limitations. Mammography, a form of X-ray imaging, is considered the gold standard for breast cancer screening in high-income countries because of its ability to detect cancer at an early stage. However, its effectiveness may be limited by the radiographic density of the breasts; In dense breasts, noncalcified cancers are more likely to be missed [2]. This limitation is particularly significant because the diagnostic accuracy is largely dependent on breast density and denser breasts, which pose a challenge for clear imaging [3]. Ultrasound, on the other hand, uses sound waves to create images of breast tissue [4]. It has been shown to have a high detection sensitivity, especially in younger women and those with dense breasts, where mammography may not be as effective.

Computer-assisted diagnosis (CAD) systems have shown significant achievements in improving breast cancer detection and provide complementary tools to traditional diagnostic methods [5]. The effectiveness of the CAD system is emphasized by its ability to detect and classify breast cancer with high precision, sensitivity, and specificity, as demonstrated in various studies. Recent research has shown that the precision of breast cancer detection is greatly influenced by the characteristics of mammography and ultrasound. Although mammography is widely used, its sensitivity is limited, especially in dense breast tissue, which can conceal the presence of tumours [6]. Contrast-enhanced mammography (CEM) is introduced to address some of these limitations, introducing a similar performance in the detection of mammography-occult diseases with higher sensitivity than conventional mammography and higher specificity than ultrasound [7].

The development and application of computer-aided diagnosis (CAD) systems in the detection and classification of breast cancer by mammography and ultrasound have significantly benefitted from various feature extraction methods. In mammography, several feature extraction techniques have been used in different studies. For example, an approach involves the extraction of 16 geometrical

characteristics from regions of interest (ROI) in mammograms, which are then analysed using machine learning algorithms to classify mammograms into four classes [8]. Another mammography method was to calculate 271 characteristics in various categories, including shapes, textures, contrasts, and other characteristics, and to calculate additional characteristics derived from the dilated segment [9]. Some methods have been used in ultrasound. A study extracted 855 characteristics, including shapes, contours, and texture, from breast ultrasound images [10]. Another method uses the pyramid of orientated gradient descriptor histograms to obtain a characteristic vector without prior processing of tumour region selection [11]. Automated contouring and morphological analysis were used to calculate 19 practical morphological characteristics [12]. The various forms of features used affect the performance of a classification model.

Some cases that often occur in the learning process using a single classification algorithm are noise, bias, and low accuracy. This is due to the use of non-uniform data samples and the presence of overlapping classes [13]. One method of reducing these issues is to implement the concept of ensemble learning, which is designed to improve the stability and accuracy of machine learning algorithms. The concept of ensemble learning is a paradigm of learning that uses a combination of several models synergistically to improve the quality of predictions collectively [14], [15]. The objective of Ensemble learning is to reduce the bias, variance, or error that often occurs in individual models [16], [17], [18]. The main goal is to achieve predictive results that are more accurate, stable, and generally better than the results of a single model.

Furthermore, one of the keys to successful learning is the use of the right features. Feature extraction techniques have evolved [19]. In general, features consist of colour, shape, texture, and others. Some previous research related to medical imaging, especially breast cancer, uses features of colour and texture [20]. Deep feature extraction uses pre-trained convolutional neural networks such as VGG-19, SqueezeNet, ResNet-18, and GoogLeNet to distinguish between benign and malignant tumour types in ultrasound images [21],[22] Therefore, it is necessary to experiment with the use of several image feature extraction schemes using convolutional layers.

We propose the use of an ensemble learning concept for breast cancer classification using two medical images, namely mammography and ultrasound. This ensemble learning model uses five machine learning algorithms, including a supervised vector machine, random forest, nearest neighbours k, decision tree, and logistic regression. The ensemble learning model was trained using mammography and ultrasound images that have been feature extracted using four different convolution layers, including simple convolution, feature fusion convolution, feature fusion depthwise convolution, and feature fusion depthwise separable convolution. The four model schemes were evaluated for accuracy, precision, recall, ROC curve, specificity, F1 score, kappa and Matthews correlation coefficient. Some of the main contributions of this research are summarised in the following.

- We propose a novel ensemble learning model for breast cancer classification using mammography and ultrasound.
- The feature extraction technique uses four different convolution layers, namely, simple convolution (SC), feature fusion convolution (FFC), feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC).
- The ensemble learning model uses soft voting with five machine learning algorithms, namely the support vector machine, the random forest, the k closest neighbours, the decision tree, and logistic regression.
- We investigate the use of different feature extraction techniques in the ensemble learning model by evaluating its performance.

The structure of this paper consists of several sections; after the introduction in Section I, Section II describes the proposed method, for Sections III and IV explain the results and discussion on the use of the proposed method. Section V explains the conclusions and future research and ends with acknowledgments.

II. METHODS

A. Datasets

This study used primary data from Sardjito Hospital and Kotabaru Yogyakarta Cancer Clinic that have been identified and diagnosed by radiologists. The approval for the private data set was obtained from the Ethics Committee (Ref. No.: KE/FK/1229/EC/2023). The results of the image reading were initially classified into five class categories in BIRADS. The researchers grouped the data from the BIRADS standard into two classes, namely benign and malignant. BIRADS 2 and BIRADS 3 are benign classes, while BIRADS 4 and BIRADS 5 are malignant classes. Details of the amount of data in each benign and malignant class for both types of images can be seen in Table I.

TABLE I. THE INITIAL DATA SET (BEFORE AUGMENTATION)

Dataset	Class		Total
	Malignant	***Benign***	
Ultrasound	177	244	421
Mammography	236	203	439

B. Proposed Method

We propose an ensemble learning model using two different modes of mammography and ultrasound data with four convolution layer schemes for feature extraction. In the classification process, five machine learning algorithms (support vector machine, random forest, KNN, decision tree, and logistic regression) are run together in one iteration by voting. In general, we propose a model that consists of four main processes: data preprocessing, feature extraction, classification, and performance evaluation. The use of the model aims to compare the use of four convolutional layer schemes as feature extraction using voting on the determination of classification results. Details can be seen in Fig. 1.

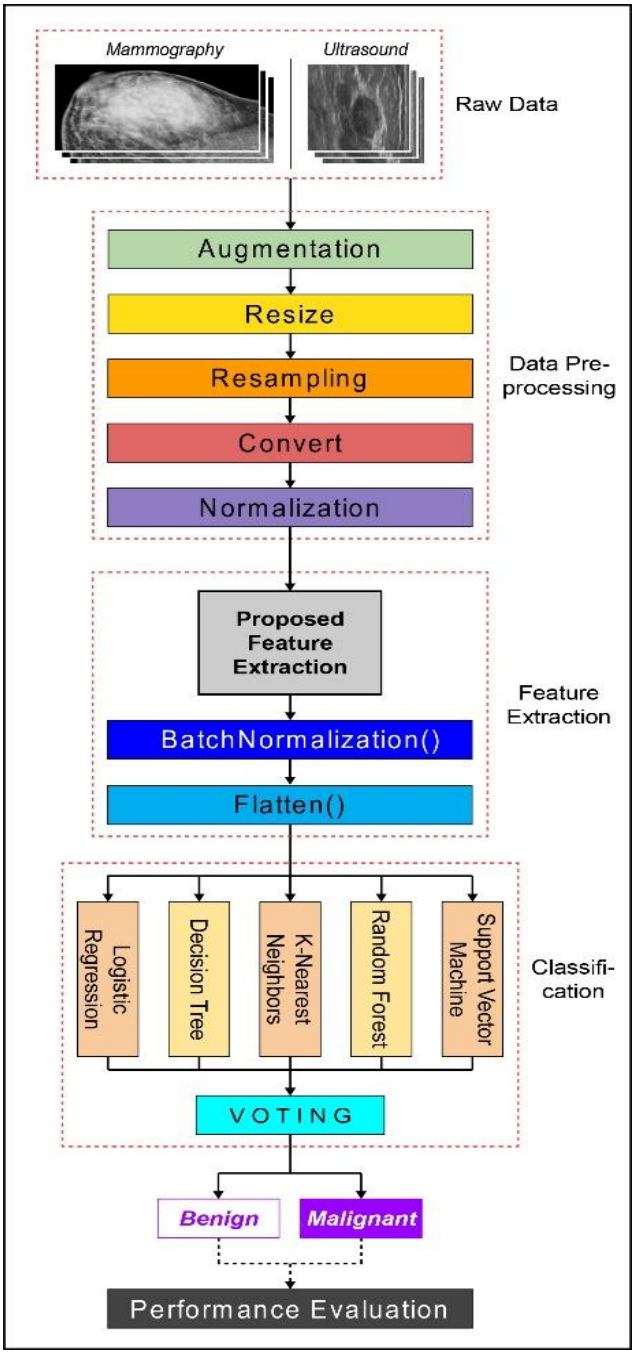


Fig. 1. Proposed ensemble learning model with four different feature extraction schemes.

Details of the convolution layers for feature extraction for the four schemes are shown in Fig. 2. The first scheme uses four convolution layers. The second scheme uses feature-fusion convolution in feature extraction with the same layer configuration as in the first scheme. Feature fusion is a technique that combines predictions from multiple machine learning models to create a combination of more than one feature that is used to generate the final prediction [19], [20]. The third scheme uses a depth-wise convolution of the feature fusion, and the last scheme uses a separable depth-wise feature fusion convolution.

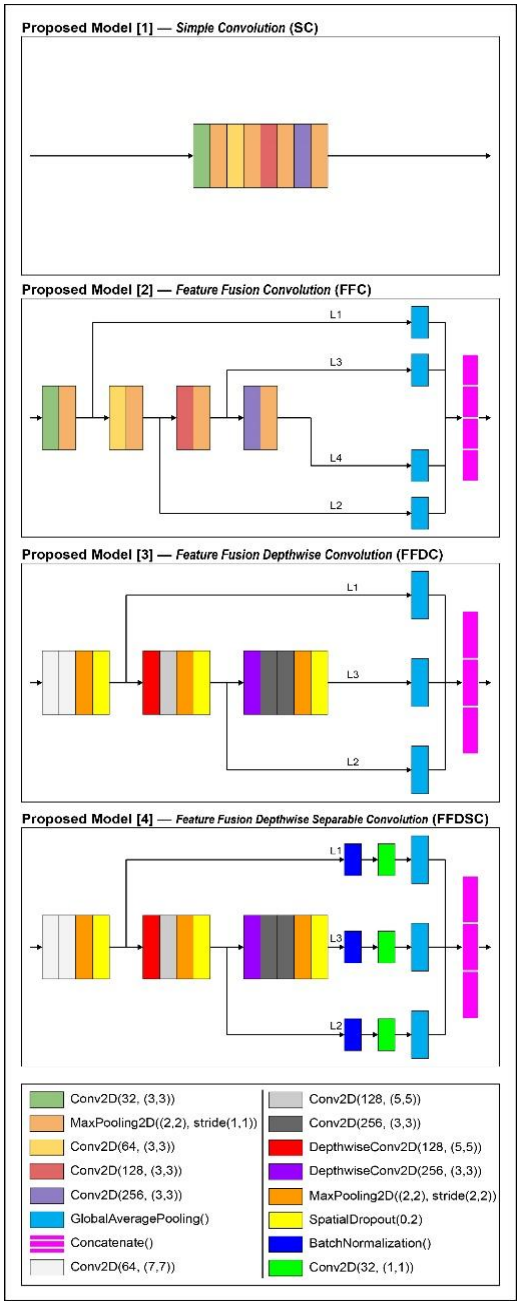


Fig. 2. Convolution layer scheme as feature extraction.

1) Pre-processing: This preprocessing stage consists of five image improvement techniques before feature extraction, including: augmentation, resize, resampling, convert, and normalisation. Some augmentation techniques in medical images that have been shown to improve classification accuracy are translation, shear, and rotation [5]. Translation is done by shifting the image by a maximum of 15 pixels either towards the positive or negative poles of the x-axis and the y-axis with a probability of 1. Shear is done by shearing the image by a maximum of 15 degrees toward the positive or negative poles of the x-axis and y-axis with a probability of 1. Rotation is done by rotating the image by a maximum of 25 degrees clockwise or counterclockwise. After image

enhancement, resize the ultrasound image to 224×224 pixels and the mammography image to 448 x 224 pixels so that the entire data set has a uniform resolution. The total data amounted to 1,000 images for each modal image with various predefined parameters. After that, the data are converted into a numpy array to convert the colour data in the image into numerical data using the numpy library. Labels or classes are extracted using numpy by taking the folder name used in each class. The purpose of converting data into numeric form is to allow the model to recognise the data to be trained. The last step in preprocessing is image normalisation using a rescaling technique with a normalisation factor of 1/255 to change the range from 1-255 to 0-1. The preprocessing data are shown in Table II.

TABLE II. DATASET AFTER IMAGE PRE-PROCESSING

Dataset	Class		Size [Height×Width] (pixel)	Total
	Malignant	*Benign.Benign.*		
Ultrasound	500	500	224 × 224	1.000
Mammography	500	500	448 × 224	1.000

2) Feature extraction: This research uses four different CNN architecture schemes in the convolution layer in the feature extraction stage, namely: simple convolution (SC), feature fusion convolution (FFC) [23], feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC). The first scheme with simple convolution consists of four layers of convolution or Conv2D and 32 filters in the first layer, 64 filters in the second layer, 128 filters in the third layer, and 256 filters in the fourth layer. The kernel size used in the convolution layer is 3x3. The pooling process in each layer uses a kernel pool of the maximum value of MaxPooling2D pixels with a size of 2x2. Each convolution layer, Conv2D, uses the Rectified Linear Activation (ReLU) activation function and the 'same' value in the padding parameter so that the output image is the same size and does not cut at the edges. The pooling layer, MaxPooling2D, also uses the 'same' value for the padding parameter.

The second scheme with FFC, consists of four convolution blocks that are run cumulatively. The first block consists of a Conv2D layer with 32 filters and a kernel size of 3×3. Then a Maxpooling2D layer with a kernel size of 2×2 and standard strides of 1×1 pixels. The second block consists of the layers of the first block plus Conv2D with filter 64 and kernel size 3×3. Then a Maxpooling2D layer with a kernel size of 2×2 and standard strides of 1×1 pixels. The third block consists of the layers of the second block plus Conv2D with filter 128 and kernel size 3×3. Then the Maxpooling2D layer with a kernel size of 2×2 and standard strides of 1×1 pixels. The fourth block consists of the layers of the third block plus Conv2D with filter 256 and kernel size 3×3. Then a Maxpooling2D layer with a kernel size of 2×2 and standard strides of 1×1 pixels. Each Conv2D convolution layer uses the Rectified Linear Activation (ReLU) activation function and the 'same' value in the padding parameter so that the output image is the same size and not cut at the edges. The pooling layer, MaxPooling2D, also uses the 'same' value for the padding parameter.

The third scheme, FFDC, consists of three convolution blocks that are run cumulatively. The first block consists of two Conv2D layers with 64 filters and a kernel size of 7×7. Then a Maxpooling2D layer with a kernel size of 2×2 and strides of 2×2 pixels. Next, we have the SpatialDropout2D layer with a probability of 0.2. The second block consists of the layers of the first block plus DepthwiseConv2D with a kernel size of 3×3. Then two Conv2D layers with 256 filters and 3×3 kernel size. Next, we have a Maxpooling2D layer with a kernel size of 2×2 and a shift of 2×2 pixels. Then comes the SpatialDropout2D layer with a probability of 0.2. The third block consists of the layers of the second block plus DepthwiseConv2D with a kernel size of 5×5. Then Conv2D with filter 128 and kernel size 5×5. Next, we have the Maxpooling2D layer with a kernel size of 2×2 and a shift of 2×2 pixels. Subsequently, the SpatialDropout2D layer was populated with a probability of 0.2. Each Conv2D and Depthwise2D uses the Rectified Linear Activation (ReLU) activation function and the 'same' value in the padding parameter so that the output image is the same size and does not cut the edges. In layer-pooling, MaxPooling2D also uses the 'same' value for the padding parameter.

The fourth scheme with FFDSC has the same architecture as FFDC. The difference between the two types of architecture lies at the end of the process of each block; for the FFDSC architecture, a batch normalisation layer is added, then Conv2D with a 32 filter and a kernel size of 1×1 which is often referred to as Pointwise Convolution. In the first scheme, after passing through the data convolution layer, a batch normalisation layer is added to overcome problems caused by changes in input distribution that occur during the training process.

The results of the training process are entered into the flatten layer to change the data dimension from three dimensions to one dimension to facilitate the classification process. However, in the second, third, and fourth schemes before the BatchNormalization and Flatten layers, the GlobalAveragePooling2D (GAP) layer is added with the aim of reducing dimensions, accelerating computation, and reducing overfitting. After passing through the convolution layer in each block, the data are merged in the concatenate layer. The next stage is the process of compiling the model with Adam parameters in the optimiser and binary cross-entropy in loss.

3) Classification: In this research, the ensemble learning used is voting on the classifier. Voting is a technique in which various predictive models vote or weigh their predictions, and the final result is taken based on the majority of these votes or weights [24]. The classification process using five algorithms, namely: SVM [25], Random Forest [26], KNN, Decision Tree, and Logistic Regression are used as voting-based classifiers in the designed model. The hyperparameters used for each classifier are the result of a series of experiments that have been carried out to optimise model performance. The SVM

used is the SupportVectorClassifier with a linear kernel and a penalty value of 3.0. Random Forest uses an estimator of 100. KNN uses a neighbour count value of 3, uses distance as a weight assessment scheme, and cosine as a distance metric value. Decision Tree uses the entropy value as a criterion to measure split quality. Logistic regression uses L1 as penalty value, liblinear as optimisation algorithm, and a maximum iteration value of 750. In voting classifiers, SVM, regression are trained together in one iteration. The "soft" parameter is used in voting with the aim that the classification results used later are based on the average probability of each classifier, not on the number of dominant classifiers.

4) Perfomance evaluation: We used 5-fold cross-validation for training and model evaluation to calculate the average over five iterations. Some common indicators to determine classification system performance include true positive (TP), true negative (TN), false positive (FP), and false negative (FN). These four basic indicators can be used to determine eight other metrics such as accuracy, precision (TPR), recall, ROC curve, specivicity, F1 score, kappa and matthews correlation coefficient (MCC), which are defined in Eq. (1)-Eq. (8):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP} * 100\% \quad (2)$$

$$Recall = \frac{TP}{TP+FN} * 100\% \quad (3)$$

$$ROC\ Curve\ (FPR) = \frac{FP}{(FP+TN)} \quad (4)$$

$$Specificity = \frac{TN}{(TN+FP)} \quad (5)$$

$$\text{F1 Score} = \frac{2x(\text{PrecisionxRecall})}{(\text{Precision}+\text{Recall})} \quad (6)$$

$$Kappa = \frac{p_o - p_e}{1-p_e} \quad (7)$$

$$Kappa = \frac{TN\ x\ TN - FP\ x\ FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (8)$$

III. RESULTS

Ultrasound and mammography data were trained separately using a cross-validation fold selection model, KFold, with five data folds. Training data are evaluated by testing the label of the predicted results against the label of the actual data using a confusion matrix. The confusion matrix results for both modal data show that the model has fairly accurate results characterised by the match between the predicted data label and the actual data label having a very dominant data membership in each class. The validation process of the training data is also analysed using several metrics such as accuracy, precision, recall, RoC curve, specificity, F1 score, Cohen's Kappa (Kappa) and Matthews correlation coefficient (MCC). Information about the data metric: The values of the training validation results can be seen in Table III.

Both graphs in Fig. 3 show the accuracy graph using Eq. (1). Both graphs compare the performance of the four feature extraction schemes. The FFC scheme consistently has the most superior feature extraction capability in ultrasound and mammography images. This is reflected in the higher average validated accuracy compared to the other feature extraction schemes, which is 89.90% on the ultrasound image and 98.20% on the mammography image. Meanwhile, Fig. 4 shows that the FFC scheme has the lowest loss among others for mammography and ultrasound, 3.6404 and 0.6488, respectively.

The Receiver Operating Characteristics (ROC) curve serves to evaluate the performance of binary classification models with a focus on the trade-off between sensitivity (recall) and specificity. Both graphs in Fig. 5 show that the FFC scheme performs better than the other three schemes with AUC values of 0.90 and 0.98 on ultrasound and mammography images, respectively. The precision recall curve serves to evaluate the performance of binary classification models, especially in unbalanced data. Fig. 6. shows the results of using four feature extraction schemes for both images. Similarly to the RoC curve, the AUC value also shows that the FFC scheme has the best feature extraction capability of 0.93 and 0.98 in ultrasound and mammography images. The results of the comparison of feature extraction techniques with four different schemes show that scheme 2 (FFC) has the best performance in ensemble models using five machine learning classification algorithms. SVM, Random Forest, KNN, Decision Tree, and Logistic Regression. The feature extraction process with scheme 4 (FFDSC) is designed to reduce the number of parameters and increase the efficiency of deep learning so that it is not optimally applied to classification algorithms using conventional machine learning (non-deep learning).

TABLE III. THE RESULTS OF THE MEAN METRICS FOR 5-FOLD OF EACH MODAL

Dataset	Proposed Feature Extraction	Accuracy (%)	Precision (%)	Recall (%)	ROC curve (%)	Specificity (%)	F1 score (%)	Kappa (%)	MCC (%)
Ultrasound	Simple Convolution	0.8240	0.8241	0.8240	0.1760	0.8240	0.8240	0.6480	0.6481
	FF Convolution	0.8990	0.8995	0.8990	0.1010	0.8990	0.8990	0.7980	0.7985
	FF Depthwise Convolution	0.8930	0.8930	0.8930	0.1070	0.8930	0.8930	0.7860	0.7860
	FF depthwise separable convolution	0.8820	0.8822	0.8820	0.1180	0.8820	0.8820	0.7640	0.7642
Mammography	Simple Convolution	0.9220	0.9222	0.9220	0.0780	0.9220	0.9220	0.8440	0.8442
	FF convolutionconvolution	0.9820	0.9821	0.9820	0.0180	0.9820	0.9820	0.9640	0.9641
	FF Depth-wiseDepth-wise Convolution	0.9630	0.9634	0.9630	0.0370	0.9630	0.9630	0.9260	0.9264
	FF Separable Convolutio in Depth	0.9500	0.9507	0.9500	0.0500	0.9500	0.9500	0.9000	0.9007

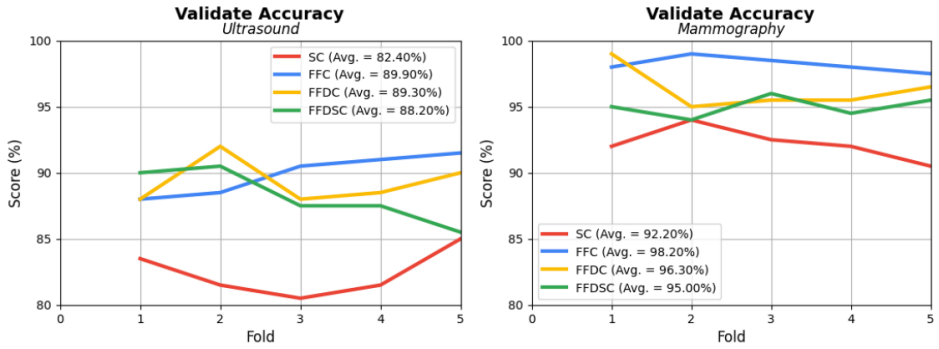


Fig. 3. Accuracy graph in ultrasound and mammography.

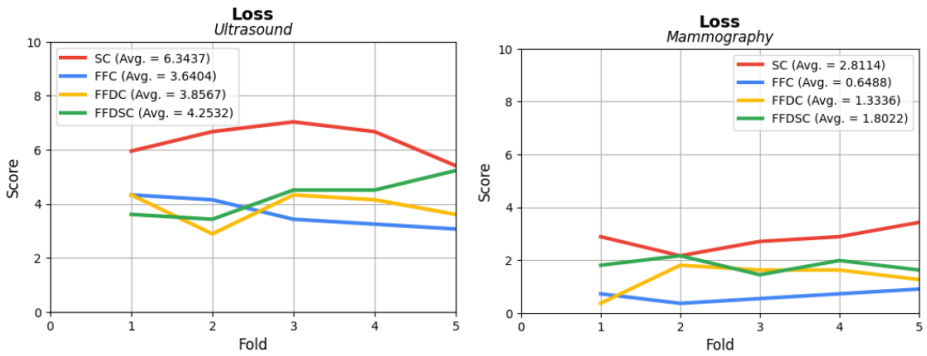


Fig. 4. Graph of loss score on ultrasound and mammography images.

IV. Discussion

Several studies focus on the classification of breast cancer using several types of medical imaging. The limited number of data sets in medical images, especially hispatological images, is a challenge for deep learning techniques for feature extraction. The use of deep convolutional neural networks (DCNN) and SVM techniques for classification [27]Techniques to overcome the imbalance in the amount of data in a particular class are a challenge in itself, and there are several ways to overcome these limitations, one of which is the use of image enhancement techniques. Image augmentation techniques have been shown to improve classification system performance compared to the original dataset [28]. It has also performed feature extraction using deep learning with several types of CNN architectures (ResNet-18, ResNet-34, ResNet-50, and EfficientNet-B0). The study used mammography images and the classification process based on multiple instances learning [29] In contrast to what was done by [30] with other types of images, namely thermal imaging with classification based on convolutional neural networks. The collaborative use of two image models has also been done by [31] using hispatology and ultrasound images based on transfer learning.

This research has conducted eight different scenarios with four different feature extraction techniques. The test results show that the use of the feature extraction technique of the feature fusion convolution type has been shown to give the best performance compared to the other three feature extraction techniques. For the classification stage, this research uses five machine learning algorithms that are combined into one unit in an ensemble machine learning system with a soft voting classifier. Table IV compares the results of the proposed method with those of other methods from previous studies for the classification of breast cancer.

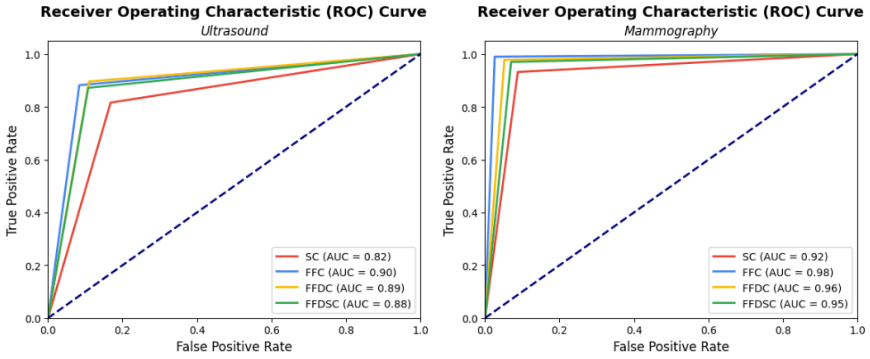


Fig. 5. Receiver Operating Characteristic (ROC) curve in ultrasound and mammography images.

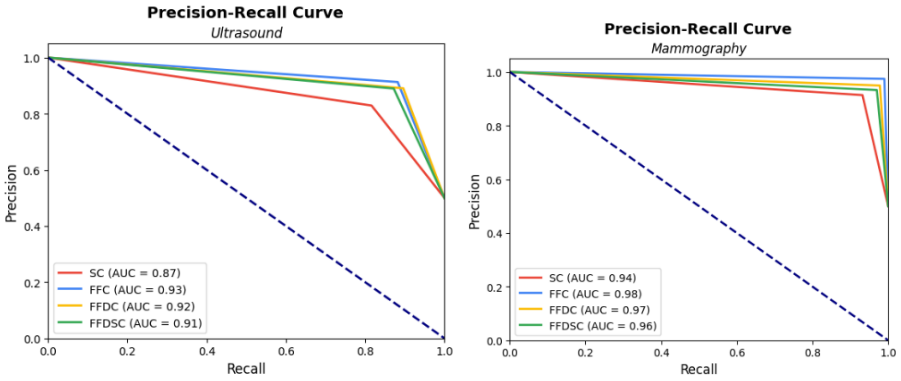


Fig. 6. Precision recall curve on ultrasound and mammography images.

TABLE IV. COMPARISON OF THE PERFORMANCE OF PROPOSED METHOD FOR THE CLASSIFICATION OF BREAST CANCER

Reference	Feature Extraction	Classification	Image	Parameter			
				Acc (%)	Loss	RoC	Precision-Recall
Kulkarni & Rabidas [28]	U-Net model		Mammography	96.81	-	-	-
Hassan et al. [26]	Deep convolutional neural networks (DCNNs)	SVM	Hispatology	99.24	-	-	-
Bobowicz et al.[28]	ResNet-18, ResNet-34, ResNet-50 and EfficientNet-B0	Multiple instance learning-based	Mammography	81.6	-	0.90	-
Roslidar et al. [30]	Convolutional neural networks		Thermal Imaging	-	-	-	-
Aaroj et al. [[30]	Transfer Learning		Hispatology and ultrasound	99.35	-	-	-
Proposed	Simple convolution (SC), feature fusion convolution (FFC), feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC).	Ensemble machine learning	Mammography	98.20	0.6488	0.98	0.98
			Ultrasound	89.90	3.6404	0.90	0.93

V. CONCLUSIONS AND FUTURE WORK

This study has compared ensemble learning models with four convolution layer schemes for feature extraction in breast cancer classification using ultrasound and mammography images separately. The four convolution layer schemes used are simple convolution (SC), feature fusion convolution (FFC), feature fusion depthwise convolution (FFDC), and feature fusion depthwise separable convolution (FFDSC). Classification is performed using an ensemble learning soft voting classifier with SVM, Random Forest, KNN, Decision Tree, and Logistic Regression algorithms. The experimental results show that the FFC convolutional layer scheme achieves the best performance for both datasets. In the ultrasound data set, FFC achieved a value of 0.90 in each of the Val-Acc, TPR, Recall, TNR and F1-Score metrics. In the mammography data set, the FFC reached a value of 0.98 on each of the same metrics. These results emphasize the effectiveness of feature fusion in improving classification performance. Future research can focus on exploring more complex fusion techniques, such as using multimodal data or combining classification with deep learning.

ACKNOWLEDGMENT

The authors gratefully acknowledge the sponsorship of the institute for research and community service UIN Sunan Kalijaga Yogyakarta based on SK Rector No: B-1913/Un.02/L3/TL/05/2024.

REFERENCES

[1] C. Y. Chou, T. T. Shen, W. C. Wang, and M. P. Wu, “Favorable breast cancer mortality-to-incidence ratios of countries with good human development index rankings and high health expenditures,” Taiwan J Obstet Gynecol, vol. 63, no. 4, pp. 527–531, Jul. 2024, doi: 10.1016/j.tjog.2023.11.012.

[2] A. Glechner et al., “Mammography in combination with breast ultrasonography versus mammography for breast cancer screening in women at average risk,” Cochrane Database of Systematic Reviews, vol. 2023, no. 3, Mar. 2023, doi: 10.1002/14651858.CD009632.pub3.

[3] S. Aslani et al., “Enhancing cancer prediction in challenging screen-detected incident lung nodules using time-series deep learning,” Computerized Medical Imaging and Graphics, vol. 116, Sep. 2024, doi: 10.1016/j.compmedimag.2024.102399.

[4] L. Xiaoming et al., “Application of combined preoperative indocyanine green lymphography and ultrasonography for low-pressure vein localization in secondary lymphedema surgery for breast cancer,” Asian J Surg, vol. 47, no. 1, pp. 289–295, Jan. 2024, doi: 10.1016/j.asjsur.2023.08.121.

[5] M. Hamiane and F. Saeed, “SVM classification of MRI brain images for computer-assisted diagnosis,” International Journal of Electrical and Computer Engineering, vol. 7, no. 5, pp. 2555–2564, 2017, doi: 10.11591/ijece.v7i1.pp2555-2564.

[6] J. Liu et al., “An overview of artificial intelligence in medical physics and radiation oncology,” Journal of the National Cancer Center, vol. 3, no. 3, pp. 211–221, Sep. 2023, doi: 10.1016/j.jncc.2023.08.002.

[7] A. Grażyńska et al., “BIRADS 4 – Is it possible to downgrade lesions that do not enhance on recombinant contrast-enhanced mammography images?,” Eur J Radiol, vol. 167, Oct. 2023, doi: 10.1016/j.ejrad.2023.111062.

[8] N. M. Hassan, S. Hamad, and K. Mahar, “Mammogram breast cancer CAD systems for mass detection and classification: a review,” Multimed

Tools Appl, vol. 81, no. 14, pp. 20043–20075, Jun. 2022, doi: 10.1007/s11042-022-12332-1.

[9] R. Song, T. Li, and Y. Wang, "Mammographic Classification Based on XGBoost and DCNN with Multi Features," IEEE Access, vol. 8, pp. 75011–75021, 2020, doi: 10.1109/ACCESS.2020.2986546.

[10] W. C. Shia, L. S. Lin, and D. R. Chen, "Classification of malignant tumours in breast ultrasound using unsupervised machine learning approaches," Sci Rep, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-81008-x.

[11] Y. L. Huang, D. R. Chen, Y. R. Jiang, S. J. Kuo, H. K. Wu, and W. K. Moon, "Computer-aided diagnosis using morphological features for classifying breast lesions on ultrasound," Ultrasound in Obstetrics and Gynecology, vol. 32, no. 4, pp. 565–572, Sep. 2008, doi: 10.1002/uog.5205.

[12] S. Beneddine, "Nonlinear input feature reduction for data-based physical modeling," J Comput Phys, 2023, [Online]. Available: https://www.elsevier.com/open-access/userlicense/1.0/

[13] L. Liu, W. Gao, H. Yu, and D. E. Keyes, "Overlapping multiplicative Schwarz preconditioning for linear and nonlinear systems," J Comput Phys, vol. 496, Jan. 2024, doi: 10.1016/j.jcp.2023.112548.

[14] A. Sharma, D. Goyal, and R. Mohana, "An ensemble learning-based framework for breast cancer prediction," Decision Analytics Journal, vol. 10, Mar. 2024, doi: 10.1016/j.dajour.2023.100372.

[15] L. Alzubaidi et al., "MEFF – A model ensemble feature fusion approach for tackling adversarial attacks in medical imaging," Intelligent Systems with Applications, vol. 22, Jun. 2024, doi: 10.1016/j.iswa.2024.200355.

[16] K. Tembhare, T. Sharma, S. M. Kasibhatla, A. Achalere, and R. Joshi, "Multi-ensemble machine learning framework for omics data integration: A case study using breast cancer samples," Inform Med Unlocked, vol. 47, Jan. 2024, doi: 10.1016/j.imu.2024.101507.

[17] I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," 2022, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2022.3207287.

[18] M. H. A. M. H. Himel, P. Chowdhury, and M. A. M. Hasan, "A robust encoder decoder based weighted segmentation and dual staged feature fusion based meta classification for breast cancer utilizing ultrasound imaging," Intelligent Systems with Applications, vol. 22, Jun. 2024, doi: 10.1016/j.iswa.2024.200367.

[19] R. Massafra et al., "Radiomic feature reduction approach to predict breast cancer by contrast-enhanced spectral mammography images," Diagnostics, vol. 11, no. 4, Apr. 2021, doi: 10.3390/diagnostics11040684.

[20] S. Uyun and L. Choridah, "Feature selection mammogram based on breast cancer mining," International Journal of Electrical and Computer Engineering, vol. 8, no. 1, pp. 60–69, Feb. 2018, doi: 10.11591/ijece.v8i1.pp60-69.

[21] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros, and T. Ganslandt, "Transfer learning for medical image classification: a literature review," Dec. 01, 2022, BioMed Central Ltd. doi: 10.1186/s12880-022-00793-7.

[22] M. Ilyas et al., "Deep Learning based Classification of Thyroid Cancer using Different Medical Imaging Modalities: A Systematic Review," VFAST Transactions on Software Engineering, vol. 9, no. 4, pp. 1–17, 2021.

[23] I. U. Haq, H. Ali, H. Y. Wang, C. Lei, and H. Ali, "Feature fusion and Ensemble learning-based CNN model for mammographic image classification," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 6, pp. 3310–3318, Jun. 2022, doi: 10.1016/j.jksuci.2022.03.023.

[24] G. Żabiński, J. Gramacki, A. Gramacki, E. Mińta-Jakubowska, T. Birch, and A. Disser, "Multi-classifier majority voting analyses in provenance studies on iron artefacts," J Archaeol Sci, vol. 113, Jan. 2020, doi: 10.1016/j.jas.2019.105055.

[25] V. Jain, A. Jain, A. Chauhan, S. S. Kotla, and A. Gautam, "American Sign Language recognition using Support Vector Machine and Convolutional Neural Network," International Journal of Information Technology (Singapore), vol. 13, no. 3, pp. 1193–1200, Jun. 2021, doi: 10.1007/s41870-021-00617-x.

[26] Dhiyaussalam and S. 'Uyun, "Optimization of Random Forest Hyperparameters with Genetic Algorithm in Classification of Lung Cancer," in 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2023, pp. 1–7.

[27] A. M. Hassan, A. Yahya, and A. Aboshosha, "A framework for classifying breast cancer based on deep features integration and selection," Neural Comput Appl, vol. 35, no. 16, pp. 12089–12097, Jun. 2023, doi: 10.1007/s00521-023-08341-2.

[28] S. Kulkarni and R. Rabidas, "Fully convolutional network for automated detection and diagnosis of mammographic masses," Multimed Tools Appl, vol. 82, no. 29, pp. 44819–44840, Dec. 2023, doi: 10.1007/s11042-023-14757-8.

[29] M. Bobowicz et al., "Attention-Based Deep Learning System for Classification of Breast Lesions—Multimodal, Weakly Supervised Approach," Cancers (Basel), vol. 15, no. 10, May 2023, doi: 10.3390/cancers15102704.

[30] R. Roslidar et al., "A Review on Recent Progress in Thermal Imaging and Deep Learning Approaches for Breast Cancer Detection," 2020, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/ACCESS.2020.3004056.

[31] S. Arooj et al., "Breast Cancer Detection and Classification Empowered With Transfer Learning," Front Public Health, vol. 10, Jul. 2022, doi: 10.3389/fpubh.2022.924432.

Design and Application of a TOPSIS-Based Fuzzy Algorithm

A Case Study from Tourism Attraction Evaluation

Fei Liu
School of Humanities and Tourism, Zhejiang Institute of Economics and Trade, Hangzhou, 310018 Zhejiang, China

***Abstract*—The study aims to evaluate the tourism attractiveness of different tourist attractions in the same region through the TOPSIS model in the perspective of culture and tourism integration, so as to provide theoretical and practical support for the tourism development of the region. On the basis of the concept of culture and tourism integration and its importance in tourism development, the evaluation index system of tourism attraction is constructed, including the indicators of tourism resources, tourism infrastructure, etc. Finally, the entropy weighting method and TOPSIS model are used for the comprehensive evaluation of these indicators, and the weight of each indicator and the comprehensive score of the tourism attraction of a certain place are derived through calculation. The results show that through the analysis of the TOPSIS model, the advantages and shortcomings of the region in terms of tourism resources and cultural characteristics can be clearly understood, and recommendations can be targeted, including strengthening tourism infrastructure construction, excavating and protecting cultural characteristics, and so on. These suggestions can help to further improve and enhance the tourism attractiveness of a certain place, so as to attract more tourists and promote the development of the local economy. Meanwhile, the methodology and framework of this study also provide reference and reference for other regions to carry out similar tourism attractiveness evaluation. In the context of cultural and tourism integration, this study expands the perspective of tourism evaluation and provides new ideas and methods for local tourism development.**

***Keywords*—Cultural and tourism integration; attractiveness; TOPSIS; entropy weight method**

I. INTRODUCTION

The integration of culture and tourism refers to the deep integration of culture and tourism, through the excavation and use of local cultural resources, to create a destination with cultural charm and tourism attraction [1]. In order to achieve a win-win situation of cultural heritage and tourism in today's society, the integration of culture and tourism has become a hot topic in the development of the tourism industry, for the evaluation of tourism attraction under the integration of culture and tourism, academics and the tourism industry have carried out a lot of research [2, 3].

In the study of tourism attractiveness evaluation under the perspective of culture and tourism integration, scholars are committed to constructing the evaluation index system, determining the weights of the indexes and selecting the comprehensive evaluation method, so as to comprehensively consider the impact of culture and tourism integration on tourism attractiveness [4, 5]. First of all, in terms of the evaluation index system, the index system usually includes indicators of tourism resources, infrastructure, service quality, cultural characteristics and other aspects, in order to comprehensively evaluate the tourism attractiveness of the destination [6, 7]. For example, the indicators of tourism resources can include natural landscape, historical sites, folk customs, etc.; the indicators of infrastructure can include the degree of accessibility, accommodation facilities, catering services, etc.; the indicators of service quality can include the attitude of tourism services, the quality of tour guides, and tourism safety [8]. Meanwhile, in order to accurately evaluate tourism attractiveness [9], it is necessary to determine the weight of each indicator, and under the perspective of cultural and tourism integration, the importance of different indicators may change, so it is necessary to determine the weight by appropriate methods [10]. Researchers usually adopt methods such as hierarchical analysis (AHP) [11] or principal component analysis (PCA) [12,13] in order to determine the weights of the indicators so as to reflect the importance of each indicator more accurately in the comprehensive evaluation. Meanwhile, evaluation method is an important part of the evaluation model, and different scholars have tried different methods to evaluate the attractiveness, such as fuzzy theory-based tourism attractiveness evaluation [14.15], MBO method [16], the integrated method of MDS and ANP method [17], and many other evaluation methods [18, 19].

Under the perspective of culture and tourism integration, the evaluation of tourism attractiveness not only needs to build a good indicator system, but also for the selection of the evaluation model is also crucial, while the increase in the number of tourists in recent years shows that how to choose good tourist attractions for travellers has become a point of concern, and the use of fuzzy comprehensive evaluation and other methods of selection of multiple attractions is a little bit of heavy workload, and computationally redundant [20]. The TOPSIS model, as a commonly used multi-indicator comprehensive evaluation method, has been widely used in various fields of evaluation, the model can be used in the evaluation and selection of tourist destinations through the comprehensive calculation of the scores and weights of the indicators to arrive at a comprehensive evaluation of tourism attractiveness of the destination [21, 23]. Therefore, this paper builds the evaluation index system on the accumulation of literature, and at the same time adopts the entropy weight method and the improved TOPSIS method to construct the evaluation model, which is used to study the evaluation of tourist attractiveness of tourist attractions under the perspective of

culture and tourism fusion, with a view to providing an in-depth exploration and expansion of the concept of culture and tourism fusion, promoting the continuous improvement and innovation of the theory of culture and tourism fusion, and at the same time, providing the scientific evaluation method and decision-making reference for the development of local tourism.

II. Evaluation Model Construction

A. Cultural and Tourism Integration

Culture and tourism integration refers to the in-depth integration of the two industries of culture and tourism to achieve the goal of mutual promotion and common development. In the integration of culture and tourism, several key points can be summarised as follows through relevant research:

- Cultural inheritance and innovation: The integration of culture and tourism requires that traditional culture be inherited and protected, while also focusing on innovation and giving traditional culture modern connotations to make it more in line with the aesthetics and needs of contemporary people.
- Integration and development of tourism resources: Fully integrate local tourism resources such as natural scenery and humanistic landscapes, carry out planning and development, and create distinctive and quality tourism products and services.
- Integration of cultural and tourism projects: Through cultural and artistic performances, cultural and creative bazaars, and cultural themed exhibitions, culture and tourism projects are combined to enrich the experience and feelings of tourists.
- Cultural and tourism industry linkage: Encouraging the in-depth integration of the cultural industry with the tourism industry, promoting the interactive development of cultural and creative products and tourism products, and forming an industry linkage effect.
- Culture and tourism branding: With the help of cultural resources and tourist attractions, create a culture and tourism brand with local characteristics and cultural connotations, and enhance the image and popularity of the place.
- Cultural and tourism marketing and promotion: Use new media and social platforms to promote cultural and tourism products and attract more tourists and cultural enthusiasts.
- Public services for culture and tourism: Upgrading cultural and tourism service facilities and public services to provide visitors and residents with a better cultural experience and tourism environment.

Many studies consider the above seven key points to be effectively integrated and implemented, and the integration of culture and tourism will be able to realise the complementary advantages of culture and tourism and promote the sustainable development of local economy and society. The author's research believes that the evaluation of attractiveness of tourist attractions is closely related to the integration of culture and tourism, and that cultural heritage and tourism experience directly affect the attractiveness of tourist attractions; excellent cultural heritage and rich tourism experience can enhance the attractiveness of attractions and attract more tourists to come to visit them, while creative design and sustainable development can add new charms to attractions and make them stand out in the highly competitive tourism market. Therefore, the key points of cultural and tourism integration are closely linked to the relationship between the attractiveness evaluation of tourist attractions, which promotes the enhancement of the cultural value and attractiveness of tourist destinations and injects new vitality into the development of the tourism industry. Therefore, this paper will carry out the tourism attractiveness evaluation based on cultural and tourism integration through different tourist attractions, to explore the important influencing factors of cultural and tourism integration and their degree of influence on the attractions.

B. Constructing the Evaluation Index System

Existing studies believe that the construction of tourism attraction evaluation index system under the perspective of culture and tourism integration should contain multiple dimensions, such as natural environment, cultural heritage, tourism infrastructure, service quality, etc., in order to comprehensively assess tourism attraction. At the same time, the evaluation indicators should be comparable to facilitate comparison between different regions or attractions, and the meaning of the indicators should also take into account the regional characteristics and differences; in view of the fact that the current indicators are more subjective, the evaluation indicators should be quantifiable, so that it is easy to collect and analyse the data, based on which to ensure the objectivity of the evaluation results.

Through reviewing the literature, this paper argues that the indicator system needs to take into account the cultural, natural environment and historical characteristics of different regions, i.e., the evaluation indicators should have a certain degree of geographical adaptability, avoiding the simple application of standard indicators. Given that most of the research objects become attractions have been solidified, the index referentiality needs to be sufficient, and should be based on the status quo more easily accessible data and information, to avoid difficult to obtain or inaccurate data. At the same time, taking into account that the tourism market serves the market, so the evaluation index system should adapt to the changes in the tourism market and environment. Finally, this paper considers that the evaluation indexes should be operable, so that managers and decision-makers can take corresponding management measures according to the evaluation results. Based on this, this paper finally constructs the indicator system as shown in Table I.

TABLE I. TOURISM ATTRACTION EVALUATION INDEX SYSTEM BASED ON THE CONCEPT OF CULTURAL AND TOURISM INTEGRATION

Level 1 indicators	Secondary indicators	Indicator assessment content
Indicators of cultural content	Number of cultural heritages	Measures the amount of important cultural heritage, such as historical buildings, monuments and artefacts that a destination possesses.
	Status of protection of cultural heritage	Assessment of the protection and restoration of cultural heritage at the destination, including restoration work, regulatory measures, etc.
	Richness of cultural activities	Measure the number and quality of cultural festivals, exhibitions, performances and other events organized by the destination.
	Development of cultural industries	Assess the scale and contribution of the destination's cultural industries, including cultural and creative products, cultural and artistic performances, and so on.
Tourism infrastructure indicators	Accessibility	Evaluate how well the destination is connected to the transport network, including accessibility by road, rail and air.
	Accommodation facilities	Measures the number and quality level of accommodation facilities such as hotels, B&Bs and resorts in the destination.
	Tourism reception capacity	Evaluate the destination's capacity and level of service to tourists, including guides, interpreters, reception facilities, etc.
	Service level of guided tours	Measure the professionalism and service attitude of the destination guide.
Natural landscape indicators	Quality of the natural environment	Measurement of natural environmental indicators such as air quality, water quality and environmental cleanliness of the destination.
	Natural landscape features	Evaluate the unique natural landscape features of the destination, such as mountains, lakes and forests.
	Ecotourism resources	Measures the richness of the destination's ecotourism resources, including wildlife, eco-farms, etc.
Indicators for cultural and tourism integration projects	Development of Cultural and Creative Industries	Assess the extent and impact of the cultural and creative industries in the destination.
	Culture and Tourism Integration Construction	To measure the construction of demonstration projects and demonstration areas in the field of cultural and tourism integration in destinations.
	Cultural and Tourism Integration Activities Organised	To assess the number and quality of cultural and tourism integration activities organised by the destination, such as cultural and arts festivals and cultural and creative bazaars.
Indicators of tourist satisfaction	Traveller's comment	Gather travellers' ratings and opinions about the destination, including attractions, services, transport and more.
	Tourist return rate	To assess the proportion of tourists who choose the destination again to travel, reflecting the return and loyalty of the destination's visitors.
	Word-of-mouth communication	Evaluate the destination's word-of-mouth and user experience through online and social media channels.

It can be seen that in the evaluation of the attractiveness of tourism regions, the indicator system based on the integration of culture and tourism is considered more comprehensively, and as we all know, the concept of integration of culture and tourism refers to the integration of culture and tourism to create attractions or projects with cultural connotations and tourism attractiveness. Therefore, it can be seen that the following characteristics are presented when constructing tourism attraction indicators:

1) Cultural connotation: Whether the attraction or project has rich cultural connotation, such as historical heritage, folk customs, artistic performance, etc., which can be reflected by indicators of cultural heritage, history and culture, folk customs, etc., and is compatible with the first-level indicator of cultural connotation indicators.

2) Tourism facilities: Whether the attraction or project provides perfect tourism facilities and services, such as convenient transport, accommodation and catering, and guided tour services. This can be reflected through the indicators of convenient transportation, accommodation and catering support, which corresponds to the tourism infrastructure indicator in this paper.

3) Experience and interaction: Whether the attraction or project provides rich experience and interactive projects, such as cultural experience, handicraft production, interactive performances, etc., which can be reflected through the indicators of experiential projects and interactive activities, which corresponds to the indicator of cultural and tourism integration projects in this paper.

4) Visitors' perception: Whether the attraction or project is innovative and sustainable, such as the innovative application of the concept of cultural and tourism integration, the richness of tourism resources, and the follow-up evaluation of tourists, etc. This point is mirrored by the indicators of natural landscape and the indicators of tourists' satisfaction in this paper, and this paper considers that although there is a little bit of linkage between the two, there is also a certain degree of discrepancy, so they are listed separately to be analysed.

C. Coefficient of Variation Method

Entropy weighting method is a multi-indicator decision-making method to determine the importance of each indicator in decision-making by calculating the entropy value and weight of each indicator [24]. Entropy weighting method can consider the correlation and importance between indicators by calculating the information entropy of the indicators, so as to reflect more accurately the influence of each indicator on the decision-making, and at the same time, compared with the traditional method of subjective assignment, the entropy weighting method does not require subjective evaluation, but rather, it determines the weights through the amount of information of the data itself, which has the objectivity and scientific nature [25].

This paper considers that the entropy weight method can consider the importance of multiple indicators comprehensively, which can well reflect the multiple factors and indicators involved in the evaluation of tourism attractiveness under the concept of cultural-tourism fusion, and at the same time, the process of adopting the entropy weight method to use information entropy to measure the differences and uncertainties between indicators and to determine the weights of the indicators can better reflect the correlations and influences among the factors of tourism attractiveness under the concept of cultural-tourism fusion. In general, the entropy weight method is comprehensive, objective and widely applicable in the evaluation of tourism attractiveness under the concept of cultural and tourism integration, and can effectively calculate the weights and reflect the impact of cultural and tourism integration on tourism attractiveness. The calculation steps are as follows.

1) Indicator data forwarding: Considering that the indicator weight data in this paper all belong to the larger and more important indicators, therefore, this kind of indicator data does not need to be forwarded.

$$x'_{ij} = x_{ij} \tag{1}$$

Where: x'_{ij} - Data matrix elements after normalisation;

x_{ij} -Initial element of the data matrix.

2) Data standardisation

$$r_{ij} = \frac{x'_{ij} - min\left(x'_j\right)}{max\left(x'_j\right) - min\left(x'_j\right)} \tag{2}$$

Where: r_{ij} - Normalised data matrix elements;

x'_{ij} -After normalisation the data matrix elements in this paper are the initial matrix elements.

3) Calculating information entropy: Noting the matrix obtained from the normalised post-processing as $R = \left(r_{ij}\right)_{m\times n}$, the information entropy is E_j for a given indicator r_j.

$$E_j = -\frac{1}{lnm}\sum_{i=1}^{m} p_{ij} lnp_{ij} \tag{3}$$

Included among these:

$$p_{ij} = \frac{r_{ij}}{\sum_{j=1}^{n} r_{ij}} \tag{4}$$

Where: E_j - information entropy;

r_{ij} -Normalised data matrix elements.

If the information entropy of an indicator is smaller, it indicates that the degree of variation of its indicator value is larger, and the amount of information provided is also larger, and it can be considered that the indicator plays a greater role in the comprehensive evaluation.

4) Calculation of weights

$$\omega_{ij} = \frac{1 - E_j}{\sum_{j=1}^{n}\left(1 - E_j\right)} \tag{5}$$

Where: E_j - information entropy.

D. Improvement of the TOPSIS Model

Based on Table I of this paper, it can be seen that the evaluation of tourism attractiveness based on culture and tourism integration is a process that involves a wide range of indicators, and at the same time the assessment content is redundant, while the TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) model can comprehensively consider a number of indicators, including culture, tourism, economy and other indicators, so as to more comprehensively evaluate the tourism attractiveness. The TOPSIS (Technique for Order of Preference by Similarity to Ideal Solution) model can consider multiple indicators, including cultural, tourism, economic and other indicators, to evaluate tourism attractiveness in a more comprehensive way. At the same time, the TOPSIS model is applicable to different types of indicators and different ranges of data, and can be flexibly adjusted and applied according to the characteristics of cultural and tourism resources in different regions. In addition, the advantage of the TOPSIS model is that it can compare the evaluation object with the ideal solution and the negative ideal solution, so as to determine the optimal solution, which is conducive to the improvement of the objectivity and accuracy of the evaluation results. The purpose of this paper is that through the evaluation of TOPSIS model, the advantages and shortcomings of cultural and tourism integration can be better found, and provide scientific basis for the development of cultural and tourism integration, of course, there are some limitations of TOPSIS method, such as the determination of the weights is more subjective, taking this into account, this paper adopts the objective weighting method - entropy weighting method for the determination of weights, and at the same time, considering that the TOPSIS model is lower than the perception of the evaluation results, so it is improved. At the same time, considering that the TOPSIS model has a low perception of the evaluation results in the evaluation, it is improved to carry out the evaluation of tourism attractiveness based on the improved TOPSIS model, and its calculation process is as follows.

1) Calculate the weighted data matrix

$$e_{ij} = \omega_j r_{ij} \tag{6}$$

Where: e_{ij} - weighted matrix elements;

ω_{ij} -weight vector;

r_{ij} -Normalised data matrix elements.

2) Calculate the distance between the weighting matrix and the most value

After processing you can form a data matrix

$$M = \left(e_{ij}\right)_{m\times n} \tag{7}$$

Define the maximum value of each indicator, i.e., each column, ase_j^+

$$e_j^+ = max\left(e_{1j}\cdots e_{nj}\right) \tag{8}$$

Define the maximum value of each indicator, i.e., each column, ase_j^-

$$e_j^- = max\left(e_{1j}\cdots e_{nj}\right) \tag{9}$$

Define the ith object to be at a distance from the maximum asd_i^+ (positive ideal solution)

$$d_i^+ = \sqrt{\sum_{j=1}^{n}\left(e_j^+ - r_{ij}\right)^2} \tag{10}$$

Define the distance of the ith object from the maximum asd_i^- (negative ideal solution)

$$d_i^- = \sqrt{\sum_{j=1}^{n}\left(e_j^- - r_{ij}\right)^2} \tag{11}$$

3) Calculation of scores

$$S_i = \frac{d_i^-}{d_i^- + d_i^+} \tag{12}$$

$$Score_i = \frac{S_i}{maxS_i + 0.1} \tag{13}$$

Where: S_i -TOPSIS calculations;

$Score_i$ -Evaluation score.

III. Example Validation of Tourism Attractiveness Evaluation in the Perspective of Culture and Tourism Integration based on TOPSIS Model

A. Overview of the Study Area

The study area belongs to Mabian County in Sichuan Province, which is located under the jurisdiction of Leshan City, a county full of history, culture and natural scenery. At present, tourism in Mabian County is in a rapid development stage, attracting more and more tourists for sightseeing, holiday and leisure. First of all, Mabian County has rich historical and cultural resources, including ancient religious buildings and traditional folk culture. Guanyin Yan, Chongtianlou, Taiping, Ma'er Mountain Forest Park, and the First Walled City of Liangshan Mountain in the county attract a large number of tourists. In addition, there are some traditional folk activities in the county, such as Sichuan Opera, Sichuan Cuisine and Sichuan Tea, which attract culture lovers to experience. Secondly, Mabian County is blessed with natural scenery resources, beautiful scenery, with excellent outdoor sports and adventure environment, such as hiking, rock climbing, rafting and other projects are highly favoured by tourists. In addition, it is understood that Mabian County is also actively developing rural tourism, combining local farming culture and idyllic scenery with tourism, providing tourists with a richer tourism experience. In general, the current status of tourism development in Mabian County is good, with abundant tourism resources, attracting a large number of tourists, but how attractive the attractions are based on the view of cultural and tourism integration is still an unknown, so this paper will evaluate and analyse the five tourism zones in Mabian County (Guanyinyan, Chongtianlou, Taiping, Ma'er Mountain Forest Park, and the First Walled Village of Liangshan Mountain, which is hereinafter represented by Attractions 1 - Attractions 5).

B. Entropy Weight Method Weight Calculation

In this paper, the above research object to collect data is shown in Table II, while based on the collection of data according to the Eq. (1-5) calculated to get the entropy weight method of the relevant values and weights are shown in Table III.

C. Analysis of Weighting Results

In this paper, according to the results of weight calculation in Table III, the comprehensive single-indicator weight analysis chart and the analysis chart of hierarchical weight results are drawn by sorting as shown below Fig. 1 and Fig. 2.

Combined with the Fig. 1 and Fig. 2 , it can be seen that for the first-level indicators, cultural connotation indicators and natural landscape indicators accounted for 25.97% and 20.59% respectively, which shows that in the evaluation of tourism attractiveness of culture and tourism integration, the tilt towards culture is obvious, and the reason why cultural connotation indicators and natural landscape indicators are more important is that they represent the history of the people and the environment, which are the two important components of the tourism resources respectively. First of all, the cultural connotation indicator represents the richness of a place's history, traditions, arts and customs, etc. Places with rich cultural connotations tend to attract the interest of tourists, who want to feel and experience the local cultural charms by visiting cultural monuments, participating in traditional festivals, tasting local cuisine, etc. Therefore, the cultural connotation indicator is crucial to the attractiveness of culture and tourism integration tourism. The natural landscape indicator represents the attractiveness of a place's natural environment, geographical landscape and ecological resources, etc. Places with beautiful natural landscape can often make tourists feel the wonder and beauty of nature, and they want to enjoy the beauty of nature through viewing natural scenery, participating in outdoor activities, experiencing ecological farming, etc. Therefore, the natural landscape indicator is also very important for the attractiveness of the integrated tourism of culture and tourism. Therefore, the natural landscape indicator is also very important for the attraction of cultural tourism integration. This also shows that when a place has both rich cultural connotations and

beautiful natural landscapes, it tends to attract more tourists to come to experience and explore.

TABLE II. DATA ON THE STUDY POPULATION

Secondary indicators	Attractions 1	Attractions 2	Attractions 3	Attractions 4	Attractions 5
Number of cultural heritages	85	82	81	70	78
Status of protection of cultural heritage	90	87	86	76	87
Richness of cultural activities	75	70	73	80	70
Development of cultural industries	75	70	72	87	76
Accessibility	80	83	86	70	78
Accommodation facilities	78	81	88	76	87
Tourism reception capacity	86	89	87	83	75
Service level of guided tours	88	91	90	81	70
Quality of the natural environment	87	81	78	89	70
Natural landscape features	90	86	87	82	81
Ecotourism resources	78	73	75	87	75
Development of Cultural and Creative Industries	87	90	75	70	75
Culture and Tourism Integration Construction	70	80	80	70	72
Cultural and Tourism Integration Activities Organised	76	80	78	81	86
traveller's comment	80	75	83	89	88
Tourist return rate	75	75	81	91	82
Word-of-mouth communication	80	76	89	87	87

TABLE III. RESULTS OF WEIGHTING CALCULATIONS

Secondary indicators	Information entropy	weights	Tiered weighting	Level 1 indicators	Weights
Number of cultural heritages	0.8514	0.0589	0.2268	Indicators of cultural content	0.2597
Status of protection of cultural heritage	0.8458	0.0611	0.2354		
Richness of cultural activities	0.9216	0.0311	0.1197		
Development of cultural industries	0.7262	0.1086	0.4181		
Accessibility	0.8793	0.0479	0.2169	Tourism infrastructure indicators	0.2207
Accommodation facilities	0.8820	0.0468	0.2120		
Tourism reception capacity	0.8872	0.0447	0.2027		
Service level of guided tours	0.7950	0.0813	0.3684		
Quality of the natural environment	0.8214	0.0708	0.3440	Natural landscape indicators	0.2059
Natural landscape features	0.8542	0.0578	0.2809		
Ecotourism resources	0.8053	0.0772	0.3751		
Development of Cultural and Creative Industries	0.8825	0.0466	0.3215	Indicators for cultural and tourism integration projects	0.1449
Culture and Tourism Integration Construction	0.8463	0.0609	0.4206		
Cultural and Tourism Integration Activities Organised	0.9058	0.0374	0.2579		
traveller's comment	0.8688	0.0520	0.3081	Indicators of tourist satisfaction	0.1688
Tourist return rate	0.9229	0.0306	0.1812		
Word-of-mouth communication	0.7826	0.0862	0.5107		

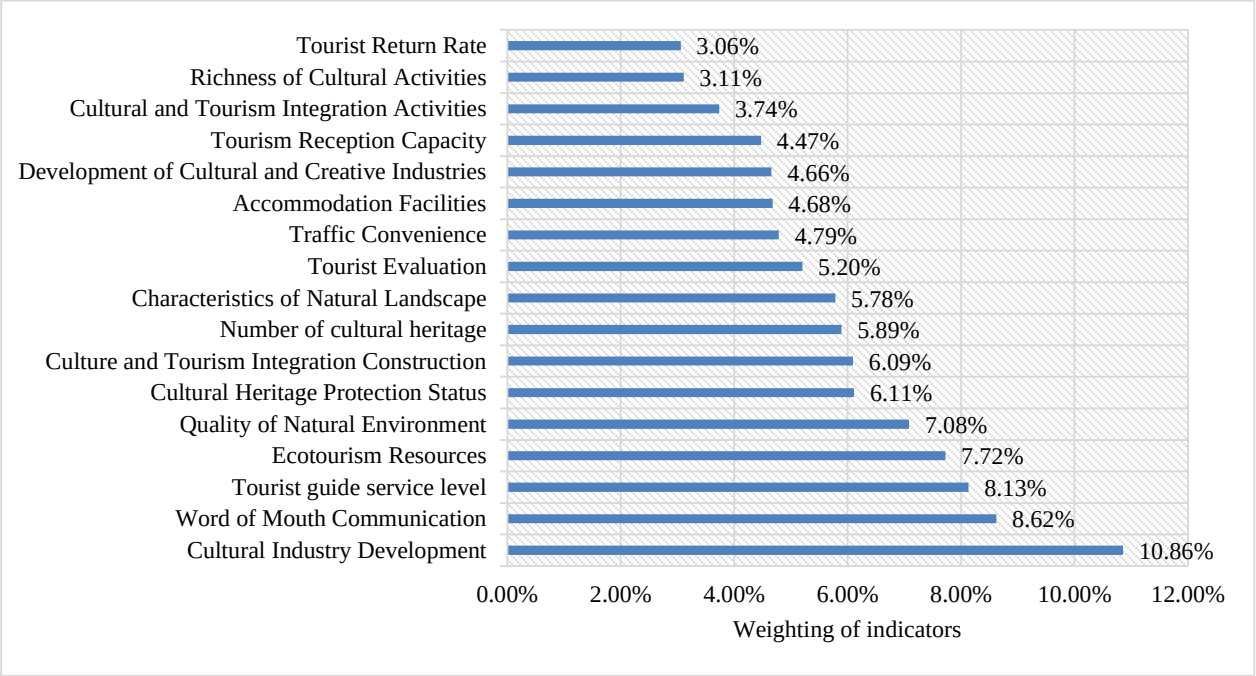


Fig. 1. Histogram of the ranking of weighting results.

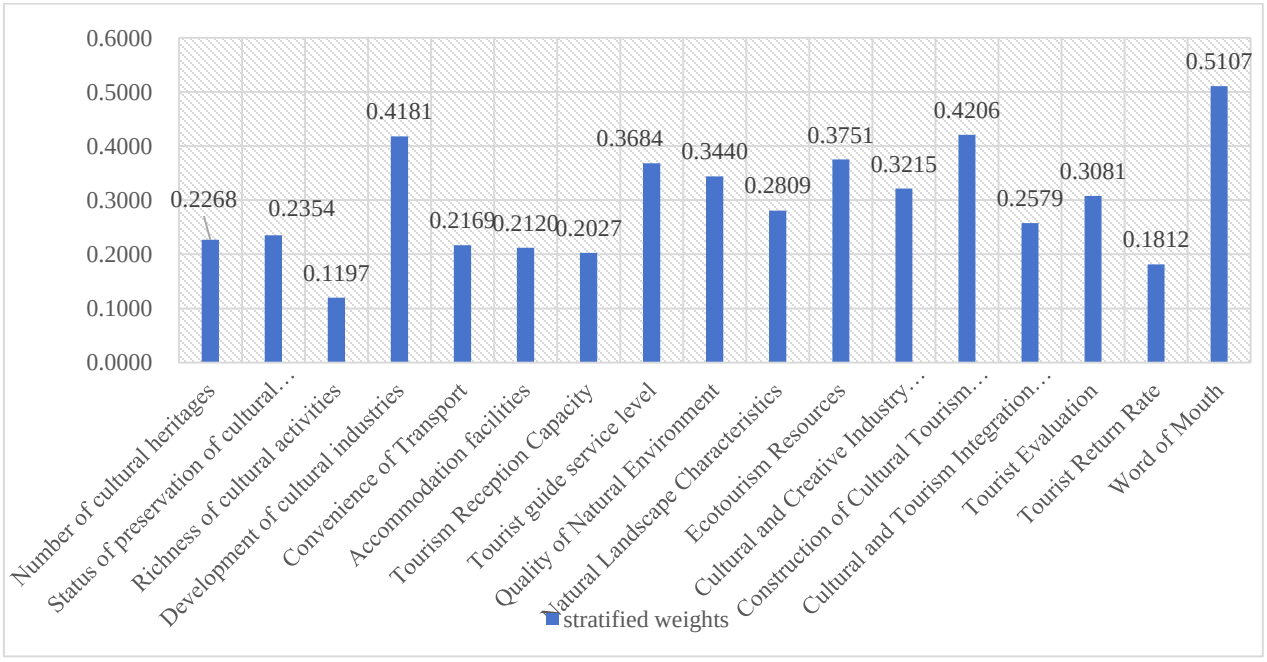


Fig. 2. Histogram of stratified weights.

And according to the calculation results of single indicators and bar charts, it can be seen that the development of cultural industry, word-of-mouth communication, tour guide service level, eco-tourism resources, and the quality of the natural environment are in the forefront, accounting for 10.86 %, 8.62 %, 8.13 %, 7.72 %, 7.08 %, which can be seen to belong to the indicators of cultural connotation, tourists' satisfaction, tourism infrastructure, and the natural landscape, respectively. This shows that even though the first-level indicators are not considered, the single indicator weights still indicate the importance of the indicators under the first-level indicators, which are analysed as follows: in the evaluation of tourism attractiveness of culture and tourism integration, the cultural industry development, word-of-mouth transmission, tour guide service level, ecotourism resources and natural environment quality indicators are all more important because they each represent different aspects of tourism purposes, and play an important role in enhancing the tourist The reason is that they represent different aspects of tourism purposes and play an important role in enhancing tourist experience and attracting tourists. Firstly, the cultural industry development indicator represents the development level of a place in terms of cultural

creative industries, cultural festivals and cultural product development. A place with rich cultural industries can provide tourists with diversified cultural experiences and cultural products, enhance their sense of participation and interactivity, and thus improve tourism attractiveness. Secondly, word-of-mouth (WOM) indicators represent tourists' satisfaction and experience of a destination. Good WOM means that tourists praise and recommend the destination, which can attract more tourists to come. Word-of-mouth communication is especially important in today's social media era, and tourists' favourable comments and sharing can directly influence other potential tourists' choices. The tour guide service level indicator represents the level of tourism service and the quality of tourists' experience in the destination. Quality tour guide service can provide tourists with a better travelling experience and enhance their knowledge and understanding of the destination, thus increasing their satisfaction and loyalty to the destination. Finally, the ecotourism resources indicator and the natural environment quality indicator represent the natural environment and ecological protection of the destination. In today's context of environmental protection and sustainable development [22], rich ecotourism resources and good quality of the natural environment can attract more tourists who are concerned about environmental protection, and also meet the contemporary demand for healthy, leisure and environmentally friendly tourism.

The results of the above weighting analyses show that important factors need to be considered comprehensively in tourism development and evaluation in order to enhance the competitiveness and attractiveness of tourism destinations.

D. The TOPSIS Model

1) Model evaluation: The TOPSIS model evaluation of the study area was carried out according to the Eq. (6)-Eq. (13), and the results of the calculation of the first-level indicators of the five study objects and their attractiveness evaluation are now presented in Table IV.

TABLE IV. EVALUATION RESULTS FOR THE STUDY AREA

place of interest (tourism)	Indicators of cultural content			Tourism infrastructure indicators		
	d+vector	*d-vector*	*score*	*d+vector*	*d-vector*	*score*
Attractions 1	0.0297	0.0288	68.3038	0.0152	0.0391	69.0666
Attractions 2	0.0427	0.0205	44.9665	0.0089	0.0468	80.6586
Attractions 3	0.0378	0.0195	47.1028	0.0029	0.0476	90.4069
Attractions 4	0.0258	0.0423	86.1356	0.0315	0.0231	40.5971
Attractions 5	0.0296	0.0225	59.8747	0.0448	0.0160	25.2612
place of interest (tourism)	**Natural landscape indicators**			**Indicators for cultural and tourism integration projects**		
	d+vector	*d-vector*	*score*	*d+vector*	*d-vector*	*score*
Attractions 1	0.0198	0.0364	72.0567	0.0306	0.0295	54.3214
Attractions 2	0.0343	0.0221	43.5494	0.0444	0.0086	89.3314
Attractions 3	0.0335	0.0181	38.9272	0.0269	0.0294	51.0423
Attractions 4	0.0118	0.0470	88.8821	0.0072	0.0446	14.8057
Attractions 5	0.0463	0.0043	36.2356	0.0177	0.0337	36.7036
place of interest (tourism)	**Indicators of tourist satisfaction**			**attractiveness rating**		
	d+vector	*d-vector*	*score*	*d+vector*	*d-vector*	*score*
Attractions 1	0.0413	0.0102	19.0468	0.0117	0.0146	80.6141
Attractions 2	0.0321	0.0249	41.8058	0.0154	0.0140	69.3502
Attractions 3	0.0249	0.0264	49.2807	0.0130	0.0144	76.3585
Attractions 4	0.0027	0.0461	90.4190	0.0119	0.0170	85.4721
Attractions 5	0.0206	0.0314	57.8215	0.0166	0.0097	53.6475

2) Analysis of results: Guanyinyan, Chongtianlou, Taiping, Maer Mountain Forest Park, Liangshan First Walled City scored 80.6141, 69.3502, 76.3585, 85.4721, 53.6475 in this evaluation, it can be seen that Chongtianlou and Liangshan First Walled City scored relatively low, based on the results of the attractiveness evaluation, based on the rating results of the two attractions mentioned above, this paper carries out the secondary survey analysed after the visit. First of all, the main reason for Chongtianlou is that the cultural connotation is not deep enough and there are certain problems in dissemination, Chongtianlou as a cultural attraction, its history, traditions, folklore and other aspects of cultural connotation is not rich enough and insufficient publicity, through the evaluation of the first level of indicators can be seen, its cultural connotation indicators score is very low, so it is necessary to strengthen the publicity and properly enrich their own advantages as a cultural classic. Mabian County Liangshan First Walled City attractions scored low mainly because of the scenic area's infrastructure is not perfect, the level of service is not high, including car parks, toilets, guide services, transportation, etc., and at the same time, the scenic area's cultural and tourism integration projects are not diversified enough and innovative enough, and are more conventional, if the Liangshan First Walled City continues to lack of innovative projects with local culture, traditional industries and other fusion, will affect the attractiveness of the scenic area to tourists, therefore, it is necessary to improve the

cultural connotation indicators. Attractive to tourists, so to improve the rating of the scenic area, it is necessary to improve the infrastructure, the introduction of innovative cultural and tourism integration projects.

IV. CONCLUSION

This study evaluates the tourism attractiveness in the perspective of culture and tourism integration based on the TOPSIS model, and the results show that the model can effectively assess the attractiveness of tourism destinations and provide a scientific basis for tourism development, and the results of the evaluation of tourism attractiveness through the perspective of culture and tourism integration show that the integration of culture and tourism resources plays an important role in enhancing the attractiveness of tourism destinations, which provides a It provides theoretical support for the development of culture and tourism integration. At the same time, this study also found that under the perspective of culture and tourism integration, the evaluation of tourism attractiveness needs to consider the integration of cultural heritage and tourism experience, in order to achieve the sustainable development of tourism destinations to further improve the attractiveness of existing tourism resources, and at the same time, it is necessary to strengthen the cultural and artistic activities of attractions with cultural heritage, and to make a unique tourism area.

In view of the demand for the development of cultural and tourism integration, more field research and case analyses can be considered in the future to explore the evaluation methods of tourism attractiveness in different regions and cultural backgrounds, so as to enrich the research results. At the same time, we can also combine big data and artificial intelligence technology to build a more accurate tourism attractiveness evaluation model under the perspective of culture and tourism integration, so as to achieve dynamic data collection and dynamic real-time evaluation, and to provide more scientific support for tourism development as well as suggestions for development decisions.

REFERENCES

[1] Canavan B. Tourism culture: nexus, characteristics, context and sustainability[J]. Tourism management, 2016, 53: 229-243.

[2] Tang, C., Liu, Y., Wan, Z., & Liang, W. (2023). Evaluation system and influencing paths for the integration of culture and tourism in traditional villages. Journal of Geographical Sciences, 33(12), 2489-2510. Journal of Geographical Sciences, 33(12), 2489-2510.

[3] Ma, Y., & Chen, Y. (2023). The Inspiration of the Fusion of Chinese and Western Cultures for the Development of Macau City. Journal of Sociology and Ethnology, 5(11), 162-166.

[4] zhu, h., liu, j. m., tao, h., & zhang, j. (2015). Evaluation and spatial analysis of tourism resources attraction in Beijing based on the internet information. Journal of Natural Resources, 30(12),. 2081-2094.

[5] Kenedi, K., Sukmawan, I., & Laksana, A. (2022). Evaluation of the economic potential of coastal tourism strategic area of anyer tourism-cinangka. Jurnal Ekonomi, 11(01), 611-618.

[6] Cherapanukorn, V., & Sugunnasil, P. (2022). Tourist attraction satisfaction factors from online reviews. a case study of tourist attractions in Thailand. journal of environmental management & amp; Tourism, 13(2), 379-390.

[7] Cao, Q., Sarker, M. N. I., Zhang, D., Sun, J., Xiong, T., & Ding, J. (2022). Tourism competitiveness evaluation: evidence from mountain tourism in China. Frontiers in Psychology, 13, 809314.

[8] Wang, S., Wang, J., Shen, W., & Wu, H. (2023). The evaluation of tourism service facilities in Chinese traditional villages based on the living protection concept: Theoretical framework and empirical case study. Journal of Asian Architecture and Building Engineering, 22(1), 14-31.

[9] Meng Jia. Research on the Attractiveness Evaluation of County Tourism of Old Revolutionary Sites [D]. Hebei Normal University,2023.DOI:10.27110/d.cnki.ghsfu.2023.000446.

[10] Yang, S., & Kong, X. (2022). Evaluation of Rural Tourism Resources Based on AHP-Fuzzy Mathematical Comprehensive Model. Mathematical Problems in Engineering, 2022 (1), 7196163.

[11] Gu, X., Hunt, C. A., Jia, X., & Niu, L. (2022). Evaluating nature-based tourism destination attractiveness with a Fuzzy-AHP approach. Sustainability, 14(13), 7584.

[12] HUA Zhiqiang,ZHANG Chunsheng,CHEN Liying,et al. Comprehensive evaluation of tourism resource attractiveness based on principal component analysis method[J]. Journal of Hubei College of Nationalities (Natural Science Edition),2015,33(04):399-401.Gu, X., Hunt, C. A., Jia, X., & Niu, L. (2022). Evaluating nature-based tourism destination attractiveness with a Fuzzy-AHP approach. sustainability, 14(13), 7584.

[13] Hoang, H. T., Truong, Q. H., Nguyen, A. T., & Hens, L. (2018). Multicriteria evaluation of tourism potential in the central highlands of vietnam: combining geographic information system (GIS), analytic hierarchy process (AHP) and principal component analysis (PCA). Sustainability, 10(9), 3097.

[14] Gu, X., Hunt, C. A., Jia, X., & Niu, L. (2022). Evaluating nature-based tourism destination attractiveness with a Fuzzy-AHP approach. Sustainability, 14(13), 7584.

[15] Hou Xiaomin. Research on the Evaluation of Tourism Attractiveness of Destinationless Cruise Products [D]. Guizhou Normal University, 2023.Vinogradova, M. V., Larionova, A. A., Maloletko, A. N., & Kaurova, O. V. (2016).

[16] The use of MBO (management of objectives) method of attraction and evaluation of effectiveness of investments to the tourism and hospitality. International review of management and Marketing, 6(2), 241-246.

[17] Saputro, K. E. A., Hasim, Karlinasari, L., & Beik, I. S. (2023). Evaluation of Sustainable Rural Tourism Development with an integrated approach using MDS and ANP methods: case study in Ciamis, West Java, Indonesia. Sustainability, 15(3), 1835.

[18] Lu, W. A. N. G., Ziruo, H. U. A. N. G., Le, Y. U., & Zhizhong, N. I. N. G. (2024). Evaluation of the Development Resilience of Tourist Attractions under the Influence of Major Public Health Events. Journal of Resources and Ecology, 15 (3), 698-710.

[19] Huang Lan. Research on Tourism Attraction Evaluation of West Lake Famous Scenic Area Based on IPA Analysis [D]. Northwest Normal University,2022.DOI:10.27410/d.cnki.gxbfu.2022.002177.

[20] Wang, Yumei, Peide Liu, and Yiyu Yao. "BMW-TOPSIS: A generalised TOPSIS model based on three-way decision." information sciences 607 (2022): 799-818.

[21] Ramakrishnan, Krishnapuram Ravi, and Shankar Chakraborty. "A cloud TOPSIS model for green supplier selection. "Facta Universitatis, Series. Mechanical Engineering 18.3 (2020): 375-397.

[22] Zhao, Ding-Yi, Yu-Yu Ma, and Hung-Lung Lin. "Using the entropy and TOPSIS models to evaluate sustainable development of islands: a case in China." Sustainability 14.6 (2022): 3707.

[23] Luyen, Le Anh, and Nguyen Van Thanh. "Logistics service provider evaluation and selection: hybrid servqual-fahp-topsis model." Processes 10.5 (2022): 1024.

[24] HUANG Tianyuan,WANG Shunsheng. Evaluation of science popularisation capacity of water conservancy scenic area based on entropy power method-AHP empowerment[J]. Water Resources Planning and Design,2024,(06):94-98+120.

[25] Li, Z., Luo, Z., Wang, Y., Fan, G., & Zhang, J. (2022). Suitability evaluation system for the shallow geothermal energy implementation in region by Entropy Weight Method and TOPSIS method. Renewable Energy Renewable Energy , 184, 564-576.

Enhanced Butterfly Optimization Algorithm for Task Scheduling in Cloud Computing Environments

Yue ZHAO
College of Computer, Cangzhou Jiaotong College, Cangzhou 061199, China

***Abstract*—Cloud computing is transforming the provision of elastic and adaptable capabilities on demand. A scalable infrastructure and a wide range of offerings make cloud computing essential to today's computing ecosystem. Cloud resources enable users and various companies to utilize data maintained in a distant location. Generally, cloud vendors provide services within the limitations of Service Level Agreement (SLA) terms. SLAs consist of various Quality of Service (QoS) requirements the supplier promises. Task scheduling is critical to maintaining higher QoS and lower SLAs. In simple terms, task scheduling aims to schedule tasks to limit wasted time and optimize performance. Considering the NP-hard character of cloud task scheduling, metaheuristic algorithms are widely applied to handle this optimization problem. This study presents a novel approach using the Butterfly Optimization Algorithm (BOA) for scheduling cloud-based tasks across diverse resources. BOA performs well on non-constrained and non-biased mathematical functions. However, its search capacity is limited to shifted, rotated, and/or constrained optimization problems. This deficiency is addressed by incorporating a virtual butterfly and improved fuzzy decision processes into the conventional BOA. The suggested methodology improves throughput and resource utilization while reducing the makespan. Regardless of the number of tasks, better results are consistently produced, indicating greater scalability.**

***Keywords*—Cloud computing; resource utilization; task scheduling; Butterfly Optimization Algorithm; fuzzy decision strategy**

I. INTRODUCTION

The Internet of Things (IoT) has evolved from the exponential proliferation of smart sensors in recent years and the demand for inter/interconnections between devices [1, 2]. IoT opens up broad medical, manufacturing, and logistics opportunities, necessitating high reliability, durability, flexibility, adaptability, and control levels [3]. Furthermore, IoT devices are limited in resources and equipped with specialized chips configured with various rules [4]. In this way, conventional networks become more complex owing to the specific requirements of IoT applications. Software-Defined IoT (SD-IoT) aims to apply Software-Defined Networking (SDN) to IoT to bring elasticity to managing resources and networks in traditional networks. SDN is regarded as a critical paradigm for next-generation networking [5].

A group of networked computers with several shared computing resources is referred to as the cloud. Recently, cloud computing has developed rapidly, enabling globally distributed data centers to develop and scale up to provide high-quality and reliable services [6]. Cloud computing has emerged as an effective model for providing computational resources on a "pay-per-use" basis. It brings uniformity and transformation to IT enterprises [7]. Cloud computing has significant prospects and poses several problems in conventional IT evolution due to its expanding uses and promotion [8]. In recent years, cloud computing has become an alternative online strategy to empower users. It offers access to shareable and customizable resources on demand, quickly allocated and released with little management or collaboration from the cloud provider [9]. This invention offers several advantages, including enhanced economic benefits related to time, cost, inventory management, and storage. This breakthrough enables all programs to operate on a virtual platform, with resources allocated across Virtual Machines (VMs) [10].

An effective and dynamic task scheduler is essential when multiple users simultaneously request services from the cloud environment, particularly from diverse and heterogeneous resources [11]. An optimal and adaptable task scheduler is critical in the cloud paradigm. Moreover, it must function under the workload submitted to the cloud platform [12]. An inefficient scheduling process in the cloud environment causes diminished service quality from cloud service providers, eroding confidence and adversely affecting corporate operations [13]. Hardware virtualization is the basis for distributing cloud resources. Numerous VMs are hosted on an individual computing server to support multiple users executing concurrent processes. VMs running in cloud data centers are given thousands of tasks. Consequently, employing an effective scheduler inside the cloud framework is advantageous for cloud providers and customers, allowing mutual benefits.

Task scheduling assigns cloud tasks to VMs to shorten makespan and enhance resource usage. Due to the NP-hard characteristics of this issue, conventional scheduling techniques have challenges regarding scalability and efficiency, especially in dynamic cloud settings [14]. Metaheuristic algorithms, such as the Butterfly Optimization Algorithm (BOA), have been shown to help tackle complicated optimization problems [15]. These algorithms leverage techniques like random walks and graph-based embeddings to improve search efficiency and adaptability in diverse optimization contexts [16]. BOA is suitable for such tasks due to its simplicity, excellent balance between exploration and exploitation, and adaptability to diverse optimization landscapes. Besides, its computational efficiency and the ability to converge on high-quality solutions make it a competitive choice for improving cloud task scheduling. This study presents an improved BOA, including a fuzzy decision method and an innovative virtual butterfly idea to maximize the algorithm's search efficacy and flexibility in

cloud job scheduling. As a summary, this study made the following contributions:

- A novel variant of the BOA is introduced, incorporating a fuzzy logic model and a virtual butterfly design to improve the algorithm's search efficiency and adaptability in cloud computing environments.
- A fuzzy logic model is implemented to continuously alter the balance between BOA's exploration and exploitation phases, allowing for better adaptation to varying optimization conditions.
- A virtual butterfly agent is developed that aggregates information from all butterflies, directing the swarm to promising areas in the search area, thereby avoiding premature convergence and improving the overall solution quality.
- The enhanced algorithm is evaluated using CloudSim with GoCJ and HCSP datasets, demonstrating its superior performance in minimizing makespan, improving resource utilization, and enhancing throughput compared to other metaheuristic methods like PSO and standard BOA.

The remaining portion of the paper is laid out in the following arrangement. Section II summarizes related research on cloud-based task scheduling. Section III defines the task scheduling problem and presents the challenges in optimizing makespan and resource utilization. Section IV introduces the proposed enhanced BOA, detailing its fuzzy decision strategy and virtual butterfly concept. Section V reports the findings and analyzes the efficiency of the developed algorithm. Section VI discusses key findings and outlines the limitations of current work. Lastly, Section VII offers a conclusion and recommends areas for further research.

II. Literature Review

Metaheuristic algorithms have been extensively adopted for scheduling tasks in cloud computing. Recent research efforts have focused on improving these algorithms to address local optima and poor convergence challenges. Some research studies emphasize the importance of parallel calculations in maximizing task scheduling efficiency. The need to balance exploration with exploitation has resulted in hybrid and adaptable methodologies. Nevertheless, several current methodologies encounter constraints when used in extensive, diverse cloud settings.

Dubey and Sharma [17] proposed the Chemical Reaction Partial Swarm Optimization (CR-PSO) method for distributing several independent jobs among VMs in a cloud computing context. This hybrid methodology integrates the merits of Chemical Reaction Optimization (CRO) and Particle Swarm Optimization (PSO), producing an optimum task scheduling sequence that accounts for both job requirements and deadlines. Hybridization optimizes makespan, decreases costs, and diminishes energy use. Comprehensive simulations were performed with the CloudSim tools, illustrating the algorithm's efficacy. The comparative examination of several scenarios, varying quantities of VMs and tasks, demonstrates a decrease in execution time between 1–6% and, at times, exceeding 10%. Furthermore, the CR-PSO algorithm boosted makespan by 5–12%, decreased costs by 2–10%, and improved energy efficiency by 1–9%. These findings validate the algorithm's capacity for enhanced resource management and scheduling in cloud systems.

Mangalampalli, et al. [18] developed a task scheduling system using Firefly Optimization, prioritizing jobs and VMs to guarantee precise scheduling. This methodology utilizes synthetic datasets with diverse distributions and workloads from NASA and HPC2N for assessment. This methodology, implemented in the CloudSim simulation environment, is contrasted with baseline methods like genetic, Ant Colony Optimization (ACO), and PSO algorithms. The simulation outcomes indicate that the firefly-based method substantially surpasses these benchmarks in minimizing makespan, augmenting resource availability, increasing the success rate, and decreasing turnaround time, thus producing a more dependable and effective scheduling solution for cloud environments.

Bezdan, et al. [19] suggested an enhanced bat algorithm to tackle multi-objective job scheduling in cloud settings. The strategy seeks to optimize efficiency while minimizing search duration. The methodology was assessed with the CloudSim toolbox on both regular and synthetic parallel workloads. The findings revealed that the hybridized bat algorithm surpasses conventional metaheuristic methods, highlighting its significant potential for enhancing job scheduling effectiveness.

Wu and Xiong [20] created an innovative job scheduling approach for cloud computing with PSO algorithm. Initially, the resource scheduling issue in a cloud computing ecosystem is simulated, and a task execution duration function is established. The updated PSO approach is then implemented to coordinate application activities and improve load distribution. It relies on the Copula algorithm to explore the correlation between variables and probability while defining the attractor component to prevent the objective function from being ensnared in local optimums. The analysis indicates that the proposed resource allocation and scheduling methodology may enhance cloud computing resource usage and decrease job completion time.

Mangalampalli, et al. [21] offered a multiple-objective task scheduling method based on the Grey Wolf Optimization (MOTSGWO) algorithm by optimizing scheduling options dynamically depending on resource availability and anticipated demand requirements. This approach allocates resources to meet customer budgets and work priorities. The MOTSGWO methodology is executed through the Cloudsim toolkit, with workloads generated via the development of datasets with varied task distributions and sequences sourced from NASA and HPC2N distributed repositories. The comprehensive evaluation findings reveal that MOTSGWO is superior to previous benchmark strategies and improves critical metrics.

Saif, et al. [22] presented a multi-goal GWO algorithm aimed at minimizing the QoS targets of latency and energy usage implemented inside the fog broker, which is crucial to job distribution. The experimental observations confirm the efficacy of the MGWO algorithm relative to contemporary algorithms in minimizing delay and energy consumption.

III. PROBLEM DEFINITION

As shown in Fig. 1, a cloud data center comprises numerous Physical Machines (PMs), each capable of providing distinct end-user services. PMs can generate thousands of VMs dynamically. Alternatively, multiple host machines can collaborate to support a single VM. Cloud service providers offer VMs with different performance and pricing options, meeting a wide range of user requirements. This study explores the problem of allocating VMs to incoming, independent tasks. Each task is assumed to run exclusively on one VM and cannot be partitioned into smaller segments. Managing task scheduling in such an environment featuring varying capabilities is a complex challenge, represented by the sets of tasks and VMs in Eq. (1) and Eq. (2).

$$T = \{t_1, t_2, t_3, \dots, t_n\} \quad (1)$$

$$VM = \{vm_1, vm_2, vm_3, \dots, vm_m\} \quad (2)$$

The set T represents tasks, each defined by a specific number of instructions. At the same time, VM denotes a set of VMs, each with defined computational power determined by Millions of Instructions Per Second (MIPS). In most cases, the workload volume surpasses the available VMs. These arrays act as inputs to the scheduling algorithm, which seeks to derive an optimal mapping of tasks to VMs. This mapping outlines the assignment of tasks to VMs, as expressed in Eq. (3).

$$Map = \{(t_2, vm_1), (t_1, vm_3), (t_3, vm_2), \dots, (t_n, vm_m)\} \quad (3)$$

In the mapping solution, each task (represented as the first element of a tuple) is assigned uniquely to a VM, while a VM can be associated with multiple tasks. This implies that each task is allocated to a particular VM, while a VM may handle several assignments. The execution time (ET) for a specific task t_i on a VM vm_j is calculated using Eq. (4).

$$ET_{task_i vm_j} = Number\ of\ instruction\ in\ t_i\ /vm_j\ MIPS \quad (4)$$

It is assumed that each VM processes multiple tasks sequentially without interruption. Eq. (5) defines the overall completion time (CT) for all tasks allocated to a particular VM. Notably, faster VMs will complete their assigned tasks faster than slower ones.

$$CT_{vm_j} = \sum_{i=1}^{n} number\ of\ instrucions\ in\ t_i\ /vm_j\ MIPS \quad (5)$$

In metaheuristic algorithms, the arrangement of tasks on VMs is continuously adjusted to optimize their fitness values. As such, the task assignments on each VM may change during the algorithm's execution. If task t_x is replaced with task t_y on a VM, the completion time for multiple tasks executed in the VM is computed using Eq. (6).

$$CT_{vm_j} = CT_{vm_j} - (number\ of\ instrucions\ in\ t_x\ /vm_j\ MIPS) + (number\ of\ instrucions\ in\ t_y\ /vm_j\ MIPS) \quad (6)$$

A critical metric in task scheduling is the makespan, which refers to the total time required to complete all tasks across the available VMs. The makespan is determined using Eq. (7), representing the maximum completion time among all VMs involved in the scheduling process.

$$Makespan = \max\left(CT_{vm_j}\right)\ \forall j \epsilon 1,2, \dots, k \quad (7)$$

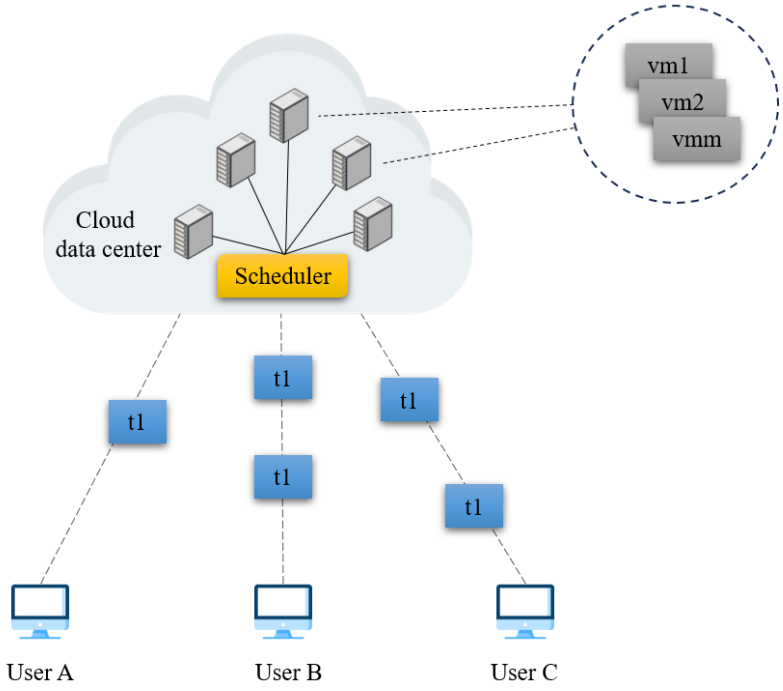


Fig. 1. Task scheduling process in cloud computing.

Throughout this study, makespan, completion time, and execution time are measured in seconds. Maximizing resource utilization during task scheduling is advantageous, ensuring a resource is fully utilized before allocating another instance on the cloud. Eq. (8) calculates the Average Resource Utilization (ARU) for PMs. This involves summing up the completion times of all VMs and dividing by the total number of VMs (m), followed by dividing the result by the makespan.

$$ARU = \left(\sum_{i=1}^{m} CT_{vm_j}/m\right)/makespan \quad (8)$$

System efficiency is further assessed through throughput, defined as the number of tasks processed per unit of time, calculated using Eq. (9). It is determined by dividing the total number of tasks by the makespan, resulting in a throughput value measured in terms of tasks accomplished per second.

$$Throughout = Total\ number\ of\ tasks /makespan \quad (9)$$

Additionally, Response Time (RT) is a key metric representing the duration between the scheduling decision and the initiation of task execution on a VM. Multiple VMs may operate on the identical PM, and a single VM may perform several tasks. Eq. (10) describes the mean response time for all tasks across multiple VMs. This calculation involves dividing the total initiation intervals for multiple tasks into the overall task count to determine the mean response time for each VM. The overall average response time is then obtained by averaging the response times across all VMs.

$$RT = \left(\sum_{j=1}^{m}\sum_{i=1}^{n} RT_i\right)/m \quad (10)$$

IV. PROPOSED METHOD

BOA mimics the foraging behavior of butterflies, driven by their sense of smell (olfaction). It aims to resemble how butterflies locate food sources (flowers) by navigating their environment using global and local search capabilities. In cloud computing and task scheduling, BOA facilitates allocating tasks to VMs by finding near-optimal solutions through iterative improvements. BOA is characterized by three central concepts: olfactory modality, global and local search, and scent intensity.

In BOA, butterflies are represented as solutions, and their sense of smell is modeled to guide their movement. Each butterfly (solution) is attracted to the most promising areas (best solutions), allowing for effective exploitation and exploration of the search area. The algorithm alternates between global search (exploration) and local search (exploitation). Global search occurs when butterflies move towards a global highest-quality solution. The solution space can thus be explored in new ways. Local search is triggered when butterflies move toward other butterflies nearby, allowing solutions to be fine-tuned.

The effectiveness of butterflies' movement depends on the intensity of the scent, which changes based on their position in the search space. The scent is calculated using fitness functions, which assess how effective a given solution is concerning the objective by cutting down makespan or optimizing resource utilization. The scent intensity S_i of each butterfly i is calculated using Eq. (11):

$$S_i = c.f_i^a \quad (11)$$

where c represents a sensory modality that affects scent strength, f_i^a stands for the fitness value of butterfly i, and a controls the nonlinearity of scent. The movement of a butterfly i towards a global best solution (global search) is given by:

$$x_i^{t+1} = x_i^t + r.S_g.(g^t - x_i^t) \quad (12)$$

Where x_i^t is the position of the butterfly i at iteration t, r stands for random number in the range [0, 1], S_g refers to the scent intensity of the best solution found so far (global best), and g^t specifies the global best position at iteration t. The movement of a butterfly towards another butterfly (local search) is represented as:

$$x_i^{t+1} = x_i^t + r.S_j.\left(x_j^t - x_i^t\right) \quad (13)$$

where S_j is the scent intensity of butterfly j, and x_j^t is its position at iteration t. The balance between global and local search is controlled using a probability parameter p. A random number $r \in [0,1]$ is compared to p to determine whether a butterfly moves towards the global best or another butterfly as follows:

- If $r<p$, the butterfly follows the global search strategy.
- If $r \geq p$, it engages in local search.

The task scheduling problem is treated as an optimization challenge in cloud computing. The objective is to reduce makespan, minimize energy consumption, and improve resource utilization by finding the best allocation of tasks to VMs. Each butterfly represents a potential solution, where:

- A solution is a specific mapping of tasks to VMs.
- The fitness of a solution is rated in terms of makespan, energy consumption, and other QoS metrics.

During each iteration, BOA adjusts butterflies' positions according to their scent intensities and the best solutions determined so far. Continuous iterations allow the algorithm to converge towards an optimal or near-optimal task allocation strategy.

The conventional BOA faces three primary challenges: (1) a fixed exploration-to-exploitation ratio controlled by the parameter p, leading to rigidity in the search process; (2) the possibility of being stuck in local optima due to a fixed global best attraction; and (3) pairwise interaction between butterflies, which limits search efficiency in complex optimization problems. The Fuzzy Butterfly Optimization Algorithm FBOA) overcomes these drawbacks by introducing (a virtual butterfly and fuzzy decision-making strategy. Fig. 2 compares conventional BOA and FBOA in terms of their exploration and exploitation strategies. FBOA integrates fuzzy logic to adjust the transition between these strategies.

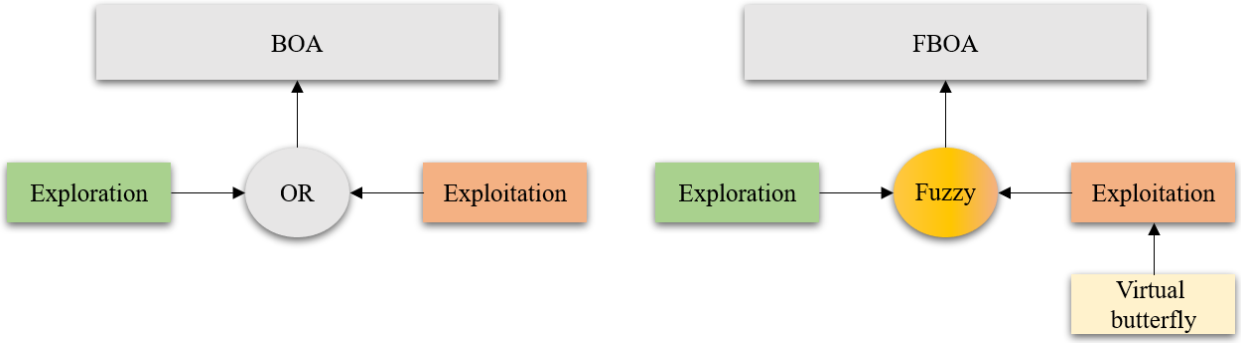


Fig. 2. Structure of BOA and FBOA.

To improve the adaptability of the BOA, FBOA employs a nine-rule fuzzy decision-making strategy. This strategy dynamically adjusts the tendency factor τi for each butterfly, which dictates its balance between exploitation and exploration based on the current optimization context. A novel concept called the virtual butterfly X^v is introduced, which serves as a guiding agent in navigation. Unlike the standard BOA's random pairwise interactions, the virtual butterfly uses the information from the best solutions and adjusts its direction based on the current problem's objective function. The Normalized Objective Function (NOF) is calculated to estimate the relative effectiveness of each butterfly's position:

$$NOF_i = \frac{f(X_i) - f(g^*)}{f(X^{worst}) - f(g^*) + \mu}, \quad i = 1,2,\dots,M \qquad (14)$$

Where $f(X_i)$ is the objective function value of the i^{th} butterfly, $f(g^*)$ is the fitness value of the current global best butterfly, $f(X^{worst})$ is the fitness value of the worst butterfly, μ is a small positive scalar to avoid division by zero, and M is the population size. The fuzzy decision system updates the tendency factor τi using the NOF and predefined fuzzy rules.

$$\tau_i = \omega_\tau + \Delta\tau_i \qquad (15)$$

where ω_τ indicates the origin of the tendency factor, $\Delta\tau_i$ is the adjustment value derived through the fuzzy inference process. Membership functions (as shown in Fig. 3) are used to categorize NOF into linguistic variables such as Small (S), Medium (M), and Large (L). The output values are adjusted based on the rules provided in Table I.

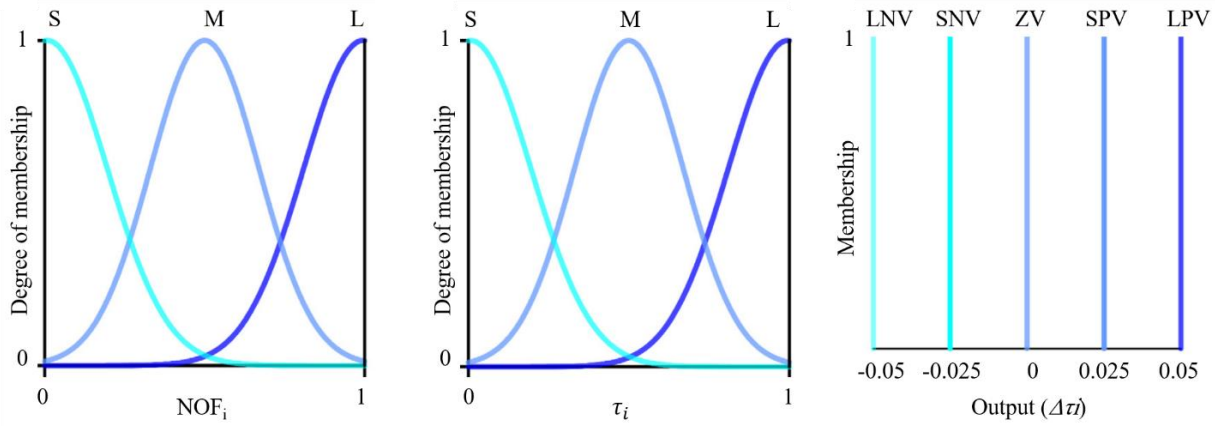


Fig. 3. Membership functions.

TABLE I. FUZZY RULES

Rules	Inputs		Output
	τi	NOF_i	
1	L	L	LNV
2	M	L	ZV
3	S	L	LPV
4	L	M	SNV
5	M	M	ZV
6	S	M	SPV
7	L	S	ZV
8	M	S	SPV
9	S	S	SNV

The updated movement rule in FBOA is formulated to incorporate the virtual butterfly and the fuzzy-adjusted tendency factor:

$$X_i^{t+1} = X_i^t + q.(\tau_i^2.g^* - X_i^t) + a.r.(\tau_i^2.X^v - X_i^t) \quad (16)$$

where g^* is the global best position, X^v is the position of the virtual butterfly, q is a random number in the range [0, 1], and α is the coefficient determining the impact of the virtual butterfly. The value of α is determined by the fitness comparison between X^v and X_i:

$$a=1 \quad if\, f(X^v) < f(X_i), \quad a = -1 \;\; if\; f(X^v) \geq f(X_i) \quad (17)$$

The virtual butterfly X^v aggregates information from the entire population using the weighted average of butterfly positions:

$$X^v = \sum_{i=1}^{M} X_i^v c_i^v \quad (18)$$

where c_i^v is a normalized weight defined as:

$$c_i^v = \frac{exp(\xi_i^v)}{\sum_{j=1}^{M} exp(\xi_j^v)} \quad (19)$$

ξ_i^v is the normalized fitness difference for butterfly i:

$$\xi_j^v = \frac{f(X_i^v) - f(X^{worst})}{f(X^{worst}) - f(g^*) + \mu} \quad (20)$$

FBOA addresses the task scheduling challenge in cloud computing environments by dynamically balancing exploration and exploitation through a fuzzy decision-making system. By adjusting the tendency factor for each butterfly using fuzzy logic, FBOA adapts its search behavior to optimize the allocation of tasks to VMs. This adaptability ensures that FBOA can efficiently handle the complex solution area of task scheduling, improving resource utilization, shortening makespan, and reducing energy usage. The integration of a virtual butterfly concept further enhances the algorithm's potential to overcome local optima, leading to more effective and balanced scheduling solutions. Through iterative adjustments, FBOA ensures that cloud resources are optimally allocated, providing better service quality and meeting the diverse demands of cloud users.

V. Performance Evaluation

The tests were performed on an Intel Core i5-12400 system featuring a 2.50 GHz processor and 16 GB of RAM. The effectiveness of the suggested FBOA was examined using the CloudSim 3.0.3 simulation toolkit and datasets from the Heterogeneous Computing Scheduling Problem (HCSP) and Google Cloud Jobs (GoCJ). The data center comprises a single entity equipped with 12000 MIPS processing capacity. It hosts three types of machines, each with varying cores, quad-core, hexa-core, and octa-core, and supports a memory range of 512 MB to 14436 MB. For the GoCJ dataset, the configuration includes ten VMs with MIPS capacities ranging from 400 to 12000 MIPS. For the HCSP instances, 32 VMs are distributed. This setup allows for a comprehensive analysis of task scheduling performance across diverse resource capacities and task complexities.

In cloud computing, makespan is a critical metric as it measures the total time required to complete all tasks on a set of VMs. A lower makespan indicates more efficient scheduling, allowing servers to process workloads quickly. As shown in Fig. 4, FBOA effectively minimizes makespan, particularly as the number of tasks increases, through its optimized task-to-VM mapping strategy. Compared with other methods such as Whale Optimization Algorithm (WOA) [23], Random Matrix Particle Swarm Optimization (RMPSO) [24], Security- and Energy-Aware (SAEA) [25], and Genetic Algorithm with MapReduce (GAMR) [26], FBOA demonstrates the smallest average increase in makespan (4.3%), illustrating its superior scalability in handling the rising number of tasks across 19 GoCJ instances. This makes FBOA highly suitable for dynamic cloud environments where workload demands fluctuate. For HCSP tasks, a similar trend was observed, with FBOA consistently achieving a reduced makespan across varying instances, as depicted in Fig. 5. This indicates that the algorithm adapts well to heterogeneous task categories, achieving efficient resource allocation.

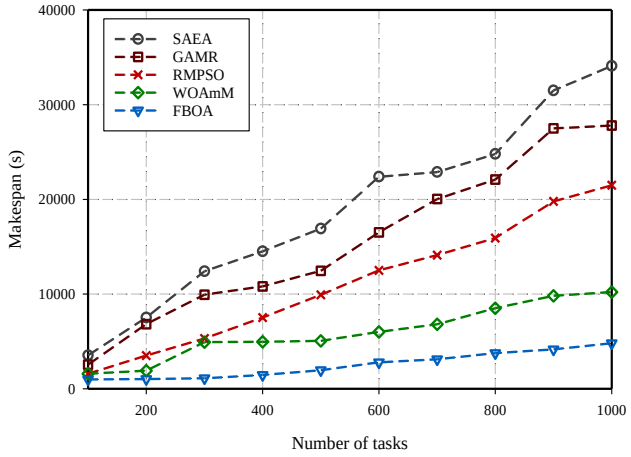


Fig. 4. Makespan for GoCJ dataset.

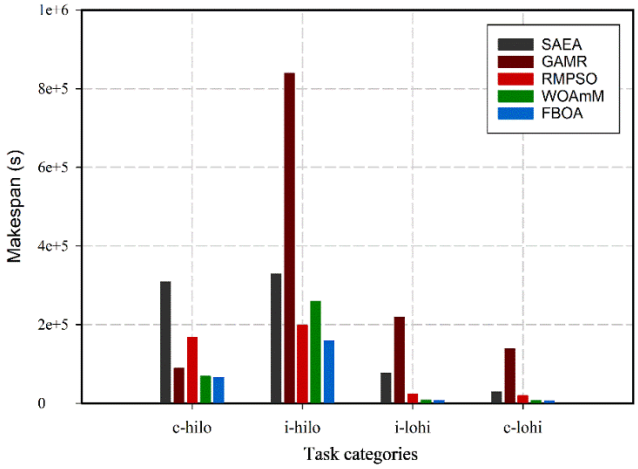


Fig. 5. Makespan for HCSP dataset.

Effective resource utilization ensures that the cloud infrastructure makes the most efficient use of available resources. FBOA achieved the best resource utilization over

other algorithms benchmarked, such as WOA, MRMPSO, and GAMR, as illustrated in Fig. 6 and Fig. 7. By dynamically adjusting its search mechanisms, FBOA efficiently managed VMs, reducing idle time and enhancing overall resource allocation. This is crucial in cloud environments where minimizing waste and optimizing VM usage can save significant costs. Among all methods, FBOA stood out for its ability to balance the workload, resulting in consistent resource usage, even with complex task distributions.

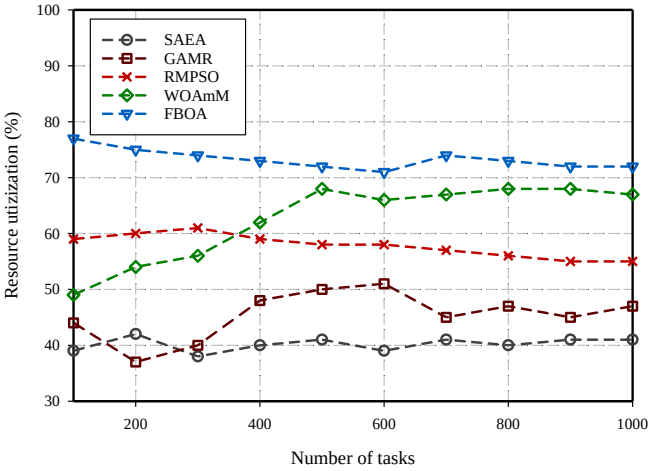


Fig. 6. Resource utilization for GoCJ dataset.

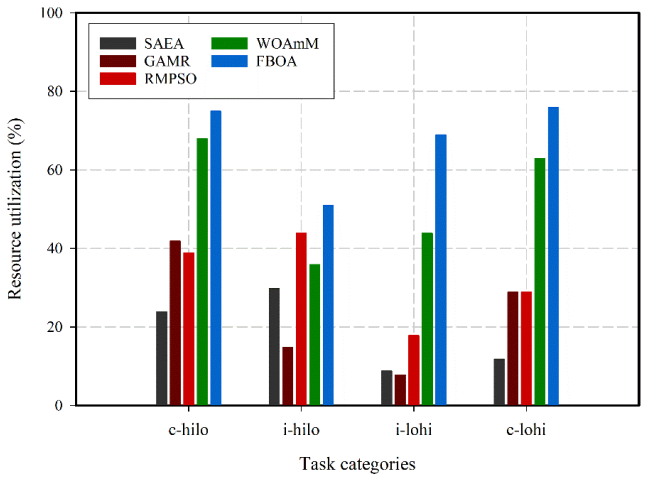


Fig. 7. Resource utilization for HCSP dataset.

Throughput, representing the number of tasks accomplished in a given period of time, serves as a measure of system efficiency. The results showed that FBOA achieved the highest throughput, which can be attributed to its parallel processing capabilities and refined task scheduling mechanism. Fig. 8 and Fig. 9 show a marked improvement in throughput for FBOA, especially when handling tasks with varying complexities. The enhanced throughput rates indicate that FBOA can efficiently manage large-scale task scheduling, making it well-suited for cloud infrastructures with fluctuating demands.

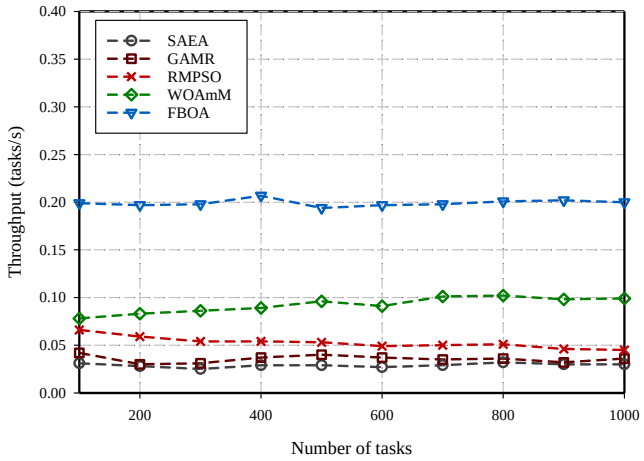


Fig. 8. Throughput utilization for GoCJ dataset.

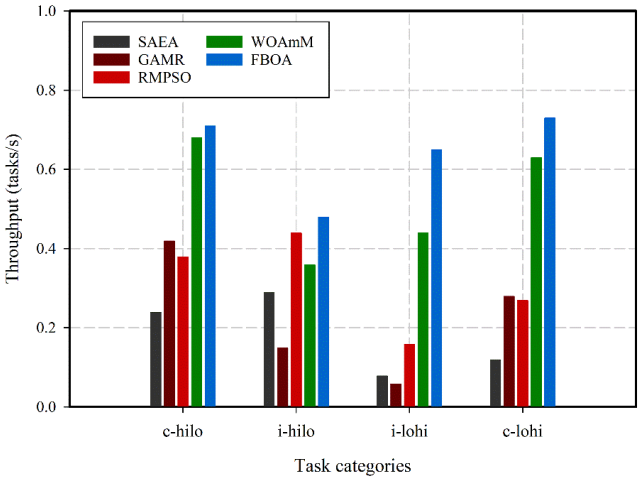


Fig. 9. Throughput utilization for HCSP dataset.

VI. Discussion

Compared to existing alternatives, the suggested FBOA significantly improves task scheduling problems. This success in the FBOA highlights the novelties achieved in coupling enhancements, a fuzzy decision strategy, and the idea of a virtual butterfly when overcoming deficiencies in BOA.

The fuzzy decision-making mechanism dynamically balances the phases of exploration and exploitation. Such adaptability allows the algorithm to cope with complex search spaces and prevent it from converging too early, guaranteeing robust optimization even for more complicated tasks. A proper example is that FBOA improves the makespan results on both GoCJ and HCSP datasets, proving that it can efficiently map tasks into VMs with good scalability.

The introduction of the concept of the virtual butterfly enables the central agent to gather and disseminate information in the swarm regarding the ongoing search process, orienting the same toward the most promising regions. This feature enhances

the solution quality and convergence rates, which is evident in the superior performance of FBOA in terms of the minimization of resource wastage and higher throughput value. FBOA maximizes the efficiency of a cloud environment since it ensures equal workload distribution and optimal utilization of resources.

These results hold important implications for cloud computing. With reduced makespan, improved resource utilization, and higher throughput, FBOA can contribute to creating efficient and scalable task scheduling solutions. For instance, such enhancements in makespan, resource utilization, and throughput shall enable service providers to reduce costs and improve user reliability, rendering cloud computing infrastructures more viable and competitive.

Furthermore, the adaptiveness of FBOA towards dynamic workloads and heterogeneous resources makes it likely to enable dynamic and real-time applications such as IoT ecosystems and high-performance computing. Therefore, contributions in this paper open new paths for further research in advanced cloud management intelligence and thus set a roadmap for optimizations in this rapidly changing field.

These benefits of FBOA are partially outweighed by several shortcomings. First, the virtual butterfly mechanism may lead to a significant increase in computation overhead in scenarios with extremely high heterogeneity among tasks or VMs. Further, this algorithm has only been experimented with in simulations with certain datasets. The real implementation could bring in unforeseen challenges, as scalability and adaptability may arise in infrastructures that are more dynamic or involve multi-clouds.

VII. Conclusion

In this study, we proposed FBOA to cope with the complexity of task allocation in cloud computing setups. By integrating fuzzy logic into the standard BOA, FBOA dynamically balanced exploration and exploitation, adapting to varying workload demands and resource availability. This adaptability ensured tasks were scheduled efficiently across VMs, minimizing makespan, reducing resource wastage, and improving overall system performance. The effectiveness of the proposed FBOA was validated through comprehensive simulations using the CloudSim toolkit, employing diverse datasets. The results demonstrated that FBOA consistently outperformed other metaheuristic algorithms such as WOA, RMPSO, MRMPSO, SAEA, and GAMR across critical metrics, including execution time, response time, throughput, resource utilization, and makespan. Notably, FBOA achieved lower increases in makespan, better resource allocation, and higher throughput, making it a robust and scalable solution for cloud environments. The superior performance of FBOA results from its ability to adjust the search behavior using a fuzzy decision-making mechanism and the introduction of the virtual butterfly concept, which helps avoid premature convergence and improves solution diversity. These features allow FBOA to effectively respond to the varying demands of cloud computing workloads, ensuring efficient resource use while meeting SLAs. Future research could explore further enhancements to FBOA, such as hybridizing it with other optimization techniques or applying it to emerging cloud paradigms like edge and fog computing to extend its applicability and effectiveness further.

References

[1] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," Journal of Network and Computer Applications, vol. 97, pp. 23-34, 2017.

[2] M. Shoeibi, A. E. Oskouei, and M. Kaveh, "A Novel Six-Dimensional Chimp Optimization Algorithm—Deep Reinforcement Learning-Based Optimization Scheme for Reconfigurable Intelligent Surface-Assisted Energy Harvesting in Batteryless IoT Networks," Future Internet, vol. 16, no. 12, p. 460, 2024, doi: https://doi.org/10.3390/fi16120460.

[3] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy - efficient data fusion methods in the Internet of Things," Concurrency and Computation: Practice and Experience, vol. 34, no. 15, p. e6959, 2022.

[4] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective," Sustainability, vol. 15, no. 4, p. 3317, 2023.

[5] A. Rahman et al., "Impacts of blockchain in software - defined Internet of Things ecosystem with Network Function Virtualization for smart applications: Present perspectives and future directions," International Journal of Communication Systems, p. e5429, 2023.

[6] V. Hayyolalam, B. Pourghebleh, M. R. Chehrehzad, and A. A. Pourhaji Kazem, "Single - objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," Concurrency and Computation: Practice and Experience, vol. 34, no. 5, p. e6698, 2022.

[7] I. Behera and S. Sobhanayak, "Task scheduling optimization in heterogeneous cloud computing environments: A hybrid GA-GWO approach," Journal of Parallel and Distributed Computing, vol. 183, p. 104766, 2024.

[8] B. Godavarthi, N. Narisetty, K. Gudikandhula, R. Muthukumaran, D. Kapila, and J. Ramesh, "Cloud computing enabled business model innovation," The Journal of High Technology Management Research, vol. 34, no. 2, p. 100469, 2023.

[9] A. Katal, S. Dahiya, and T. Choudhury, "Energy efficiency in cloud computing data centers: a survey on software technologies," Cluster Computing, vol. 26, no. 3, pp. 1845-1875, 2023.

[10] J. Zhou et al., "Comparative analysis of metaheuristic load balancing algorithms for efficient load balancing in cloud computing," Journal of cloud computing, vol. 12, no. 1, p. 85, 2023.

[11] V. Hayyolalam, B. Pourghebleh, A. A. Pourhaji Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," The International Journal of Advanced Manufacturing Technology, vol. 105, pp. 471-498, 2019.

[12] S. Gupta and S. Tripathi, "A comprehensive survey on cloud computing scheduling techniques," Multimedia Tools and Applications, vol. 83, no. 18, pp. 53581-53634, 2024.

[13] M. Yadav and A. Mishra, "An enhanced ordinal optimization with lower scheduling overhead based novel approach for task scheduling in cloud computing environment," Journal of Cloud Computing, vol. 12, no. 1, p. 8, 2023.

[14] B. Pourghebleh, A. Aghaei Anvigh, A. R. Ramtin, and B. Mohammadi, "The importance of nature-inspired meta-heuristic algorithms for solving virtual machine consolidation problem in cloud environments," Cluster Computing, vol. 24, no. 3, pp. 2673-2696, 2021.

[15] S. Arora and S. Singh, "Butterfly optimization algorithm: a novel approach for global optimization," Soft computing, vol. 23, pp. 715-734, 2019.

[16] E. Bozorgi, S. Soleimani, S. K. Alqaiidi, H. R. Arabnia, and K. Kochut, "Subgraph2vec: A random walk-based algorithm for embedding knowledge graphs," arXiv preprint arXiv:2405.02240, 2024, doi: https://doi.org/10.48550/arXiv.2405.02240.

[17] K. Dubey and S. C. Sharma, "A novel multi-objective CR-PSO task scheduling algorithm with deadline constraint in cloud computing," Sustainable Computing: Informatics and Systems, vol. 32, p. 100605, 2021.

[18] S. Mangalampalli, G. R. Karri, and A. A. Elngar, "An efficient trust-aware task scheduling algorithm in cloud computing using firefly optimization," Sensors, vol. 23, no. 3, p. 1384, 2023.

[19] T. Bezdan, M. Zivkovic, N. Bacanin, I. Strumberger, E. Tuba, and M. Tuba, "Multi-objective task scheduling in cloud computing environment by hybridized bat algorithm," Journal of Intelligent & Fuzzy Systems, vol. 42, no. 1, pp. 411-423, 2022.

[20] Z. Wu and J. Xiong, "A novel task-scheduling algorithm of cloud computing based on particle swarm optimization," International Journal of Gaming and Computer-Mediated Simulations (IJGCMS), vol. 13, no. 2, pp. 1-15, 2021.

[21] S. Mangalampalli, G. R. Karri, and M. Kumar, "Multi objective task scheduling algorithm in cloud computing using grey wolf optimization," Cluster Computing, vol. 26, no. 6, pp. 3803-3822, 2023.

[22] F. A. Saif, R. Latip, Z. M. Hanapi, and K. Shafinah, "Multi-objective grey wolf optimizer algorithm for task scheduling in cloud-fog computing," IEEE Access, vol. 11, pp. 20635-20646, 2023.

[23] G. Narendrababu Reddy and S. P. Kumar, "Multi objective task scheduling algorithm for cloud computing using whale optimization technique," in Smart and Innovative Trends in Next Generation Computing Technologies: Third International Conference, NGCT 2017, Dehradun, India, October 30-31, 2017, Revised Selected Papers, Part I 3, 2018: Springer, pp. 286-297.

[24] X. Tang, C. Shi, T. Deng, Z. Wu, and L. Yang, "Parallel random matrix particle swarm optimization scheduling algorithms with budget constraints on cloud computing systems," Applied Soft Computing, vol. 113, p. 107914, 2021.

[25] B. M. H. Zade, N. Mansouri, and M. M. Javidi, "SAEA: A security-aware and energy-aware task scheduling strategy by Parallel Squirrel Search Algorithm in cloud environment," Expert Systems with Applications, vol. 176, p. 114915, 2021.

[26] Z. Peng, P. Pirozmand, M. Motevalli, and A. Esmaeili, "Genetic Algorithm - Based Task Scheduling in Cloud Computing Using MapReduce Framework," Mathematical Problems in Engineering, vol. 2022, no. 1, p. 4290382, 2022.

Leveraging Large Language Models for Automated Bug Fixing

Shatha Abed Alsaedi[1, *], Amin Yousef Noaman[2], Ahmed A. A. Gad-Elrab[3], Fathy Elbouraey Eassa[4], and Seif Haridi[5]
Department of Computer Science-Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia[1, 2, 3, 4]
Department of Computer Science-College of Computer Science and Engineering, Taibah University, Yanbu 46421, Saudi Arabia[1]
School of Electrical Engineering and Computer Science, KTH, Sweden[5]

***Abstract*—Bug fixing, which is known as Automatic Program Repair (APR), is a significant area of research in the software engineering field. It aims to develop techniques and algorithms to automatically fix bugs and generate fixing patches in the source code. Researchers focus on developing many APR algorithms to enhance software reliability and increase the productivity of developers. In this paper, a novel model for automated bug fixing has been developed leveraging large language models. The proposed model accepts the bug type and the buggy method as inputs and outputs the repaired version of the method. The model can localize the buggy lines, debug the source code, generate the correct patches, and insert them in the correct locations. To evaluate the proposed model, a new dataset which contains 53 Java source code files from four bug classes which are Program Anomaly, GUI, Test-Code and Performance has been presented. The proposed model successfully fixed 49 out of 53 codes using gpt-3.5-turbo and all 53 using gpt-4-0125-preview. The results are notable, with the model achieving accuracies of 92.45% and 100% with gpt-3.5-turbo and gpt-4-0125-preview, respectively. Additionally, the proposed model outperforms several state-of-the-art APR models as it fixes all 40 buggy programs in QuixBugs benchmark dataset.**

***Keywords*—Bug fixing; automated program repair; large language models; software debugging; software maintenance; machine learning**

I. Introduction

Automated Program Repair (APR) is considered one of the most ideal tasks in automated software engineering, with the potential to reduce the costs associated with software development and maintenance [32]. APR refers to the automated fixing of bugs or defects in the source code by a software tool [31]. It aims to minimize human intervention in the debugging process through the development and implementation of intelligent algorithms capable of automatically detecting and fixing errors in the source code.

Historically, APR techniques have relied on a variety of approaches, ranging from genetic algorithms [12-14] to symbolic execution [37], each with its strengths and limitations. However, the advent of Large Language Models (LLMs) in artificial intelligence has opened new directions for research and application in software engineering tasks [18, 33, 34], offering promising prospects for enhancing the accuracy, efficiency, and scope of automated bug fixes.

This paper introduces a novel model for APR that leverages the capabilities of two state-of-the-art LLMs, gpt-3.5-turbo and gpt-4-0125-preview, to automate the bug-fixing process across various bug types. The proposed model distinguishes itself by not only localizing bugs with high accuracy but also generating and inserting the correct patches into the source code. Accordingly, it significantly reduces the time and resources traditionally required for bug fixing, thereby accelerating the software development lifecycle, and enhancing developer productivity.

To validate the effectiveness of the proposed model, a new dataset consisting of 53 Java source code files, categorized into four distinct bug classes: Program Anomaly, GUI, Test-Code, and Performance was constructed. The performance of the proposed model was evaluated against this dataset to reveal its capability in bugs repair, with results indicating an impressive accuracy rate. These findings not only support the potential of integrating LLMs into the APR process but also open new directions for future research in the field.

The proposed model distinguishes itself from other APR models due to its ability to successfully debug code, localize buggy lines, generate correct patches, and insert them in the appropriate locations. It achieves this by only requiring the buggy code and the bug type, without needing additional user input, information about test cases, or prior knowledge of patch attempts. This makes it a more efficient, practical, and novel model.

The main contributions of this paper are as follows:

- Presenting a novelty in the use of gpt-4-0125-preview for APR: To the best of our knowledge, we present the first-of-its-kind model that leverages gpt-4-0125-preview for automated program repair, enabling the repair of multi-hunk and multi-fault bugs simultaneously. This significantly extends the capabilities of current APR models.
- Presenting a self-contained repair mechanism: Our approach does not rely on external test cases, prior patch knowledge, or feedback loops. Instead, it operates only with the buggy method and bug type, making it more efficient and less resource-intensive than traditional APR techniques.

- Illustrating an advanced fault localization: The proposed model identifies and fix buggy lines of code without requiring identification of a statement-level bug localization, reducing the manual effort typically needed in bug fixing.
- Building a new dataset: A new dataset comprising 53 Java programs across four bug categories: Program Anomaly, GUI, Test-Code, and Performance has been generated in this research.
- Presenting a comparison with state-of-the-art models: This paper has demonstrated that the proposed model not only achieves a higher bug fix rate than state-of-the-art APR models, but also it achieved this with fewer dependencies on external tools and inputs, making it a more scalable and practical solution for real-world use.

The rest of this paper is structured as follows: Section II outlines the motivation behind this research; Section III reviews existing literature on automated bug fixing algorithms; Section IV describes the proposed model; Section V details the experimental study; Section VI presents the experimental results; Section VII discusses the implications of these results; and Section VIII concludes the paper.

II. Motivation

To fix a bug in a software program, the location of the bug must first be identified. This includes the buggy class, the buggy method, and the buggy statement. Our previously proposed bug localization model [22] employs an information-retrieval-based approach to identify the buggy class and method within the class; however, it does not localize the buggy statement. On the other hand, spectrum-based fault localization (SBFL) can identify more precise locations, such as the buggy statement [19]. However, it requires a large number of passing and failing test cases with test oracles, which poses some limitations [19]. Furthermore, the performance of APR is influenced more by the quality of test cases than their quantity [19]. LLMs have shown significant improvements in software engineering, demonstrating exceptional performance in tasks such as code and document generation [6]. Consequently, this research proposes an automated bug fixing model that leverages a specific model of large language models, the gpt-4-0125-preview model. The main novelty in the proposed approach that it does not require test cases or oracles, nor does it require prior statement-level bug localization. It utilizes our previously proposed bug prediction model [23] to detect the bug type, identifies the buggy method via our localization model [22], and outputs the fixed version of the method, enhancing the accuracy of existing bug fixing models. The overall bug management system is presented in Fig. 1.

III. Related Work

Automated bug fixing, also known as automated program repair (APR), is a prominent research topic that has attracted significant interest from many researchers in the field. The literature includes many existing studies that present algorithms for automated bug repair. As illustrated in Fig. 2, these studies can be classified into four main categories: bug reports-based APR, constraint-based APR, search-based APR, and learning-based APR. More details about each category will be presented in the following paragraphs.

A. Bug reports-Based APR

Liu et al. [1] proposed R2Fix model for automatically generating bug-fixing patches using bug reports. Their model uses machine learning techniques, semantic patch generation techniques, and past fix patterns to automatic bug fixing. They evaluated their model for three bug types which are buffer overflows, null pointer bugs, and memory leaks. In the evaluation, they used three projects, the Linux kernel, Mozilla, and Apache. Their model generated 57 correct patches.

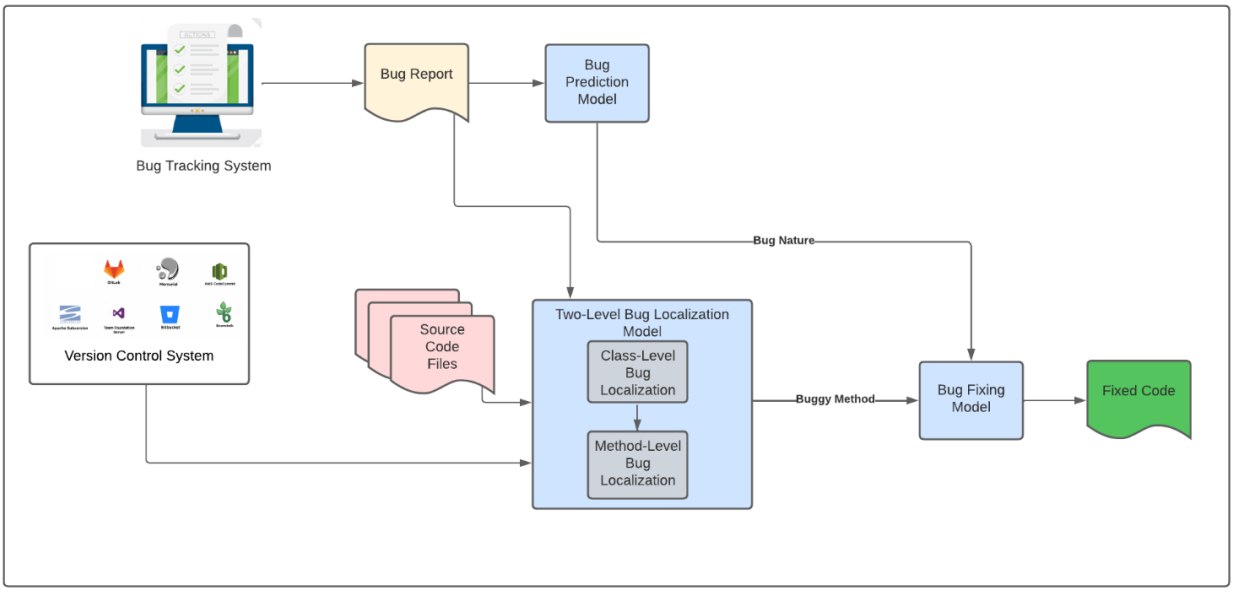


Fig. 1. Overall architecture of the bug management system.

Koyuncu et al. [2] proposed a novel model, namely iFixR, for bug localization and repair. In fact, their model is a bug repair system driven by bug reports. Their model uses bug reports as input. The main steps in their model are as follow: first, the bugs reports are fed to an information retrieval-based bug localizer; second, using fix pattern, patches are generated and validated by regression testing; third, patches are ordered by their priority for the developers. Their model did not have any assumption on the availability of test cases. To evaluate their model, they use and re-organize De-fects4J benchmark dataset and found that their proposed model can generate and recommend priority correct (and more plausible) patches for a wide range of issues reported by users.

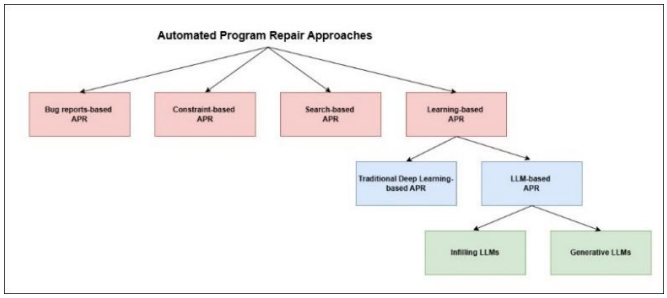


Fig. 2. Categories of existing APR approaches.

B. Constraint-Based APR

Constraint-based APR techniques use formal specifications and constraint solvers to transform the program repair problem into a constraint satisfaction problem. By focusing on expression-level variations and quickly pruning infeasible parts of the search space, these techniques can efficiently generate patches that satisfy the desired program behavior as specified by test cases or formal constraint [15].

Xuan et al. [3] proposed Nopol, a model for repairing faulty conditional statements, their model uses test cases as implicit program specifications. It generates patches by identifying expression-level changes that satisfy the constraints derived from the test cases.

Nguyen et al [4] introduced SemFix, a constraint-based method for APR, their method uses test cases to guide the patch synthesis process. The key features of SemFix include the use of symbolic execution and constraint solving to generate patches that ensure the program meets the desired behavior as specified by the test cases.

C. Search-Based APR

Goues et al. [14] presented GenProg, a generic and automated method for software repair, utilizing Genetic Programming to evolve and generate effective patches for a wide range of software defects. In the experiment, their method successfully fixed bugs in 16 programs written in C language which contain eight types of bugs. In their method, they presented three main novel ideas: Firstly, their method searches, in the program, at the statement level of the abstract syntax tree (AST). Secondly, to repair the bugs, their method does not introduce new code and uses only statements from the program. Lastly, genetic operators are localized to statements that are executed on the failing test case. However, their method does not support multi-hunk or multi-language program repair [15].

Yuan et al. [12] presented ARJA, a genetic programming-based approach for automated program repair in Java. ARJA suggests a new way of approaching automated program repair by treating it as a multi-objective optimization problem. They used a multi-objective optimization approach to minimize the weighted failure rate and patch size simultaneously, employing a multi-objective genetic algorithm (NSGA-II) to search for simpler repairs. The key innovation lies in designing a more detailed representation of patches, where the search spaces for likely-faulty locations, operation types, and ingredient statements are separated. This separation is intended to enhance the effectiveness of the genetic programming algorithm in finding appropriate solutions for fixing bugs in Java programs. The authors conducted a large-scale experimental study on both seeded and real-world bugs, demonstrating the effectiveness of ARJA in generating correct patches for a significant number of bugs compared to other repair approaches.

Yang et al. [13] presented a novel approach that uses similar bug fix information to automatic bug repair based on Genetic programming (GP). In their model, the candidate patches are generated by applying GP utilizing similar bug repair information. Then, the candidate patches are verified, by using a fitness function based on given test cases, if they are adoptable or not. Finally, the model generates the patch to fix the buggy code.

D. Learning-Based APR

Learning-based APR can be categorized further into two subcategories which are traditional deep learning-based APR, and Large Language Models-based (LLMs-based) APR.

1) Traditional deep learning-Based APR: Li et. al. [10] used deep learning algorithms in APR and proposed a two-level model, DLFix, which is based on deep learning algorithm that applies code transformation learning which learns from prior bug repairs and the surrounding code contexts of the fixes. The first tier contains an RNN model that is used to learn the context of bug fixes and the second tire uses the result from the previous tire as an additional weighting input learn the code transformations of the bug-fixing.

Lutellier et al. [16] presented a novel model, CoCoNuT, utilizes a new context-aware neural machine translation (NMT) architecture and an ensemble deep learning model which consists of combination of convolutional neural networks (CNNs) to automatically fix defects in multiple languages. The faulty source code and its surrounding context is separately represented using NMT. Their model utilizes CNNs in the hierarchical features extraction. However, their model cannot be used for multi-hunk bug fixing.

Huq et al. [17] proposed a novel sequence-to sequence model, Review4Repair, a deep learning-based approach that uses a neural machine translation (NMT) in APR. Their model uses the code review information to increase the performance of the automated bug fixing process.

DLFix, CoCoNuT, and Review4Repair all utilize context-aware strategies for patch generation and have the ability to fix multi-type bugs. However, all of them do not concern with the design of fault localization and work with the help of existing fault localization tools, and all do not have the capability of multi-hunk and multi-fault repair [15].

2) LLMs-Based APR: Recent advancements in large pre-trained LLMs offer a new direction for developing novel program repair models that do not rely on historical bug fixes [9]. Recently, some researchers proposed APR models based on LLMs. These models can be divided further into two sub-categories Infilling LLM-based APR, and Generative LLM-based APR.

Xia et al. [8] utilized CodeBERT [7], a pre-trained bimodal model designed for both programming languages (PL) and natural languages (NL), to propose a novel model for automated program repair (APR) called AlphaRepair. This model, which does not require retraining or fine-tuning on historical bug fixes datasets, represents the first cloze-style APR approach. Unlike traditional NMT tasks, AlphaRepair handles repair tasks as cloze tasks [15], aiming to predict the correct code based on its surrounding context [9]. Evaluation results showed that AlphaRepair can outperform state-of-the-art APR approaches. Mashhadi and H. Hemmati [20] presented a novel APR model which used CodeBERT [7] and fine-tuned it on ManySStuBs4J small and large datasets to generate fix patches for the buggy code. Evaluation results showed that their model can generate correct fixes in 19-72% of the cases.

Prenner & Robbes [21] presented a study which investigates the performance of Codex [11], a GPT language model that fine-tuned on GitHub code, in bug localization and fixing tasks. They used a dataset of 40 bugs in Java and Python and found that although Codex is not specifically trained for automated program repair task, it is effective for this task. Their observations also found that it is more effective at fixing Python bugs than Java bugs.

Xia and Zhang [5] presented ChatRepair, a conversational approach to automatically fixing bugs using ChatGPT. Their model takes as input relevant test failure information, enhancing bug-fixing capabilities of ChatGPT by learning from the failures and successes of previous patching attempts on the same bug. The model successfully fixed 162 out of 337 bugs from the Defects4j dataset.

Sobania et al. [18] presented an analysis of using ChatGPT for automated bug repair. In their study, they utilized 40 programs written in the Python programming language from the publicly available QuixBugs dataset to explore capabilities of ChatGPT in the bug fixing process. They found that providing ChatGPT with additional information (i.e., hints) about the bugs significantly improved its performance, resulting in a success rate that reached fixing 31 out of 40 bugs from the QuixBugs dataset. This performance outperforms several state-of-the-art APR models.

Despite the extensive development of various APR approaches, the field still faces significant challenges in improving the precision and generalizability of bug fixes across various programming environments. Existing models often rely heavily on extensive test suites, historical bug fixes, or detailed bug reports, which may not always be available or sufficiently comprehensive. Furthermore, many of these models struggle with complex bug fixes that require understanding detailed programming contexts or generating comprehensive patches. This paper aims to overcome some of current limitations by implementing a novel APR model that utilizes the advanced capabilities of LLMs such as gpt-3.5-trubo and gpt-4-0125-preview. Our approach reduces dependency on traditional inputs like test suites and historical fixes by directly interpreting the context of buggy code and generating appropriate fixes. By achieving this, the proposed model not only aims to enhance the accuracy and applicability of automated bug fixes but also to improve the repair process, making it more efficient and less reliant on extensive manual inputs. Thus, this paper aims to provide a foundation for future research in employing large language models to refine and expand automated repair algorithms.

IV. METHODOLOGY

This section illustrates the methodology that applied to develop the bug fixing model.

A. The Architecture of the Proposed Model

The architecture of the proposed model is illustrated in Fig. 3. This model takes the bug type and the buggy method as input and outputs the fixed version of the method. The bug types include Program Anomaly, GUI, Test-Code, and Performance. Program Anomaly bug refers to bugs in the source code files that occurs due to problems in the code [35] such as logical and syntax errors [28], GUI category refers to any bugs in the code that are related to the design of graphical user interface design or event handling [35], Test-Code bugs are occurs due to any problem which are related to the test code [35], while Performance bugs are related to the problem in the source code that are affect performance issues such as memory usage and memory leaks [35]. The two inputs are fed into the LLM using the prompt which is presented in the next subsection. In this research, two LLMs were tested to determine which one produces the highest accuracy in bug fixing for the generated dataset. These LLMs are gpt-3.5-turbo [24] and gpt-4-0125-preview [25]. As will be detailed in Section VI, gpt-4-0125-preview proved to be more accurate than gpt-3.5-turbo in terms of bug fixing and was therefore chosen for the proposed model.

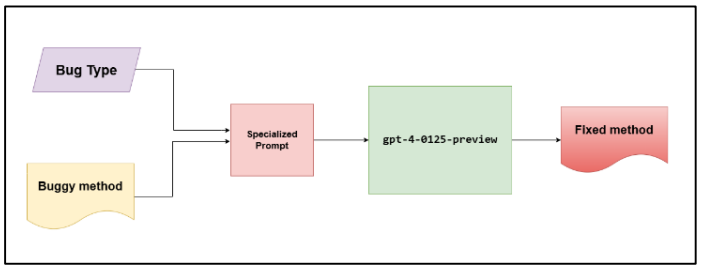


Fig. 3. High-level architecture of the proposed bug fixing model.

B. Large Language Models

LLMs are advanced deep learning algorithms capable of understanding, summarizing, translating, predicting, and generating content by leveraging large datasets [36]. In this research two LLMs from GPT family are used: the gpt-3.5-trubo model from GPT-3.5 and gpt-4-0125-preview model from GPT-4.

gpt-3.5-turbo [24]: is a variant of the GPT-3.5 language model. It is designed for enhanced performance and efficiency. This model is a group of models that improves upon GPT-3.5, enhances the ability to understand and generate both natural language and code.

gpt-4-0125-preview [25]: is a variant of the GPT-4 model. It is designed to give users a preview of the capabilities of the GPT-4 architecture. gpt-4-0125-preview performs tasks such as code generation more thoroughly than previous preview models. Additionally, this model is intended to reduce instances of "laziness" where the model does not fully complete a task [30].

C. Prompt Engineering

Models based on GPT utilize the paradigm of learning through prompts [26]. A prompt can be defined as a set of instructions that are given to the LLM, such as gpt-4-0125-preview, that program it by customizing it and/or improving or refining its capabilities [26]. Additionally, it is used to instruct the LLM to automate process or enforce rules [27]. Prompt engineering involves crafting clear, specific, and bias-mitigated prompts to guide the behaviour and output of AI models like GPT. It requires defining tasks, providing examples, and iterating on prompt design based on feedback to optimize results. Domain-specific knowledge may be necessary, and evaluation metrics help evaluate the effectiveness of the designed prompt.

The main contribution of this work lies in the application of LLMs for APR, and prompt engineering is an essential part of this process. The design of the prompt significantly influences the performance of the utilized LLM [6]. In the proposed model, the prompt is structured to instruct the LLM to accurately fix a buggy method. This prompt explicitly informs the LLM of the buggy method and categorizes the bug type. In the proposed model for automated bug fixing, a specific prompt template is utilized to guide LLMs, like gpt-3.5-turbo and gpt-4-0125-preview, to generate fixed versions of buggy Java methods. The template consists of several components: an objective statement, a placeholder for the incorrect Java function ({buggy_function}), and a detailed reason description placeholder ({reason_description}) that provides context about the type of bug, including categories such as GUI, Program Anomaly, Test-Code, and Performance. This context helps the model to better understand the issue and generate an appropriate fix. The template is structured as follows:

- **Objective Statement**: "Your goal is to correct the given Java function. The function will have a Javadoc that describes the function's purpose, parameters, and return value."
- **Function Placeholder**: {buggy_function}, which includes the buggy method which may has a Javadoc. The proposed model works effectively regardless of whether the method includes a Javadoc description.
- **Reason Description Placeholder:** {reason_description}, which explains the specific bug type, such as GUI issues related to layout problems or program anomaly involving logical errors.

The final prompt ends with "The correct Java function is:" prompting the model to generate the fixed function based on the provided details. This structured approach helps in automating the repair of code by leveraging the advanced capabilities of LLMs, thus enhancing the accuracy and efficiency of APR without requiring additional user input or test cases. This template plays a critical rule in achieving high accuracy rates in bug fixing, as demonstrated in the experiments conducted in this study.

V. Experimental Details

This section outlines the experimental setup and describes the implementation of the proposed bug fixing model. The experiments were conducted on a computer equipped with an x64 Intel® Core™ i7-10510U CPU @ 1.80GHz and 16.0 GB of RAM, running a 64-bit Windows Operating System. The model was developed using the Python programming language on Google Colab.

A. Dataset

To demonstrate the effectiveness of the proposed model, it is evaluated through two experiments. The first experiment involves evaluating of the proposed model using a new dataset. While in the second experiment, the proposed model is evaluated using a publicly available dataset to compare its performance against several state-of-the-art APR models.

1) Data collection and labeling: There are many reasons behind generating a new dataset in our research. First, there is no available dataset that contains source code files labeled in the same bug categories that are predicted by our bug prediction model. Secondly, we want to accurately evaluate our proposed model in a way that avoids information leakage from an existing dataset that might have already been seen by the LLM.

The presented dataset contains 53 Java source code files. These source code files were classified into four categories: Program Anomaly, GUI, Test-Code, and Performance. The Program Anomaly category contains bugs in the source code that occur due to logical and syntax errors. The GUI category refers to any bugs related to the design of graphical user interfaces or event handling. Test-Code bugs occur due to problems related to the test code. Performance bugs are related to issues in the source code that affect performance, such as memory usage, memory leaks, and missed performance improvements. The source code snippets used in this study were generated or collected from various sources available online. This involved identifying, collecting, and inspecting code snippets that contained bugs. The primary focus was on extracting examples that clearly demonstrated typical software

defects across different categories. These sources included open-source projects, coding forums, and educational resources. To ensure the quality and relevance of the dataset, each code snippet was carefully reviewed and categorized into one of the four main bug types mentioned above. This labeling process involved a detailed examination of each code snippet to identify the specific type of bug it represented. For instance, snippets involving inefficient coding practices or resource management issues were labeled as Performance Bugs, while those affecting the visual or functional aspects of the user interface were categorized as GUI Bugs. Similarly, errors within the test cases themselves were classified as Test-Code Bugs, and various coding errors leading to abnormal behavior or crashes were labeled as Program Anomaly Bugs.

This collaborative effort in data collection and labeling ensured a comprehensive and well-organized dataset, providing a solid foundation for analyzing common patterns and implications of different bug types in software engineering. Table I shows the number of Java source code files in each category. The Program Anomaly category contains 22 Java source code files, the GUI category contains 10 Java source code files, the Test-Code category includes 10 Java source code files, and the Performance category includes 11 Java source code files.

2) Analysis of the generated the dataset: The generated dataset consists of faulty source code files categorized into four main bug types: Performance, GUI, Test-Code, and Program Anomaly. Each category has distinct characteristics and common issues that are significant in understanding software bugs. A detailed analysis is presented for each bug type, their common patterns, and their effects in the following subsections.

a) Performance bugs: Performance bugs are mostly related to inefficient coding practices that reduce the runtime performance of applications. In the generated dataset, common issues include:

- **Inefficient String Operations:** Examples include using += in loops for string concatenation and creating new String objects unnecessarily.
- **Resource Management Issues**: Such as memory leaks from not clearing lists and inefficient use of wrapper classes leading to unnecessary boxing.
- **Control Flow Issues:** Infinite loops and concurrent modification exceptions which can cause applications to hang or crash.
- **Recursion Problems**: Recursive methods without proper termination can lead to stack overflow errors.
- **Other Issues:** Incorrect access patterns that can severely impact performance.

Effects: Performance bugs can cause significant degradation in application efficiency, leading to higher resource consumption and potential application failure. Identifying and optimizing these areas is crucial for enhancing software performance.

b) GUI Bugs: GUI bugs affect the usability and visual consistency of the application interface. Common GUI issues in the generated dataset include:

- **Visibility and Layout Problems:** Issues such as frames not being visible when they should be, incorrect component placements, and duplicate buttons in layouts.
- **Event Handling and Updates:** Bugs like missing initialization of components leading to null pointer exceptions, and incomplete status bar updates.
- **Drawing and Redrawing Inefficiencies**: Inefficient methods for redrawing components and missing drawing code.

Effects: GUI bugs can lead to poor user experience by making the interface confusing or non-functional. Proper testing and validation of the user interface components are essential to ensure smooth user interaction.

c) Test-Code Bugs

Test-Code bugs are errors within the test cases themselves which reduce the reliability of testing. Common issues in the generated dataset include:

- **Duplicate and Incorrect Test Methods:** Duplicate methods and logical errors in the methods being tested.
- **Uninitialized Variables:** Leading to compilation errors or incorrect test execution.
- **Incorrect Assertions:** Logical errors in assertions which lead to incorrect test outcomes.

Effects: Bugs in test code can lead to false positives or negatives, giving a misleading picture of the software quality. Ensuring the correctness of test code is as important as the application code itself.

d) Program anomaly bugs: Program Anomaly bugs involve a wide range of coding errors that result in abnormal behaviour or crashes. Common issues include:

- **Programming Errors:** Syntax errors and logical errors.
- **Control Flow Issues:** Infinite loops and incorrect loop conditions leading to unexpected behaviour.
- **Null Pointer and Type Safety Issues:** Potential null pointer exceptions and type mismatches.
- **Algorithmic and Recursive Errors:** Issues in algorithm implementation causing incorrect results or infinite recursion.

Effects: Program Anomaly bugs are critical as they often lead to crashes or incorrect program behaviour, significantly affecting the reliability and correctness of the software code. Comprehensive code review and rigorous testing are necessary to detect and fix these issues.

TABLE I. STATISTICS OF THE GENERATED DATASET

Bug Type Category	Total Number of Java Source Code Files in the category
GUI	10
Program Anomaly	22
Test-Code	10
Performance	11

3) QuixBugs benchmark dataset: The second experiment evaluates the proposed model using a publicly available benchmark dataset, QuixBugs[40], a collection of programs, each containing a specific bug, designed to test and evaluate the effectiveness of APR tools. The dataset includes 40 distinct algorithmic problems, such as sorting, graph traversal, and dynamic programming, implemented in both Java and Python. By providing a standardized set of challenges, QuixBugs allows researchers to consistently compare the performance of different APR methods. This dataset is widely used in academic research to advance the field of automated bug fixing. In the experiment, we use source code files written in Java only.

B. Evaluation Metrics

To evaluate the performance of the proposed model, the following evaluation metrics have been used.

- **Number of Repaired Defects**: It counts the defects which were successfully fixed by the repair algorithm. This metric is useful for demonstrating the capability of the algorithm. The diversity of defect classes in benchmark programs are crucial for the accuracy of this metric [29].
- **Repaired Defect Class:** Identifies specific classes of defects that the repair algorithm can successfully address. This helps in understanding the scope and specialization of the repair algorithm [29].
- **Success Rate (%):** The Success Rate of an automated program repair algorithm is quantitatively defined as the percentage of buggy programs that were successfully repaired by the algorithm out of the total number of buggy programs subjected to repair attempts. It is calculated using the Formula (1).

$$Success\ Rate\ (\%) = \left(\frac{Total\ Number\ of\ Successfully\ Fixed\ Programs}{Total\ Number\ of\ Buggy\ Programs}\right) \times 100 \quad (1)$$

VI. RESULTS

This section shows the results of the two experiments of the proposed bug fixing model on the generated dataset and on the benchmark, QuixBugs, dataset.

A. Results of the Proposed Model on the Generated Dataset

In the first experiment, the investigation of the achieved results by the proposed model explored the effectiveness of leveraging advanced language models, gpt-3.5- turbo and gpt-4-0125-preview, in automatically identifying and fixing bugs in software code. The analysis was structured around four main categories of bugs: GUI, Program Anomaly, Test-Code, and Performance. Table II shows the effectiveness of the proposed model in accurately repairing code from the used dataset.

The performance of the proposed model was evaluated based on its ability to correctly identify and fix these bugs. The proposed model can effectively localize the buggy lines, generate the correct patches, and insert them in the correct locations. The total accuracy of the proposed models (Success Rate) was calculated as follows: For the gpt-3.5-turbo model, the accuracy was 92.45%, while the gpt-4-0125-preview model achieved an accuracy of 100%. Fig. 4 shows an overview of the bug fixing results achieved by the proposed model.

TABLE II. RESULTS OF THE PROPOSED BUG FIXING MODEL

Bug Type	Correct Fixes by the Proposed Model (leveraging gpt-3.5-turbo)	Correct Fixes by the Proposed Model (leveraging gpt-4-0125-preview)
GUI	10 out of 10 (100%)	10 out of 10 (100%)
Program Anomaly	21 out of 22 (95.45%)	22 out of 22 (100%)
Test-Code	10 out of 10 (100%)	10 out of 10 (100%)
Performance	8 out of 11 (72.73%)	11 out of 11 (100%)
Total Fixes (Success Rate)	49 out of 53 (92.45%)	53 Out of 53 (100%)

B. Results of the Proposed Model on the QuixBugs Dataset

The proposed model that leverages gpt-4-0125-preview model has been evaluated using QuixBugs dataset. In more detail, the model achieved significant success, effectively repairing all 40 programs. This assessment covered a range of algorithmic problems, such as sorting, graph traversal, and dynamic programming. Utilizing advanced language model, gpt-4-0125-preview, our proposed APR model accurately detected and fixed bugs, demonstrating its robustness and efficiency.

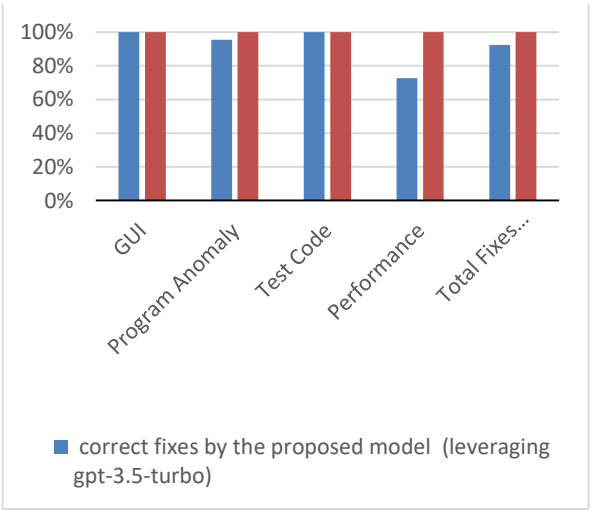


Fig. 4. An overview of the bug-fixing results achieved by the proposed model on the generated dataset.

VII. Analysis and Discussion

This section presents a comprehensive analysis and discussion of the results achieved by the proposed bug-fixing model in both experiments. Additionally, in this section, we answer the following research questions that address the improvements made by the proposed model in APR field.

- **Research Question 1**: What is the effect of different LLMs on the proposed bug fixing model on fixing different categories of programming bugs?
- **Research Question 2:** How does the proposed bug fixing model compare to several state-of-the-art APR models?
- **Research Question 3**: What are the practical improvements of integrating LLMs into the software development lifecycle for bug fixing?

A. Discussion of the First Experiment

To answer the first research question, we analyze the results of the first experiment.

Research Question 1: What is the effect of different LLMs on the proposed bug fixing model on fixing different categories of programming bugs?

Observations: Overall, the proposed model which leverages gpt-4-0125-preview demonstrates a 100% success rate, correctly fixing all 53 buggy source code files in the dataset. In contrast, the proposed model which leverages gpt-3.5-turbo achieves a 92.45% accuracy, successfully repairing 49 out of 53 buggy source code files. In the Program Anomaly category, the proposed model that leverages gpt-3.5-turbo fixed 21 out of 22 buggy source code files, while the proposed model that leverages gpt-4-0125-preview fixed all 22. The slight improvement with the latter model might be attributed to enhanced understanding or processing capabilities of gpt-4-0125-preview, possibly due to larger training data or improved algorithms. The most notable difference is observed in the performance category, where the proposed model that leverages gpt-3.5-turbo fixed 8 out of 11 bugs, whereas the proposed model that leverages gpt-4-0125-preview fixed all 11. This improvement may reflect advancements in the ability of gpt-4-0125-preview model to understand and optimize code for performance, a complex task that often requires deep under-standing and nuanced changes. Both models perform equally well in fixing GUI and Test-Code bugs, achieving a 100% success rate. This indicates robust capabilities in addressing issues within these specific categories, suggesting that the models have effectively learned patterns or solutions relevant to these types of bugs. Fig. 5 shows a buggy method in the dataset which is from GUI category. Fig. 6 and Fig. 7 show the fixed versions of this method using the proposed model (leveraging gpt-3.5-turbo model and gpt-4-0125-Preview model, respectively).

From this example, it is clear that both proposed models aim to address the issues found in the original code of the buggy method, but they do so with different ways. The buggy method has a bug in the loop that results in displaying incorrect keypad layout (i.e. it displays numbers from 1 to 12 instead of numbers from 1 to 9).

```
/**
 * This method  for GUI that resembles a telephone keypad.
 *
 *
 */

public void telephone() {
   JFrame frame = new JFrame();
   frame.setDefaultCloseOperation(JFrame.EXIT_ON_CLOSE);
   frame.setSize(new Dimension(250, 200));
   frame.setTitle("Telephone");

   frame.setLayout(new BorderLayout());

   // main phone buttons
   JPanel centerPanel = new JPanel(new GridLayout(4, 3));
   for (int i = 1; i <= 12; i++) {
   centerPanel.add(new JButton("" + i));
   }
   centerPanel.add(new JButton("0"));
   centerPanel.add(new JButton("0"));
   centerPanel.add(new JButton("0"));
   frame.add(centerPanel, BorderLayout.CENTER);

   // south status panel
   JPanel southPanel = new JPanel(new FlowLayout());
   southPanel.add(new JLabel("Number to dial: "));
   southPanel.add(new JTextField(10));
   frame.add(southPanel, BorderLayout.SOUTH);
  frame.setVisible(true);
   }
```

Fig. 5. Buggy method from GUI category.

```
/**
 * This method creates a GUI that resembles a telephone keypad.
 *
 * @param none
 * @return void
 */
public void createTelephoneGUI() {
    JFrame frame = new JFrame();
    frame.setDefaultCloseOperation(JFrame.EXIT_ON_CLOSE);
    frame.setSize(new Dimension(250, 200));
    frame.setTitle("Telephone");

    frame.setLayout(new BorderLayout());

    // main phone buttons
    JPanel centerPanel = new JPanel(new GridLayout(4, 3));
    for (int i = 1; i <= 9; i++) {
        centerPanel.add(new JButton("" + i));
    }
    centerPanel.add(new JButton("0"));
    centerPanel.add(new JButton("*"));
    centerPanel.add(new JButton("#"));
    frame.add(centerPanel, BorderLayout.CENTER);

    // south status panel
    JPanel southPanel = new JPanel(new FlowLayout());
    southPanel.add(new JLabel("Number to dial: "));
    southPanel.add(new JTextField(10));
    frame.add(southPanel, BorderLayout.SOUTH);

    frame.setVisible(true);
}
```

Fig. 6. Fixed method from GUI using the proposed model (leveraging gpt-3.5-turbo model).

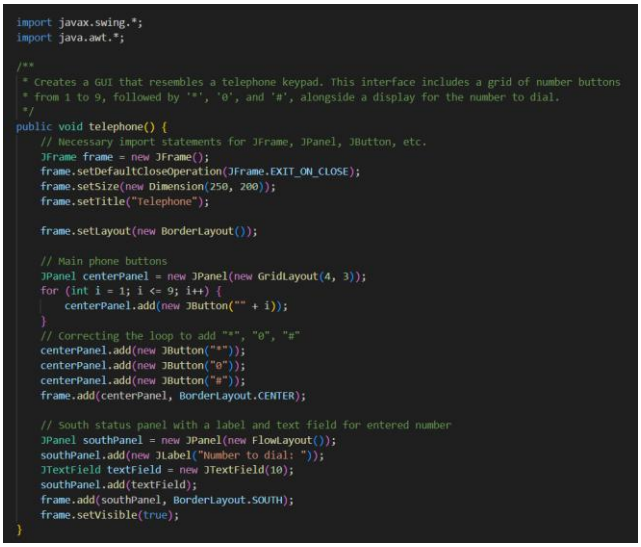


Fig. 7. Fixed method from GUI using the proposed model (leveraging gpt-4-0125-preview model).

Additionally, it incorrectly adds three "0" buttons instead of the conventional "*", "0", and "#" buttons found on a telephone keypad. Fig. 6 shows that the proposed model that leverages gpt-3.5-turbo model modifies the name of method to create TelephoneGUI, aligning with Java naming conventions by clearly describing the action performed by the method. This change emphasizes the importance of readability and maintainability in code. Additionally, it fixes the keypad layout to include "1" through "9", followed by "0", "*", and "#". Although the bug in the telephone keypad GUI has been successfully resolved, the arrangement of the buttons on the bottom row remains incorrect. The expected order of the buttons is * 0 #; however, the current implementation places them in the order 0 * #. Additionally, this model lacks the inclusion of import statements necessary for a standalone application, assuming prior context or placement within a larger source code. This approach addresses the discovered bugs and improves the naming of the method for better clarity. However, it fails to make the code independently usable by omitting necessary import statements. Fig. 7 shows the fixing code of the same buggy method using the proposed model that leverages gpt-4-0125-preview model which offers a comprehensive repairing by not only fixing the keypad layout but also adding essential import statements and a javadoc comment at the beginning of the method. This fixing assumes no prior context, aiming to make the code snippet independently compliable and understandable. The fixed version of the code follows best practices by including a detailed method description and fixing the keypad issue, which matches the standard layout of a telephone keypad. This model provides a more thorough fixing by ensuring that the code is both correct and self-sufficient. It addresses not only the initial logical issues but also enhances the readability and reusability of the code by adding documentation and necessary technical details for compilation.

Another example of bug fixing ability of the proposed model can be illustrated by Fig. 8-10. Fig. 8 presents a buggy code from the Test-Code category. The original buggy code presents a simple divide method intended to perform division operations while handling a potential division-by-zero error through an exception. However, the accompanying test cases contain several issues: incorrect assertions that do not match the expected outcomes of the division operation, duplication of test method names (which is not allowed in Java), and lack of a test case to explicitly check for the division by zero scenario.

Fig. 9 and Fig. 10 present the fixed versions of the buggy code by the proposed model in the two experiments, the first one when proposed fixing model leveraging gpt-3.5-turbo model and the second one when the proposed fixing model leveraging gpt-4-0125-preview model. Both models address the core issues in the original code, but they do so with varying levels of detail and adherence to best practices in software testing and documentation. Fig. 9 shows that leveraging gpt-3.5-turbo model in the proposed model fixes the logical bugs in the test cases and suggests a structured approach to handle different test scenarios, including dividing by zero. It provides clear, separate test methods with descriptive names that reflect their purpose: one for a successful division operation and another for testing division by zero. Improves readability and maintenance of the test suite by clearly separating test cases. Additionally, it correctly identifies the need for and implements a test case for division by zero, enhancing the robustness of the tests. However, in comparison with Fig. 10, the fixing is somewhat basic and does not explicitly address best practices in exception handling within test methods (e.g., using the expected annotation attribute).

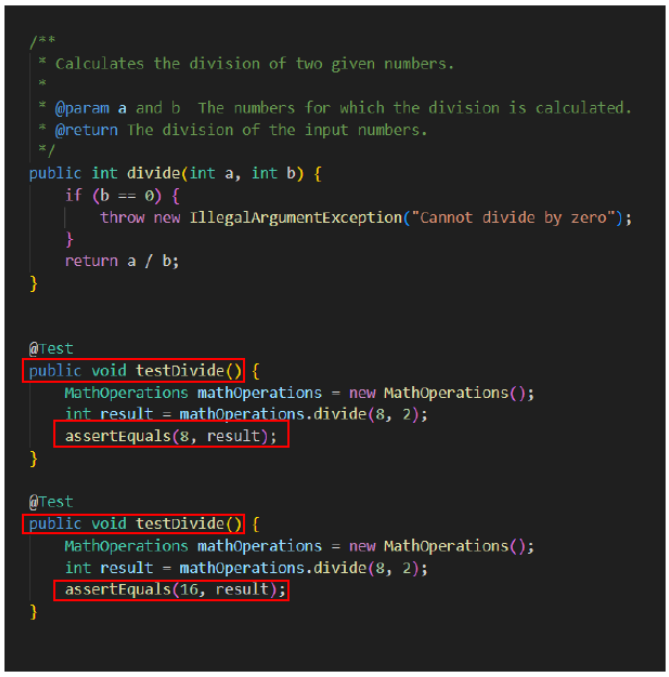


Fig. 8. Buggy code from Test-Code category.

Fig. 10 shows that leveraging gpt-4-0125-preview model in the proposed model offers a comprehensive fixing by not only repairing the test cases but also refining the testing approach to align with best practices. It fixes the bugs,

addresses method name duplication, and includes a detailed approach to testing exceptions using the expected attribute of the *@Test* annotation. This enhances the readability and maintainability of the test suite by demonstrating advanced testing techniques. However, the fixing assumes familiarity with the expected attribute for testing exceptions, which may require additional context for less experienced Java developers.

```java
/**
 * Calculates the division of two given numbers.
 *
 * @param a The numerator of the division.
 * @param b The denominator of the division.
 * @return The division of the input numbers.
 * @throws IllegalArgumentException if the denominator is zero.
 */
public int divide(int a, int b) {
    if (b == 0) {
        throw new IllegalArgumentException("Cannot divide by zero");
    }
    return a / b;
}

@Test
public void testDivide() {
    MathOperations mathOperations = new MathOperations();
    int result = mathOperations.divide(8, 2);
    assertEquals(4, result);
}

@Test
public void testDivideByZero() {
    MathOperations mathOperations = new MathOperations();
    assertThrows(IllegalArgumentException.class, () -> {
        mathOperations.divide(8, 0);
    });
}
```

Fig. 9. Fixed code from Test-Code category using the proposed model (leveraging gpt-3.5-turbo model).

```java
import static org.junit.Assert.assertEquals;
import org.junit.Test;

public class MathOperationsTest {

    /**
     * Calculates the division of two given numbers.
     *
     * @param a The dividend.
     * @param b The divisor.
     * @return The quotient from dividing a by b.
     */
    public int divide(int a, int b) {
        if (b == 0) {
            throw new IllegalArgumentException("Cannot divide by zero");
        }
        return a / b;
    }

    @Test
    public void testDivide_SuccessfulDivision() {
        MathOperations mathOperations = new MathOperations();
        int result = mathOperations.divide(8, 2);
        assertEquals(4, result); // Correcting the expected result to match 8 divided by 2.
    }

    // Example of how to test for division by zero scenario, if needed.
    @Test(expected = IllegalArgumentException.class)
    public void testDivisionByZero_ThrowsException() {
        MathOperations mathOperations = new MathOperations();
        mathOperations.divide(8, 0); // This should throw an IllegalArgumentException.
    }
}
```

Fig. 10. Fixed code from Test-Code category using the proposed model (leveraging gpt-4-0125-preview model).

The overall evaluation results provide a high-level view of the effectiveness of each model in the experiment in automated program repair. The best score achieved by the proposed model that leverages gpt-4-0125-preview is particularly significant, suggesting a notable advancement in the capability of the proposed model in understanding and fixing a wide range of software bugs.

The difference in accuracy between the two models (92.45% for the model leveraging gpt-3.5-turbo vs. 100% for the model leveraging gpt-4-0125-preview) could be attributed to several factors inherent to the evolution of generative pre-trained transformer models, such as increased model size, more diverse and extensive training data, or refined training techniques that improve understanding and generation capabilities.

B. *Discussion of the Second Experiment*

To answer the second research question, we analyze the results of the second experiment.

Research Question 2: How does the proposed bug fixing model compare to several state-of-the-art APR models?

Observations: To indicate the effectiveness of the proposed model, a comparison with the performance of several state-of-the-art APR models using the same dataset is crucial. Therefore, the achieved results of the proposed model on the QuixBugs dataset are compared to four studies which are:

- The results of the study presented by authors in the paper [39], involve an empirical study on automated bug repair using QuixBugs benchmark dataset.
- AlphaRepair [8] represents the first cloze-style APR approach, and it handles repair tasks as cloze tasks to predict the correct code based on its surrounding context.
- CoCoNut [16] which utilizes a new context-aware neural machine translation architecture and an ensemble deep learning model to fix buggy code.
- CURE [38] which is a novel APR tool that focuses on resolving software bugs through a sophisticated, context-aware neural machine translation (NMT) approach. Its main goal is to enhance the accuracy and effectiveness of bug fixes by utilizing detailed contextual information and robust neural network models. This allows CURE to generate more accurate patches by understanding not only the buggy code but also the surrounding context, significantly improving the quality and reliability of the fixes.

Table III illustrates the total number of QuixBugs programs out of 40 that are correctly fixed by the mentioned state of the art APR models and our proposed bug fixing model which leverages gpt-4-0125-preview model.

TABLE III. COMPARISON BETWEEN SEVERAL APR ON QUIXBUGS DATASET

Reference	Total Number of Correctly Fixed Programs	Success Rate
Ye et al. [39]	16 out of 40	40%
AlphaRepair [8]	28 out of 40	70%
CoCoNut [16]	13 out of 40	32.5%
CURE [38]	26 out of 40	65%
The Proposed Model	**40 out of 40**	**100%**

As observed from this table, our proposed model achieves the best results, successfully fixing all buggy programs in the dataset. Additionally, our proposed model does not need additional information about test cases, previous knowledge of patch attempts, nor additional follow up conversations which make it more efficient, practical and has its own novelty among other APR models as it only needs the buggy code and the bug type to debug the code, localized the buggy lines, generate the correct patches, and insert them in the right locations. Therefore, our proposed model outperforms all these state-of-the art APR models.

Furthermore, it was observed from this experiment that the proposed model not only fixed the identified bug, but also effectively enhanced the code by addressing the edge cases, aligning with Java naming conventions, and following better programming practices. In more details, Fig. 11 and Fig. 12 show the buggy version of TO_BASE.java program of QuixBugs dataset and its fixed version by the proposed model. Fig. 11 shows the buggy version of this program. The purpose of this Java program is to convert a given integer (num) into a string representation of that number in a specified base (b). The base can range from 2 to 36, allowing for conversion to binary, octal, decimal, hexadecimal, and other bases up to base 36, which includes digits 0-9 and letters A-Z. The main bug in this program is related to how the result string is constructed. In more detail, the code appends each new digit to the end of the result string, which results in the digits being in reverse order. Furthermore, there is no handling for the edge case where the input number is zero. From Fig. 12, the improvements made by the proposed model can be observed, which are:

- It fixes the bug; the fixed method reverses the resulting string ensuring the digits are in the correct order.
- It adds base validation; the fixed method includes a validation check to ensure the base is between 2 and 36. If the base is invalid, it returns an empty string.
- It adds zero handling; the fixed method correctly handles the case when num is zero, returning "0".
- It uses StringBuilder for efficiency; the fixed method uses StringBuilder for string concatenation, which is more efficient than repeatedly creating new string objects.
- It increases the clarity of the code; the fixed method is well-documented, explaining the purpose, parameters, and return value, making it easier to understand and maintain.

Therefore, the proposed model has significantly improved the original buggy method by adding necessary validations, handling edge cases, and improving performance through efficient string manipulation. These changes enhance the robustness of the method, efficiency, and readability, ensuring it works correctly across a wide range of inputs.

Furthermore, it was observed that the proposed model has the ability of multi-hunk and multi-fault repair. In more detail, the proposed model shows an ability for multi-hunk and multi-fault repair in software code. It can navigate and rectify issues spread across different parts of the source code (multi-hunk) and address several distinct types of faults within a single execution (multi-fault).

```java
import java.util.*;

/*
 * To change this template, choose Tools | Templates
 * and open the template in the editor.
 */

/**
 *
 * @author derricklin
 */
public class TO_BASE {
    public static String to_base(int num, int b) {
        String result = "";
        String alphabet = "0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZ";
        int i;
        while (num > 0) {
            i = num % b;
            num = num / b; // floor division?
            result = String.valueOf(alphabet.charAt(i))+result;
        }

        return result;
    }
}
```

Fig. 11. Buggy TO_BASE.java from QuixBugs.

```java
/**
 * Converts an integer to its representation in a specified base.
 *
 * @param num The number to convert.
 * @param base The base of the target number system (must be between 2 and 36 inclusive).
 * @return A string representation of the number in the specified base, or an empty string if the base is out of the valid range.
 */
public class TO_BASE {
    public static String to_base(int num, int base) {
        // Validate the base
        if(base < 2 || base > 36) {
            return "";  // Return an empty string for invalid base values
        }

        // Handle the case when num is 0
        if (num == 0) {
            return "0";
        }

        StringBuilder result = new StringBuilder();
        String alphabet = "0123456789ABCDEFGHIJKLMNOPQRSTUVWXYZ";

        while (num > 0) {
            int remainder = num % base;
            num = num / base;
            result.append(alphabet.charAt(remainder));
        }

        // Since the conversion process builds the result string in reverse order,
        // we need to reverse it before returning.
        return result.reverse().toString();
    }
}
```

Fig. 12. Fixed TO_BASE.java by the proposed model.

The proposed model that leverages gpt-4-0125-preview model, in particular, demonstrates a higher level of sophistication in applying best practices, suggesting an advanced understanding and capability in handling complex repair scenarios efficiently. This analysis supports the

utilization of LLMs in automating comprehensive code repair tasks, highlighting their role in supporting and enhancing software development and maintenance processes.

Research Question 3: What are the practical improvements of integrating LLMs into the software development lifecycle for bug-fixing?

Observations: Integrating LLMs into the software development lifecycle for bug fixing has several practical improvements include:

- **Increased Productivity:** Developers can focus on higher-level tasks, as LLMs handle routine bug fixing, leading to increased productivity and efficiency.
- **Improved Code Quality:** Automated bug fixing can lead to more consistent and reliable code quality, as LLMs can systematically apply best practices and coding standards.
- **Reduced Time-to-Market:** Faster bug detection and fixing reduce the overall development cycle time, enabling faster releases and update**s.**
- **Enhanced Collaboration:** LLMs can help in bridging the gap between different team members (e.g., developers and testers) by providing clear and actionable bug debugging and fixes.

VIII. CONCLUSION

This paper introduced a novel approach to APR, utilizing LLMs to automate the bug-fixing process. Through the leveraging of gpt-3.5-turbo and gpt-4-0125-preview models, a significant leap forward in the field of software engineering has been demonstrated, particularly in enhancing software reliability and developer productivity. The proposed model presents an ability to accurately localize bugs across various code segments, debug the source code, generate correct patches, and integrate these fixes into the appropriate locations within the source code. The evaluation of the proposed model, conducted on a diverse dataset comprising 53 Java source code files categorized into four distinct bug categories, confirms the efficiency of the proposed model. The results show that the gpt-3.5-turbo model achieved an impressive success rate, successfully repairing 49 out of 53 source code files, equivalent to an accuracy of 92.45%. In contrast, the gpt-4-0125-preview model exhibited an exceptional performance, achieving 100% success rate in bug fixing. Additionally, the proposed model was evaluated using the QuixBugs benchmark dataset, and it can correctly fix all its Java buggy programs. The proposed model was compared to several state-of-the-art APR models and outperformed them. Such outcomes not only highlight the robustness of the proposed model in handling many types of bugs but also, reflect the advancements in large language-based program repair techniques. Furthermore, the comparative analysis offers valuable insights into the evolution of AI capabilities, particularly in the context of software debugging and maintenance. The superior performance of the gpt-4-0125-preview model, characterized by its ability to execute multi-hunk and multi-fault repairs with a higher degree of accuracy, points towards a promising future where the boundaries of automated software engineering can be expanded significantly. As the field continues to evolve, future research can uncover more advanced models and methodologies, further enhancing the scope and accuracy of automated bug fixing.

ACKNOWLEDGMENT

Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU), Jeddah, Saudi Arabia, funded this project under grant no. (KEP-PhD-102-611-1443). The authors, therefore, acknowledge DSR for the financial support.

REFERENCES

[1] C. Liu, J. Yang, L. Tan, and M. Hafiz, "R2Fix: Automatically Generating Bug Fixes from Bug Reports," in *Proc. 2013 IEEE Sixth Int. Conf. Softw. Testing, Verification and Validation*, 2013, doi: 10.1109/ICST.2013.24.

[2] A. Koyuncu et al., "IFIXR: Bug Report Driven Program Repair," in *Proc. 2019 27th ACM Joint Meeting on European Softw. Eng. Conf. and Symp. on the Foundations of Softw. Eng.*, Aug. 2019, doi: 10.1145/3338906.3338935.

[3] J. Xuan, M. Martinez, F. Demarco, M. Clement, S. L. Marcote, T. Durieux, D. Le Berre, and M. Monperrus, "Nopol: Automatic repair of conditional statement bugs in java programs," IEEE Transactions on Software Engineering, vol. 43, no. 1, pp. 34–55, 2016.

[4] D. T. Nguyen, D. Qi, A. Roychoudhury, and S. Chandra, "Semfix: Program repair via semantic analysis," in *2013 35th International Conference on Software Engineering (ICSE)*, pp. 772–781, IEEE, 2013.

[5] H. S. Xia and L. Zhang, "Keep the Conversation Going: Fixing 162 out of 337 Bugs for $0.42 Each Using ChatGPT," *arXiv preprint arXiv:2304.00385*, 2023. [Online]. Available: https://arxiv.org/pdf/2304.00385.pdf [Accessed: Feb. 8, 2024].

[6] W. Ma et al., "LLMs: Understanding Code Syntax and Semantics for Code Analysis," 2024, *arXiv:2305.12138*. [Online]. Available: https://arxiv.org/pdf/2305.12138.pdf [Accessed: Mar. 25, 2024].

[7] Z. Feng et al., "Codebert: A Pre-Trained Model for Programming and Natural Languages," 2020, *arXiv:2002.08155*. [Online]. Available: https://arxiv.org/abs/2002.08155 [Accessed: Jan. 29, 2024].

[8] C. S. Xia and L. Zhang, "Less Training, More Repairing Please: Revisiting Automated Program Repair Via Zero-Shot Learning," in *Proceedings of the 30th ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering* , pp. 959–971, 2022.

[9] C. S. Xia, Y. Wei, and L. Zhang, "Automated Program Repair in the Era of Large Pre-Trained Language Models," in *Proceedings of the 45th International Conference on Software Engineering (ICSE 2023). Association for Computing Machinery*, May 2023, doi: 10.1109/ICSE48619.2023.00129.

[10] Y. Li, S. Wang, and T. N. Nguyen, "DLFix: Context-Based Code Transformation Learning for Automated Program Repair," in *Proceedings of the 42nd ACM/IEEE International Conference on Software Engineering (ICSE'20). 602 614.*, pp. 602–614, 2020.

[11] M. Chen et al., "Evaluating Large Language Models Trained on Code," 2021. [Online]. Available: http://arxiv.org/abs/2107.03374 [Accessed: Mar. 20, 2024].

[12] Y. Yuan and W. Banzhaf, "ARJA: Automated Repair of Java Programs via Multi-Objective Genetic Programming," *IEEE Trans. Softw. Eng.*, vol. 46, pp. 1040–1067, 2020, doi: 10.1109/TSE.2018.2874648.

[13] G. Yang, Y. Jeong, K. Min, J. Lee, and B. Lee, "Applying Genetic Programming with Similar Bug Fix Information to Automatic Fault Repair," *Symmetry*, vol. 10, p. 92, 2018, doi: 10.3390/sym10040092.

[14] C. Le Goues, T. Nguyen, S. Forrest, and W. Weimer, "GenProg: A Generic Method for Automatic Software Repair," *IEEE Trans. Softw. Eng.*, vol. 38, pp. 54–72, 2012.

[15] K. Huang et al., "A Survey on Automated Program Repair Techniques," 2023, *arXiv:2303.1*. [Online]. Available: https://arxiv.org/abs/2303.1 [Accessed: Mar. 17, 2024].

[16] T. Lutellier et al., "Coconut: Combining Context-Aware Neural Translation Models Using Ensemble for Program Repair," in *Proceedings of the 29th ACMSIGSOFT International Symposium on Software Testing and Analysis*, pp. 101–114, Jul. 2020, doi: 10.1145/3395363.3397369.

[17] F. Huq, M. Hasan, M. M. A. Haque, S. Mahbub, A. Iqbal, and T. Ahmed, "Review4Repair: Code Review Aided Automatic Program Repairing," *Inf. Softw. Technol.*, vol. 143, p. 106765, 2022.

[18] D. Sobania, M. Briesch, C. Hanna, and J. Petke, "An Analysis of the Automatic Bug Fixing Performance of ChatGPT," in *Proc. 2023 IEEE/ACM Int. Workshop on Automated Program Repair (APR)*, Melbourne, Australia, 2023, pp. 23-30, doi: 10.1109/APR59189.2023.00012.

[19] J. Zhang et al., "Revisiting Test Cases to Boost Generate-and-Validate Program Repair," in *Proc. 2021 IEEE Int. Conf. Softw. Maintenance and Evolution (ICSME)*, Sep. 2021, doi: 10.1109/ICSME52107.2021.00010.

[20] E. Mashhadi and H. Hemmati, "Applying Codebert for Automated Program Repair of Java Simple Bugs," in *Proc. 2021 IEEE/ACM 18th Int. Conf. Mining Softw. Repositories (MSR)*, Mar. 2021, doi: 10.1109/MSR52588.2021.00063.

[21] J. A. Prenner and R. Robbes, "Automatic Program Repair with OpenAI's Codex: Evaluating QuixBugs," 2021, *arXiv preprint arXiv:2111.03922*. [Online]. Available: https://arxiv.org/abs/2111.03922 [Accessed: Jan. 31, 2024].

[22] S. Alsaedi, A. A. Gad-Elrab, A. Noaman, and F. Eassa, "Two-Level Information-Retrieval-Based Model for Bug Localization Based on Bug Reports," *Electronics*, vol. 13, p. 321, 2024.

[23] S. A. Alsaedi, A. Y. Noaman, A. A. Gad-Elrab, and F. E. Eassa, "Nature-based prediction model of bug reports based on Ensemble Machine Learning Model," *IEEE Access*, vol. 11, pp. 63916–63931, 2023.

[24] OpenAI, "GPT-3.5 Turbo Model Documentation," OpenAI Platform, 2024. [Online]. Available: https://platform.openai.com/docs/models/gpt-3-5-turbo [Accessed: Apr. 15, 2024].

[25] OpenAI, "GPT-4 Turbo and GPT-4 Model Documentation," OpenAI Platform, 2024. [Online]. Available: https://platform.openai.com/docs/models/gpt-4-turbo-and-gpt-4 [Accessed: Apr. 15, 2024].

[26] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig, "Pretrain, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing," *ACM Comput. Surv.*, vol. 55, pp. 1–35, 2023.

[27] J. White et al., "A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT," 2023, *arXiv preprint arXiv:2302.11382*. [Online]. Available: https://arxiv.org/abs/2302.11382 [Accessed: Mar. 21, 2024].

[28] R. M. Karampatsis and C. Sutton, "How Often Do Single-Statement Bugs Occur? The ManySStuBs4J Dataset," in *Proc. 17th Int. Conf. Mining Softw. Repositories*, Jun. 2020, pp. 573–577, doi: 10.1145/3379597.3387491.

[29] Y. Qi, W. Liu, W. Zhang, and D. Yang, "How to Measure the Performance of Automated Program Repair," in *Proc. 2018 5th Int. Conf. Information Sci. and Control Eng. (ICISCE)*, Jul. 2018, doi: 10.1109/ICISCE.2018.00059.

[30] OpenAI, "New Embedding Models and API Updates," OpenAI Blog, 2024. [Online]. Available: https://openai.com/blog/new-embedding-models-and-api-updates [Accessed: Apr. 15, 2024].

[31] A. Zirak and H. Hemmati, "Improving Automated Program Repair with Domain Adaptation," *ACM Trans. Softw. Eng. Methodol.*, vol. 33, pp. 1–43, 2024, doi: 10.1145/3631972.

[32] L. Gazzola, D. Micucci, and L. Mariani, "Automatic Software Repair: A Survey," *IEEE Trans. Softw. Eng.*, vol. 45, pp. 34–67, 2017.

[33] J. Wang et al., "Software Testing with Large Language Models: Survey, Landscape, and Vision," *IEEE Trans. Softw. Eng.*, doi: 10.1109/TSE.2024.3368208.

[34] X. Du et al., "Evaluating Large Language Models in Class-Level Code Generation," in *Proc. IEEE/ACM 46th Int. Conf. Softw. Eng.*, Apr. 2024, doi: 10.1145/3597503.3639219.

[35] H. A. Ahmed, N. Z. Bawany, and J. A. Shamsi, "CaPBug-A Framework for Automatic Bug Categorization and Prioritization Using NLP and Machine Learning Algorithms," *IEEE Access*, vol. 9, pp. 50496–50512, 2021, doi: 10.1109/ACCESS.2021.3069248.

[36] NVIDIA, "What Are Large Language Models?" NVIDIA Glossary. [Online]. Available: https://www.nvidia.com/en-us/glossary/large-language-models/ [Accessed: Apr. 20, 2024].

[37] T. Nazir and M. Pinzger, "SymDefFix—Sound Automatic Repair Using Symbolic Execution," 2022, *arXiv preprint arXiv:2209.03815*. [Online]. Available: https://arxiv.org/abs/2209.03815 [Accessed: Apr. 17, 2024].

[38] T. Jiang, T. Lutellier, and L. Tan, "Cure: Code-aware neural machine translation for automatic program repair," in *Proc. 2021 IEEE/ACM 43rd Int. Conf. Softw. Eng. (ICSE)*, Madrid, Spain, 22–30 May 2021, IEEE: New York, NY, USA, 2021, pp. 1161–1173.

[39] H. Ye, M. Martinez, T. Durieux, and M. Monperrus, "A comprehensive study of automatic program repair on the QuixBugs benchmark," *J. Syst. Softw.*, vol. 171, p. 110825, 2021.

[40] J. Koppel, "QuixBugs," [Online]. Available: https://github.com/jkoppel/QuixBugs/tree/master/ [Accessed: Jun. 2, 2024].

Towards Secure Internet of Things Communication Through Trustworthy RPL Routing Protocols

Rui LI
Shaanxi Technical College of Finance and Economics, Xianyang 712000, China

***Abstract*—The Internet of Things (IoT) refers to a network of connected objects for autonomous data exchange and processing. With the increasing growth in IoT, ensuring data transmission integrity and security is essential, as data is subject to many attacks. Currently, the routing protocol for low-power lossy networks is RPL and finds wide deployment in IoT deployments. It also provides a framework to define characteristics related to low-power consumption and resilience to specific routing attacks. RPL trust-based routing protocols improve RPL security by introducing a threshold for Minimum Acceptable Trust, permitting only trusted nodes with a sufficient level of obtained trust to participate in routing. This mechanism is designed to reduce malicious activities and to establish secure communications. This paper will provide an overall review of trustworthy RPL routing methods in IoT and discuss the trust metrics of these approaches and their limitations. To the best of our knowledge, this is the first survey focusing on trust-based RPL protocols in IoT, offering valuable insights into the performance of protocols and possible improvements.**

***Keywords*—Internet of Things; routing; trust; data transmission**

I. Introduction

A. Context

The Internet of Things (IoT) is one of the main transformative technological developments experienced today, involving billions of devices connected, communicating, and exchanging data over different networks [1]. As it has been implemented into daily life, IoT involves general applications and particular industries, including healthcare, transportation, agriculture, and manufacturing [2]. This interdependence has forged unparalleled motivation for efficiency and innovation, where devices can analyze the collected data and make decisions in real-time [3]. On the inverse side, however, the rapid proliferation of IoT networks and billions of devices operating today motivated numerous problems, not least of which pertain to security and data privacy. This sets up a situation in which exponential growth in connected devices makes the need for integrity and safety within IoT environments increasingly central [4].

Routing protocols constitute the central heart of IoT networks, facilitating efficient and reliable data delivery across various connected devices [5]. Considering the challenge posed by the diversity of IoT deployment scenarios, which often involve several resource-constrained devices, such protocols are expected to choose optimal paths for communication while keeping energy consumption as low as possible but preserving network performance [6]. Scalability for large and dynamic networks, energy efficiency to prolong the operable life of battery-powered devices, secure communications that protect data from possible threats, and, above all, autonomy to enable them to self-manage without constant outside control are some of the critical design elements for IoT routing protocols. The Routing Protocol for Low-power and Lossy Networks (RPL) has since become the de-facto solution to these issues due to a solid framework and suitability to meet low-power and lossy network requirements. By prioritizing energy efficiency, adaptability has confidence in RPL as a significant enabler of smooth and secure operations for IoT networks, mainly in applications that feature resource scarcity [7].

B. Problem Statement

RPL is specifically designed to meet the peculiar needs of IoT networks, mainly those constrained by limited power, processing capability, and unreliable communication links [8]. RPL organizes the network into a hierarchical framework called a Destination-Oriented Directed Acyclic Graph (DODAG), wherein nodes establish paths to a common destination, typically an Internet gateway or a data sink [9]. Each node in DODAG receives a rank, reflecting its relation to others. The lower this rank, the closer the node is to the study [10]. RPL makes efficient routing decisions since rank computation is based on routing metrics like link quality, throughput, and latency. Technical adaptability and energy efficiency make the RPL useful in innumerable critical domains, ranging from healthcare monitoring systems to smart city infrastructure and industrial automation. Each of these domains needs reliable data transmission with optimum utilization.

Despite RPL's efficiency and adaptability, it is subjected to various security threats like blackhole, rank, and wormhole attacks. Under the Blackhole attack, any malicious node advertises itself with an optimum path to intercept and drop packets to disrupt communications. In Rank attacks, an adversary can alter its rank to change network topology, which might lead to inefficient routing or partitioning of the network. The wormhole attack establishes a virtual tunnel between two distant parts of the network to deceive nodes, which may result in the interception or modification of data. This attack on RPL is critical, as such vulnerabilities can compromise essential IoT applications, such as healthcare and smart city systems, leading to data loss, service disruption, or even risks to human safety. Proper security can protect data integrity, preserve smooth communication, and maintain users' confidence in such IoT systems.

C. Motivation and Contribution

Trust-based RPL routing protocols have emerged as promising to enhance RPL security. These protocols incorporate

mechanisms that evaluate the trustworthiness of network nodes based on their behavior, record of communications, or reliability. By assigning a trust score, only nodes that meet predefined trust thresholds can participate in the routing process, thereby reducing the possibility of malicious activities. Essential trust metrics include packet forwarding ratios, energy consumption, and reputation scores distinguishing trusted and compromised nodes. Based on these trust metrics, the trust-based RPL protocols monitor for attacks, including blackhole or rank manipulation, to ensure data is forwarded to reliable nodes. This approach significantly improves the integrity, trust, and reliability of communications in IoT and enhances general network resilience against diverse security threats.

Kamgueu, et al. [11] presented various RPL protocol enhancements, emphasizing security and mobility. They classified existing solutions and discussed their effectiveness in mitigating RPL's inherent vulnerabilities. However, they have not discussed trust-based security mechanisms for RPL. Yang, et al. [12] presented an overview of possible RPL security vulnerabilities in RPL-based IoT networks and a discussion about possible countermeasures. While this study addresses most security-related issues, it provides only a limited discussion on trust-based routing protocols and lacks a critical analysis of trust metrics and their implementation mechanisms.

Sobral, et al. [13] presented a detailed survey of the routing protocol for Low-Power and Lossy Networks. This paper considers the evaluation of different protocols with various performance metrics like energy efficiency and scalability. The security issues are briefly discussed, but a detailed study on trust-based approaches has not been drawn. Bang, et al. [14], on the other hand, presented a thorough analysis of enhancements in security, scalability, and energy efficiency to RPL. They also discussed various attacks on RPL and their respective mitigation strategies. However, the trust-based routing protocol discussion falls short and thus demands more research into trust metrics and their impact on network performance. Shah, et al. [15] addressed the challenges and solutions of routing protocols in mobile IoT environments. They emphasized the challenges in mobility management and energy constraints but gave little attention to trust-based routing mechanisms, specifically on RPL.

Although much research has been done on RPL and its security challenges, a literature gap regarding the need for comprehensive surveys on trust-based RPL routing protocols is evident. Most related works are contributed piecemeal, leaving room for comprehensive analysis and comparison among those trust-based mechanisms. This, therefore, requires a review of the same protocols; indeed, trust-based approaches have promised a solution for improving IoT network security and preventing malicious activities while ensuring reliable communication. Such a review will shed light on the strengths and limitations of the current trust-based methods and highlight areas for further improvements to guide future research. This gap gives us a reduced understanding of the potentials available with the trust-based RPL protocols and retardation in further development of secured and resilient IoT systems.

This study primarily attempts to evaluate trust-based RPL routing protocols, examine the various trust measures used, and outline their limits and issues. Additionally, the study aims to identify and propose potential future research avenues to address existing shortcomings and increase security in IoT networks. This study is distinctive as it represents the inaugural comprehensive survey focused on trust-based RPL protocols. It offers a thorough analysis that addresses a significant gap in the literature and establishes a foundation for researchers and practitioners aiming to develop secure and efficient IoT communication systems.

II. Background

This section describes the basic principles of IoT and routing protocols, including the general architecture and operation based on the protocol RPL and the principle of trust-based routing in IoT networks.

A. Definition and Impact of IoT

IoT is ubiquitous across virtually every aspect of human activity in modern life because of intelligent gadgets or interconnected systems. Companies, governments, and organizations increasingly employ independent devices to increase productivity, enhance quality of service, and spur economic growth [16]. IoT, in return, with its rapid advancement in various fields of health and energy management for military purposes, agriculture, and smart city infrastructure, among others, is trying to make the world much more useful and linked. It has been described by many as the "system of systems" or even a "network of networks," with IoT communication protocols capable of making devices self-configuring and strictly limiting access to authorized and trusted users [17].

However, as IoT networks scale to billions of devices, the issues of security, scalability, and resource consumption become critical. Each IoT service model differs in security, architecture, and implementation, making standardization and integration of new services even more challenging [18]. Fig. 1 illustrates the complexity of the ecosystem that needs to be managed, starting with exceptionally diverse applications and network models and proceeding to considerations about security and privacy. Security remains a significant barrier to IoT's expansion, especially given the resource limitations of many IoT devices, making them vulnerable to various threats. These include Intrusion Detection Systems (IDS), trust-based models, and cryptographic techniques. The general idea of adapting to various IoT applications while implementing routing algorithms is highly trust-based models, which can support detecting and isolating malicious nodes.

B. Overview of the RPL Protocol

RPL stands for Routing Protocol for Low-power and Lossy Networks, designed to meet the unique requirements of IoT networks, which, in turn, feature limited power, processing capabilities, and unreliable communication links. The RPL protocol is one of the proactive protocols, which immediately establishes routing paths once the network becomes operative. It organizes the network into a DODAG structure, illustrated in Fig. 2. This structure has one root node that coordinates network communication. Due to such a structure, RPL can efficiently decide routing paths by ranking each node, as illustrated in Fig. 2. For the computation of the ranks, local metrics and constraints are considered so that the network has no cycle and performance is optimized.

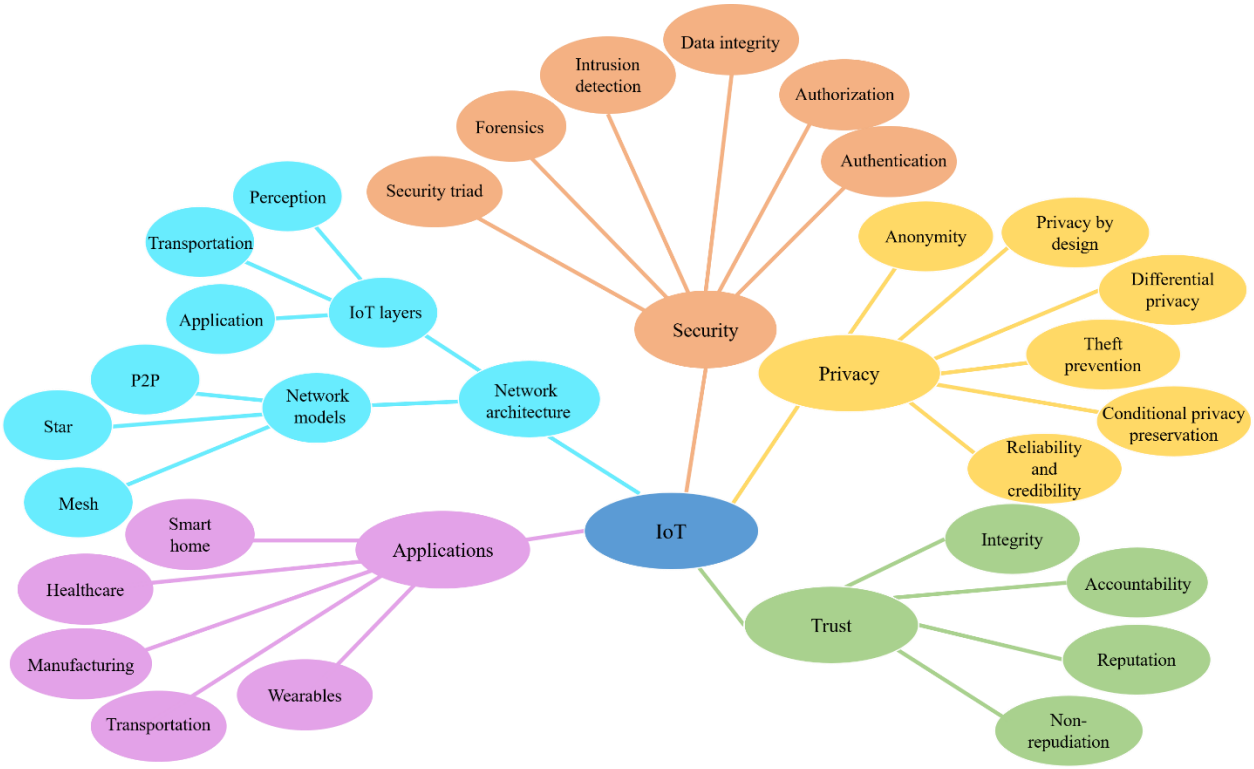


Fig. 1. An overview of IoT paradigm.

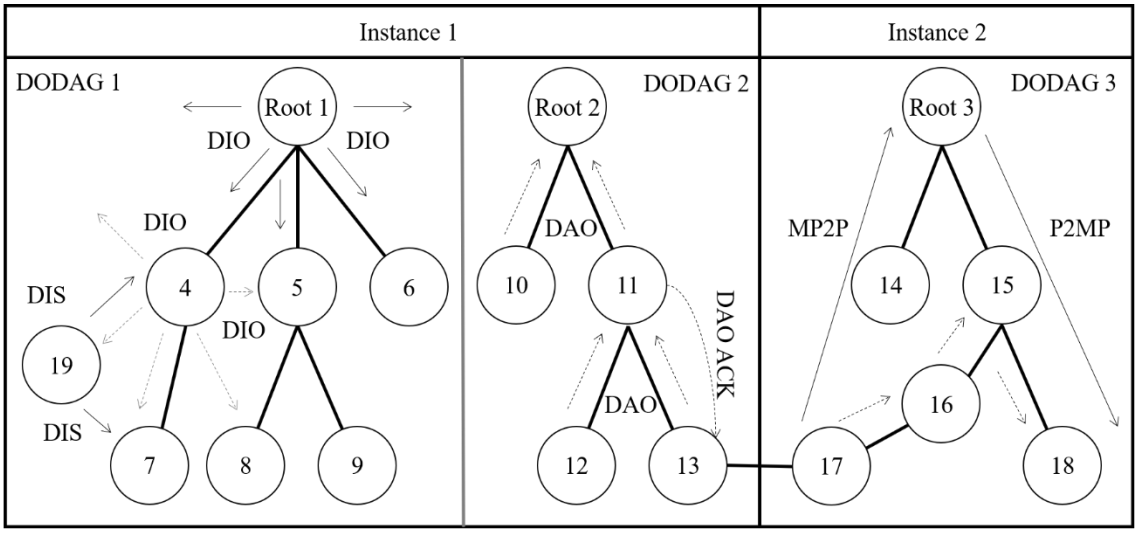


Fig. 2. DODAG structure.

The formation of DODAG in RPL starts with the root node broadcasting DIO messages. The nodes hear these messages, calculate their rank, and choose a parent node based on routing metrics. The nodes ensure the acyclicity of the DODAG by ensuring that the rank of its parent is less than that of itself. Techniques like trickle timers prevent extra overhead by controlling message traffic. The nodes can join the network by sending the DODAG Information Solicitation message. RPL also has different modes of operation for root to upward routes and downward routes based on Destination Advertisement Object (DAO) messages to communicate efficiently in the network.

C. Trust-based Routing in IoT

The widespread deployment of IoT devices in diverse environments increases the risk of security breaches, such as attacks that drain a device's energy or disrupt network operations. Given these threats, designing adequate safeguards is crucial, particularly considering the limited resources of many IoT devices. A promising approach is trust-based routing, which detects and isolates malicious nodes using appropriate trust assessment strategies. Trust among network nodes builds up over time, allowing the system to discriminate effectively between honest and malicious participants. However, attackers

can manipulate trust metrics either by falsely downgrading reliable nodes or by falsely promoting threatening ones, undermining the effectiveness of trust-based solutions and disrupting the stability of networks.

In other words, trust-based routing protocols ensure participation in route formation and data exchange through nodes that can be adjudged trusted depending on their behavior and past interactions. Isolation strategies will be applied to protect the network against malicious or selfish behavior by nodes. Trust attributes assessment is one of the most challenging barriers at the network level. Thus, trust-based approaches are recommended for several IoT applications.

D. Performance Metrics

This section presents some of the most essential metrics considered in the literature to evaluate the performance and security of RPL-based IoT networks. Each metric is briefly explained, and the relevant equations are presented.

Malicious node containment: This evaluates the protocol's effectiveness in isolating malicious nodes, expressed as follows.

$$Acc = \frac{No.of\ detected\ malicious\ nodes}{Total\ no.of\ malicious\ nodes} \quad (1)$$

Where Acc represents the detection accuracy of the proposed solution.

Network throughput: Measured in kilobits per second, indicates the data transmission rate over a given period.

$$Throughput = \frac{Total\ data\ transmitted\ (in\ kilobits)}{Total\ time\ (in\ seconds)} \quad (2)$$

Where Total Data Transmitted is the amount of data successfully sent over the network, and Total Time is the duration over which the data transmission occurs. This formula provides the data transmission rate in kilobits per second (kbps).

Storage cost: Nodes maintain lists of neighbors' behaviors, increasing storage requirements at the node level.

$$SC = N \times S_b \quad (3)$$

Where N represents the number of neighboring nodes for which behavior information is maintained, and S_b is the storage size required (in bytes) to store the behavior information of a single node.

Median packets dropped: This metric indicates the percentage of packets dropped during attacks, calculated as:

$$D_r = \frac{Drop_r}{\sum_{k=1}^{n} TAk} \quad (4)$$

$$E[Dpkt] = Median\ (D) \quad (5)$$

Packet Delivery Ratio (PDR): PDR is the median value of the packet delivery ratio across multiple repetitions, calculated as:

$$S_r = \frac{Rcvd}{\sum_{k=1}^{n} P_k} \quad (6)$$

$$E[DRpkt] = Median\ (S) \quad (7)$$

Communication overhead: This metric assesses the additional communication costs due to control messages exchanged during security events.

$$CO = \frac{CM}{TD} \times 100 \quad (8)$$

Where *CM* is the number of control messages transmitted during security events, and *TD* denotes the total data packets transmitted over the network.

Misdetection rate: This rate measures the error in detecting malicious nodes, accounting for false positives and negatives.

$$MR = \frac{FP + FN}{TP + TN + FP + FN} \quad (9)$$

Where *FP* denotes the number of false positives (benign nodes incorrectly identified as malicious), *FN* represents the number of false negatives (malicious nodes incorrectly identified as benign), *TP* is the number of true positives (malicious nodes correctly identified), and *TN* is the number of true negatives (benign nodes correctly identified).

Trust values: A trust value is a feeling of confidence in a node's predictable behavior and honesty. Trust values are essential for routing to ensure that only trustworthy nodes participate in packet forwarding, especially faraway nodes. The trust can be computed as follows:

$$RT(N_i, N_m) = DT(N_i, N_j) \times DT(N_j, N_m) \quad (10)$$

Where $RT(N_i, N_m)$ is the recommended trust value from node *i* to node *m*, and $DT(N_i, N_j)$ and $DT(N_j, N_m)$ are the direct trust values between nodes.

Reliability: Reliability is based on a predefined trust threshold. It assesses the trustworthiness and dependability of nodes, ensuring only reliable nodes participate in routing. Trust ratings are similar to assigning ranks to nodes based on trust indices.

$$R = \frac{N_t}{N} \quad (11)$$

Where *R* is the Reliability of the network, N_t represents the number of nodes with trust values above the trust threshold, and *N* is the total number of nodes in the network.

Key generation time: It calculates the time a key is generated for secure data transmission. High key generation times cause IoT devices to consume extra resources and experience delays. Thus, reducing them is critical for efficient data communication.

$$KGT = E_t - S_t \quad (12)$$

Where E_t denotes the end time when the key generation process is completed and S_t represents the start time when the key generation process begins.

Energy consumption: This metric quantifies the energy used during data transmission in the network. It is determined by the difference between the total energy T_e and the remaining energy R_e, given by:

$$EC = T_e - R_e \quad (13)$$

Delay: It refers to the time taken for data packets to travel from the source nodes to the root destination node in the network. Thus, it can be measured as a difference between packet transmission time, $P_{r,t}$, and packet received time, $P_{t,t}$, given as:

$$D = \sum_{i=0}^{n} \left(P_{r,t} - P_{t,t}\right) \quad (14)$$

III. TRUST-BASED RPL ROUTING PROTOCOLS

The taxonomy of trust-based RPL protocols can be divided into: 1) Trust-based detection and isolation mechanisms, which primarily detect and isolate malicious nodes by relying on trust scores generated from node behavior and recorded communications; 2) Energy-efficient and lightweight models of trust, which essentially reduce computational and energy overhead due to trust evaluations-small enough for resource-constrained IoT devices; 3) Advanced models of trust for dynamic environments for the adaptation of trust mechanisms to the respective mobility, heterogeneity, and time-varying conditions in IoT networks; 4) Attack-specific mitigation strategies that aim to identify unique trust metrics coupled with countermeasures developed for particular threats, such as blackhole, rank, and wormhole attacks. Together, these categories provide solutions for crucial challenges to RPL security by enhancing detection, reducing resource consumption, and improving network resilience against various threats.

A. Trust-Based Detection and Isolation Mechanisms

Airehrour, et al. [19] proposed SecTrust-RPL, which uses trust assessments to segregate malicious nodes from the network while determining optimal routing paths. Successful packet exchanges thus contribute to node reliability and build-up throughput. However, complex attacks like combinations of rank, blackhole, and Sybil would not be defended, and the integration of trustworthy nodes into the network needs to be addressed. Airehrour, et al. [20] developed a trust-based routing protocol for low-power networks whose efficiency was checked with RPL classic through Minimum Rank with Hysteresis Objective Function (MRHOF).

Rakesh [21] introduced the concept of SecRPL-MS to secure RPL-based IoT networks through authentication and security measures. This minimizes energy consumption in rank, Sybil, blackhole, and man-in-the-middle attacks. However, it also has some disadvantages since it neglects DoS/DDoS threats entirely and provides no satisfactory trust-based verification process. Ioulianou, et al. [22] presented SRF-IoT, which was developed by integrating external IDS and trust-based methods against rank and blackhole attacks. Decreasing the extra parent switches to the minimum will provide higher network efficiency, though it suffers from limitations in identifying unidentified attacks and indentation.

Patel, et al. [23] proposed a trust-based intrusion detection solution, FSTIDS, to reduce the impact of topology-based selective forwarding attacks against RPL networks. Selective forwarding is a hard attack to detect because it manifests by merely dropping control or data packets. FSTIDS performs computation at the sink node for trust values to reduce overhead, embedding a threshold and uncertainty factor to maintain accuracy.

Airehrour, et al. [24] proposed a trust-based routing protocol that efficiently protected against blackhole attacks and improved network efficiency without extra traffic overhead. This scheme addresses the challenge caused by compromised sensor nodes in IoT networks. Compromised nodes affect routing integrity by issuing false control information, dropping packets, and even introducing false data during aggregation or obstructing data forwarding.

Djedjig, et al. [25] analyzed the problem of trust management in RPL networks, illustrating that trusting only the Trusted Platform Module (TPM) is not enough to make nodes reliable. They pointed out that nodes may be infected internally or selfishly and thus still build the RPL topology. Their solution proposes trust values as a main routing criterion derived from a node's behavior to create a robust trust mechanism. Airehrour, et al. [26] proposed a comprehensive trust-based RPL protocol that could efficiently prevent black hole attacks. The authors evaluated the performance of the proposed protocol with standard RPL using the MRHOF and IETF's Contiki RPL implementation to analyze its capability in mitigating black hole threats.

These protocols enhance the security of RPL networks, which expose malicious nodes through trust evaluations and isolate them. Some protocols, such as SecTrust-RPL and SecRPL-MS, as shown in Table I, detect dangerous nodes based on trust metrics for their isolation to perform secure routing. However, their procedures are different when various types of attacks need to be handled.

Though SecTrust-RPL is efficient at threat node isolation without dealing with sophisticated attacks, SecRPL-MS detects many threats but cannot mitigate DDoS attacks. Similarly, RPL SRF-IoT maintains an external IDS to provide additional security features against rank and blackhole attacks. However, again, it fails to identify undetected threats. In addition, protocols like FSTIDS and TrustedRPL are more focused on efficient computation of trust and self-organization, ignoring factors related to node mobility and energy consumption in IoT.

B. Energy-Efficient and Lightweight Trust Models

Subramanian, et al. [27] has proposed HTmRPL++, which enhances the trust between fog nodes without compromising network speed. Though effective against BSA, it does not consider node mobility and addresses only one attack type. Mehta and Parmar [28] have proposed an energy-efficient strategy against wormhole and gray hole attacks using lightweight trust mechanisms. The technique is practical, yet extra or combined attacks are not considered.

Ul Hassan, et al. [29] suggested CTrust-RLP, which introduced a control layer that detects and isolates blackhole attacks with efficient energy conservation. Still, it reduces the processing overhead without scalability and cannot cope with multiple threats, such as Sybil and rank attacks.

TABLE I. AN OVERVIEW OF TRUST-BASED DETECTION AND ISOLATION MECHANISMS

Reference	Key features	Strengths	Limitations
[19]	Uses trust to isolate malicious nodes and optimize routing by examining successful packet exchanges	High throughput, effective at detecting and isolating suspicious nodes	Does not account for complex collusive attacks (e.g., rank/blackhole, rank/Sybil) and lacks strategies for integrating trustworthy nodes
[20]	Secures against black hole and selective forwarding attacks without adding traffic overhead	Effective against critical attacks and maintains network efficiency	Ignores energy consumption and does not address a broader range of attacks
[21]	Employs authentication and security to mitigate rank, Sybil, blackhole, and man-in-the-middle attacks, with a focus on delay and energy reduction	Comprehensive attack defense reduces packet loss and delay	Overlooks DDoS/DoS threats and lacks a robust trust verification mechanism
[22]	Combines external IDS and trust-based mechanisms to counter rank and blackhole attacks	Reduces parent switches and enhances network efficiency	Limited in detecting additional threats and lacks a solution for indentation attacks
[23]	Uses trust-based intrusion detection for selective forwarding attacks, computing trust at the sink node to reduce overhead	Efficient in minimizing processing load and precise trust computation	Does not consider mobility scenarios, restricting effectiveness in dynamic environments
[24]	Protects against black hole attacks without increasing network traffic using a trust-based mechanism	Secures against compromised nodes and maintains traffic efficiency	Does not consider energy criteria or other types of attacks
[25]	Uses trust values as the primary routing criterion to create a self-organized network	Focuses on internal threats and enables trust-based self-organization	Lacks simulation for empirical validation and does not analyze energy or routing overhead
[26]	Compares trust-based RPL to standard RPL and IETF Contiki RPL, focusing on mitigating black hole attacks	Strong protection against black hole attacks, validated through comparison	Insufficient focus on selective forwarding, ranking, and Sybil attacks; does not address energy and network lifespan

Djedjig, et al. [30] developed a Management Trust Scheme (MRTS) to enhance the security of RPL networks through a distributed and collaborative trust model. This model evaluates nodes' behavior to compute a trust-based value called the Extended RPL Node Trustworthiness (ERNT) measure. MRTS leverages these trust assessments to ensure that only reliable nodes participate in routing, enabling the self-organization of a secure network based on trust status.

Sisodiya, et al. [31] proposed a multicast trust-based RPL management scheme to enhance network security. This approach achieves its goal by continuously monitoring and isolating untrusted nodes that would destroy data integrity or delay and corrupt messages in the protocol. The protocol enables nodes to infer the degree of trust of neighboring nodes in establishing a network topology. It is more effective than broadcast-based transmissions. It provides greater energy efficiency, high throughput, and low dead node ratios. This schema detects malicious nodes before route establishment, allowing only trustworthy nodes to participate in secure multicast.

All these protocols target the optimization of the RPL security mechanism, not to compromise network performance for energy conservation. Most protocols listed in Table II, such as HTmRPL++ and CTrust-RPL, target lightweight design and will be suitable for application in resource-constrained environments. HTmRPL++ addresses specific attacks, such as Ballot Stuffing Attacks, without considering the node's mobility. As a result, it presents robust mutual trust among fog nodes. CTrust-RPL detects a suspicious node or source that can create an attack by using a control layer to reduce overhead processing efficiently; this technique has poor scalability and resistance against sophisticated threats.

TABLE II. AN OVERVIEW OF ENERGY-EFFICIENT AND LIGHTWEIGHT TRUST MODELS

Reference	Key features	Strengths	Limitations
[27]	Trust mechanism tailored for fog nodes, designed to maintain network speed and performance; tested for reliability, delay, and ballot stuffing attacks	Efficient communication for resource-constrained fog nodes and lightweight design	Does not account for node mobility, limited to testing against a single attack
[28]	Defends against wormhole and grayhole attacks using an energy-efficient trust-based approach	Reduces packet loss, isolates rogue nodes and improves performance	Fails to address combined or additional attacks beyond wormhole and grayhole
[29]	Control layer-based trust mechanism that conserves energy while stopping black hole attacks; calculates trust based on packet exchanges	Low processing and storage overhead, improved network longevity, and effective against black hole attacks	Not scalable or distributed, lacks defenses against threats like Sybil and ranking attacks, and limited evaluation of trust mechanism
[30]	Distributed trust model using extended RPL node trustworthiness for secure routing; allows self-organization based on trust status	Enables secure, self-organized network, and behavior-based trust evaluation	Overlooks energy consumption and routing/security overhead, not tested with Cooja-Contiki simulator
[31]	Uses a trust-based approach to identify and isolate malicious nodes, enabling efficient multicast transmission	More energy-efficient and reliable compared to Broadcast RPL, enhances network longevity and throughput	Does not address performance under various attack types, limited platform testing restricts versatility

C. Advanced Trust Models for Dynamic Environments

Muzammal, et al. [32] offered SMTrust, a mobility-based trust model that mitigates attacks such as black holes, rank, and gray holes. It provides optimization of routing in both static and mobile environments but needs to address energy consumption and other potential threats comprehensively. Al-Jumeily, et al. [33] came up with a hybrid trust approach against Sybil attacks, which they called THC-RPL. While this can help with network longevity, reducing packet loss significantly, this protocol has yet to be tested in practical situations, the relevance of which is thus limited.

Addressing one of the biggest challenges in military usage of COTS IoT devices, Thulasiraman and Wang [34] produced a lightweight trust-based per-node secure data transmission architecture based on routing in mobile IoT networks. This design enhanced security for the RPL IoT routing algorithm by bringing in nonce identity values, timestamps, and whitelisting. The modified protocol selects routing paths based on computed node trust values and the Average Received Signal Strength Indicator (ARSSI).

Hassan, et al. [35] introduced the Jini Index approach, a trust-based security framework designed to effectively detect and manage Sybil attacks, one of the most challenging internal threats in IoT networks. Their architecture employs a tiered design comprising layers of devices and fog nodes, enhancing overall network security and energy efficiency by offloading processing tasks from individual nodes. This approach significantly improves Sybil attack detection, reducing latency and energy consumption.

Savitha and Basarkod [36] introduced TEMGTO-RPL, which applies Gorilla Forces Optimization for node selection while balancing trust and energy efficiency. According to the simulation results, the protocol performs well but needs to consider various attack methods. Therefore, the TEMGTO-RPL protocol has an improvement space.

Muzammal, et al. [37] proposed the SMTrust model, which involves an extension of RPL with a trust factor criterion based on mobility in IoT networks. The SMTrust approach resists general RPL attacks such as Blackhole, grey hole, rank, and version number attacks. It considers the sink and sensor node mobility and only allows trustworthy nodes to participate in the network. It is designed for sensor nodes to provide confidentiality, integrity, and availability during routing and communicating data.

Jiang and Liu [38] proposed a defensive technique that effectively addresses sophisticated selective forwarding attacks in RPL-based IoT networks. They design and implement a set of energy-efficient attacks that can flexibly select the type and proportion of packets to be broadcast to maximize the impact, increasing the number of missed packets while keeping stealthy. They brought the right to neutralize these threats, a simple trust-based security mechanism using a beta trust model with asymmetric forgetting rates and decaying trust values.

The protocols reviewed address securing RPL networks in highly mobile or dynamic scenarios. According to Table III, protocols like SMTrust and THC-RPL integrate mobility-based trust metrics into their trust computations to ensure both fixed and mobile nodes. SMTrust can protect against various RPL attacks and perform better than the existing schemes. The significant drawbacks of this protocol lie in its failure to focus on energy consumption and protection against a wide range of threats. THC-RPL adopts a hybrid approach to Sybil attack detection to extend the network lifetime, yet it is waiting for validation in a natural environment.

In Thulasiraman and Wang's Lightweight Security Architecture, countermeasures against DoS and Sybil attacks are robustly provided in mobile IoT networks. Energy efficiency problems may be shown when the number of mobile nodes increases. The proposed Jini Index method by Hassan and Tariq detects Sybil attacks with minimum latency and energy utilization efficiently. However, more is needed to cover broader recognition of attacks using machine learning. TEMGTO-RPL proposes a trust model based on optimization that strikes a balance between energy and trust considerations without concerns about multiple attack vectors. These models show how trust mechanisms will adapt to IoT environments, which will be dynamic and balance mobility, energy efficiency, and security.

D. Attack-Specific Mitigation Strategies

Kim, et al. [39] proposed PITrust, which utilizes the RSSI and a centralized mechanism for trust to enhance the detection accuracy of Sybil attacks. While this method is effective, it targets only one type of attack with no extensive security scope. Lahbib, et al. [40] have proposed LT-RPL, which secures RPL networks under blackhole and grayhole attacks while still assured QoS guarantees. While this is efficient, it does not consider other types of attacks and protocol testing in various application settings. Karkazis, et al. [41] contributed to TXPFI by enhancing the routing with minimal transmission to enhance efficiency.

These approaches focus their contribution on protocols developed for targeted threats to optimize security measures against certain kinds of attacks in RPL networks. As summarized in Table IV, PITrust uses RSSI and a centralized trust mechanism to find Sybil attacks with high accuracy, but it does not cover other attack types. LT-RPL uses an ETX-based trust model to prevent black and gray hole attacks and ensure efficient and energy-conscious routing.

However, it is not validated in diverse application scenarios and does not cancel additional threats. TXPFI proposes a metric that minimizes message transmissions, considering retransmissions and lost links, and significantly reduces communication overhead. It provides a deficiency in performance evaluation during an attack and does not provide comprehensive security.

TABLE III. AN OVERVIEW OF ADVANCED TRUST MODELS FOR DYNAMIC ENVIRONMENTS

Reference	Key features	Strengths	Limitations
[32]	Incorporates mobility-based trust metrics for both fixed and mobile nodes; defends against blackhole, rank, and version number attacks	Outperforms MRHOF, SecTrust, and MRTS; improves performance for mobile and static nodes	Does not address energy efficiency or consider additional attack types
[33]	Hybrid trust model that uses Direct Trust (DT) and Indirect Trust (IDT) to detect Sybil nodes; relays trust data to the root node	Reduces packet loss, extends network lifespan, effective against Sybil attacks	Not tested in real-world environments, limiting practical applicability
[34]	Trust-based routing for mobile IoT using nonce, timestamp, and ARSSI; designed for COTS IoT devices	Protects against DoS and Sybil attacks, high PDR, lightweight and efficient	Does not consider energy consumption or handle a large number of mobile nodes efficiently
[35]	Layered trust model using fog devices to manage Sybil attacks; reduces detection latency and energy consumption	Effectively detects Sybil attacks, reduces energy use and latency	Does not integrate ML algorithms for broader attack mitigation, limited to Sybil attack recognition
[36]	Uses Gorilla Forces Optimization (GTO) for trust and energy-efficient routing; considers trust value, energy ratio, node distance, and PDR	Balances trust and energy efficiency, selects optimal nodes for routing	Does not address multiple attack types, lacks comprehensive security measures
[37]	Focuses on mobility and trust criteria to secure RPL networks against blackhole, greyhole, and rank attacks	Ensures confidentiality, integrity, and availability of sensor nodes; energy-efficient design	Limited in addressing combined attacks, needs efficiency improvements in detection and data transfer
[38]	Uses a beta trust model with asymmetric forgetting rates to detect and mitigate selective forwarding attacks	Efficient against selective forwarding, energy-efficient attack model	Ignores mobile node dynamics, does not address other attack vectors

IV. FUTURE RESEARCH DIRECTIONS

Although several trust-based RPL routing protocols have recently been developed for securing IoT networks, several research gaps and challenges remain to be addressed. Below are future research directions to enhance these protocols' effectiveness, scalability, and robustness.

Most protocols proposed to date target specific attacks, like black holes, Sybil, and rank attacks. Sophisticated and combined attack strategies will attack IoT networks. Future research should develop comprehensive solutions to guard against multiple and simultaneous attack types, such as DDoS, wormhole, and other advanced collusive attacks. Hybrid approaches using machine learning-based anomaly detection and trust evaluation mechanisms may be promising.

These trust-based protocols should take advantage of energy consumption in IoT environments when the resources are constrained. Novel approaches to be explored that tend to minimize energy consumption while sustaining a high level of security. Adaptive computation of trust, energy-aware routing decisions, and lightweight cryptographic methods may lead to less energy-consuming solutions. The protocol design should embed energy consumption models so that a careful evaluation of trade-offs involving security against resource management can be estimated.

IoT networks are highly dynamic, with nodes frequently joining and leaving a network. Most protocols still lack adaptiveness in large-scale or high-mobility scenarios. Trust mechanisms to handle node mobility, changing network topology, and variable device densities using scalable and efficient methods need more research attention. The development of distributed systems for trust management or decentralized usage of blockchain might allow increased scalability and resilience of RPL networks.

Most proposed protocols have been tested only with simulations with minimal scenarios or synthetic datasets. Implementation and testing in diverse and realistic real-world environments are essential in studying practical applicability and effectiveness. Field tests and experiments in intelligent cities or industrial automation based on these solutions may reveal unforeseen challenges and performance issues that might not become evident at a simulation level.

The selection and computation of appropriate trust metrics are critical factors in accurately detecting malicious nodes. In the future, refinement in trust evaluation methods should incorporate context-aware metrics, analysis of historical data, and adaptive trust thresholds. Machine learning algorithms improve trust evaluations by recognizing complex patterns and making better decisions in trust-based routing, making even more enhancement possible.

TABLE IV. AN OVERVIEW OF ATTACK-SPECIFIC MITIGATION STRATEGIES

Reference	Key features	Strengths	Limitations
[39]	Trust-based mechanism using RSSI and a centralized trust model for Sybil attack detection	High detection accuracy, improved routing performance compared to conventional protocols	Limited to addressing only Sybil attacks; does not consider other types of security threats
[40]	Trust management method integrated with ETX-based MRHOF to secure routing topology from black hole and gray hole attacks	Effective at identifying and isolating malicious nodes, provides QoS for energy-efficient routing	Does not address other attack types, lacks validation across various application scenarios and platforms
[41]	Routing metric that minimizes message transmissions by considering frame retransmissions and authenticating lost links	Reduces the number of messages transfers, improves efficiency in data delivery	Does not evaluate protocol under attack conditions, lacks comprehensive security measures and analysis of network lifetime impact

Limited computing and storage resources characterize most IoT devices. Thus, lightweight, efficient security architecture developments are of prime importance. Future research should investigate novel architectures that combine low overhead with solid security features. For example, offloading trust-related intensive computational tasks to fog or edge computing could achieve a good trade-off between performance and security in resource-constrained environments.

Integrating the IoT with blockchain, edge computing, and AI opens a new avenue for IoT network security. Blockchain will offer trust management in a decentralized and tamper-proof way, whereas AI can enhance features related to anomaly detection and adaptive security responses. The research study will focus on integrating such emerging technologies with trust-based RPL protocols to achieve secure and intelligent IoT ecosystems.

In light of the diverse IoT devices and networks, interoperability within the variant trust-based protocols poses a challenge. Any future work, therefore, should seek to develop standardized frameworks and protocols for easy interoperability across different platforms and communication standards. Industry, academia, and standardization bodies are encouraged to work together to enable the adoption of secure and interoperable mechanisms of trust.

Finally, there are issues regarding user data, which could raise some privacy concerns. It is essential to consider techniques during the design of research that will maintain users' privacy while still allowing trust evaluations. This might be a balance between security, trust management, and data privacy using techniques like differential privacy, homomorphic encryption, or federated learning.

Future protocols can adapt to changes in network conditions, such as varying device behavior, environmental factors, or application-specific requirements. Adaptive trust models will adjust the parameters of trust computation on the fly and enhance the robustness of RPL Networks. In this regard, context-aware security mechanisms can ensure that protection is appropriate for the current network state and the criticality of the data transmitted.

V. Conclusion

Despite increasing and diversifying cyber threats, IoT network security remains an open challenge. This review has presented a critical analysis of several trust-based RPL routing protocols proposed to enhance network security. The protocols discussed in this study utilize trust-based mechanisms to detect and isolate malicious nodes to ensure data integrity and optimize routing efficiency. Whereas some have been quite effective in eliminating specific attacks, such as blackhole, Sybil, and rank attacks, there are significant lacunae concerning their comprehensively addressing complex, multifaced threats and adaptation issues related to dynamic network environments. The evaluation of performance metrics across protocols presented a variety of tradeoffs between security and energy efficiency versus network performance. For example, some have high detection accuracy with low communication overhead; however, energy consumption and scalability in large networks should be considered. While designed for energy efficiency, others do not offer protection against complex or combined attacks. These observations clearly dictate why future research should concentrate on holistic and adaptive solutions, including performance and resource constraints. Others pertain to real-world validation and the introduction of cutting-edge technologies such as AI, blockchain, and edge computing. When the IoT network becomes widely adopted, most challenges in energy management, scalability, and user privacy will be resolved.

References

[1] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," Journal of Network and Computer Applications, vol. 97, pp. 23-34, 2017.

[2] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective," Sustainability, vol. 15, no. 4, p. 3317, 2023.

[3] M. Soori, B. Arezoo, and R. Dastres, "Internet of things for smart factories in industry 4.0, a review," Internet of Things and Cyber-Physical Systems, vol. 3, pp. 192-204, 2023.

[4] A. Morchid, R. El Alami, A. A. Raezah, and Y. Sabbar, "Applications of internet of things (IoT) and sensors technology to increase food security and agricultural Sustainability: Benefits and challenges," Ain Shams Engineering Journal, vol. 15, no. 3, p. 102509, 2024.

[5] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," Cluster Computing, vol. 23, no. 2, pp. 641-661, 2020.

[6] W. Mao, Z. Zhao, Z. Chang, G. Min, and W. Gao, "Energy-efficient industrial Internet of Things: Overview and open issues," IEEE transactions on industrial informatics, vol. 17, no. 11, pp. 7225-7237, 2021.

[7] V. Hayyolalam, B. Pourghebleh, and A. A. Pourhaji Kazem, "Trust management of services (TMoS): investigating the current mechanisms," Transactions on Emerging Telecommunications Technologies, vol. 31, no. 10, p. e4063, 2020.

[8] A. Verma and V. Ranga, "Security of RPL based 6LoWPAN Networks in the Internet of Things: A Review," IEEE Sensors Journal, vol. 20, no. 11, pp. 5666-5690, 2020.

[9] E. Bozorgi, S. Soleimani, S. K. Alqaiidi, H. R. Arabnia, and K. Kochut, "Subgraph2vec: A random walk-based algorithm for embedding knowledge graphs," arXiv preprint arXiv:2405.02240, 2024, doi: https://doi.org/10.48550/arXiv.2405.02240.

[10] P. Chithaluru et al., "An enhanced opportunistic rank-based parent node selection for sustainable & smart IoT networks," Sustainable Energy Technologies and Assessments, vol. 56, p. 103079, 2023.

[11] P. O. Kamgueu, E. Nataf, and T. D. Ndie, "Survey on RPL enhancements: A focus on topology, security and mobility," Computer Communications, vol. 120, pp. 10-21, 2018.

[12] W. Yang, Y. Wang, Z. Lai, Y. Wan, and Z. Cheng, "Security Vulnerabilities and Countermeasures in the RPL-based Internet of Things," in 2018 international conference on cyber-enabled distributed computing and knowledge discovery (CyberC), 2018: IEEE, pp. 49-495.

[13] J. V. Sobral, J. J. Rodrigues, R. A. Rabêlo, J. Al-Muhtadi, and V. Korotaev, "Routing protocols for low power and lossy networks in internet of things applications," Sensors, vol. 19, no. 9, p. 2144, 2019.

[14] A. O. Bang, U. P. Rao, P. Kaliyar, and M. Conti, "Assessment of routing attacks and mitigation techniques with RPL control messages: A survey," ACM Computing Surveys (CSUR), vol. 55, no. 2, pp. 1-36, 2022.

[15] Z. Shah, A. Levula, K. Khurshid, J. Ahmed, I. Ullah, and S. Singh, "Routing protocols for mobile Internet of things (IoT): A survey on challenges and solutions," Electronics, vol. 10, no. 19, p. 2320, 2021.

[16] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy-efficient data fusion methods in the Internet of Things," Concurrency and Computation: Practice and Experience, vol. 34, no. 15, p. e6959, 2022.

[17] M. Shoeibi, A. E. Oskouei, and M. Kaveh, "A Novel Six-Dimensional Chimp Optimization Algorithm—Deep Reinforcement Learning-Based Optimization Scheme for Reconfigurable Intelligent Surface-Assisted Energy Harvesting in Batteryless IoT Networks," Future Internet, vol. 16, no. 12, p. 460, 2024, doi: https://doi.org/10.3390/fi16120460.

[18] B. Pourghebleh, K. Wakil, and N. J. Navimipour, "A comprehensive study on the trust management techniques in the Internet of Things," IEEE Internet of Things Journal, vol. 6, no. 6, pp. 9326-9337, 2019.

[19] D. Airehrour, J. A. Gutierrez, and S. K. Ray, "SecTrust-RPL: A secure trust-aware RPL routing protocol for Internet of Things," Future Generation Computer Systems, vol. 93, pp. 860-876, 2019.

[20] D. Airehrour, J. Gutierrez, and S. K. Ray, "A trust-aware RPL routing protocol to detect blackhole and selective forwarding attacks," Journal of Telecommunications and the Digital Economy, vol. 5, no. 1, pp. 50-69, 2017.

[21] B. Rakesh, "Novel authentication and secure trust based RPL routing in mobile sink supported Internet of Things," Cyber-Physical Systems, vol. 9, no. 1, pp. 43-76, 2023.

[22] P. P. Ioulianou, V. G. Vassilakis, and S. F. Shahandashti, "A trust-based intrusion detection system for RPL networks: Detecting a combination of rank and blackhole attacks," Journal of Cybersecurity and Privacy, vol. 2, no. 1, pp. 124-153, 2022.

[23] B. Patel, J. Vasa, and P. Shah, "Forwarding Neighbor Based Sink Reputed Trust Based Intrusion Detection System to Mitigate Selective Forwarding Attack in RPL for IoT Networks," SN Computer Science, vol. 4, no. 4, p. 420, 2023.

[24] D. Airehrour, J. Gutierrez, and S. K. Ray, "Securing RPL routing protocol from blackhole attacks using a trust-based mechanism," in 2016 26th International telecommunication networks and applications conference (ITNAC), 2016: IEEE, pp. 115-120.

[25] N. Djedjig, D. Tandjaoui, and F. Medjek, "Trust-based RPL for the Internet of Things," in 2015 IEEE Symposium on Computers and Communication (ISCC), 2015: IEEE, pp. 962-967.

[26] D. Airehrour, J. Gutierrez, and S. K. Ray, "A testbed implementation of a trust-aware RPL routing protocol," in 2017 27th International Telecommunication Networks and Applications Conference (ITNAC), 2017: IEEE, pp. 1-6.

[27] N. Subramanian, S. M. GB, J. P. Martin, and K. Chandrasekaran, "HTmRPL++: a trust-aware RPL routing protocol for fog enabled Internet of Things," in 2020 International Conference on COMmunication Systems & NETworkS (COMSNETS), 2020: IEEE, pp. 1-5.

[28] R. Mehta and M. M. Parmar, "Trust based mechanism for securing iot routing protocol rpl against wormhole &grayhole attacks," in 2018 3rd International Conference for Convergence in Technology (I2CT), 2018: IEEE, pp. 1-6.

[29] T. ul Hassan, M. Asim, T. Baker, J. Hassan, and N. Tariq, "CTrust-RPL: A control layer-based trust mechanism for supporting secure routing in routing protocol for low power and lossy networks-based Internet of Things applications," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 3, p. e4224, 2021.

[30] N. Djedjig, D. Tandjaoui, F. Medjek, and I. Romdhani, "New trust metric for the RPL routing protocol," in 2017 8th International Conference on Information and Communication Systems (ICICS), 2017: IEEE, pp. 328-335.

[31] M. Sisodiya, V. Dahima, and S. Joshi, "Trust based Mechanism using Multicast Routing in RPL for the Internet of Things," in 2020 12th International Conference on Computational Intelligence and Communication Networks (CICN), 2020: IEEE, pp. 392-397.

[32] S. M. Muzammal, R. K. Murugesan, N. Z. Jhanjhi, M. Humayun, A. O. Ibrahim, and A. Abdelmaboud, "A trust-based model for secure routing against RPL attacks in internet of things," Sensors, vol. 22, no. 18, p. 7052, 2022.

[33] D. Al-Jumeily, D. Arshad, N. Tariq, T. Baker, H. Tawfik, and M. Asim, "A Lightweight Trust-enabled Routing in RPL-based IoT Networks Against Sybil Attack," PLoS One, vol. 17, no. 7, 2022.

[34] P. Thulasiraman and Y. Wang, "A lightweight trust-based security architecture for RPL in mobile IoT networks," in 2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC), 2019: IEEE, pp. 1-6.

[35] M. Hassan et al., "Gitm: A gini index-based trust mechanism to mitigate and isolate sybil attack in rpl-enabled smart grid advanced metering infrastructures," IEEE Access, vol. 11, pp. 62697-62720, 2023.

[36] M. Savitha and P. Basarkod, "A Trust and Energy based Routing using Gorilla Troops Optimization in RPL Networks," in 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2023: IEEE, pp. 1-7.

[37] S. M. Muzammal, R. K. Murugesan, N. Z. Jhanjhi, and L. T. Jung, "SMTrust: Proposing trust-based secure routing protocol for RPL attacks for IoT applications," in 2020 International Conference on Computational Intelligence (ICCI), 2020: IEEE, pp. 305-310.

[38] J. Jiang and Y. Liu, "Secure IoT routing: selective forwarding attacks and trust-based defenses in RPL network," arXiv preprint arXiv:2201.06937, 2022.

[39] J.-D. Kim, M. Ko, and J.-M. Chung, "Physical identification based trust path routing against sybil attacks on RPL in IoT networks," IEEE Wireless Communications Letters, vol. 11, no. 5, pp. 1102-1106, 2022.

[40] A. Lahbib, K. Toumi, S. Elleuch, A. Laouiti, and S. Martin, "Link reliable and trust aware RPL routing protocol for Internet of Things," in 2017 IEEE 16th International Symposium on Network Computing and Applications (NCA), 2017: IEEE, pp. 1-5.

[41] P. Karkazis, I. Papaefstathiou, L. Sarakis, T. Zahariadis, T.-H. Velivassaki, and D. Bargiotas, "Evaluation of RPL with a transmission count-efficient and trust-aware routing metric," in 2014 IEEE International Conference on Communications (ICC), 2014: IEEE, pp. 550-556.

Cybersecurity Awareness in Schools: A Systematic Review of Practices, Challenges, and Target Audiences

Abdulrahman Abdullah Arishia[1], Nazhatul Hafizah Kamarudinb[2], Khairul Azmi Abu Bakarc[3], Zarina Binti Shukurd[4], Mohammad Kamrul Hasan[5]

Master's Degree in Cybersecurity, Universiti Kebangsaan Malaysia (UKM), 43600 UKM Bangi, Selangor[1]
Research Coordinator, CYBER, Universiti Kebangsaan Malaysia (UKM)[2]
Assistant Dean (Teaching and CITRA), Universiti Kebangsaan Malaysia (UKM)[3]
APEL Q Coordinator, Universiti Kebangsaan Malaysia (UKM)[4]
Head of Network & Communication Technology (NCT) Lab CYBER, Universiti Kebangsaan Malaysia (UKM)[5]

***Abstract*—This systematic literature review examines cybersecurity awareness in schools, focusing on effective practices, challenges, and future directions. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, peer-reviewed publications in English were sourced from ACM Digital Library, IEEE Xplore, ScienceDirect, SpringerLink, and Emerald, covering the period from 2019 to 2024. Studies were included if they focused on cybersecurity awareness in primary and secondary educational settings, excluding those unrelated to educational contexts or published before 2019. A total of 816 records were identified, of which 220 were duplicates and removed. After screening and eligibility assessments, 14 studies met the inclusion criteria. Risk of bias was minimized by adhering to strict inclusion criteria, such as limiting the review to high-quality, peer-reviewed studies, and ensuring consistency in the data extraction process. The review highlights effective practices such as using serious games, mobile apps, and tailored programs to enhance cybersecurity awareness. Challenges include inconsistent curricula, insufficient parental involvement, and resource limitations. These results emphasize integrating cybersecurity education across school curricula and regularly updating content to reflect evolving threats. Limitations include the exclusion of non-English and non-peer-reviewed studies. Future research should consider broader contexts and additional sources.**

Keywords*—*Cybersecurity awareness; threats; awareness programs; education; school security

I. Introduction

In the digital age, educational institutions have progressively embraced technology, integrating it extensively into classroom instruction and administrative management. While this integration offers numerous advantages, it also significantly increases schools' exposure to cybersecurity risks, threatening the integrity of educational delivery and the privacy of student and teacher data [1,2]. The vulnerability of schools to cyber threats is particularly exacerbated by limited resources, which leave educational networks open to unauthorized access and data theft [3]. Additionally, the swift transition to e-learning platforms, particularly during the COVID-19 pandemic, has expanded the attack surface, providing cybercriminals with more opportunities to exploit outdated or unpatched systems [1,4].

Historically, responses to cybersecurity threats within educational settings have tended to be reactive rather than proactive. There is now a critical and urgent need to enhance cybersecurity awareness and improve practices across all stakeholders, particularly as the Internet becomes an indispensable part of educational infrastructure [5,6]. Effective cybersecurity awareness programs must encompass a wide audience, including students, who require adequate guidance and supervision to navigate online risks safely [7]. The current systematic literature review aims to meticulously examine the state of cybersecurity awareness in schools, delineating effective practices and identifying ongoing challenges. The review addresses the growing complexity and scale of cyber threats and underscores the necessity for comprehensive strategies that integrate public education, technology updates, and community involvement [8,9]. The outcomes of this review are designed to inform educational policymakers, school IT administrators, and educators, providing them with strategic insights to develop and implement robust, effective cybersecurity education and awareness programs.

Research Questions:

1) What are effective practices for improving cybersecurity awareness in schools?

2) What are schools' current problems and challenges in implementing cybersecurity awareness programs?

3) Who is the target audience for the current cybersecurity awareness assessment?

By addressing these questions, the research aims to contribute significantly to the body of knowledge on cybersecurity in education, facilitating the development of secure learning environments that effectively leverage digital technologies.

II. Literature Review

Cybersecurity awareness has become an essential aspect of contemporary education systems, driven by the widespread

integration of digital technologies in schools. As students, educators, and administrators increasingly depend on digital platforms, educational institutions face heightened exposure to cyber threats. This literature review examines the practices, challenges, and target audiences in fostering cybersecurity awareness in school environments.

Yuliana (2022) underscores the importance of raising cybersecurity awareness among children, particularly in the context of online schooling, which increases their vulnerability to cyberattacks and malware. The study demonstrates that digital literacy training can significantly enhance children's awareness of cybersecurity. It advocates teaching children how to avoid risky online behaviour, such as falling victim to phishing scams, pornography, cyberbullying, identity theft, and privacy breaches. Additionally, it stresses the importance of educating children on password security and fostering caution when engaging in online gaming [10].

Ondrušková and Pospíšil (2023) argue that the increasing use of the Internet necessitates adequate cybersecurity awareness to mitigate the risks and dangers associated with the online environment. Their findings reveal only a moderate level of cybersecurity awareness during initial testing and show that one-off training sessions have an insignificant impact on improving online behavior. The study concludes that one-time interventions are insufficient and recommends integrating cybersecurity awareness education throughout the entire educational process to effectively enhance online safety skills [11].

Nehrar and Deepanshi (2023) highlight the critical role of cybersecurity awareness and education programs in an era characterized by pervasive digital connectivity and cyber threats. They emphasize that such programs are effective in enhancing knowledge, fostering behavioral changes, and ensuring long-term impact on cybersecurity practices [12].

Prümmer et al. (2024) emphasize the growing significance of cybersecurity in mitigating financial losses, productivity disruptions, and reputational damage caused by cyberattacks. Their research highlights the pivotal role of end-user behavior in achieving robust cybersecurity within organizational settings. Comprehensive training programs are identified as an effective means of improving cybersecurity behavior, with most studies reporting positive outcomes regardless of the training method or topic. Notably, game-based training methods are frequently employed, demonstrating their effectiveness in engaging participants and enhancing cybersecurity practices [13].

The integration of technology into education has brought both opportunities and challenges, particularly in the realm of cybersecurity. Buyu & Ogange (2021) notes that the rapid adoption of digital tools in schools has introduced significant cybersecurity risks, including malware, ransomware, phishing, and denial-of-service attacks, which disrupt teaching, compromise data security, delay assessments, and erode trust between teachers and students. These challenges are exacerbated in developing countries, where schools often lack the funding, infrastructure, and expertise to effectively address these threats. Moreover, teachers frequently lack the necessary cybersecurity knowledge to protect themselves and their students, while young learners remain particularly vulnerable due to limited awareness of online risks and poor adherence to safe practices [14].

The transition to online education, particularly during the COVID-19 pandemic, has further amplified these challenges. Al-Fatlawi (2024) highlights how the rapid shift to digital learning exposed vulnerabilities, as many educators lacked the skills and resources to secure online environments. Students with low cybersecurity awareness became frequent targets of phishing, malware, and data breaches. The lack of structured training for teachers has compounded these issues, preventing them from adequately guiding students in safe online behaviours. Additionally, disparities in motivations and levels of understanding between different demographics, including educators and students, complicate the development of comprehensive and effective cybersecurity strategies [15].

Efforts to incorporate cybersecurity awareness into school curriculums face significant hurdles. According to Triplett (2023), while methods such as game-based learning show promise in engaging students and raising awareness, they often fall short of addressing the full range of skills and knowledge required to tackle real-world cybersecurity threats. A shortage of trained mentors and educators specializing in cybersecurity further diminishes the effectiveness of these initiatives. This gap underscores the need for sustained investment in teacher training and the integration of cybersecurity education across all levels of schooling [16].

The growing reliance on the Internet in education has also led to an increase in cyber-related issues such as cyberbullying, online fraud, racial abuse, pornography, and gambling, as noted by Sareen and Jasaiwal (2021). These problems arise from a lack of awareness among internet users and disproportionately affect children, especially in the context of online education. Limited teacher expertise, inadequate funding, and insufficient resources hinder schools' ability to implement robust cybersecurity education programs. Addressing these challenges requires a collaborative effort among educators, parents, policymakers, and media platforms to cultivate a culture of cybersecurity awareness from an early age [17].

The literature review underscores the critical need for robust cybersecurity awareness initiatives in contemporary education systems. As digital technologies become integral to learning, schools face escalating cyber threats that compromise safety, disrupt teaching, and erode trust. Studies highlight the importance of integrating comprehensive and sustained cybersecurity education across all levels of schooling to address vulnerabilities among students and educators. Effective approaches, such as digital literacy training and game-based learning, show promise but require adequate resources, skilled mentors, and long-term strategies to ensure meaningful impact. Addressing these challenges necessitates collaboration among educators, parents, policymakers, and media, fostering a culture of cyber safety that empowers learners and mitigates the risks of an increasingly connected educational landscape.

III. METHODOLOGY

To deliver an exhaustive review of cybersecurity awareness within school and educational settings and to bridge identified research gaps, this study is grounded in the PRISMA ("Preferred

Reporting Items for Systematic Reviews and Meta-Analyses") framework [18].

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement is an authoritative and updated guideline that enhances transparency in the reporting of systematic reviews. It addresses the rationale, methods, and outcomes of reviews [19]. The choice of the PRISMA framework was driven by its standardization, which ensures consistency and reliability in the systematic review process; and its comprehensiveness [20], which is crucial to addressing the multi-faceted issues of cybersecurity comprehensively and practically in the education sector [21]. The review process was initiated using the PRISMA framework, which consists of four phases: identification, screening, eligibility, and inclusion. The flow diagram, providing detailed information and statistics, is shown in Fig. 1. The four phases are explained in detail below. Data collection was conducted by a single reviewer to maintain consistency in the extraction process. Four instructors provided guidance and oversight, ensuring the integrity and accuracy of the data collection.

A. Identification Phase

1) Selecting database: In the initial phase of the literature review for a systematic review of cybersecurity awareness in schools, a range of scholarly and normative databases were carefully selected to ensure comprehensive coverage of the relevant literature. Selected databases include ACM Digital Library, IEEE Xplore, ScienceDirect; SpringerLink, and Emerald .These platforms were strategically selected to access diverse, high-quality academic resources essential for the comprehensive exploration of current studies.

2) Selecting keywords: In the keyword selection phase of the Systematic Literature Review (SLR) on cybersecurity awareness in schools, specific research questions were identified to guide the research process.

- To explore practices for improving cybersecurity awareness in schools, (“cybersecurity” OR “security” OR “cybersecurity”) AND (“awareness” OR “learning”) AND (“framework” OR “model”), and Added Educational Context More detailed (“Education Sector”, “Academy”, “School” or “Educational Institution”) were used.
- Regarding the current problems and challenges that schools face in implementing cybersecurity awareness programs, the keywords (“challenges of cybersecurity education” OR “barriers to cybersecurity training” OR “cyber threats”) and (“education sector”) were used.
- For the target audience in cybersecurity awareness assessments, the keywords (“cybersecurity” OR “security” OR “cybersecurity”) AND (“awareness” OR “learning”) AND (“assessment” OR “evaluation”) were used.

These keywords were carefully selected to ensure comprehensive coverage and relevance to the study's specific areas of investigation.

3) Initial search :In the initial search stage of the systematic literature review (SLR) following the PRISMA framework, a comprehensive search was conducted across the selected databases. This search, based on the keywords identified in the previous stage, aimed to retrieve relevant scientific papers and yielded a total of 816 records.

4) Inclusion/exclusion criteria: To maintain the integrity and relevance of the systematic review, rigorous inclusion and exclusion criteria have been applied focused on impurity removal. The review is confined to studies published within the past five years (from 2019 to 2024), ensuring the data's freshness and pertinence to current contexts. Sources have been restricted to peer-reviewed articles which are critical for upholding high research standards and credibility. Additionally, all studies included in the review are published in English, setting a necessary linguistic criterion that aids in uniform comprehension and analysis. Furthermore, only studies classified as Q1 or Q2 are included to ensure the inclusion of high-quality research. On the exclusion front, any studies published before 2019 are omitted to reinforce the recency of the included research. Non-peer-reviewed materials are also excluded to preserve the scientific integrity and quality of the review. Studies not available in English or not meeting the quality classification are systematically excluded.

5) Duplicate removal: All the duplicate records were eliminated, which summed up to 220. It resulted in many duplicate records. Finally, the total remaining articles after the identification phase was .596

B. Screening Phase

Following the identification phase, the screening phase involved a detailed review of the title and abstract of each selected article 5 96articles met the criteria during this phase and were advanced to the subsequent eligibility phase. The screening was conducted using new inclusion and exclusion criteria specifically tailored for this stage, as outlined below.

Inclusion and Exclusion Criteria: During this phase of the PRISMA systematic review, criteria to refine the selection of articles were applied. Only studies that specifically focused on cybersecurity awareness within educational settings were included, as articles that did not directly focus on cybersecurity awareness within educational settings were excluded. For example, the study by Renaud & Ophoff, which explored cybersecurity implementation in SMEs rather than educational institutions, was excluded [22]. Additional exclusions included works such as those reported in Reference, Although it presents a cybersecurity awareness framework based on the behaviour change wheel, did not specifically target educational settings [23]. When information was missing or unclear, assumptions were made based on the available data. For instance, if a study did not specify the age range of students but was conducted in primary schools, it was assumed that the participants were aged 5–12 years, based on common educational structures. Similarly, for studies with incomplete intervention details, those elements were excluded from the analysis unless sufficient information justified their inclusion. This criterion led to the exclusion of 252 records that solely addressed information security awareness,

cybersecurity awareness, or other general security issues without a direct educational context.

C. Eligibility Phase

A total of 344 articles that successfully passed the second screening stage were chosen for in-depth full-text review during the eligibility stage. All articles were accessed, and the established filtering criteria were applied to evaluate each article at this stage.

Inclusion and Exclusion Criteria: During the eligibility phase of the PRISMA systematic review, stringent criteria were applied to refine the selection of articles, focusing on studies specifically within the context of schools to ensure alignment with the research aim. For example, the framework proposed in [24] for assessing cybersecurity maturity in higher education institutes (HEIs) in the UK, despite its comprehensive approach and use of recognized Capability Maturity Model (CMM) methodology, was excluded due to its specific focus on higher education. Rather than the specific educational context (schools), this careful selection process resulted in a significant reduction in eligible articles. This drastic decline stems from the overwhelming number of studies focusing on cybersecurity in educational settings, particularly schools, highlighting the challenge of finding research that focuses exclusively on the educational context, especially in schools. A careful selection and exclusion process ensures that the final set of articles strongly represents the most relevant and up-to-date research regarding cybersecurity awareness in schools, thus providing a solid foundation for identifying effective strategies and challenges specific to this educational context. In addition, a group of articles was excluded due to a lack of access to the full text, which will be illustrated in Fig. 1. Given that the studies evaluated during the Q1-2 phases are peer-reviewed, validated articles that have been thoroughly evaluated and reviewed by referees, issues of bias due to missing data were not encountered. This rigorous selection process ensures the reliability and integrity of the systematic review findings, providing a solid foundation for identifying effective strategies and challenges specific to cybersecurity awareness in schools.

D. Included Phase

All articles that successfully passed the third eligibility stage were carried forward into the final in-depth analysis of this study. Consequently, 14 articles that met all the criteria of this search were included for detailed examination.

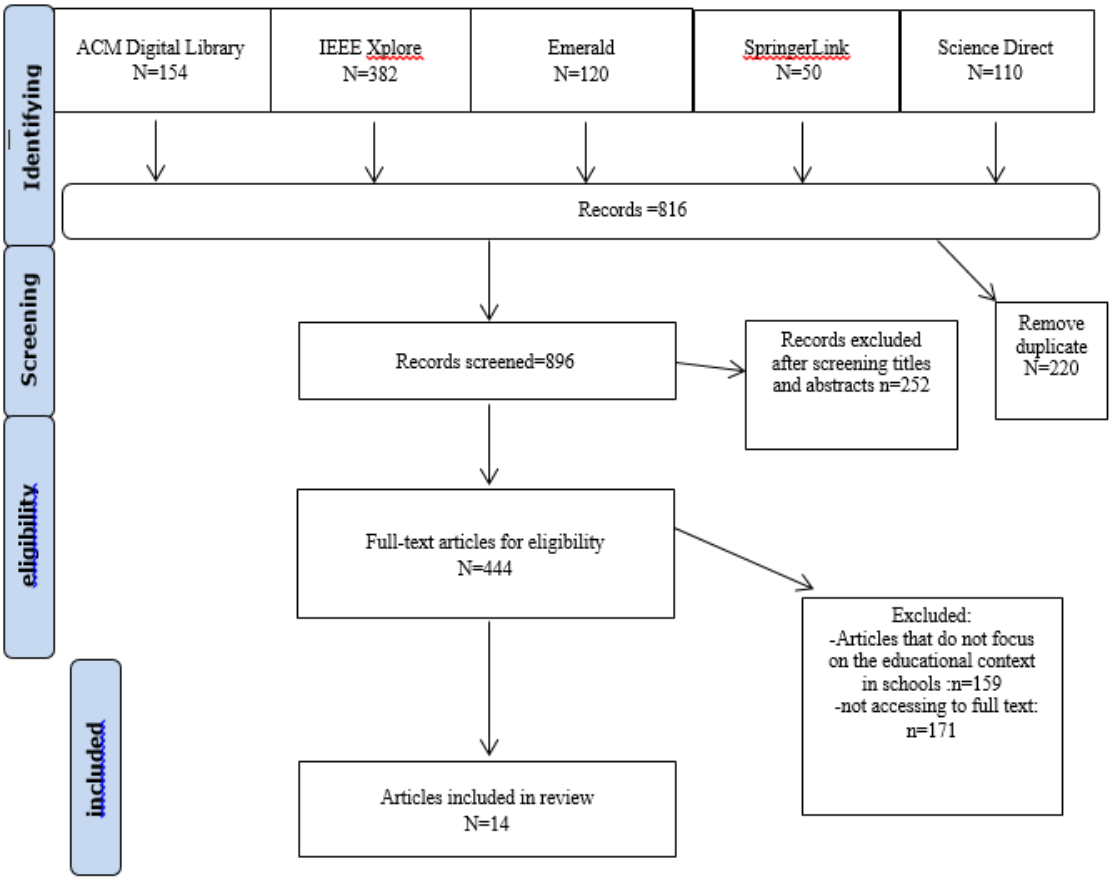


Fig. 1. Prisma flow chart.

IV. Results and Analysis

The results of this systematic review are presented based on the research questions through the analysis of the included studies (n = 14). Fig. 2 shows the distribution of the studied articles over the years in which they were published, these studies span from 2020 to 2024.

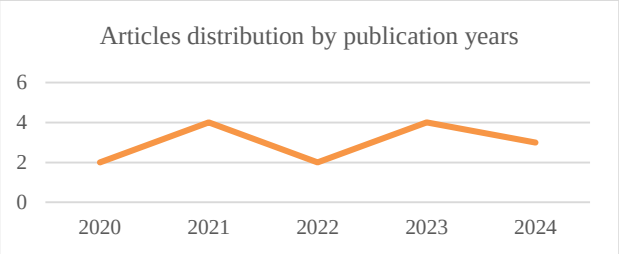


Fig. 2. Articles distribution by publication years.

While Fig. 3 shows the classification of the types of studied articles:

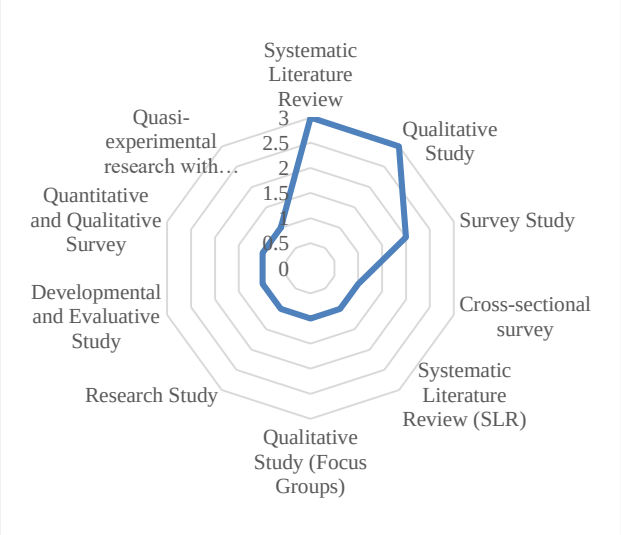


Fig. 3. The classification of the types of studied articles.

Addressing bias in systematic reviews

- Consultation with Experts: Collaboration with experts in the field of cybersecurity education was done to uncover any overlooked or insufficiently reported elements of cybersecurity awareness initiatives.
- The comprehensiveness of Reporting: Each study's reporting thoroughness was meticulously evaluated. Any discrepancies were looked for where expected results or data were absent and unaccounted for in the reports.

RQ1: what are effective practices to improve Cybersecurity awareness in schools?

Digital comics, serious games, and mobile applications are extensively used to foster engaging and interactive cybersecurity education, enhancing learning experiences by employing real-world scenarios and relevant challenges. This approach is particularly effective in engaging children and enhancing their understanding of cybersecurity concepts [25].Techniques such as the use of a web-based Learning Content Management System (LCMS) and mobile apps provide game-based cybersecurity education that captivates students' interest through interactive gameplay, which is crucial in maintaining student engagement and reinforcing cybersecurity [26]. Additionally, the gamified framework "Cyber-Hero" incorporates narrative techniques with serious games to offer an immersive learning experience that emphasizes critical cybersecurity skills, such as creating robust passwords and understanding online threats [27].

Tailored educational programs and comprehensive integration of cybersecurity into curricula are fundamental to creating a robust educational foundation. Segmented educational programs cater to specific demographic groups such as students, teachers, and parents, ensuring that each group receives relevant and effective cybersecurity knowledge [28]. Furthermore, the integration of cybersecurity topics throughout all K-12 levels ensures that students develop a comprehensive understanding from an early age, which is crucial for building a solid cybersecurity foundation [29].

Comprehensive cybersecurity education also involves the integration of national cybersecurity strategies into school curricula to align educational efforts with broader national security objectives [30]. Narrative-based learning techniques, which use storytelling and problem-solving to enhance engagement and understanding, are also highlighted as effective methods in making cybersecurity education more relatable and effective [11] [31].

Enhanced training programs for teachers and parents are critical to equipping them with the necessary skills to effectively impart cybersecurity knowledge and oversee children's online activities. These training programs not only enhance teachers' ability to teach cybersecurity effectively but also empower parents to supervise their children's online interactions more effectively [29]. Regular training sessions for both educators and parents help to ensure that children are adequately supervised and educated about the risks and safety measures associated with cyber activities [31] [32]. Additionally, the development of E-Safety frameworks that address ICT risks and the incorporation of feedback mechanisms into cyber wellness programs help monitor and control the effectiveness of cybersecurity education, adapting to new challenges and ensuring the safety of learners [33] [28]. Curriculum content that is framed around broad categories and uses innovative teaching methods like gamification not only enhances learning but also encourages active participation from students, making the learning process both enjoyable and educational [34]. Finally, the recommendations for interactive and practical teaching methods aim to engage students effectively and ensure that cybersecurity education evolves in response to new cybersecurity challenges, thus maintaining its relevance and effectiveness in educating young learners [11]

TABLE I. ANALYSIS OF PRACTICES OF CYBERSECURITY AWARENESS IN SCHOOL

Ref	Authors & year	study appraisal	Study type	Practices
[25]	Farzana Quayyum , Daniela S. Cruzes, Letizia Jaccheri (2021)	Q1	Systematic Literature Review	The review highlights multiple approaches to raising cybersecurity awareness among children. These include digital comics, serious games, and mobile apps focused on cybersecurity education. It emphasizes engaging children through interactive learning that covers internet fundamentals, privacy awareness, and risk management strategies.
[28]	T€urkera, Mıhcı and C¸akmakb, Ebru Kılıc¸ (2020)	Q2	Cross-sectional survey	The research utilizes detailed surveys to identify specific gaps in cyber wellness knowledge among students, teachers, and parents. Based on survey results, it recommends tailored educational interventions that address these specific knowledge gaps. Key practices include 1. Segmented Educational Programs: Developing age and role-specific educational materials that cater specifically to students, teachers, or parents. 2. Interactive Learning Modules: Implementing interactive and engaging learning modules that cover key topics like netiquette, cyberbullying, and online privacy practices. 3. Awareness Campaigns: Running targeted awareness campaigns that encourage safe and responsible internet use, focusing on the risks of internet addiction and the importance of copyright laws. 4. Feedback Mechanisms: Incorporating feedback mechanisms to continuously assess and improve the effectiveness of cyber wellness programs.
[29]	Ahmed Ibrahim, Marnie McKee, Leslie F. Sikos, Nicola F. Johnson (2022)	Q1	Systematic Literature Review (SLR)	The review suggests several targeted practices to enhance cybersecurity awareness: 1. Development of clear and consistent terminology across cybersecurity education to standardize teaching materials. 2. Incorporation of cybersecurity topics into curricula at all K-12 levels, ensuring a comprehensive understanding from a young age. 3. Enhanced teacher training programs to provide educators with the necessary skills and knowledge to effectively teach cybersecurity. 4. Use of interactive and practical learning modules to engage students and reinforce cybersecurity concepts
[32]	Stephanie Bannon, Tracy McGlynn, Karen McKenzie, Ethel Quayle (2024)	Q1	Qualitative Study (Focus Groups)	Inclusive Education Programs: Implement specialized cybersecurity education programs that account for the diverse needs of ASN students, emphasizing practical strategies to manage online risks. Parent and Teacher Training: Provide training for parents and teachers on how to support ASN students in navigating online spaces safely. Development of Tailored Online Safety Rules: Create school policies that address the specific online safety needs of ASN students, ensuring that they are comprehensible and enforceable.
[31]	Mohamed Ayyash, Tariq Alsboi, Omar Alshaikh, Isa Inuwa-Dutse, Saad Khan, and Simon Parkinson (2024)	Q1	Survey Study	Integration of Cybersecurity in Curriculum: Implement comprehensive cybersecurity education in schools. Narrative-based Learning: Emphasize storytelling, problem-solving, and video-based teaching methods to enhance engagement and understanding. Parent and Teacher Workshops: Organize regular training for parents and teachers to equip them with effective strategies to supervise and educate children on cybersecurity.
[35]	Ana Kovačević, Nenad Putnik, Oliver Tošković (2020)	Q1	Survey Study	Curriculum Improvement: Based on the study's findings, schools can enhance their curriculum to include more comprehensive cybersecurity education, focusing on practical knowledge and secure behaviors. Educational Programs: Development of targeted educational programs that address identified gaps in student knowledge and behaviors.
[34]	Rahime Belen Sağlam, Vincent Miller, Virginia N. L. Franqueira (2023)	Q1	Systematic Literature Review	- Curriculum content framed around six broad categories -Innovative teaching methods like gamification for a 'hands-on' experience. - Engaging students in curriculum design - Incorporating a 'bottom-up' approach listening to children's views.
[36]	Abel Moyo, Theo Tsokota, Caroline Ruvinga, Colletor T. Chipfumbu Kangara(2021)	Q1	Qualitative Research	- Development of an E-Safety framework to teach and safeguard learners from ICT-related risks. - Integration of cybersecurity knowledge into school curricula. - Continuous education, monitoring, and control of ICT use.
[37]	Farzana Quayyum, Jonas Bueie, Daniela S. Cruzes, Letizia Jaccheri, Juan Carlos Torrado Vida (2021)	Q1	Qualitative Study	- The study found that parents often discuss online security with their children at home, using real-world incidents reported in the media as teachable moments. - Schools in Norway enhance cybersecurity awareness by collaborating with organizations like Barnevakten, Bruk Hue, and Medietilsynet. These organizations assist in providing targeted cybersecurity training and lectures, not just for students but also for teachers and parents, fostering a comprehensive educational environment.
[27]	Hani Qusa, Jumana Tarazi (2021)	Q1	Research Study	- A new educational method is presented, the Gamification Framework (Cyber-Hero), which utilizes a specially designed gamified framework to boost early cybersecurity training in high schools. - This framework incorporates narrative instruction into serious games, which are strategically used to teach critical cybersecurity concepts, such as creating strong passwords. This method aims to make learning both engaging and effective.

				- The framework emphasizes two main dimensions—motivation and deployment. Motivational strategies include invoking emotions like fear to encourage proactive learning behaviors, while deployment strategies involve periodic gameplay and iterative learning, ensuring students can see their progress and are motivated to improve.
[30]	Saleh AlDaajeha , Heba Saleousa , Saed Alrabaeea, , Ezedin Barkaa , Frank Breitingerb , Kim-Kwang Raymond Chooc (2022)	Q1	Qualitative Analysis	- The study emphasizes the importance of integrating national cybersecurity strategies (NCSP) into educational curricula. - It suggests developing and aligning cybersecurity educational practices to meet the strategic objectives laid out by national policies.
[26]	Filippos Giannakas, Andreas Papasalouros, Georgios Kambourakis, Stefanos Gritzalis, 2023	Q2	developmental and evaluative study	- Utilization of a web-based Learning Content Management System (LCMS) and mobile application to deliver engaging, game-based cybersecurity education. - Application of the ARCS model of motivation to ensure the educational content is engaging and effective. - Flexible learning modes (standalone or client/server) to accommodate different learning environments and scenarios.
[13]	Julia Prümmer, Tommy van Steen, Bibi van den Berg (2024)	Q1	Systematic Literature Review	- Incorporate game-based training methods which were most frequently used and shown to be effective. - Employ a variety of training approaches, including simulation and presentation-based, to cater to different learning preferences and enhance effectiveness. - Use multi-method training campaigns to cover a broader range of cybersecurity topics effectively. - Ensure training includes interactive and engaging content to maintain participant interest and improve learning outcomes.
[11]	Dana Ondrušková, Richard Pospíšil, (2023)	Q2	Quasi-experimental research using time-series data to evaluate the impact of cyber security training on the online behavior of school children.	- Recommends integrating cybersecurity education into the ongoing educational curriculum rather than relying on one-off interventions. - Suggests using interactive and practical teaching methods to engage students effectively. - Recommends ongoing assessments and adaptations of educational content to ensure it meets the evolving challenges of cybersecurity.

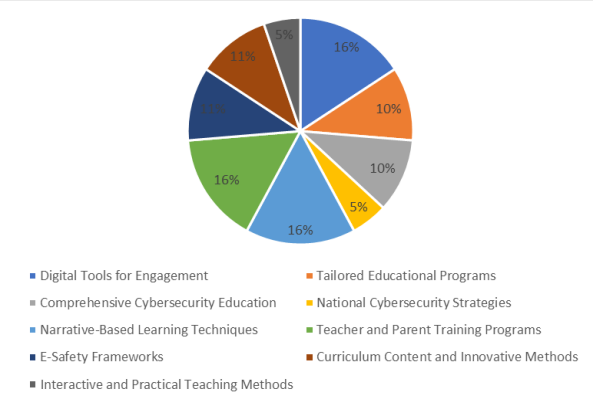


Fig. 4. Effective cybersecurity awareness practices in schools.

RQ2: What are the current issues and challenges that schools face in implementing cybersecurity awareness programs?

- Comprehensive Curriculum and Effective Educational Approaches*:*

The development of a comprehensive, consistent curriculum is critical yet challenging, necessitating curricula that not only cover a broad range of cybersecurity topics but are also adaptable to various learning needs, including those of students with Additional Support Needs (ASN). The challenge is compounded by often inconsistent educational content across different programs, which can lead to gaps in student knowledge and preparedness. Furthermore, traditional training methods have shown limitations in effectively engaging and educating students, highlighting the need for innovative approaches that cater effectively to today's diverse student bodies [25] [34] [26]. Fig. 4 shows effective cybersecurity awareness practices in schools.

- Parental Involvement and Supervision:

The role of parents in reinforcing cybersecurity education at home is crucial, yet many parents lack adequate cybersecurity awareness, which undermines their ability to support their children's learning and enforce safe online practices. Schools face the challenge of empowering parents with the necessary tools and knowledge to better manage and monitor their children's online activities. This includes providing parents with training and regular updates on cybersecurity practices and potential threats, thereby extending the learning environment beyond the classroom and fostering a holistic approach to cybersecurity education [31] [37]

- Knowledge and Awareness Gaps:

A major barrier to effective cybersecurity education is the lack of clarity in cybersecurity terminology and the absence of systematic teaching methods, particularly at the K-12 level. These gaps not only affect students but also impact educators and parents, complicating the task of achieving a comprehensive

understanding of cybersecurity across all stakeholders. The disparities in awareness based on demographic factors such as gender, class, and daily internet usage necessitate tailored educational efforts to ensure that all students, regardless of background, receive adequate and effective cybersecurity training [28] [29] [35].

- Human and Resource Limitations:

Human error is identified as a significant vulnerability in cybersecurity, exacerbated by a shortage of trained specialists within the educational sector. This shortage hampers the ability of schools to develop and deliver effective cybersecurity education, further strained by inadequate resources. Addressing these limitations requires not only more specialized training for educators but also sufficient funding to ensure that schools can afford to implement robust cybersecurity programs [27] [30].

- Engagement and Methodology Effectiveness:

Keeping students engaged in cybersecurity learning is a major challenge, particularly for younger audiences who may find traditional methods unappealing. The effectiveness of various training methodologies varies widely, and contradictory findings in research on optimal training methods create confusion about the best approaches to take. This situation calls for continuous experimentation and adaptation of teaching strategies to find what works best in different educational contexts [26] [13]

- Monitoring and Control Difficulties:

Implementing effective digital monitoring in schools involves navigating complex privacy issues and technical challenges. Schools must balance the need to monitor students' online activities to ensure their safety with the need to respect privacy rights. Effective policies and tools are needed to navigate these waters successfully, which requires ongoing dialogue among educators, parents, and policymakers [33]

- Intervention and Training Impact:

The impact of cybersecurity interventions and training programs varies widely, with some interventions showing limited effects on long-term behavioural changes and awareness among children. The success of these interventions often hinges on their duration and intensity. Well-designed, sustained programs that are integrated into the regular curriculum are more likely to have a lasting impact, highlighting the need for continuous evaluation and adaptation of these programs to meet evolving educational needs [11].

TABLE II. ANALYSIS OF CHALLENGES OF CYBERSECURITY AWARENESS IN SCHOOL

challenges	No. references
Comprehensive Curriculum and Effective Educational Approaches	[25] [34] [26]
Parental Involvement and Supervision	[31] [37]
Knowledge and Awareness Gaps	[28] [29] [35]
Human and Resource Limitations	[27] [30]
Engagement and Methodology Effectiveness	[26] [13]
Monitoring and Control Difficulties	[36]
Intervention and Training Impact	[11]

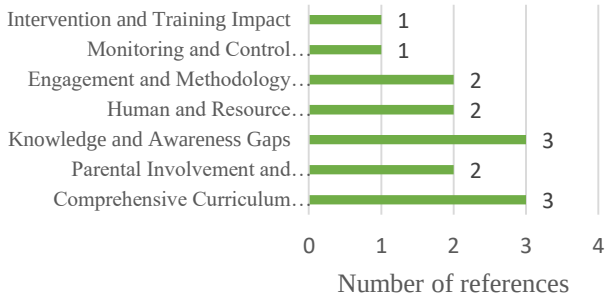


Fig. 5. Challenges of cybersecurity awareness.

RQ3: Who is the target audience in the current assessment of cybersecurity awareness?

The current assessment of cybersecurity awareness (Fig. 5) comprehensively addresses a diverse array of target audiences, each identified as critical in the ongoing efforts to enhance cybersecurity education. Primary school students are among the youngest learners, and their engagement is facilitated through digital tools such as comics, serious games, and mobile applications. These tools are essential in introducing foundational cybersecurity concepts in an age-appropriate and engaging manner, fostering early awareness and safe online behaviours [26] [25]. Secondary and high school students represent a slightly older demographic that faces more complex online risks, including cyberbullying, online privacy issues, and password security. For these students, educational programs often incorporate interactive and gamified learning experiences designed to maintain their engagement while reinforcing critical cybersecurity skills. These approaches are tailored to address the developmental and cognitive abilities of this age group, ensuring that the learning is both relevant and effective [28][33][27][30]. Teachers are a pivotal audience in this context, as they are the primary facilitators of cybersecurity education within schools. Training programs for teachers are therefore crucial, providing them with the necessary knowledge and pedagogical tools to effectively deliver cybersecurity education. Such programs also emphasize the importance of continuous professional development, enabling teachers to stay updated with the latest cybersecurity trends and threats [28][31][33]. Meanwhile, parents are recognized as the first line of defense in a child's online life. However, many parents lack the requisite knowledge and tools to effectively monitor and guide their children's online activities. Studies recommend targeted training for parents to improve their understanding of cybersecurity risks and how to communicate these risks effectively to their children, thereby extending cybersecurity education from the classroom to the home environment [31][37]

Curriculum designers and policymakers play a strategic role in shaping the educational landscape to include comprehensive cybersecurity education. Their responsibility involves integrating cybersecurity topics into the broader educational curriculum, ensuring that these efforts are aligned with national security objectives and are adaptable to the rapidly evolving digital threats [29] [11]. Additionally, in the corporate sector, corporate trainers and HR professionals are identified as key

audiences for cybersecurity training. These professionals are tasked with implementing training programs that not only educate employees about cybersecurity risks but also engage them through innovative methods such as game-based learning, which has been shown to enhance the effectiveness of these programs[13]. Finally, cybersecurity professionals are a crucial audience, as they are directly involved in protecting organizations from cyber threats.

The studies underscore the need for advanced educational frameworks tailored to these professionals, focusing on developing their skills to address the dynamic and complex nature of cyber threats. These frameworks are designed to enhance the preparedness of cybersecurity teams, ensuring they are equipped to handle real-world challenges effectively [30][13].

This comprehensive, multi-audience approach to cybersecurity awareness is critical in creating a resilient educational ecosystem that prepares all relevant stakeholders to navigate the digital world safely and responsibly.

TABLE III. ANALYSIS OF TARGET AUDIENCES

Category	Reference Numbers
Primary School Students	[26][25]
Secondary School Students	[28] [34] [36] [27]
K-12 Students	[29]
Children Aged 13-15	[11]
Special Needs Students	[32]
Teachers	[28] [36] [31]
Parents	[31] [37]
Curriculum Designers and Policymakers	[26] [29] [11]
Cybersecurity Professionals	[30] [13]

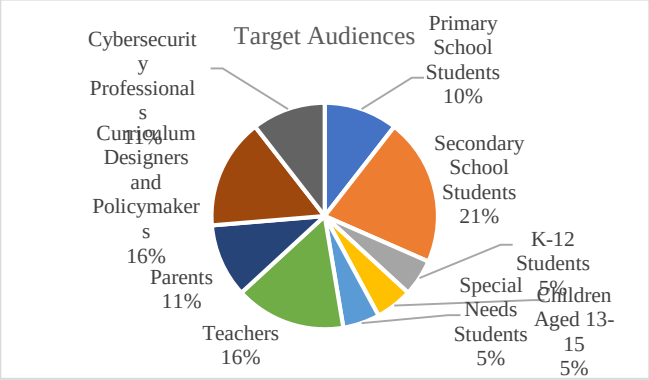


Fig. 6. Target audiences for cybersecurity awareness.

This systematic review systematically explores key practices, challenges, and target audiences (Fig. 6) in advancing cybersecurity awareness within educational institutions. The findings demonstrate that strategies such as the integration of digital tools, embedding cybersecurity within K-12 curricula, and targeted training for educators and parents significantly enhance student engagement and comprehension. However, challenges such as inconsistencies in curricular content, insufficient parental involvement, and resource constraints persist as significant barriers. The review underscores the necessity of tailoring cybersecurity practices to diverse target groups, including students, teachers, parents, and policymakers, to develop a comprehensive and resilient cybersecurity education framework. Crucially, the practices identified in the literature varied substantially depending on the target audience, leading to a wide range of specific outcomes for each group. The certainty of the evidence was evaluated, with particular attention to directness, ensuring that the findings closely aligned with the research questions and provided a robust foundation for the recommendations, despite some variations in study design and reporting rigor.

V. DISCUSSION

The findings from this systematic literature review reveal several dimensions of cybersecurity awareness in schools, emphasizing both effective practices and persistent challenges. The necessity for comprehensive and adaptable cybersecurity curricula is underscored, reflecting a critical need to bridge the gap between the evolving demands of the digital world and current educational offerings. This aligns with Blažič & Blažič, 2022 [38], who highlight the disconnect between educational content and the realities of digital threats, suggesting a pressing need for curriculum updates that mirror current technological landscapes. Innovative teaching methods, particularly those incorporating interactive tools like serious games and mobile apps, have proven effective. These methods resonate with findings from Jin et al., 2018, who confirm the efficacy of game-based learning in engaging students and enhancing their understanding of complex cybersecurity concepts [39]. However, the review also points to significant barriers in the development and implementation of such curricula, notably the inconsistency across educational programs which leads to gaps in cybersecurity knowledge and preparedness, as discussed by Lehto, 2022, this inconsistency can leave students underprepared for the cybersecurity challenges they face [40]. Furthermore, the review brings attention to the necessity of inclusive education strategies, particularly for students with Additional Support Needs (ASN). Ensuring that cybersecurity education is accessible to all students is crucial, a sentiment echoed by Ondrušková & Pospíšil, 2023 , who stress the importance of tailored approaches for students with special needs to safely navigate the digital world [11]. Parental engagement emerges as another vital component. The effectiveness of cybersecurity education often hinges on the support and knowledge of parents, yet many lack the necessary skills to guide their children in safe online behaviors, a gap highlighted by Al-Naser et al., 2019 emphasizes the need for programs that not only educate students but also empower parents with the necessary cybersecurity knowledge[41].

Additionally, the review identifies widespread gaps in knowledge among students, teachers, and parents, exacerbated by unclear cybersecurity terminology and a lack of systematic teaching approaches. This challenge is supported by O'Brien, 2019, who notes that the absence of standardized cybersecurity curricula in public schools underscores a widespread educational deficiency [42].

Demographic factors such as age, gender, and internet usage habits also influence cybersecurity awareness, necessitating more personalized educational interventions to address these disparities, as noted by Fatokun et al., 2019, this approach ensures that cybersecurity education is effective and inclusive, catering to the diverse needs of all students [43].

Moreover, the scarcity of trained cybersecurity specialists and adequate funding further constrains the ability of schools to offer effective cybersecurity education, a situation highlighted by Catota et al., 2019 argue for increased investment in both human resources and financial support to enable the development of comprehensive cybersecurity programs [44]. Issues of privacy and the need for effective digital monitoring are also discussed, pointing out the increasing complexity of maintaining privacy in an age of widespread digital surveillance[45]. This complexity necessitates clear policies and careful decision-making to balance safety and privacy rights effectively. Lastly, the variable impact of cybersecurity interventions is noted, with some programs significantly enhancing awareness and behavior, while others show limited long-term effects. Nasir, 2023 supports this observation, suggesting that the effectiveness of cybersecurity education programs can vary greatly depending on their design and implementation, highlighting the need for well-structured, sustainable initiatives[46].

Based on the discussion, this systematic literature review outlines strategies to boost cybersecurity awareness across educational levels through adaptable curricula incorporating key cybersecurity concepts and evolving through expert collaboration. It emphasizes creating targeted programs to enhance parental awareness, designing inclusive curricula with interactive tools like comics and games, and engaging the community through regular outreach. Recommendations include ongoing professional development for teachers, regular curriculum assessments, and partnerships with cybersecurity organizations to keep educational content current. Additionally, establishing clear online safety policies is advocated to ensure a secure learning environment. These measures aim to prepare students, teachers, and parents to confidently navigate digital challenges, underscoring the need for continuous research to keep educational practices up-to-date with the digital landscape.

VI. Conclusion

The systematic literature review on cybersecurity awareness in schools reveals critical issues and effective practices necessary for improving cybersecurity education. Key challenges include the lack of a consistent and comprehensive curriculum, inadequate parental supervision and knowledge, and the diverse needs of students, especially those with additional support needs (ASN). Effective practices identified include using interactive educational tools, tailored interventions, gamification, and narrative-based learning. Additionally, community involvement and continuous improvement of cybersecurity programs are essential. Recommendations include developing standardized curricula, enhancing parental education, tailoring programs for diverse needs, using interactive tools, implementing innovative teaching methods, fostering community involvement, providing ongoing professional development for educators, conducting regular assessments, establishing partnerships with cybersecurity organizations, and promoting safe online practices. Implementing these recommendations can significantly enhance cybersecurity awareness among students, educators, and parents, creating a safer online environment.

VII. Study Limitations

Despite the comprehensive approach taken in this systematic literature review, several limitations must be acknowledged. The scope was confined to studies published in the past five years and limited to peer-reviewed articles in English, potentially excluding valuable insights from earlier research, non-English publications, and grey literature. The selected databases, although extensive, may not encompass all relevant studies, especially those in lesser-known journals or emerging sources, leading to an incomplete representation of the current state of cybersecurity awareness in schools. The PRISMA framework relies heavily on the quality and reporting standards of included studies, so any deficiencies directly impact the reliability and generalizability of the findings. Additionally, the exclusion of articles due to lack of full-text access or lower quality thresholds might have omitted pertinent studies. Potential bias introduced by the subjective nature of the inclusion and exclusion criteria, despite consistent application efforts, could influence the final selection of articles. Furthermore, the review did not deeply analyze the diverse educational contexts and varying levels of technological infrastructure across different regions and countries, which means the applicability of certain findings and recommendations may be limited or require adaptation to fit specific local contexts. Addressing these limitations in future research will be essential to developing a more comprehensive and nuanced understanding of cybersecurity awareness in schools, enhancing the effectiveness of educational strategies and interventions across diverse settings.

- Funding statement:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. All resources utilized were provided by the authors' own efforts.

- Registration Information:

The review was not registered in any systematic review protocol registry.

- Review Protocol Access:

No protocol was prepared for this systematic review.

- Amendments to Protocol or Registration:

As no protocol was prepared or registered, there were no amendments related to registration or protocol.

- Competing Interests

There are no competing interests to declare. No financial, personal, or professional conflicts influenced the conduct, results, or reporting of this systematic review.

- Availability of Data, Code, and Other Materials:

This systematic review exclusively involves the inclusion and analysis of previously published studies. No new data,

analytic code, or other materials were generated for this review. Therefore, there are no additional materials available for public access. All relevant data have been extracted from the included studies and are presented within the manuscript.

REFERENCES

[1] M. Bada and J. R. C. Nurse, "Chapter 4 - The social and psychological impact of cyberattacks," V. Benson and J. B. T.-E. C. T. and C. V. Mcalaney, Eds. Academic Press, 2020, pp. 73–92. doi: https://doi.org/10.1016/B978-0-12-816203-3.00004-6.

[2] B. Pranggono and A. Arabo, "COVID-19 pande.. cybersecurity issues," Internet Technol. Lett., vol. 4, no. 2, p. e247, Mar. 2021, doi: https://doi.org/10.1002/itl2.247.

[3] M. Muthuppalaniappan and K. Stevenson, "Healthcare cyber-attacks and the COVID-19 pandemic: An urgent threat to global health," Int. J. Qual. Heal. Care, vol. 33, no. 1, pp. 1–12, 2021, doi: 10.1093/intqhc/mzaa117.

[4] S. A. Jawaid, "Cyber Security Threats to Educational Institutes: A Growing Concern for the New Era of Cybersecurity," Int. J. Data Sci. Big Data Anal., vol. 2, no. 2, 2022, doi: 10.51483/ijdsbda.2.2.2022.11-17.

[5] S. M. Zurkarnain and A. A. Abdulsahib, "Investigating The Correlation between The Five Major Personality Traits and a Student's Online Habits," Asia-Pacific J. Inf. Technol. Multimed., vol. 12, no. 01, pp. 57–69, 2023, doi: 10.17576/apjitm-2023-1201-04.

[6] A. Arifin, U. Mokhtar, Z. Hood, S. Tiun, and D. Jambari, "Parental Awareness on Cyber Threats Using Social Media," J. Komun. Malaysian J. Commun., vol. 35, pp. 485–498, Jun. 2019, doi: 10.17576/JKMJC-2019-3502-29.

[7] M. J. A. Rahman, M. I. Hamzah, M. H. M. Yasin, M. M. Tahar, Z. Haron, and N. K. E. Ensimau, "The UKM Students Perception towards Cyber Security," Creat. Educ., vol. 10, no. 12, pp. 2850–2858, 2019, doi: 10.4236/ce.2019.1012211.

[8] B. M. Dioubate, W. D. W. Norhayate, Z. F. Anwar, S. Fauzilah, H. M. Faiz, and L. O. Hai, "The Role of Cybersecurity on the Performance of Malaysian Higher Education Institutions," J. Pengur., vol. 67, pp. 31–41, 2023, doi: 10.17576/pengurusan-2022-67-03.

[9] S. S. I. Rahim, M. I. M. Huda, S. Sa'ad, and R. Moorthy, "Cyber Security Crisis/Threat: Analysis of Malaysia National Security Council (NSC) Involvement Through the Perceptions of Government, Private and People Based on the 3P Model," vol. 1, no. 2, pp. 4–6, 2024.

[10] Y. Yuliana, "THE IMPORTANCE OF CYBERSECURITY AWARENESS FOR CHILDREN," Lampung J. Int. Law, vol. 4, pp. 41–48, Jun. 2022, doi: 10.25041/lajil.v4i1.2526.

[11] D. Ondrušková and R. Pospíšil, "The good practices for implementation of cyber security education for school children," Contemp. Educ. Technol., vol. 15, no. 3, 2023.

[12] S. Nehra and M. Deepanshi, "Cybersecurity Awareness and Education Programs : A Review of Effectiveness," Int. J. Creat. Res. Thoughts, vol. 11, no. 7, pp. 504–508, 2023.

[13] J. Prümmer, T. van Steen, and B. van den Berg, "A systematic review of current cybersecurity training methods," Comput. Secur., vol. 136, no. July 2023, p. 103585, 2024, doi: 10.1016/j.cose.2023.103585.

[14] W. Buyu and B. Ogange, "Cybersecurity in Online Learning: Innovations for Teacher Training and Empowerment," p. 6, 2021.

[15] H. H. M. A. Al-Fatlawi, "Awareness of cyber security aspects in distance education," J. Pedagog. Sociol. Psychol., vol. 6, no. 1, 2024, doi: 10.33902/jpsp.202424403.

[16] W. Triplett, "Addressing Cybersecurity Challenges in Education," Int. J. STEM Educ. Sustain., vol. 3, pp. 47–67, Jan. 2023, doi: 10.53889/ijses.v3i1.132.

[17] D. A. Sareen and S. Jasaiwal, "Need of cyber security education in modern times," Int. J. Multidiscip. Trends, vol. 3, no. 2, pp. 188–191, 2021, doi: 10.22271/multi.2021.v3.i2c.179.

[18] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering – A systematic literature review," Inf. Softw. Technol., vol. 51, no. 1, pp. 7–15, 2009, doi: https://doi.org/10.1016/j.infsof.2008.09.009.

[19] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.," BMJ, vol. 372, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[20] Li T, Higgins JPT, and Deeks JJ (editors)., "Chapter 5: Collecting data | Cochrane Training," in Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021)., 2021.

[21] R. Hamdi, "CYBERSECURITY AWARENESS IN SAUDI ARABIA: A SYSTEMATIC LITERATURE REVIEW," 14th International Conference on Education and New Learning Technologies. pp. 4805–4815, 2022.

[22] K. Renaud and J. Ophoff, "A cyber situational awareness model to predict the implementation of cyber security controls and precautions by SMEs," Organ. Cybersecurity J. Pract. Process People, vol. 1, no. 1, pp. 24–46, 2021.

[23] M. Alshaikh, H. Naseer, A. Ahmad, and S. B. Maynard, "Toward sustainable behaviour change: an approach for cyber security education training and awareness," 2019.

[24] A. Aliyu et al., "A holistic cybersecurity maturity assessment framework for higher education institutions in the United Kingdom," Appl. Sci., vol. 10, no. 10, p. 3660, 2020.

[25] F. Quayyum, D. S. Cruzes, and L. Jaccheri, "Cybersecurity awareness for children: A systematic literature review," Int. J. Child-Computer Interact., vol. 30, p. 100343, 2021, doi: 10.1016/j.ijcci.2021.100343.

[26] F. Giannakas, A. Papasalouros, G. Kambourakis, and S. Gritzalis, "A comprehensive cybersecurity learning platform for elementary education," Inf. Secur. J., vol. 28, no. 3, pp. 81–106, 2019, doi: 10.1080/19393555.2019.1657527.

[27] H. Qusa, "Cyber-Hero : A Gamification framework for Cyber Security Awareness for High Schools Students," pp. 677–682, 2021.

[28] P. Mıhcı Türker and E. Kılıç Çakmak, "An Investigation of Cyber Wellness Awareness: Turkey Secondary School Students, Teachers, and Parents," Comput. Sch., vol. 36, no. 4, pp. 293–318, 2019, doi: 10.1080/07380569.2019.1677433.

[29] A. Ibrahim, M. McKee, L. F. Sikos, and N. F. Johnson, "A Systematic Review of K-12 Cybersecurity Education Around the World," IEEE Access, vol. 12, pp. 59726–59738, 2024, doi: 10.1109/ACCESS.2024.3393425.

[30] S. Aldaajeh, H. Saleous, S. Alrabaee, E. Barka, F. Breitinger, and K. R. Choo, "The Role of National Cybersecurity Strategies on the Improvement of Cybersecurity Education," 2022.

[31] M. Ayyash, T. Alsboui, O. Alshaikh, I. Inuwa-Dute, S. Khan, and S. Parkinson, "Cybersecurity Education and Awareness among Parents and Teachers: A Survey of Bahrain," IEEE Access, vol. 12, no. May, pp. 86596–86617, 2024, doi: 10.1109/ACCESS.2024.3416045.

[32] S. Bannon, T. McGlynn, K. McKenzie, and E. Quayle, "The internet and young people with Additional Support Needs (ASN): Risk and safety," Comput. Human Behav., vol. 53, pp. 495–503, 2014, doi: 10.1016/j.chb.2014.12.057.

[33] A. Moyo, T. Tsokota, C. Ruvinga, and C. T. Chipfumbu, "An E _ safety Framework for Secondary Schools in Zimbabwe," Technol. Knowl. Learn., no. 0123456789, 2021, doi: 10.1007/s10758-021-09545-y.

[34] R. Belen and N. L. Virginia, "Kent Academic Repository A Systematic Literature Review on Cyber Security Education for Children," vol. 66, 2023, doi: 10.1109/TE.2022.3231019.

[35] A. N. A. Kovačević, N. Putnik, and O. Tošković, "Factors Related to Cyber Security Behavior," vol. 8, 2020, doi: 10.1109/ACCESS.2020.3007867.

[36] A. Moyo, T. Tsokota, C. Ruvinga, and C. T. Chipfumbu, "An E - safety Framework for Secondary Schools in Zimbabwe," Technol. Knowl. Learn., no. 0123456789, 2021, doi: 10.1007/s10758-021-09545-y.

[37] F. Quayyum, J. Bueie, D. S. Cruzes, L. Jaccheri, and J. Carlos, "Understanding parents ' perceptions of children ' s cybersecurity awareness in Norway," vol. 1, no. 1, pp. 1–6, 2021.

[38] B. J. Blažič and A. J. Blažič, "Cybersecurity Skills among European High-School Students: A New Approach in the Design of Sustainable Educational Development in Cybersecurity," Sustain., vol. 14, no. 8, 2022, doi: 10.3390/su14084763.

[39] G. Jin, M. Tu, T.-H. Kim, J. Heffron, and J. White, "Evaluation of Game-Based Learning in Cybersecurity Education for High School Students," J.

Educ. Learn., vol. 12, no. 1, pp. 150–158, 2018, doi: 10.11591/edulearn.v12i1.7736.

[40] M. Lehto, Development Needs in Cybersecurity Education : Final report of the project, no. 96. 2022.

[41] A. E. Al-Naser, A. Bushager, and H. Al-Junaid, "Parents' awareness and readiness for smart devices' cybersecurity," IET Conf. Publ., vol. 2019, no. CP758, pp. 0–6, 2019, doi: 10.1049/cp.2019.0226.

[42] C. O'Brien, "TEACHERS' PERCEPTIONS ABOUT USE OF DIGITAL GAMES AND ONLINE RESOURCES FOR CYBERSECURITY BASICS EDUCATION: A CASE STUDY," no. January, pp. 1–19, 2019.

[43] F. B. Fatokun, S. Hamid, A. Norman, and J. O. Fatokun, "The Impact of Age, Gender, and Educational level on the Cybersecurity Behaviors of Tertiary Institution Students: An Empirical investigation on Malaysian Universities," J. Phys. Conf. Ser., vol. 1339, no. 1, 2019, doi: 10.1088/1742-6596/1339/1/012098.

[44] F. E. Catota, M. Granger Morgan, and D. C. Sicker, "Cybersecurity education in a developing nation: The Ecuadorian environment," J. Cybersecurity, vol. 5, no. 1, pp. 1–19, 2019, doi: 10.1093/cybsec/tyz001.

[45] D. J. Power, C. Heavin, and Y. O'Connor, "Balancing privacy rights and surveillance analytics: a decision process guide," J. Bus. Anal., vol. 4, no. 2, pp. 155–170, Jul. 2021, doi: 10.1080/2573234X.2021.1920856.

[46] S. Nasir, "Exploring the Effectiveness of Cybersecurity Training Programs : Factors , Best Exploring the Effectiveness of Cybersecurity Training Programs : Factors , Best Practices , and Future Directions," no. August, 2023, doi: 10.22624/AIMS/CSEAN-SMART2023P18.

Integrating Multi-Agent System and Case-Based Reasoning for Flood Early Warning and Response System

Nor Aimuni Md Rashid[1]*, Zaheera Zainal Abidin[2], Zuraida Abal Abas[3]
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM)[1, 2, 3]
College of Computing, Informatics and Mathematics, Universiti Teknologi Mara, Malacca Branch, Malacca, Malaysia[1]

***Abstract*—This research addresses the limitations of current Multi-Agent Systems (MAS) in Flood Early Warning and Response Systems (FEWRS), focusing on gaps in risk knowledge, monitoring, forecasting, warning dissemination, and response capabilities. These shortcomings reduce the system's reliability and public trust, highlighting the need for better flood preparedness and learning mechanisms. To tackle these issues, this study proposes a new conceptual framework combining Case-Based Reasoning (CBR) with MAS, aimed at enhancing flood prediction, learning, and decision-making. CBR enables the system to learn from past flood events by retrieving and adapting cases to improve future predictions and responses, while MAS allows for decentralized and collaborative decision-making among various agents within the system. This integration fosters a dynamic, real-time system that adapts to changing conditions and improves over time through continuous feedback. The framework's effectiveness is evaluated using the quadruple helix model, addressing social, economic, environmental, and governance aspects. Socially, the system increases community resilience through improved early warnings. Economically, it reduces flood impacts by enabling faster and more accurate responses. Environmentally, it enhances monitoring and preservation of ecosystems. In governance, the framework improves coordination between agencies and the public. The CBR-MAS framework significantly improves intelligent detection, decision-making speed, and community resilience, offering substantial improvements over traditional FEWRS. This adaptive approach promises to build a more reliable, trust-worthy system capable of handling the complexities of flood risks in the future.**

***Keywords*—Flood; multi-agent system; flood early warning system; case-based reasoning; quadruple helix; flood risk**

I. Introduction

Floods are a prevalent natural phenomenon that can significantly impact human settlements and the surrounding ecosystem, often resulting in substantial socio-economic consequences. These impacts include property destruction, infrastructure impairment, and the interruption of vital services [1]. As flood risks grow more severe due to climate change and other factors, there has been a notable shift towards flood risk mitigation strategies, especially when traditional flood defense methods are perceived as ineffective or impractical [2]. Therefore, understanding the vulnerability of communities to flood impacts and developing comprehensive strategies for prevention, mitigation, and management is crucial [3].

Flood disaster management requires a multi-faceted approach that spans across different stages of a flood event. This is typically divided into three main categories: pre-disaster, during the disaster, and post-disaster, encompassing four phases: (i) prevention/mitigation, (ii) preparedness, (iii) response, and (iv) recovery [4], [5], [6] as shown in Fig. 1. A key component of flood preparedness is the Flood Early Warning System (FEWRS), which provides prompt and reliable data on potential flood occurrences. A well-designed FEWRS allows authorities to proactively monitor and detect potential hazards, enabling early intervention and preparation to mitigate the flood's impact [5].

According to the United Nations Office for Disaster Risk Reduction (UNDRR), a comprehensive early warning system integrates hazard monitoring, forecasting, disaster risk assessment, communication, and preparedness activities [7]. This enables communities, governments, businesses, and other entities to take timely action before a hazardous event occurs. As described by the World Meteorological Organization (WMO) in 2011 [8], flood forecasting and warning systems serve as a bridge between accurate rainfall forecasting, hydrometric data collection, real-time flood forecasting models, and issuing early warnings.

The Sendai Framework for Disaster Risk Reduction (SFDRR) [9] also identifies FEWRS as a high-priority tool for flood risk management, essential for mitigating the increasing flood risks posed by climate change in both industrialized and developing countries [10]. A comprehensive FEWRS as shown in Fig. 2, includes four key components: (i) risk knowledge, (ii) monitoring and forecasting, (iii) warning dissemination and communication, and (iv) response capabilities [11]. Each component plays a critical role in ensuring the effectiveness of early warning systems. For instance, risk knowledge encompasses understanding exposure, hazard, and vulnerability, while monitoring and forecasting address the uncertainties of hydrodynamic and meteorological factors. Any deficiency in these components can jeopardize the entire system's functionality [10], [12].

In recent years, Multi-Agent Systems (MAS) have emerged as a valuable tool in flood management. MAS offers a dynamic approach to modeling complex and distributed domains, improving decision-making, flood forecasting, risk assessment, and response capabilities [13]. MAS has been successfully applied in various areas, including reservoir flood control

optimization [14], traffic simulation during floods [15], and assessing flood losses and household responses [16]. By integrating physical and social aspects of flood risk, agent-based models and MAS provide a promising approach to address the complexities of flood management [17].

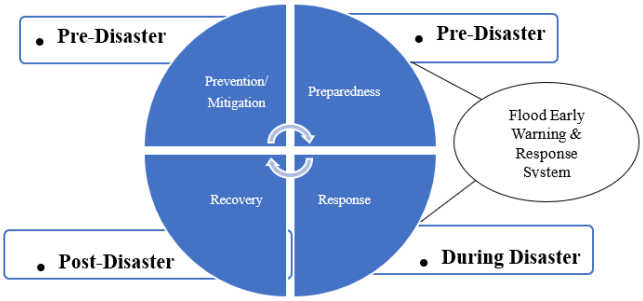


Fig. 1. Flood disaster management phases.

Fig. 2. FEWRS main components.

Current flood warning systems confront issues such as limited flexibility, a lack of collaborative behavior, and insufficient utilization of real-time input, all of which impede efficient crisis decision-making. The research seeks to answer key questions: How can CBR and MAS enhance the adaptability and accuracy of flood warning systems? What role does stakeholder collaboration, through the quadruple helix model in social, economic, environmental, and governance perspectives play in improving disaster management? How can the proposed framework address real-world challenges in flood response? This study tackles these difficulties by investigating how to combine Case-Based Reasoning (CBR) with Multi-Agent Systems (MAS) to develop a more flexible and collaborative framework for flood control. The goal is to improve forecast accuracy and decision-making by incorporating important stakeholders in the quadruple helix model in the area of social, economic, environmental, and governance perspectives. This research explores the application of MAS in addressing flood-related hydrological issues and systematically classifies MAS approaches in hydrologic modeling and prediction. It aims to demonstrate how these sophisticated techniques can enhance flood early warning systems and decision-making processes.

The paper is structured as follows: Section II focuses on materials and methods which discuss on the flood-related MAS modelling and reviewing various methods. Section III introduces a conceptual framework for MAS-FEWRS, while Section IV analyzes and evaluates the research findings. Finally, Section V concludes by summarizing the main findings and highlighting the significance of MAS-FEWRS in improving flood disaster management practices.

II. Material and Methods

Given the gradual increase in complexity of the contemporary world, it is imperative to acknowledge that flood prediction processes are also becoming increasingly intricate in tandem with the changing global climate. Consequently, it is imperative to develop, examine, and construct models that exhibit higher levels of complexity to effectively capture the interactions between the system and its growing complexity [13]. The escalation of complexity on a global scale may suggest that traditional models may not be sufficient in accurately depicting these intricate transformations. Hence, the utilization of MAS can effectively address complex problems that may prove challenging or unfeasible for a single agent or a monolithic system to resolve. Therefore, intelligence can encompass systematic, functional, procedural, or algorithmic methods for searching, discovering, and processing information [18].

A. Data Collection

Our research methodology involved a literature review to identify relevant articles for this study on developing a flood early warning system using multi-agent approaches. Initially, we collected a total of 76 articles from Scopus, Web of Science (WOS), and Institute of Electrical and Electronics Engineers (IEEE) databases. To refine our selection, we first removed any duplicate articles, resulting in 61 unique articles. Next, we scanned the titles and abstracts of these articles to assess their relevance to this research topic. By excluding 20 articles that did not align with our research objectives or did not address flood early warning systems or agent-based approaches, we were left with 41 articles.

Moving forward, we obtained the full texts of the 41 selected articles and performed a thorough reading and analysis. During this process, we carefully evaluated each article based on predefined inclusion criteria. After a comprehensive assessment, we excluded 26 articles that did not meet these criteria, leaving us with a final set of 15 articles that are directly pertinent to our research topic. The 15 relevant articles will serve as the foundation for the methods section of our research paper. They will contribute valuable insights into existing methodologies, techniques, and findings on flood early warning systems and agent-based/multi-agent approaches. By leveraging the knowledge gained from these articles, we will be able to propose and develop our own flood early warning system using multi-agent techniques, building upon the existing literature in the field. Fig. 3 depicts the review process.

B. Review Analysis

Based on the 15 articles, a comparative analysis was conducted. The relevant research papers were reviewed, and a comparison was made based on three criteria. The first criterion

focuses on the components of the FEWRS. This includes risk knowledge, monitoring and forecasting, warning dissemination, and response capabilities. Consequently, the second criterion examines the basic features of MAS, such as leadership, decision function, heterogeneity, agreement parameter, delay consideration, data transmission frequency, mobility, reasoning, perception, communication, and negotiation methods. The third criterion is based on key aspects of designing MAS models, including coordination control, Multi-Agent Learning System (MAL), fault detection, task allocation, localization, organization, and security.

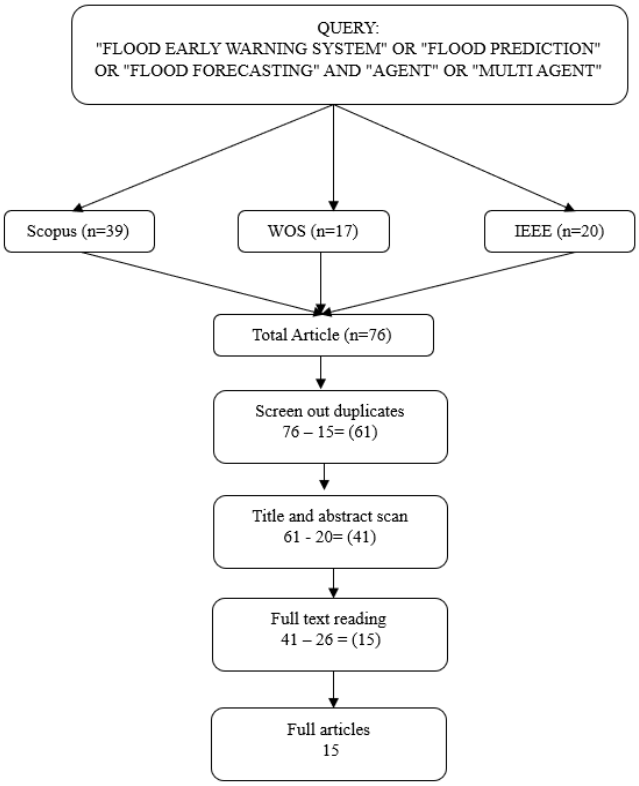


Fig. 3. Review analysis.

Through this comparative analysis, the strengths and limitations of existing approaches in flood early warning systems using agent-based techniques are evaluated, and potential areas for improvement in the design of such systems are identified. Table I depicts the mapping of existing MAS-based flood-related modeling with FEWRS components. In contrast, Tables II compare the existing MAS-based flood modeling with the basic features and key aspects of designing a MAS-based model in complex systems.

Fig. 4 illustrates the observed trends in FEWRS components implementation over the years. Between 2003 and 2009, there was a consistent presence of two key themes: "Risk Knowledge" and "Monitoring & Forecasting." This indicates an initial focus on comprehending risks and monitoring procedures. Nevertheless, it appears that the activities related to "Warning Dissemination" and "Response Capabilities" were relatively inactive during this timeframe, suggesting a potential emphasis on acquiring information rather than the prompt implementation of measures. The year 2011 witnessed a notable transition characterized by an increase in the practice of "Monitoring & Forecasting" and the emergence of "Warning Dissemination," indicating a proactive stance towards mitigating potential risks. By 2014, the domain of "Monitoring & Forecasting" had established a firm position, whereas the domain of "Risk Knowledge" experienced a decline in its level of prominence. However, there has been a noticeable shift towards prioritizing the dissemination of warnings and occasional displays of response capabilities, indicating an increasing emphasis on prompt and efficient responses.

The years 2015 and 2021 demonstrated significant advancements, as evidenced by the consistent prevalence of "Warning Dissemination" and "Risk Knowledge" and the emerging recognition of the importance of "Response Capabilities." By 2022, the domains of "Monitoring & Forecasting," "Warning Dissemination," and "Response Capabilities" had attained a state of strong establishment, thereby highlighting the adoption of a comprehensive and proactive strategy for the development of the system. The capabilities of the MAL System have demonstrated significant advancement over time. Initially, the system relied on the collaborative ANYTIME Multi-Agent System (AMAS) theory. However, it has since progressed to more sophisticated methodologies such as Deep Q-Network (DQN) and Twin Delayed-Deep Deterministic Policy Gradient. In general, the observed patterns suggest a gradual and flexible evolution within the system, demonstrating a continuous dedication to enhancing its capacities and ability to withstand potential obstacles.

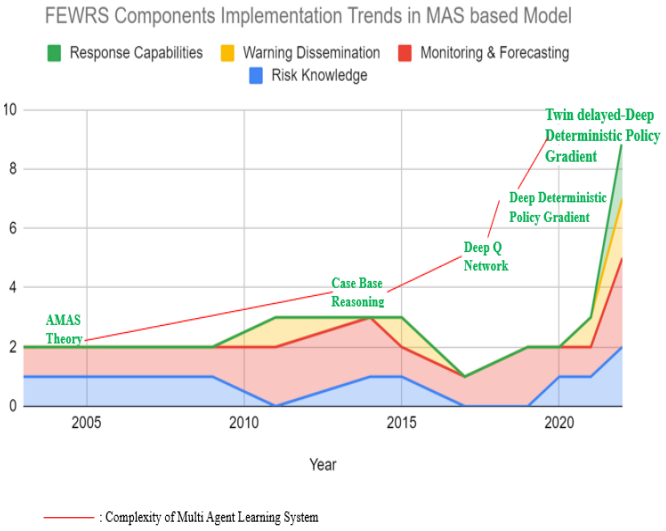


Fig. 4. FEWRS observed trends of MAL implementation.

TABLE I. EXISTING MODEL ANALYSIS

Research/ Existing model	Year	FEWRS Components				Input Parameter	Output Parameter
		Risk Knowledge	Monitoring & Forecasting	Warning Dissemination	Response Capabilities		
[20]	2003	/	/	X	X	Rainfall Level, River Level	Hourly Water Level
[21]	2009	/	/	X	X	Rainfall Level, River Level	Hourly Water Level
[22]	2011	X	/	X	X	River flow, River level, Precipitation	River flow, Warning code
[23]	2011	X	/	/	X	Rainfall, Water level	River water level, Flood alert (mild, critical, dangerous)
[24]	2014	X	/	X	X	Rainfall, Flow velocity, Water level	Sensor data classification to valid and invalid
[18]	2014	/	/	X	X	Rainfall	Water Level
[25]	2015	/	/	/	X	Rainfall, Runoff, Water Level	Estimate time for flood
[26]	2017	X	/	X	X	Rainfall, Flow velocity, Water level	Sensor data classification to valid and invalid
[27]	2019	X	/	X	X	Rainfall	Flood prone area
[28]	2019	X	/	X	X	Processed optical image, local state of the swarm	Flood prone area
[29]	2020	/	/	X	X	Rainfall	Water level
[30]	2021	/	/	/	X	Rainfall, Runoff, Water Level	Estimate time for flood
[31]	2022	X	/	/	/	Satellite photos, Meteorological data	Flood status (yes/no)
[32]	2022	/	/	X	X	Rainfall, flow rate	Discharge rate of the dam
[33]	2022	/	/	/	X	Rainfall, Water level, streamflow	Flood status (yes/no)

TABLE II. EXSITING MODEL ANALYSIS BASED ON MAS DESIGN ASPECT

Research/ Existing Model	MAS Design Key Aspects Consideration						
	Coordination Control	Multi-Agent Learning (MAL) System	Fault Detection	Task Allocation	Localization	Organization	Security
[20]	/	AMAS Theory (collaborative)	X	Decentralized	Not dynamic	Team	X
[21]	/	AMAS Theory (collaborative)	X	Decentralized	Not dynamic	Team	X
[22]	/	X	X	Decentralized	Not dynamic	Hierarchical	X
[23]	/	X	X	Decentralized	Not dynamic	flat	X
[24]	/	X	/	Decentralized	Dynamic	Team	X
[18]	/	Use Case-Based Reasoning	X	Decentralized	Dynamic	Team	X
[25]	/	X	X	Decentralized	Not dynamic	Team	X
[26]	/	X	/	Decentralized	Dynamic	Team	X
[27]	/	X	X	Decentralized	Dynamic	Swarm	X
[28]	/	Deep Q-Network	X	Decentralized	Dynamic	Swarm	X
[29]	/	Deep Deterministic Policy Gradient	/	Decentralized	Dynamic	Hierarchical	X
[30]	/	X	X	Decentralized	Not dynamic	Team	X
[31]	/	X	X	Not defined	Not defined	Hierarchical	X
[32]	/	Twin delayed-Deep Deterministic Policy Gradient	X	Decentralized	Not defined	Hierarchical	X
[33]	/	X	X	Decentralized	Not dynamic	Team	X

III. RESULT

The conceptual framework presented in this research paper (Fig. 5) is derived from a combination of previous concepts related to FEWRS, MAS design principles, and a thorough review of existing models. By building upon these foundations, our framework aims to improve the effectiveness and efficiency of FEWRS in mitigating flood hazards. The framework's development starts with examining FEWRS, serving as the basis for understanding the core components and requirements of flood early warning systems. Hence, by analyzing the strengths and limitations of existing FEWRS models, our framework incorporates advancements and novel approaches to address critical challenges in flood management.

Drawing on the principles of MAS, our framework introduces a multi-agent architecture comprising different specialized agents. These agents, including the Monitoring Agent, Forecasting Agent, Warning Dissemination Agent, Response Agent, and Learning Agent, work collaboratively to enhance the overall performance of the flood early warning system. Additionally, this framework incorporates active learning and inference techniques within a Reasoning/Inference Agent. This agent leverages data and information the system collects to make informed decisions and predictions regarding flood events. Thus, this active learning approach enhances the system's adaptability and predictive capabilities.

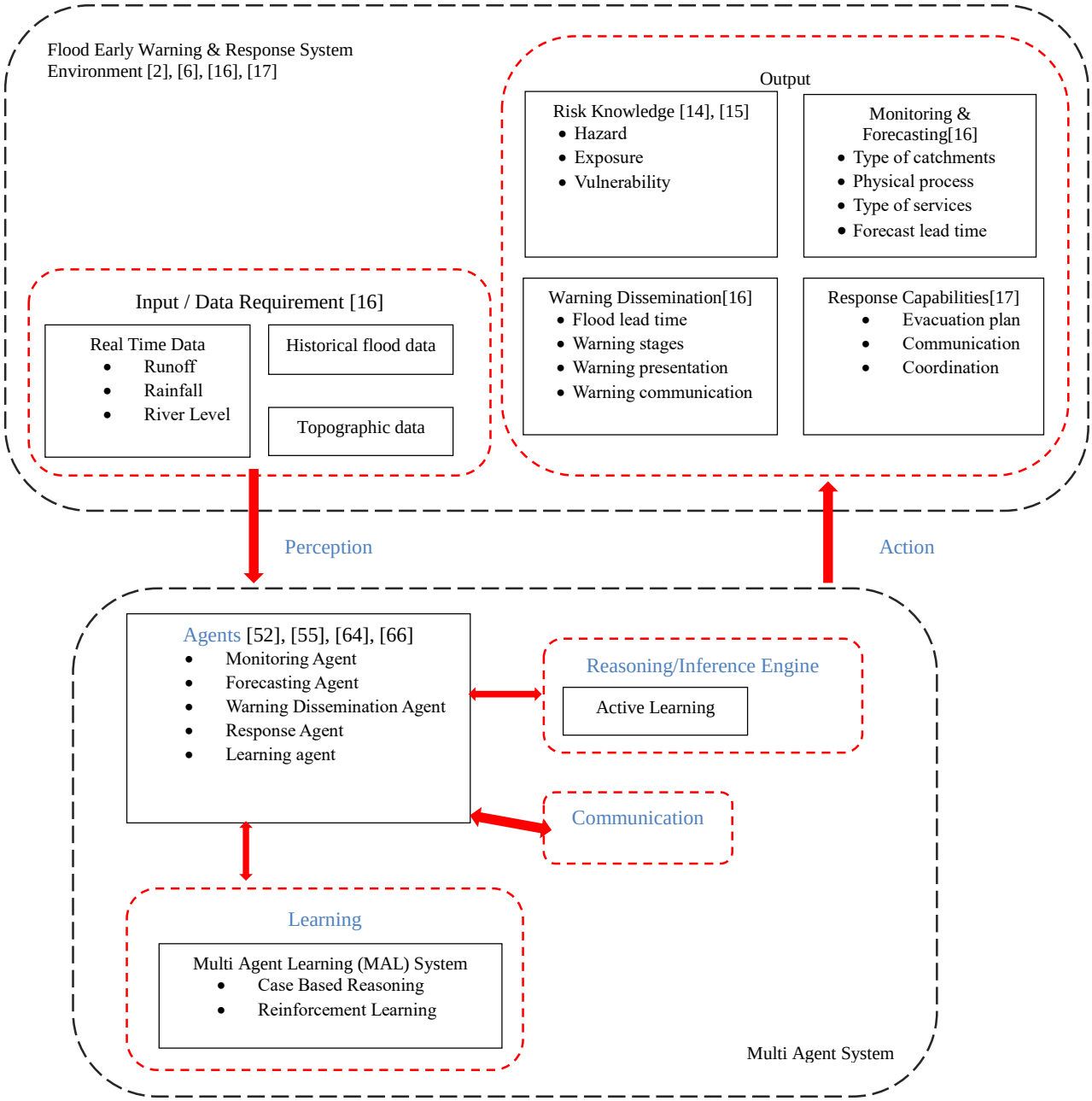


Fig. 5. Proposed MAS-FEWRS conceptual framework.

Other than that, communication is another integral component of the framework. It enables seamless information exchange among the agents, ensuring coordination and synchronization in decision-making. Effective communication mechanisms are designed to facilitate real-time data sharing, forecast dissemination, warning communication, and response coordination. Furthermore, the Perception component acts as a bridge between the MAS and the FEWRS environment. It encompasses collecting various data types, such as real-time data (river level, runoff, rainfall), historical data, and topographic data. These inputs are fed into the system for analysis, modeling, and decision-making processes.

The action component represents the output of the MAS, directed toward the FEWRS environment. It encompasses the four key components of FEWRS: Risk Knowledge, Monitoring and Forecasting, Warning Dissemination, and Response Capabilities. The framework provides a comprehensive approach to flood management by integrating these components. Moreover, this framework introduces a MAL System: Case-Based Reasoning (CBR) technique. This learning system enhances the overall adaptability of the agents by leveraging past experiences.

In summary, our conceptual framework is derived from previous concepts related to FEWRS, MAS design principles, and a review of existing models. By combining these elements, our framework introduces a comprehensive approach to flood early warning systems, emphasizing collaboration among agents, active learning, effective communication, and a past experience learning system. It emphasizes the importance of data collection, analysis, communication, and coordinated response for effective flood management.

IV. Discussion

Our discussion is structured into four main components, each addressing a distinct aspect of our research; (i) Understanding FEWRS and Identifying Key Challenges, (ii) Comparative Analysis of Existing Flood Early Warning System Models, (iii) Conceptual Framework: Building upon the insights gained from the comparative analysis and (iv) Impact of the Conceptual Framework using Quadruple Helix Model.

A. Understanding Flood Early Warning Systems (FEWRS) and Identifying Key Challenges

FEWRS is crucial in minimizing the damage and casualties caused by floods [34, 35, 36]. However, the efficiency of FEWRS in flood disasters is limited by various factors. Factors such as system quality, information quality, user satisfaction, service quality, use, perceived usefulness, intention to use, net benefits, perceived ease of use, compatibility, user experience, relative advantage, complexity, perceived risks, educational quality, and confirmation have been identified as significant factors affecting the effectiveness of FEWRS. Moreover, accurate intelligence is essential for issuing early warnings and responding effectively to floods. A structured review of the FEWRS literature identified twenty-seven types of key intelligence required in the flood cycle. This intelligence can be captured using technological solutions at various stages of a flood event to support decision-making for early warnings and response.

Implementing effective FEWRS is crucial in reducing losses and casualties caused by floods. The SFDRR emphasizes the need for multi-hazard warning systems and disaster risk information, including FEWRS, to be available to the community by 2030. The increased losses from floods can be attributed to population growth and rapid urbanization. To improve the effectiveness of FEWRS, it is important to enhance the employment performance of government agencies through technological innovation. FEWRS should have effective usability features and strategic information access and display to provide accurate and timely information to stakeholders. However, existing FEWRS often fail to effectively provide information on flood disasters, highlighting the need for improvement.

B. Comparative Analysis of Existing Flood Early Warning System Models

In this comparative discussion, we analyze and compare the information presented in three different areas: FEWRS components, MAS features, and key aspects for designing MAS models. These areas provide insights into developing and implementing flood early warning systems and MAS. By examining the details within each area, we can better understand the advancements, variations, and considerations in these domains.

The first area of analysis is FEWRS components. This highlights the research and existing models, the year of development, and the specific components involved in these systems. The components include risk knowledge, monitoring and forecasting, warning dissemination, and response capabilities. Furthermore, the input parameters vary from rainfall and river levels to river flow, precipitation, flow velocity, and sensor data. The output parameters also differ, ranging from hourly water levels to flood alerts and estimated times for flooding. This highlights the importance of collecting comprehensive data and providing timely warnings for effective flood management.

The second area of analysis is MAS features, which play a crucial role in the design and functioning of MAS models. The review provides insights into various features, including leadership, decision function, heterogeneity, agreement parameter, delay consideration, data transmission frequency, mobility, reasoning, perception, communication, and negotiation method. Moreover, these features reflect the diversity and flexibility of MAS models in adapting to different contexts and objectives. The variations observed in MAS features compatibility demonstrate the range of strategies employed to facilitate coordination, decision-making, information exchange, and negotiation among agents within the system.

The third area of analysis focuses on the key aspects of designing MAS models. This review explores coordination control, MAL systems, fault detection, task allocation, localization, organization, and security. Coordination control mechanisms can be centralized or decentralized, depending on the distribution of decision-making authority. Hence, the choice of coordination control mechanism directly influences the dynamics and efficiency of multi-agent collaboration. MAL systems encompass various theories and algorithms, enabling

agents to learn and improve their performance through environmental interactions. Fault detection mechanisms ensure system robustness, while task allocation strategies optimize overall performance. Localization techniques enhance agents' awareness, organization mechanisms define system structure, and security measures protect system integrity.

Comparing these three areas reveals the interconnectedness and interdependence of flood early warning systems and MAS models. The MAS features and key aspects of designing MAS models directly contribute to the effectiveness and efficiency of flood early warning systems. Moreover, MAS models provide a framework for integrating and coordinating the diverse components of flood early warning systems, enabling real-time monitoring, accurate forecasting, efficient warning dissemination, and prompt response capabilities.

Furthermore, the variations observed in MAS features and key aspects highlight the need for adaptive and context-specific approaches. Depending on their unique environmental, social, and organizational factors, different flood-prone regions may require different coordination control mechanisms, learning algorithms, fault detection strategies, and localization techniques.

By considering these insights, researchers and practitioners can enhance the design and implementation of flood early warning systems by incorporating MAS principles and utilizing suitable MAS features. Therefore, this holistic approach can lead to improved risk management, effective coordination, and timely response in mitigating the impact of floods and other natural disasters.

C. *Conceptual Framework: Building upon the Insights Gained from the Comparative Analysis*

Developing an effective conceptual framework for flood early warning systems requires a deep understanding of the underlying principles, existing models, and innovative approaches. In this regard, our research is guided by four key hypotheses supported by compelling evidence and critical analysis. These hypotheses serve as the foundation for our conceptual framework, enabling us to address the limitations and challenges identified in existing models and enhance the effectiveness of flood management practices.

Hypothesis 1 suggests that the integration of a MAS architecture within the FEWRS environment enhances the overall effectiveness and efficiency of flood management. The evidence supporting this hypothesis lies in the specialized agents involved, such as the Monitoring Agent, Forecasting Agent, Warning Dissemination Agent, Response Agent, and Learning Agent. By incorporating these agents, the system benefits from improved coordination and task distribution, leading to more informed decision-making and response capabilities. MAS design principles, which emphasize collaboration, adaptability, and distributed intelligence, are well-suited to address the complexity and uncertainty associated with flood events.

Hypothesis 2 proposes that active learning and inference techniques within the Reasoning/Inference Agent enhance the accuracy and reliability of flood predictions and decision-making. The evidence supporting this hypothesis lies in the capability of active learning techniques to continuously update the system's knowledge and models based on incoming data. This continuous learning process leads to improved predictive capabilities. Additionally, integrating inference techniques allows for extracting meaningful insights from various data sources, facilitating more informed real-time decision-making.

Hypothesis 3 asserts that effective communication mechanisms among the agents within the MAS contribute to the timely and accurate dissemination of flood warnings and response coordination. The evidence supporting this hypothesis lies in the seamless communication facilitated by the MAS architecture. Hence, real-time data sharing, forecast dissemination, and coordinated response actions are made possible, ensuring stakeholders receive timely and relevant information for informed decision-making. This robust communication mechanism enables efficient resource allocation, evacuation procedures, and overall flood management.

Hypothesis 4 suggests that the use of the MAL System, which is the CBR technique, enhances the adaptability and learning capabilities of the agents in response to changing flood conditions. CBR is a problem-solving approach involving solving new problems by retrieving and adapting solutions from similar cases. In the context of flood prediction, CBR can be utilized to improve the accuracy and reliability of flood forecasts by leveraging historical flood events and their associated data. Several experiments demonstrate the performance of CBR in various domains. For example, in the field of childhood disease diagnosis, a study compared rule-based reasoning and CBR and discovered that CBR had the best accuracy, achieving 92% accuracy [19].

By considering and validating these four hypotheses, our conceptual framework aims to address the limitations and challenges identified in existing models, enhance the effectiveness of flood early warning systems, and contribute to more efficient flood management practices. Through rigorous analysis, testing, and the integration of compelling evidence, we seek to provide practical insights and innovative solutions for real-world flood scenarios.

D. *Impact of the Conceptual Framework using Quadruple Helix Model*

The conceptual framework we have developed, which integrates a MAS architecture, active learning, effective communication, and a hybrid learning system, holds great potential for significantly impacting flood management practices. To analyze the impact of our framework, we will utilize the Quadruple Helix model, emphasizing the collaboration and interaction among academia, industry, government, and society. Fig. 6 illustrates the basic Quadruple Helix model that we are referring to.

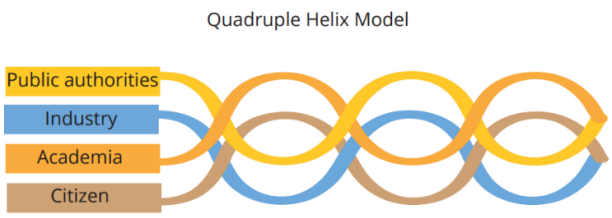


Fig. 6. Quadruple helix model.

1) Academia: In the context of academia, our conceptual framework offers an opportunity for further research and academic advancement. By exploring the integration of MAS architecture, active learning, and hybrid learning systems, we contribute to the theoretical understanding of flood management and the development of innovative approaches. As a result, the findings and insights from our research can enrich the academic literature and serve as a foundation for future studies in the field of flood early warning systems.

2) Industry: Implementing our conceptual framework has practical implications, particularly in developing and improving flood early warning systems. Integrating specialized agents within the MAS architecture enhances coordination and task distribution, improving decision-making and response capabilities. Furthermore, incorporating active learning and inference techniques improves the accuracy and reliability of flood predictions. Effective communication mechanisms ensure the timely dissemination of flood warnings and facilitate response coordination. Moreover, combining CBR and RL, the hybrid learning system enhances adaptability and learning capabilities. These advancements can potentially revolutionize the industry's approach to flood management, resulting in more effective and efficient systems.

3) Government: Our conceptual framework offers valuable insights for government agencies responsible for managing and mitigating flood risks. By adopting the MAS architecture, government entities can enhance their coordination and response capabilities in flood events. Moreover, integrating active learning and inference techniques improves the accuracy of flood predictions, enabling more informed decision-making in real-time. In addition, effective communication mechanisms ensure timely dissemination of warnings and support coordinated response actions. The learning system enables government agencies to adapt their strategies and responses to changing flood conditions, leading to improved flood management practices. Implementing our framework can enhance the government's ability to safeguard lives and property, reduce flood impacts, and ensure the overall resilience of communities.

4) Society: The ultimate impact of our conceptual framework is on society as a whole. By incorporating advanced technologies and methodologies, we aim to enhance the effectiveness of flood early warning systems, ultimately reducing the negative impacts of floods on society. The timely dissemination of accurate flood warnings can help individuals and communities make informed decisions, evacuate if necessary, and prepare for potential flood events. Note that effective communication mechanisms facilitate the coordination of response actions, ensuring that resources are allocated efficiently and evacuation procedures are well-coordinated. Hence, adopting our framework can improve public safety, reduce property damage, and increase resilience in the face of flood events.

By considering the impact of our conceptual framework through the Quadruple Helix model, we recognize the collaborative efforts and interactions among academia, industry, government, and society. The framework's implementation has the potential to drive positive change, transform flood management practices, and create a significant impact on the well-being and safety of individuals and communities affected by floods.

V. CONCLUSION

This research paper has explored the integration of MAS and MAL techniques to enhance FEWRS. By conducting a comprehensive review and comparative analysis of existing MAS models related to flood early warning systems, we have gained valuable insights into the system components, basic features of MAS, and key aspects of designing MAS models.

Our findings highlight the potential of MAS in addressing the complexity and dynamism associated with flood early warning systems. Other than that, the comparative analysis revealed the strengths and limitations of different MAS approaches, providing a basis for developing an improved conceptual framework. This framework combines the FEWRS components with MAL techniques, particularly CBR.

The proposed conceptual framework offers several advantages, including enhanced adaptability, scalability, and the ability to learn from past experiences. By incorporating CBR-RL, the framework enables the system to reason and make decisions based on historical cases, further improving the accuracy and timeliness of FEWRS.

The theoretical contribution of this study lies in its integration of MAS and MAL within the FEWRS framework, providing a novel approach that bridges technological advancements and disaster management. By incorporating the quadruple helix model, the framework fosters collaboration between academia, industry, government, and society, ensuring a comprehensive and stakeholder-driven approach to flood management. This integration not only advances theoretical understanding but also lays the groundwork for practical, scalable solutions.

However, this research has limitations that warrant future exploration. The framework's evaluation relied on historical data and simulated feedback, which may not fully capture real-world complexities. Future research should prioritize real-world deployment and validation in diverse geographical and climatic conditions. Additionally, exploring its applicability to other disaster types can provide insights into the framework's scalability and versatility.

In conclusion, this research paper contributes to the field of disaster management by presenting a comparative analysis of MAS models related to flood early warning systems and proposing a conceptual framework that integrates MAS, FEWRS components, and MAL techniques. The proposed framework can potentially advance the accuracy, efficiency, and adaptability of flood early warning systems, ultimately improving disaster preparedness and response. Correspondingly, further research and implementation efforts are encouraged to validate and refine the proposed framework, leading to real-world applications and positive outcomes in flood-prone regions.

ACKNOWLEDGMENT

Thank you to Kementerian Pendidikan Tinggi Malaysia (KPT) for awarding scholarships SLAB-SLAI program, College of Computing, Informatics and Mathematics, UiTM Melaka, and Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka (UTeM).

REFERENCES

[1] C. Luu, Q. D. Bui, and J. von Meding, "Mapping direct flood impacts from a 2020 extreme flood event in Central Vietnam using spatial analysis techniques," Int J Disaster Resil Built Environ, vol. 14, no. 1, pp. 85–99, Jan. 2023, doi: 10.1108/IJDRBE-07-2021-0070.

[2] D. L. T. Hegger et al., "Toward more flood resilience: Is a diversification of flood risk management strategies the way forward?," Ecology and Society, vol. 21, no. 4, p. art52, 2016, doi: 10.5751/ES-08854-210452.

[3] D. Lowe, K. Ebi, and B. Forsberg, "Factors Increasing Vulnerability to Health Effects before, during and after Floods," Int J Environ Res Public Health, vol. 10, no. 12, pp. 7015–7067, Dec. 2013, doi: 10.3390/ijerph10127015.

[4] S. A. H. B. S. Muzamil et al., "Proposed Framework for the Flood Disaster Management Cycle in Malaysia," Sustainability (Switzerland), vol. 14, no. 7, Apr. 2022, doi: 10.3390/su14074088.

[5] W. A. Hammood, R. A. Arshah, S. M. Asmara, H. Al Halbusi, O. A. Hammood, and S. Al Abri, "A systematic review on flood early warning and response system (FEWRS): A deep review and analysis," Jan. 01, 2021, MDPI AG. doi: 10.3390/su13010440.

[6] I. M. Yusoff, A. Ramli, M. Mohamad, and N. M. Nasir, "Exploring the Managing of Flood Disaster: A Malaysian Perspective," 2018. doi: 10.17576/geo-2018-1403-03.

[7] D. Perera, J. Agnihotri, O. Seidou, and R. Djalante, "Identifying Societal Challenges in Flood Early Warning Systems," 2020. doi: 10.1016/j.ijdrr.2020.101794.

[8] World Meteorological Organization, Manual on Flood Forecasting and Warning, 2011th ed. Geneva 2: World Meteorological Organization, 2011.

[9] U. Nations Office for Disaster Risk Reduction, "Sendai Framework for Disaster Risk Reduction 2015 - 2030," 2015. [Online]. Available: https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030\

[10] S. Samansiri, T. Fernando, and B. Ingirige, "Critical Failure Factors of Flood Early Warning and Response Systems (FEWRS): A Structured Literature Review and Interpretive Structural Modelling (ISM) Analysis," Geosciences (Switzerland), vol. 13, no. 5, May 2023, doi: 10.3390/geosciences13050137.

[11] United Nations Development Programme (UNDP), "Five approaches to build functional early warning systems," 2018.

[12] W. A. Hammood, S. M. @Asmara, R. A. Arshah, O. A. Hammood, H. Al Halbusi, and M. A. Al-Sharafi, "Factors influencing the success of information systems in flood early warning and response systems context," Telkomnika (Telecommunication Computing Electronics and Control), vol. 18, no. 6, pp. 2956–2961, Dec. 2020, doi: 10.12928/TELKOMNIKA.v18i6.14666.

[13] J. Simmonds, J. A. Gómez, and A. Ledezma, "The role of agent-based modeling and multi-agent systems in flood-based hydrological problems: A brief review," Journal of Water and Climate Change, vol. 11, no. 4, pp. 1580–1602, 2020, doi: 10.2166/wcc.2019.108.

[14] A. Almeida and D. López-de-Ipiña, "A Distributed Reasoning Engine Ecosystem for Semantic Context-Management in Smart Environments," Sensors, vol. 12, no. 8, pp. 10208–10227, Jul. 2012, doi: 10.3390/s120810208.

[15] Y. Wang, X. Chen, L. Wang, and G. Min, "Effective IoT-Facilitated Storm Surge Flood Modeling Based on Deep Reinforcement Learning," IEEE Internet Things J, vol. 7, no. 7, pp. 6338–6347, Jul. 2020, doi: 10.1109/JIOT.2020.2969959.

[16] L. E. Yang, J. Scheffran, D. Süsser, R. Dawson, and Y. D. Chen, "Assessment of Flood Losses with Household Responses: Agent-Based Simulation in an Urban Catchment Area," Environmental Modeling & Assessment, vol. 23, no. 4, pp. 369–388, Aug. 2018, doi: 10.1007/s10666-018-9597-3.

[17] W. Kellens, T. Terpstra, and P. De Maeyer, "Perception and Communication of Flood Risks: A Systematic Review of Empirical Research," Risk Analysis, vol. 33, no. 1, pp. 24–49, Jan. 2013, doi: 10.1111/j.1539-6924.2012.01844.x.

[18] B. Linghu and F. Chen, "An intelligent multi-agent approach for flood disaster forecasting utilizing case based reasoning," in Proceedings - 2014 5th International Conference on Intelligent Systems Design and Engineering Applications, ISDEA 2014, Institute of Electrical and Electronics Engineers Inc., Dec. 2014, pp. 182–185. doi: 10.1109/ISDEA.2014.48.

[19] I. Werdiningsih, R. Hendradi, P. Purbandini, B. Nuqoba, and E. Anna, "The Efficient Distance Weighted Case Base Rule (DW-CBR) for Early Childhood Diseases Diagnosis," 2021. doi: 10.47839/ijc.20.2.2174.

[20] J.-P. Georgé and M.-P. Gleizes, "Real-time simulation for flood forecast: an adaptive multi-agent system staff ÂGIR-ÂGe, Innovation sociale et Reflexivité View project." [Online]. Available: http://www.irit.fr/SMAC

[21] J.-P. Georgé, S. Peyruqueou, C. Régis, and P. Glize, "Experiencing Self-adaptive MAS for Real-Time Decision Support Systems," 2009, pp. 302–309. doi: 10.1007/978-3-642-00487-2_32.

[22] A. M. Matei, "Multi-Agent System for Monitoring and Analysis Prahova Hydrographical Basin," 2011.

[23] S. S. Weerawardhana and G. B. Jayatilleke, "Web service based model for inter-agent communication in multi-agent systems: A case study," in 2011 11th International Conference on Hybrid Intelligent Systems (HIS), IEEE, Dec. 2011, pp. 698–703. doi: 10.1109/HIS.2011.6122191.

[24] E. M. Marouane, E. Mostafa, and E. Mohamed, "Intelligent data classification and aggregation in wireless sensors for flood forecasting system," in Proceedings of 2014 Mediterranean Microwave Symposium (MMS2014), IEEE, Dec. 2014, pp. 1–8. doi: 10.1109/MMS.2014.7088991.

[25] M. El Mabrouk, M. Ezziyyani, Z. A. Sadouq, and M. Essaaidi, "New Expert System for Short, Medium and Long-Term Flood Forecasting and Warning," J Theor Appl Inf Technol, vol. 20, no. 2, 2015, [Online]. Available: www.jatit.org

[26] M. El Mabrouk and S. Gaou, "Proposed Intelligent Pre-Processing Model of Real-Time Flood Forecasting and Warning for Data Classification and Aggregation," International Journal of Online Engineering (iJOE), vol. 13, no. 11, p. 4, Nov. 2017, doi: 10.3991/ijoe.v13i11.7382.

[27] W. Satria Aji, "Simulasi dengan Multi Agent Prediksi Banjir Berdasarkan Intensitas Hujan Menggunakan Particle Swarm Optimization," 2019. [Online]. Available: http://ejournal.urindo.ac.id/index.php/TI

[28] D. Baldazo, J. Parras, and S. Zazo, "Decentralized Multi-Agent Deep Reinforcement Learning in Swarms of Drones for Flood Monitoring," in 2019 27th European Signal Processing Conference (EUSIPCO), IEEE, Sep. 2019, pp. 1–5. doi: 10.23919/EUSIPCO.2019.8903067.

[29] S. M. Saliba, B. D. Bowes, S. Adams, P. A. Beling, and J. L. Goodall, "Deep reinforcement learning with uncertain data for real-time stormwater system control and flood mitigation," Water (Switzerland), vol. 12, no. 11, pp. 1–19, Nov. 2020, doi: 10.3390/w12113222.

[30] E. M. Marouane, "Towards a Real Time Distributed Flood Early Warning System," 2021. [Online]. Available: www.ijacsa.thesai.org

[31] A. Rafanelli, S. Costantini, and G. De Gasperis, "A Multi-Agent-System framework for flooding events," 2022. [Online]. Available: https://www.epa.gov/climate-indicators/weather-climate

[32] R. ARAKAWA and P. CHUN, "Multi-reservoir flood control using deep reinforcement learning," Artificial Intelligence and Data Science, vol. 3, no. 3, pp. 46–53, 2022, doi: 10.11532/jsceiii.3.3_46.

[33] J. A. Simmonds Sheppard, "Multi-agent system for flood forecasting in Tropical River Basin," Universidad Carlos III de Madrid, 2022.

[34] D. Perera, O. Seidou, J. Agnihotri, A. Wahid, and M. Rasmy, "Flood Early Warning Systems: A Review Of Benefits, Challenges And Prospects Environmentally Sustainable Water Storage Management and Operation in Sri Lanka View project WASCAL (West African Service Science Center on Climate Change and Adapted Land Used) View project", doi: 10.13140/RG.2.2.28339.78880.

[35] IPCC, : Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change. NY: Cambridge University Press, 2012.

[36] H. Kreibich et al., "Adaptation to flood risk: Results of international paired flood event studies," Earths Future, vol. 5, no. 10, pp. 953–965, Oct. 2017, doi: 10.1002/2017EF000606.

Multi-Source Consistency Deep Learning for Semi-Supervised Operating Condition Recognition in Sucker-Rod Pumping Wells

Jianguo Yang[1], Bin Zhou[2]*, Muhammad Tahir[3]*, Min Zhang[4], Xiao Zheng[5], Xinqian Liu[6]
School of Mechanical Engineering, Shandong University of Technology, Zibo, China[1]
School of Computer Science and Technology, Shandong University of Technology, Zibo, China[2, 4, 5, 6]
Department of Computer Science-Faculty of Engineering Science and Technology (FEST), Iqra University, Main Campus, Defense View, Karachi 75500, Sindh, Pakistan[3]

Abstract*—How making full use of the multiple measured information sources obtained from the sucker-rod pumping wells based on deep learning is crucial for precisely recognizing the operating conditions. However, the existing deep learning-based operating condition recognition technology has the disadvantages of low accuracy and weak practicality owing to the limitations of methods for handling single-source or multi-source data, high demand for sufficient labeled data, and inability to make use of massive unknown operating condition data resources. To solve these problems, here we design a semi-supervised operating condition recognition method based on multi-source consistency deep learning. Specifically, on the basis of the framework of WideResNet28-2 convolutional neural network (CNN), the multi-head self-attention mechanism and feedforward neural network are first used to extract the deeper features of the measured dynamometer cards and the measured electrical power cards, respectively. Then, the consistency constraint loss based on cosine similarity measurement is introduced to ensure the maximum similarity of the final features expressed by different information sources. Next, the optimal global feature representation of multi-source fusion is obtained by learning the weights of the feature representations of different information sources through the adaptive attention mechanism. Finally, the fused multi-source feature combined with the multi-source semi-supervised class-aware contrastive learning is exploited to yield the operating condition recognition model. We test the proposed model with a dataset produced from an oilfield in China with a high-pressure and low permeability thin oil reservoir block. Experiments show that the method proposed can better learn the critical features of multiple measured information sources of oil wells, and further improve the operating condition identification performance by making full use of unknown operating condition data with a small amount of labeled data.

Keywords*—Operating condition recognition of sucker-rod pumping wells; multi-source consistency learning; semi-supervised learning; CNN; attention mechanism

I. INTRODUCTION

In the process of oil production, the sucker-rod pumping well production system can obtain a large amount of operating condition data from multiple information sources, such as dynamometer cards, electrical parameters, etc. Meanwhile, massive unknown operating condition data can also be obtained. It is crucial to timely and accurately recognize the operating conditions by effectively using and processing these different high-dimensional heterogeneous information sources. Recently, deep learning has become a research hotspot in oil production engineering and information technology due to its powerful ability to handle complex data structures and extract deep hidden features, thus improving the accuracy of model classification [1-2]. However, deep learning usually relies mainly on a great amount of labeled data for model training in order to maintain high accuracy. But the acquisition of such a large number of labeled data is impractical, and this will also cause a serious waste of massive unlabeled data resources. Therefore, exploring the integration of multiple information sources, massive unlabeled operating condition data, and deep learning to optimize the operating condition recognition with few labeled data is much of important for further improving the accuracy and extending the application of operating condition identification.

Existing deep learning-based sucker-rod pumping well operating condition recognition technology is mainly achieved by dynamometer cards or electrical parameters alone [3-4], or the effective combination of the two kinds of information as well as the oil well production information with artificial intelligence methods [5-6]. These studies have achieved good results, but also hold some limitations: (1) Relying on a single data source, which may easily cause false positives in electromechanical hydraulic coupling complex systems. (2) Poor fusion strategy and low robustness when integrating multiple information sources, which can degrade the performance and lower the robustness due to the uncertainty of production statistical data [7]. (3) Massive labeled samples requirement [8-9] and few unlabeled samples considered [10-11] during model training, which can greatly weaken the practical value of operating condition recognition technologies. (4) The feature extraction strategy is mainly realized by combining mechanism analysis [12-14], which will cause precision errors in the calculation of model feature parameter values and affect the recognition effect. Besides, the problems of "zero division" (appears during the electrical parameter timing signal recognition) and damping coefficient (appears during the pump dynamometer card recognition) will also increase the accuracy error.

Multi-source information fusion [15] is an information integration technology that can achieve more comprehensive and accurate feature representation by integrating features from

different information sources. By combining the unique advantages of each data source, it overcomes the limitation of a single data source, reduces the prediction error, and therefore significantly improves the robustness. Compared with traditional multi-feature connection strategy [16], multi-source information fusion can smartly integrate multi-source features and effectively address the challenge caused by simple connection fusion, such as higher feature dimension, information redundancy and inability to capture complex relationships among features. In the fusion process, following the consistency principle [17], multi-source information fusion can keep the logical consistency of the fused features and reduce information redundancy and conflict. Further, taking the cosine similarity measure [18] as a consistency measure, a more accurate measurement instead of the traditional Euclidean distance, can further improve the reliability and effectiveness of fusion results. In order to realize efficient multi-source information fusion, a powerful feature extractor is also needed to capture the key features of each information source. With wide network structure, WideResNet28-2 (WRN28-2) [19] can capture more abundant and diversified feature information, thus providing basic support for subsequent fusion. Referring to the encoder design in Transformer architecture [20], the introduction of multi-head self-attention mechanism and feedforward neural network can facilitate the deep interaction between multi-source features, significantly enhance the representation capability of multi-source features, and provide more abundant feature information for the fusion process. Studies have shown [21] that the introduction of an attention mechanism can further improve the effect of feature fusion. The attention mechanism can automatically learn and dynamically adjust the weights of features of various data sources and focus its attention on more critical information in the classification task, thus effectively improving the classification performance and overall effect.

Semi-supervised deep learning [22] trains the model by using a few labeled data and lots of unlabeled data, with the aim of enhancing the generalization ability of the model. It has four main learning strategies: pseudo-label learning [23], consistency regularization [24], deep learning-based generation [25], and graph-based strategy [26]. In pseudo-label learning, the model uses its own predictions of unlabeled samples to expand the training data. Compared with other methods, it has high flexibility and strong generalization, making it applicable for a variety of practical application scenarios, therefore is widely used in semi-supervised deep learning. Contrastive learning [27] is an effective unsupervised learning method. It learns distinguishing feature representations by comparing sample pairs, therefore the learned features achieve improved generalization on unlabeled data. Studies have shown that the combination of class-aware contrastive learning, pseudo-label learning and consistency regularization learning [28] can effectively solve the common confirmation bias problem in semi-supervised deep learning and improve accuracy. By minimizing the distance of samples of the same category in feature space and maximizing the distance between samples of different categories, this combination strategy can not only enhance differentiation between categories, but also significantly improve the accuracy of semi-supervised learning for tasks such as image recognition.

As described above, to solve the limitations of existing deep learning-based operating condition recognition research, this paper proposes a semi-supervised sucker-rod pumping well operating condition recognition method based on multi-source consistency deep learning. Specifically, in order to avoid damping coefficient, "division by zero" and statistical data uncertainty, we select the measured dynamometer cards and the measured electrical power cards (both of which are high-dimensional binary images) as multi-source data. We adopt the framework of WRN28-2 convolutional neural network, and achieve the maximum consistency interactive learning based on features complements of different information sources through the combination of multi-head self-attention mechanism, feedforward neural network and cosine similarity measurement. Then, we apply the adaptive attention mechanism to achieve the optimal global feature representation of multi-source fusion by learning the weights of feature representations of different information sources, thus improving the existing operating condition feature extraction technology by capitalizing on all the multi-source information. The combination of multi-source fusion features and semi-supervised class-aware contrastive learning can further utilize the massive unlabeled data and a few multi-source labeled data to improve the operating condition recognition accuracy and extend its applications.

The contributions of this paper are as follows:

1) An efficient multi-source feature fusion deep learning algorithm for the operating condition is proposed. Based on the framework of the WRN28-2 convolutional neural network, the algorithm uses a multi-head self-attention mechanism and feedforward neural network to extract the deep image features of each measured information source respectively, and introduces consistency constraint loss based on cosine similarity measurement to ensure the maximum similarity of the final feature representations from different information sources. Then, the optimal global feature representation of multi-source fusion is obtained by using an adaptive attention mechanism to learn the weights of feature representations of different information sources.

2) A multi-source semi-supervised class-aware contrastive learning operating condition recognition method is proposed, which effectively integrates the multi-source feature fusion deep learning algorithm with the semi-supervised class-aware contrastive algorithm. Its objective function consists of four parts: supervised loss function, multi-source information consistency constraint loss function, unsupervised loss function and class-aware contrastive loss function. The recognition bias of the same category in multi-source feature space is reduced by class-aware contrastive learning, and the discriminant ability of the semi-supervised deep learning model is highly enhanced.

3) The effectiveness and practicability of the proposed method is verified experimentally. Compared with state-of-the-art multi-source semi-supervised learning recognition methods, traditional single-source recognition methods or multi-feature connection recognition methods, different attention mechanism learning recognition methods, and different contrastive learning

recognition methods, the proposed method exhibits improved recognition accuracy in different proportions of labeled training data, especially with 10\% increase. In addition, key hyperparameter analysis and ablation experiments further verify the effectiveness and practicability of the proposed method.

II. Related Works

In this part, we first give a short introduction to the information sources of the sucker-rod pumping well production system in terms of the operating principle of the system, and then we present the related works on operating condition recognition research respectively, using dynamometer cards, electrical parameters and multiple information sources.

A. Information Source Analysis of the Sucker-Rod Pumping Well Production System

The architecture diagram of the sucker-rod pumping well production system [29] is shown in Fig. 1.

It can be observed that the system is composed of 15 parts: (1) Standing valve; (2) Pump barrel; (3) Traveling valve; (4) Plunger; (5) Dynamic fluid level; (6) Sucker rod; (7) Casing; (8) Tubing; (9) Beam hanger; (10) Horse head; (11) Beam; (12) Link rod; (13) Crank; (14) Induction motor; (15) Electronic control cabinet.

The operating principle of the system is as follows: First, the high-speed rotating motion of the induction motor will be converted into the mechanical up-and-down swing of the beam by the crank connecting to the link rod. Then, the sucker rod connected to the beam hanger suspended from the horse head drives the pump plunger in the underground wellbore to move up and down through the mechanical swing of the beam. Finally, the mixed fluid in the wellbore is pumped to the surface through the tubing.

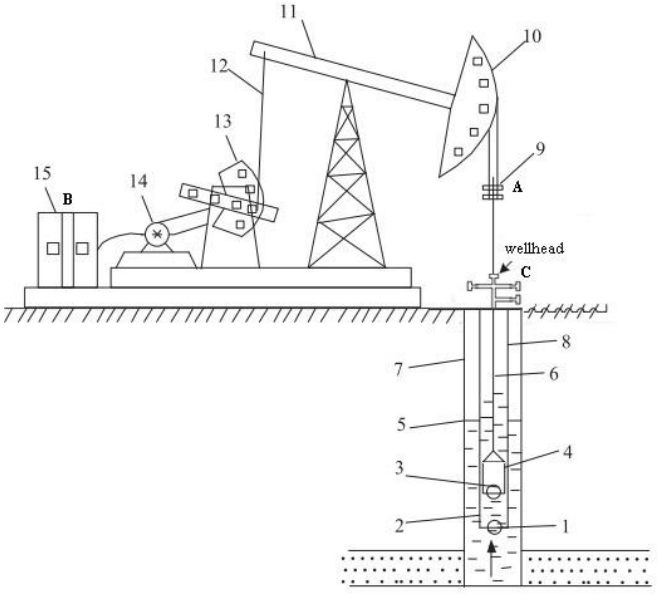


Fig. 1. The architecture diagram of the sucker-rod pumping well production system.

As the key downhole production equipment, the oil well pump mainly consists of three parts: pump barrel, plunger, and valves (see (1) and (3) in Fig. 1). One operation of the oil well pump includes valves opening and closing during one stroke.

In the sucker-rod pumping well production system, the measured information mainly comes from three sources: dynamometer cards (see A in Fig. 1), electronic parameters (see B in Fig. 1), wellhead data (wellhead temperature, wellhead pressure, etc., see C in Fig. 1). The measured dynamometer cards are binary images composed of the displacement and load of the polish rod (see (9) in Fig. 1), which are usually with high dimension and strong noise, and can reflect the situation of the wellbore and the stratum in real-time. The existing operating condition recognition technology using dynamometer cards mainly uses the pump dynamometer card data (transformed from the measured dynamometer card data after denoising, but may occur the damping coefficient problem). The measured electrical parameters are time-series signals, which can reflect the situation of the ground and the stratum in real-time. It is worth mentioning that the "division by zero" problem may occur when using the electronic parameters for operating condition recognition.

In the production process, there are also some non-measured information sources, such as dynamic fluid level (see (5) in Fig. 1), liquid-producing capacity, working time, etc. These statistical production data can also be regarded as information sources. However, the present research on operating condition recognition usually focuses on effectively utilizing the aforementioned two types of information sources.

B. The Operating Condition Recognition Technology Review

In general, the research on the operating condition recognition technology of the sucker-rod pumping well production system is mainly focused on three aspects: dynamometer card-based, electronic parameter-based and multi-source-based.

1) The dynamometer card-based recognition technology: The dynamometer card-based recognition method is the most widely used one, which is mainly implemented by ground or pump dynamometer cards combined with some artificial intelligence technology. For example, Li et al. [30] employed the moment feature and the “four-point method” to extract pump dynamometer card features and used SVM combined with particle swarm optimization to recognize working conditions. Zheng et al. [31] used the moment feature and the astr polygon decomposition centroid localization algorithm to extract ground dynamometer card features and used HMM combined with clonal selection optimization to recognize working conditions. Zhang et al. [32] extracted the feature of dynamometer cards by the fast discrete curvelet transform and utilized a sparse multi-graph regularized extreme-learning machine to recognize working conditions. The above working condition recognition methods involve a large number of complex feature parameter calculation and need a large amount of labeled data for training, thus greatly affecting the working

condition recognition performance. To this end, recently, deep learning is applied to extract the features in order to improve the recognition performance. Ye et al. [3] proposed an improved CNN to automatically obtain the deep feature expression of the dynamometer card data, overcoming the limitations of traditional feature extraction methods and achieving better recognition accuracy. But this method requires lots of labeled working condition data for model training. In the case of few-shot learning, He et al. [10] proposed to compress the raw working condition data through a 4-dimensional time-frequency feature extraction method, and then optimize parameters of the convolutional shrinkage neural network through meta-learning. This method overcomes the limitation of a large number of labeled data requirements for training, but it reveals the neglect of the massive unknown working condition data. In addition, there are some precision errors in the calculation of the characteristic data value based on the mechanism analysis.

2) The electrical parameter-based recognition technology: The electronic parameter-based operating condition recognition technology works with lower cost and higher reliability. Zheng et al. [33] proposed to extract feature data of the measured electric power signal based on mechanism analysis and accomplish the operating condition recognition by HMM. In addition, Chen et al. [34] deduced the electric power card model by considering the angular velocity of the crank and the friction and inertia of the 4-bar linkage and implemented the operating condition recognition by establishing a feature atlas of electric power cards. In order to further improve the working condition recognition effect based on electrical parameter sequence signals, researchers also introduced deep learning technology to extract deeper features of electrical parameter data. Wei et al. [4] realized a deep and broad learning system by motor power data, utilizing CNN to extract features of the electric power signal and employing the broad learning system to recognize the operating condition. However, because of " division by zero", the recognition accuracy based on electronic parameters is far from satisfactory. In addition, the current deep learning-based working condition recognition technology using electrical parameter data also needs a lot of labeled working condition training data.

3) The multiple information source-based recognition technology: Although it is easy to classify operating conditions using single-source data, it is also prone to error. Therefore, researchers started to develop operating condition recognition models based on multiple information sources. Zheng et al. [35] applied seven feature data such as sucker-rod specifications, pump diameter, stroke, speed, moisture content, gas-liquid ratio and working condition to a HMM model. Zhang et al. [36] adopted feature data such as dynamometer cards, power, dynamic fluid level, and liquid-producing capacity. Liu et al. [37] selected feature data such as dynamometer cards, work day time, and liquid-producing capacity. However, all the above operating condition recognition methods use non-measured production statistics data, which will affect the robustness of the recognition model. Zhou et al. [38] put forward a semi-supervised operating condition recognition technology integrating four measured data, namely, ground dynamometer cards, electric power signal, wellhead temperature, and wellhead pressure. The technical model has good performance and wide application, but the recognition accuracy of multi-classification conditions needs to be improved. Deep learning technology can improve model performance by deeply mining multi-source fusion feature expressions. Li et al. [5] developed an improved deep learning-based operating condition recognition method by the fusion of Fourier descriptor features and graphic features based on dynamometer cards. Abdurakipov et al. [7] used pump dynamometer cards, pump inlet pressure and temperature production data to establish an operating condition identification model with transformer. In summary, at present, the multi-source operating condition recognition technology based on deep learning mostly relies on lots of labeled data, and the multi-source fusion technology is mostly traditional. Recently, in the multi-source semi-supervised deep learning classification recognition research, Wu et al. [39] proposed a multi-view semi-supervised learning method based on graph convolutional network to improve model performance under label-scarce situations. This method maps data from multiple views into a low-dimensional space and uses Laplacian embedding technology to preserve the structural information between data, thereby achieving the fusion and representation learning of multi-view data. In addition, this method implies that we can try to explore the best consistent connection relationship between multiple views to obtain better recognition effects.

To sum up, the research on the operating condition recognition technology has achieved remarkable results, especially with the introduction of deep learning technology, which overcomes the limitations of traditional manual design features. However, in order to establish an efficient and practical operating condition recognition model based on deep learning, limitations such as the requirement of lots of labeled data, the neglect of massive unknown data, and the improvement of the multi-source fusion strategy remain for further handling.

III. Multi-Source Consistency for Semi-Supervised Deep Learning

This section introduces our proposed multi-source consistency deep learning for semi-supervised operating condition recognition method from sucker-rod pumping wells, emphasizing the integration of a multi-source consistency strategy, which includes WRN28-2 convolutional neural network backbone, multi-head self-attention mechanism, feedforward neural network, and an adaptive attention-based feature fusion strategy. Furthermore, this section describes how our model employs the semi-supervised learning technique to make efficient use of limited labeled data alongside a large number of unlabeled data, thereby enhancing recognition performance.

A. Problem Definition

Before introducing our method, the key symbols and variables used in this paper are clearly described in Table I.

TABLE I. NOTATION DESCRIPTION

Notation	Description
$\mathcal{V}$	the number of information sources
D	multi-source data set
L^v	labeled data set of the v^{th} data source
U^v	unlabeled data set of the v^{th} data source
l	the number of labeled data
m	the number of unlabeled data
A_w	weak data enhancement method
A_s	strong data enhancement method
F	backbone network model
G	global feature after multi-source feature fusion
H	multi-source data feature representation after multi-source information consistency learning
ω^v	the attention weight of the data feature of the v^{th} data source
S^v	cosine similarity of data from the v^{th} data source
$\|\cdot\|_2$	the l_2 norm

Here, we study a semi-supervised operating condition recognition task based on multi-source data. Suppose there are $\mathcal{V}$ different data sources, and each data source contains the C class. Multi-source data set D consists of labeled data set L and unlabeled data set U, i.e. $D = \{L, U\}$. For labeled data set L, it is defined as $L = \{L^v | v = 1,2,\dots,\mathcal{V}\} = \{(x_i^v, y_i) | i = 1,2,\dots,l\}$, where L^v is the labeled data set of the v^{th} data source, x_i^v is the i^{th} labeled image of the v^{th} data source, $y_i \in \{1,2,\dots,C\}$ is the label corresponding to the i^{th} image, and l is the number of labeled data. Unlabeled data set U is defined as $U = \{U^v | v = 1,2,\dots,\mathcal{V}\} = \{u_i^v | i = l+1, l+2,\dots,l+m\}$, where U^v is the unlabeled data set of the v^{th} data source, u_i^v is the i^{th} unlabeled image of the v^{th} data source, m is the number of unlabeled data. In order to increase the data diversity and improve the accuracy of semi-supervised learning, this paper adopts two data enhancement methods: weak enhancement $A_w(\cdot)$ and strong enhancement $A_s(\cdot)$. Weak enhancement usually includes small-angle rotation, translation, scaling, etc. Strong enhancement includes cropping, color transformation, etc. For labeled data set L and unlabeled data set U, the weak enhanced data are represented as $A_w(L)$ and $A_w(U)$, and the strong enhanced data are represented as $A_s(L)$ and $A_s(U)$, respectively. In this paper, WRN28-2 model is used as backbone network $F(\cdot)$ to learn feature representations of multi-source data, which are used for subsequent consistency learning and semi-supervised learning tasks.

B. Model Overview

This paper proposes a semi-supervised sucker-rod pumping well operating condition classification method based on multi-source information consistency deep learning. The model structure is shown in Fig. 2. The algorithm consists of the following two key steps: multi-source information consistency learning and multi-source semi-supervised learning. In multi-source information consistency learning, firstly, the WRN28-2 convolutional neural network is used to extract the feature representation of multi-source data. For the multi-source feature representation of the same sample, the dependence relationship in the feature representation is captured by multi-head self-attention mechanism, and the deeper feature is extracted by non-linear transformation with feedforward neural network. In order to ensure the consistency of multi-source information, this paper introduces a multi-source consistency constraint loss function based on the cosine similarity to strengthen the correlation among different information sources. In addition, the attention mechanism is used to learn the weights of feature representations from different information sources to achieve an effective fusion of multi-source feature representations, so as to obtain the optimal global feature representation of each sample. In multi-source semi-supervised learning, two enhancement strategies, weak enhancement and strong enhancement, are applied to the unlabeled data respectively, and the corresponding predicted values are generated, and the unlabeled data with high confidence are selected to generate pseudo labels. To improve the prediction accuracy, consistency regularization is used to calculate the cross-entropy loss of the predicted value corresponding to the strong enhancement and the pseudo-label corresponding to the weak enhancement. In view of the possibility of pseudo label method being affected by confirmation bias in the training process, this paper introduces the class-aware contrastive learning method [28], aggregates samples of the same category in feature space, and disperses samples of different categories to reduce confirmation bias and improve the discriminant ability.

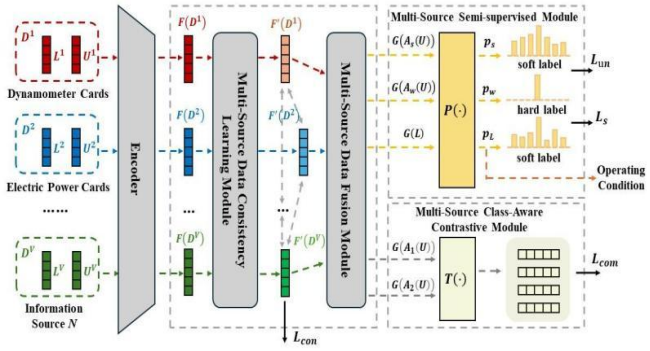


Fig. 2. Model structure diagram.

C. Consistency Learning of Multi-source Information

In this paper, the WRN28-2 backbone network is used to learn the feature representation $F(D)$ of multi-source data set D, where $F(D) = \{F(L), F(U)\}$, and the expressive power of the feature representation is enhanced by multi-head self-attention mechanism. Then, the multi-source consistency constraint loss function based on cosine similarity is used to ensure that the feature representation of the multi-source data is consistent in the feature space, thereby improving the generalization ability. The multi-source information consistency learning model structure is shown in Fig. 3.

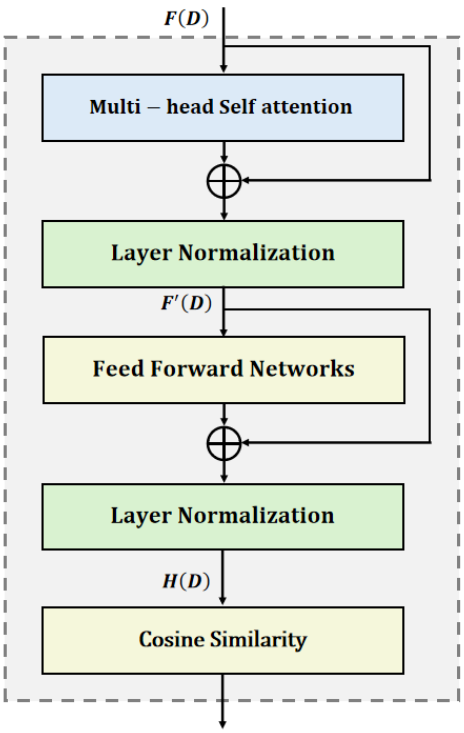


Fig. 3. Structure diagram of multi-source information consistency learning model.

Firstly, the feature representation of the multi-source data is sent into the multi-head self-attention module (MHSAM), and the feature representation ability of the multi-source data is enhanced by capturing the information of the input feature representation in their respective subspaces. Then, we implement residual connections between the input feature $F(D)$ and output feature $F'(D)$ of the MHSAM and carry out Layer Normalization (LN), as shown in Eq. (1):

$$F'(D) = LN\left(F(D) + MHSAM\left(F(D)\right)\right) \quad (1)$$

Then, the deeper features are extracted by nonlinear transformation through the feedforward network module (FFNM). Similarly, the residual connections and layer normalization are used to stabilize the training process, as shown in Eq. (2):

$$H(D) = LN\left(F'(D) + FFNM\left(F'(D)\right)\right) \quad (2)$$

where the FFNM consists of two linear mapping layers and Relu activation function, the former of which is $FFNM(x) = Relu(W_1 * W_2 * x)$, W_1 and W_2 are the weights of the linear mapping layer. $H(D)$ is the feature representation of multi-source data D after consistency learning of multi-source information. The above multi-source information consistency learning method is inspired by the Transformer network structure, and the Transformer coding layer is used to realize the relevant calculation.

There is an intrinsic consistency between different feature representations of multi-source data. To make full use of the consistency and improve the generalization of the model, we design a multi-source consistency constraint loss function based on cosine similarity to maximize the cosine similarity between feature representations of different sources, so as to ensure their consistency in feature space.

Specifically, for the multi-source feature representation $H(D)$ of the same sample, the l_2 norm ($\|\cdot\|_2$) is first converted into the standard feature representation $\|H(D)\|_2$ to eliminate the influence caused by the magnitude difference between the multi-source feature representation vectors. Then, the cosine similarity between the multi-source feature representation vectors is calculated to obtain the degree of consistency between the multi-source feature representations, which is calculated as in Eq. (3):

$$S^v = \|H(D^v)\|_2 \cdot \|H(D^v)\|_2^T \quad (3)$$

where S^v represents the cosine similarity of the v^{th} data source, and $(\cdot)^T$ represents the transpose of the vector.

Then, to ensure the consistency of feature representations from different data sources in feature space and enhance the generalization ability of the model, the difference matrix between any two data source similarity matrices is calculated, and the square of the norm of the difference matrix is calculated as the constraint loss function of multi-source information consistency learning, as shown in Eq. (4):

$$L_{con} = \sum_{i,j \in v} \left\| S^i - S^j \right\|_2^2 \quad (4)$$

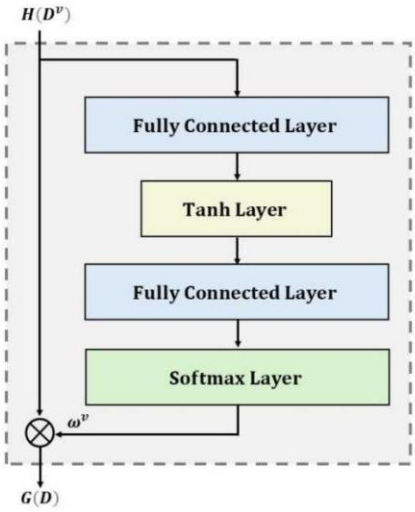


Fig. 4. Structure diagram of multi-source feature fusion model.

To further improve the performance, we further apply a multi-source feature fusion method based on adaptive attention mechanism, the model structure diagram is shown in Fig. 4. The adaptive attention mechanism is used to learn the weights of different data source feature representations, and the multi-source feature representations are then fused into a unified global feature representation to obtain the optimal global characteristics of each sample. Specifically, for the v^{th} data source, the data feature is represented as $H(D^v)$. In this paper, a network structure consisting of fully connected layer, followed by $tanh$ activation function and then by fully connected layer is adopted. The nonlinear combination of $H(D^v)$ is obtained by the network structure learning and the

corresponding attention score W_{out}^{v} is generated, as shown in Eq. (5):

$$W_{out}^{v} = W_2 \cdot [tanh(W_1 \cdot H(D^v) + b_1)] + b_2 \quad (5)$$

where W_1 and W_2 represent the weight of the two fully connected layers respectively, b_1 and b_2 are the corresponding bias vectors.

Then, the SoftMax function is used to normalize the attention score W_{out}^{v} , and the attention weight ω^v of each data source feature representation is obtained. The calculation is shown in Eq. (6):

$$\omega^v = softmax(W_{out}^{v}) = \frac{exp(W_{out}^{v})}{\sum_{n=1}^{N} exp(W_{out}^{v})} \quad (6)$$

where exp is the natural exponential function, and $\sum_{v=1}^{\mathcal{V}} \omega^v = 1$.

Finally, according to the normalized attention weight ω^v, the feature representation of each data source is fused into a unified global feature representation, and the fusion process is achieved by weighted summation, as shown in Eq. (7):

$$G(D) = \sum_{v=1}^{\mathcal{V}} \omega^v \cdot H(D^v) \quad (7)$$

where $G(D)$ is the global feature representation after fusion, and $G(D) = \{G(L), G(U)\}$.

D. Multi-Source Data Semi-supervise Module

After completing the multi-source information consistency learning and the multi-source feature fusion, this paper implements pseudo-label-based semi-supervised learning of multi-source data by referring to Fix-Match [40]. Two enhancement strategies, weak enhancement $G\big(A_w(U)\big)$ and strong enhancement $G\big(A_s(U)\big)$, are applied to unlabeled data, respectively. Then, corresponding predicted values p_w and p_s are generated using a fully connected layer $P(\cdot)$, which are calculated as shown in Eq. (8):

$$p_w = P\left(G\big(A_w(U)\big)\right)$$
$$p_s = P\left(G\big(A_s(U)\big)\right) \quad (8)$$

Then, weakly enhanced unlabeled data with high confidence predicted value are screened out, and the corresponding pseudo-label q_w are obtained using $argmax$ function, as shown in Eq. (9) :

$$q_w = argmax(\hat{p}_w) \quad (9)$$

where $\hat{p}_w$ represents the part of p_w, $\max(\hat{p}_w) > t$, $\max(\cdot)$ is the maximum function and t is the confidence threshold.

Finally, to increase the prediction accuracy, the consistency regularization is adopted to compute the cross-entropy loss between the predicted value of the strong enhancement and the pseudo-label of the weak enhancement. The detailed calculation of the semi-supervised loss L_{un} is shown in Eq. (10):

$$L_{un} = -\sum_{c=1}^{C} q_w^c \log(p_s^c) \quad (10)$$

where c represents the c^{th} class, and C is the total number of sample classes.

The supervised loss function L_s uses the cross-entropy loss of multi-source labeled data L, which is calculated as shown in Eq. (11):

$$L_s = -\sum_{c=1}^{C} y^c \log(p_L) \quad (11)$$

where p_L is the predicted value of labeled samples generated by $P(\cdot)$, $p_L = P\big(G(L)\big)$.

Pseudo-label based semi-supervised learning methods need to generate pseudo-labels independently in the process of semi-supervised learning, and are susceptible to confirmation bias in the training process. To solve this, this paper introduces a multi-source data class-aware contrastive learning method, aggregates samples of the same category in feature space, and drives samples of different categories away from each other. By the contrastive learning, the robustness of the model is increased and the impact of confirmation bias is reduced.

First, the unlabeled data are enhanced by two different data enhancement methods $A_1(\cdot)$ and $A_2(\cdot)$ respectively, obtaining $A(U) = [A_1(U), A_2(U)]$, and then the high-dimensional feature representation is mapped to a low-dimensional embedding vector through the projection layer $T[\cdot]$ composed of two fully connected layers. The specific calculation is shown in Eq. (12):

$$T[G(A(U))] = \{T[G(A_1(U))], T[G(A_2(U))]\} \quad (12)$$

The class-aware contrastive learning loss consists of two parts: narrowing the aggregation distance of enhanced data from the same category samples and widening the aggregation distance of enhanced data from the different category samples. The specific calculation is shown in Eq. (13):

$$L_{com} = \sum_{i \in I} \left[log \sum_{a \in O(i)} \left(exp\left(\frac{T\left[G\left(A^i(U)\right)\right] \cdot T\left[G\left(A^a(U)\right)\right]}{\tau} \right) - exp\left(\frac{T\left[G\left(A_1^i(U)\right)\right] \cdot T\left[G\left(A_2^i(U)\right)\right]}{\tau} \right) \right) \right] \quad (13)$$

where $I = \{1,2,\dots,M\}$, M is the number of unlabeled samples. $O(i)$ is the index of $M-1$ unlabeled samples except the index i. $T\left[G(A_1^i(U))\right]$ and $T\left[G(A_2^i(U))\right]$ are a set of positive sample pairs that are output from the i^{th} unlabeled sample after two data enhancement transformations and then through the projection layer. $T[G(A^i(U))]$ and $T[G(A^a(U))]$ are the corresponding negative sample pairs. τ is the temperature coefficient.

E. Objective Function

The loss function set includes four parts: supervised loss function L_s of labeled data, multi-source information consistency constraint loss function L_{con}, unsupervised loss function L_{un} and class-aware contrastive loss function L_{com}. The total loss function L is calculated using Eq. (14):

$$L = L_s + \lambda_{con} L_{con} + \lambda_{un} L_{un} + \lambda_{com} L_{com} \quad (14)$$

where λ_{con}、λ_{un} and λ_{com} are the weights of multi-source information consistency constraint loss, unsupervised loss and class-aware contrastive loss, respectively.

IV. Experimental Results

The experimental dataset is obtained from an oilfield in China with a high-pressure and low-permeability thin oil reservoir block. The operating condition samples used in the experiment are selected strictly according to the operation records of the oil wells. The measured dynamometer cards (binary images composed of polish rod displacement and load) and measured electric power cards (binary images composed of polish rod displacement and motor power) are included in each operating condition sample. The measured dynamometer card data and electric power card data are composed of the data points collected at the actual corresponding acquisition time on the production site. The sample set is built by the accumulation of operating condition samples of 60 sucker-rod pumping wells for 3 years and includes 11 typical operating conditions, each operating condition has 150 samples, for a total of 1650 samples. The 11 typical operating conditions include normal, assist-blowing, lack of supply liquid, rod cutting, stuck pump, traveling valve failing, wax precipitation, pump leakage, tubing leakage, standing valve leakage, traveling valve leakage.

The experiment is run on Python 3.8.10/pytorch 1.10.0 and Linux 5.4.0/24 GB NVIDIA RTX3090 GPU. The training set and the test set each have 825 samples, of which each operating condition contains 75 samples. The experimental results are averaged 10 times.

To verify the effectiveness and practicability, experiments are carried out from five aspects: (1) The verification of the superiority of the basic model. The learning models based on the classical residual network are compared in the case of full labeled training samples. (2) The verification of the effectiveness of the proposed method. On the basis of the first part of the experiment, different operating condition recognition methods are compared in terms of the operating condition recognition effect with full labeled training samples. (3) The verification of the practicability. On the basis of the second part of the experiment, different operating condition recognition methods are compared in terms of the operating condition recognition effect with different proportions labeled training samples. (4) Hyperparameter analysis. The impact of the four main hyperparameters of the objective function λ_{con}、λ_{un}、λ_{com} and the number of heads of the multi-head attention mechanism are analyzed, respectively. (5) Ablation test.

A. The Basic Model Analysis

To verify the superiority of the designed infrastructure, in the case of full labeled training samples, based on the two main learning models of the depth and width of the classical residual network combined with class-aware contrastive learning (CCL), we compare the results of various operating condition recognition methods based on traditional single-source operating condition recognition methods (WRN28-2/10_CCL, Resnet18/34/50/101/152_CCL,), traditional multi-source connection operating condition recognition methods (MCWRN28-2/1_CCL, MCResnet18/34/50/101/152_CCL), multi-source interactive consistent learning operating condition recognition methods (MIWRN28-2/10_CCL), and multi-source data online augmented interactive consistent learning operating condition recognition methods (MAIWRN28-2_CCL (the proposed method in this paper), MAIWRN28-10_CCL). In the proposed method, the pseudo-label threshold τ is 0.95; the learning rate is 0.02; the number of heads of the multiple self-attention mechanisms in the transformer encoder layer is 2; the middle layer dimension of FFNM is set to 512; λ_{un} and λ_{com} are set to 1, and λ_{con} is set to 10. The comparison results of various operating condition recognition methods based on the classical residual network combined with CCL are shown in Table II.

TABLE II. The Comparison Results of Various Operating Condition Recognition Methods based on the Classical Residual Network Combined with CCL (Average Accuracy, %)

Operating condition recognition methods	Measured dynamometer cards	Measured electrical power cards	Measured dynamometer cards and electrical power cards
Resnet18_CCL	93.67	95.83	--
Resnet34_CCL	92.87	95.25	--
Resnet50_CCL	88.55	94.71	--
Resnet101_CCL	84.09	94.27	--
Resnet152_CCL	89.79	95.27	--
WRN28-2_CCL	93.72	**95.93**	--
WRN28-10_CCL	**94.34**	95.67	--
MCResnet18_CCL	--	--	96.55
MCResnet34_CCL	--	--	96.46
MCResnet50_CCL	--	--	95.39
MCResnet101_CCL	--	--	95.03
MCResnet152_CCL	--	--	94.18
MCWRN28-2_CCL	--	--	97.55
MCWRN28-10_CCL	--	--	97.23
MIWRN28-2_CCL	--	--	97.96
MIWRN28-10_CCL	--	--	97.41
MAIWRN28-2_CCL(Ours)	--	--	**98.79**
MAIWRN28-10_CCL	--	--	98.34

From Table II:

1) Among the operating condition recognition methods based on the classical residual network combined with CCL, recognition methods based on WideResNet have significantly better effects than those based on Resnet, which reflects that WideResNet can obtain better performance by increasing network width and reducing network depth.

2) Among the traditional single-source operating condition recognition methods, in terms of recognition effects, recognition methods based on the measured electrical power cards are obviously superior to those based on the measured dynamometer cards, which shows that the measured electrical power cards adopted in this paper can better overcome the influence of "zero division" on the operating condition recognition accuracy. In addition, the traditional single-source

operating condition recognition methods based on WideResNet are more accurate than those based on ResNet. Meanwhile for the measured dynamometer cards, WRN28-10_CCL has the best recognition effect, and for the measured electrical power cards, WRN28-2_CCL has the best recognition accuracy.

3) Among the traditional multi-source connection operating condition recognition methods, MCWRN28-2/10_CCL, MCResnet18/34_CCL have better recognition accuracy than those based on traditional single-source methods, but MCResnet50/101/152_CCL are the opposite, which indicates the technology limitations of the traditional multi-source connection operating condition recognition methods. In addition, MCWRN28-2/10_CCL have better recognition effects than MCResnet18/34_CCL.

4) Among the operating condition recognition methods based on multi-source consistency learning, MIWRN28-2/10_CCL are more accurate than those recognition methods based on traditional single source and traditional multi-source connection, which reflects the advantage of the multi-source consistency learning method proposed in this paper. Further, this paper carries out the online augmented processing on the multi-source consistency learning to further improve the recognition accuracy. MAIWRN28-2/10_CCL well reflect this strategy for the operating condition recognition, of which MAIWRN28-2_CCL performs best and is selected and used in subsequent experiments in this paper.

B. Multi-Source Learning Operating Condition Recognition with Full Labeled Training Samples

To verify the effectiveness, on the basis of the first part of the experiment, the method proposed in this paper is compared with the interpretable multi-view graph convolutional network recognition method (IMvGCN) [39], and the multi-source consistency learning operating condition recognition methods based on different contrastive learning (MAIWRN28-2_InfoCL), based on different attention mechanism fusion (MAIWRN28-2_SELF_CCL, MAIWRN28-2_SE_CCL). The comparison results of different multi-source learning operating condition recognition methods with full labeled training samples are shown in Table III.

From Table III:

1) Compared with the multi-source learning operating condition recognition method based on InfoNCE contrastive learning, our method improves the recognition accuracy by about 1.2%, showing the superiority of CCL adopted in this paper.

2) Compared with the multi-source learning operating condition recognition methods based on SE attention mechanism and self-attention mechanism fusion, the proposed method improves the recognition accuracy by about 0.5% and 0.6%, respectively, showing the advantage of fusion learning of the attention mechanism adopted in this paper.

3) Compared with IMvGCN, our method improves the recognition accuracy by about 2.2%, showing the superiority of the multi-source consistency learning adopted in this paper.

4) Compared with the multi-source learning operating condition recognition methods based on different multi-source learning algorithms, different contrastive learning algorithms and different attention mechanism fusion algorithms, our method has higher recognition accuracy, thus verifying the effectiveness of the proposed method.

TABLE III. THE COMPARISON RESULTS OF VARIOUS OPERATING CONDITION RECOGNITION METHODS BASED ON THE CLASSICAL RESIDUAL NETWORK COMBINED WITH CCL (AVERAGE ACCURACY, %)

Methods	Measured dynamometer cards and measured electrical power cards
MAIWRN28-2_InfoCL	97.63
MAIWRN28-2_SELF_CCL	98.17
MAIWRN28-2_SE_CCL	98.32
IMvGCN	96.57
MAIWRN28-2_CCL(Ours)	**98.79**

C. Multi-Source Learning Operating Condition Recognition with Different Proportions of Labeled Training Samples

To verify the practicability, on the basis of the two parts of the experiment, we add the multi-source consistency learning semi-supervised CCL recognition method based on CoMatch [41] (MAIWRN28-2_CoMatch_CCL), and compare the recognition results in 5 groups of labeled training samples with different proportions (1%, 10%, 30%, 50%, 70%). The comparison results of different multi-source learning operating condition recognition methods with different proportions of labeled training samples are shown in Table IV.

TABLE IV. COMPARISON RESULTS OF DIFFERENT MULTI-SOURCE LEARNING OPERATING CONDITION RECOGNITION METHODS WITH DIFFERENT PROPORTIONS OF LABELED TRAINING SAMPLES (AVERAGE ACCURACY, %)

Methods	1% (n=11)	10% (n=88)	30% (n=25)	50% (n=41)	70% (n=58)
MAIWRN28-2_InfoCL	80.64	91.03	93.64	95.45	96.86
MAIWRN28-2_SELF_CCL	81.17	92.45	95.55	96.08	97.18
MAIWRN28-2_SE_CCL	81.45	92.85	94.81	96.53	97.52
MAIWRN28-2_CoMatch_CCL	79.88	91.12	94.34	95.84	97.56
IMvGCN	**83.79**	92.68	93.65	95.81	96.35
MAIWRN28-2_CCL(Ours)	82.98	**94.13**	**96.18**	**97.20**	**98.25**

From Table IV:

1) In five groups of different proportions labeled training samples, compared with various multi-source semi-supervised learning recognition methods (MAIWRN28-2_SE_CCL, MAIWRN28-2_CoMatch_CCL, MAIWRN28-2_SELF_CCL, MAIWRN28-2_InfoCL, IMvGCN), our method has obtained higher accuracy (except 1%), reflecting the advantage of the semi-supervised learning algorithm adopted.

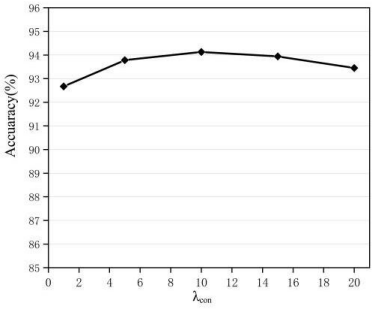
2) Among the recognition results of five groups of labeled training samples with different proportions, 10% (8 labeled training samples for each class) has a particularly significant recognition effect, which shows that our method can effectively

use a large number of unlabeled training samples to further improve the recognition accuracy, thus verifying the practicability of our method.

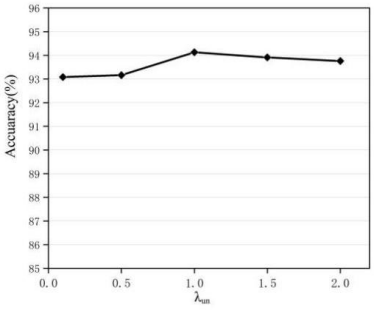
D. Key Hyperparameter Analysis

To explore the influence of hyperparameters on the performance of the operating condition recognition model, an extensive hyperparameter analysis of the proposed method is carried out with 10% proportion labeled training samples. We focus on four major hyperparameters: multi-source information consistency constraint loss weight λ_{con}, unsupervised loss weight λ_{un}, class-aware contrastive loss weight λ_{com}, and the number of heads of multi-head attention in Transformer coding layer. The value range of λ_{con} is [1,5,10,15,20], the value range of λ_{un} is [0.1, 0.5, 1,1.5, 2], the value range of λ_{com} is [0.1, 0.5, 1,1.3, 1.5], the value range of the head number of multi-head attention is [1,2,4,8,16].

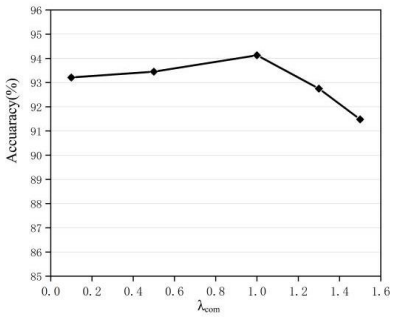
Fig. 5 shows the classification accuracy on the selected data set in different hyperparameter settings. It can be seen from Fig. 5 (a) and (b) that our method shows high stability within certain parameter ranges of λ_{con} and λ_{un}, that is, when λ_{con} , λ_{un} change within a certain ranges, the model accuracy fluctuates only slightly. It can be seen from Fig. 5(c), with the increase of λ_{com}, the impact on the model gradually increases. It can be seen from Fig. 5(d), when the number of heads of attention varies within the value range, the accuracy shows a trend of first increasing and then decreasing, and when the head=2, the accuracy is optimal. By selecting the optimal hyperparameter configuration, the overall performance of our proposed method can be effectively improved.



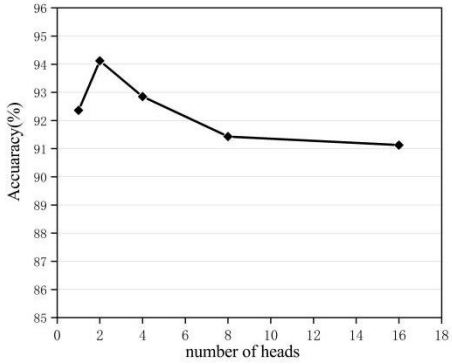
(a) Accuracy with different λ_{con}, λ_{un}=1.0, λ_{com}=1.0, multi-head attention with 2 heads.



(b) Accuracy with different λ_{un} , λ_{con}=10, λ_{com}=1.0, multi-head attention with 2 heads.



(c) Accuracy with different λ_{com} , λ_{con}=10, λ_{un}=1.0, multi-head attention with 2 heads.



(d) Accuracy with different heads of multi-head attention, λ_{com}=1.0, λ_{con}=10, λ_{un}=1.0.

Fig. 5. Key hyperparameter analysis of the proposed method in 10% proportion labeled training samples.

E. Ablation Experiments

Ablation experiments are conducted from three aspects with 10% proportion of labeled training samples: (1) the verification of the effectiveness of the multi-feature fusion algorithm based on attention mechanism. We compare the operating condition recognition results with the pooling learning multi-feature fusion method (MAIWRN28-2_Pooling_CCL) and the fixed-weight multi-feature fusion method (MAIWRN28-2_Fixed-weight_CCL). (2) the verification of the effectiveness of the contrastive learning algorithm. We compare with the multi-feature fusion method based on non-contrastive learning (MAIWRN28-2). (3) the verification of the effectiveness of online augmented data processing. We compare with the multi-feature fusion method without online augmented data processing (MIWRN28-2_CCL). The comparison results of ablation experiments are shown in Table V.

TABLE V. THE COMPARISON RESULTS OF VARIOUS OPERATING CONDITION RECOGNITION METHODS BASED ON THE CLASSICAL RESIDUAL NETWORK COMBINED WITH CCL (AVERAGE ACCURACY, %)

Methods	10% labeled (n=88)
MIWRN28-2_CCL	91.27
MAIWRN28-2	90.18
MAIWRN28-2_Fixed-weight_CCL	90.46
MAIWRN28-2_Pooling_CCL	92.18
MAIWRN28-2_CCL(Ours)	**94.13**

From Table V:

1) Compared with the multi-feature fusion method based on pooling learning and fixed weights, the multi-feature fusion method proposed in this paper can obtain better global feature representation by adaptive adjustment of the feature weights of different information sources through the attention mechanism, thus improving the condition recognition effect.

2) Compared with the multi-feature fusion method based on non-contrastive learning, the proposed method can effectively solve the common confirmation bias problem in semi-supervised deep learning through class-aware contrastive learning, thus further improving the model performance.

3) By using the augmented technique based on geometry operations such as rotation, scaling and expansion, our method can significantly expand the size and diversity of the training set without the need for collecting additional data, thereby improving the model's performance on unknown data.

V. Conclusion and Future Work

In order to solve the limitations of operating condition recognition in the context of big data oil production, such as the bottleneck of traditional single or multi-source operating condition recognition technology, high demand for lots of labeled operating condition data, and inability to make use of massive unknown operating condition data resources, we propose a semi-supervised sucker-rod pumping well operating condition recognition method based on multi-source consistency deep learning. The proposed approach first draws on the design idea of the Transformer coding layer and extracts the deep features of the measured dynamometer cards and the measured electrical power cards, respectively, through a multi-head self-attention mechanism and feedforward neural network. Then, combined with cosine similarity measurement, it realizes the maximum consistency of interactive learning on the complementary feature information of different information sources. Further, the optimal global feature representation of multi-source fusion is obtained by self-learning the weights of the feature representations of different information sources by the attention mechanism. Finally, the multi-source fusion feature is combined with the multi-source semi-supervised class-aware contrastive learning to build the operating condition recognition model and carry out the operating condition recognition. A large number of operating condition recognition comparison experiments show our method can not only mine the deep characteristics of multiple measured information source data but also effectively utilize massive unknown operating condition data in a small amount of labeled operating condition data to further improve the recognition effect and engineering practicability. Key hyperparameter analysis and ablation experiments further verify the effectiveness of the proposed method.

Future Work

The future directions of this paper worth further exploration are: (1) The study of the influence of heterogeneous data from multiple sources on the universality and generalization of the operating condition recognition model. In this paper, two kinds of image information sources, dynamometer cards and electrical power cards, are used to carry out research by combining the adaptive interactive learning fusion method based on the attention mechanism. The sucker-rod pumping well production system is a complex nonlinear system with mechanical-electrical-hydraulic coupling, and operating conditions are complex and changeable. Therefore, it is worth studying more comprehensive and appropriate information sources and more effective fusion learning methods. (2) The exploration of more efficient semi-supervised deep learning methods to further optimize the operating condition recognition model performance in insufficient labeled samples.

Acknowledgment

This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR2021MF031.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] Z. Huang, K. Li, C. Ke, H. Duan, M. Wang, S. Bing, "An intelligent diagnosis method for oil-well pump leakage fault in oilfield production Internet of Things system based on convolutional attention residual learning," Engineering Applications of Artificial Intelligence, vol. 126, p., 106829, 2023.

[2] W. Wu, X. Xing, H. Wei, B. Li, X. Wang, "Fault diagnosis of pumping system based on multimodal attention learning (CBMA Learning)," Journal of Process Control, vol. 128, p. 103006, 2023.

[3] Z. W. Ye, Q. J. Yi, "Working-condition diagnosis of a beam pumping unit based on a deep-learning convolutional neural network," Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, vol. 236, no. 5, pp, 2559-2573, 2022.

[4] J. L. Wei, X. W. Gao, "Fault diagnosis of sucker rod pump based on deep-broad learning using motor data," IEEE Access, vol. 8, pp. 222562-222571, 2020.

[5] J. N. Li, J. Shao, W. Wang, W. H. Xie, "An evolutional deep learning method based on multi-feature fusion for fault diagnosis in sucker rod pumping system," Alexandria Engineering Journal, vol. 66, pp. 343-355, 2023.

[6] Y. He, Z. Guo, X. Wang, W. Abdul, "A Hybrid Approach of the Deep Learning Method and Rule-Based Method for Fault Diagnosis of Sucker Rod Pumping Wells," Energies, vol. 16, no. 7, p. 3170, 2023.

[7] S. S. Abdurakipov, M. Dushkin, D. Del'tsov, E. B. Butakov, "Diagnostics of Oil Well Pumping Equipment by Using Machine Learning," Journal of Engineering Thermophysics, vol. 33, no. 1, pp. 39-54, 2024.

[8] H. Li, H. Niu, Y. Zhang, Z. Yu, "Research on indirect measuring method of dynamometer diagram of sucker rod pumping system based on long-short term memory neural network," Journal of Intelligent & Fuzzy Systems, vol. 45, no. 3, pp. 4301-4313, 2023.

[9] X. Wang, Y. He, F. Li, Z. Wang, X. Dou, H. Xu, et al., "A working condition diagnosis model of sucker rod pumping wells based on big data deep learning," International petroleum technology conference (IPTC), Beijing, China, 2019, pp. 317–326.

[10] Y. P. He, C. Z. Zang, P. Zeng, M. X. Wang, Q. W. Dong, G. X. Wan, et al., "Few-shot working condition recognition of a sucker-rod pumping system based on a 4-dimensional time-frequency signature and meta-learning convolutional shrinkage neural network," Petroleum Science, vol. 20, no. 2, pp. 1142-1154, 2023.

[11] Z. Ma, Y. Chen, Y. Fan, X. He, W. Luo, J. Shu, "An improved AoT-DCGAN and T-CNN hybrid deep learning model for intelligent diagnosis of PTCs quality under small sample space," Applied Sciences, vol. 13, no. 15, p. 8699, 2023.

[12] D. Z. Hao and X. W. Gao, "Unsupervised Fault Diagnosis of Sucker Rod Pump Using Domain Adaptation with Generated Motor Power Curves," Mathematics, vol. 10, no. 8, p. 1224 ,2022.

[13] Y. P. He, H. B. Cheng, P. Zeng, C. Z. Zang, Q. W. Dong, G. X. Wan, et al., "Working condition recognition of sucker rod pumping system based on 4-segment time-frequency signature matrix and deep learning," Petroleum Science, vol. 21, no.1, pp. 641-653, 2024.

[14] R. Zhao, C. Wang, H. Zhao, C. Xiong, J. Shi, X. Zhang, et al., "Research and Application of Rod Pump Working Condition Diagnosis and Virtual Production Metering Based on Electric Parameters," In SPE Middle East Oil and Gas Show and Conference , SPE, 2021.

[15] P. Zhang, T. Li, G. Wang, C. Luo, H. Chen, J. Zhang et al., "Multi-source information fusion based on rough set theory: A review," Information Fusion, vol. 68, pp. 85-117, 2021.

[16] J. Huang, Z. Chen, Q. J. Wu, C. Liu, H. Yuan, W. He, "CATFPN: Adaptive feature pyramid with scale-wise concatenation and self-attention," IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 12, pp. 8142-8152, 2021.

[17] A. Kumar, P. Rai, H. Daume, "Co-regularized multi-view spectral clustering," Advances in neural information processing systems, vol. 24, 2011.

[18] G. Qian, S. Sural, Y. Gu, S. Pramanik, "Pramanik. Similarity between Euclidean and cosine angle distance for nearest neighbor queries," In Proceedings of the 2004 ACM symposium on Applied computing, 2004, pp. 1232-1237.

[19] S. Zagoruyko, N. Komodakis, "Wide Residual Networks," In British Machine Vision Conference 2016. British Machine Vision Association, 2016, arxiv preprint arxiv:1605.07146.

[20] Y. Liu, Y. Zhang, Y. Wang, F. Hou, J. Yuan, J. Tian, et al. "A survey of visual transformers," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 6, pp. 7478 - 7498, 2023.

[21] Z. Wang, J. Xuan, T. Shi, "Multi-source information fusion deep self-attention reinforcement learning framework for multi-label compound fault recognition," Mechanism and Machine Theory, vol. 179, p. 105090, 2023.

[22] X. Yang, Z. Song, I. King, Z. Xu, "A survey on deep semi-supervised learning," IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 9, pp. 8934-8954, 2022.

[23] B. Zhang, Y. Wang, W. Hou, H. Wu, J. Wang, M. Okumura, et al., "Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling," Advances in Neural Information Processing Systems, vol. 34, pp. 18408-18419, 2021.

[24] Y. Fan, A. Kukleva, D. Dai, B. Schiele, "Revisiting consistency regularization for semi-supervised learning," International Journal of Computer Vision, vol. 131, no. 3, pp. 626-643, 2023.

[25] W. Ma, F. Cheng, Y. Xu, Q. Wen, Y. Liu, "Probabilistic representation and inverse design of metamaterials based on a deep generative model with semi-supervised learning strategy," Advanced Materials, vol. 31, no. 35, p. 1901111, 2019.

[26] Z. Song, X. Yang, Z. Xu, I. King, "Graph-based semi-supervised learning: A comprehensive review," IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 11, pp. 8174-8194, 2022.

[27] Y. Gan, H. Zhu, W. Guo, G. Xu, G. Zou, "Deep semi-supervised learning with contrastive learning and partial label propagation for image data," Knowledge-Based Systems, vol. 245, p. 108602, 2022.

[28] F. Yang, K. Wu, S. Zhang, G. Jiang, Y. Liu, F. Zheng, et al., "Class-aware contrastive semi-supervised learning," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 14421-14430.

[29] A. Zhang, X. W. Gao, "Supervised dictionary-based transfer subspace learning and applications for fault diagnosis of sucker rod pumping systems," Neurocomputing, vol. 338, pp. 293-306, 2019.

[30] K. Li, X. W. Gao, Z. Tian, "Using the curve moment and the PSO-SVM method to diagnose downhole conditions of a sucker rod pumping unit," Petroleum Science, vol. 10, pp. 73–80, 2013.

[31] B. Y. Zheng, X. W. Gao, "Sucker rod pumping diagnosis using valve working position and parameter optimal continuous hidden Markov model," Journal of Process Control, vol. 59, pp. 1–12, 2017.

[32] A. Zhang, X. W. Gao, "Fault diagnosis of sucker rod pumping systems based on Curvelet Transform and sparse multi-graph regularized extreme learning machine," International journal of computational intelligence systems, vol. 11, pp. 428–437, 2018.

[33] B. Y. Zheng, X. W. Gao, X. Y. Li, "Fault detection for sucker rod pump based on motor power," Control Engineering Practice, vol. 86, pp. 37–47, 2019.

[34] D. C. Chen, R. Q. Zhou, H. X. Meng, Y. Peng, F. Chang, D. Jiang et al., "Fault diagnosis model of the variable torque pumping unit well based on the power-displacement diagram," IOP Conference Series: Earth and Environmental Science, 2019.

[35] B. Y. Zheng, X. W. Gao, X. Y. Li, "Diagnosis of Sucker Rod Pump based on generating dynamometer cards," Journal of Process Control, vol. 77, pp. 76–88, 2019.

[36] R. Zhang, Y. Yin, L. Xiao, "A real-time diagnosis method of reservoir-wellbore-surface conditions in sucker-rod pump wells based on multidata combination analysis," Journal of Petroleum Science and Engineering, vol. 198, p. 108254, 2021.

[37] S. Liu, C. S. Raghavendra, Y. Liu, K. Yao, O. Balogun, L. Olabinjo, et al., "Automatic early fault detection for rod pump systems," SPE Annual Technical Conference and Exhibition: OnePetro, Denver, Colorado, USA, 2011.

[38] B. Zhou, R. Niu, S. Yang, J. G. Yang, W. W. Zhao, "Multisource working condition recognition via nonlinear kernel learning and p-Laplacian manifold learning," Heliyon, vol. 10, no. 5, p. E26436, 2024.

[39] Z. Wu, X. Lin, Z. Lin, Z. Chen, Y. Bai, S. Wang, "Interpretable graph convolutional network for multi-view semi-supervised learning," IEEE Transactions on Multimedia, vol. 25, pp. 8593-8606, 2023.

[40] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, et al., "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," Advances in neural information processing systems, vol. 33, pp. 596-608, 2020.

[41] I. Hernandez-Sequeira, R. Fernandez-Beltran, Y. Xu, P. Ghamisi, F. Pla, "Semi-supervised classification for remote sensing datasets," International Conference on Image Analysis and Processing, Cham: Springer Nature Switzerland, 2023, pp. 463-474.

Development of a Smart Water Dispenser Based on Object Recognition with Raspberry Pi 4

Dani Ramdani[1], Puput Dani Prasetyo Adi[2]*, Andriana[3], Tjahjo Adiprabowo[4],
Yuyu Wahyu[5], Arief Suryadi Satyawan[6], Sally Octaviana Sari[7], Zulkarnain[8], Noor Rohman[9]
Department of Electrical Engineering, Universitas Langlangbuana, Bandung, Indonesia[1, 3, 4, 7, 8, 9]
National Research and Innovation Agency (BRIN), Indonesia[2, 5, 6]

***Abstract*—In this project, we develop and apply a Smart Water Dispenser system, which is combined with object recognition and fluid level control supported by Ultrasonic Sensors, Raspberry Pi, and also DC Motors. The essence of this system is to develop a system using the Raspberry Pi 4 Model B with other components that have been integrated and interrelated Hardware and programming using OpenCV, YOLO V8, and other components, the point is that the cup can be detected, and water filling is done precisely and automatically. The process carried out is the detection of cups automatically using Raspberry Pi which is in charge of controlling the DC Motor and also the Ultrasonic sensor (HC-SR04) and detecting based on the volume of water with precision. The dispenser functions to pump water based on the condition of the volume of water in the glass and stop pumping if the volume of the glass has been fulfilled, aka not spilling with a percentage of 90%. In the scenario process, the cup search process is first carried out by scanning three times until a cup is found, if a cup is found, then the sensor component and the valve for the release of water in the hose will stop right at the position of the cup and the water will fill the cup automatically. Otherwise, the system will move backward and the system will be turned off. The first testing process has been successful and shows the effectiveness of the system in the process of finding cups and managing water levels. This innovation shows hope for improving user comfort, especially for disabilities, utilizing advanced technology in object recognition, and of course, saving water usage. In the testing process obtained 95% to 97% accuracy in object detection with different types of cups.**

Keywords*—Smart water dispenser; object recognition; Raspberry Pi 4; YOLO VB; ultrasonic sensor

I. Introduction

The demand for automation solutions has significantly increased in recent years, driven by advancements in Internet of Things (IoT) technologies and smart devices. IoT-based systems enable the interaction of interconnected electronic devices that can be automatically controlled and monitored, transforming various sectors, including smart homes, industries, and offices, toward enhanced convenience and efficient resource use. One of the promising applications of smart technology is in the development of automated water dispenser systems, which provide water efficiently and accurately, reducing the need for manual methods that are prone to wastage and contamination [1].

Traditional water dispensers rely heavily on manual operation, which often leads to inefficient water usage. These conventional systems cannot detect the user's needs or monitor the fluid level, which presents several limitations, particularly in environments that demand precise control of water use. Consequently, the integration of smart water dispensers equipped with sensors and object recognition technology has become an ideal solution. These systems can automatically detect the presence of objects like cups and dispense the appropriate amount of water, minimizing direct contact and saving users time. Research conducted by Raspberry Pi (2023) demonstrates that an automatic water bottle filling system using Raspberry Pi can detect bottles and adjust the water flow according to the object's needs, showcasing the practical potential of such technology [2]. In addition, the important conclusions of this research are: a) the importance of this study: the development of this research will be able to automate the process of filling water, especially cups. And can be applied to patients with disabilities and can be applied for other purposes according to their needs and development using sophisticated technology following novelty research. b) Recent advances in technologies applied to this research are how to use object recognition technology using artificial intelligence using the Convolutional Neural Network (CNN) method and a combination with other methods in Machine Learning and Deep Learning. Technology is evolving rapidly with various approaches such as Image Recognition [25-30] and Convolution Neural Network (CNN). Some studies use Image Recognition with various approach methods [31-40] and also hardware and Raspberry Pi 4 b Microprocessor as a processor.

II. Literature Review

A. Object Recognition Technology in Water Dispensers

Object recognition has emerged as a core element in the development of smart devices, especially for systems such as water dispensers that require accurate detection of specific objects before initiating any action. Object recognition technology enables the system to "see" and interpret the surrounding environment so that water is dispensed only when a suitable object, like a cup, is detected. This feature not only reduces water waste but also contributes to better user convenience and hygiene. One of the widely used frameworks for object recognition is OpenCV, paired with YOLO (You Only Look Once) to enhance detection accuracy and speed in real-time applications [3], [4]. In the context of water dispensers, a camera connected to the Raspberry Pi leverages these algorithms to detect the presence of a cup accurately and in a short time, allowing for efficient control of water flow [5].

*Corresponding Author

B. The Role of Raspberry Pi in IoT Applications

Raspberry Pi, particularly the Raspberry Pi 4 model, has become a popular choice in developing IoT applications due to its affordability, compact size, and adequate processing capabilities. The Raspberry Pi 4 is capable of handling image processing and real-time object detection operations, which are critical for applications such as smart water dispensers that must respond promptly to the presence of objects. For applications that require object detection, the Raspberry Pi can be programmed to process inputs from the camera and send signals to other devices, such as water pumps and motors, to control fluid dispensing automatically [6]. This capability is particularly valuable in smart water dispenser systems, as it enables the device to function autonomously with minimal user intervention.

C. Sensor Technology and Fluid Dispensing Control

In addition to object recognition, sensor technology plays an essential role in the functionality of smart water dispensers. Sensors are used to detect the presence of objects, monitor fluid levels, or gauge environmental conditions surrounding the device. A recent article on Instructables (2023) explored a water level monitoring system using Raspberry Pi, which could be integrated into smart water dispensers to control water flow. This setup allows for controlled water dispensing according to the capacity of the container or cup being used, thereby reducing the risk of spills or underfilled containers [7]. While traditional dispensers may release excess water or require manual monitoring, sensors in smart dispensers ensure the precise amount of water is dispensed every time, supporting both user convenience and sustainability.

D. The Need for Automation in Daily Life

Automation has become increasingly relevant in modern households, primarily through smart home devices that enable a range of tasks to be conducted without manual intervention. Automation also promotes environmental sustainability by optimizing resource use, such as water. Smart water dispenser systems are designed to address two main concerns: efficient water usage and enhanced user convenience. These systems also reduce direct contact with the device, making them a hygienic solution in both public and private settings. A fully automated water dispenser that uses object recognition technology exemplifies the extent to which IoT and sensor-based automation can improve daily life by conserving resources and minimizing the risk of contamination [8].

E. Benefits of the Study and Contributions to Smart Home Technology

This research contributes to the field of smart home technology by developing a device that not only functions automatically but also interacts intelligently with the user. A system utilizing Raspberry Pi and object recognition technology can expand the capabilities of smart homes in terms of automation, providing a sustainable solution for water usage. The implementation of this technology is not only beneficial for household users but can also be applied in offices, schools, and public facilities where efficiency and reduced physical contact are priorities. By offering a self-contained water dispensing solution, this study bridges the gap between basic automated dispensers and advanced IoT-integrated devices that respond to specific user needs [9].

III. METHOD

This study employs an experimental method to design, implement, and test a smart water dispenser system utilizing object recognition technology and automated control. Below are the methodology steps applied in this research, complemented with relevant references.

A. System Design

1) Component selection: This study uses the Raspberry Pi 4 Model B as the main controller, the Raspberry Pi Camera V2 for object recognition, and a DC motor to drive the water flow mechanism. Other key components include a water pump and relay to control water flow as needed [10], [11].

2) Object recognition integration with OpenCV and YOLO: The system employs the OpenCV library and YOLO (You Only Look Once) algorithm for real-time object detection. YOLO allows high-speed and accurate object detection, making it suitable for detecting the presence of glass in the filling area in real-time [12], [13].

3) Sensor setup and water flow control: The DC motor and relay are connected to the Raspberry Pi to automatically control water flow. When a glass is detected, the Raspberry Pi activates the water pump through the relay and drives the motor to fill the glass to the desired level [14]. The Raspberry Pi controls the movement of the DC motor, which carries a camera to scan for the presence of a glass. When a glass is detected, the Raspberry Pi then checks the water level using an ultrasonic sensor. If the glass is empty or the water level is insufficient, the water pump is activated to fill the glass until the desired level is reached.

B. Implementation Phase

1) Raspberry Pi programming: The Raspberry Pi is programmed using Python to run the OpenCV and YOLO V8 libraries. The program processes the image input from the camera to detect the presence of a glass. When a glass is detected, the Raspberry Pi sends a signal to the relay to activate the water pump and control the DC motor [15], and The water level using an ultrasonic sensor.

2) Camera configuration and object detection algorithm: The Raspberry Pi Camera V2 is positioned at a specific angle to scan the dispenser area. The YOLO algorithm is implemented in Python on the Raspberry Pi for real-time object detection. A threshold setting in YOLO is applied to ensure that only objects meeting certain criteria (such as the shape and size of glass) and Moved by a DC motor are identified as targets [16].

3) Initial testing and system calibration: After the hardware and software integration is complete, initial testing is conducted to ensure all components function as designed. Testing includes camera calibration or positioning to detect objects accurately and setting the Pump activation timing based on the measurement of the water level based on ultrasonic sensor readings [17]. Pump activation timing based on

measurement of the water level based on ultrasonic sensor readings.

C. Testing and Data Collection Phase

1) Object detection accuracy testing: The system is tested to detect the presence of glasses of various shapes and sizes. This test aims to ensure that the YOLO and OpenCV algorithms can detect objects accurately under different lighting conditions and positions [18].

2) Reliability testing of automatic water dispensing: The system is tested with different glass sizes to ensure the automatic water dispensing matches the glass's capacity. Data is collected to evaluate the system's reliability in stopping the water flow once the desired volume is reached [19].

3) System stability testing: Stability testing is performed to ensure the system can operate consistently over a long period without failure. This test includes monitoring power consumption, sensor durability, and detection accuracy over multiple usage cycles [20].

D. Data Analysis

1) Object detection accuracy analysis: The object detection test results are analyzed based on the success rate of detection relative to the number of trials. This analysis aims to determine the system's accuracy level in recognizing the presence of a glass under various conditions [21].

2) Automatic water dispensing effectiveness analysis: The effectiveness of automatic water dispensing is analyzed by measuring the precision of the dispensed water volume according to the glass size. The error percentage in the dispensed water volume is calculated to evaluate the system's precision [22].

3) Overall system evaluation: An overall evaluation of the system is conducted by considering data from object detection accuracy testing, water dispensing reliability, and system stability. The analysis results will demonstrate the system's effectiveness in reducing water waste and enhancing user convenience [23].

E. Conclusion and Development Recommendations

The data analysis results will be summarized to conclude the success of implementing this smart water dispenser system. Recommendations for further development will also be included, particularly in improving detection accuracy and water dispensing efficiency [24].

This diagram provides a visual representation of the workflow and integration of each component, helping readers understand the process from start to finish within the system.

F. Flowchart Smart Water Dispenser

The smart water dispenser flowchart and explanation are depicted in Fig. 1.

Explanation of Components in the Block Diagram

1) Raspberry Pi 4 Model B:

- Serves as the main controller, handling both object recognition and control of hardware.
- Integrates with the Raspberry Pi Camera V2 using OpenCV and YOLO v8 for real-time glass detection.
- Controls the DC Motor with L298N Motor Driver to move the conveyor.

2) Camera and Detection:

- The Raspberry Pi Camera scans for the presence of a glass as it moves on the conveyor.
- If a glass is detected, Raspberry Pi stops the motor, aligns the spout, and activates the pump.

3) Water Pump and Ultrasonic Sensor:

- Raspberry Pi activates the relay to start the water pump and then uses the Ultrasonic Sensor HC-SR04 to monitor the water level in the glass.
- Once the water level reaches 90% of the glass height, the relay turns off the pump. And then activate the relay to start the water pump.

4) Process Flow:

- The DC motor moves the conveyor for glass detection.
- When the glass is detected, the motor stops, the water pump is activated, and the water level is monitored by the ultrasonic sensor.
- Once the required water level is reached, the pump stops, and the motor returns to its initial position.

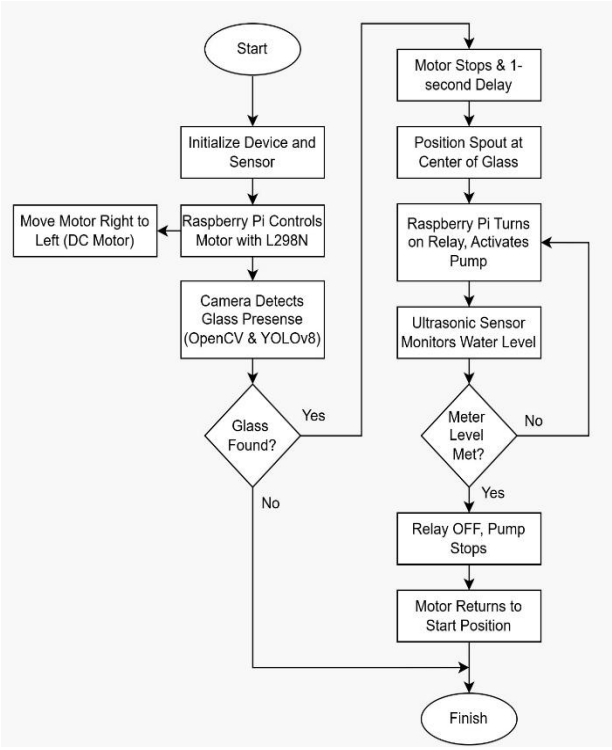


Fig. 1. Flowchart smart water dispensers.

Moreover, Fig. 3, 4, and 5 are a frame of hardware that outlines the efficient workflow of the smart dispenser system controlled entirely by Raspberry Pi, handling both the detection and water dispensing process.

Moreover, The Hardware and Software Design:

1) Raspberry Pi 4 model B: Used for object recognition programs utilizing the Raspberry Pi Camera V2. The Raspberry Pi Camera V2 is programmed with the OpenCV application and YOLO V8 algorithm.

2) Control system: Control Systems include a 17-stepper motor, L298N motor driver, HC-SR04 ultrasonic sensor, and a relay for turning the water pump on and off. The ultrasonic sensor, camera, and water hose are mounted together on a conveyor system, whose movement is synchronized with the rotation of the stepper motor.

Fig. 2 shows the system configuration built in this research. There are several sets of hardware and software installed and placed according to their functions and uses. Moreover. The method taken in this research is Object Detection and retrieval and equalization on the dataset side, such as Convolutional Neural Network (CNN) by utilizing a smart camera as hardware connected to a Raspberry Pi 4b and also the Python programming language, such as the YOLO and OpenCV platforms. The goal is to build a smart system on the dispenser Therefore, it can fill water automatically.

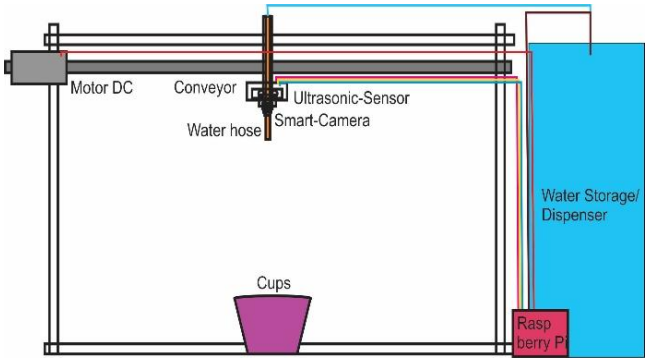


Fig. 2. Configuration of the built tool.

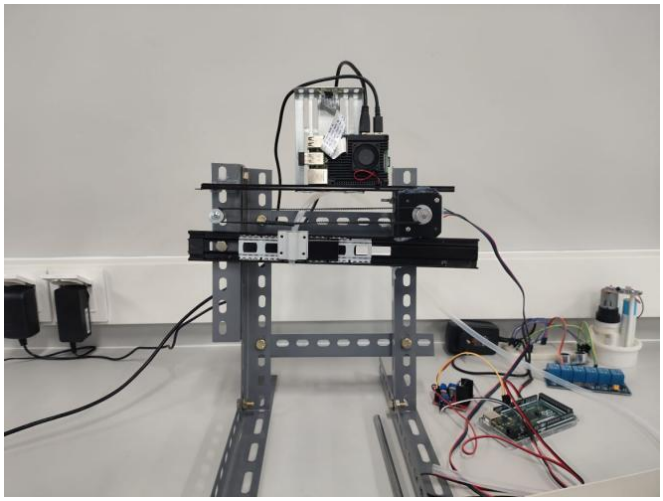


Fig. 3. A series of hardware in this system.

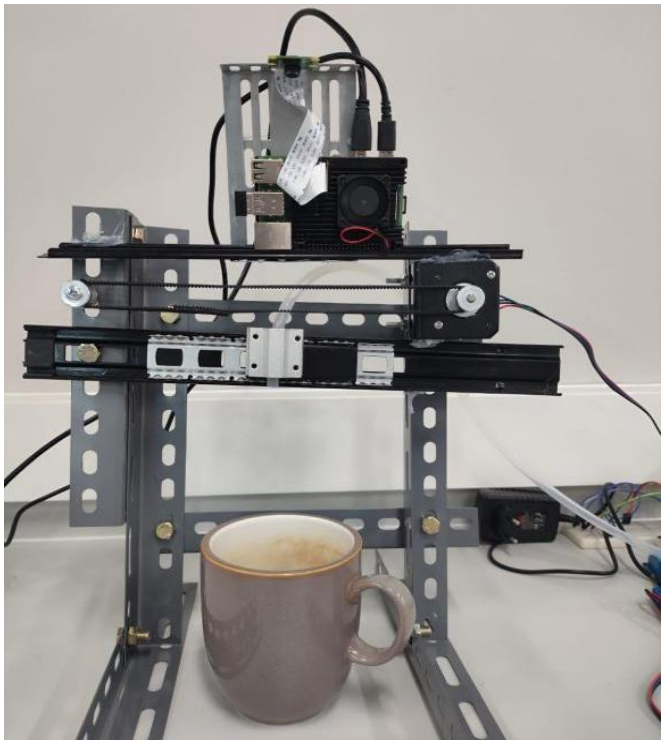


Fig. 4. Frame of hardware.

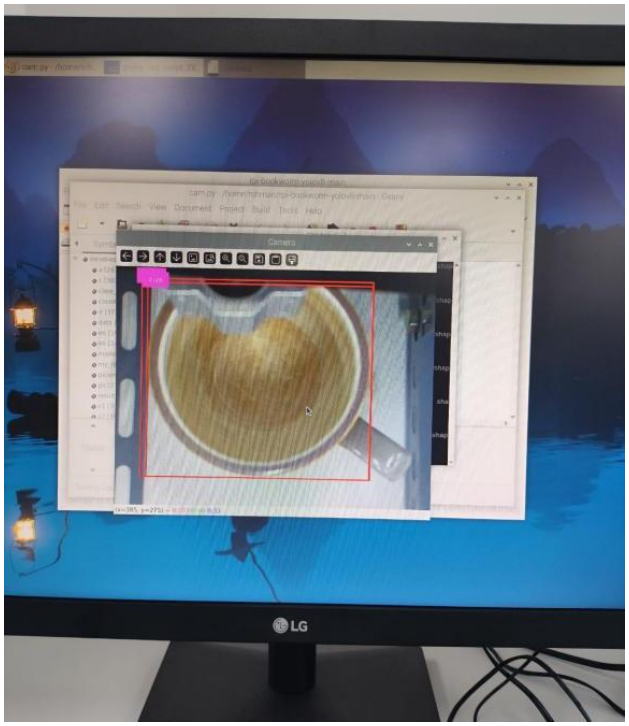


Fig. 5. Glass object detection system.

The configuration system is also carried out on the DC Motor in running the Conveyor so that the camera and other devices such as the Ultrasonic Sensor and the Mini Pump can work precisely.

IV. Result and Discussion

This section presents the outcomes of testing conducted to evaluate the performance of the smart water dispenser system in terms of object detection accuracy, reliability of automatic water dispensing, and overall system stability. The discussion is based on data obtained from various testing scenarios to assess system effectiveness and identify potential improvements. The innovation proposed in this research is how to produce an invention that can have intelligence in terms of automatic water filling, specifically in terms of detecting glass objects and classifying them. The camera moves to detect the presence of a glass. Once a glass is detected, water will flow, and the level is monitored using an ultrasonic sensor. When the water level reaches the desired point, the filling process stops, and the motor returns to its initial position. If the camera does not detect a glass by the end of its path, it will return and continue scanning until a glass is found, or the user turns off the device.

Steps of Operation:

1) Initialization and Conveyor Movement: When the device is powered on, the Arduino Mega commands the motor to move the conveyor from right to left. This movement shifts the camera's position to scan for the presence of a glass.

2) Object Detection by the Camera: The Raspberry Pi Camera V2, running an OpenCV and YOLO V8 program, detects objects. When a glass (cup) is identified, the motor stops after a 1-second delay. This delay ensures that the water hose is positioned directly above the center of the glass.

3) Water Pump Activation: When the motor halts (assuming the hose is centered over the glass), the Arduino Mega triggers the relay to turn on the water pump, starting the filling process.

Monitoring Water Level: The ultrasonic sensor monitors the water level until it reaches a predetermined height (e.g., 90% of the glass height). Once the desired level is reached, the relay is turned off, and the pump stops. Return to Initial Position The motor moves back to the initial position (left side) after completing the filling process. Moreover, the Pseudocode 1 will explain in detail how this system works.

A. Test Results

1) Object detection accuracy: The initial test focused on the system's ability to detect the presence of glass using the YOLO algorithm integrated with OpenCV. Tests were conducted with glasses of different sizes and shapes under various lighting conditions. The results showed that the system achieved an accuracy rate of 94% under optimal lighting and 89% in low-light conditions. Some inaccuracies were observed with objects that had high reflectivity, such as metal or transparent glasses, which affected detection accuracy.

1. **Import library**, GPIO, Camera, OpenCV, YOLO
2. Setup GPIO
3. Function definition (motor movement, Ultrasonic distance reading, water filling)
4. The motor starts moving forward
- Record start moving time (initial time set) 4.

5. LOOP for True:
- Take a picture from the Pi camera
- Object detection using YOLO on the image

A. If the detected object is a "*cup*":
- Stop the motor
- Record the time after the motor stops (set stop time)
- Call the water fill () function to start filling the water
- Calculate the rewind duration (stop time - start time)
- Rewind the motor for the calculated duration
- Stop the motor

B. If the motor runs for more than 20 seconds without detecting the glass:
- Stop the motor
- Rewind the motor for 20 seconds (to return to the starting position)
- Stop the motor
- Display the detection result image

--------- *Pseudocode 1*----------

2) Reliability of automatic water dispensing: The second test assessed the system's reliability in filling glasses to their respective capacities. The system was tested with three different glass sizes (small, medium, and large). Results indicated that the system consistently dispensed water with a margin of error of ±5% from the target capacity. Automatic filling stopped precisely when the volume matched the glass capacity, demonstrating the system's effectiveness in preventing water spillage and overuse. +10% and -16%. With a target water level of 5cm, the water level varies between 4.2 to 5.5cm. error range + 0,5cm and - 0,8cm. This is due to the uneven water surface when filling the water.

3) System stability: Stability tests were conducted to evaluate the system's ability to operate consistently over multiple cycles. The system was tested across 50 continuous usage cycles without a reset, and results showed no performance degradation or component failure. This test demonstrated that both the hardware and software components of the system are stable and reliable for long-term use.

B. Discussion

The test result and analysis are explained in Table I.

1) Analysis of object detection accuracy: Test results revealed that high detection accuracy can be achieved with the YOLO algorithm running on a Raspberry Pi, especially under well-lit conditions. However, accuracy slightly decreased in low-light scenarios, highlighting the importance of lighting in detection performance. To improve accuracy for reflective or transparent objects, additional techniques such as image preprocessing or adjusting the camera angle to reduce reflection effects can be considered.

2) Evaluation of automatic water dispensing reliability: The system's success in filling glasses accurately demonstrates the effectiveness of its automated control mechanisms. A small margin of error indicates that the system provides precise control over water volume, which is crucial for water conservation and spill prevention. The use of relays and DC motors controlled by the Raspberry Pi proved capable of

delivering rapid and stable responses for automatic water dispensing. For applications requiring higher precision, additional sensors to measure water levels within the glass could enhance control accuracy.

3) System stability evaluation: Stability testing indicated that the system operates consistently over extended periods without failure. This highlights the durability of the hardware, including the Raspberry Pi, camera, and control components. Such reliability is critical for household or industrial applications where repeated use is expected. The stability test also demonstrated that the system functions efficiently with minimal power consumption, avoiding overheating or technical issues.

4) Implications and potential for further development: The test results suggest that the smart water dispenser system holds significant potential for household or public applications. With reliable object detection, precise water dispensing, and high system stability, it offers an effective solution for water conservation and user convenience. Potential improvements include integrating additional sensors for water level detection, employing enhanced filters for transparent object detection, and optimizing the algorithm to improve accuracy under diverse lighting conditions.

TABLE I. TESTING RESULT FOR SMART WATER DISPENSER SYSTEM

No.	Test Type	Parameter	Test Result	Description
1.	Object Detection Accuracy	Optimal Lighting Condition	94 % Object detection accuracy	High accuracy under optimal lighting conditions
		Low Lighting Condition	89 % Object detection accuracy	Reduced accuracy in low-lighting
		Transparent/Reflective Objects	75 % Object detection accuracy	Transparent and reflective objects are harder to detect
2.	Automatic Water Dispensing Reliability	Small Glass (150 ml)	95% accuracy (±5 % margin of error)	Accurate dispensing for small glass capacity
		Medium Glass (250 ml)	97% accuracy (±5 % margin of error)	Accurate dispensing for medium glass capacity
		Large Glass (500 ml)	96% accuracy (±4 % margin of error)	Accurate dispensing for large glass capacity
3.	System Stability	Continuous Operation Cycles	50 cycles without failure	The system operates stably over 50 cycles without interruptions
		Power Consumption	5V, 2.5 A (Rasberry Pi Standard)	The system operates with minimal power consumption
		Operating Temperature	40-45° (no overheating)	Temperature remains stable during prolonged operation

Explanation of the Table I Testing:

1) Object detection accuracy: Tests were conducted to measure the system's accuracy in detecting the presence of a glass under various conditions. Under optimal lighting, the system achieved 94% accuracy, while low lighting conditions slightly reduced accuracy to 89%. Transparent or reflective objects, such as glass, resulted in lower detection accuracy (75%). In optimal lighting conditions or less light, glass detection can still run well. but if the size of the glass used is not standard, it will affect the detection accuracy, especially if the glass is transparent without color, the accuracy will decrease even more.

2) Automatic Water Dispensing Reliability: This test measured the system's accuracy in dispensing water according to the glass capacity. Three different glass sizes were tested, showing a low margin of error across all sizes (small, medium, and large), indicating high accuracy in controlling the water volume dispensed.

3) System Stability testing: It was conducted to ensure the system could operate continuously over multiple cycles. The system showed consistent performance over 50 cycles without failure, maintained stable operating temperatures (40-45°C), and consumed minimal power (5V, 2.5A).

V. CONCLUSION

This study successfully developed and implemented a smart water dispenser system using Raspberry Pi and object recognition technology to automate water dispensing based on the detection of a glass. The system incorporates OpenCV and YOLO algorithms for real-time object detection and controls a DC motor and water pump to ensure precise and reliable water filling. Testing demonstrated that the system achieved a high detection accuracy of 94% under optimal lighting conditions and maintained stable performance over multiple cycles, with no operational failures observed.

The system's ability to accurately detect different glass sizes and automatically adjust the water volume demonstrates its effectiveness in reducing water waste and enhancing user convenience. Moreover, the stable operating temperature and minimal power consumption indicate that the system is suitable for continuous use in real-world applications, whether in household settings or public spaces.

In summary, the smart water dispenser system shows promising potential for improving the efficiency and sustainability of water use in smart homes and IoT applications. Future improvements, such as adding sensors for enhanced water level detection and optimizing algorithms for detecting transparent or reflective objects, could further enhance the system's robustness and broaden its range of applications.

The strength of this research is the process of filling water in the glass automatically, by involving the Convolutional Neural Network (CNN) method, so that the water filling process has its intelligence and can be applied to patients with disabilities, but this research has a weakness, namely still using a scanning system on the cup at point A to point B statically, not having a Degree of Freedom (DOF) like the Robot Arm.

ACKNOWLEDGMENT

Gratitude is extended to the National Research and Innovation Agency (BRIN) for their support, ideas, and guidance in the development of this research. Appreciation is also given to the research institution for providing facilities and technical assistance throughout the research process. Special thanks to the Faculty of Engineering and the Department of Electrical Engineering for providing funding, facilities, and necessary infrastructure. The support from all parties has been invaluable to the smooth conduct of this research. It is hoped that the results of this research will contribute to the advancement of science and technology in the future. Thanks also to other universities who helped collaborate in improving this manuscript until it was accepted.

FUTURE RESEARCH

The development of this Smart Water Dispenser is still static, so it is necessary to make changes to the side of the water filling hose and Degree of Freedom (DOF) as in the Robot Arm, so that the position of the cup that is anywhere in the X, Y, Z coordinates, can fill water with great precision.

REFERENCES

[1] Mukherjee, A. G., Wanjari, U. R., Chakraborty, R., Renu, K., Vellingiri, B., George, A., ... & Gopalakrishnan, A. V. (2021). A review on modern and smart technologies for efficient waste disposal and management. Journal of Environmental Management, 297, 113347.

[2] Gnann, N., Baschek, B., & Ternes, T. A. (2022). Close-range remote sensing-based detection and identification of microplastics on water assisted by artificial intelligence: a review. Water Research, 222, 118902.

[3] Wahab, F., Ullah, I., Shah, A., Khan, R. A., Choi, A., & Anwar, M. S. (2022). Design and implementation of real-time object detection system based on single-shoot detector and OpenCV. Frontiers in Psychology, 13, 1039645.

[4] Gheorghe, C., Duguleana, M., Boboc, R. G., & Postelnicu, C. C. (2024). Analyzing Real-Time Object Detection with YOLO Algorithm in Automotive Applications: A Review. CMES-Computer Modeling in Engineering and Sciences, 141(3), 1939-1981.

[5] Omar, A. A. (2022). Development of Water Surface Robot System for Lake Sanitation and Sampling (Master's thesis, University of Malaya (Malaysia)).

[6] Aluma, I. (2019). IGN & implementation of smart tap system using Raspberry PI (Case study: Primary schools) (Doctoral dissertation, Kampala International University, School of Engineering and Applied Sciences (SEAS) Department of Electrical, Telecom and Computer Engineering).

[7] Müller, B. J. (2020). The design process of coffee creamer portion packaging concepts that can be correctly sorted during the recycling process (Bachelor's thesis, University of Twente).

[8] Bautista, P. S. (2024). Iot-based smart cage with water filtering monitoring system for cats and dogs.

[9] Ahmed, S. (2022). Application of Emerging Digital Technologies in Self-monitoring of Health and Water Usage.

[10] Islam, M. J., Islam, M. Z., & Dudaib, S. A. Fabrication of a prototype model of a solar-powered smart water level monitor system.

[11] Mamun, A. T., Sadik, M. D., Shefat, T. S., Adnan, A. H., & Bhuyan, M. H. (2023). IOT-BASED ROBOTIC CAR WITH LIVE STREAMING SYSTEM (Doctoral dissertation, Faculty of Engineering, American International University–Bangladesh).

[12] Sokolov, O., Iakovets, A., Andrusyshyn, V., & Trojanowska, J. (2024). Development of a Smart Material Resource Planning System in the Context of Warehouse 4.0. Eng, 5(4), 2588-2609.

[13] Guo, C., Qiu, S., Ni, T., Wang, B., & Liu, Q. (2023). Fast Phase Recognition of Mechanical Helical Phased Array Antenna Element Based on Line-Scan Machine Vision. IEEE Transactions on Instrumentation and Measurement.

[14] Islam, M. J., Islam, M. Z., & Dudaib, S. A. Fabrication of a prototype model of a solar-powered smart water level monitor system.

[15] Ahmed, O. (2023). Remote control and monitoring of a solar water pumping system using a cellular network for Sukkur Pakistan (Doctoral dissertation, Memorial University of Newfoundland).

[16] Paz, P. R. M. D. (2024). Enhancing Sustainable Recycling with Automated Identification of Colored Glass Bottles Using YOLO Object Detection (Doctoral dissertation, University of Applied Sciences Technikum Wien).

[17] Naz, S. A., Doeven, E. H., Adams, S., Kouzani, A., & Guijt, R. M. (2023). Closed-loop control systems for pumps used in portable analytical systems. Journal of Chromatography A, 1695, 463931.

[18] Silva, J., Coelho, P., Saraiva, L., Vaz, P., Martins, P., & López-Rivero, A. (2024). Validating the Use of Smart Glasses in Industrial Quality Control: A Case Study. Applied Sciences, 14(5), 1850.

[19] Hernandez-Lerena, G., Torres, E., Parham, C. F., & Leung, T. (2023). Portable Drinking Water Cooler and Dispenser.

[20] Kaur, R., & Singh, S. (2023). A comprehensive review of object detection with deep learning. Digital Signal Processing, 132, 103812.

[21] Chen, Y., Xu, H., Zhang, X., Gao, P., Xu, Z., & Huang, X. (2023). An object detection method for bayberry trees based on an improved YOLO algorithm. International journal of digital earth, 16(1), 781-805.

[22] Qiao, Y., Wu, M., Song, N., Ge, F., Yang, T., Wang, Y., & Chen, G. (2024). Automated pretreatment of environmental water samples and non-targeted intelligent screening of organic compounds based on machine experiments. Environment International, 109072.

[23] Odoi-Yorke, F. (2024). A systematic review and bibliometric analysis of electric cooking: evolution, emerging trends, and future research directions for sustainable development. Sustainable Energy Research, 11(1), 24.

[24] Mudryk, R., & Mitchenko, T. (2023). Shared automatic drinking water treatment and dispensing systems and methods of their optimization. Water and water purification technologies. Scientific and technical news, 35(1), 9-25.

[25] P.D.P. Adi, & Yuyu, W. (2022). Performance evaluation of ESP32 camera face recognition for various projects. Iota, 2(1), Internet of Things and Artificial Intelligence Journal. https://doi.org/10.31763/iota.v2i1.512.

[26] Meddeb, H., et.al. (2023). Development of surveillance robot based on face recognition using Raspberry-PI and IOT. Microprocessors and Microsystems. Volume 96, February 2023, 104728. Doi. 10.1016/j.micpro.2022.104728.

[27] Rakesh, M.D. et.al. (2025). Implementation of real time root crop leaf classification using CNN on raspberry-Pi microprocessor. Smart Agricultural Technology. Volume 10, March 2025, 100714. doi: 10.1016/j.atech.2024.100714.

[28] Mathe, S.E. (2024). A comprehensive review on applications of Raspberry Pi. Computer Science Review. Volume 52, May 2024, 100636. doi. 10.1016/j.cosrev.2024.100636.

[29] Fan, H. et.al. 2023. Raspberry Pi-based design of intelligent household classified garbage bin. Internet of Things. Volume 24, December 2023, 100987. doi. 10.1016/j.iot.2023.100987.

[30] Sajjad, M. Et.al. 2020. Raspberry Pi assisted face recognition framework for enhanced law-enforcement services in smart cities. Future Generation Computer Systems. Volume 108, July 2020, Pages 995-1007. doi. 10.1016/j.future.2017.11.013.

[31] Lu, J. et.al. 2024. Analyzing the structure-activity relationship of raspberry polysaccharides using interpretable artificial neural network model. International Journal of Biological Macromolecules. Volume 264, Part 1, April 2024, 130354. doi.10.1016/j.ijbiomac.2024.130354.

[32] Abas, et.al. 2023. A Raspberry Pi based blockchain application on IoT security. Expert Systems with Applications. Volume 229, Part A, 1 November 2023, 120486. doi. 10.1016/j.eswa.2023.120486.

[33] Anand, G. et.al. 2021. Object detection and position tracking in real time using Raspberry Pi. Materials Today: Proceedings. Volume 47, Part 11, 2021, Pages 3221-3226. doi.10.1016/j.matpr.2021.06.437.

[34] Ntizikira, E. et.al. 2025. Enhancing IoT security through emotion recognition and blockchain-driven intrusion prevention. Internet of Things. Volume 29, January 2025, 101442. doi.10.1016/j.iot.2024.101442.

[35] Yanan Sun, Y. et.al.2021. Pulse-spouted microwave freeze drying of raspberry: Control of moisture using ANN model aided by LF-NMR. Journal of Food Engineering. Volume 292, March 2021, 110354. doi.10.1016/j.jfoodeng.2020.110354.

[36] Tanque.,M, & Bradford, P. (2023). Chapter Six - Virtual Raspberry Pi-s with blockchain and cybersecurity applications. Advances in Computers. Volume 131, 2023, Pages 201-232. doi. 10.1016/bs.adcom.2023.04.005.

[37] Zhang, J. et.al. (2024). ISMSFuse: Multi-modal fusing recognition algorithm for rice bacterial blight disease adaptable in edge computing scenarios. Computers and Electronics in Agriculture. Volume 223, August 2024, 109089. doi.10.1016/j.compag.2024.109089.

[38] Abdellatif, M.M. et.al. (2023). A low cost IoT-based Arabic license plate recognition model for smart parking systems. Ain Shams Engineering Journal. Volume 14, Issue 6, June 2023, 102178. doi. 10.1016/j.asej.2023.102178.

[39] Maschler, B. (2020). et.al. Distributed Cooperative Deep Transfer Learning for Industrial Image Recognition. Procedia CIRP. Volume 93, 2020, Pages 437-442. doi. 10.1016/j.procir.2020.03.056.

[40] Kumar, S.V. et.al. 2020. Smart driver assistance system using raspberry pi and sensor networks. Microprocessors and Microsystems. Volume 79, November 2020, 103275. doi. 10.1016/j.micpro.2020.103275.

Machine Learning as a Tool to Combat Ransomware in Resource-Constrained Business Environment

Luis Jesús Romero Castro, Piero Alexander Cruz Aquino, Fidel Eugenio Garcia Rojas
Department of Computer Science, Universidad Peruana de Ciencias Aplicadas, Lima, Perú

***Abstract*—Ransomware has emerged as one of the leading cybersecurity threats to microenterprises, which often lack the technological and financial resources to implement advanced protection systems. This study proposes a cybersecurity model based on machine learning, designed not only for the detection and mitigation of ransomware attacks but also as a scalable and adaptable solution that can be integrated into business infrastructures across various sectors. By leveraging advanced techniques to identify malicious behavior patterns, the system alerts businesses before significant damage occurs. Moreover, this approach provides complementary measures such as automated updates and backups, enhancing resilience against cyber threats in resource-constrained environments. This research aims not only to protect critical data but also to contribute to the development of accessible cybersecurity models, improving operational continuity and promoting sustainability in the digital landscape.**

Keywords*—Ransomware; cybersecurity; machine learning; microenterprise; threat detection

I. Introduction

The exponential growth of ransomware attacks poses a serious threat to the cybersecurity of various organizations, including microenterprises, which often lack the resources to implement advanced protection systems. These attacks, characterized by their ability to lock or encrypt critical data in exchange for a ransom, have evolved in complexity and frequency, significantly impacting the operational and financial stability of these businesses.

Recent reports project that ransomware attacks occur every 11 seconds, with associated costs potentially reaching $20 billion by 2021. These alarming figures underscore the urgent need to develop advanced detection and prevention strategies to mitigate these devastating attacks in the microenterprise sector [1], which is especially vulnerable due to its limited cybersecurity investment capacity.

The analysis of leaked ransomware source codes, such as the detailed study of the Conti ransomware [9], has provided valuable insights into the attack techniques employed by these threats. However, despite progress, many studies have primarily focused on identifying and characterizing attacks without offering comprehensive solutions addressing all stages of ransomware defense. While machine learning algorithms and advanced detection techniques have proven effective in identifying threats [1][2], these approaches often lack proactive and adaptive measures to anticipate and neutralize new, evolving attack vectors. Additionally, the lack of integration between detection solutions and mitigation strategies limits their effectiveness in real-world microenterprise environments. It is, therefore, crucial to develop a cybersecurity model that not only detects but also prevents and dynamically responds to ransomware attacks, ensuring more robust and sustainable protection for these businesses.

This study proposes implementing a cybersecurity model based on machine learning techniques specifically designed to strengthen the cybersecurity posture of microenterprises against ransomware. This model will integrate early detection and neutralization strategies inspired by recent research [3][4], aiming to develop a proactive system capable of identifying and mitigating threats before causing significant damage.

In addition to exploring advanced detection approaches, complementary measures such as automated backups and regular software updates will be considered to establish a comprehensive defense against ransomware attacks [5], [7]. This proposal aims not only to protect the critical data of microenterprises but also to lay the foundation for a holistic cybersecurity approach capable of adapting and effectively responding to emerging threats.

This study will provide a practical roadmap for implementing an advanced cybersecurity system in microenterprises, contributing to the effective protection of critical data and systems against the growing threats of ransomware. By integrating machine learning-based approaches for ransomware detection and response, as proposed by previous studies [1], [3], [4], this research aims to significantly enhance the ability of these businesses to anticipate and neutralize such attacks. Moreover, complementary strategies, such as neutralizing ransomware techniques through format-preserving encryption [6] and implementing dynamic and static analysis to identify and mitigate new ransomware variants [9], will be considered. These integrated measures will enable microenterprises to not only effectively respond to current threats but also adapt to future challenges in the field of cybersecurity.

This study is organized as follows: Section II presents prior research related to the topic. Section III focuses on the contribution of the proposed cybersecurity model and its key components, while Section IV discusses the results and their implications. Discussion is given in Section V. Finally, Section VI provides conclusions and recommendations for future work.

II. Related Work

Various studies on ransomware management and cybersecurity have been identified, offering a wide range of approaches and solutions to address these threats. For instance, [1] focuses on applying advanced machine learning algorithms

for ransomware detection and mitigation, while [2] emphasizes the importance of dynamic analysis combined with machine learning for early detection. Additionally, other works such as [3], [4], [5], and [6] explore different approaches, ranging from decision tree-based detection to ransomware neutralization through encryption techniques. The detailed analysis of leaked Conti ransomware source code, as described in study [9], also provides crucial insights into the attack techniques employed. These studies represent just a sample of the available research but demonstrate the diversity of approaches that can be applied in the context of microenterprises to strengthen their defenses against ransomware and effectively protect their data.

Most previous works on ransomware detection through machine learning have focused on specific aspects, such as initial threat detection and static behavior analysis, leaving aside an integrated approach that combines early detection, proactive mitigation, and adaptability to resource-constrained environments. For instance, recent studies like the analysis of Conti ransomware have provided valuable insights into attack techniques but failed to address how these solutions can be tailored to microenterprises with limited infrastructure.

This study proposes a comprehensive model that bridges the existing gap by combining:

- Adaptability in Resource-Constrained Environments: Designing a lightweight architecture that does not rely on expensive hardware or advanced technical expertise.
- Early Detection and Mitigation: Leveraging advanced machine learning algorithms to identify malicious patterns and act before irreversible damage occurs.
- Dynamic Updating: Incorporating a continuous learning system based on static and dynamic analysis to tackle new ransomware variants.

These proposals not only fill the gap in designing and implementing effective solutions for microenterprises but also highlight critical areas for future research, such as improving machine learning model interpretability and optimizing performance under real-world operating conditions.

Source: [1]: Detection and prevention of ransomware, [2]: Enhancements in ransomware detection, [3]: Evaluation of ransomware detection, [4]: Dynamic ransomware detection, [5]: Detection based on decision tree algorithms, [6]: Ransomware detection and neutralization, [7]: Impact of cybersecurity research, [8]: Decision-making regarding ransomware payments, [9]: Threat case analysis.

In the field of cybersecurity and threat management, various studies and models have been developed to address the protection of critical data and the mitigation of risks associated with ransomware. These works provide a valuable framework for designing and implementing effective cybersecurity strategies in microenterprise environments. Table I shows related works on cybersecurity and ransomware detection.

TABLE I. RELATED WORKS ON CYBERSECURITY AND RANSOMWARE DETECTION

Method	Evaluation Method	Main Outcome	Source
Advanced Machine Learning	Evaluation with real data	Implementation of advanced machine learning algorithms to identify and mitigate ransomware threats	[1]
Machine Learning and Dynamic Analysis	Experimental evaluation	Improved early ransomware detection through dynamic analysis combined with machine learning	[2]
Machine Learning Evaluations	Algorithm comparison	Assessment of the effectiveness of various machine learning algorithms in ransomware detection	[3]
Dynamic Analysis and Machine Learning	Experimental evaluation	Use of dynamic analysis and machine learning to enhance real-time ransomware detection	[4]
Random Forest Algorithm	Accuracy comparison	Use of the Random Forest algorithm to improve accuracy in ransomware detection	[5]
Encryption and Machine Learning	Experimental evaluation	Development of ransomware neutralization techniques through format-preserving encryption	[6]
Research on Ransomware Impact	Impact analysis	Proposals to enhance the impact of cybersecurity research through collaboration and multidisciplinary approaches	[7]
Theoretical Models and Decision Analysis	Case analysis	Analysis of victims' payment decisions and their impact on the proliferation of ransomware	[8]
Analysis of Leaked Source Codes	Detailed case analysis	Gaining critical insights into ransomware attack techniques through analysis of leaked source codes	[9]
Cybersecurity Impact on SMEs	Case analysis	SMEs in the Balearic Islands lose an average of €30,000 due to cyberattacks	[10]
Ransomware Growth Analysis	Case analysis	Study of the 81% increase in ransomware attacks	[11]

Relevant studies include research focused on ransomware detection using machine learning algorithms [3]. This approach offers an important perspective on detection techniques based on behavior analysis and specific characteristics of ransomware attacks. Additionally, investigations explore dynamic ransomware detection methods through behavior analysis and machine learning [4], highlighting the importance of proactive approaches to addressing these threats.

In the realm of enterprise architecture, a detailed analysis of Conti ransomware is proposed, emphasizing the importance of understanding leaked source codes to develop effective countermeasures [9]. This study underscores the necessity of applying multidisciplinary approaches that combine behavioral and forensic analysis techniques to strengthen microenterprise security against ransomware.

Furthermore, these studies provide a comprehensive perspective on anti-ransomware research strategies, outlining a roadmap for enhancing the impact of research in this field [7]. This investigation highlights the importance of tackling ransomware from multiple dimensions, including risk assessment, technological innovation, and best practices in cybersecurity.

These studies represent only a glimpse of the broad spectrum of available research on ransomware management and cybersecurity. Analyzing and applying these approaches in the context of microenterprises will offer valuable insights to strengthen their defenses and effectively protect sensitive data.

III. CONTRIBUTION

The aim of this research is to propose a cybersecurity model based on machine learning techniques to address ransomware in Peruvian microenterprises, which face resource limitations in implementing advanced protection systems. This model seeks to provide robust and efficient protection against cyber threats by enabling early ransomware detection using advanced machine learning algorithms, as employed in previous studies [3][4].

Behavioral analysis is performed in real-time to identify suspicious patterns and anomalous behaviors, complemented by an automated system response that isolates suspicious files and notifies security personnel to mitigate risks immediately. Additionally, the model incorporates continuous updates based on detailed analyses of ransomware source codes [9], enabling the adaptation and enhancement of defenses against new attack variants.

Finally, the system generates detailed and periodic reports on detected threats and actions taken, providing a comprehensive and up-to-date view of the security status, as suggested in related studies [6], [7]. This contribution represents a significant advancement in the protection of critical data for microenterprises, establishing a robust and adaptable standard for cybersecurity in this sector (see Fig. 1).

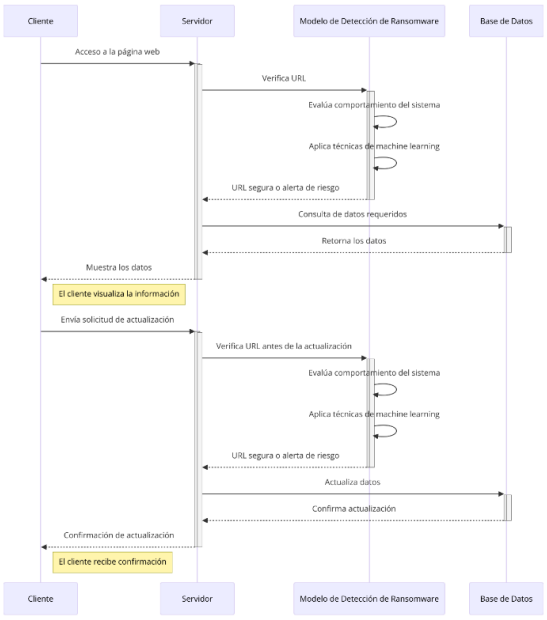


Fig. 1. Microenterprise client working within the company's web page.

A. Proposed Machine Learning Model for Ransomware

1) Continuous updates based on source codes: The objective of this strategy is to ensure that the model remains dynamic and adaptive, staying effective against new ransomware variants. To achieve this, source code mining is utilized, automating the extraction of unique features from new ransomware variants through both static and dynamic analysis. Static analysis identifies malware signatures by reviewing the source code, while dynamic analysis observes behavior in real-time by executing the ransomware in a controlled environment (sandbox). For instance, these techniques can be instrumental in detecting new file access paths or encryption methods used by modern variants such as LockBit or REvil.

Additionally, the model undergoes retraining using transfer learning. This approach allows the reuse of parts of the previously trained model, incorporating new patterns without losing acquired knowledge, thereby conserving computational resources by avoiding full retraining from scratch. Controlled simulations are also conducted using platforms like MITRE Caldera, where real attacks are replicated in secure environments. From these simulations, the model is adjusted to respond to observed tactics, such as data exfiltration or mass encryption. Synthetic data based on these attacks is also generated, reinforcing the model's training.

2) Model pipeline: Data collection will focus on capturing system logs that include access events, file modifications, network traffic, and security-related events tied to the applications used by microenterprises. These logs will be complemented with public and private datasets of known ransomware, such as WannaCry, Locky, and CryptoLocker, along with normal behavior data. During preprocessing, tasks such as data cleaning and normalization will be performed to remove duplicates, handle null values, and standardize formats. Additionally, feature engineering will be conducted to create key attributes, such as massive directory changes, unusual spikes in resource usage (CPU, network, or memory), and a high number of failed access attempts.

For model training, the data will be split into training (70%), validation (20%), and testing (10%) sets. The model will be validated using metrics such as precision, recall, F1-score, and a confusion matrix. These metrics will assess the model's performance, ensuring a balance between accurate detection and reduced false positives. Subsequently, the model will be deployed as a microservice integrated into the IT infrastructure of microenterprises, generating automatic alerts with details such as the affected file, detection time, and suggested actions.

The detection and response processes for ransomware in microenterprises were modeled and characterized, analyzing both their current state (AS-IS) and the desired future state (TO-BE). This analysis included threat identification, updating the machine learning model, and generating security reports. These steps ensure a clear and effective implementation of the proposed model, tailored to the specific needs of microenterprises and guaranteeing an efficient response to potential attacks (see Fig. 2 and Fig. 3).

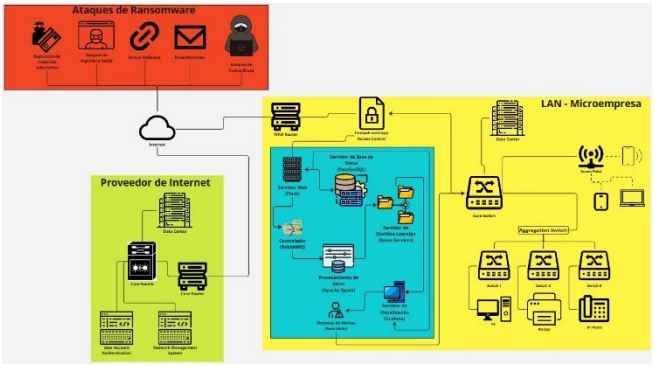


Fig. 2. Integrated architecture.

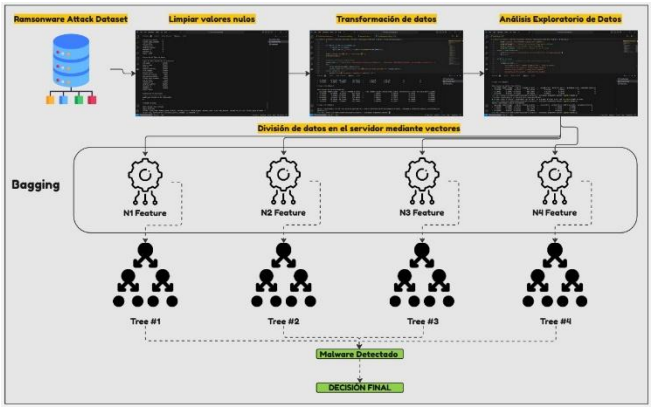


Fig. 3. Detection model.

3) Key features of the model: The designed model incorporates key features that ensure its effectiveness and adaptability to the dynamic environments of microenterprises.

a) Scalability: The model is designed to scale efficiently, enabling deployment across multiple microenterprises using shared infrastructure. Tools such as Apache Spark, which facilitates distributed data processing, and TensorFlow, which handles large data volumes without compromising performance, are employed. This ensures that the model can grow with the businesses' needs without losing effectiveness.

b) Efficiency: The model's hyperparameters, such as the number of trees in Random Forest algorithms or the kernel coefficient in Support Vector Machines (SVM), have been optimized. These optimizations minimize the occurrence of false positives and false negatives, increasing the accuracy of ransomware detection. This not only enhances the model's effectiveness but also strengthens user confidence in the implemented solution.

c) Portability: The model is designed to be highly portable, supporting deployment both on local servers and cloud solutions such as AWS, Azure, and Google Cloud. Furthermore, it is compatible with heterogeneous environments, eliminating the need for retraining when switching platforms. This guarantees flexible and adaptable implementation across various microenterprise infrastructures

4) Expected results: The proposed model aims to deliver tangible outcomes that significantly enhance microenterprises' ability to counter ransomware attacks, promoting effective and sustainable protection.

a) Reduction in response times: The model is expected to detect ransomware in real time, with response times measured in milliseconds. This capability will enable immediate actions, such as isolating malicious files or disconnecting affected devices from the network, minimizing the risk of propagation and additional damage.

b) Reduction in financial impact: The model will help reduce the financial impact associated with ransomware attacks by preventing the need for ransom payments and mitigating operational losses. These losses, currently estimated at millions of dollars annually for affected microenterprises, will be significantly reduced thanks to the system's preventive and reactive capabilities.

c) Increase in awareness: To foster greater user awareness, the model will provide clear and detailed visual reports. These reports will include information on the types of threats detected, the system's efficiency in protection tasks, and the automatic actions taken to mitigate risks. This transparency will help users better understand the system's functionality and its value in protecting against ransomware attacks.

B. System Requirements

This section outlines the components necessary for the development and implementation of the machine learning-based cybersecurity model for microenterprises. The requirements are divided into hardware and software (see Table II and Table III) to ensure a scalable and efficient infrastructure capable of supporting data processing and analysis demands.

TABLE II. HARDWARE REQUIREMENTS

Hardware	Description
Memoriy (RAM)	16 GB of RAM to handle large data volumes and perform complex analysis operations.
storage	500 GB SSD storage to ensure fast and efficient data access.
Processor	Intel Core i7 for optimal performance in data processing.
GPU	NVIDIA with CUDA support, required to accelerate the training of machine learning models.

Details: [1]: Programming Languages and Environments, [2]: Machine Learning Libraries and Frameworks, [3]: Databases and Storage, [4]: Infrastructure and Development Tools, [5]: Data Management and Visualization Tools.

TABLE III. SOFTWARE REQUIREMENTS

Details	Software	Description
[1]	Python, Jupyter Notebook	Development of the machine learning model.
[2]	Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch	Data manipulation and model development.
[3]	PostgreSQL, Redis, Apache Spark	Structured storage and distributed processing.
[4]	Docker, Flask, Visual Studio Code	Containerization and application development.
[5]	Elasticsearch, Kibana, Grafana, Tableau, Power BI	Visualization and monitoring of security data.

C. Model Implementation

The implementation follows a series of structured steps to ensure effectiveness in protecting the critical data of microenterprises:

Process (Table IV): [1]: Process Initiation, [2]: Data Retrieval, [3]: Command for Review, [4]: Data Processing, [5]: Machine Learning Data Analysis, [6]: Analysis Results, [7]: Visualization and Alerts, [8]: Real-Time Notifications, [9]: Process Completion.

TABLE IV. PROCESS DESCRIPTION FOR MICROENTERPRISE DETECTION SYSTEM

	Description
[1]	The worker accesses the interface through the Web Server (Flask) to interact with the system.
[2]	The Web Server (Flask) retrieves and stores data in the Database, ensuring that the data needed for analysis is properly retrieved and managed.
[3]	The Database sends a command to the Controller to initiate the processing of the retrieved data.
[4]	The Controller works with Apache Spark to extract and process the data, performing the necessary operations to handle large volumes of data.
[5]	The data processed by Apache Spark is sent to the Machine Learning (ML) module, where advanced algorithms analyze the data and generate results.
[6]	The results of the analysis are sent from the Machine Learning (ML) module to the Visualization and Alerts System.
[7]	The Visualization and Alerts System presents the results to the worker in real time and issues relevant notifications if necessary.
[8]	Real-time notifications are managed throughout the process to ensure that the worker is informed of any critical changes or events.
[9]	Once the tasks are completed and the results displayed, the system returns to the starting point to handle new requests.

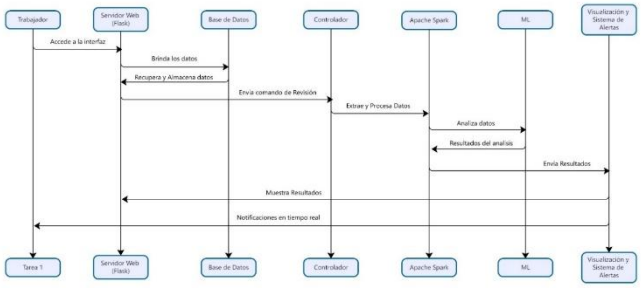


Fig. 4. Internal system usage process for the microenterprise.

IV. VALIDATION AND RESULTS

The validation of the proposed cybersecurity model is a critical stage to ensure its effectiveness and robustness in detecting and mitigating ransomware attacks in microenterprises. This process involves a series of tests and evaluations designed to verify the proper integration of system components, the accuracy of the machine learning model, and its performance under various operational conditions typical of microenterprises. Fig. 4 shows internal system usage process for the microenterprise.

Through integration testing, feature parameterization, load and performance testing, model training and evaluation, security testing, and continuous monitoring, the goal is to ensure that the system meets the security requirements of microenterprises and responds effectively to cyber threats. These validations enable the system to adapt to the infrastructure and resource limitations characteristic of microenterprises, ensuring efficient and effective implementation.

A. Validation Objective

The objective of this project is to develop an advanced ransomware detection and control model using machine learning algorithms and robust technologies to effectively protect microenterprises from cyber threats. This system will employ a comprehensive cybersecurity approach that not only detects malicious behaviors but also acts automatically to contain and mitigate the effects of attacks.

To achieve this objective, technologies such as Flask for web server development, RabbitMQ for message queue management, PostgreSQL as the database, Redis for caching, and Apache Spark for processing large volumes of data will be integrated. Additionally, machine learning models will be implemented using Scikit-learn, and data visualizations will be developed with Grafana to provide users with a clear and efficient interface.

The model validation will include a comprehensive process encompassing various types of tests to ensure the model's accuracy, efficiency, and security. These tests include:

- Unit Testing: Each system component will be verified to ensure the correct implementation of individual functions using tools like unittest in Python.
- Integration Testing: Different modules will be tested for correct interaction, validating integrations between services such as RabbitMQ and Apache Spark.
- Functional Testing: The system's compliance with functional requirements will be validated using tools like Selenium for user interface testing.
- Security Testing: Vulnerabilities will be identified, and system data and resources will be protected through vulnerability analysis and secure session management.
- Performance Testing: The system's capacity under various load conditions will be evaluated with Apache JMeter.
- Usability Testing: The system's ease of use by end users will be analyzed using heuristic evaluations and user testing.
- User Acceptance Testing (UAT): Scenarios reflecting real-world use cases will confirm that the system meets the end user's expectations.

Regarding the machine learning algorithms, a Support Vector Machine (SVM) model will be implemented, with data transformed and normalized to optimize performance. Multiple algorithms will be evaluated, and hyperparameters will be fine-tuned to enhance the system's accuracy in detecting ransomware. The MITRE Caldera framework will be employed to simulate cyberattacks, strengthening the system's response capabilities against real-world threats.

This comprehensive validation approach, combined with advanced algorithms, will ensure the system not only detects ransomware but also acts proactively to secure the operational continuity of microenterprises in a safe digital environment.

B. Evaluation Metrics for ML Model Performance

1) Area Under the ROC Curve (ROC AUC)

Definition: The Area Under the Receiver Operating Characteristic (ROC) Curve measures the model's ability to distinguish between positive and negative classes across different classification thresholds.

Formula: There is no closed-form formula to calculate the ROC AUC, as it is determined by integrating the ROC curve.

Interpretation:

- Value 0.5: The model has no discriminative ability, equivalent to random guessing.
- Value close to 1: The model has excellent ability to distinguish between classes.
- Value close to 0: The model misclassifies the classes entirely.

Observation: The ROC AUC is useful for evaluating models on imbalanced datasets, as it considers all possible classification thresholds and is unaffected by class distribution. However, it may be less informative when classes are extremely imbalanced, in which case metrics such as the Area Under the Precision-Recall Curve might be more appropriate.

2) Matthews Correlation Coefficient (MCC)

Definition: The Matthews Correlation Coefficient evaluates the quality of predictions in binary classification problems, considering true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

Formula:

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})(\mathrm{TP}\times\mathrm{TN})}}$$

Interpretation:

- Value 1: Perfect prediction.
- Value 0: Predictions are equivalent to random guessing.
- Value -1: Perfect inverse prediction (all predictions are incorrect).

Observation: MCC is a balanced metric suitable for datasets with imbalanced classes. It provides a more comprehensive view of model performance compared to Accuracy, as it takes all components of the confusion matrix into account.

3) Balanced Accuracy

Definition: Balanced Accuracy is the mean of the true positive rate (Recall) and the true negative rate (Specificity).

Formula:

$$Balanced\ Accuracy = \frac{(Recall + Specificity)}{2}$$

Where:

$$Specificity = \frac{TN}{(TN + FP)}$$

Interpretation: Balanced Accuracy provides a measure that accounts for the model's ability to correctly detect both positive and negative classes, thus balancing the impact of imbalanced classes.

Observation: This metric is valuable for imbalanced datasets, offering an evaluation that does not favor the majority class. It is especially relevant in applications like medical diagnosis or fraud detection, where both classes are equally critical.

4) Logarithmic Loss (Log loss)

Definition: Logarithmic Loss measures the uncertainty of the model's predictions based on the probabilities assigned to the correct classes.

Formula:

$$\boldsymbol{Log\ Loss} = -\frac{\mathbf{1}}{\boldsymbol{N}}\sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} = [\boldsymbol{y_i} log(pi) + (1 - y_i) log(1 - pi)]$$

Where:

- y_i is the true label (0 or 1).
- p_i is the predicted probability for the positive class.
- N is the number of instances.

Interpretation:

- Value 0: Perfect predictions.
- Higher values: Indicate greater errors in assigned probabilities.

Observation: Log Loss is particularly useful when class probabilities matter, such as in applications requiring confidence estimates for predictions. It penalizes incorrect but confident predictions more severely.

5) Cohen´s Kappa

Definition: Cohen's Kappa measures the agreement between the model's predictions and the actual observations, adjusting for agreement expected by chance.

Formula:

$$\kappa = \frac{po - pe}{1 - pe}$$

Where:

- po is the observed agreement proportion.
- pe is the expected agreement proportion by chance.

Interpretation:

- Value 1: Perfect agreement.
- Value 0: Agreement equal to random chance.
- Negative values: Agreement worse than random.

Observation: Cohen's Kappa is especially useful in multi-class classification tasks and in scenarios where accounting for random agreement is important. It is more robust than Accuracy in contexts where class imbalance might inflate perceived model performance.

6) Area Under the Precisión-Recall Curve (PR AUC)

Definition: PR AUC measures the trade-off between precision and recall across different classification thresholds, focusing on the positive class's performance.

Formula: Similar to ROC AUC, it does not have a closed formula and is computed by integrating the Precision-Recall curve.

Interpretation: A higher PR AUC indicates better model performance in terms of precision and recall. It is particularly useful for imbalanced datasets where the positive class is the primary focus of interest.

Observation: PR AUC is especially valuable in datasets with significant class imbalance, as it focuses on the performance of the minority (positive) class without being influenced by the majority class's abundance. It provides a more informative evaluation than ROC AUC in such scenarios. Table V shows evaluation metrics.

Metric: [1]: ROC AUC, [2]: MCC, [3]: Balanced Accuracy, [4]: Log Loss, [5]: Cohen´s Kappa, [6]: PR AUC.

TABLE V. EVALUATION METRICS

Metric	Ventajas	Desventajas
[1]	Considers all thresholds; useful for imbalanced datasets.	Less informative in cases of extreme class imbalance.
[2]	Considers all elements of the confusion matrix; balanced.	More complex to interpret compared to Accuracy.
[3]	Balances performance across classes; useful for imbalanced datasets.	May be less intuitive than traditional metrics.
[4]	Accounts for prediction confidence; penalizes incorrect predictions.	Harder to interpret compared to class-based metrics.
[5]	Adjusts agreement for chance; useful for multiple classes.	Can be less intuitive and more complex to calculate.
[6]	Focused on the positive class; useful for highly imbalanced datasets.	Does not account for the negative class, which may be a limitation.

C. Comparative Analysis of Ransomware Detection Algorithms

In this analysis, we compare three Machine Learning algorithms used to detect ransomware: Neural Networks, Random Forest, and Support Vector Machine (SVM). They will be evaluated in terms of accuracy, recall, F1-score, and performance in validation and test trials.

Algorithm: [A]: Accuracy (Training), [B]: Loss (Training), [C]: Accuracy (Validation), [D]: Loss (Validation).

TABLE VI. SUMMARY OF TRAINING AND VALIDATION

Algorithm	[A]	[B]	[C]	[D]
Neural Networks	99.94%	0.0033	100 %	0.0023
Random Forest	100%	-	100 %	-
(SVM)	100%	1.293939977335144e-05	100 %	1.293939977335144e-05

Table VI compares the accuracy and loss of the algorithms during training and validation. All models have shown perfect performance in validation. The low losses in Neural Networks and SVM indicate an excellent ability to minimize error during training.

Algorithm: [A]: Class, [B]: Precision (Validation), [C]: Recall (Validation), [D]: F1-Score (Validation), [E]: Support (Validation)

[A]: Class: 0: no ransonware, 1: ransonware

TABLE VII. VALIDATION RESULTS

Algorithm	[A]	[B]	[C]	[D]	[F]
Neural Networks	0	100%	100%	100%	1048326
	1	80%	80%	89%	1249
Random Forest	0	100%	100%	100%	209459
	1	100%	100%	100%	256
(SVM)	0	100%	100%	100%	209459
	1	100%	100%	100%	256

Table VII details the validation results for each class (No Ransomware and Ransomware) in terms of Precision, Recall, F1-Score, and Support. While both Random Forest and SVM achieved perfect results across all metrics, Neural Networks showed slightly lower precision in the Ransomware class, though they maintained excellent overall performance.

Algorithm: [A]: Class, [B]: True Positives (TP), [C]: False Positives (FP), [D]: False Negatives (FN), [E]: True Negatives (TN)

[A]: Class: 0: no ransonware, 1: ransonware

TABLE VIII. CONFUSION MATRIX (VALIDATION)

Algorithm	[A]	[B]	[C]	[D]	[E]
Neural Networks	0	1048326	0	0	1048326
	1	1249	0	159831	1249
Random Forest	0	209459	0	0	209459
	1	256	0	0	256
(SVM)	0	209459	0	0	209459
	1	256	0	0	256

This confusion matrix (Table VIII) illustrates each algorithm's ability to correctly classify ransomware and non-ransomware instances. Neural Networks exhibited some false negatives in the ransomware class, whereas Random Forest and

SVM achieved perfect classification with no false positives or false negatives.

Algorithm: [A]: Class, [B]: Precision (Test), [C]: Recall (Test), [D]: F1-Score, [E]: Support (Test). Table IX shows test results.

[A]: Class: 0: no ransonware, 1: ransonware

TABLE IX. TEST RESULTS

Algorithm	[A]	[B]	[C]	[D]	[E]
Neural Networks	0	100%	100%	100%	1048326
	1	80%	80%	89%	1249
Random Forest	0	100%	100%	100%	640000
	1	100%	100%	100%	100000
(SVM)	0	100%	100%	100%	799032
	1	100%	100%	100%	968

In the test results, both Random Forest and SVM maintained perfect precision and recall across both classes. Neural Networks, while highly accurate in the No Ransomware class, exhibited lower precision in the Ransomware class, suggesting potential challenges in detecting these instances.

Algorithm: [A]: Class, [B]: True Positives (TP), [C]: False Positives (FP), [D]: False Negatives (FN), [E]: True Negatives (TN)

[A]: Clase: 0: no ransonware, 1: ransonware

TABLE X. CONFUSION MATRIX (TEST)

Algorithm	[A]	[B]	[C]	[D]	[E]
Neural Networks	0	799032	0	0	799032
	1	968	0	0	968
Random Forest	0	639999	1	0	639999
	1	100000	0	0	100000
(SVM)	0	799032	0	0	799032
	1	968	0	0	968

The confusion matrix (Table X) for the test phase confirms the validation results, with both Random Forest and SVM demonstrating perfect classification. Neural Networks also showed no false positives or false negatives in the test phase, consistent with their high performance in prior validations.

V. DISCUSSION

The proposed ransomware detection model in this study relies on the analysis of both quantitative and qualitative data. Quantitative data includes statistics on the frequency and economic impact of ransomware attacks, while qualitative data encompasses descriptions of malicious behavior patterns and common technological limitations in microenterprises. These inputs informed the design of an approach tailored to the specific needs of this sector, based on principles of scalability, adaptability, and precision.

According to recent reports, ransomware attacks have increased by 81% over the past year, with estimated losses amounting to billions of dollars globally. This highlights the urgent need for proactive and accessible solutions to detect threats before they cause significant damage. The proposed model utilizes machine learning algorithms such as Random Forest and Support Vector Machine, which have demonstrated performance exceeding 90% in controlled environments. However, it is essential to consider the limitations of implementing these solutions in resource-constrained infrastructures.

The model's approach is mixed: quantitative in its ability to measure performance metrics such as accuracy, response time, and detection rates, and qualitative in evaluating ease of integration into microenterprises with non-technical staff. The goal is to provide a tool that is not only effective but also operationally and economically viable.

From a state-of-the-art perspective, most existing studies on ransomware detection focus on scenarios involving large corporations with advanced infrastructure. For example, research by Dobbertin and Leiva (2020) identified common attack patterns but did not address how to adapt these to microenterprises. This study bridges that gap by proposing a lightweight and adaptable model specifically designed for resource-constrained environments. Additionally, a dynamic updating system is incorporated based on the analysis of new ransomware variants, a feature frequently overlooked in prior research.

The goal of the model is to ensure the operational continuity of microenterprises, significantly reducing the financial and operational impact of cyberattacks. Complementary measures such as automated backups and an early warning system are included. In simulated tests, these features have shown promising results, achieving a 70% reduction in response time to simulated attacks.

This discussion highlights the integration of quantitative and qualitative data into the design of the proposed model, emphasizing its contribution to the state of the art in cybersecurity for microenterprises. Initial results are encouraging, but real-world testing is required to fully validate its effectiveness and adaptability. Additionally, the model presents opportunities for future research, such as exploring hybrid approaches that combine machine learning with emerging technologies like blockchain.

VI. CONCLUSION

The evaluation of cybersecurity technologies demonstrated that machine learning algorithms, such as Random Forest and Support Vector Machine, are effective in detecting ransomware, meeting the criteria of scalability and adaptability for resource-constrained microenterprises. The designed model, based on malicious behavior patterns and supported by an integrated architecture, achieved threat detection accuracy exceeding 90% in controlled environments. Controlled tests validated that the model and network structure are efficient, achieving rapid response times and precise ransomware detection in simulated scenarios.

However, it is important to acknowledge certain limitations and challenges encountered during the study. Although the results in controlled environments were promising, implementation in real-world scenarios may vary due to the diversity and complexity of microenterprise technological infrastructures. The model's adaptability to different configurations and its performance in operational conditions require further validation. Additionally, the performance of machine learning models heavily depends on the quality and quantity of available data; insufficient or poor-quality data can lead to ineffective models. The lack of interpretability in these models can cause distrust and hinder their adoption in certain contexts. Moreover, the cyber threat landscape is constantly evolving, with a significant increase in the frequency and sophistication of ransomware attacks. For instance, an 81% increase in ransomware attacks has been observed in the past year [10] implying that models must be continuously updated to remain effective. Finally, implementing machine learning-based solutions can be challenging for microenterprises due to a lack of technological and financial resources. Small and medium-sized enterprises (SMEs) have reported significant losses due to cyberattacks; for example, in the Balearic Islands, SMEs have lost an average of 30,000 euros due to cyberattacks [11].

Despite these limitations, the developed continuity plan includes clear protocols for incident management, automated backups, and regular model updates, ensuring operational resilience for microenterprises against cyber threats. It is recommended to conduct pilot tests in microenterprises across various sectors to assess the model's effectiveness and adaptability in operational conditions, implement a periodic model update process to incorporate new attack techniques and ensure its relevance against emerging threats, and develop training programs for microenterprise staff, focusing on identifying and responding to security incidents, thereby complementing the effectiveness of the technological model.

This study reaffirms the importance of investing in innovative and accessible technologies to safeguard critical data and improve the cybersecurity posture of microenterprises. While the proposed model has demonstrated effectiveness in controlled environments, future research should prioritize validating its performance in real-world scenarios. Pilot studies in microenterprises across various sectors are essential to assess its adaptability to conditions such as heterogeneous networks, limited technical expertise, and budget constraints. These tests will help identify necessary adjustments to maximize its effectiveness and usability.

Future work should also explore the integration of advanced machine learning techniques, such as recurrent neural networks (RNNs) and deep learning, to enhance the detection of complex malicious patterns and advanced ransomware variants designed to bypass traditional defenses. Additionally, the model's economic impact must be assessed by developing metrics to measure the cost-benefit ratio, balancing the savings from attack prevention against the implementation and maintenance costs.

Privacy and ethical considerations are crucial, given the need to access enterprise data for training and improving accuracy. Techniques like federated learning, which allows local model training without sharing sensitive data, could ensure data privacy. Furthermore, scalability and portability should be evaluated by testing the model's performance on cloud platforms to enable efficient implementation across diverse technological infrastructures, from local servers to distributed cloud environments.

The incorporation of emerging technologies, such as blockchain, represents another promising avenue. Blockchain could enhance the model by improving data integrity and traceability of security events, especially in managing real-time security alerts. These advancements, combined with continuous refinement to address emerging cyber threats, will ensure that microenterprises remain resilient in an evolving digital landscape. By focusing on these areas, future research can contribute to the development of more robust, efficient, and accessible solutions, advancing the field of cybersecurity for microenterprises.

REFERENCES

[1] Hammadeh, K. and Kavitha, M. (no date a) Unraveling ransomware: Detecting threats with Advanced Machine Learning Algorithms, International Journal of Advanced Computer Science and Applications (IJACSA). Recovered from: https://doi.org/10.14569/IJACSA.2023.0140952

[2] (2020) Two-stage ransomware detection using dynamic analysis and Machine Learning Techniques | Request PDF. Recovered from: https://doi.org/10.1007/s11277-020-07166-9

[3] Seong Il Bae, Gyu Bin Lee y Eul Gyu Im. (2020) Ransomware detection using machine learning algorithms. Concurrency and Computation Practice and Experience. Recovered from: https://www.scopus.com/record/display.uri?eid=2-s2.0-85068027073&origin=reflist

[4] Urooj, U, Al-rimy, B, Zainal, A, Ghaleb, F, Rassam, M. (2022). Ransomware Detection Using the Dynamic Analysis and Machine Learning: A Survey and Research Directions. Applied Sciences (Switzerland). Recovered from: https://doi.org/10.3390/app12010172

[5] Khammas, Ban Mohammed. (2020). Ransomware Detection using Random Forest Technique. ICT Express. Recovered from: https://doi.org/10.1016/j.icte.2020.11.001

[6] Lee, J. et al. (2023) Neutralization method of ransomware detection technology using format preserving encryption, MDPI. Recovered from: https://doi.org/10.3390/s23104728

[7] Jamie Pont, Osama Abu Oun, Calvin Brierley, Budi Arief, and Julio Hernandez-Castro. 2019. A roadmap for improving the impact of anti-ransomware research. In Nordic Conference on Secure IT Systems. Springer, 137–154. Timothy Mclntosh (2022) Ransomware mitigation in the modern Era: A Comprehensive Review, Research Challenges, and Future Directions. ACM Computing Surveys. Recovered from: https://doi.org/10.1145/3479393

[8] Cartwright, E., Hernandez Castro, H., Cartwright, A., 2019. To pay or not: game theoretical models of Ransomware. J. Cybersecur. 5 (1), 1–12. Alena Yuryna Connolly y Hervé Borrion (2022) Reducing Ransomware Crime: Analysis of Victims' Payment Decisions. Computers and Security. Recovered from: https://doi.org/10.1016/j.cose.2022.102760

[9] Saleh A., Yang X., Wei S. (2022) An Analysis of Conti Ransomware Leaked Source Codes. IEEE Acces. Recovered from: https://doi.org/10.1109/ACCESS.2022.3207757

[10] BTR Consulting. (2024, December 11). *81% increase in ransomware attacks in one year*. TyN Magazine. Retrieved from https://tynmagazine.com/aumento-del-81-de-ataques-ransomware-en-un-ano

[11] Tchernokojev, P. (2024, September 25). *Small and medium-sized enterprises in the Balearic Islands lose an average of 30,000 euros due to cyberattacks*. Cadena SER. Retrieved from https://cadenaser.com/baleares/2024/09/25/las-pymes-de-baleares-pierden-30000-euros-de-media-por-ciberataques-radio-mallorca

Traffic Speed Prediction Based on Spatial-Temporal Dynamic and Static Graph Convolutional Recurrent Network

YANG Wenxi, WANG Ziling, CUI Tao, LU Yudong, QU Zhijian
School of Computer Science and Technology, Shandong University of Technology, Country Zibo, China

***Abstract*—Traffic speed prediction based on spatial-temporal data plays an important role in intelligent transportation. The time-varying dynamic spatial relationship and complex spatial-temporal dependence are still important problems to be considered in traffic prediction. In response to existing problems, a Dynamic and Static Graph Convolutional Recurrent Network (DASGCRN) model for traffic speed prediction is proposed to capture the spatial-temporal correlation in the road network. DASGCRN consists of Spatial Correlation Extraction Module (SCEM), Dynamic Graph Construction Module (DGCM), Dynamic Graph Convolution Recurrent Module (DGCRM) and residual decomposition. Firstly, the improved traditional static adjacency matrix captures the relationship between each time step node. Secondly, the graph convolution captures the overall spatial information between the road networks, and the dynamic graph isomorphic network captures the hidden dynamic dependencies between adjacent time series. Thirdly, spatial-temporal correlation of traffic data is captured based on dynamic graph convolution and gated recurrent unit. Finally, the residual mechanism and the phased learning strategy are introduced to enhance the performance of DASGCRN. We conducted extensive experiments on two real-world traffic speed datasets, and the experimental results show that the performance of DASGCRN is significantly better than all baselines.**

***Keywords*—Intelligent transportation; traffic speed prediction; spatial-temporal correlation; dynamic graph; graph convolution recurrent network**

I. INTRODUCTION

Traffic prediction is a crucial component of intelligent transportation systems, aiming to forecast future traffic conditions based on historical observational data. The systems optimize the use of traffic resource and enhance efficiency [1]. Traditional univariate time series prediction methods, including Historical Average (HA) [2], Vector Auto-Regression (VAR) [3], Support Vector Regression (SVR) [4] and Auto-Regressive Integrated Moving Average (ARIMA) [5], mainly focus on time dependence. There methods ignore the spatial-temporal correlation between nodes, resulting in poor prediction accuracy. Spatial-temporal feature fusion models involve data modeling across spatial and temporal dimensions, and have been widely used to solve traffic prediction challenges due to their significant versatility [6, 7].

In recent years, spatial-temporal graph neural networks have attracted the attention of researchers in the field of traffic prediction because of their excellent performance. Spatial-temporal graph neural network is a typical application of Graph Convolution Network (GCN) in spatial-temporal domain. At present, most popular traffic prediction methods are based on GCN, which uses predefined graph structure to capture spatial features between nodes, and uses Convolutional Neural Networks (CNN) [8] or Recurrent Neural Networks (RNN) [9,10] to extract temporal features [11, 12]. However, these methods rely heavily on predefined static graph structures, which directly affect the predictive performance of the model. To address the limitations of static graph structures, researchers have proposed a data-driven approach to adaptively generate adjacency matrices [13]. These methods improve the performance of the model by learning the parameters of the adjacency matrix in the training process and then calculating the similarity between the embeddings of learnable nodes [14].

However, the traffic prediction data show strong dynamic spatial-temporal correlation. The adjacency matrix generated by static adaptation is difficult to capture the complex dynamic characteristics of the road network. Researchers are increasingly interested in modeling dynamic nonlinear spatial-temporal correlations inherent in traffic data. Recent literature on STG-NCDE [15] combines adaptive graphs with neural controlled differential equations to further improve the performance of the model. STGODE [16] captures the dynamic spatial-temporal correlation of traffic flow through tensor-based ordinary differential equations, and combines semantic adjacency matrices and time-extended convolutional structures to capture long-distance spatial-temporal correlation. ST-GDN [17] uses multi-resolution traffic transformation information and local-global regional dependencies for prediction. DSTAGNN [18] directly mines historical traffic flow data to extract spatial-temporal correlations, and effectively capture the dynamic attributes of spatial associations between nodes. Traffic flow probability graphs [19] employ reinforcement learning to generate dynamic graph for extracting spatial-temporal features. DAGCRN [20] captures the spatial-temporal dependencies in traffic data through dynamic adjacency matrix and graph convolutional recurrent network, combined with spatial-temporal relationship extraction, adjacency matrix update and global temporal attention module. MHSRN [21] captures the moving features between timestamps through a hybrid convolution module, and designs a spatially aware multi-attention module to capture global and local spatial-temporal features. DSTGRNN [22] captures the spatial-temporal dependence in traffic data through the dynamic graph convolution module and generator, combines spatial-temporal relationship extraction, node embedding and dynamic feature

update, and integrates dynamic and static features to improve the accuracy of traffic flow prediction.

Considering the complex spatial-temporal correlation of traffic data, great progress has been made in traffic forecasting. However, there are still some challenges to be solved in the integration of dynamic spatial-temporal features. Firstly, the predefined adjacency matrix and adaptive adjacency matrix, which reflect the static structure of the traffic network, are static and cannot capture the dynamic characteristics of the actual traffic network over time. Therefore, this highlights the need for more sophisticated modeling of dynamic features in traffic prediction topologies. Secondly, the static distance graph and the dynamic attribute graph provide different perspectives on the topology of the traffic network. Their effective integration can provide more comprehensive and accurate topological information, to better capture spatial dependencies. Finally, most traffic forecasting methods generally do not distinguish between normal and abnormal traffic conditions, despite the prevalence of unexpected events such as accidents and traffic control measures. Therefore, a more in-depth exploration of anomalous traffic patterns is essential for enhancing the accuracy of traffic predictions.

In response to the identified challenges, a traffic speed prediction model that integrates spatial-temporal dynamic-static graph convolutional recurrent networks (DASGCRN) is proposed. Firstly, the learning aid self-recurrent unit matrix and trainable parametric matrix are added to the traditional adjacency matrix by the Spatial Correlation Extraction Module (SCEM). The pre-defined adjacency matrix is reconstructed to enhance the expressiveness of the traffic graph. Secondly, based on SCEM, the Dynamic Graph Construction Module (DGCM) introduces a hypernetwork based on gate control mechanism and sparse connection layer to construct dynamic adjacency matrices. Through dynamic node feature propagation, DGCM improves the stability and effectiveness of graph structures in dynamic environments. In addition, a dynamic graph generation method is proposed to capture the correlation between nodes while considering the periodic and dynamic changes of the traffic network. Furthermore, a dynamic graph convolutional recurrent model (DGCRM) based on RNNs integrates static and dynamic graphs to capture the spatial-temporal dependencies in traffic networks. Finally, in order to further improve the prediction performance of DASGCRN, a residual mechanism and a phased learning strategy are introduced.

The rest of this paper is organized as follows. The traffic prediction problem is formulated in Section II. Motivated by the challenges, we introduce the details of our solutions in Section III. After that, we evaluate our model by two real-world traffic datasets and derive the parameter studies and experimental results in Section IV. Studying the ablation experiments in Section V. Finally, we conclude our paper in Section VI.

II. Problem Formulation

Based on the connectivity of the actual road network among sensors in the given datasets, a corresponding road network topology graph $G = (V, E, A)$ can be generated. $V = \{v_1, \dots, v_N\}$ represents the set of all nodes in the network and N denotes the total number of nodes, with each node representing a traffic sensor deployed at the roadside responsible for recording traffic information at its location. E denotes the set of edges that represents the spatial connectivity between nodes. The adjacency matrix $A \in R^{N \times N}$ is employed to depict the spatial adjacency relationships between nodes, where a_{ij} is an element of matrix A indicating the spatial connection status between nodes v_i and v_j. If $v_i, v_j \in V$ is connected to $(v_i, v_j) \in E$, then $a_{ij} = 1$; otherwise $a_{ij} = 0$. The relationship between road connectivity and the road network topology is illustrated in Fig. 1.

In this context, $X_t = [x_t^1, \dots, x_t^N]^T \in R^{N \times L \times C}$ denote the traffic speed at the N nodes of topology graph G at time t, L represent the total length of each time series, and C indicate the total number of node feature types. The traffic speed prediction problem is defined as shown in Eq. (1): Given the speed sequence $X_{1:P} = [X_1, \dots, X_P]^T \in R^{P \times N \times L}$ and topology graph G, the objective is to learn a mapping function F to predict the future sequence $\tilde{X}_{(P+1)(P+Q)} = \left[\tilde{X}_{(P+1)}, \dots, \tilde{X}_{(P+Q)}\right]^T \in R^{Q \times N \times L}$. Here, P represents the length of the historical speed data sequence, and Q denotes the length of the speed sequence to be predicted.

$$[X_{1:P}, G] \xrightarrow{F} \tilde{X}_{(P+1)(P+Q)} \tag{1}$$

III. The Proposed DASGCRN Model

A. Model Architecture

The architecture and various modules of the proposed DASGCRN model are illustrated in Fig. 2. The model consists of four main modules: Spatial Correlation Extraction Module (SCEM), Dynamic Graph Construction Module (DGCM), Dynamic Graph Construction Recurrent Module (DGCRM), Residual Decomposition and Model Training Strategy. SCEM models the spatial relationship between nodes based on static adjacency matrix. DGCM module is responsible for modeling the dynamic spatial-temporal relationship between nodes and edges, effectively capturing the potential dynamic correlation in the traffic network, and generating the dynamic adjacency matrix. The DGCRM module integrates dynamic and static adjacency matrices to model long-term dependencies between historical and future time steps by focusing on connections between different nodes in the transportation network, capturing time dependencies from a global perspective. In addition, residual traffic improves the training process of the model by adding skip connections between different layers of the network. Finally, piecewise learning training strategy is used to promote model convergence.

B. Spatial Correlation Extraction Module (SCEM)

1) In traffic road networks, traditional static adjacency matrices are typically used to represent the static connectivity relationships between nodes, as shown in Eq. (2). Where A_{v_i,v_j} denotes the edge weight between sensors v_i and v_j, d_{v_i,v_j} represents the road network distance from node v_i to v_j. Additionally, σ indicates the standard deviation of the distance, and κ denotes the sparsity threshold.

$$A_{v_i,v_j} = \begin{cases} exp\left(-\frac{d_{v_i,v_j}}{\sigma^2}\right), v_i \neq v_j, d_{v_i,v_j} \leq \kappa \\ 0, otherwise \end{cases} \tag{2}$$

2) The static adjacency matrix reflects the static connection state of the road. While the central road node is restricted to processing only the information from neighboring nodes, thus ignoring the data from distant nodes with similar traffic flow patterns [23]. The static adjacency matrix only considers the spatial distance factor, ignoring the influence of time factor on the relationship between nodes. In the real traffic road network, the spatial relationship between nodes is affected by time change [24], which limits the representation of dynamic spatial relationship in the traditional static adjacency matrix. Inspired by DAGCRN [20], the improved static adjacency matrix incorporates an identity matrix and a parametric matrix, as shown in Eq. (3).

$$\tilde{A} = A + A_{par} + I_N \tag{3}$$

3) where $\tilde{A}$ represents the improved static adjacency matrix, A denotes the traditional static adjacency matrix, and I_N is the identity matrix with diagonal elements equal to 1, indicates that each node is connected to itself (self-recurrent). This mechanism.

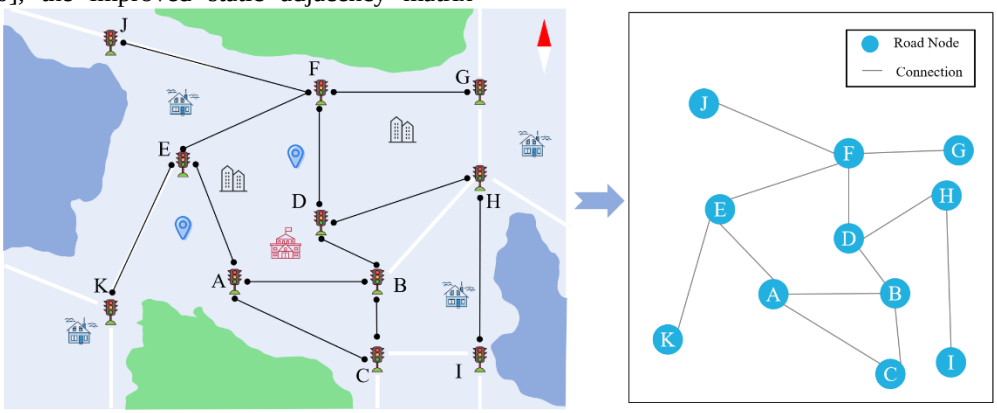


Fig. 1. The corresponding relationship between road connection and road network topology diagram.

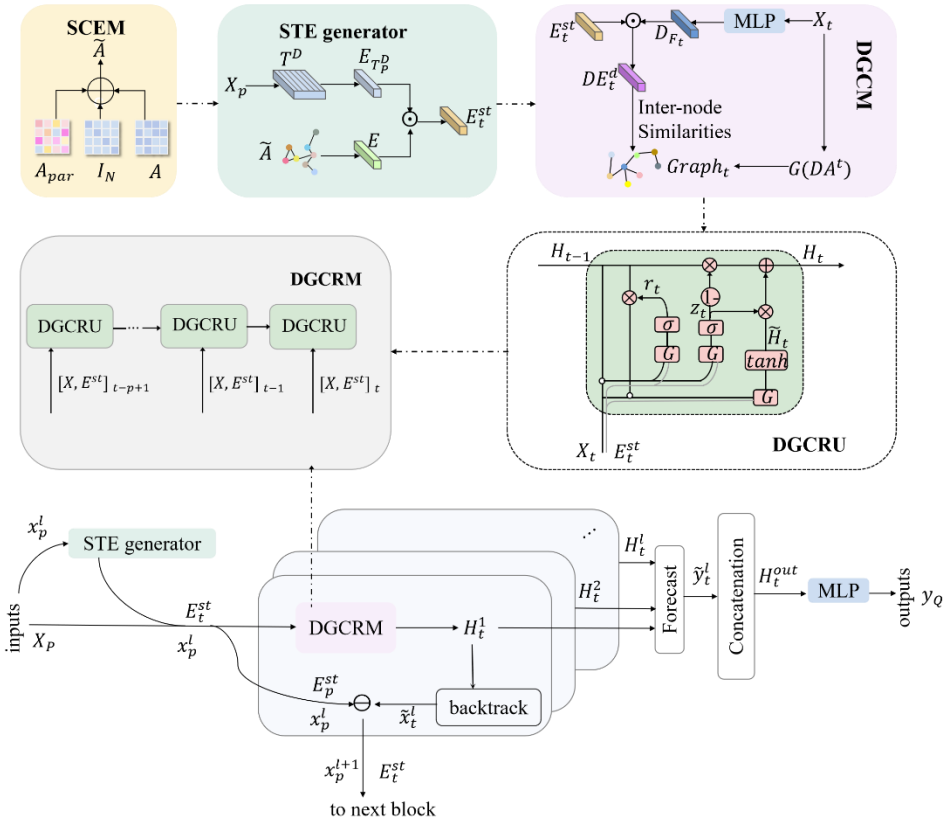


Fig. 2. The architecture and modules of DASGCRN.

4) It allows nodes to retain their own information while considering the impact of their state on neighboring nodes at each update step. The goal is to enhance the model's ability to capture long-term dependencies related to changes in traffic speed variations. A_{par} is the parametric matrix, initialized randomly and optimized through continuous iterations. As shown in Fig. 3, A_{par} enables the dynamic learning of the weight relationships between the central node (pink pentagon) and surrounding nodes (brown and beige circles). Allows adaptive adjustment of the connection based on these weights. This flexibility helps to explore the dynamic dependence of traffic speed over time in static network space. After A_{par} is optimized across all training samples, its combination with A and I_N forms $\tilde{A}$, enabling $\tilde{A}$ to more accurately represent the dynamic spatial relationships in the traffic network.

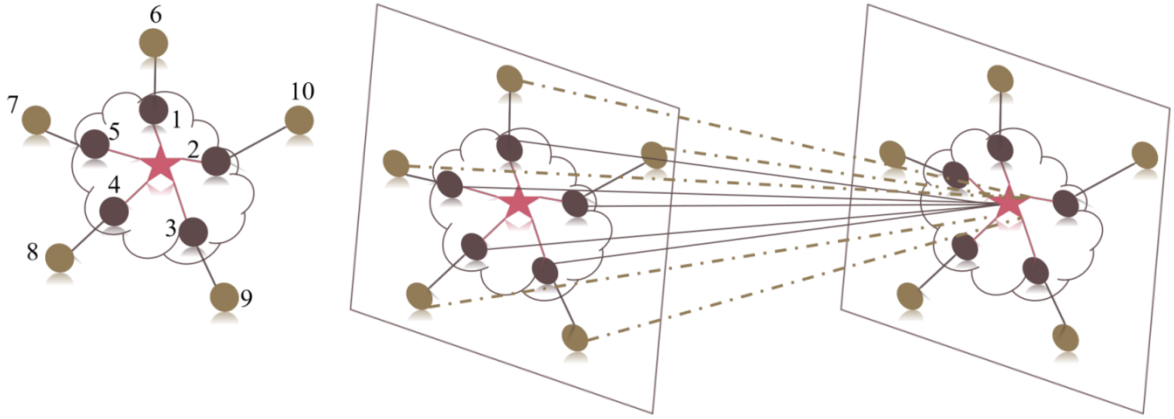


Fig. 3. Parametric matrix diagram.

C. Dynamic Graph Construction Module (DGCM)

The Dynamic Graph Construction Module (DGCM) has been carefully designed to efficiently capture spatial dependencies in traffic speed sequences, as well as temporal variations in these sequences. Specifically, the algorithm first builds a Spatial-Temporal Embedding generator (STE generator), which consists of two basic components: a spatial embedding generator and a temporal embedding generator. These components work together to extract and represent the spatial and temporal features present in the traffic speed sequence, thereby improving the overall understanding of its dynamics.

The core function of the spatial embedding generator is to learn the graph structure from input $\tilde{A}$, after which the learned node information is processed through two fully connected layers to produce the spatial embedding E. The primary task of the temporal embedding generator is to determine the daily encoding embedding $E_{T_P^D}$ that corresponds to the current traffic speed sequence X_P. Subsequently, $E_{T_P^D}$ and E are combined using an element-wise multiplication operation to obtain the new spatial-temporal embedding E_t^{st}. The implementation process of the spatial-temporal embedding is shown in Eq. (4).

$$E_t^{st} = E_{T_P^D} \odot E \tag{4}$$

where $E_{T_P^D} \in R^{P \times N \times D}$ is the daily embedding composed of consecutive time steps $P = [t - p + 1, \dots, t]$, and the p denotes the input sequence length. Additionally, $\odot$ signifies the element-wise multiplication operation. The input at the current time step x_t is processed through a multi-layer perceptron MLP to extract dynamic features, as shown in Eq. (5).

$$D_{F_t} = MLP(x_t) \tag{5}$$

where, $D_{F_t} \in R^{N \times D}$ represents the dynamic features of the current time step x_t after transformation. An element-wise multiplication operation is performed between D_{F_t} and E_t^{st}, and the resulting product is normalized to generate the dynamic graph embedding DE_t^d, as shown in Eq. (6).

$$DE_t^d = tanh\left(D_{F_t} \odot E_t^{st}\right) \tag{6}$$

The dynamic adjacency matrix DA^t at time t is calculated using the similarity between nodes, as shown in Eq. (7). This dynamic adjacency matrix expresses the evolving connectivity and changing trends among the network nodes. Furthermore, α is a hyperparameter used to control the saturation rate of the activation function.

$$DA^t = ReLU\left(tanh\left(\alpha\left(DE^d DE^{d^T}\right)\right)\right) \tag{7}$$

D. Dynamic Graph Convolutional Recurrent Module (DGCRM)

The Dynamic Graph Convolutional Recurrent Module (DGCRM) primarily integrates the improved static adjacency matrix $\tilde{A}$ with the generated dynamic adjacency matrix DA^t. This integration aggregates the information between the nodes of the traffic network and their neighbors, and effectively captures the dynamic spatial-temporal dependencies in the traffic network. Both $\tilde{A}$ and DA^t reflect the correlations among nodes from different perspectives. The DGCRM module effectively combines these matrices to provide a comprehensive view of the traffic road network for predictive modeling, accommodating the dynamic characteristics of the graph structure over time, as detailed in Eq. (8) to Eq. (12).

DGCRM utilizes X_P, $\tilde{A}$, and DA^t as inputs for the graph convolutional layer, and computes a weighted average to the outputs of the graph convolution.

$$\widetilde{H}_t^{(k)} = \alpha H_t^{in} + \beta H_t^{(k-1)} DA^t + \gamma H_t^{(k-1)} \tilde{A} \quad (8)$$

$$H_t^{(k)} = ReLU\left(\left(\mu H_t^{in} + (1-\mu)\widetilde{H}_t^{(k)}\right) W_P^{(k)}\right) \quad (9)$$

$$H_t^{out} = \sum_{i=0}^{K} H_t^{(k)} W^{(k)} \quad (10)$$

$$H_t^{(0)} = H_t^{in} \quad (11)$$

$$H_t^{out} = \Theta_{*G}\left(H_t^{in}, DA^t, \tilde{A}\right) \quad (12)$$

where α, β and γ are hyperparameters that adjust the weights of different components. $W_P^{(k)}$ denotes the k-th order parameter matrix. Hyperparameter μ is used to control the retention rate of the original node information in the process of information transmission, and deeper neighborhood exploration is carried out while preserving the local structure. H_t^{in} and H_t^{out} represent the input and output node states of the dynamic graph convolution at time step t, respectively. $W^{(k)} \in R^{k \times d_{in} \times d_{out}}$ is the learnable feature transformation matrix, and k denotes the model propagation depth. $H_t^{out} = \Theta_{*G}\left(H_t^{in}, DA^t, \tilde{A}\right)$ serves as a simplified representation of the dynamic graph convolution process, where Θ indicates dynamic graph generation, and $*G$ represents dynamic graph convolution. In the context of the DGCRU module diagram, G is used to denote dynamic graph convolution.

DGCRM is composed of Dynamic Graph Convolutional Gated Recurrent Unit (DGCRU), with the hidden state H_t^l of the final DGCRU serving as the output of the DGCRM. DGCRU is formulated by substituting the matrix multiplication operations in the GRU with dynamic graph convolution modules, as detailed in Eq. (13) to Eq. (16).

$$z_t = \sigma(\Theta_{*G}^z [X_t || H_{t-1}, E_t^{st}]) \quad (13)$$

$$r_t = \sigma(\Theta_{*G}^r [X_t || H_{t-1}, E_t^{st}]) \quad (14)$$

$$\widetilde{H}_t = tanh\left(\Theta_{*G}^h [X_t || r_t \odot H_{t-1}, E_t^{st}]\right) \quad (15)$$

$$H_t = (1 - z_t) \odot H_{t-1} + z_t \odot \widetilde{H}_t \quad (16)$$

where X_t and H_t represent the input and output at time step t, respectively. The symbol $\odot$ denotes the Hadamard product, while $||$ indicates concatenation operations. The function $\sigma(\cdot)$ refers to the sigmoid activation function. z_t and r_t are the update gate and reset gate at time t. Respectively, $\widetilde{H}_t$ represents the candidate state of the GRU unit. The symbol $*G$ signifies dynamic graph convolution, and Θ^z, Θ^h, and Θ^r correspond to the learnable parameters associated with the respective graph convolutions.

E. Residual Decomposition

To achieve multi-step predictions and sequence decomposition of traffic flow. An output sublayer consisting of linear layers is constructed after DGCRM. The mathematical expressions for the output sublayer are shown in Eq. (17) and Eq. (18).

$$\tilde{y}_t = Linear_{1:t}\left(H_t^l\right) \quad (17)$$

$$\tilde{x}_t = Linear_{1:t}\left(H_t^l\right) \quad (18)$$

where the predicted output $\tilde{y}_t^l \in R^{Q \times N \times C}$ denotes the prediction of y_t by the l-th block based on x_p^l, while the backward predicted output $\tilde{x}_t^l \in R^{P \times N \times C}$ represents the reverse prediction of x_p^l by the l-th block. The final output of the prediction model is given in Eq. (19) and Eq. (20).

$$y_Q = \sum_1^l \tilde{y}_t^l \quad (19)$$

$$x_p^{l+1} = x_p^l - \tilde{x}_t^l \quad (20)$$

Backward prediction can be understood as the decomposition process of traffic speed sequence. The aim is to eliminate the part of the information that the model has learned, while retaining the unlearned part for further processing by reconstructing the velocity sequence. Specifically, $\tilde{x}_t^l$ can be viewed as the reconstruction of the input speed sequence x_p^l, which incorporates the information that H_t^l has learned from x_p^l. By utilizing a residual decomposition mechanism, the learned information from x_p^l is removed, and the unlearned components of x_p^{l+1} are preserved for modeling in the subsequent block. The outputs from each block are then summed to produce the final prediction result.

F. Model Training

In order to improve the training efficiency of the model, a phased learning strategy is designed. In the initial phase of training, instead of training all blocks at the same time, only the first few blocks are trained. As the number of training layers increases, additional blocks are gradually incorporated into the training process. This approach significantly reduces the time and memory overhead required in the early stages of training. The training process is illustrated in Algorithm 1.

Algorithm 1: Training Algorithm of the DASGCRN Model.

Input: The static graph $G = (V, E, A)$, traffic speed sequence $X_p \in R^{P \times N \times D}$, current time step $X_t \in R^{T \times N \times D}$, time embedding T^D, spatial embedding E, training epochs $epochs$;

```
set turn = 1, i = 1
repeat
    initialize hidden state H_t^0, randomly select a batch (input X ∈ R^{P×N×D}, output Y ∈ R^{Q×N×D}, time of day T ∈ R^{Q×N×D}) from X_p;
    if turn%s == 0 and i < K then
        i = i + 1
    end if
        for p in 0,1,...,P − 1 do
            Calculate ỹ_t^l and x̃_t^l according to Eq.16, Eq.17 ;
    end for
        sum ỹ_Q = Σ_1^l ỹ_t^l
        Compute loss L = loss(ỹ_Q, y_Q)
        Back propagation and update parameters according to L;
        turn = turn + 1;
    Until the model reaches a stable state;
```

Output: the leaned DASGCRN model.

IV. EXPERIMENTS

A. Datasets

The experiment used two real-world traffic history datasets to verify the performance of different prediction models, as detailed in Table I.

- Los-loop Dataset

The dataset, from the City of Los Angeles, records speed information collected by loop detectors from March 1-7, 2012. It includes 207 road nodes and 1,313 road connections, collecting speed data every five minutes.

- SZ-taxi Dataset.
- The dataset comes from taxi tracks in Shenzhen City from January 1-31, 2015, covering 156 major roads in Luohu District, including 266 road connection relationships, with speed data collected every 15 minutes.

TABLE I. DETAILED INFORMATION OF TRAFFIC DATASETS FOR EXPERIMENT

Dataset	Number of Sensors	Edge	Time Interval	Time Range
SZ-taxi	156	266	15min	2015.1.1-2015.1.31
Los-loop	207	1313	5min	2012.3.1-2012.3.7

B. Parameter Settings

According existing methods, we split the datasets into training and testing sets in the same way as the baseline, i.e. 8:2 on the Los-loop and SZ-taxi datasets. Historical traffic speed data from the past hour is used to predict traffic speed in the future hour. The time step for the Los-loop dataset is set to 12 (sampling for 5 minutes), while the time step for the SZ-taxi dataset is set to 4 (sampling for 15 minutes). The experiment was repeated for 5 times and the average value of the evaluation index was reported.

The model employed the Adam optimizer with a learning rate of 0.003, 64 hidden units, an embedding dimension of 10, a batch size of 64, and an epoch value of 300. Mean Absolute Error (MAE) was used as the training loss function, along with an early stopping strategy to prevent overfitting.

C. Baseline Methods

HA: The Historical Average method (HA) predicts future speeds by analyzing the average values of historical speed data.

SVR: Support Vector Regression (SVR) is one of the classical time series analysis models that employs linear support vector machines to perform regression tasks.

Graph Convolutional Network (GCN): GCNs are capable of considering the spatial structural characteristics of graphs, extending convolution operations to graph structures.

Gated Recurrent Unit (GRU): GRUs can capture significant data dependencies over large gaps in time series.

T-GCN [11]: The T-GCN model simultaneously captures spatial and temporal dependencies using GCN and GRU.

A3T-GCN [25]: The A3T-GCN model introduces an attention mechanism to the T-GCN framework to extract weight information for each time step.

NA-DGRU [26]: The NA-DGRU model utilizes two GRUs to extract correlations between speed and time from both the original features and aggregated neighborhood features.

MTA-CN [27]: The MTA-CN model transforms long time single feature historical data into short time multi-feature data and incorporates a two-stage attention mechanism to capture the significance of features in different time periods and time steps.

FD-TGCN [28]: The spatial module introduces a novel Dynamic Convolution Matrix (DCM) to learn the characteristics of dynamic road structures, while the temporal module employs a Fast Temporal Convolution Network (FTCN) to model long-term temporal relationships.

D. Experimental Evaluation Metrics

The experiment employs five commonly used evaluation metrics to assess the performance of various prediction methods:

1) Mean Absolute Error (MAE): This index intuitively reflects the actual forecast error, as shown in Eq. (20).

2) Root Mean Squared Error (RMSE): RMSE is sensitive to outliers and is often used as a standard to measure the predictive performance of deep learning models, as shown in Eq. (21).

3) Accuracy (ACC): This measure describes the degree of fit between the predicted value and the true value; The closer the value is to 1, the better the prediction performance, as shown in Eq. (22).

4) Coefficient of Determination (R2): This statistic evaluates the goodness of fit of regression models; higher values indicate better predictive accuracy, as defined in Eq. (23).

5) Explained Variance Score (var): This score measures the model's ability to explain the variance in the data; values closer to 1 indicate a higher explanatory power of the model, as defined in Eq. (24).

$$MAE = \frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}\left|y_{ij} - \tilde{y}_{ij}\right| \tag{20}$$

$$RMSE = \sqrt{\frac{1}{MN}\sum_{j=1}^{M}\sum_{i=1}^{N}\left(y_{ij} - \tilde{y}_{ij}\right)^2} \tag{21}$$

$$ACC = 1 - \frac{Y - \tilde{Y}_F}{Y_F} \tag{22}$$

$$var = 1 - \frac{Var\{Y - \tilde{Y}\}}{Var\{Y\}} \tag{23}$$

$$R^2 = 1 - \frac{\sum_{j=1}^{M}\sum_{i=1}^{N}\left(y_{ij} - \tilde{y}_{ij}\right)^2}{\sum_{j=1}^{M}\sum_{i=1}^{N}\left(y_{ij} - \bar{Y}\right)^2} \tag{24}$$

In the above equations, y_{ij} and $\tilde{y}_{ij}$ represent the actual and predicted traffic speeds on road i-th at time t, respectively. M denotes the sample size of the traffic series, while N represents the set of roads. Y and $\tilde{Y}$ denote the sets of y_{ij} and $\tilde{y}_{ij}$, with $\bar{Y}$ being the average of Y. Both MAE and RMSE describe error

values, where smaller values indicate better model performance. On the contrary, the accuracy reflects the correctness of the prediction, and the higher the value, the better the prediction effect.

E. Experimental Results

Table II presents a comparison of the DASGCRN model with nine baseline models on the Los-loop and SZ-taxi datasets. The metrics underlined in the table indicate the best results, while the models in bold italic type denote those whose prediction results are directly referenced from the published papers. A "/" indicates that the baseline model did not provide that metric at the time of publication. It is evident from the experimental results that the proposed DASGCRN model outperforms all other models on all evaluation indicators of the two datasets. Statistical methods and traditional machine learning models have high requirements for data stationarity and usually focus only on temporal correlation. This leads to challenges in meeting these requirements for traffic data, resulting in poor predictive performance from traditional methods. In contrast, compared with GCN and GRU, the DASGCRN model improved on all evaluation measures. This shows that the model effectively captures the spatial topological features of urban road network and the temporal changes of traffic state. In addition, other baseline models also achieve better predictive performance compared to traditional methods, highlighting the strong spatial-temporal correlations in traffic data. Capturing these spatial-temporal features enhances traffic prediction accuracy.

TABLE II. COMPARISON DASGCRN AND BASELINE MODELS ON LOS-LOOP AND SZ-TAXI

T	**Datasets**	**Los-loop**									
	Methods	HA	SVR	GCN	GRU	***TGCN***	***A3T-GCN***	***NA-DGRU***	***MTA-GN***	***FD-TGCN***	DSAGCRN
15min	MAE	4.0162	3.7285	5.3525	3.0602	3.1802	3.1365	3.0281	3.1004	3.083	2.7888
	RMSE	7.5323	6.0084	7.7922	5.2182	5.1264	5.0904	5.1348	5.1058	5.133	5.0462
	ACC	0.8715	0.8977	0.8673	0.9109	0.9127	0.9133	0.9126	/	0.9119	0.9140
	R2	0.7083	0.8123	0.6843	0.8576	0.8634	0.8653	/	0.8670	/	0.8675
	Var	0.7084	0.8146	0.6844	0.8577	0.8634	0.8653	/	0.8679	/	0.8682
30min	MAE	4.4136	3.7248	5.6118	3.6505	3.7466	3.6610	3.6692	3.6041	3.712	3.0812
	RMSE	8.3204	6.9588	8.3353	6.2802	6.0598	5.9974	6.1358	5.8462	5.904	5.8436
	ACC	0.8581	0.8815	0.8581	0.8931	0.8968	0.8979	0.8955	/	0.8960	0.9004
	R2	0.6408	0.7492	0.6402	0.7957	0.8098	0.8173	/	0.8148	/	0.8205
	Var	0.6409	0.7523	0.6404	0.7958	0.8100	0.8173	/	0.8148	/	0.8220
45min	MAE	4.7898	4.1288	5.9534	4.0915	4.1158	4.1712	4.0567	3.9483	/	3.2735
	RMSE	9.0213	7.7504	8.8036	7.0343	6.7065	6.6840	6.7604	6.5731	/	6.3597
	ACC	0.8462	0.8680	0.8500	0.8801	0.8857	0.8861	0.8851	/	/	0.8916
	R2	0.5783	0.6899	0.5999	0.7446	0.7679	0.7694	/	0.7726	/	0.7869
	Var	0.5783	0.6947	0.6001	0.7451	0.7684	0.7705	/	0.7732	/	0.7894
60min	MAE	5.1504	4.5036	6.2892	4.5186	4.6021	4.2343	4.4256	4.0154	4.535	3.4260
	RMSE	9.6602	8.4388	9.2657	7.6621	7.2677	7.0990	7.2776	6.8749	6.983	6.7397
	ACC	0.8354	0.8562	0.8421	0.8694	0.8762	0.8790	0.8764	/	0.8760	0.8852
	R2	0.5070	0.6336	0.5583	0.6980	0.7283	0.7407	/	0.7492	/	0.7605
	Var	0.5070	0.5593	0.5593	0.6984	0.7290	0.7415	/	0.7495	/	0.7637
T	Datasets	SZ-taxi									
	Metric	HA	SVR	GCN	GRU	***TGCN***	***A3TGCN***	***NA-DGRU***	***MTA-CN***	***FD-TGCN***	DSAGCRN
15min	MAE	2.7842	2.6233	4.2367	2.5955	2.7145	2.6840	2.7387	2.6105	2.667	2.4964
	RMSE	4.2991	4.1455	5.6596	3.9994	3.9825	3.9989	4.0587	4.0440	4.036	3.9716
	ACC	0.7005	0.7012	0.6107	0.7149	0.7195	0.7218	0.7173	/	0.7187	0.7233
	R2	0.8305	0.8423	0.6654	0.8329	0.8539	0.8512	/	0.8526	/	0.8554
	Var	0.8305	0.8424	0.6655	0.8329	0.8539	0.8512	/	0.8530	/	0.8559
30min	MAE	2.8191	2.6875	4.2647	2.6906	2.7522	2.7038	2.7280	2.6158	2.778	2.5064
	RMSE	4.3508	4.1628	5.6918	4.0942	4.0317	4.1749	4.0683	4.0684	4.083	3.9899
	ACC	0.6969	0.7100	0.6085	0.7184	0.7167	0.7202	0.7166	/	0.7154	0.7221

	R2	0.8264	0.8410	0.6616	0.8249	0.8451	0.8493	/	0.8507	/	0.8541
	Var	0.8264	0.8413	0.6617	0.8250	0.8451	0.8493	/	0.8512	/	0.8546
45min	MAE	2.8488	2.7359	4.2844	2.7743	2.7645	2.7261	2.7393	2.6954	/	2.5157
	RMSE	4.3916	4.1885	5.7142	4.1534	4.0910	4.2461	4.0777	4.1168	/	4.0034
	ACC	0.6941	0.7082	0.6069	0.7143	0.7155	0.7186	0.7159	/	/	0.7211
	R2	0.8231	0.8391	0.6589	0.8198	0.8436	0.8474	/	0.8412	/	0.8531
	Var	0.8231	0.8397	0.6590	0.8199	0.8436	0.8474	/	0.8415	/	0.8536
60min	MAE	2.8754	2.7751	4.3034	2.7712	2.7860	2.7391	2.7487	2.6396	2.762	2.5254
	RMSE	4.4302	4.2156	5.7361	4.0747	4.1299	4.2707	4.0851	4.1637	4.104	4.0167
	ACC	0.6914	0.7063	0.6054	0.7197	0.7142	0.7169	0.7154	/	0.7141	0.7202
	R2	0.8199	0.8370	0.6564	0.8266	0.8421	0.8454	/	0.8434	/	0.8521
	Var	0.8199	0.8379	0.6564	0.8267	0.8421	0.8454	/	0.8440	/	0.8526

The experiment further divided the data of one day from the Los-loop and SZ-taxi test sets. It drew the prediction curves of DASGCRN and T-GCN models within the interval of 5 minutes and 60 minutes, and compared them with the Ground Truth, as shown in Fig. 4. As can be seen from the dotted box in Fig. 4, the DASGCRN model is more sensitive to capturing data with sudden increases or decreases in speed than the T-GCN, so the predictions are more accurate. This improved performance is attributed to the integration of time series modeling in DASGCRN's dynamic graph structure, which enables the model to focus on specific patterns of dynamic change, resulting in faster and more accurate predictions. As highlighted by the pink dotted boxes in Fig. 4(c) and Fig. 4(d), DASGCRN showed superior performance in both short (5 minutes) and long (60 minutes) predictions.

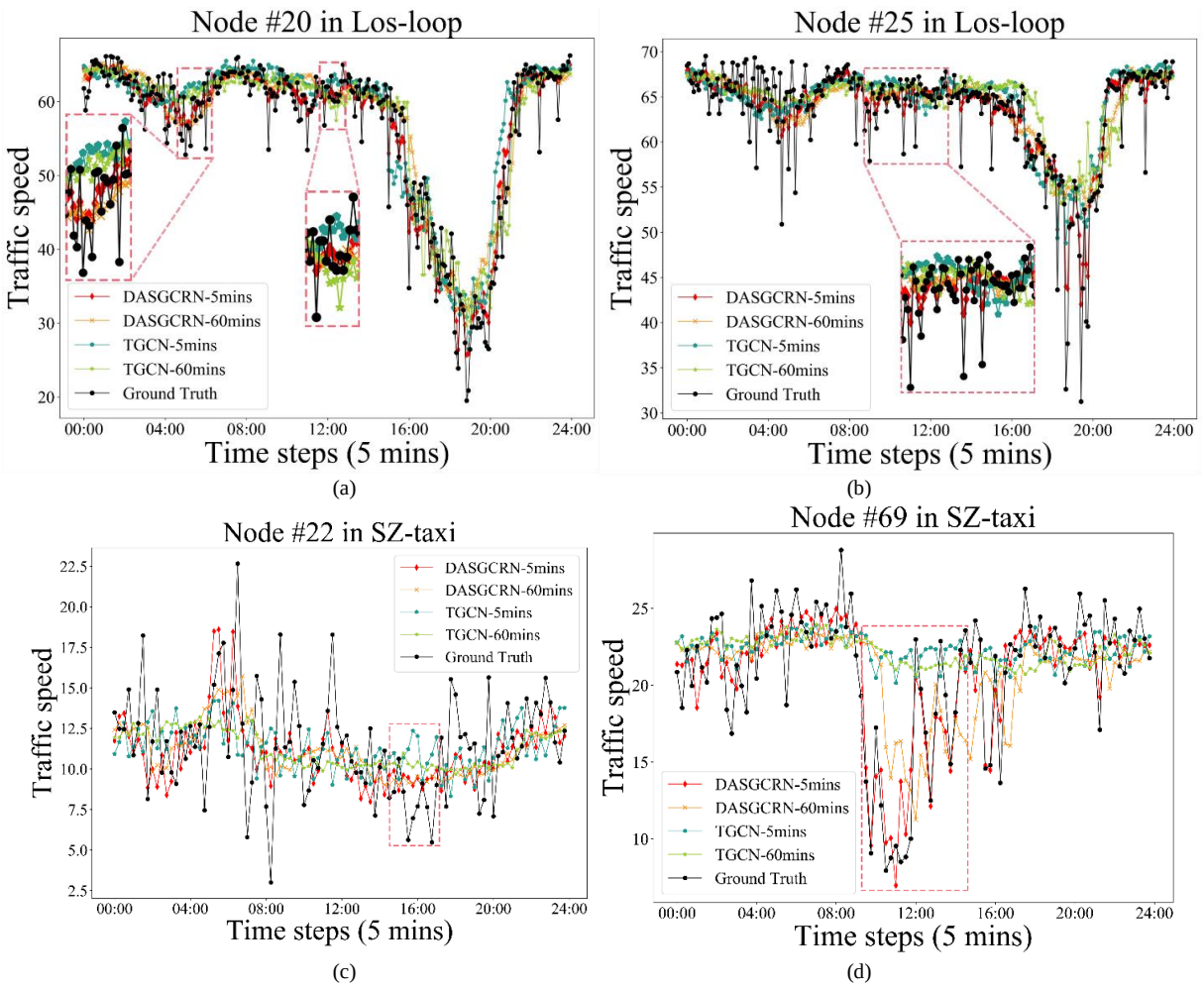


Fig. 4. The prediction curves of two datasets.

Fig. 5 illustrates the absolute error heatmaps of predicted values versus actual values for the DASGCRN model across two datasets. Due to the large size of the dataset, we selected only the first 60 time steps of the 12 roads in each dataset and generated heatmaps for each dataset with four different prediction ranges. Generally, the prediction performance of the model tends to decline with the increase of the prediction range. However, the heatmap shows that DASGCRN has maintained good performance across both the short-term and long-term forecast ranges. This is due to the dynamic graph convolution module used in the DASGCRN model, which replace the matrix multiplication in GRU. It can allow for more flexible control over feature information transmission and enhance the model's ability to capture long-term dependencies.

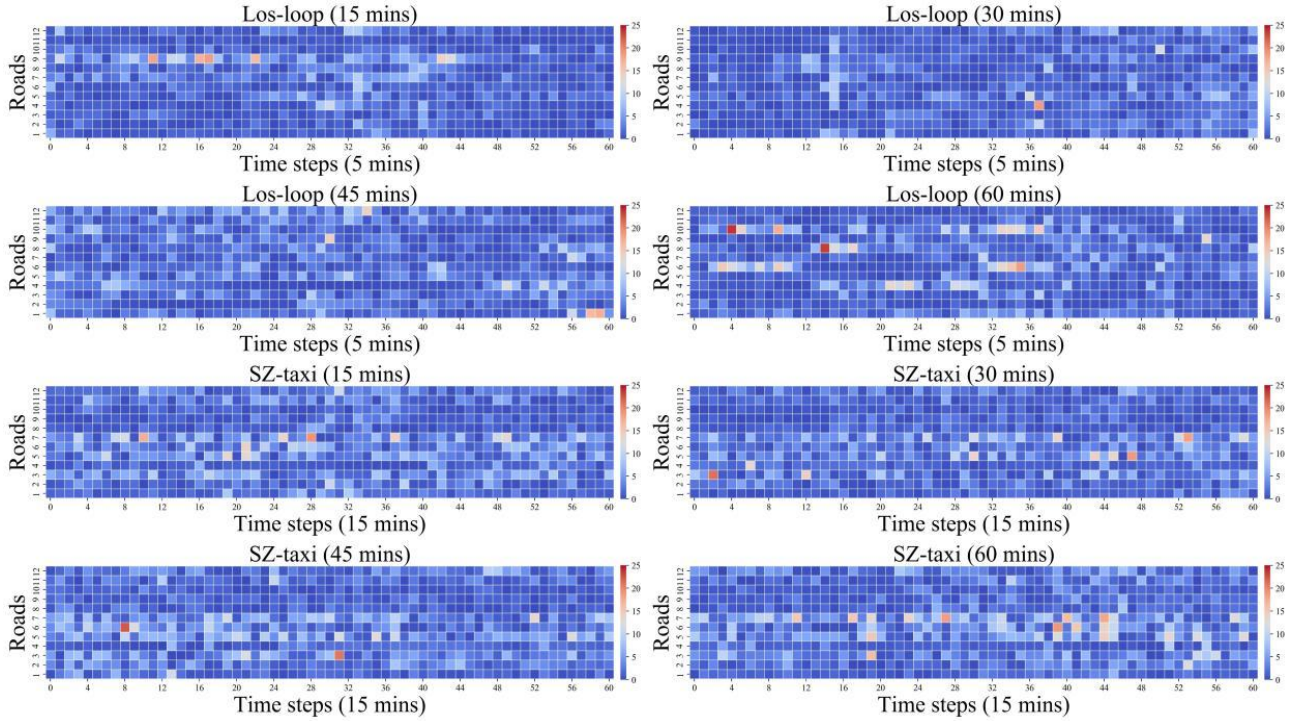


Fig. 5. The heatmaps of different forecasting granularity of DASGCRN two datasets.

V. Ablation Study

To validate the effectiveness of key components in the DASGCRN model, nine variants of the DASGCRN model were designed:

w/o I_N: DASGCRN without the self-loop adjacency matrix.

w/o A_{par}: DASGCRN without the learnable parameter matrix.

w/o DA_t: DASGCRN without the dynamic adjacency matrix.

w/o FCL: Replacing dynamic graph convolution with a simple fully connected layer in DASGCRN.

w/o T^D: DASGCRN without temporal embeddings, using only spatial embeddings to generate dynamic embeddings.

w/o DA_tI_N: DASGCRN without dynamic graph convolution and the self-loop adjacency matrix.

w/o DA_tA_{par}: DASGCRN without dynamic graph convolution and the learnable parameter matrix.

w/o Mix: DASGCRN without residual connections in dynamic graph convolution, using the last hop's output as the output of dynamic graph convolution.

w/o TS: DASGCRN without segmented training.

TABLE III. Comparison of Ablation Results of DASGCRN and Variant on Los-loop Dataset

T	Datasets	Los-loop									
	Methods	DASGCRN	w/o I_N	w/o A_{par}	w/o DA_t	w/o FCL	w/o T^D	w/o DA_tI_N	w/o DA_tA_{par}	w/o Mix	w/o TS
15min	MAE	2.7888	3.6346	3.6731	9.1303	7.7888	3.4029	11.5724	13.6317	3.4096	2.9895
	RMSE	5.0462	6.3577	6.0787	13.7764	11.0462	5.5153	15.6325	17.0738	5.4673	5.3043
	ACC	0.8715	0.8977	0.8673	0.9109	0.9127	0.9133	0.9126	/	0.9119	0.9140
	R2	0.8675	0.7937	0.8106	0.0516	0.8075	0.8430	0.0427	0.0572	0.8452	0.8541
	Var	0.8682	0.7992	0.8178	0.0529	0.8082	0.8462	0.0429	0.0573	0.8453	0.8553

30min	MAE	3.0812	3.8443	3.9975	9.1325	8.0812	3.7666	11.9750	14.0136	3.8089	3.3411
	RMSE	5.8436	6.9098	6.9040	13.7807	11.8436	6.2646	16.4364	17.6247	6.2148	6.1151
	ACC	0.9004	0.8822	0.8823	0.7651	0.7004	0.8932	0.5946	0.5062	0.8941	0.8958
	R2	0.8205	0.7546	0.7528	0.0492	0.7205	0.7951	0.0375	0.0489	0.7979	0.8037
	Var	0.8220	0.7601	0.7595	0.0505	0.7220	0.7978	0.0374	0.0490	0.7980	0.8050
45min	MAE	3.2735	4.0607	4.2960	9.1320	8.2735	4.1217	12.5917	14.9371	4.1324	3.6399
	RMSE	6.3597	7.4254	7.5872	13.7751	11.3597	6.8776	16.7614	18.4267	6.7719	6.7389
	ACC	0.8916	0.8735	0. 8707	0.7652	0.6916	0.8828	0.5348	0.4371	0.8846	0.8852
	R2	0.7869	0.7143	0.6984	0.0478	0.7869	0.7509	0.0325	0.0427	0.7582	0.7597
	Var	0.7894	0.7196	0.7047	0.0490	0.7894	0.7538	0.0325	0.0429	0.7584	0.7612
60min	MAE	3.4260	4.2945	4.5812	9.1300	8.4260	4.4270	14.3057	15.7381	4.1324	3.9130
	RMSE	6.7397	7.9381	8.2088	13.7699	11.7397	7.3856	18.1420	18.9570	6.7719	7.2682
	ACC	0.8852	0.8648	0.8602	0.7654	0.7852	0.8742	0.4733	0.3502	0.8846	0.8762
	R2	0.7605	0.6702	0.6434	0.0462	0.6605	0.7109	0.0274	0.0391	0.7582	0.7185
	Var	0.7658	0.6753	0.6491	0.0473	0.6637	0.7137	0.0274	0.0394	0.7584	0.7202
T	Datasets	SZ-taxi									
	Metric	DASGCRN	*w/o* I_N	*w/o* A_{par}	*w/o* DA_t	*w/o* FCL	*w/o* T^D	*w/o* DA_tI_N	*w/o* DA_tA_{par}	*w/o* Mix	*w/o* TS
15min	MAE	2.4964	2.5311	2.7136	7.8361	6.5102	3.5257	9.6270	12.1762	2.5381	2.6109
	RMSE	3.9716	3.9888	4.1893	9.9781	8.0270	4.0046	11.3872	13.7591	4.0457	4.1046
	ACC	0.7233	0.7221	0.7040	0.3049	0.6143	0.7210	0.2371	0.2169	0.7182	0.7141
	R2	0.8554	0.8541	0.6733	0.0873	0.6588	0.8530	0.0674	0.0593	0.8498	0.8454
	Var	0.8559	0.8541	0.6733	0.0873	0.6592	0.8537	0.0680	0.0595	0.8500	0.8454
30min	MAE	2.5064	2.6485	2.7365	7.8344	6.3132	3.5600	10.3725	12.8530	2.5531	2.6485
	RMSE	3.9899	4.1580	4.3476	9.9757	8.6667	4.0576	11.3774	14.6328	4.0684	4.1580
	ACC	0.7221	0.7103	0.6714	0.3050	0.6034	0.7173	0.2046	0.1928	0.7166	0.7103
	R2	0.8541	0.8414	0.6573	0.0878	0.6521	0.8491	0.0592	0.0578	0.8482	0.8414
	Var	0.8546	0.8414	0.6573	0.0878	0.6524	0.8498	0.0592	0.0579	0.8484	0.8414
45min	MAE	2.5157	2.6530	2.7942	7.8392	6.1271	3.5883	11.9260	13.6430	2.5858	2.6530
	RMSE	4.0034	4.1669	4.6217	9.9800	8.0709	4.0908	12.8517	15.0647	4.1209	4.1669
	ACC	0.7211	0.7098	0.6350	0.3049	0.5965	0.7151	0.1794	0.1744	0.7130	0.7098
	R2	0.8531	0.8408	0.6114	0.0874	0.6067	0.8467	0.0516	0.0538	0.8444	0.8408
	Var	0.8536	0.8408	0.6112	0.0874	0.6070	0.8473	0.0510	0.0542	0.8445	0.8408
60min	MAE	2.5254	2.6635	2.8272	7.8342	6.6996	3.6056	11.7428	13.6430	2.5981	2.6635
	RMSE	4.0167	4.1821	4.9261	9.9778	6.3843	4.1166	14.2693	15.0647	4.1365	4.1821
	ACC	0.7202	0.7085	0.5073	0.3050	0.5512	0.7133	0.1274	0.1744	0.7119	0.7087
	R2	0.8521	0.8397	0.5800	0.0882	0.5856	0.8448	0.0375	0.0538	0.8432	0.8397
	Var	0.8526	0.8397	0.5800	0.0882	0.5860	0.8453	0.0374	0.0542	0.8435	0.8397

Table III presents the ablation study results for the DASGCRN variants on the Los-loop and SZ-taxi datasets, indicating that each design module performs as expected. To visually compare the performance of DASGCRN and its variants. Fig. 6 further illustrates the comparative results for both datasets. From the results, we can observe:

1) When the dynamic adjacency matrix DA_t is removed, the model's predictive performance declines sharply, demonstrating the necessity of capturing dynamic information within the road network.

2) The learnable parameter matrix A_{par} can adaptively explore potential spatial relationships between nodes, contributing to improved predictive performance.

3) The removal of temporal embeddings T^D results in a decrease in overall model performance. This indicates that temporal embeddings have a significant impact on the predictive capability of DASGCRN.

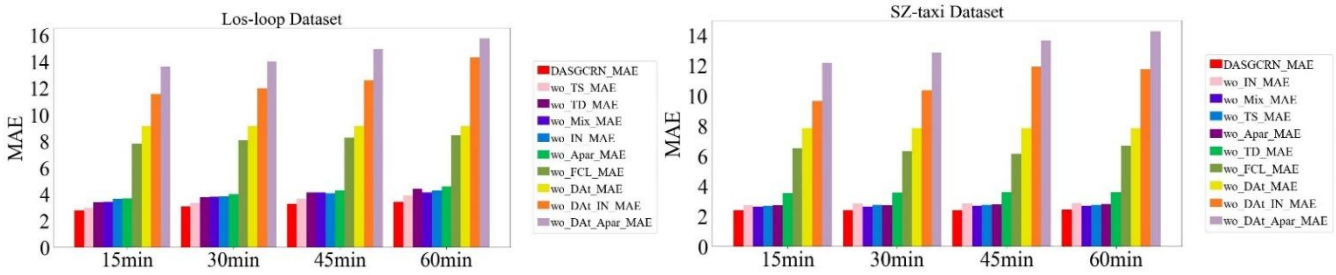


Fig. 6. Ablation experimental results on Los-loop and SZ-taxi

VI. Conclusion

A traffic speed prediction model DASGCRN based on dynamic spatial-temporal information is proposed. Firstly, the parameterized matrix and self-recurrent adjacency matrix are used to represent the spatial relationship between nodes effectively, which surpasses the traditional static adjacency matrix. Secondly, the dynamic graph is generated by using spatial-temporal feature embedding and traffic speed sequence, and the effective combination of dynamic graph and predefined graph is realized by step generation process. This method reduces the dependence on the prior knowledge of the road network and enhances the capture of the dynamic spatial-temporal dependence in the traffic network. Finally, the residual decomposition with skip connection is used to facilitate the transfer of feature information. It improves the training process of the model, and further improve the performance of the model.

In the future, we will further integrate more real-time datasets associated with traffic flow prediction, such as Internet of Things devices, autonomous vehicles, and mobile sensors, the model can capture the dynamic variations of the traffic network more accurately. Meanwhile, we will also enhance our attention to external factors such as weather and road conditions, and incorporate them as key features into the model, thereby enhancing the model's expressive capacity and predictive accuracy in a data-driven manner.

Acknowledgment

All the data mentioned in this article can be available at https://github.com/pritamBikram/Traffic_Dataset.

We would like to thank the research team that provided the dataset.

Funding

This work was supported by the Outstanding Youth Innovation Teams in Higher Education of Shandong Province(2019KJN048).

References

[1] Yin X, Wu G, Wei J, et al. Deep learning on traffic prediction: Methods, analysis, and future directions[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(6): 4927-4943.

[2] Liu J, Guan W. A summary of traffic flow forecasting methods[J]. Journal of highway and transportation research and development, 2004, 21(3): 82-85.

[3] Kamarianakis Y, Prastacos P. Forecasting traffic flow conditions in an urban network: Comparison of multivariate and univariate approaches[J]. Transportation Research Record, 2003, 1857(1): 74-84.

[4] Montesinos López O A, Montesinos López A, Crossa J. Support vector machines and support vector regression[M]//Multivariate Statistical Machine Learning Methods for Genomic Prediction. Cham: Springer International Publishing, 2022: 337-378.

[5] Yao R, Zhang W, Zhang L. Hybrid methods for short-term traffic flow prediction based on ARIMA-GARCH model and wavelet neural network[J]. Journal of Transportation Engineering, Part A: Systems, 2020, 146(8): 04020086.

[6] MENG Chuang, WANG Hui, LIN Hao, LI Kecen, WANG Xinpeng. Review of Research on Road Traffic Flow Data Prediction Methods[J]. Computer Engineering and Applications, 2023, 59(14): 51-61.

[7] CUI Jian-xun, YAO Jia, ZHAO Bo-yuan. Review on short-term traffic flow prediction methods based on deep learning[J]. Journal of Traffic and Transportation Engineering, 2024, 24(2): 50-64.

[8] Yang D, Li S, Peng Z, et al. MF-CNN: traffic flow prediction using convolutional neural network and multi-features fusion[J]. IEICE TRANSACTIONS on Information and Systems, 2019, 102(8): 1526-1536.

[9] Lu S, Zhang Q, Chen G, et al. A combined method for short-term traffic flow prediction based on recurrent neural network[J]. Alexandria Engineering Journal, 2021, 60(1): 87-94.

[10] Belhadi A, Djenouri Y, Djenouri D, et al. A recurrent neural network for urban long-term traffic flow forecasting[J]. Applied Intelligence, 2020, 50: 3252-3265.

[11] Zhao L, Song Y, Zhang C, et al. T-GCN: A temporal graph convolutional network for traffic prediction[J]. IEEE transactions on intelligent transportation systems, 2019, 21(9): 3848-3858.

[12] Zhu J, Wang Q, Tao C, et al. AST-GCN: Attribute-augmented spatial-temporal graph convolutional network for traffic forecasting[J]. IEEE Access, 2021, 9: 35973-35983.

[13] YANG Ping, LI Chengxin, LIU Yicheng, et al. A spatio-temporal graph model for multi-step prediction of dynamic traffic flow[J]. Computer Engineering and Design, 2024,45(04): 1195-1201.

[14] Wu Z, Pan S, Long G, et al. Graph WaveNet for Deep Spatial-Temporal Graph Modeling[C]//proceedings of the International Joint Conference on Artificial Intelligence. 2019: 1907-1913.

[15] Choi J, Choi H, Hwang J, et al. Graph neural controlled differential equations for traffic forecasting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(6): 6367-6374.

[16] Fang Z, Long Q, Song G, et al. Spatial-temporal graph ode networks for traffic flow forecasting[C]//Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining. 2021: 364-373.

[17] Zhang X, Huang C, Xu Y, et al. Traffic flow forecasting with spatial-temporal graph diffusion network[C]//Proceedings of the AAAI conference on artificial intelligence. 2021, 35(17): 15008-15015.

[18] Lan S, Ma Y, Huang W, et al. DSTAGNN: Dynamic spatial-temporal aware graph neural network for traffic flow forecasting[C]//International conference on machine learning. PMLR, 2022: 11906-11917.

[19] Li F, Feng J, Yan H, et al. Dynamic graph convolutional recurrent network for traffic prediction: Benchmark and solution[J]. ACM Transactions on Knowledge Discovery from Data, 2023, 17(1): 1-21.

[20] Shi Z, Zhang Y, Wang J, et al. DAGCRN: Graph convolutional recurrent network for traffic forecasting with dynamic adjacency matrix[J]. Expert Systems with Applications, 2023, 227: 120259.

[21] Zhang X, Wen S, Yan L, et al. A hybrid-convolution spatial–temporal recurrent network for traffic flow prediction[J]. The Computer Journal, 2024, 67(1): 236-252.

[22] Xia Z, Zhang Y, Yang J, et al. Dynamic spatial–temporal graph convolutional recurrent networks for traffic flow forecasting[J]. Expert Systems with Applications, 2024, 240: 122381.

[23] Fang S, Zhang Q, Meng G, et al. GSTNet: Global spatial-temporal network for traffic flow prediction[C]//IJCAI. 2019: 2286-2293.

[24] Wang X, Ma Y, Wang Y, et al. Traffic flow prediction via spatial temporal graph neural network[C]//Proceedings of the web conference 2020. 2020: 1082-1092.

[25] Bai J, Zhu J, Song Y, et al. A3t-gcn: Attention temporal graph convolutional network for traffic forecasting[J]. ISPRS International Journal of Geo-Information, 2021, 10(7): 485.

[26] Tian X, Zou C, Zhang Y, et al. NA-DGRU: A Dual-GRU Traffic Speed Prediction Model Based on Neighborhood Aggregation and Attention Mechanism[J]. Sustainability, 2023, 15(4): 2927.

[27] Wang C H, Cai J, Ye Q, et al. A two-stage convolution network algorithm for predicting traffic speed based on multi-feature attention mechanisms[J]. Journal of Intelligent & Fuzzy Systems, 2023 (Preprint): 1-16.

[28] Sun L, Liu M, Liu G, et al. FD-TGCN: Fast and dynamic temporal graph convolution network for traffic flow prediction[J]. Information Fusion, 2024, 106: 102291.

Enhanced Aquila Optimizer Algorithm for Efficient Stance Classification in Online Social Networks

Na LI
Department of Information Technology, ZhengZhou Vocational College of Finance and Taxation,
Zhengzhou, Henan, 450048, China

***Abstract*—Stance classification in Online Social Networks (OSNs) is essential to comprehend users' standpoints on various issues relating to social, political, and commercial aspects. However, traditional methods applied to large datasets and complex text structures usually face several challenges. This study introduces the Enhanced Aquila Optimizer (EAO), a metaheuristic algorithm designed to improve convergence and precision in stance classification tasks. EAO incorporates three new strategies: Opposition-Based Learning (OBL) to improve the exploration, Chaotic Local Search (CLS) to escape from the local optima, and a Restart Strategy (RS) to rejuvenate the search process. Experimental assessments on benchmark OSN datasets prove the superiority of EAO in terms of accuracy, precision, and computational efficiency compared to state-of-the-art methods. These findings position EAO as a potential revolution for stance classification and other large-scale text analysis tasks by offering a robust solution that can be used in real-time for complex OSN scenarios.**

***Keywords*—Stance classification; online social networks; opposition-based learning; chaotic local search; Aquila Optimizer**

I. Introduction

In the past couple of years, the rapid development and growth of Online Social Networks (OSNs) have been driven by the remarkable evolution of our economic and social lives due to the internet [1]. OSNs represent powerful communication platforms for sharing interests, exchanging ideas, or creating communities. Schneider, et al. [2] defined OSNs as networks of individuals with common interests in activities and mutual relationships. Most OSNs enable users to share messages, photos, and videos, rate objects, or discuss any topic, from health problems to political opinions [3]. Another social network is Facebook, estimated to have 2.38 billion monthly registered users as of March 2019, and Twitter has around 330 million monthly active users. These are dynamic and real-time sources of information wherein users actively post their views on certain topics. Thus, OSNs have become a significant means of understanding trends in public opinion and have valuable insights into how ideas originate and spread in complex social networks [4].

The challenges for position classification in OSNs are enormous, considering the immense volume and complexity of information exchanged daily. OSNs generate vast amounts of unstructured text data covering diverse subjects, from sixteen social issues to product reviews, often including various linguistic expressions, slang, abbreviations, and dialects of different regions [5]. This makes developing a model that can generalize for all types of content pretty tough. Moreover, the language used in OSNs is generally informal, with humor, sarcasm, and implicit cues included, making their correct classification challenging. Traditional machine learning techniques, including Neural Networks (NNs) and Support Vector Machines (SVM), often need to scale up efficiently for big dynamic data sets [6, 7]. These methods can only grab those subtle expressions reliant upon context and reduce the accuracy of finding users' positions. More advanced and scalable methods are needed to process the inherent complexity present in OSN data and improve the accuracy of position classification [8].

Metaheuristic optimization algorithms have gained considerable attention due to their efficiency in exploring large-sized search spaces associated with complex classification tasks [9]. Unlike the conventional methods of optimization, which suffer from high dimensionality and run the risk of getting trapped in local optima, metaheuristic algorithms have shown flexibility in navigating through such complex landscapes [10]. All these algorithms are population-based and explore a range of possible solutions at every iteration, with a better chance of converging to a global optimum. Their adaptability predisposes them to be good performers across various scenarios, especially in data-rich dynamic environments like OSNs. Metaheuristic algorithms explore the balance between exploration and exploitation in high-dimensional classification tasks and complex data to reach high accuracy while reducing computational costs [11]. This flexibility and strength of position makes metaheuristic optimization a valuable tool in solving the problems of position classification in OSNs.

Aquila Optimizer (AO) is a newly introduced metaheuristic algorithm. The algorithm is inspired by the hunting strategy of the Aquila bird, one of the best algorithms in balancing exploration and exploitation to find its prey. AO simulates four different phases of the hunting behavior of the Aquila bird for moving between diverse explorations to the search space and focused exploitation around promising solutions. So far, this approach has successfully solved several optimization tasks, positioning AO as one of the more promising choices for challenging high-dimensional problems. Like many other metaheuristics, however, AO has ample limitations. In complex problem spaces, convergence can be slow, and this method may get stuck in suboptimal regions. These all relate to the ability of the algorithm to balance the trade-off between exploration and exploitation, which is the central challenge in the efficient exploration of massive, complex Spatiotemporal data environments characterizing OSNs. This makes enhancements

of AO of paramount importance and assures better performance of such robust applications on classification.

The complexity of position classification in OSNs, marked by vast data volumes, high dimensionality, and nuanced language, demands highly efficient optimization methods. Although AO performed so well in optimization performance in many scenarios, the efficiency of the algorithms is crippled when dealing with the scale and intricacies of OSN data. Convergence to optimality can take a long time and sometimes even get stuck in local optima, which will make this algorithm deteriorate more for the performance of large-scale classification tasks where timely and proper analysis is required. Based on these findings, we will represent an Enhanced Aquila Optimizer (EAO) by implementing extra mechanisms into AO to enhance its capability.

Metaheuristic optimization algorithms have drawn much interest since they can handle high dimensions and the inherent complexity of OSN's stance classification tasks. It is still challenging to efficiently map these complicated data structures into user and linguistic patterns. Recent breakthroughs in graph-based embeddings, such as Subgraph2vec, confirmed the power of random walk-based methods for meaningful representation extraction from structured data, including knowledge graphs [12]. Motivated by such techniques, our study focuses on optimizing the exploration and exploitation of high-dimensional data with the EAO.

EAO unifies three central strategies, namely Opposition-Based Learning (OBL), Chaotic Local Search (CLS), and a Restart Strategy (RS). OBL enhances the exploration capability by considering the opposite solutions; hence, it increases the diversity in exploiting the search space for high possibilities of converging to global optima. CLS contributes controlled randomness to the search process, which helps the algorithm maintain its distance from local optima and promotes diversity in potential solution regions. Finally, it uses RS in its last stage to reset the stagnated solution and reinitialize the diverted search agents into suboptimal areas, rejuvenating the exploration process and accelerating the convergence speed.

II. Related Work

Several works have addressed position classification in OSNs using machine learning and optimization techniques. Parimi and Rout [13] envisaged a paradigm based on similarity to spread rival and counter-rumors. They introduced a probabilistic score-based system for determining whether a user should support a rumor or a counter-rumor. This paper uses a neighborhood analysis-based propagation methodology to examine the effects of rumor and counter-rumor cascades in OSNs. Determining the minimum user count that will start the counter-rumor and reducing communication costs in the application is another complex problem this study attempts to tackle.

A comprehensive Hybrid Clustered Shuffled Frog-Leaping Algorithm-Particle Swarm Optimization (HCSFLA-PSO) algorithm was proposed by Hu, et al. [14] to quickly and continuously suppress rumors spread in OSNs. First, a novel scheme for refuting rumors and an inventive depiction of trust levels are presented by dissecting social interactions and examining intimacy, independence, and credibility. Second, a thorough HCSFLA-PSO algorithm is developed, utilizing the PSO algorithm's quick convergence and the SFLA's local clustering to refute rumors. This comprises the CP-HCSFLA-PSO component for real-time rumor refutation during truth evolution and the CNP-HCSFLA-PSO sub-algorithm for timely rumor refutation, adapted to various social relationships with differing levels of trust.

Saeidi [15] provided a straightforward, uncomplicated, and efficient approach to determining trust relationships between different OSN members. Consequently, four novel approaches for assessing the trustworthiness of users are developed and evaluated with the Anderson-Darling statistical hypothesis and Kolmogorov-Smirnov analyses to choose and verify the most suitable model using 20,613 empirical data from 4,552 volunteers in social networks. A metaheuristic algorithm based on the Artificial Bee Colony (ABC) optimization approach was designed to address the temporal complexity of the issue and identify the optimal model fit.

Fatehi, et al. [16] have developed a hybrid model that integrates graph-based and artificial intelligence methodologies to enhance the coverage and precision of online social networks. This method uses a distributed learning automaton rather than established graph-based search methods like breadth-first search, which can identify all reliable associations without limitations. Simulation findings conducted on an actual dataset from Epinions.com show enhanced accuracy and coverage relative to leading algorithms. The suggested approach demonstrates an accuracy of 0.939, indicating a 6% improvement over similar algorithms.

Mallick, et al. [17] devised a collaborative deep-learning algorithm for detecting fake news. The suggested method employs user input to assess the trustworthiness of communications, with message ranking established according to these evaluations. Messages of lower rank are preserved for linguistic analysis to verify their authenticity, whereas highly ranked information is acknowledged as legitimate communications. A Convolutional Neural Network (CNN) transforms user input into rankings inside the deep learning framework. Messages with negative ratings are returned to the system for CNN model training.

Vaghefi, et al. [18] investigated personal disclosure and brand perception on online networks like Foursquare and Twitter. Based on social and hyper-personal information processing theories, relationships between peers, distance, and advertising messages are examined. An integrated dataset reveals that self-disclosure is significantly influenced by checking in with friends and their proximity. Especially when interacting with well-informed peers, individuals prefer to ignore inquiries that reflect poorly on their health.

Bangyal, et al. [19] investigated sentiment analyses and the detection of fake news using machine learning and deep learning in different challenges, and significant data volumes arose and became dynamic over Twitter. They presented for the first time an innovative proposed method of detecting COVID-19-related false information in deep learning models, particularly the BiGRU model, for which the obtained accuracy was scored very impressively at around 91%. Bangyal, et al.

[20] compared classical machine learning classifiers such as Support Vector Machines and Random Forests. The results proved efficient in the competition and improved the robustness of even simpler models. Bangyal, et al. [21] extended the work by including semantic models and the TF-IDF and compared eight machine-learning classifiers with four deep-learning classifiers. The valuable outcome provided a comparison between accuracy and computational cost. While an advance in sentiment analysis, these works pointed to the need for scalable algorithms of high precision that handle stance classification in large and complex datasets. This motivated the development of the EAO.

Position classification in OSNs is an attractive topic since there has been a high demand from users to make effective use of people's opinions and to control information dissemination effectively. Various state-of-the-art machine learning and optimization techniques have been studied for the works shown in Table I regarding rumor suppression, trust evaluation, sentiment analysis, and fake news detection. These works span classical machine learning model applications, leveraging the latest metaheuristics algorithms and deep learning frameworks, each with different advantages and limitations. Despite such advances, most previous approaches also either suffer from issues like scalability and computation efficiency or are very poorly applicable to OSN data, which by default is complex and time-variant. This extends earlier work and bridges previous gaps in the literature that resulted in the development of the proposed EAO, providing for a state-of-the-art scalable and efficient stance classifier.

TABLE I. AN OVERVIEW OF PREVIOUS STUDIES

Reference	Methodology/algorithm	Dataset	Key findings	Limitations
[13]	Neighborhood analysis-based propagation and multi-objective genetic algorithm	Actual datasets for rumor and counter-rumor analysis	Efficiently reduced communication costs and minimized counter-rumor users	Limited scalability for larger datasets
[14]	Hybrid clustered shuffled frog-leaping algorithm-particle swarm optimization	Simulated datasets for real-time and evolving rumor refutation	Effectively suppressed rumors, accounting for trust and timeliness	High computational complexity for real-time adaptation
[15]	Artificial bee colony optimization	Facebook dataset (20,613 data points from 4,552 users)	Identified optimal trust models with reduced temporal complexity	Requires significant computational resources
[16]	A graph-based and distributed learning automaton	Epinions.com dataset	Enhanced accuracy and coverage by 6% and 10%, respectively	Does not address dynamic changes in network topology
[17]	Collaborative deep learning with CNN	Simulated datasets for fake news detection	Achieved 98% accuracy in detecting fake news	Dependent on user input for initial rankings
[18]	Integrated social and hyper-personal information processing theories	Integrated dataset (Foursquare and Twitter)	Revealed self-disclosure influenced by peer proximity and advertising	Limited to health-related and brand perception studies
[19]	BiGRU deep learning model	COVID-19-related false information dataset	Achieved 91% accuracy in detecting COVID-19-related fake news	Focused only on COVID-19-related data
[20]	Classical machine learning classifiers (SVM, Random Forest)	Twitter datasets	Proved efficient for simpler models in sentiment analysis	Simpler models are less effective for nuanced stance classification
[21]	Semantic models with TF-IDF, multiple classifiers	Twitter datasets with semantic enrichment	Provided a trade-off comparison between accuracy and computational cost	Needs scalable algorithms for complex stance classification tasks

III. PROPOSED METHOD

A. Problem Definition

Position classification is a task aimed at estimating the stance of users concerning particular topics in OSNs, including opinion classification as favourable, unfavourable, or neutral towards various entities or topics. Unlike general sentiment analysis, which broadly adjectives text into positive, negative, or neutral areas, position classification is a closer sentiment analysis. It pinpoints users' opinions concerning clearly defined subjects, such as political figures, social problems, products, or happenings, offering a better perspective on public opinion. This fact renders this task especially hard due to the dynamism of the OSN data, which are often unstructured, different in linguistic styles, and very massive.

One of the big questions is how to classify user opinions efficiently and accurately from a large dataset with high precision. OSNs generate magnificent volumes of textual data daily, from explicit endorsement to subtle criticism or neutral observations. Such data is usually unstructured and informal, including colloquialisms, abbreviations, and context-dependent meanings, which make their correct classification challenging. Moreover, the data in OSN is typically full of mixed emotions,

sarcasm, and idiomatic expressions, complicating their classification.

This vast amount of data is bound together with complexity, and traditional machine learning with statistical methods faces scalability issues while handling this immense data. While the dataset grows, conventional models like support vector machines lose much of their accuracy and efficiency, or even simple neural networks take up much computational time and resources for training and inference. Moreover, these models must be revised to represent complicated relationships and subtleties specific to the context in the OSN language and limit generalization across a wide range of topics with users' expressions.

Metaheuristic optimization is one of the ways of taming this. Considering position classification as an optimization problem, metaheuristic algorithms will handle high-dimensional feature spaces to arrive at the uppermost solution, a highly performant classification with limited computational overhead. Within such a paradigm, the optimization algorithms must be solid and flexible enough to meet the peculiar demands of the data in OSNs. In this regard, it may likely involve the development of classical algorithms, such as Aquila Optimizer, for better exploration, adaptability, and higher accuracy regarding OSN position classification.

B. Enhanced Aquila Optimizer

AO is a newly formulated population-based optimization algorithm inspired by Aquila birds' predatory strategies proposed by Abualigah, et al. [22]. Aquila birds originate from the Northern Hemisphere and are considered one of the most famous predators, distinguished by their agility, strong talons, and strong feet. Thus, they may catch a wide range of prey, from squirrels and rabbits to marmots and hares. The proposed AO algorithm draws inspiration from four different foraging behaviors of Aquila birds that oscillate between exploration and exploitation during hunting. The AO algorithm starts with a randomly generated population of candidate solutions that can be mathematically expressed as follows:

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,j} & \cdots & X_{1,D} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,j} & \cdots & X_{2,D} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{n-1,1} & X_{n-1,2} & \cdots & X_{n-1,j} & \cdots & X_{n-1,D} \\ X_{n,1} & X_{n,2} & \cdots & X_{n,j} & \cdots & X_{n,D} \end{bmatrix} \quad (1)$$

Each element $X_{i,j}$ in this matrix represents the position of an agent, calculated as follows:

$$X_{i,j} = rand \times \left(UB_j - LB_j\right) + LB_j, \;\; i = 1, \dots, n, \;\; j = 1, \dots, D \quad (2)$$

Where *rand* represents a random number between 0 and 1, UB_j and LB_j are the upper and lower bounds for each dimension, *n* denotes the population size, and *D* is the number of decision variables. The AO algorithm consists of four unique stages, facilitating the balance between exploration and exploitation. These stages are triggered under the following conditions:

$$\begin{cases} Perform\ exploration, & if\ t \le \frac{2}{3} \times T \\ Perform\ exploitation, & otherwise \end{cases} \quad (3)$$

Where T refers to the total number of iterations and t is the current iteration.

1) Expanded exploration: At this stage, Aquila searches over a large area to find its prey by performing a high dive and then a soaring flight. In this phase, the position of each agent is updated using Eq. (4).

$$X_1(t+1) = X_{best}(t) \times \left(1 - \frac{t}{T}\right) + (X_M(t) - X_{best}(t) \times rand) \quad (4)$$

Xbest represents the best position, and XM(t) gives the average position of the current population generation.

2) Narrowed exploration: This process is Aquila's most commonly adopted hunting strategy. It involves a short gliding flight and contour-following maneuver. Eq. (5) updates the agent's position.

$$X_2(t+1) = X_{best}(t) \times Levy(D) + X_R(t) + (\gamma - x) \times rand \quad (5)$$

Where Levy(D) is a Levy flight distribution, D represents the number of dimensions, and XR indicates an agent's position at random. The Levy flight is defined by Eq. (6).

$$Levy(D) = s \times \frac{u \times \sigma}{|v|^{1/\beta}} \quad (6)$$

Where β=1.5, u, and v are random values, and s=0.01 is a scaling factor.

3) Expanded exploitation: In the third stage, the search scope is narrowed further; the agent is prepared for an attack through a low-flight preliminary assault. The position update is done as follows:

$$X_3(t+1) = \left(X_{best}(t) - X_M(t)\right) \times \alpha - rand + \left((UB - LB) \times rand + LB\right) \times \delta \quad (7)$$

Where α and δ are exploitation parameters set to 0.1.

4) Narrowed exploitation: In this final stage, Aquila chases and attacks the prey using an escape trajectory. Eq. (8) updates the agent's position.

$$X_t(t+1) = QF \times X_{best}(t) - (G_1 \times X(t) \times rand) - G_2 \times Levy(D) + rand \times G_1 \quad (8)$$

Here, QF represents the quality factor, where $QF(t) = \frac{t}{2 \times rand - 1} / (1 - r)^2$, and $G_1 = 2 \times rand - 1$ and $G_2 = 2 \times \left(1 - \frac{t}{T}\right)$ manage different motion and attack angle aspects.

C. Opposition-based Learning

The OBL strategy was first proposed by Tizhoosh [23], and since then, it has been implemented for several swarm optimization algorithms to improve their performance

significantly. Many researchers have practiced this technique on swarm optimization algorithms to enhance the exploration and convergence potential of the swarm optimization algorithm. For instance, OBL is combined with the SSA framework in [24] to avoid local optimization problems. In [25], the Harris Hawks Optimization (HHO) algorithm combines the OBL concept with a chaotic local search to substantially improve its exploration capability. Zhang, et al. [26] used OBL to enhance the algorithm's performance in arithmetic optimization. The OBL works based on comparing the original solution's fitness value with its opposition. The opposition solution of an integer x in the bounds $[lb, ub]$ can be computed using Eq. (9).

$$\bar{x} = ub + lb - x \tag{9}$$

For a vector x, the opposite value of each component can be determined as follows.

$$\bar{x_j} = ub_j + lb_j - x_j \tag{10}$$

Where lb_j and ub_j represent the lower and upper bounds for the j^{th} dimension.

D. Chaotic Local Search

CLS is one of the well-known algorithms applied to various swarm optimization techniques, such as the Jaya Algorithm [27], brainstorm optimization [28], and WOA [29]. The CLS approach is usually implemented by using a logistic map in the following:

$$o^{s+1} = Co^s(1 - o^s) \tag{11}$$

Where s denotes the iteration and C is a control parameter, typically set to 4. Values of o^1 can be initialized at 0.25, 0.50, or 0.75. CLS focuses on local searches around the optimal solution found so far, aiming to enhance accuracy within that neighborhood. The CLS-generated values C_s in iteration i are computed as follows:

$$C_s = (1 - \mu) \times T + \mu \hat{C_i}, \;\; i = 1,2, \dots, n \tag{12}$$

Where $\hat{C_i}$ is calculated as:

$$\hat{C_i} = LB + C_i \times (UB - LB) \tag{13}$$

Here, μ represents a shrinking factor, determined by:

$$\mu = \frac{T - t + 1}{T} \tag{14}$$

E. Restart Strategy

In the optimization process, some agents will be entrapped or stuck in a particular local optimal and fail to obtain the best performance. Such agents cannot contribute to improving the search but consume additional computing resources. Zhang, et al. [30] proposed RS that restarts or relocates such stagnant agents. The RS strategy traces the improvement frequency of each agent. If an agent does not find newer and better solutions, a trial will increase in value. When it reaches a certain threshold predefined, the position of the agent resets according to the following equations:

$$X(t + 1) = lb + \text{rand} \cdot (ub - lb) \tag{15}$$

$$X(t + 1) = \text{rand} \cdot (ub + lb) - X(t) \tag{16}$$

F. Proposed Algorithm

EAO, the improvisation proposed in this paper, tries to overcome its bottleneck with the help of standard AO by applying strategies like OBL, RS, and CLS. OBL applied during initialization and in the position updates ensures that the optimizer is initialized with a robust set of agents and explores a good amount of solution space. CLS fine-tunes the best solution in every iteration to ensure an enhanced search precision neighborhood. Finally, RS re-positions the stagnant agents, which stirs the exploration process. Fig. 1 shows the pseudocode of EAO. To obtain the computational complexity of EAO, one can focus on the three phases separately: initialization, assessment, and position update. This gives, in total, the complexity *O(EAO)*:

$$O(EAO) = O(\text{Initialization}) + O(\text{Assessment}) + O(\text{Position update}) + O(\text{CLS} + \text{OBL} + \text{RS}) \tag{17}$$

Assuming T as the total iterations, N as the population size, and D as the number of dimensions, each component has the following complexity:

- Initialization: *O(N)*
- Assessment: $O(N \times T)$
- Position update: $O(N \times D \times T)$
- RS and OBL: $O(N \times D \times T)$
- CLS: $O(N \times T)$

The overall complexity of the EAO can be expressed as follows:

$$O(EAO) = O(N) + O(N \times T) + O(N \times D \times T) + O(N \times D \times T) = O(N \times D \times T) \tag{18}$$

```
Initialize the population matrix X for the AO
Calculate the opposition-based set X̄ and select the top N solutions from X ∪ X̄
Set the initial parameters for the AO
while (iteration t < maximum iterations T) do
    Evaluate the objective function for each solution
    Identify the best agent X_best
    for each agent i from 1 to N do
        Update the mean position of the current solution set
        Compute the parameters y, x, G1, G2, and Levy(D)
        if current iteration t ≤ 2/3T then
            if (random value ≤ 0.5) then
                Update the current position using Eq. 4
                Compute the opposite position using Eq. 10
            else
                Update the current position using Eq. 5
                Compute the opposite position using Eq. 10
        else
            if (Fitness value rand ≤ 0.5) then
                Update the current position using Eq. 7
                Compute the opposite position using Eq. 10
            else
                Update the current position using Eq. 8
                Compute the opposite position using Eq. 10
        end if
    end for
Apply the restart strategy using Eqs. 15 and 16
Apply CLS strategy using Eq. 12
```

Fig. 1. Enhanced Aquila Optimizer.

IV. Performance Evaluation

Stance detection involves an automated method for determining how a writer expresses support, opposition, or neutrality toward a particular argument or topic. The scope of analysis in this field is extensive and may include subjects such as individuals, organizations, governmental policies, social movements, or commercial products. For instance, a detailed examination of Barack Obama's speeches could be conducted to ascertain his position on regulating guns in the U.S. Individuals convey their opinions on various issues through various platforms, including Facebook, YouTube, Twitter, and online forums. This approach applies to several areas, such as stand-alone stance classification, information retrieval, automatic text summarization, and text inference. Over the past decade, significant research has focused on modeling stances in digital media. This study utilized four datasets from an existing database, comprising training and testing data derived from tweets. These datasets cover over two million stance expressions across four topics.

In this work, stance classification is analyzed as an optimization problem. The dataset is initially converted into structured data for analysis. Such datasets used to perform experimentation include Hillary Clinton, the Legalization of Abortion, Atheism, and the Feminist Movement, which constitute an optimization space. Firstly, documents are generated for each dataset after the preprocessing steps. Since the number of tweets is taken across rows in the document array, the total volume of words obtained by preprocessing defines the dimensionality of a single tweet. If a word from a set is featured in a given tweet, its column will be assigned a value equal to 1. In other cases, when some word from a set does not appear in the text of a certain tweet, the respective column will be recorded with 0. Consequently, the document matrix is only composed of zeros and ones.

A vector of word weights is also constructed based on word weights within documents. The "maximum word passes" denotes the highest quantity of records where a single word occurs. The weight for each word is then calculated as the ratio of the word's frequency to the maximum occurrence across all documents. These calculated weights collectively form the word weight vector used in the optimization analysis.

In this research, the similarities among the population participants (potential solutions) and the document structure are critical for constructing an effective classification system. Specifically, correlations between document matrix components and possible solutions are carefully evaluated. Various methods exist for measuring similarity, including Overlap, Dice, Jaccard, and Cosine. This study found that the Jaccard similarity measure yielded the best results, and a modified Jaccard similarity measure quantified correlations across texts. This approach determines standard features by calculating the proportion of shared features over the entire text.

Mathematically, for an individual $X_i = (X_{i1}, X_{i2}, \ldots, X_{iM})$ in the population and a corresponding line $D_j = (D_{j1}, D_{j2}, \ldots, D_{jM})$ Jaccard similarity is calculated as follows:

$$J(D_j, X_i) = \frac{X_i \cap D_j}{X_i \cup D_j} \quad (19)$$

The extended Jaccard similarity is computed using Eq. (20).

$$J(D_j, X_i) = \frac{\sum_{k=1}^{M} D_{jk} X_{ik}}{\sum_{k=1}^{M} (D_{jk}^2) + \sum_{k=1}^{M} (X_{ik}^2) - \sum_{k=1}^{M} D_{jk} X_{ik}} \quad (20)$$

Furthermore, to account for the significance of word frequency, tweet word counts relative to the total document word count were included. This ratio is calculated using Eq. (21).

$$ratio_j = \frac{No.of\ words\ in\ the\ j^{th}\ comment}{Total\ No.of\ words\ in\ the\ document} \quad (21)$$

This ratio helps assign appropriate word weights based on their occurrence within the preprocessed dataset. The new similarity measure, incorporating both the Jaccard similarity and the word frequency ratio, is defined as:

$$Similarity_{ij} = \alpha \times jaccard + \beta \times ratio_j \quad (22)$$

Where α and β are coefficients whose sum equals one, allowing for a balanced weighting scheme to optimize the model's performance.

The similarity of each X_i within the population was evaluated based on all tweets included in the dataset. Classification of tweets was conducted by comparing the similarity value to a predefined threshold: tweets were classified as either above or below this threshold. To determine an individual's fitness in similarity analysis, Eq. (23) was used:

$$F = \alpha \times \frac{TP}{length} + \beta \times \frac{TP}{FP+TP} + \gamma \times \frac{TN}{FP+TN} + \omega \times \frac{TP}{FN+TP} \quad (23)$$

Where False Negative (FN) refers to the number of instances in which the rule incorrectly identified as negative when they belong to the positive class, True Negative (TN) stands for the number of cases accurately recognized as negative by the rule, False Positive (FP) signifies the count of instances incorrectly labeled as positive by the rule, even though they belong to the negative class, and True Positive (TP) represents the number of cases correctly identified by the rule as belonging to the positive class.

The coefficients $\alpha, \beta, \gamma, and\ \omega$ are weight values that must collectively sum to one. These weights are customizable and can be adjusted to optimize the performance of the fitness function, allowing the algorithm to be tailored for the best possible results in a given context.

The proposed EAO algorithm's effectiveness in addressing the stance detection problem, a complex task in online social network analysis, was evaluated and compared against several classifiers. The experiments were performed using MATLAB R2018b on a system equipped with an Intel Core i5-12400F CPU running at 2.5 GHz and 8 GB of RAM. The comparative analysis of the algorithms' results was based on the following classification metrics.

$$F - measure = \frac{2TP}{FN+FP+2TP} \quad (24)$$

$$Recall = \frac{TP}{FN+TP} \quad (25)$$

$$Precision = \frac{TP}{FP+TP} \quad (26)$$

$$Correctly\ labaled\ data = \frac{TN+TP}{FN+TN+FP+TP} \quad (27)$$

Tables II to V present a comparative analysis of EAO, ACO, and AO algorithms applied to stance detection datasets on various social issues. Each table illustrates the performance of multiple classification algorithms, focusing on F-measure, recall, precision, accuracy, and correctly labeled data metrics. For the Feminist Movement dataset (Table II), EAO performs better than others, achieving an accuracy rate of 61.387%, while NaiveBayes and staking also show competitive results. Similarly, in the Atheism dataset (Table III), EAO achieves the highest accuracy of 64.593%, followed closely by random forest.

In the analysis of the Legalization of Abortion dataset (Table IV), EAO emerges as the top-performing classification algorithm with an impressive accuracy rate of 70.201%, significantly higher than the other contenders. Finally, for the Hillary Clinton dataset (Table V), EAO shows better accuracy at 83.921%, indicating enhanced performance than the other algorithms.

TABLE II. RESULTS FOR THE FEMINIST MOVEMENT DATASET

Algorithm	F-measure	Recall	Precision	Accuracy (%)	Correctly labeled data
EAO	0.574	0.598	0.613	61.387	173
NaiveBayes	0.561	0.552	0.601	55.112	158
Stacking	0.522	0.563	0.521	56.845	164
QDA	0.522	0.563	0.521	56.845	163
REPTree	0.518	0.569	0.548	56.849	163
Random forest	0.512	0.517	0.505	51.586	148
Random tree	0.509	0.514	0.502	51.581	147
Extra tree	0.459	0.462	0.475	45.619	129

TABLE III. RESULTS FOR THE ATHEISM DATASET

Algorithm	F-measure	Recall	Precision	Accuracy (%)	Correctly labeled data
EAO	0.621	0.578	0.654	64.593	143
Random forest	0.619	0.628	0.623	62.784	139
Extra tree	0.559	0.571	0.569	57.705	127
Random tree	0.588	0.592	0.586	59.084	131
NaiveBayes	0.556	0.617	0.648	61.375	136
QDA	0.506	0.578	0.584	57.714	128
REPTree	0.501	0.528	0.577	52.716	117
Stacking	0.501	0.528	0.577	52.716	117

TABLE IV. RESULTS FOR THE LEGALIZATION OF ABORTION DATASET

Algorithm	F-measure	Recall	Precision	Accuracy (%)	Correctly labeled data
EAO	0.691	0.686	0.695	70.201	197
NaiveBayes	0.676	0.672	0.688	67.135	187
Random forest	0.666	0.678	0.682	67.845	191
Extra tree	0.643	0.651	0.647	65.342	182
Stacking	0.599	0.625	0.611	62.742	175
Random tree	0.597	0.601	0.605	60.254	168
QDA	0.578	0.635	0.578	63.572	177
REPTree	0.576	0.572	0.641	57.098	159

TABLE V. RESULTS FOR THE HILLARY CLINTON DATASET

Algorithm	F-measure	Recall	Precision	Accuracy (%)	Correctly labeled data
EAO	0.811	0.853	0.856	83.921	248
NaiveBayes	0.813	0.821	0.808	82.372	242
Random forest	0.801	0.818	0.799	82.371	242
REPTree	0.783	0.835	0.822	83.715	243
Extra tree	0.779	0.782	0.776	78.306	230
Random tree	0.727	0.733	0.721	73.546	216
QDA	0.632	0.677	0.719	67.786	199
Stacking	0.628	0.587	0.702	58.306	173

V. CONCLUSION

This paper proposed and evaluated EAO on the OSN stance detection problem, a complex high-dimensional classification problem. The performance of EAO was rigorously compared to ACO and classical AO on multiple datasets about social issues. In this regard, a comparative analysis was conducted, where the EAO outperforms its competitors on the classification accuracy of F-measure, recall, and precision for most of them, thus proving the more remarkable ability of EAO in exploring the search space in most cases. Our results suggest that embedding superior methodologies such as OBL and CLS contributes to robustness and efficiency while handling complex problems with EAO. EAO effectively resolves the challenges posed by stance detection in OSNs through improvements in convergence speed while sustaining diversity among solutions. This work may be extended by further optimization or hybridization strategies, including adaptation of EAO for other tasks related to social media analysis and deep learning frameworks for better performance. The proposed algorithm EAO has considerable potential for wild applications in dynamic and data-heavy settings, serving as a valid tool for large-scale social network data analysis.

REFERENCES

[1] H. Zhihong and L. Tao, "Presenting a Novel Method for Identifying Communities in Social Networks Based on the Clustering Coefficient," International Journal of Advanced Computer Science and Applications, vol. 14, no. 8, 2023.

[2] F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger, "Understanding online social network usage from a network perspective," in Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, 2009, pp. 35-48.

[3] M. Bhattacharya, S. Roy, S. Chattopadhyay, A. K. Das, and S. Shetty, "A comprehensive survey on online social networks security and privacy issues: Threats, machine learning-based solutions, and open challenges," Security and Privacy, vol. 6, no. 1, p. e275, 2023.

[4] M. R. Kondamudi, S. R. Sahoo, L. Chouhan, and N. Yadav, "A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches," Journal of King Saud University-Computer and Information Sciences, vol. 35, no. 6, p. 101571, 2023.

[5] G. Jethava and U. P. Rao, "Exploring security and trust mechanisms in online social networks: An extensive review," Computers & Security, p. 103790, 2024.

[6] S. Shin, Z. Jiang, R. E. Lim, and J. Lyu, "Forecasting the Spread of Sustainability Movement: Computational Analysis on Social Media Messages Promoting Climate Actions," Journal of Current Issues & Research in Advertising, vol. 45, no. 3, pp. 282-300, 2024.

[7] G. V. Kumar, M. I. Bellary, and T. B. Reddy, "Prostate cancer classification with MRI using Taylor-Bird squirrel optimization based deep recurrent neural network," The Imaging Science Journal, vol. 70, no. 4, pp. 214-227, 2022.

[8] P. Rajesh, M. Ismail. Ismail. B, M. Alam, and M. Tahernezhadi, "Network forensics investigation in virtual data centers using elk," in 2021 International Symposium on Electrical, Electronics and Information Engineering, 2021, pp. 175-179.

[9] M. Shoeibi, M. M. S. Nevisi, S. S. Khatami, D. Martín, S. Soltani, and S. Aghakhani, "Improved IChOA-Based Reinforcement Learning for Secrecy Rate Optimization in Smart Grid Communications," Computers, Materials and Continua, vol. 81, no. 2, pp. 2819-2843, 2024, doi: https://doi.org/10.32604/cmc.2024.056823.

[10] V. Nyemeesha, M. Kavitha, and B. Mohammed Ismail, "Detection and classification of skin cancer using unmanned transfer learning based probabilistic multi-layer dense networks," International Journal of Computational Intelligence and Applications, vol. 21, no. 04, p. 2250027, 2022.

[11] M. Shoeibi, M. M. S. Nevisi, R. Salehi, D. Martín, Z. Halimi, and S. Baniasadi, "Enhancing Hyper-Spectral Image Classification with Reinforcement Learning and Advanced Multi-Objective Binary Grey Wolf Optimization," Computers, Materials and Continua, vol. 79, no. 3, pp. 3469-3493, 2024, doi: https://doi.org/10.32604/cmc.2024.049847.

[12] E. Bozorgi, S. Soleimani, S. K. Alqaiidi, H. R. Arabnia, and K. Kochut, "Subgraph2vec: A random walk-based algorithm for embedding knowledge graphs," arXiv preprint arXiv:2405.02240, 2024, doi: https://doi.org/10.48550/arXiv.2405.02240.

[13] P. Parimi and R. R. Rout, "Genetic algorithm based rumor mitigation in online social networks through counter-rumors: a multi-objective optimization," Information Processing & Management, vol. 58, no. 5, p. 102669, 2021.

[14] X. Hu, X. Xiong, Y. Wu, M. Shi, P. Wei, and C. Ma, "A hybrid clustered SFLA-PSO algorithm for optimizing the timely and real-time rumor refutations in online social networks," Expert Systems with Applications, vol. 212, p. 118638, 2023.

[15] S. Saeidi, "A new model for calculating the maximum trust in Online Social Networks and solving by Artificial Bee Colony algorithm," Computational Social Networks, vol. 7, no. 1, p. 3, 2020.

[16] N. Fatehi, H. S. Shahhoseini, J. Wei, and C.-T. Chang, "An automata algorithm for generating trusted graphs in online social networks," Applied Soft Computing, vol. 118, p. 108475, 2022.

[17] C. Mallick, S. Mishra, and M. R. Senapati, "A cooperative deep learning model for fake news detection in online social networks," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 4, pp. 4451-4460, 2023.

[18] M. S. Vaghefi, D. L. Nazareth, S. P. Nerur, and K.-Y. Chen, "Self-disclosure in online social networks: An empirical study of location-based check-ins and impression management," Information & Management, vol. 61, no. 7, p. 104017, 2024.

[19] W. H. Bangyal, S. Amina, R. Shakir, G. Ubakanma, and M. Iqbal, "Using Deep Learning Models for COVID-19 Related Sentiment Analysis on Twitter Data," in 2023 International Conference on Human-Centered Cognitive Systems (HCCS), 2023: IEEE, pp. 1-6.

[20] W. H. Bangyal, M. Iqbal, A. Bashir, and G. Ubakanma, "Polarity Classification of Twitter Data Using Machine Learning Approach," in 2023 International Conference on Human-Centered Cognitive Systems (HCCS), 2023: IEEE, pp. 1-6.

[21] W. H. Bangyal et al., "Detection of Fake News Text Classification on COVID-19 Using Deep Learning Approaches," Computational and mathematical methods in medicine, vol. 2021, no. 1, p. 5514220, 2021.

[22] L. Abualigah, D. Yousri, M. Abd Elaziz, A. A. Ewees, M. A. Al-Qaness, and A. H. Gandomi, "Aquila optimizer: a novel meta-heuristic optimization algorithm," Computers & Industrial Engineering, vol. 157, p. 107250, 2021.

[23] 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.

[24] A. G. Hussien, "An enhanced opposition-based salp swarm algorithm for global optimization and engineering problems," Journal of Ambient Intelligence and Humanized Computing, vol. 13, no. 1, pp. 129-150, 2022.

[25] A. G. Hussien and M. Amin, "A self-adaptive Harris Hawks optimization algorithm with opposition-based learning and chaotic local search strategy for global optimization and feature selection," International Journal of Machine Learning and Cybernetics, vol. 13, no. 2, pp. 309-336, 2022.

[26] Y.-J. Zhang, Y.-F. Wang, Y.-X. Yan, J. Zhao, and Z.-M. Gao, "LMRAOA: An improved arithmetic optimization algorithm with multi-leader and high-speed jumping based on opposition-based learning solving engineering and numerical problems," Alexandria Engineering Journal, vol. 61, no. 12, pp. 12367-12403, 2022.

[27] J. Zhao, Y. Zhang, S. Li, Y. Wang, Y. Yan, and Z. Gao, "A chaotic self-adaptive JAYA algorithm for parameter extraction of photovoltaic models," Math. Biosci. Eng, vol. 19, no. 6, pp. 5638-5670, 2022.

[28] Y. Yu, S. Gao, S. Cheng, Y. Wang, S. Song, and F. Yuan, "CBSO: a memetic brain storm optimization with chaotic local search," Memetic Computing, vol. 10, pp. 353-367, 2018.

[29] H. Chen, Y. Xu, M. Wang, and X. Zhao, "A balanced whale optimization algorithm for constrained engineering design problems," Applied Mathematical Modelling, vol. 71, pp. 45-59, 2019.

[30] H. Zhang et al., "Ensemble mutation-driven salp swarm algorithm with restart mechanism: Framework and fundamental analysis," Expert Systems with Applications, vol. 165, p. 113897, 2021.

Math Role-Play Game Using Lehmer's RNG Algorithm

Chong Bin Yong[1], Rajermani Thinakaran[2], Nurul Halimatul Asmak Ismail[3], Samer A. B. Awwad[4]
Faculty of Data Science and Information Technology, INTI International University, Negeri Sembilan, Malaysia[1, 2]
Department of Computer Science-Applied College, Princess Nourah bint Abdulrahman University,
Riyadh, Kingdom of Saudi Arabia[3]
Quality, Risk and Business Continuity Department, Imam Muhammad Ibn Saud Islamic University, Kingdom of Saudi Arabia[4]

***Abstract*—Due to the COVID-19 pandemic, schools in Malaysia have been physically closed for more than 40 weeks and the students have to learn online. As Malaysia transitions to endemicity, many younger students struggle to keep up with their education due to significant learning loss caused by school closures and the challenges of virtual classes, including distractions and reduced engagement. This study aims to address these issues by developing an educational application that integrates gaming elements, focusing on arithmetic for Year 6 primary school students. The application engages students through interactive gameplay, requiring them to solve math problems to progress, thereby promoting a fun and effective way to enhance their arithmetic skills and mitigate learning loss.**

***Keywords*—Lehmer's RNG algorithm; online educational; gamification**

I. Introduction

COVID-19 pandemic has changed to being endemic and become more common, students start to return to school and learn face-to-face in physical class. Before that, the schools are closed and students conducted their classes with online learning. Some teachers will ask the students to open the webcam but most of them did not know what the students doing behind the screen. After a long time of school closure, a huge problem was occurred and be discovered which is 'Learning Loss' [1]. Due to the lengthy summer break, when kids did not learn for a while and were unfamiliar, learning loss first occurred. Even though pupils had been learning online, there was still a problem with learning loss when the school closed. Numerous academics have provided compelling data and statistics to support their claims that this issue exists, showing that student test scores are lower than they were prior to the epidemic [2-5].

The issue of learning loss is crucial and demands attention since it has a significant negative impact on Malaysian students, particularly those from less-educated or ignorant parents. Some of the B40 (the bottom 40% of income earners) family's youngsters find it challenging to participate in online learning in Malaysia [1]. The goal of online learning is to help students finish their coursework on time, yet they learn less and find it difficult to make up on their learning after 40 weeks of break from class. One of the reasons is that it is difficult to maintain elementary school children' interest in online learning since they lack the capacity to seek out information on their own. Additionally, because they are not being watched, most pupils rarely study independently when under lockdown. The pupils in primary school may have not be able to achieve the desired outcomes, and this problem may continue to secondary school [5].

The objective of this paper is to create a learning game for Malaysian Year 6 primary school pupils. The application, which will be deployed on a personal computer (PC) platform, will concentrate on the math curriculum's arithmetic topic. The math for grade six comprises fraction and decimal multiplication and division. There are many degrees of difficulty for the application, and English is the in-game language. Playing the game and responding to the questions helps students gain arithmetic proficiency. A captivating tale will be created for the Role-Playing Game (RPG). In order to save the princess, students will take on the role of an explorer and attempt to rebel against the demon king. To harm foes, students must correctly respond to the question they are provided by the adversaries within the allotted time. Students will be attacked by the opponents if they fail to respond to the question or take too long. The student either wins the game by killing the final boss or loses the game when their health reaches 0. To get pupils to play the game, a compelling tale is necessary. The game is designed to integrate educational content seamlessly with engaging gameplay, ensuring students remain motivated to learn while having fun. Built-in feedback mechanisms will help students track their progress and identify areas for improvement. The adaptive difficulty levels ensure that the game challenges students based on their proficiency, promoting continuous learning. Java and the NetBeans IDE will be utilized in this study's application development. Additionally, usability testing will be conducted to evaluate the game's effectiveness in improving students' arithmetic skills.

II. Literature Review

A. Current Situation of Education

Today, every child must have access to education. Every element of life is inspired by education, which also provides the road for personal growth. From their education, students may assure a bright future, develop knowledge, and boost their confidence [6]. According to the United Nations International Children's Fund (UNICEF), the closing of schools across the world has an impact on, puts students at danger, or causes them to lag behind [7]. Additionally, more than 1.5 billion children were affected by the epidemic, and the most vulnerable kids suffered grave consequences, according to a United Nations

assessment [8]. The number of students worldwide who will be impacted by school closings in 2020 is shown in Fig. 1.

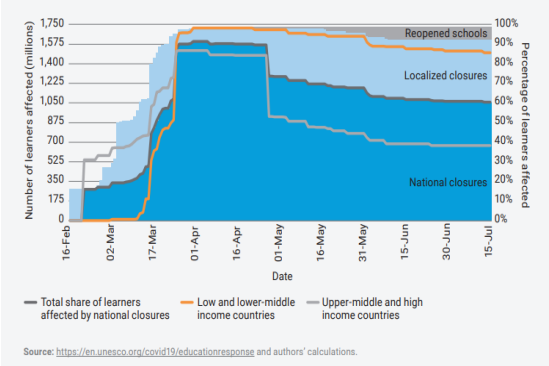


Fig. 1. Number of children affected by school closures globally.

In Malaysia, the government has mandated that students remain at home and do their coursework online [9]. The Malaysian government has made significant contributions to education by giving B40 families and students free internet connection so they may do their coursework online [10]. Although a great alternative to traditional classroom instruction, online learning still has several drawbacks, like restricted Internet accessibility, difficulty in interacting with others, a lack of learning materials, and more. These problems have an impact on students' educational experiences, particularly in some rural parts of Malaysia [11] brought the learning loss problem in result.

Although not a new issue, learning loss manifests itself more frequently in another way. Learning loss originally happened in the summer and was referred to as "Summer learning loss" since some countries, including the United States and Canada, have extensive summer vacations. The prolonged summer break disrupts the instructional pattern, and pupils risk forgetting what they have learned [12]. Despite the COVID-19 epidemic, students were still able to learn online, but there were still issues with learning loss. The next sections will cover several studies and publications that use a lot of statistics to support their arguments. Due to the closing of the schools, these two learning deficits have comparable features. Online learning needs self-discipline because pupils can't focus on the screens [13]. Additionally, it may be challenging for kids from low-income households to access the Internet, which has an impact on their ability to study [14]. The problem of learning loss during the pandemic was created by the closing of schools and a shift in the teaching approach.

Numerous studies using extensive data have found the impacts of learning loss. The majority of the researchers that analysed the test scores of pupils before and after the shutdown of the schools discovered evidence of student learning loss. Additionally, several research noted the rise in inequity, with some student groups suffering greater learning losses than others [15]. In order to evaluate the arithmetic, spelling, and reading exam results for primary students in grades 8 to 11 in the Netherlands, research was done, according to the Proceedings of the National Academy of Sciences of the United States of America (PNAS).

Despite learning remotely for eight weeks when schools were closed, the pupils' arithmetic and spelling test scores fell by 2.15 and 0.76 percentile points, respectively (Fig. 2). According to estimates based on the aforementioned finding, pupils generally underperformed by 3.16 percentiles, or 0.08 standard deviations (SDs). The impact of learning loss is 60% worse for pupils from less-educated parents where it is concentrated [2].

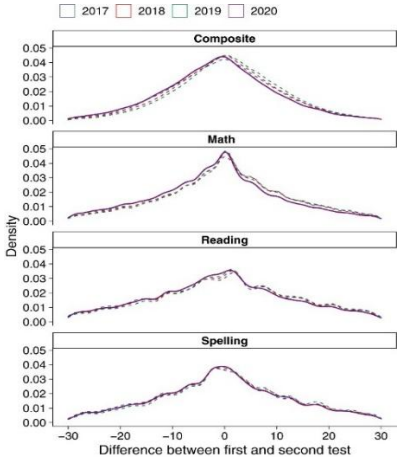


Fig. 2. Difference in test scores of years 2017 to 2020.

Additionally, another study used data from 5.4 million American kids in grades 3 through 8 who took arithmetic and reading tests during the first two years of the epidemic to conduct research. Between autumn 2021 and fall 2019, students' average math exam scores declined from 0.20 to 0.27 SDs. The reading test scores are 0.09 to 0.18 SDs lower (Fig. 3), and the difference in performance between kids attending primary schools with low and high levels of poverty has grown by 0.10-0.20 SDs [16].

Although the school closures took place in the early stages of the epidemic, this study did not take into account whether the pupils had engaged in remote learning. The arithmetic score has decreased for two years, while the reading score has decreased between the fall of 2020 and the fall of 2021 [3].

According to another news from UNICEF, school closures resulted in significant learning losses in math and reading plus the learning losses are roughly proportional to the length of closures in some countries. Students aged 10 to 15 has significant learning losses in math and reading from the results of two states in Mexico. Many other countries also faced this problem, for instance, South Africa, India, Pakistan and Brazil [8]. Since the schools have been closed for about 40 weeks in Malaysia, the learning loss on Malaysia students must be more serious [1].

The causes of learning loss might vary depending on the circumstances. It is clear from this experiment that the COVID-19 pandemic's propagation is its primary cause. To protect pupils from the pandemic's risks, the majority of governments have stated that schools would be closed. The solution is then put into place to allow the students to continue their education—remote learning or online learning. The aforementioned elements are closely related to one another. Zhdanov et al. (2022) categorized the factors impacting the learning losses as

the following themes: “change in teaching methods”, “opportunities to reach education”, “less time for learning”, “less control / feedback” and “emotional factors” (Fig. 4) [17].

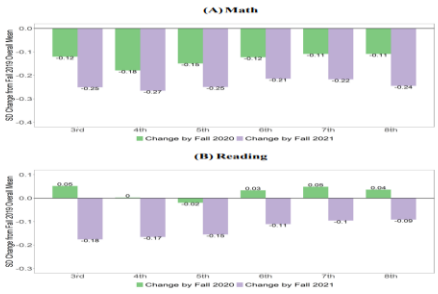


Fig. 3. Trends in test scores in fall 2020 and fall 2021 compared with fall 2019.

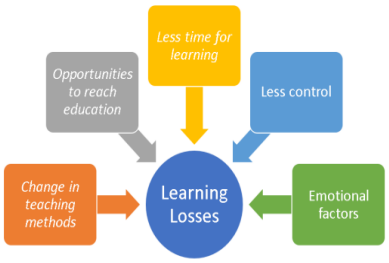


Fig. 4. Affecting factors to learning losses during COVID-19.

B. Maintaining the Integrity of the Specifications

Many schools, colleges, and universities have also begun to employ e-learning apps or platforms in the education in order to reduce the pandemic-related learning loss and improve the learning experiences. The students may always remain linked, from any location. As a result, its importance and utility have increased as a result of the COVID-19 epidemic.

Additionally, a lot of schools, colleges, and universities have started implementing e-learning platforms or applications in the classroom to reduce the learning loss caused by the epidemic and enhance the learning experiences. The students can always keep in touch, from any location. As a result, the COVID-19 pandemic's emergence made it more important and valuable. When it is discovered that the pupils are having academic difficulties, it will already be too late because there is no single evaluation to examine the students until form 5 since the two public examinations, Primary School Achievement Test (UPSR) and Form Three Assessment (PT3) have been abolishment [18].

Numbers and their operations are the subject of the scientific discipline known as mathematics [19]. Galileo defined mathematics as "the language in which God has written the cosmos" in his definition. He believes that learning math is necessary to comprehend the universe and that it is an essential language [20]. In addition, as previously noted, learning loss has a significant influence on the arithmetic subject in basic education. As a result, in the current context, minimizing and reversing the learning loss is crucial.

C. Gamification with Education

A game is a playing activity that is used for fun or learning. People are under stress as a result of the COVID-19 outbreak since they are unable to leave their homes. They begin playing a game for enjoyment and relaxation. Game and gamification are two distinct concepts. Gamification is the process of applying game concepts and components to non-game activities [21]. Combining education with games creates gamification in education, which mostly occurs in the context of online learning. As a result, it is an appropriate response to the current circumstance, and pupils are accustomed to this kind of instruction. Gamification in education may enhance the learning process, grab students' attention, and make the subject matter more entertaining [22].

In Indonesia, Udjaja, Guizot, and Chandra (2018) use gamification to enhance learning in elementary math. They develop an interactive learning game by fusing gaming aspects with math resources. They seek to make math learning more engaging so that children can readily comprehend the arithmetic information. Consequently, both teachers and students find the game-based learning approach to be engaging. Most pupils are now more interested in learning math, and they see the value in doing so [23].

D. Game Genres

Table I categorizes various game genres and provides a brief description of their key features. Each genre emphasizes different gameplay mechanics, offering unique player experiences. For example, action games focus on combat and intensity, while puzzle games emphasize critical thinking and problem-solving. Role-playing games immerse players in characters and storylines, while sandbox games allow open-world exploration and creativity. These genres highlight the diversity in game design, catering to a wide range of player preferences and objectives [24-25].

TABLE I. GENRES AND DESCRIPTION OF GAMES

Game Genres	Description
Action	Games with action intensity and combat as its main attraction.
Adventure	Games that focus on exploration and puzzle solving.
Multiplayer online battle area (MOBA)	Games that place the focus on player control, map and resource management.
Puzzle	Games with puzzles as its key game mechanic.
Real-time strategy (RTS)	Games that include real-time reasoning and problem solving.
Role-playing	Games that allow players to take control of a character and immerse in their situation.
Sandbox	Games that associate with player choice, open environments and non-linear gameplay.
Simulation	Games that focus on creating a game world and it is matching with the real-world situation.
Shooters	Games with different weapons for aiming and shooting at targets.
Survival	Games that focus on resource management to keep the player character alive.
Platformer	Games with 2D side-scrollers and simple controls.

However, in order to implement the gamification in education, it is important to know that not every game type is suitable to the teaching environment. Amory et al. (1999) have conducted research to discover the most suitable game type and game elements. They tested on the adventure, RTS, shooting and simulation games. They found out that the students prefer playing the adventure and RTS rather than the other two. This may be caused by the user interface and game play. The research also shown that login, memory, visualization, mathematics, reflexes and problem solving are the important game elements [26].

Instead, eleven game types identified by [24] and [25], role-playing game and sandbox game has been implemented in the education. Role-playing game has been used to teach vocabulary [27] and Japanese [28]. Also, it is proved that the role-playing games are able to improve the knowledge and encourage students to learn. The game was utilized as a learning material in the class, provide a good environment and experience to students. The example of the sandbox game is education edition of Minecraft. It allows the users to gather resource and use the resource to create anything they want. It motivates the creativity of students and can be used to teach various subject, for instance, math, history and visual arts. It is proved that the math ratings of 4th grade students in Australian schools have increased when using Minecraft for math's learning [29].

Furthermore, Hussein et al. (2019) conduct research of the past relevant literature between 2006 to 2017 years to find out the effects of educational gaming in teaching science of primary levels. They discovered that RPG is the most popular game genre. Students are able to control the game avatar and communicate with Non-Player Character (NPC) which allow the students immerse in the learning environment [30]. Another research also conducts research of the past relevant literature but from 2016 to 2020 finds out that adventure games is the most preferred game genre while sport and simulation, RPG and puzzle games are often used [31] (Fig. 5).

Hassan, Mailok and Hashim (2019) conducts research on the relationship between gender and game genres by collecting data from Diploma students in Malaysia. In overall, the most popular one is Adventure games. Adventure and puzzle games are the most popular game genre for female while action and strategy games are the most popular game genre for male. Male do not prefer music or dance games while female do not prefer cross genre games [32]. Fig. 6 shows the table of the game genres selection based on gender.

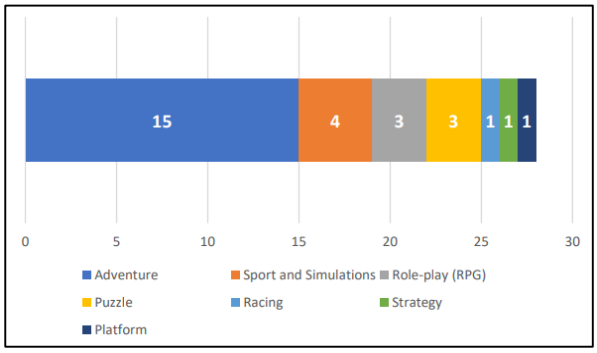


Fig. 5. Frequency of game genres.

No	Game Genres	Female	Percentages (%)	Male	Percentages (%)	Total	Percentages (%)
1	Adventure	26	63.4	57	76.0	83	71.6
2	Board/card	17	41.5	27	36.0	44	37.9
3	Puzzle	26	63.4	27	36.0	53	45.7
4	Platform	20	48.8	34	45.3	54	46.6
5	Strategy	23	56.1	58	77.3	81	69.8
6	Sport	15	36.6	42	56.0	57	49.1
7	Action	21	51.2	58	77.3	79	68.1
8	Shooting	24	58.5	51	68.0	74	64.7
9	Simulation	15	36.6	30	40	45	38.8
10	RPG	10	24.4	42	56.0	52	44.8
11	Music/Dance	23	56.1	22	29.3	45	38.8
12	Cross Genre	9	22.0	29	38.7	38	32.8

Fig. 6. Game genre selection based on gender.

To develop a gaming application, the problem statement, objective and scope must be clearly determined. Then, developer has to create the game design or prototype. Developer can start coding the application when all the details are finalized. The application will then test by the developers and external testers. Finally, the application can be released after testing [33]. On the other hand, [23] implement the game development life cycle (GDLC) to develop the math education game. It consists of initiation, pre-production, production, testing and release. Initiation phase is to create the game concept with simple description. Pre-production phase is to create and review the game design plus create a game prototype. Production phase is to produce the game assets and coding. Testing phase is to test the functionality and usability of game internally. Release phase is the final stage which the game is ready to be released to public. Tan and Suparjoh (2022) also implement GDLC to create the educational game for mobile but it has one more, beta phase before the application is published. After the internal testing, the game will be tested by third-party or external tester in beta phase [33-34].

E. Game Algorithm

Other than that, algorithms can be used in the production phase in order to make the gaming application more effective. Developers may use algorithms and implement them into the application. Candra et al. (2021) used A-star (A*) algorithm to design the pathfinding and develop a tree planting game. A* algorithm uses the heuristics function to plan a path between multiple nodes efficiently. It uses the best-first search and finds the shortest path by calculating the cost from initial node to destination node [35]. Galam et al. (2019) conduct research on the performance A* algorithm and Dijkstra's algorithm. Dijkstra's algorithm is also a pathfinding algorithm that use to search for the shortest path but it has to build the shortest paths to all other nodes in the graph which is not efficient. Therefore, 68 uses the simpler A* algorithm to create the tower-defence game because he or she thinks A* algorithm can be considered as a low-fill version of Dijkstra's algorithm. In conclusion, the A* algorithm is frequently used in game development [39].

Also, Lehmer's Random Number Generation (RNG) algorithm is used to generate a simplified initial world state in a role-playing game [36]. The researcher compares the RNG algorithm with the random number generators inside the C++ library and discover that the average speed of the algorithm is faster than the other methods (Fig. 7). Thus, it will be better to include the RNG algorithm in order to improve the software performance.

Lehmer's RNG algorithm is based on a general formula which is 'Xk+1 = a · Xk mod m'. The multiplier 'a' is an element

of high multiplicate order of modulo 'm' and the modulus 'm' should be a prime number or power of prime number. The seed 'X0' is coprime to 'm' and has to be set at the beginning along with other variables. Fig. 8 shows the code example of Lehmer's RNG algorithm in C#.

Alhassan et al. (2019) and da Silva and Villela (2016) use the breadth-first search (BFS) algorithm to generate and expand nodes of a tree data structure. It includes a simple first in first out (FIFO) queue which mean the parents of same level will be checked before moving to their children. All the node of level n will be expanded before any node of level n+1. Alhassan et al. (2019) uses it to create a puzzle game whereas da Silva and Villela (2016) uses it to create a mobile game [37-38].

Zain et al. (2020) uses Minimax search algorithm and Alpha-Beta pruning algorithm to develop a tic-tac-toe game. Minimax search algorithm is a backtracking algorithm based on the game theory, decision theory, statistics and philosophy [40]. It uses to find the optimal path and maximize the player's chance to win whereas minimize the opponent's chance to win. Alpha-Beta pruning algorithm is an optimization algorithm for the minimax algorithm [40]. It reduces the computation time by removing all the unnecessary bunches when it finds a better path. In minimax algorithm, it will continue evaluating all the nodes until it reaches the root node. Monte Carlo algorithm is used to develop card game and puzzle game [41-42]. Monte Carlo algorithm is a searching algorithm that conducts simulations to evaluate unknown patterns or unknown actions to know their values by using a rollout policy [41]. It requires a lot of information to make the simulation.

Method	Avg. Time
rand()	0.001160
std::random	0.067793
Lehmer	0.000269

Fig. 7. Average speed of RNG methods.

```
public class LehmerRng
{
  private const int a = 16807;
  private const int m = 2147483647;
  private const int q = 127773;
  private const int r = 2836;
  private int seed;
  public LehmerRng(int seed)
  {
    if (seed <= 0 || seed == int.MaxValue)
      throw new Exception("Bad seed");
    this.seed = seed;
  }
  public double Next()
  {
    int hi = seed / q;
    int lo = seed % q;
    seed = (a * lo) - (r * hi);
    if (seed <= 0)
      seed = seed + m;
    return (seed * 1.0) / m;
  }
}
```

Fig. 8. Code example of Lehmer's RNG algorithm.

Table II shows the summary of algorithms that can be used for game development. It includes the name and the basic function of the game development algorithms. Hence, specifically tailored to address key challenges in game development. These algorithms ensure efficiency, improve gameplay mechanics, and enhance user experience by solving complex problems like shortest path searches, and probabilistic outcomes in a structured and reliable manner.

TABLE II. USES OF ALGORITHMS IN GAME DEVELOPMENT

Game Development Algorithm	Description
A* algorithm	To search the shortest path between the initial and final state.
Lehmer's RNG algorithm	To get a random number by using the equation. ($X_{k+1} = a \cdot X_k \bmod m$)
Breadth-first search (BFS) algorithm	To search a node in the tree structure level by level.
Dijkstra's algorithm	To find the shortest path between nodes in a graph.
Minimax search algorithm	To find the optimal move by decision making and game theory.
Alpha-Beta pruning algorithm	To optimize the minimax algorithm by decrease the evaluation times.
Monte Carlo algorithm	To estimate the possible outcomes of an uncertain events.

III. METHODOLOGY

There are three phases in the game education application development. The first phase is on collecting data from the user stakeholders. Questionnaire will be used as quantitative research method to get the basic information. The target audiences for the questionnaire include the year 6 students, educators, parents and guardians in Malaysia. Moreover, interview will be conducted as qualitative research method with selected students to have in-depth understanding. The collected data will be used to identify the learning experience, learning performance, the needs of learning materials and more.

The second phase will be based on software development. Waterfall methodology will be used to develop the application. It consists of five phases which are requirements analysis, design, implementation, testing, deployment, and maintenance. The author will start analyzing the requirements after data collection and create an application design. Then, the author will code and test the application to ensure its performance. The application will then deploy to the target audience and be maintained.

The third phase is to test the application with both functional testing and usability testing. Functional testing is used to check the errors or bugs within the application by implementing the black box testing. The output will be compared with expected result and they must be same. Usability testing is used to test whether the application can help the students to improve their math skills. Testers will have to do a short math test before and after they play the game. Both testing will be conducted during the testing phase in waterfall methodology.

F. Phase 1 – Data Collection

Quantitative research is a type of research that uses to collect information that can be measured and analyzed. It focuses on data that is structured and can be represented numerically [43]. This study will use questionnaire as the research approach to collect data from the persons that related to the students, such as, educators, parents and guardians. The questionnaire is estimated to have 12 questions. Qualitative research is a type of research that aims to find out about people's experiences and help to understand what is important for people. It concerns with subjective 'meanings' rather than objective 'facts' [44]. This study will use interview method as the research approach and conduct interview with three Year 6 primary school students.

Different perspective can be gotten from the interviewees about the learning performance of students.

G. Phase 2 – Waterfall Methodology for Software Development

Waterfall methodology consists of five phases which are requirements analysis, design, implementation, testing, deployment and maintenance. It is one of the software development life cycle models that allow us to plan the development process in a systematic manner.

- Phase 1 – Requirements analysis focus on capturing the requirements or problems faced by the target audience. The objective is to understand all the requirements which includes scope, costs, timelines and limitation. This phase involves gathering input from stakeholders, teachers and students, to ensure the application meets their needs. Additionally, it identifies potential challenges, such as technical constraints or user accessibility, to ensure a clear development roadmap.
- Phase 2 – During design phase, programmer has to decide a plan for the solution. An application design will be created in this study.
- Phase 3 – Implementation is the phase that the coding start. Software design has been confirmed in the previous phase and programmer start to develop the application based on that design. The application use Java as the programming language and NetBeans IDE as the platform for development.
- Phase 4 – The aim of testing phase is to test the application so that it will achieve the expected result. Also, the application will be tested to find out there is any bug or not. Programmer has to debug if there are errors or bugs in the application.
- Phase 5 – The application must deploy to the target audience and need to be maintained in order to have improvement in the future if necessary.

H. Phase 3 – Software Testing

Functional testing is a type of software testing that use to verify and validate the functional requirements of the software. It involves black box testing and every function will be tested by entering the input and comparing the output with the expected result. The application should able to generate the question automatically and know the answer entered by the user is right or wrong. Usability testing is used to test whether the application can improve the knowledge. Hence, the target audience is Year 6 primary school student. If the students are unable to learn from the application, then the expected result of the study cannot be achieved.

IV. Design and Development

The target audience, users, and stakeholders were surveyed by the author to determine the needs and expectations for the system. The data gathered can be used to comprehend fundamental needs and identify potential issues. Functional requirements, non-functional requirements, and user requirements can all be improved after examining the replies. The design of the suggested educational gaming application is then displayed using UML diagrams, including use case diagrams, activity diagrams, sequence diagrams, and more.

Questionnaire and interview are conducted before the design and development. Four components make up the questionnaire. The target audience, year 6 primary school pupils, and stakeholders, such as teachers, parents, and guardians, were all included in the first section's demographic information gathering. In the second portion, the author discusses how pupils learn and how they perform before and after school closures. The third segment is where the student's experiences and facts regarding the gaming are gathered. Respondents are prompted to react with their thoughts on the educational gaming applications in the last section.

Besides that, like the questionnaire, the interview questions are divided into four pieces. The interviewees' demographic data is gathered in the first phase. It is intended to identify the interviewees' learning performance and experience in the second portion. The third piece asks the interviewers about their gaming history, while the last section collects their system needs and expectations. Initial requirements for the game education application are recognized and defined based on questionnaire and interview responses as well as the author's gaming history. Table III shows the functional requirements, non-functional requirements, and usability requirements.

TABLE III. Application Requirements

Functional Requirements	• Main menu with play, load, control and quit buttons • In-game setting menu with save game, control and exit game functions • Character control • Turn-based battle system with math arithmetic questions • Inventory system • Level-based progression with health point (HP) and experience point (XP) • Display game scenes • Difficulty levels • Game clearance timer • Random questions with random number generation (RNG) algorithm
Non-functional Requirements	• Provide accurate questions and examine the answers correctly • Provide fast response on player control • User-friendly
Usability Requirements	• Basic animation in combat scene • Basic examples of questions in menu

I. Design

(Fig. 9-11) illustrate aid in defining the gaming education application's architecture, features, and functionalities.

J. Development

There are many scenes in the game education application and the author uses state engine to differentiate them. The state engine includes ‘titlestate’, ‘playstate’, ‘optionstate’,

'dialoguestate', 'combatstate', 'characterstate', 'victorystate' and 'gameoverstate'.

Title screen is the first screen when the users open the application. Users can start a new game by pressing 'New Game' button, continue previous game by pressing 'Load Game' button, find out the game control and question examples by pressing 'How to Play' button and close the application by pressing 'Quit' button. The users require to select difficulty after pressing 'New Game' button. Difficulty levels will affect the question complexity and the time given for users to answer a question. 'Easy' is 90 seconds, 'Normal' is 135 seconds, and 'Hard' is 180 seconds.

After pressing the 'New Game' button or 'Load Game' button, the application will change to play state. In this state, the application will render and display the tile map, character, non-player character (NPC) and more. Option screen are shown when the state engine change to option state by pressing 'Esc' on keyboard. It will open the interface of options menu and allow the users to save, control the volume of music and sound effect, change difficulty level, read game story, game control, question examples and end game. When the users press the buttons of 'Difficulty Level', 'Game Story', 'Game Control' and 'Question Examples', it will open the same interfaces as the title screen.

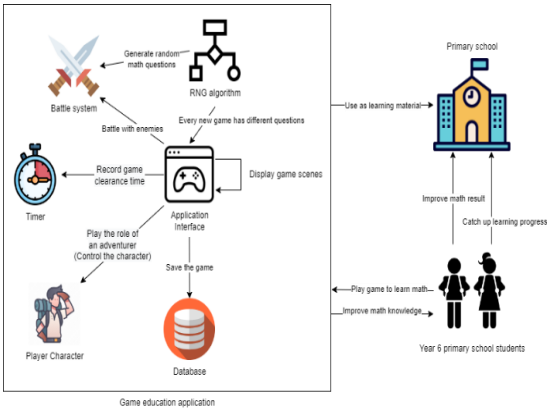


Fig. 9. Rich picture diagram.

```
public float uniform()
{
        System.out.println(r_seed);

    // highest and lowest for a rar
    long hi = r_seed / q;
    long lo = r_seed - q * hi;
      long lo = r_seed % q;

    // calculate random number
    long t = (a * lo) - (r * hi);

    // if positive
    if (t > 0)
    {
        r_seed = t;
    }
    else
    {
        r_seed = t + m;
    }

      System.out.println(r_seed);
    return r_seed;
}
```

Fig. 10. RNG algorithm.

Moreover, users can enter the dialogue screen by pressing 'E' on keyboard while the character is beside the NPC. The state engine will change to combat state when users can press 'Enter' on keyboard while the character is beside the enemy. The health points of enemy and character are shown in the combat screen. Users can choose to 'Attack', 'Skill', 'Item' or 'Run' in the combat. When the users press the 'Run' button, it will back to the map and change to the play state.

The RNG algorithm's Java code is seen in Fig. 10 The most crucial component of the application that enables it to produce unique numbers each time is this one. A variety of questions will be constructed using the generated numbers once they have been put in an array. Users will get the important objects after defeating the enemies. Users must keep answering the questions correctly until the health points of enemy goes to zero. Furthermore, users can enter the character screen by pressing 'Q' on keyboard in the play screen. The interface will display the level, health point, attack damage, current experience, experience to level up and inventory.

When the users clear the game, the state engine will change to victory state and show the interface of victory screen. The game clearance time will be displayed to let the users know how much time they spend to clear the game. Users can press the 'Quit' button to go back to the title screen. When the health point of character become zero because users did not answer the question correctly or overtime, the state engine will change to game over state and show the interface of game over screen. Users are able to press the 'Retry' button to go back to the play screen or press the 'Quit' button to go to the title screen.

In addition, the application is connected to the MySQL database with phpMyAdmin at localhost by using XAMPP. The application will store the data of current date, time, game clearance time, victory or game over, difficulty level, number of questions answered correctly, total number of questions generated in the game. The number of questions is separated according to the question types in the database. Fig. 11 depicted below shows the game interface of the application.

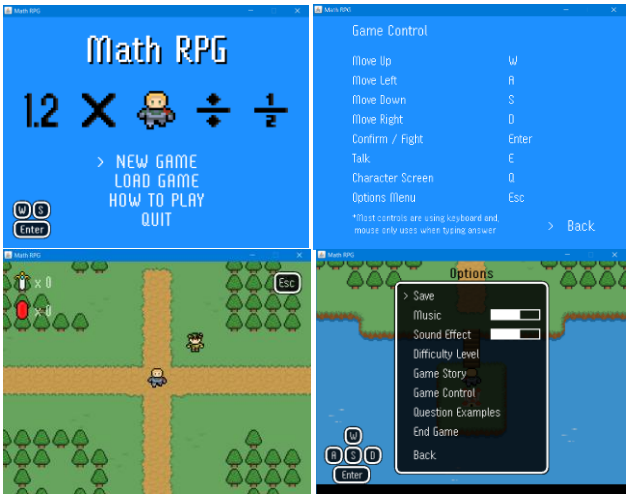


Fig. 11. Game interface.

V. TESTING

In order to make sure the game education application can fulfill the aim and criteria; the author will carry out functional and usability testing on it.

A. Functional Testing

Black box testing is used in functional testing to identify errors or bugs within the application, ensuring the predicted and actual outcomes match. Test cases enable developers to verify that various system features are functioning as expected and that the system meets the requirements. Table IV – VII presents the test cases conducted using the application, highlighting the thorough approach taken to ensure system reliability. This method ensures that critical features are tested rigorously, providing confidence in the system's performance. Additionally, the testing process helps in uncovering hidden issues that could impact user experience.

TABLE IV. TEST CASES

Tester	Chong Bin Yong
Test Date	11/20/2022
Application Developer	Chong Bin Yong

TABLE V. TEST CASE 1

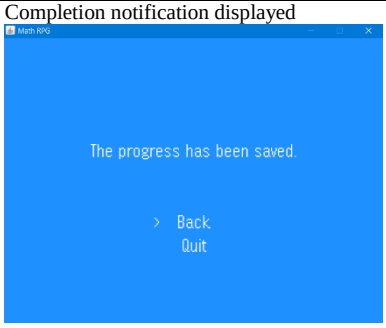
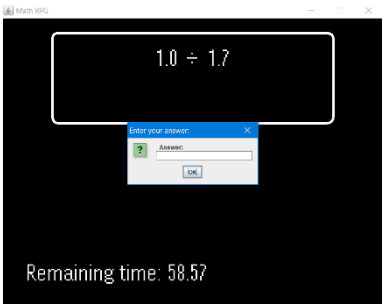
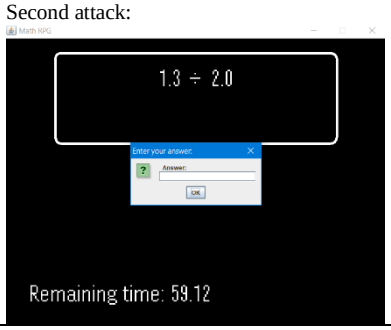
Test Objective	Validate the 'Save' button in options menu
Potential Test Inputs	1. Press 'Save' button
Expected Test Outputs	Completion notification displayed
Test Procedures	1. Enter the game 2. Press 'Esc' to open options menu 3. Press 'Save' button 4. View the screen displayed
Actual Test Results	Completion notification displayed 

TABLE VI. TEST CASE 2

Test Objective	Validate the 'Load Game' button in title screen
Potential Test Inputs	1. Press 'Load Game' button
Expected Test Outputs	Enter the game with previous saved process
Test Procedures	1. Press 'Load Game' button 2. View the screen displayed
Actual Test Results	The saved process is loaded while entering the game 

TABLE VII. TEST CASE 3

Test Objective	To test the random questions generator
Potential Test Inputs	1. Press 'Attack' button in combat scene
Expected Test Outputs	Different questions are generated each time
Test Procedures	1. Enter the game 2. Combat with one of the enemies 3. Press 'Attack' button 4. View the screen displayed
Actual Test Results	The questions are different every time the user wants to attack First attack:  Second attack: 

B. Usability Testing

Usability testing is performed to determine whether a game-based educational application can assist students in developing their mathematical abilities. After playing the game, test subjects were asked to fill out a questionnaire, and some testers had to do a quick arithmetic exam both before and after playing the game. The usability testing involved four Year Six primary school children as participants. Each participant completed the questionnaire, and two of them conducted the short math test for both pre-test and post-test evaluations.

The questionnaire focused on assessing the user interface, engagement level, and overall game effectiveness in enhancing learning. The pre-test and post-test results were analyzed to determine the improvement in arithmetic skills after using the application. These results provided valuable insights into the application's impact on learning outcomes and highlighted areas for further refinement. Fig. 12 illustrates the usability testing process for Participant 1, showcasing the steps taken during the evaluation.

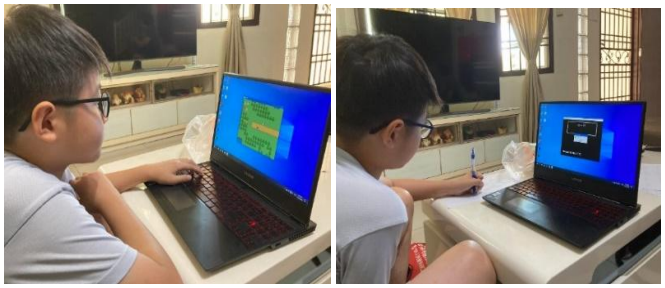


Fig. 12. Usability testing of participant 1.

The results of the post-test indicated an improvement in mathematical skills after interacting with the game. Table VIII – IX show the pre-test and post-test for the participants that have been using the application.

TABLE VIII. PRE-TEST

Questions	*Answers*	*Participant 1*	*Participant 2*
Q1. 5.2 × 2.7	A1. 14.04	14.04 (Correct)	14.04 (Correct)
Q2. 6.2 ÷ 9.8	A2. 0.63	No answer (Wrong)	No answer (Wrong)
Q3. $\frac{5}{9} \times \frac{1}{7}$	A3. $\frac{5}{63}$	$\frac{5}{63}$ (Correct)	No answer (Wrong)
Q4. $\frac{8}{9} \div \frac{3}{10}$	A4. $\frac{80}{27}$	$\frac{80}{27}$ (Correct)	No answer (Wrong)
Q5. 4.5 × 2.5 × 2.0	A5. 22.5	3.65 (Wrong)	No answer (Wrong)
Q6. 5.7 ÷ 4.6 ÷ 4.9	A6. 0.25	No answer (Wrong)	No answer (Wrong)
Q7. $\frac{5}{7} \times \frac{5}{4} \times \frac{9}{4}$	A7. $\frac{225}{112}$	No answer (Wrong)	No answer (Wrong)
Q8. $\frac{10}{9} \div \frac{1}{3} \div \frac{5}{8}$	A8. $\frac{16}{3}$	No answer (Wrong)	No answer (Wrong)

After the game, a brief arithmetic exam is administered to participants to determine their math proficiency. Eight arithmetic problems covering fraction and decimal multiplication and division were also included in the quick math test. Although the questions alter from the ones on the pre-test math test, the question format remains the same. The participant is given the same amount of time—5 minutes—to do the brief math test. Table IX shows the results of post-test.

TABLE IX. POST-TEST

Questions	Answers	Participant 1	Participant 2
Q1. 7.7 × 3.7	A1. 28.49	28.49 (Correct)	7.6 (Wrong)
Q2. 8.6 ÷ 2.2	A2. 3.91	4.0 (Wrong)	0.3 (Wrong)
Q3. $\frac{6}{4} \times \frac{5}{4}$	A3. $\frac{15}{8}$	$\frac{15}{8}$ (Correct)	$\frac{13}{12}$ (Wrong)
Q4. $\frac{1}{8} \div \frac{2}{4}$	A4. $\frac{1}{4}$	$\frac{1}{4}$ (Correct)	No answer (Wrong)
Q5. 7.5 × 4.1 × 10.0	A5. 307.5	347.5 (Wrong)	No answer (Wrong)
Q6. 6.6 ÷ 5.5 ÷ 1.5	A6. 0.8	0.8 (Correct)	No answer (Wrong)
Q7. $\frac{10}{3} \times \frac{4}{7} \times \frac{5}{10}$	A7. $\frac{20}{21}$	$\frac{20}{21}$ (Correct)	No answer (Wrong)
Q8. $\frac{3}{7} \div \frac{6}{1} \div \frac{9}{1}$	A8. $\frac{1}{126}$	$\frac{18}{7}$ (Wrong)	No answer (Wrong)

Participant 1 answered 5 out of 8 questions correctly and his answers of questions 2 and 5 are also similar with the correct answers. Participant 2 did not answer any questions correctly. For participant 1, he is able improve the arithmetic skill after playing the game. For participant 2, he will try to answer the question after playing even though the answers are wrong.

In the opinion of author, participant 1 has the basic arithmetic knowledge and know how to do the questions before playing the game but participant 2 did not know how to do because of poor math knowledge. Therefore, participant 1 improved a lot after playing whereas participant 2 did not have much improvements.

A questionnaire is given for the participants to answer after playing the game in order to get their opinion or feedback related to the application. There are 25 questions in the questionnaire which first 20 questions are scale questions (1-5) and last 5 questions are open-ended questions. For the scale questions, 1 means 'Strongly disagree', 2 means 'Disagree', 3 means 'Neutral', 4 means 'Agree' and 5 means 'Strongly agree' while question 20 uses the scale as the ratings for the game application. Table X shows the questions and Table XI shows the results of the questionnaire.

TABLE X. QUESTIONNAIRE

Question	Description
Q1	I think that I would like to use the application.
Q2	I found the application unnecessarily complex.
Q3	I thought the application was easy to use.
Q4	I think that I would need the support of a technical person to be able to use this application.
Q5	I found the various functions in the application were well integrated.
Q6	I found there was too much inconsistency in this application.
Q7	I would imagine that most people would learn to use this application very quickly.
Q8	I found the application very awkward to use.
Q9	I felt very confident using the application.
Q10	I needed to learn a lot of things before I could get going with this application.
Q11	I am able to learn math quickly using this application.
Q12	I believe I score high marks after using this application
Q13	The application tells me my weakness of math.
Q14	The information provided with this application is clear.
Q15	It is easy to learn math using this application.
Q16	The information provided with the application is effective in helping me learn the math.
Q17	The interface of this application is pleasant.
Q18	This application has all the functions and capabilities I expect it to have.
Q19	I am satisfied with this application.
Q20	Rate this application as learning material.
Q21	What was the best thing about this application?
Q22	What was the worst thing about this application?
Q23	Do you understand what the game story is talking about?
Q24	How long do you think a math question will take to complete?
Q25	How many math questions do you think you have to do to learn it?

TABLE XI. RESULTS OF QUESTIONNAIRE

Questions	Participant 1	Participant 2	Participant 3	Participant 4
Q1	4	2	3	3
Q2	2	2	3	4
Q3	4	5	5	4
Q4	2	3	2	3
Q5	3	4	5	5
Q6	2	2	1	1
Q7	4	4	5	3
Q8	3	1	2	3
Q9	3	1	4	4
Q10	4	3	3	3
Q11	3	4	4	3
Q12	4	4	4	3
Q13	4	1	5	4
Q14	3	3	4	3
Q15	3	5	4	4
Q16	4	4	4	4
Q17	3	5	4	3
Q18	3	3	5	4
Q19	4	3	4	4
Q20	4	5	4	4
Q21	No opinion	The interface is nice	The application is easy to understand	Fun to play
Q22	No opinion	Questions are difficult	No opinion	A little hard to learn to use the application
Q23	Yes, roughly understand	Do not know much	Yes, roughly understand	Yes, able to understand
Q24	Between 1 minute and 1 and a half minutes	Around 3 minutes	Around 2 minutes	Around 1 and a half minutes
Q25	4 – 6 questions	20 questions	12 questions	10 questions

Overall, the participants are eager to use the game-based learning tool. They consider the application to be well-designed and simple to use. The software received a 4.25 average rating. Most users also think they can improve their arithmetic skills and test scores after using the application. They may learn math with the aid of the application, which also identifies users' areas of difficulty. In this study, the participants have a favorable attitude toward the game-based educational application. The amount of time they estimate it will take to finish a math problem is directly related to how many problems they estimate it will take to master it. By looking at the results of the pre- and post-tests for participants 1 and 2, it can be seen that these features are likewise influenced by the participants' arithmetic proficiency.

VI. CONCLUSION AND FUTURE WORK

After the application development and testing, it has been discovered that the application is not good enough because it is not very effective if the student is not good on doing arithmetic questions. The limitation of the application can be considered as future work. The author has planned the future work of the game education application. It is important to have a plan for the application development so that the users will continue to use the application and more users will be attracted to use it.

The authors will include application programming interface (API) into the system so that it can be run on the web browser and students can play the game online. A registration and login system will be built for students and educators to create their accounts and each account has its own ID. Students can play the game whereas educators can check the results according to the ID number of students after logging in. Besides that, the functions of the application will be improved and more functions will be added into the application. For example, users are allowed to interact to more game scenes and items in order to make the game more interesting. The users will traverse in different maps and they will be allowed to control the character by using both mouse and keyboard controls since the application has mostly used the keyboard controls so far.

Furthermore, the authors will add more game settings so the users will have more controls on the application. They can choose how long they want to spend on the question by separating the required completion time from the difficulty levels. Better game teaching will be also added instead of showing the game control only. More save slots and full screen mode will be added for users to choose in the application. The author will find more participants to conduct the usability testing so as to get more data related to the application performance. Feedback will be collected from the users to improve the application whereas any bugs found will be fixed in the updates.

Moreover, better math teaching will be added into the application despite showing the question examples only. More detailed steps will be shown in order to give the users a better understanding of solving the arithmetic questions. Most math homework is calculated to allow students to multiply or divide exactly in primary school. Therefore, the users may be confused when doing the arithmetic questions with remainders especially division questions. It becomes important to let the user understand how to solve this kind of questions or allow the application to generate the question that can be calculated exactly.

The authors believes that the educational math role-playing game using Lehmer's RNG algorithm could provide a solution to students to assist their learning.

REFERENCES

[1] A. Amalina, "Impact COVID-19 on Education and Socioeconomic Mobility," Astro Awani, 2021. [Online]. Available: https://www.astroawani.com/berita-malaysia/impact-covid19-education-and-socioeconomic-mobility-322706.

[2] P. Engzell, A. Frey, and M. D. Verhagen, "Learning Loss Due to School Closures During the COVID-19 Pandemic," Proc. Natl. Acad. Sci., vol. 118, no. 17, 2021.

[3] M. Nwea, J. Soland, and K. Nwea, "Test Score Patterns Across Three COVID-19-impacted School Years," 2022. https://doi.org/10.26300/ga82-6v47.

[4] J. E. Maldonado and K. De Witte, "The effect of school closures on standardized student test outcomes," Br. Educ. Res. J., 2020. https://doi.org/10.1002/berj.3754.

[5] X. Zhao, S. Toonghai, X. Zhang, and Y. Liu, "The Influencing Factors of Game Brand Loyalty," Heliyon, vol. 10, no. 10, pp. 1001-1015, 2024. https://doi.org/10.1016/j.heliyon.2024.10.e01092.

[6] A. Bhardwaj, "Importance of Education in Human Life: A Holistic Approach," Int. J. Sci. Conscious., vol. 2, no. 2, pp. 23-28, 2016.

[7] UNICEF, "Education and COVID-19," 2020. [Online]. Available: https://data.unicef.org/topic/education/covid-19/.

[8] World Bank, "Education and COVID-19: Challenges and Policy Responses," 2021. [Online]. Available: https://www.worldbank.org/en/topic/education/coronavirus.

[9] Berita Harian, "Pelajar IPT mulakan pergerakan pulang ke kampung hari ini," 2020. [Online]. Available: https://www.bharian.com.my/berita/nasional/2020/04/682052/pelajar-ipt-mulakan-pergerakan-pulang-ke-kampung-hari-ini.

[10] NST Online, "Internet allowance for B40 group is vital," New Straits Times, 2020. [Online]. Available: https://www.nst.com.my/opinion/letters/2020/06/599900/internet-allowance-b40-group-vital.

[11] M. Selvanathan, N. A. M. Hussin, and N. A. N. Azazi, "Students learning experiences during COVID-19: Work from home period in Malaysian Higher Learning Institutions," Teaching Public Administration, p. 014473942097790, 2020. doi: 10.1177/0144739420977900.

[12] C. Harris, "Summer Learning Loss: The Problem and Some Solutions," [Online]. Available: https://www.ldonline.org/ld-topics/teaching-instruction/summer-learning-loss-problem-and-some-solutions. [Accessed: Aug. 23, 2024].

[13] P. Gautam, "Advantages And Disadvantages Of Online Learning," [Online]. Available: https://elearningindustry.com/advantages-and-disadvantages-online-learning. [Accessed: Aug. 23, 2024].

[14] S. Fowler, "Effects of Poverty on Education During Distance Learning," [Online]. Available: https://www.adoptaclassroom.org/2020/06/23/effects-of-poverty-on-education-during-distance-learning/. [Accessed: Aug. 23, 2024].

[15] R. Donnelly and H. A. Patrinos, "Learning loss during Covid-19: An early systematic review," Prospects, 2021. doi: 10.1007/s11125-021-09582-6.

[16] World Bank, "Pandemic threatens future earnings and job prospects of an entire generation," [Online]. Available: https://www.worldbank.org/en/news/press-release/2021/12/06/pandemic-threatens-future-earnings-and-job-prospects-of-an-entire-generation. [Accessed: Aug. 23, 2024].

[17] S. P. Zhdanov, K. M. Baranova, N. Udina, A. E. Terpugov, E. V. Lobanova, and O. V. Zakharova, "Analysis of Learning Losses of Students During the COVID-19 Pandemic," Contemporary Educational Technology, vol. 14, no. 3, p. ep369, 2022. doi: 10.30935/cedtech/11812.

[18] N. A. M. Radhi, "(Updated) PT3 2021 cancelled, UPSR abolished," New Straits Times, 2021. [Online]. Available: https://www.nst.com.my/news/nation/2021/04/686114/updated-pt3-2021-cancelled-upsr-abolished. [Accessed: Aug. 23, 2024].

[19] R. Fatima, "Role of Mathematics in the Development of Society," Nat. Meet Celebr. Nat. Year Math., NCERT, New Delhi, 2012, pp. 1-12.

[20] A. M. Helmenstine, "Why Mathematics Is a Language," [Online]. Available: https://www.thoughtco.com/why-mathematics-is-a-language-4158142. [Accessed: Aug. 23, 2024].

[21] G. Kiryakova, N. Angelova, and L. Yordanova, "Gamification in education," in Proc. 9th Int. Balkan Education and Science Conf., 2014.

[22] G. C. Nistor and A. Iacob, "The Advantages of Gamification and Game-Based Learning," [Online]. Available: https://www.proquest.com/docview/2038227087?pq-origsite=gscholar&fromopenview=true. [Accessed: Jun. 17, 2022].

[23] Y. Udjaja, V. S. Guizot, and N. Chandra, "Gamification for elementary mathematics learning in Indonesia," Int. J. Electr. Comput. Eng. (IJECE). vol. 8, no. 6, 2018.

[24] L. Grace, "Game Type and Game Genre," [Online]. Available: http://aii.lgracegames.com/documents/Game_types_and_genres.pdf. [Accessed: Aug. 23, 2024].

[25] D. Pavlovic, "Video Game Genres Everything You Need To Know," [Online]. Available: https://www.hp.com/us-en/shop/tech-takes/video-game-genres. [Accessed: Aug. 23, 2024].

[26] A. Amory, K. Naicker, J. Vincent, and C. Adams, "The use of computer games as an educational tool: identification of appropriate game types and game elements," Br. J. Educ. Technol., vol. 30, no. 4, pp. 311-321, 1999.

[27] A. A. Rahman and A. Angraeni, "Empowering Learners with Role-Playing Game for Vocabulary Mastery," Int. J. Learn. Teach. Educ. Res., vol. 19, no. 1, pp. 60–73, 2020. [Online]. Available:

http://www.ijlter.net/index.php/ijlter/article/view/492. [Accessed: Jun. 17, 2022].

[28] Y. Udjaja, S. Renaldi, K. Tanuwijaya, and I. K. Wairooy, "The Use of Role Playing Game for Japanese Language Learning," Procedia Comput. Sci., vol. 157, pp. 298–305, 2019. doi: 10.1016/j.procs.2019.08.170.

[29] Anon., "New Study: Understanding the Impact of Minecraft in the Math Classroom," [Online]. Available: https://education.minecraft.net/en-us/blog/new-study-understanding-the-impact-of-minecraft-in-the-math-classroom. [Accessed: Aug. 23, 2024].

[30] M. H. Hussein, S. H. Ow, L. S. Cheong, M.-K. Thong, and N. Ale Ebrahim, "Effects of Digital Game-Based Learning on Elementary Science Learning: A Systematic Review," IEEE Access, vol. 7, pp. 62465–62478, 2019. doi: 10.1109/access.2019.2916324.

[31] N. Kara, "A Systematic Review of the Use of Serious Games in Science Education," Contemp. Educ. Technol., vol. 13, no. 2, p. ep295, 2021.

[32] H. Hassan, R. Mailok, and M. Hashim, "Gender and Game Genres Differences in Playing Online Games," J. ICT Educ., vol. 6, pp. 1–15, 2019. doi: 10.37134/jictie.vol6.1.2019. doi: 10.30935/cedtech/9608.

[33] S. T. Tan and S. Suparjoh, "Development of Mobile Game Application: 'Saving Endangered Flora'," Appl. Inf. Technol. Comput. Sci., vol. 3, no. 1, pp. 204–216, 2022. [Online]. Available: https://publisher.uthm.edu.my/periodicals/index.php/aitcs/article/view/2460. [Accessed: Jun. 17, 2022].

[34] R. Ramadan and Y. Widyani, "Game development life cycle guidelines," in Proc. 2013 Int. Conf. Adv. Comput. Sci. Inf. Syst. (ICACSIS), IEEE, 2013, pp. 95-100.

[35] A. Candra, M. A. Budiman, and R. I. Pohan, "Application of A-Star Algorithm on Pathfinding Game," J. Phys.: Conf. Ser., vol. 1898, no. 1, p. 012047, 2021. doi: 10.1088/1742-6596/1898/1/012047.

[36] E. Lindholm, "Procedurally generating an initial character state for interesting role-playing game experiences," [Online]. Available: https://www.diva-portal.org/smash/get/diva2:1453256/FULLTEXT01.pdf. [Accessed: Jun. 17, 2022].

[37] T. Q. Alhassan, S. S. Omar, and L. A. Elrefaei, "Game of Bloxorz solving agent using informed and uninformed search strategies," Procedia Comput. Sci., vol. 163, pp. 391-399, 2019.

[38] P. V. F. da Silva and S. M. Villela, "Applying pathfinding techniques on the development of an android game," in Proc. SBGames 2016, 2016, pp. 73-80.

[39] G. T. Galam, T. P. Remedio, and M. A. Dias, "Viral infection genetic algorithm with dynamic infectability for pathfinding in a tower defense game," in Proc. 2019 18th Brazilian Symp. Comput. Games Digit. Entertainment (SBGames), IEEE, 2019, pp. 198-207.

[40] A. M. Zain, C. W. Chai, C. C. Goh, B. J. Lim, C. J. Low, and S. J. Tan, "Development of Tic-Tac-Toe Game Using Heuristic Search," in IOP Conf. Ser.: Mater. Sci. Eng., vol. 864, no. 1, p. 012090, May 2020, IOP Publishing.

[41] J. Niklaus, M. Alberti, R. Ingold, M. Stolze, and T. Koller, "Challenging Human Supremacy: Evaluating Monte Carlo Tree Search and Deep Learning for the Trick Taking Card Game Jass," in Proc. Int. Conf. Artif. Intell. Soft Comput., Cham: Springer, Oct. 2020, pp. 505-517.

[42] C. L. Lu, C. H. Hsu, and S. R. Kuo, "Hybrid Multi-Patterns With Monte Carlo Search Approach For The Puzzle Game Solver." Int. J. Adv. Eng. Manag. Res., vol. 4. pp. 21-30

[43] M. J. Goertzen, "Introduction to Quantitative Research and Data," Library Technology Reports, vol. 53, no. 4, pp. 12–18, 2017. [Online]. Available: https://journals.ala.org/index.php/ltr/article/view/6325. [Accessed: Jun. 17, 2022].

[44] D. Silverman, Qualitative Research, SAGE, 2020. [Online]. Available: https://books.google.com.my/books?id=dJbvDwAAQBAJ&pg=PT8&dq=qualitative+research&lr=&source=gbs_selected_pages&cad=2#v=onepage&q=qualitative%20research&f=false. [Accessed: Jun. 17, 2022].

The Impact of Malware Attacks on the Performance of Various Operating Systems

Maria-Mădălina Andronache[1], Alexandru Vulpe[2], Corneliu Burileanu[3]
Research Institute "CAMPUS", National University of Science and Technology Politehnica Bucharest, Bucharest, Romania[1]
Telecommunication Department, National University of Science and Technology Politehnica Bucharest, Bucharest, Romania[2]
Speech and Dialogue Research Lab, National University of Science and Technology Politehnica Bucharest, Bucharest, Romania[3]

***Abstract*—Latest research in the field of cyber security concludes that a permanent monitoring of the network and its protection, based on various tools or solutions, are key aspects for protecting it against vulnerabilities. So, it is imperative that solutions such as firewall, antivirus, Intrusion Detection System, Intrusion Prevention System, Security Information and Event Management to be implemented for all networks used. However, if the attack has reached the network, it is necessary to identify and analyze it in order to be able to assess the damage, to prevent similar events from happening and to build an incident response adapted to the network used. This work analyzes the impact of malicious and benign files that have reached a network. Thus, during the work, various analysis methods (both static and dynamic) of real malicious software will be developed, in two different operating systems (Windows 10 and Ubuntu 22.04). Thereby, both the malware and benign files and their impact on various operating systems will be analyzed.**

***Keywords*—Cybersecurity; network security; network monitoring; incident analysis; incident response**

I. Introduction

Network security includes a large number of technologies and devices that must work together based on a predefined set of rules. These rules are primarily intended to protect sensitive information in a system. However, it must be considered that security aspects cannot work with the same set of rules indefinitely because the threat environment is constantly changing, attackers are always trying to find new vulnerabilities, and network architectures are increasingly complex and different. This is the main reason why network security management tools or applications are also constantly updated.

In this security context, malicious files are an extremely difficult aspect to ignore. So, they are represented by malicious software that is created to produce, according to study [1], various exfiltration of information or to cause various interruptions. Their general purpose is to obtain ideas, to damage the reputation of a company or a system, or material gains. Malicious files can achieve these aspects precisely because of the possibility of exploiting some vulnerabilities or due to the negligence of certain people who perform various actions unfavorable to the security context (disabling the antivirus, deactivating the firewall). From study [2], malicious files and cyber-attacks have expanded their action exponentially in recent years, being encountered more and more often both within companies and for ordinary users. Their greatest impact is given by affecting critical systems such as the health area, the financial-banking area, the area of government attacks or the industrial area. Although malicious files and new types of attacks appear daily, a main part of them are based on the skeletons of older malware to which various code improvements are added. Thus, it is imperative to analyze the existing malware and understand their characteristics because they can generate patterns of future attacks. To perform this type of analysis, it is important to distinguish between static and dynamic analysis of files. This differentiation is made considering various sources such as [3], [4] and [5]. Static analysis of a malicious file involves testing it without executing it. So, this involves the analysis of the source code and other aspects such as the magic number or the hash of the file in order to identify whether it is malicious or not. Dynamic analysis is how the file is analyzed after it is executed to observe how it affects various files or various system registries. Given the fact that this method also involves the execution of the file, it is necessary for this to be done in a closed and controlled environment. Following these two types of analysis, it will be determined whether the file is malicious. The most frequently encountered types of malware present in a system are those known as zero-day. However, the static and dynamic analysis methods cannot detect this type of attack if it does not have a known pattern. For all the other types of malicious files: ransomware, trojan, virus, worm, backdoor, this analysis can be performed for the purpose of documentation and for the purpose of identifying the main characteristics necessary to prevent a possible subsequent attack. These characteristics are also considered based on the literature in [6], [7], [8] and [9].

The purpose of this work is to perform a comparative analysis of how different types of files are executed within two different types of operating systems. For this purpose, a test environment is created, which includes both static and dynamic analysis methods. The contributions of this work consist in creating a test environment (which is similar to a regular user, within a company and does not rely on existing sandboxes), choosing the most recent files from a public malware database (that are not predefined and used in another labeled database), analyzing various events and logs (after the execution of the files) and making a comparison on the key characteristics of these files.

This paper is organized as follows. Section II provides an evaluation of the specialized literature and research related to the analysis of malicious files. Section III provides the area of background work. Experiments and results are presented in Table IV. Discussion is given in Section V and finally, Section VI concludes the paper.

II. Related Work

To perform a complete analysis of malware files, a thorough review of the research activity and tools required in the process is required. Through this section, a deeper understanding of the existing research in the field will be achieved, an aspect that will also lead to the framing of this work in the current security context.

Taking into consideration the paper [10], it is found that it introduces a study related to the dynamic analysis of malicious files, evaluating an open source SIEM (Security Information and Event Management) system called Elastic Stack. Thus, by capturing the event logging mode in Windows, a complete description of the events within the system could be achieved. Malware analysis included in the paper contains a Dynamic Analysis. This type of analysis is also used in the current paper, but compared to the work [10], this paper includes experiments within both Linux and Windows operating system.

The vulnerabilities of a system are an extremely sensitive subject for traditional detection methods because they contain extremely complex functions and algorithms, which cannot be interpreted by them. In the paper [11], detection of these vulnerabilities is carried out in a binary code, by means of neural network algorithms and by means of the NDSS18 database. This database contains CVEs (Common Vulnerabilities and Exposures) for both Windows and Linux operating systems. The results indicate a good performance of the model given by machine-learning algorithms. In this paper, both operating systems will be used to see the interaction of some malware files with the default processes of these operating systems.

In paper [12], a method for detecting the behavior of malicious Android activities is proposed through a hybrid, static and dynamic approach with automatic learning. The results of the work indicate an accuracy of 97% for the detection of anomalies. The advantages of the work would be the reduced use of some permission functions and the consumption of resources, which improve the efficiency of the system. In the present work, the absence of experiments in the area of the Android operating system is identified, as it is not a desktop environment. However, the work in [12] indicates various similarities between the operating modes of malicious files in the desktop area and in the mobile area, the static analysis being carried out identically.

According to study [13], IoT devices occupy a special place in the area of malware detection because they have different characteristics depending on the environment and the platform studied, making it extremely difficult to identify. The proposed analysis method includes both static analysis, against software shells, and dynamic analysis, through nine different sandboxes. The analysis method required the creation of a database and, for this, samples from the Padawan sandbox, VirusTotal [33] and other open-source areas were taken into account. The malware detection accuracy presented in the solution, evaluated using XGBoost, SHAP and Scikit-Learn exceeds 98%. In the current work, compared to the work [13], the sandbox used for dynamic analysis is not a commercial one, but is given by common virtual machines of common users of a company. Thus, the tools used for the experiments are also different, but the key aspects pursued are similar.

In the paper [14], a solution for detecting malicious files using the YARA tool is presented. Thus, during the work, five rules were developed for malware detection in the static analysis area. From the expressed results, it is stated that the presented solution reduces the identification time, improving the detection efficiency. In this work, Yara rules will be also used for static analysis.

Considering [15], the paper presents a way of detection and prevention of malicious files before they corrupt the test system. Thereby, a Virtual Box type environment is used in which a static analysis of malicious files is carried out. The experiments were carried out by means of IDAPro, a tool that will be used in this work as well, and the malicious files were of the type of Trojan horses. The results indicated various functions, strings, imports and exports made by the malware program, and their detection was done through Reverse Engineering.

Taking into account the experiments made in study [16], it is concluded that a static analysis is carried out on a malicious file, with the aim of detecting its behavior. The way to detect malicious files is by extracting the APIs and checking them, and the authors have developed a program that analyzes PE files. The files on which this test is performed are benign, Ransomware, Backdoor and Keylogger and this approach is also found in the present work.

From study [17], it can be concluded how ransomware works. Thus, through the specified paper, certain experiments are carried out through the Kaspersky ransomware signature database and a virtual environment. So, attacks are detected, based on rules based on the signatures of these files, and an analysis is made on their functionality and prevention. Therefore, all analyzed files are either restricted or sent to the detection area for further processing. The present work also addresses ransomware files within the experiments, but the files are different within the two papers.

In study [18], an analysis of the vulnerabilities of a software system is presented. During the work, both the main types of system vulnerabilities (from Buffer overflow to DOS or Memory Corruption) are highlighted, as well as the detection methods, which include, as in this work, static analysis, dynamic analysis and hybrid analysis. The experimental data used in the paper are the authors' own data and include historical vulnerability data. These are evaluated for vulnerability detection by means of machine-learning techniques. The essential difference between the work [18] and the current paper includes their motivation: one trains Machine Learning algorithms and the other only extracts and analyzes key features to form a complex database of malware files.

The previously cited literature sources highlight various aspects of the cyber security area and various approaches similar to the one in the present paper. A good part of them is based on machine-learning algorithms and the development of robust models through this technology. However, in the real security environment, within various companies, there is still quite a lot of skepticism regarding the area of artificial intelligence and the methods used by it. Although the advantages of these solutions are immense and solve a large part of repetitive tasks, knowledge of traditional methods is still imperative to ensure a complete and deep understanding of the field.

III. BACKGROUND WORK

In order to carry out a complete analysis of the behavior of a malicious file, both a static analysis and a dynamic analysis of it are necessary. Thus, during the experiments carried out in this work, both types of analysis will be performed on legitimate or malicious files from the Windows 10 and Ubuntu 22.04 operating systems. The goal is to identify the essential characteristics of various types of malware in order to create a complete database of characteristics. This will be able to serve, later, to create an automatic intrusion detection system within a computer network of various sizes. These types of analysis, both static and dynamic, will be implemented using various open-source tools, and the results of the study will be analyzed comparatively.

The analyzed files were downloaded from the MalwareBazaar Database [25] and include various types of files related to Windows and Linux operating systems. The method of choosing the malware samples used did not follow any specific algorithm but included the identification of similar types of malware for the two types of operating systems used. This need appeared after consulting the literature regarding similar, relevant articles in the field, finding a predilection for known or even outdated databases, which are permanently tested. Therefore, out of the desire to perform the experiments with real malware samples and not with pre-tagged databases, the MalwareBazaar Database [25] source was chosen.

Within this database there are many different malware samples, but the categories are represented by the main types of malware. Thereby, similar samples were chosen (from the same malware family, from the same category, appeared on the same day, etc.), both for Windows and for Linux. By means of this approach, the research can evaluate various ways of functioning of malicious files in its own way and can lead to unpredictable conclusions, which can contribute both in the literature and in the area of commercial applications.

A. Static Analysis

The static analysis of this work will be carried out using various open-source tools. These tools were chosen due to their popularity and effectiveness in detecting key characteristics of malicious files for both Windows and Linux operating systems.

- IDA Pro [19] is a tool that is able to disassemble the files and identify the way in which the instructions are executed in the assembly language. This tool is suitable for both Windows and Linux OS.
- PeStudio [20] is an integrated tool for static analysis of malicious software that indicates various information about files (file headers, file entropy, character strings or imported or exported functions). This tool is characteristic of the Windows operating system, but for Linux it will be an used an alternative named Malcat [26] and Detect-It-Easy [27].
- YARA rules [21] – methods of identifying malicious programs by means of commands written in a .txt program that include instructions for identifying similar sequences used or that have similar patterns.

B. Dynamic Analysis

Dynamic analysis of malicious files means executing them to be able to observe the actual behavior. So, it is found that for this aspect, it is necessary to create a safe and isolated sandbox environment. Except for this aspect, it is also necessary to introduce various tools that can be used to identify the described behavior at runtime. Also, tests will be carried out through which it will be possible to observe whether or not these types of files raise various alerts through the antiviruses specific to each OS.

The essential characteristics that must be monitored in a dynamic analysis are:

- Network traffic monitoring: To be able to track IPs and DNSs contacted for file downloads or data exfiltration.
- System file monitoring: Monitoring how certain registries are created, modified or deleted.
- CPU monitoring: To be able to observe if it becomes overloaded by certain unknown requests.
- Memory area monitoring: To be able to identify activities that are not visible.
- Code monitoring: If it can be decrypted, it provides important information about how the malicious file works.

This information can determine several aspects of the actual attack and whether the system in which it was identified is the target or only an intermediate step towards the final target.

It is mentioned that some of the tools used in the created sandboxes are characteristic of the operating system, and another part of them is common to both Windows and Linux. The chosen tools have the advantages of being open-source and, according to the literature, have increased efficiency in monitoring malicious activity.

- Regshot [22] is a dynamic analysis tool that performs a comparison of a created file with the status of registers and system keys before and after the execution of the malicious file. This tool identifies the changes made by the malicious file and is characteristic of the Windows operating system, but it can also be adapted for Linux, through a series of commands.
- Wireshark [23] is a tool that can be used to capture and analyze data packets from a network. In this work, this analyzer will be used to identify the exchange of messages between the malware file executed within the network and any IPs to which the request is made. Although this tool is normally used in the Windows operating system, adaptations can also be found for the Linux area.
- FakeNet-NG [24] is a tool that simulates Internet traffic, specially created for the malware analysis area. Thus, the malicious programs consider that the workstation is connected to the Internet and try to access various resources. These are later captured to be analyzed in the file behavior characteristics area. This tool can be implemented both in the Windows operating system area, as well as in the Linux OS area.

IV. Experiments and Results

As can be seen from the Fig. 1, the steps necessary to perform the two types of analysis are: choosing the file type, choosing the analysis method with the related implications (analyzing the file, without executing it, and analyzing it by executing it, in a sandbox environment) and choosing the appropriate tools for each analysis, separately. Considering the fact that, in the case of dynamic analysis, the execution of malicious files will involve affecting some registers or even the entire test environment, the experiments will be performed in a virtual work environment. This offers the possibility of returning to the previous settings through the Snapshot function.

This work approach will be preserved for both operating systems. In addition to the basic tools, which will be used to detect malware files, additional resources will be used, such as various functions from Microsoft or the detection of differences in the state of the CPU and memory during the attack. Also, in order to be able to analyze the involvement of anti-virus software in the experiments, this resource will also be used, and its involvement will be analyzed comparatively, depending on the operating system.

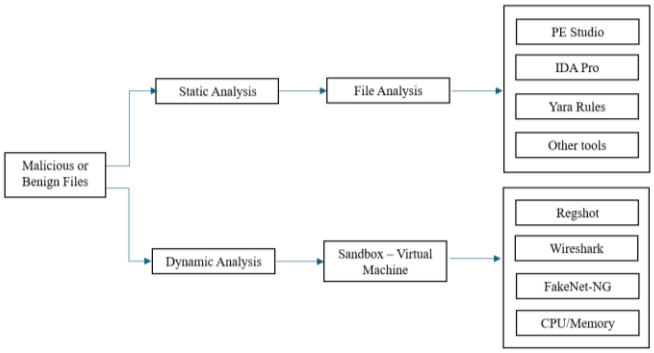


Fig. 1. Workflow diagram of the system.

A. Experiments in the Windows Operating System

a) Static analysis: In the framework of the static analysis present in the Windows operating system, all the tools that were previously presented, in Section III, will be used.

The files used for the experiments in this scenario are of different types, containing both non-malicious and malicious files downloaded from an online database or from public resources (for benign files). For the experiments carried out, the legitimate files are of various types – from executable files to document-type files, archived or photo-type files. The benign file that has .exe extension is the strings64.exe file from the Strings library [28]. Apart from benign files, several different types of malicious files were also chosen from MalwareBazaar Database [25]. Each of these was analyzed using PEStudio, strings64.exe, YARA Rules and IDA Pro. By means of these tools, it was possible to analyze the hash values of the files (MD5-Message Digest Method 5, SHA1-Secure Hash Algorithm 1, SHA256-Secure Hash Algorithm 256), their imphashes, entropy values, imported libraries or APIs. In Table I, you can find the values of file types, their imphash or MD5 hash and the entropy values for each one of them.

TABLE I. Static Analysis of Malicious files in the Windows Operating System

File Type	Characteristics	
	MD5/ Imphash	Entropy
Benign .exe	4d936b630620ff7c59da22b1206636e	6.42
Benign .doc	6698b2a4a15f86ddd4fc90ad65521cf7	7.76
Benign .txt	42631b1af161defcf4844fb1e26cfc70	4.40
Benign .jpg	2be5d32efb9c3f4b6acf94a1d1e707b4	7.96
Benign .mp4	37d7f751daa745beba4cf44b6373f2be	8.00
Benign .pdf	f9067cb2369fa0ec4e3753f67638fbca	7.34
Benign .zip	8273b4301f7a6d678c0523bb07fedd80	7.99
Benign .html	40c15f040bc8f4aeb81909fb36aa9905	4.29
Benign .iso	089a3a344f301a34dc40cc3702f2b873	7.93
Botnet .exe	5e146bf6c1ef160162ed271c0ddde908	3.54
Backdoor .dll	f34d5f2d4577ed6d9ceec516c1f5a744	4.42
Keylogger .exe	008a6a7f7e2610edadf3e2f26c73b646	7.63
Malware .exe	11ea24073ee65343ee563e3160c77fde	7.81
Ransomware .exe	914685b69f2ac2ff61b6b0f1883a054d	7.18
RAT .exe	8d5087ff5de35c3fbb9f212b47d63cad	6.59
Trojan .exe	d6d4965d7fe2d90a52736f0db331f81a	6.59
Worm .exe	2dfc2c74864b84f5530ab40a343c56d8	5.36

Imphash is, from study [31], the method by which a hash is calculated based on the libraries and APIs imported by the file. This is useful to determine if two apparently different files come from the same source or belong to the same family. Within the values presented in Table I, the imphash values are not similar and, therefore, the analyzed files are different and do not belong to the same malware family. Given the fact that the imphash value can only be calculated for executable files, for benign files, which also contain other types of files (except for .exe), the values related to the MD5 type string were added.

The entropy value gives, according to [30] the level of randomness of a file. Thus, the higher the value of the file, apart from the interval [0, 8], it can be concluded that the file is encrypted or packed and can be identified as malware. This aspect is not respected within the values in Table I because the benign files, which are also executable, have, sometimes, a higher value than malicious files such as backdoor or worm. In the case of benign files, the highest entropy values are recorded in the case of .doc, .jpg, .mp4, .pdf, .zip and .iso files.

The explanation for this phenomenon is that, in the case of .jpg or .zip files, the compression algorithms used increase the randomization of the data in order to compact them in a safe way. In the case of .mp4 type files, they encode various waveforms, which leads to a random appearance of the file. For files of type .doc or .pdf, the entropy can have a high value due to various images or text with different fonts. In the case of .iso files, they contain several types of smaller files (which can have various extensions), so its randomization index will be high.

Since the static analysis of a file, especially in the case of those considered malicious, includes aspects such as access and manipulation of memory resources, reading the source code or imported functions or libraries, a deeper look at these resources is necessary.

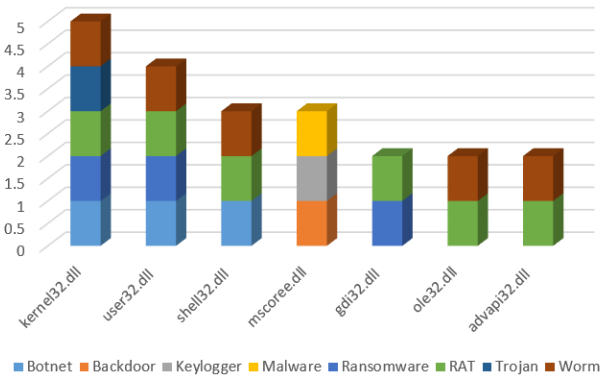


Fig. 2. DLLs imported by malicious files in Windows operating system.

Thus, bearing in mind that the files are of different types and, therefore, the imported functions are also diverse, a comparative analysis at their level would be extremely difficult to achieve. However, what can be analyzed is the domain of libraries imported by each individual file. For the current work scenario, this data can be found in Fig. 2. From this figure, you can see what type of libraries can be imported by the file, if it will be executed. All these libraries are considered legitimate libraries and compatible with the Windows operating system, having various specific functions. As an example, the kernel32.dll library deals with memory management and the control of various processes, while user32.dll considers the user interface and various inputs given by them. However, in the case of malicious files, these libraries have different functionalities. Basically, an attacker names a part of the malicious file identical to a known library but attaches it to a different directory within the file suite. This is also a direct way to detect if a library is legitimate or not. Moreover, even the fact that a file wants to create or overwrite a system library is a fairly solid indicator of the presence of malicious software.

Taking into account the fact that, through static analysis, the considered parameters can create confusion regarding the malware characteristics of a file, a dynamic analysis is also needed, which can clarify certain aspects. Therefore, the analysis of these files will have to be carried out in the next paragraph, in order to be able to observe how legitimate and malicious software functions in a real environment.

b) Dynamic analysis: Dynamic analysis is important because it allows evaluating the behavior of a software program during its execution. In the case of this scenario, the motivation for its realization is the more defined perspective on the key aspects of some legitimate files, compared to some malicious ones.

The aspects identified in the dynamic analysis will be related to the presence or absence of system changes. Thereby, by means of values in Table II, it will be identified if there are modified key values or registers (identified by the Regshot tool), if the connection to various Internet sources (Fake Net and Wireshark) is started, if there are changes to the CPU resources and memory and if the presence of the malware is identified by the antivirus.

TABLE II. DYNAMIC ANALYSIS OF MALICIOUS FILES IN THE WINDOWS OPERATING SYSTEM

File Type	Characteristics			
	Values/ Keys Modified	*Network Activity*	*CPU/ Memory*	*Detected by AV*
Benign .exe	√	X	X	X
Benign .doc	√	X	X	X
Benign .txt	√	X	X	X
Benign .jpg	√	X	X	X
Benign .mp4	√	X	√	X
Benign .pdf	√	X	√	X
Benign .zip	√	X	X	X
Benign .html	√	√	X	X
Benign .iso	√	X	X	X
Botnet .exe	√	√	√	√
Backdoor .dll	√	√	√	√
Keylogger .exe	√	√	√	√
Malware .exe	√	√	√	√
Ransomware .exe	√	√	√	√
RAT .exe	√	√	√	√
Trojan .exe	√	√	√	√
Worm .exe	√	√	√	√

During the experiments carried out in the Windows sandbox, all files were executed sequentially. After running each type of malware, the test environment needed to be replaced via the Snapshot function. This aspect was also valid for benign files, even if their execution does not endanger other files. However, it was desired that the execution of one of the files to not influence the execution of subsequent files in any way. The files that strongly affected the test environment were Worm and Ransomware. In the case of these two types of experiments, the results, given by running the malicious files, led to temporary interruptions of the virtual machines or even to the irretrievability of some data.

A slightly more atypical aspect, resulting from experiments, was in the case of the experiments carried out with the Worm type file, because the Fake Net area could not stop making recordings, having this behavior for a few minutes. During all this time, many HTTP (POST) or DNS related events were recorded, through which the malicious file tried to access ihcnogskt.biz or kkqypycm.biz multiple times.

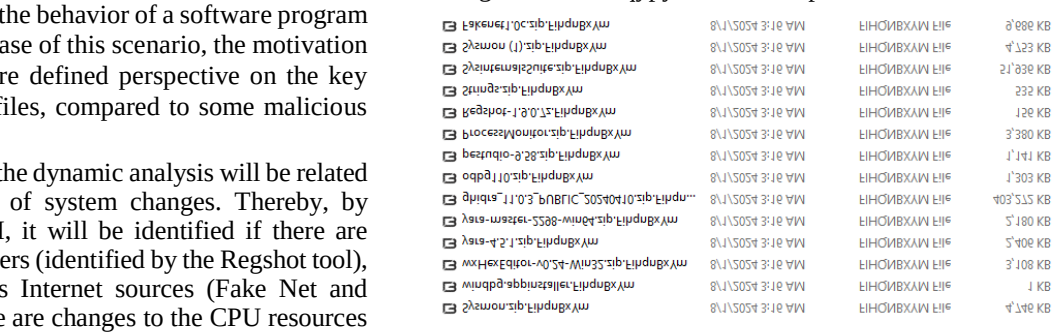


Fig. 3. Changing the file extension after conducting experiments with the malicious ransomware type file.

In the case of the experiments carried out with the Ransomware type file, after execution, it encrypted all the files of the working system, giving them a new type of file extension that can be seen in Fig. 3. As a result of this aspect, the virtual machine became inaccessible, and the files could no longer return to the previous extension type. Even after trying to return to the default settings of the virtual machine, the initial files could not be recovered. Thus, it was necessary to create another working environment.

It is mentioned that the duration of all the experiments done in the test environment did not last more than 5-10 minutes/experiment out of concern not to irreparably affecting the test environment.

B. *Experiments in the Ubuntu 22.04 Operating System*

a) Static analysis: In the case of the experiments carried out in this scenario, benign files, that are completely different from those used in Windows operating system, contain different file types, from a python executable called impelf.py [29], to some other file types corresponding to the Linux operating system. The malicious files are also different, but the way in which they were chosen was by comparing them with the previous chosen ones (be from the same malware family, be announced around the same time, etc.). Each of these files were analyzed using Malcat, Detect-it-Easy, readelf, md5sum, sha1sum, sha256sum. Since .elf and benign files have no information about their executable mode, the imphash hash could only be generated for Keylogger and Worm files, which were .exe. For the rest of the files, in Table III, their md5 hash was added.

Taking into consideration that a complete static analysis needs to include aspects related to various properties of the files (the imports made or file structures), it is also necessary to understand the malicious files in this scenario. Thus, an important amount of the files analyzed within the experiments have the extension .elf. From [32], these files are characteristic of the Linux systems, being executable files or libraries. This type of file is structurally divided into two parts: the header area and the segment area. The header contains various metadata, and the segments describe various memory operations, which are performed during execution. These segments are of various types, but the most common ones are those found in Fig. 4.

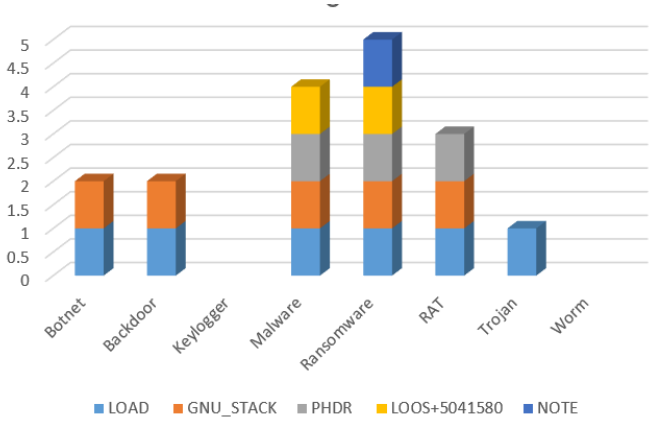


Fig. 4. Segments imported by the malicious files in the ubuntu 22.04 operating system.

Thus, as can be seen from the figure, the first type of segment is LOAD segment and it indicates the memory location where the file will be loaded and its various permissions, while the NOTE type segment includes information that can be used by the system kernel such as version, debug data, etc.

In the context of the analysis of some malicious files, although certain segments appear to be legitimate, being usually found in executable files, they can be modified to have the character of malware. Thereby, the malicious behavior can be hidden in "stuffing" segments in order to be able to pass undetected by various types of antiviruses. If these segments also have file modification or execution privileges, this behavior may indicate the presence of malware. Another detection index of these types of files is the atypically large size of some segments that are apparently legitimate.

In addition to the general segments, such as LOAD or GNU_STACK, which are present in almost all files of the .elf type, there are also segments that are a little more atypical for an ordinary file. Among these types of segments is LOOS+5041580. The fact that, even at the level of Internet resources, extremely little data about this type of segment can be identified, may indicate the presence of malware. It is emphasized that, even the searches based on the LOOS segment, without the numerical suffix, did not indicate details about a usual work segment.

Similar to the experiments carried out in the Windows operating system, the presence of malicious activity cannot be clearly defined to be able to conclude that one of the chosen files is truly legitimate or has malicious characteristics (entropy has different values, segment type files also have a legitimate character, etc.). Thus, the dynamic analysis of the files is also needed to be able to have a more explicit conclusion on the mode of operation and the influence they have on the Ubuntu 22.04 operating system.

TABLE III. STATIC ANALYSIS OF MALICIOUS FILES IN THE UBUNTU OPERATING SYSTEM

File Type	Characteristics	
	MD5/ Imphash	***Entropy***
Benign .py	b6014a53db0e1797301ec118f2625c45	4.53
Benign .txt	6045aa2bdbbfa5839a382fbc383307ac	4.75
Benign .mp4	a6b8790aeeffaa6b08b1b7dfa2b0a1f7	7.99
Benign .sh	fc331af161311d2000fb18d02764a062	5.49
Benign .tar	38544f88237f2b1184c8822289a1899d	7.99
Benign .iso	05fde34ce38913489a1a988175240f27	7.99
Benign .pdf	e9ef095f7dec56b483d2c31f915e177c	7.98
Benign .jpg	0fe826c9fad792732c9081b59bbcb613	7.85
Benign .conf	e95d5425c026ab1142a025d49bf23dc9	4.81
Backdoor .elf	9a85bf5e1b4ca4db7b5654aa48df5f2e	5.65
Botnet .elf	4d58d0cae526ee6364f7c738b83f2961	6.01
Keylogger .gz	888988a74b67d0e75f5293688ab07b71	4.12
Malware .elf	171d2a50c6d7e69281d1c3ef98d510f2	6.00
Ransomware .elf	56cabcf95add39a6feb09391ccc40dcd	6.18
RAT .elf	9f539613aae69eec04ed66550f814f6b	7.98
Trojan .elf	1655222d44cfc33dcc3d10f8a4f2e2db	1.51
Worm .tar	5a46892a133f6e380a5a2acb389c5af6	6.93

b) Dynamic analysis: Following the experiments carried out, in the Ubuntu 22.04 operating system, the same structure found in the dynamic analysis in the Windows 10 sandbox was kept. So, both the benign files and the malicious files were run one by one.

The results of the experiments can be found in Table IV, where the key elements of the detection are reproduced. Thereby, it can be observed that for all types of files, there are modified values or registry keys, and different CPU and memory values recorded during the attack. However, the situation is different for the network activity when the file is executed and the files that are detected by the ClamAV antivirus indicators. Thus, although some files are declared as malware in public databases, in our experiments, it can easily pass the security system of the virtual machine. The less favorable aspects happened in the case of running two different types of files, the one with malware and the one with ransomware. So, after running the malware file, the message "Exporting key" is recorded, after which the workstation becomes unusable, performing a restart. After the execution of the ransomware file, not even this message is displayed, but after restarting the test environment, some files are unusable. For these experiments, recording the malicious behavior was quite difficult due to the continuous restart of the virtual machine. Thereby, in order to succeed in capturing the parameters for Table IV, it was necessary to repeat the tests several times.

It is mentioned that the recording duration of the malicious events was carried out in an interval of approximately 5-10 minutes after the execution of the malware file. The values presented in Table IV reveal only the aspects recorded within this interval. It is also mentioned that, even for the benign files, after each execution, the test environment was returned to the default settings, so that the experiments are not inter-influenced.

TABLE IV. DYNAMIC ANALYSIS OF MALICIOUS FILES IN THE UBUNTU OPERATING SYSTEM

File Type	Characteristics			
	Values/ Keys Modified	***Network Activity***	***CPU/ Memory***	***Detected by AV***
Benign .py	√	**X**	√	**X**
Benign .txt	√	**X**	√	**X**
Benign .mp4	√	**X**	√	**X**
Benign .sh	√	**X**	**X**	**X**
Benign .tar	√	**X**	√	**X**
Benign .iso	√	**X**	**X**	**X**
Benign .pdf	√	**X**	**X**	**X**
Benign .jpg	√	**X**	**X**	**X**
Benign .conf	√	**X**	**X**	**X**
Backdoor .elf	√	**X**	√	√
Botnet .elf	√	**X**	√	√
Keylogger .gz	√	√	√	**X**
Malware .elf	√	**X**	√	**X**
Ransomware .elf	√	**X**	√	**X**
RAT .elf	√	**X**	√	√
Trojan .elf	√	**X**	√	√
Worm .tar	√	**X**	√	**X**

V. DISCUSSION

Taking into account the aspects analyzed in the previous section, a comparative analysis of how the benign files and the malicious software affected the test systems is necessary. Although, in Fig. 1, these analyses are treated independently, in fact, an analysis of malware files is complete only by encompassing all aspects analyzed through all experiments performed and tools used. Thus, in the static analysis, carried out in the case of both analyzed operating systems, it is found that the test files have different entropies, included in the range [3.54-8.00] for Windows and [1.51-7.99] for Ubuntu. Although the initial expectations were that malicious files have a higher entropy value (corresponding to the way in which the malicious file is encrypted and packaged to avoid detection), it is observed that, in the case of both operating systems, some of the files that do not contain malware has a higher entropy value than some files containing malware. This aspect is caused by the randomization achieved by compressing certain files (archive type), by including certain metadata such as photos or different fonts (in the case of document type files) or by processing certain different waveforms (in the case of .mp4 files).

Regarding the hash part, for both operating systems it was necessary to discover various types of hashes. Although, initially, for the tests, the representation of the data by imphash was chosen, this aspect proved to be inconsistent for files that are not Portable Executable. This aspect had an impact in both operating systems because, given the fact that the construct part of some files is not executable or is a binary one, imphash cannot be represented. Therefore, it was chosen to complete the table with the MD5 value for these types of files. A good aspect that needs to be mentioned is the purpose for which it was necessary to find out the hash of a file. Basically, according to study [3], the hash of files within the tables had two directions of development. The first direction is given by the fact that a good part of the malicious files is recognized in some tools by hashes. The second direction, which derives from the first, is the fact that, depending on various hash values (MD5, SHA, SHA256, SHA512, imphash), the characteristics of malicious files can be found in various public databases. The most used example, in this case, is through the VirusTotal tool through which, after searching for a hash, it shows extremely relevant information about the respective file. This tool can also be included in the static analysis area, being an extremely useful aspect in discovering similar files or the characteristics of a certain type of malware.

However, this approach was not necessary in the case of benign files because these files could be represented by the actual name. Their values, that are, in Table I and III, also in the form of hash have the purpose of a unitary way of data representation.

Considering that a complete static analysis must also include aspects related to the actual content of the file, it is necessary for it to include various information such as the strings used, APIs or DLLs imported or the source code, as in study [3], [4] and [10]. Therefore, in the case of both scenarios, these aspects were also analyzed. Unfortunately, due to the difference in the structure of the analyzed files, they cannot be the basis of a comparative analysis, being also extremely different. Thus, the

approach consisted in delimiting the two scenarios and performing a comparative analysis only according to the malicious files within the respective scenario. Therefore, within the Windows operating system, the most frequently encountered imported DLLs were chosen, to be able to observe if there are common characteristics between the analyzed malware files and the intentions of each malicious file, separately. It was concluded that these files import both common libraries and independent libraries. Within Fig. 2, only the common ones were included, the most frequently encountered being kernel32.dll, user32.dll, shell32.dll and mscore.dll. These have as their main characteristics file operations, privilege escalation, avoid analysis, user interface or memory management operations. The importance of these DLLs imported by malicious files can lead to conclusions about the behavior of the malicious file (as in [16]), its inclusion in a certain family, the creation of similarities with other files known to be malicious or the creation of action patterns of certain types of files.

In the Ubuntu 22.04 operating system, the analyzed files, from a static point of view, imposed the analysis of some files with .elf extensions. These are, in essence, binary files and therefore it is quite difficult to extract some essential features, in a similar way to the previous scenario. The initial analysis of these files consisted of examining the file structure (the size of the headers, their number, etc.) or the strings used, but these characteristics could not define a comparative analysis that would indicate an atypical character of the file. Therefore, the chosen feature consisted of presenting the segments loaded in memory during execution. Considering Fig. 4, it is indicated that the most frequently used segments for malicious files were LOAD, GNU_STACK and PHDR. Their essential characteristics include indications on an executable code, on working with the memory stack or on various memory locations that need to be accessed or written. In the case of malicious files, these segments include the possibility of attacks that include self-modifying, returning addresses from the memory stack or escalation of privileges. Therefore, it is extremely important for these characteristics to be known, in order to get an overview of the behavior of the malicious files. Apart from the mentioned aspects, in Fig. 4 you can see the absence of segments for two types of files - Keylogger and Worm. This aspect is due to the fact that the two types of files are not of the .elf type, but of the .gz type. Thus, they must be characterized in a different way.

Considering the aspects mentioned for static analysis and papers [3], [4], [5] and [15], it is necessary to emphasize that they are quite insufficient for a complete characterization of a file type. Therefore, a dynamic analysis would complete the unknown aspects or for which the static analysis is insufficient and would offer the possibility of an overview of the malicious file's mode of action.

Dynamic event analysis involved conducting experiments on how file execution works in a sandbox environment. However, in order to distinguish between a malicious file and a benign one, it was necessary to establish some indicators of the presence of malicious software, similar to [4], [5], [10] and [17]. Therefore, detection by OS-specific antivirus was one of the main parameters that should be considered. Also, the steep increase in machine performance, given by the CPU and memory values, is one of the necessary indicators of the presence of non-compliant activity. Other important aspects were given by the monitoring of the traffic within the network and the values or keys changed at the level of the operating system registries.

Although analyzed independently, the performance parameters could be considered insufficient (CPU and memory values increase even when running a legitimate application), in the analysis, they work as a whole. The discrepancy in results between the two operating systems is due to both the way the malicious files work and the actual file type. Not having the same applicable file available for both operating systems leads to different experimental values.

Taking into consideration all this, it was found that, if at the level of the Windows operating system, aspects are more uniform, the malware files leaving positive traces for all the mentioned indicators, in Ubuntu operating system, things are less favorable because some malicious files are not even recognized by the antivirus. Thus, they can go unnoticed and can attack systems without being detected. The worst aspect is that ransomware and malware files that, after execution, lead to the temporary inaccessibility of the virtual machine, are, initially, undetected by the antivirus. This does not mean that keylogger or worm files cannot affect systems or perform various lateral movements. Therefore, in the case of the Linux-type operating system, the fact that the antivirus used fails to effectively identify certain malicious files is an important problem.

The importance of a preliminary analysis of a file is recalled here to avoid, as much as possible, the unfavorable aspects of executing a malicious file. Thus, considering the results presented in Fig. 3 and [17], attention is drawn to the fact that, for the execution of ransomware, the working environment or even the network may become unusable. And, if the file backup area is located within the same network, it will most likely be affected as well. In all experiments performed, the performance of the virtual machine was affected by running the malicious software, with related CPU values increasing instantly, certain processes stopping or becoming temporarily inaccessible.

For benign files, there was also an increase in CPU and memory values, especially for larger files. So, in the case of the experiments carried out through the Windows operating system, the CPU and memory values increased for the .pdf and .mp4 files, while, in the case of the Ubuntu 22.04 operating system, the same values increased for the .py type files, .txt, .mp4 and tar. The only common aspect is the fact that, in both cases, the CPU and memory values increased when executing the .mp4 file because they are quite resource consuming. No conclusions can be drawn for the rest of the files, as the CPU and memory values may also increase depending on the characteristics of the file itself (its size, the information contained, etc.).

As for the area of activity of the network resources, it is different within the analyzed operating systems. Thus, if in the first scenario all malicious files register the need for access outside the network (trying to access various resources from the Internet), in the case of the second scenario, this event was registered only for the keylogger type file.

For the benign files, they did not register the need to access additional resources from the Internet, with the exception of the

html file, in Windows operating system. Although, initially, its execution can be done in an offline mode, as additional aspects are opened within it, internet resources are needed. It is very likely that, in this case too, there will be changes to the system after the working time allocated to the realization of the experiments.

Regarding the Values/Keys Modified parameter, whose value is based on the Regshot tool, but also on various processing tools such as Process Monitor, it is observed that, for all types of files analyzed (both benign and the malicious ones), there are changes at the level of various registries or keys of the operating system, regardless of whether we consider the Windows or Ubuntu 22.04 platform.

Considering all these aspects, it is necessary to open a discussion about how the execution of these files affects the performance of the analyzed test systems. Thus, taking into account both the aspects mentioned in the static analysis and those in the dynamic analysis, it is appreciated that, if a file is unknown to a user, a simple check of its hash or entropy can be a sufficient index good to see if the file is, in fact, malicious software. These aspects can be done both individually and using various tools or resources from the Internet. It is particularly important that the unknown file is not executed, if no data is known about it, in order to protect both the system through which the file was received, and the network of which it is a part, if applicable.

If the file has been executed, the way in which the performance of a system is affected includes some effective evaluation indicators. Among them, the high CPU and memory values are counted because the malicious file requires various resources to be able to increase its coverage area or to access various types of sensitive information. An immediate effect would be longer response times for applications or even the impossibility of accessing them, various crashes or even restarts of the operating systems. Another performance parameter, taken into account in the case of this work, is the access area to unknown internet resources. As a result of this fact, Internet resources can become difficult to access, and latency can increase.

Other ways in which operating systems are affected include a slower boot, changes to registries or their keys, deletion of various information or files and even their full encryption. It needs to be mentioned that since a system is compromised, it can open various backdoors for further infections.

If, when running, the file is detected by the antivirus, it will block its effective execution and delete it. But, as was observed from the experiments, sometimes antiviruses can also let malicious files pass. Therefore, increased attention is required in the case of unknown files.

Even benign files can affect the performance of operating systems. Thus, although they do not cause the same damage as malicious files, the fact that they are large files or archives with many files of various types affects the operating system. Another way in which they can impact the performance of the operating system is if they contain outdated software. These not only affect the total performance, not having any updates, but they can also create many vulnerabilities, which facilitate various types of attacks. Also, if there are files that contain errors or have modified extensions, they can also corrupt other files, even from the installer area of the operating system.

Considering all this, it is imperative that the files be verified, especially when they come from unknown sources. The verification methods can be both simple (checking the hash and comparing it with the effective extension of the file), as well as complex (which can include reverse engineering). It is also necessary to keep in mind that even benign files can affect system performance. Therefore, increased attention is needed regarding the public sources for downloading them and the way of implementation, both within the respective system and in the case of various communication networks.

In this paper, certain limitations related to the field of cybersecurity are also admitted. Among these, the method of selecting malware samples is listed, since they focused only on executables from the Windows and Linux operating systems. Also, the selection method did not have a predefined set of rules, the files being chosen according to the malware family they come from, their category and the publication date (the aim being to perform experiments with the most recent files). Thus, it is possible that files with relevant characteristics were omitted from the study. Another important limitation is given by the isolated environment in which the experiments were performed, because it was not possible to record the working mode of the malicious files at the network level and the impact that it may have on other network resources. Mentioning these limitations is important for identifying how to perform future experiments.

VI. Conclusions

In this work, a comparative analysis was presented on the way of working of various operating systems with malicious or legitimate files. Therefore, for this aspect, it was necessary to download samples of malicious files from public resources and to choose various types of legitimate files, to be able to make a comparison between them. Later, they were analyzed at the level of the Windows and Ubuntu operating systems, both through static and dynamic analysis.

The static analysis was carried out by means of open-source tools and made an identification of the malicious and benign files, without them being executed. This type of analysis was based on the identification of file hashes, their entropy, the strings used within the system, the imported libraries and the APIs called. However, given the fact that they were quite diverse and ambiguously related, they could not be the basis of a comparative analysis between the two types of operating systems used. Thus, the aspects taken into account at the level of this work were the imphash or the MD5 hash, the entropy of the files and, depending on the operating system, the imported libraries and the structural segments of the file.

In the dynamic analysis area, two sandbox virtual machines were built, with two different operating systems (Windows 10 and Ubuntu 22.04). Within these virtual machines, an isolated file execution environment was created, and the network area was simulated by means of a public tool. The analysis required the execution of the files, and this aspect led, in some cases, to the temporary unavailability of the virtual machines or even to the corruption of all the files within it.

Taking into consideration all this analysis, a discussion was also carried out on the impact that various types of files have on operating systems, either from the perspective of legitimate files or from the perspective of malicious files. The conclusion of this discussion led to underlining the importance of preliminary verification of files received from unknown authors or downloaded from various less obscure public sources, in order to prevent damage to the actual operating system or the network of which it is a part.

The most surprising conclusions that can be drawn from the experiments are given by the way in which the ransomware file, run in the Windows operating system, led to the impossibility of accessing them and by the fact that some malicious files went unnoticed by the Linux operating system antivirus. These aspects may impact the academic and work environment to carry out similar experiments to try to make antiviruses more efficient or to analyze new methods for restoring files affected by ransomware attacks.

For future work, other types of malicious files will be taken into account (adware, rootkits, bots), and the analysis will also be performed at the level of other types of operating systems (MAC OS, for desktop area, and Android and iOS, for mobile area). The final goal is to create a robust database, which contains more recent malware samples and their key features, which can be adapted to current security systems and can be considered can be considered a working basis for other experiments that will be carried out in the literature.

REFERENCES

[1] F. Ullah et al., “Data Exfiltration: A Review of External Attack Vectors and Countermeasures,” Journal of Network and Computer Applications, 2017, doi: 101. 10.1016/j.jnca.2017.10.016.

[2] A. Ghosh, “An overview article on 600% increase in Cyber Attack in 2021,” 2021, doi: 10.13140/RG.2.2.18205.52968.

[3] R. Sihwail, K. Omar, A. Zainol, A. Khairul Akram, “A Survey on Malware Analysis Techniques: Static, Dynamic, Hybrid and Memory Analysis,” 2018, doi: 8. 1662. 10.18517/ijaseit.8.4-2.6827.

[4] A. Belea, “Methods for Detecting Malware Using Static, Dynamic and Hybrid Analysis,” International Conference on Cybersecurity and Cybercrime, 10, 258–265, 2023, doi:10.19107/CYBERCON.2023.34

[5] R. Baker del Aguila, Carlos Daniel Contreras Pérez, Alejandra Guadalupe Silva-Trujillo, Juan C. Cuevas-Tello, Jose Nunez-Varela. “Static Malware Analysis Using Low-Parameter Machine Learning Models,” 2024, Computers 13, no. 3: 59, doi: 10.3390/computers13030059

[6] F. Almeida, M. Imran, J. Raik, S. Pagliarini, “Ransomware Attack as Hardware Trojan: A Feasibility and Demonstration Study,” IEEE Access, 2022, doi: 10. 44827 - 44839. 10.1109/ACCESS.2022.3168991.

[7] N. Ravichandran, T. Tewaraja, V. Rajasegaran, S. Kumar, S. Gunasekar, S. Sindiramutty, “Comprehensive Review Analysis and Countermeasures for Cybersecurity Threats: DDoS, Ransomware, and Trojan Horse Attacks,” 2024, doi:10.20944/preprints202409.1369.v1.

[8] A. Sheikh, “Trojans, Backdoors, Viruses, and Worms,” 2021, doi: 10.1007/978-1-4842-7258-9_5.

[9] B. Rajesh, P. Praveen Yadav, C. V. Chakradhar, “Malicious Computer Worms and Viruses: A Survey,” International Journal of Trend in Research and Development (IJTRD), , ISSN: 2394-9333, Special Issue | RIET-17 , December 2017, URL: http://www.ijtrd.com/papers/IJTRD13399.pdf

[10] R. Mahmoud, M. Anagnostopoulos, S. Pastrana and J. M. Pedersen, “Redefining Malware Sandboxing: Enhancing Analysis Through Sysmon and ELK Integration,” in IEEE Access, vol. 12, pp. 68624-68636, 2024, doi: 10.1109/ACCESS.2024.3400167

[11] S. Yuan et al., "Research on Vulnerability Detection Techniques Based on Static Analysis and Program Slice," 2024 6th International Conference on Electronic Engineering and Informatics (EEI), Chongqing, China, 2024, pp. 965-969, doi: 10.1109/EEI63073.2024.10696068.

[12] Y. Tian et al., “Research on Personal Privacy Security Detection Techniques for Android Applications,” 2024 9th International Conference on Electronic Technology and Information Science (ICETIS), Hangzhou, China, 2024, pp. 375-379, doi: 10.1109/ICETIS61828.2024.10593754.

[13] S. Wang et al., “A Novel Detection System for Multi-Architecture IoT Malware,” 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Tianjin, China, 2024, pp. 1758-1763, doi: 10.1109/CSCWD61410.2024.10580682.

[14] R. H. Mahdi and H. Trabelsi, “Detection of Malware by Using YARA Rules,” 2024 21st International Multi-Conference on Systems, Signals & Devices (SSD), Erbil, Iraq, 2024, pp. 1-8, doi: 10.1109/SSD61670.2024.10549308.

[15] M. F. Ismael and K. H. Thanoon, “Investigation Malware Analysis Depend on Reverse Engineering Using IDAPro,” 2022 8th International Conference on Contemporary Information Technology and Mathematics (ICCITM), Mosul, Iraq, 2022, pp. 227-231, doi: 10.1109/ICCITM56309.2022.10031698

[16] H. A. Noman, Q. Al-Maatouk and S. A. Noman, “A Static Analysis Tool for Malware Detection,” 2021 International Conference on Data Analytics for Business and Industry (ICDABI), Sakheer, Bahrain, 2021, pp. 661-665, doi: 10.1109/ICDABI53623.2021.9655866.

[17] K. Khaliq et al., “Ransomware Attacks: Tools and Techniques for Detection,” 2024 2nd International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2024, pp. 1-5, doi: 10.1109/ICCR61006.2024.10532926.

[18] H. Durgapal and D. Kumar, "Software Vulnerabilities Using Artificial Intelligence," 2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT), Greater Noida, India, 2024, pp. 1-6, doi: 10.1109/ICEECT61758.2024.10739067.

[19] Hex-rays, “IDA Pro,” Retrieved June 18, 2024 from https://hex-rays.com/ida-pro/

[20] Winitor, “pestudio” Retrieved June 19, 2024 from https://www.winitor.com/download

[21] yara, “yara v3.4.0,” Retrieved June 19, 2024 from https://yara.readthedocs.io/en/v3.4.0/gettingstarted.html

[22] GitHub, “Regshot,” Retrieved June 19, 2024 from https://github.com/Seabreg/Regshot

[23] Wirehark, “The world's most popular network protocol analyzer,” Retrieved June 19, 2024 from https://www.wireshark.org/

[24] GitHub, “flare-fakenet-ng” Retrieved June 19, 2024 from https://github.com/mandiant/flare-fakenet-ng

[25] MalwareBazaar by ABUSE, “MalwareBazaar Database” Retrieved June 20, 2024 from https://bazaar.abuse.ch/browse/

[26] MALCAT - the binary file dissector, “Malcat,” Retrieved June 20, 2024 from https://malcat.fr/index.html

[27] GitHub, “Detect-It-Easy” Retrieved June 20, 2024 from https://github.com/horsicq/Detect-It-Easy

[28] Microsoft, “Strings v2.54,” Retrieved June 21, 2024 from https://learn.microsoft.com/en-us/sysinternals/downloads/strings

[29] GitHub, “impelf,” Retrieved June 21, 2024 from https://github.com/signalblur/impelf

[30] IBM, “Analyzing files for embedded content and malicious activity”, Retrieved Nov 4, 2024 from https://www.ibm.com/docs/en/qsip/7.5?topic=content-analyzing-files-embedded-malicious-activity

[31] Chris Balles and Ateeq Sharfuddin, “Breaking Imphash”, 2019, 10.48550/arXiv.1909.07630

[32] I. Seung-Soon, “Tool interface standard (TIS) executable and linking format (ELF) specification.”, 1995.

[33] VirusTotal, “VirusTotal”, Retrieved Nov 4, 2024 from https://www.virustotal.com/gui/home/upload

A Malware Analysis Approach for Identifying Threat Actor Correlation Using Similarity Comparison Techniques

Ahmad Naim Irfan, Suriayati Chuprat, Mohd Naz'ri Mahrin, Aswami Ariffin
Universiti Teknologi Malaysia, Malaysia

***Abstract*—Cybersecurity is essential for organisations to protect critical assets from cyber threats in the increasingly digital and interconnected world. However, cybersecurity incidents are rising each year, leading to increased workloads. Current malware analysis approaches are often case-by-case, based on specific scenarios, and are typically limited to identifying malware. When cybersecurity incidents are not handled effectively due to these analytical limitations, operations are disrupted, and an organisation's brand and client trust are negatively impacted, often resulting in financial loss. The aim of this research is to enhance the analysis of Advanced Persistent Threat (APT) malware by correlating malware with its associated threat actors, such as APT groups, who are the perpetrators or authors of the malware. APT malware represents a highly dangerous threat, and gaining insight into the adversaries behind such attacks is crucial for preventing cyber incidents. This research proposes an advanced malware analysis approach that correlates APT malware with threat actors using a similarity comparison technique. By extracting features from APT malware and analysing the correlation with the threat actor, cybersecurity professionals can implement effective countermeasures to ensure that organisations are better prepared against these sophisticated cyber threats. The solution aims to assist cybersecurity practitioners and researchers in making informed decisions by providing actionable insights and a broader perspective on cyber-attacks, based on detailed information about malware tied to specific threat actors.**

***Keywords*—Malware analysis; APT group; threat actor correlation; CTI**

I. Introduction

The increasing number of cybersecurity incidents is a significant challenge faced by organisations worldwide. Throughout the year, many organisations must deal with cyber incidents involving malware. According to a report by Trend Micro, there has been a 382% increase in blocked malicious files, such as malware [1] deployed by threat actors, These threats are continuously adapting to organisations' cyber defences. Threat actors can bypass these defences due to their ever-improving modus operandi [2] [3] including the malware they use [1], which targets multiple devices, [4] such as computers and smartphones. [5] [6]. Threat actors aim to avoid detection by making it increasingly difficult to identify malicious files. This is particularly evident in cases of ransomware, where cyber-attacks are becoming more sophisticated [7]. As threat actors deploy new diversion and evasion techniques, they are able to avoid detection, highlighting the growing complexity of cyber threats [8].

Identifying malicious activities is crucial when dealing with malware found during cyber incidents, such as data breaches [12]. Cyber-attacks carried out by threat actors, particularly APT groups, are highly sophisticated and have a severe impact on victims. For example, the Lazarus Group, a state-sponsored threat actor, was reported to have attacked Automated Teller Machines (ATMs) and banks in India [13]. In addition to causing disruption, financial gain and espionage are common motivations for APT groups to execute cyber-attacks. These attacks are high-stakes because APT groups are highly motivated, skilled, and resourceful, with perpetrators often not stopping until they meet their objectives. Moreover, APT groups commonly employ stealth, anti-analysis techniques, and covert communication to evade detection, making malware analysis both difficult and time-consuming [14]. Dealing with APT groups also takes considerable time due to the scale of the cyber-attacks, which can affect targets across multiple organisations and countries.

The identification of malware, especially APT malware, is unique as it involves multiple factors and often depends on a case-by-case basis. Factors such as file type, extracted malware data, and the purpose of the malware analysis all play a role in shaping the malware analysis approach. As a result, many researchers have developed specific approaches based on these factors, such as classifying malware by type. However, current malware analysis approaches primarily focus on identifying malware itself. There is an opportunity to broaden the purpose of malware analysis by also identifying the threat actor responsible for developing the malware. Correlating malware with its associated threat actor is valuable for identifying shared features. The identification of similar features helps in discovering links between threat actors, where malware from the same actor can be correlated.

One of the factors contributing to the rising number of cybersecurity incidents is the complexity of malware analysis, which is unique and depends on a case-by-case basis. Previous research on malware analysis approaches typically focuses on factors such as the data used in experiments, features extracted, the purpose of the analysis, the medium of analysis, and how results are measured. However, current malware analysis approaches are often limited, as they primarily focus on identifying malware rather than the threat actor behind it.

Therefore, the aim of our research is to enhance an APT malware analysis approach using a similarity comparison technique to identify the threat actor.

II. Malware Analysis Approach Considerations

Malware is one of the key artefacts found in cyber-attacks and serves as a valuable source of Cyber Threat Intelligence (CTI) [9], [10], [11], data. It contains harmful code with unique signatures and behaviours [15]. Identifying these signatures and behaviours is challenging, as they are often unique to the design of the malware author. However, some malware is derived from known variants, where the signature and behavioural patterns have already been identified [16]. Current human capabilities and technologies, such as antivirus software, rely on predefined malware signature databases that require constant updates to detect new threats. The large volume of malware makes it impractical to analyse each piece manually, which is why automated technology is used to conduct these analyses.

A. Malware Group

Malware is commonly grouped by its type. Since there are many types of malware, categorising them in this way helps identify new variants within the same malware family or discover entirely new ones. Examples of common malware types include backdoors, botnets, ransomware, spyware, keyloggers, rootkits, viruses, and worms [17]. The advantage of classifying malware by type is that it enables the identification of malware families. Grouping malware in this manner improves detection accuracy, as similar types tend to share common traits. Malware type identification is achieved by analysing patterns in malware behaviour and grouping them based on these similarities.

Currently, there is a growing body of research focused on attributing malware, as it has become increasingly sophisticated. This requires analysing malware from different perspectives, such as identifying traits to group malware by platform. In addition to traditional Personal Computer (PC) malware, malware is now being developed for a wide range of platforms, including Android malware for mobile devices, Industrial Control Systems (ICS) malware for Operational Technology (OT) systems, and Internet of Things (IoT) malware for appliances connected to the internet. This approach allows malware to be grouped according to the platform it is designed to target. For example, IoT malware refers to any type of malware developed to compromise a network of connected devices, as well as the technology that facilitates communication between these devices, the cloud, and other devices within the network.

In addition to being grouped by platform, malware is also classified based on its authors, linking it to the respective threat actor. Connecting the malware to the threat actor helps gain insights into the objectives behind an attack and understand the motive of the threat actor [29]. This information is then used to build a profile of the threat actor, detailing the tools, targets, and preferred attack vectors. Having a profile of the threat actor aids in anticipating future attacks by enabling necessary preparations to enhance the organisation's cybersecurity posture. For example, incorporating known signatures and the behavioural traits of the malware into cybersecurity controls to detect or block possible threats identified [30].

The challenge in the current ecosystem is identifying specific malware groupings, rather than categorising by type, as some malware exhibits the functionalities of two or more types. For example, China Chopper is a piece of malware that displays the capabilities of a trojan, infostealer, and password brute-force attack tool, among others [18] [19]. This demonstrates that sophisticated malware has a range of functionalities, making it difficult to group strictly by type. However, despite this complexity, malware still exhibits attributes that are linked to specific threat actors, such as APT group [20]. Grouping malware by its authors enables malware analysts to attribute it to a specific threat actor group or link it to a particular threat campaign [31]. This practice enhances threat detection by providing critical insights into adversarial motives, which, in turn, facilitates proactive defence measures.

B. Malware Analysis Environment

There are various ways to build malware analysis environments, depending on the data being analysed and the specific experimental scenario. One option is to use a dedicated physical machine for performing the analysis. However, this approach is time-consuming and inflexible, as cleaning the machine and reinstalling tools after each analysis session is cumbersome. After each analysis, the machine must be cleaned, and tools need to be reinstalled. An alternative approach is to use hypervisors and preinstalled tools [21]. In this setup, the machine is simulated through virtual machines (VMs). VMs offer several advantages, including network configurations that allow for host-only connections, which prevent the machine from connecting to the internet. VMs also include a snapshot function, enabling users to capture the system's state once the machine and applications are properly configured. This snapshot is used to revert to the captured state whenever required.

C. Related Work

Related works on malware analysis approaches typically use either generic malware or APT malware for experiments. Research on APT malware often involves classification to identify APT attacks based on common features extracted from malware samples belonging to different APT groups. Additionally, features extracted from APT malware are used to distinguish between APT and non-APT malware. However, these studies do not specifically analyse PE format APT malware. Given that Windows OS is widely targeted in cyber-attacks, a dedicated extraction and analysis approach is required to gain a deeper understanding of PE-based APT malware. A comparison of related research works is presented in Table I.

TABLE I. MALWARE ANALYSIS APPROACH COMPARISON STUDY

Research Work	Malware Analysis Approach				
	Data used in Experiment	Features Extracted	Analysis Purpose	Analysis Medium	Result Measurement
Torabi, S., Dib, M., Bou-Harb, E., Assi, C., & Debbabi, M. (2021)[22]	IoT Malware (Collected using IoT-based honeypot)	Strings	Visualise Covid-related malware clusters based on strings attribute	Similarity Measurement Based on Strings Attribute to determine covid related IoT malware	Members, percentage, density
(Xu et al., 2021)[23]	APT Malware (cyber-research/APT Malware Github)	API calls	APT Malware Classification (If it is an APT malware)	Adaboost feature selection and LightGBM	Accuracy, precision, recall, F1 score
(Hu & Hsieh, 2021)[24]	APT Malware (cyber-research/APT Malware Github)	Hexadecimal and ASCII codes (APT PE samples were converted to PNG images with a fixed width of 256 pixels)	APT Malware Classification (If it is an APT malware)	Convolutional Neural Network	Avg. Train Accuracy and Max Train Accuracy
(X. Han et al., 2021)[25]	APT Malware (cyber-research/APT Malware Github)	Binary code collection and network behaviour (Grayscale image conversion)	APT Malware Classification (If it is an APT malware)	Convolutional Neural Network	Accuracy, precision, recall, F1 score
(Do Xuan & Huong, 2022)[26]	APT malware downloaded from Interactive Online Malware Sandbox (any run app)	Processes from Event ID	Classify an APT or non-APT malware	Graph Neural Network	Experimental Scenarios to measure to measure effectiveness, Accuracy, precision, recall, F1 score
Enhanced Malware Analysis Approach	Only PE file type of APT Malware (cyber-research/APT Malware Github) and vxunderground	Strings and Import Address Table	Determine features of APT malware and using similarity comparison technique to correlate with threat actor	Similarity Measurement based on strings and IAT attribute	Experiment Scenarios based on Similarity Concept

Table I presents the enhanced approach we propose, which uses the similarity comparison technique to correlate APT malware with its author. The enhanced approach is simulated through experimental scenarios designed to evaluate the results. Table I also highlights that this enhanced approach is specifically tailored for analysing PE file-type APT malware and expands the use of the similarity comparison technique, as well as the developed experimental scenarios. Therefore, this research aims to combine various techniques and methods to analyse APT malware and extract information for Cyber Threat Intelligence (CTI) purposes.

III. ENHANCED MALWARE ANALYSIS APPROACH

The enhanced approach to analysing APT malware follows the entire process flow, from feature extraction to data analysis. This approach is presented visually to demonstrate the process, which is replicable using the preferred tools and methods. The enhanced approach is illustrated in Fig. 1.

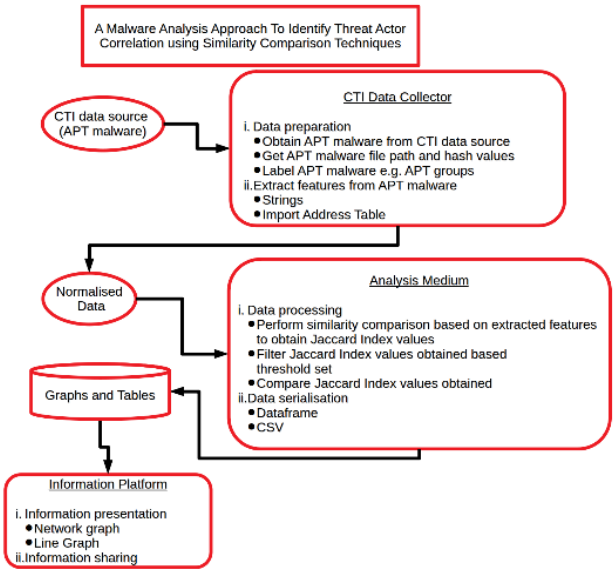


Fig. 1. Enhanced APT malware analysis approach.

Fig. 1 shows the enhanced approach for analysing APT malware to identify threat actor correlation using similarity comparison techniques. The approach takes APT malware files as input via the CTI data collector component, which performs data extraction and preparation. The normalised data is then processed and serialised by the analysis medium component. Finally, the graphs and tables generated by the analysis medium are presented and shared through the information platform component.

The first step in the approach for the CTI data collector component is to identify the APT malware source for data collection. CTI data sources are categorised into open-source, closed, or subscription-based categories. Open-source repositories, maintained by the community, include forums and websites that offer free access to content. However, the quality and availability of this content depend on how well the repository is maintained. Closed and subscription-based repositories, on the other hand, are accessible only to a select group, typically based on membership arrangements that may require payment or affiliation with an organisation. To ingest APT malware from a source, the repository typically provides a download feature to store data files locally or an API to pull data to an external storage location, such as a server or cloud.

Once APT malware is in the designated storage location, the contents of the files are extracted to obtain the relevant data. Data extraction is performed using specialised tools, which may be open-source or proprietary. Alternatively, a custom program can be written to perform the extraction process. This approach offers greater flexibility, as the features are not restricted by the limitations of open-source or proprietary data extraction tools. To develop a data extraction program, programming languages and scripting libraries typically include modules to read files and extract data based on the file format.

Initial analysis is performed on the extracted APT malware data. This step helps identify key content within the data and prepares it for subsequent processing. By reviewing the extracted data, relevant attributes are selected for analysis, and appropriate data structures are chosen based on these observations. Additionally, this step includes data labelling, which adds context to the content by tagging relevant data. Observing the data also provides an overview of its contents, allowing for the identification and removal of noise through sanitisation. Sanitisation involves filtering out irrelevant file types and removing empty rows to refine the data.

Once the data is prepared during the data collection phase, analysis mediums are used to transform it into meaningful information. At this point, the data is still in its raw format, and processing is necessary to derive actionable insights. Analysis mediums, typically built using programming languages, scripts, or available tools, perform the required processing and generate output results. Libraries and modules play a crucial role in this process, as they save time in developing the analysis mediums and allow the developer to focus more on the logic needed to process the data.

The first decision when starting data processing is to choose the appropriate algorithm for the task. In artificial intelligence, the two main options are machine learning and deep learning. The next decision is to select the data processing algorithm, which is based on the data preparation step completed earlier. A data processing algorithm is a series of instructions designed to process the data. Libraries are often used to incorporate common functions such as validation, sorting, summarising, and aggregation into the algorithm. These libraries vary across programming languages or scripts, and custom functions are written to perform tasks beyond the available scope. Pseudocode is frequently shared by the community to assist in building custom functions, and many libraries are specifically designed to support AI.

Once the data is processed, serialisation is performed to convert an object into a stream of bytes for storing the results. Common formats for storing results include CSV for tabular data and PNG for images. Serialised data are records kept for future reference when needed. Choosing the appropriate format for saving data is crucial to ensure that it is both preserved and easily shareable through the information platform.

All results obtained from the analysis are gathered on the information platform. Web-based platforms, such as blogs, wikis, and dashboards, are used to transmit and display these results. A web-based information platform is chosen because it allows for customisation in how information is presented and provides graphical tools to assist in visualising the data. Additionally, sharing features and APIs are available on web-based platforms to relay information to relevant parties.

Based on the results obtained, suitable ways to present the information include visuals such as graphs, histograms, bar charts, pie charts, and tables. These visuals help describe and interpret the data, assisting recipients in the decision-making process. Multiple visual options are available to present the information effectively, depending on the context. Key considerations when presenting information include the purpose, the recipient's background, and the structure of the information flow. To manage these visuals and considerations, a platform like a dashboard is often used to consolidate all the information in one view. A dashboard allows recipients to access information and make queries efficiently.

Before sharing information containing the analysis results, the parties who need to receive the information are identified. This step is crucial to prioritise information sharing based on the roles of personnel within the organisation. It functions as a call tree, alerting relevant personnel according to a layered, hierarchical communication model, ensuring the right people are notified of the threat. This enables the necessary preparations to be made in response to the threat. Communication methods include multiple channels, such as email, chat, SMS, and voice calls for emergencies. The information platform serves as a central medium, accessible to recipients, allowing them to pull the information when needed.

A. Malware Analysis Study

A malware analysis environment is established to experiment with and evaluate the proposed APT malware analysis approach. This environment integrates open-source tools and custom scripts to extract and compare malware features, facilitating the identification of similarities across various APT malware datasets. The environment is designed to align with the APT malware analysis methodology, simulating the analysis process using data derived from APT malware samples. The results of these experiments are assessed by reviewing the analysis outcomes. The APT malware analysis approach serves as the foundation for the environment's architecture, which is built using open-source technologies. Tools are integrated into the environment to perform similarity comparisons, offering valuable insights for

cybersecurity practitioners. These insights enable prompt actions, such as malware detection, as illustrated in Fig. 2.

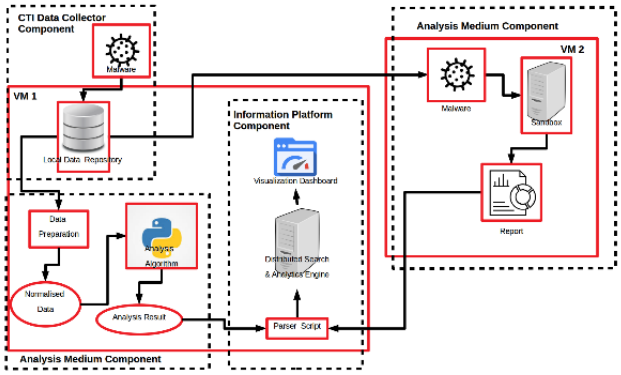


Fig. 2. Malware analysis environment.

Fig. 2 illustrates the architecture of the malware analysis environment, which implements the proposed APT malware analysis approach. Two Virtual Machines (VMs) are used in the environment to manage the integration of components separately. VM 2 runs Windows 10 for sandbox deployment, while VM 1 is used for tool installation and runs Ubuntu. The requirements and installation procedures for each tool or software used are documented on the respective tool websites. The suitability and functionality of the tools are evaluated in advance to avoid any installation or deployment issues that could hinder their operation. The tools used in the malware analysis environment are listed in Table II.

TABLE II. MALWARE ANALYSIS ENVIRONMENT

Malware Analysis Environment	Description
Host Hardware Specification	• CPU: Intel i7-13700F • RAM: 64GB (16GB assigned to VM1 and VM2)
Dataset 1 Publicly available dataset from https://github.com/cyber-research/APTMalware	The APT Malware Dataset contains over 3,500 malware samples in various file formats, such as .exe and .pdf, which are associated with 12 APT groups allegedly sponsored by five different nation-states. This dataset is primarily used for benchmarking different machine learning approaches in the context of authorship attribution. It can also serve as a valuable resource for future benchmarks or malware research.
Dataset 2 Publicly available dataset from VX Underground (https://vx-underground.org/)	The malware repository is regularly updated and maintained, with each sample properly attributed to specific cyber incidents and threat actors based on CTI reports. As such, malware from this repository serves as a suitable dataset for this research experiment. Additionally, this repository has been used as an experimental dataset in the research by Piskozub et al. (2021) [27]. The APT malware samples from 2022 and 2023 were used in this experiment.
Virtual Machine 1 (VM 1, Linux Machine)	The experiment tools installed on the virtual machine include Jupyter Lab is used to write the Python algorithms that is used for analysis durin the experiment. Elasticsearch which is a ditributed search and analysis engine is used to store experiment results obtained and the results are visualised on the Kibana (dashboard).The experiment tools installed on the virtual machine include Jupyter Lab, which is used to write the Python algorithms for analysis during the experiment, and Elasticsearch, a distributed search and analysis engine, which is used to store the experiment results. The results are then visualised on the Kibana dashboard.
Virtual Machine 2 (VM 2, Windows Machine)	Cuckoo Sandbox is a sandbox environment used for malware analysis and containment, designed to prevent outbreaks. It provides an analysis report on a given malware sample based on detection rules such as YARA (a tool commonly used in malware research and detection) to determine whether the sample is benign or malicious.

Table II describes the tools used in the malware analysis environment, which has been built according to the proposed APT malware analysis approach. This environment can be customised by replacing the existing data feeds, tools, or software with alternatives, as long as they offer the functionality outlined in the proposed approach. The tools employed in the development of the malware analysis environment are open-source and available for free download from their respective websites. Additionally, scripting is required to integrate the tools and perform tasks that require specific libraries or unique functionality. For instance, Python scripts are used to execute functions that are not provided by the selected technology providers.

B. Malware Analysis Experiment Design

The flow of the malware analysis experiment follows the enhanced malware analysis approach, which consists of six stages: data extraction from files, data preparation, data processing, data serialisation, information presentation, and information sharing. In the experiment using VM1, static analysis is performed to extract malware features such as strings and the import address table. These extracted features are then used for similarity comparison, which is conducted using the Jaccard Index, as shown in Fig. 3.

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Fig. 3. Jaccard index.

Fig. 3 shows the Jaccard Index, which is used to gauge the similarity of sample sets by measuring the ratio of the intersection to the union of the two sets (Yang et al., 2023). The Jaccard Index between two sets is calculated by dividing the number of elements in their intersection by the number of elements in their union. The value of the Jaccard Index, denoted as J(A,B), lies between 0 and 1, where $0 \leq J(A,B) \leq 1$. For example, if the intersection of sets A and B is empty, then J(A,B)=0, indicating no similarity between the two sets of assembly code [28]. This is one of the similarity comparison techniques identified for assessing similarities in malware attributes.

Dataset 1, sourced from the GitHub link mentioned in Table II, contains 3,594 APT malware samples. During the data preparation step in the CTI data collector component, 2,887 PE files are filtered out of the 3,594 APT malware samples in the dataset. For Dataset 2, which is pulled from Vxunderground, 596 PE files from 2023 and 3,446 PE files from 2022 are filtered during the same data preparation step in the CTI data collector component.

Two attributes are selected from the observations during the data preparation step. These chosen attributes—strings and the Import Address Table (IAT)—are analysed during the experiment to identify similarities in APT malware. Based on these two attributes, three experiment scenarios are designed to help security professionals, researchers, and organisations understand how to identify similarities between different malware samples. Table III describes the role of these attributes and outlines the execution of the experiment scenarios.

TABLE III. EXPERIMENT SCENARIO IMPLEMENTATION

Experiment Scenarios	Implementation
String comparison	Strings are ASCII and Unicode printable sequences of characters embedded within a file. The strings attribute refers to the human-readable text embedded within a binary file, often revealing useful information such as URLs, file paths, error messages, and even internal function names. Malware analysts frequently examine these strings to gain insights from the malware, such as its behaviour and command-and-control (C2) information. The string comparison scenario is an experiment designed to analyse APT malware and calculate the Jaccard Index between combinations of the 2,888 PE files identified. In this scenario, only Dataset 1 is used to identify the similarity between two distinct samples based on string attribute.
IAT Comparison	The Import Address Table (IAT) is part of a Windows module—either an executable or a dynamic link library (DLL)—that records the addresses of functions imported from other DLLs. The IAT provides insight into the specific system resources and APIs used by the malware, which is valuable for identifying patterns across different samples.. The IAT comparison scenario is an experiment designed to analyse APT malware and calculate the Jaccard Index between combinations of the 2,888 PE files identified. In this scenario, only Dataset 1 is used to identify the similarity between two distinct samples based on the IAT attribute.
Similarity Comparison on Dataset 1 and Dataset 2	In this experiment scenario, the strings and IAT attributes are extracted from APT malware in both Dataset 1 and Dataset 2. The similarity comparison is then performed by calculating the Jaccard Index between the APT malware samples in Dataset 1 and Dataset 2 based on these two extracted attributes. The similarity comparison is performed to identify any correlation between the two different APT malware datasets based on the string and IAT attributes. The comparison between Dataset 1 and Dataset 2 begins with the year 2023, followed by the year 2022.

Table III illustrates the experiment flow, which includes three scenarios. These experiment scenarios are designed based on the dataset and the chosen attribute similarity comparisons, which include strings and IAT, as described in detail. The scenarios are executed to demonstrate how the normalised data are analysed and represented in graphs and tables. The sample data collected are shown in Table IV.

TABLE IV. SAMPLE DATA COLLECTED

Malware Hash	Malware Label	String Attributes	IAT Attributes
00be6858156b0be404b4fa4852ffc550c25565236beaa4cb13ffe288bcb48d8e	APT 1	Syntax error! Usage: getf/putf FileName <N> Mozilla/5.0 So long! exit Shell started,wait to terminate it..... Service is running already! Service started! StartService failed! CreateProcess failed! Program started! Syntax error!	CloseServiceHandle ControlService CreateFileA CreatePipe CreateProcessA'", CreateProcessAsUserA CreateThread CreateToolhelp32Snapshot

		Usage: start </p|/s> <filename|ServiceName>	

Table IV lists sample data extracted from the APT malware. The fields identified for analysis, which are relevant to the experiment, include Hash, Label, Strings, and IAT. The comparison is then performed by calculating the Jaccard index values for the Strings and IAT attributes. This comparison, using the Jaccard index, is based on the bag of features concept, as shown in Fig. 4.

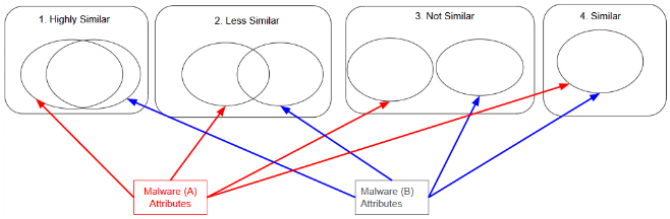


Fig. 4. Malware attributes similarity concept.

Fig. 4 illustrates the concept of attribute similarity used in malware analysis, where the similarity of malware attributes is calculated using a similarity coefficient, such as the Jaccard index. Based on the threshold of 0.8 for Jaccard index calculation: Diagram 1 represents an instance where the calculated Jaccard index value is greater than 0.8; Diagram 2 shows an instance where the Jaccard index value is less than 0.8; Diagram 3 depicts an instance where the Jaccard index value is 0. Finally, Diagram 4 illustrates an instance where the malware attributes exactly match, and the Jaccard index value is 1.

IV. Malware Analysis Results

Experiments are conducted to demonstrate the results obtained from the implementation of the enhanced approach. The results are based on three scenarios, using the malware attribute similarity concept described in Section 3B. Two distinct samples, identified by hash value, are obtained from the malware dataset and labelled as Malware 1 and Malware 2. For each comparison performed based on the designed scenario, the Jaccard Index value is calculated to identify the similarity between the compared samples.

A. Strings Similarity Comparison

The string comparison scenario in the experiment compares the string attributes of each APT malware sample. For each APT malware analysed, the string values are extracted and saved in a CSV file. A sample of the results from the similarity comparison of string attributes extracted from APT malware is presented in Table V.

TABLE V. Sample Jaccard Index Results for Strings Attribute

Malware 1 Hash	Malware 1 Label	Malware 2 Hash	Malware 2 Label	String Jaccard Index
0fbb47373b8bbefdfd9377dc2	AP	0fbb47373b8bbefdfd9377dc2	AP	1
6b6418d2738e6f688562885f4d2a1a049e4948e	T1	6b6418d2738e6f688562885f4d2a1a049e4948e copy	T1	
6c8eb3365b7fb7683b9b465817e5cb87574026e306c700f3d103eba056777720	AP T29	6c8eb3365b7fb7683b9b465817e5cb87574026e306c700f3d103eba056777720	AP T29	1

Table V shows sample Jaccard index results for the string attribute. The first row presents a sample where the Jaccard index value for the string attribute is 1. This result indicates that the malware is identical, as it represents the same sample with a similar hash value.

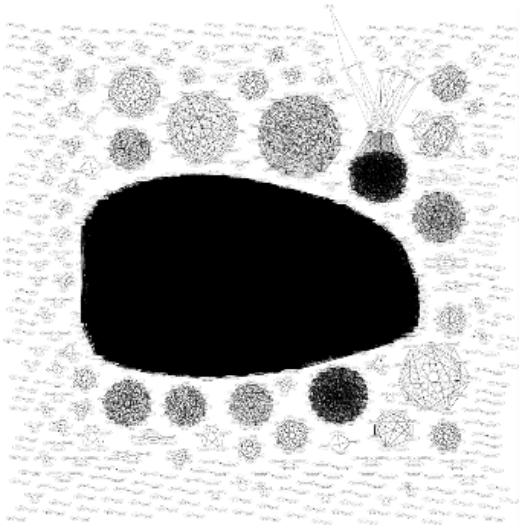


Fig. 5. Strings network graph.

Fig. 5 shows the results of the string attribute comparison between malware from different APT groups. Nodes without at least one edge are removed from the graph to reduce clutter.

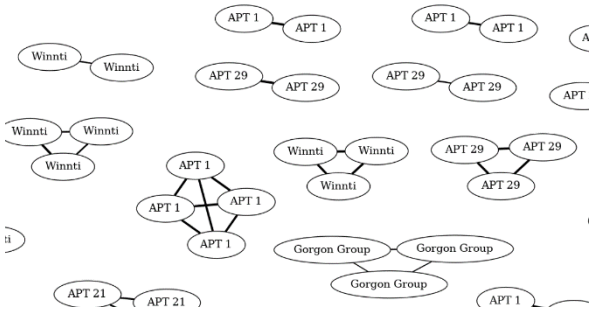


Fig. 6. Close up of strings network graph.

Fig. 6 provides a close-up of Fig. 5. The APT malware samples are grouped based on the string attribute. Observations of the results reveal that some malware, even from the same APT group, are not clustered together. Additionally, no malware from different APT groups shows any connections.

B. Imports Address Table Similarity Comparison

The IAT comparison scenario in the experiment compares the IAT attributes of each APT malware sample. For each APT malware analysed, the IAT values are extracted and saved in a CSV file. A sample of the results from the similarity comparison of IAT attributes extracted from APT malware is presented in Table VI.

TABLE VI. SAMPLE JACCARD INDEX RESULTS FOR IAT ATTRIBUTE

Malware 1 Hash	Malware 1 Label	Malware 2 Hash	Malware 2 Label	IAT Jaccard Index
0fbb47373b8bbefdfd9377dc26b6418d2738e6f688562885f4d2a1a049e4948e	APT1	0fbb47373b8bbefdfd9377dc26b6418d2738e6f68856285f4d2a1a049e4948e	APT1	1
1b3ee0274ae0ac0b83dba7f95f00e2381a5d3596d136eb1fac842a07d8d25262	APT1	6bb764f3a5ca57f9bcc72aa0c34dab64e870e22c6400f6b3f62d5986104dc68f	APT1	0.82828282828 2828
6c7e768e48b9b225b7b9f84528c53c2e6f9b639ce2e7919fe0dff9aad07ea4f5	APT29	6c8eb3365b7fb7683b9b465817e5cb87574026e306c700f3d103eba056777720	APT29	0.94845360824 7423
6c8eb3365b7fb7683b9b465817e5cb87574026e306c700f3d103eba056777720	APT29	6c8eb3365b7fb7683b9b465817e5cb87574026e306c700f3d103eba056777720	APT29	1

Table VI shows two identical samples in rows 1 and 4, which are malware from the same APT group with a Jaccard Index value of 1. Rows 2 and 3 show malware with different hashes but from the same APT group, with Jaccard Index values of 0.83 and 0.95, respectively. The closer the Jaccard Index values are to 1, the greater the similarity in the IAT attributes.

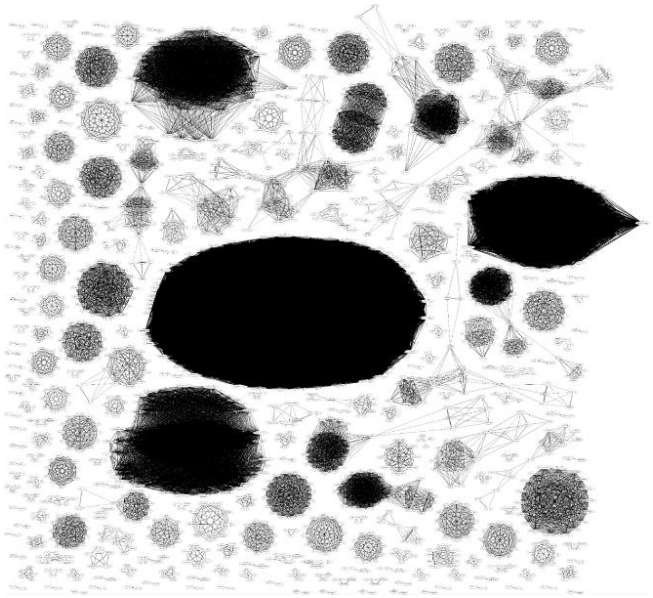


Fig. 7. IAT Network graph.

Fig. 7 shows the results of the IAT comparison between malware from different APT groups. The IAT network graph differs from the string network graph shown in Fig. 5. A close-up of the IAT network graph is presented in Fig. 8.

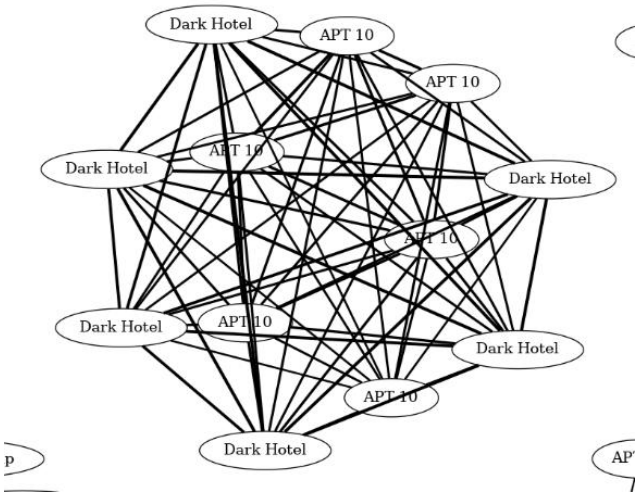


Fig. 8. Close up of IAT network graph.

In contrast to Fig. 6 from the string comparison experiment, Fig. 8 shows that in the IAT network graph, there are overlaps in IAT values for malware from different APT groups. This example demonstrates that there are overlapping attributes between malware from two different APT groups, which are grouped based on attributed threat actors.

C. Strings and Import Address Table Similarity Comparison

The similarity comparison scenario compares features from Dataset 2 (2023) and Dataset 2 (2022) with features collected from APT malware in Dataset 1. In this scenario, the strings and IAT attributes extracted from Dataset 2 are compared with the corresponding strings and IAT attributes extracted from malware samples of 12 APT groups in Dataset 1.

The similarity comparison of 596 samples from Dataset 2 (2023) with 2,887 samples from Dataset 1 took 24 minutes and 19 seconds, resulting in 1,720,652 similarity comparisons. A summary of the results is shown in Table VII.

TABLE VII. SUMMARY OF SIMILARITY COMPARISON RESULTS FOR DATASET 1 AND DATASET 2 (2023)

Similarity Comparison	Greater than 0.8	Greater than 0.5	Lower than 0.5	Lower than 0.2
Strings	0	121	1,720,531	1,715,434
IAT	9643	33673	1,684,723	1,472,266

The results of the string similarity comparison show that no Jaccard Index value exceeds 0.8. However, further examination of the results reveals that 28 samples have a Jaccard Index value greater than 0.6, with some samples exceeding 0.5. The outcome of the string similarity comparison is presented in Fig. 9.

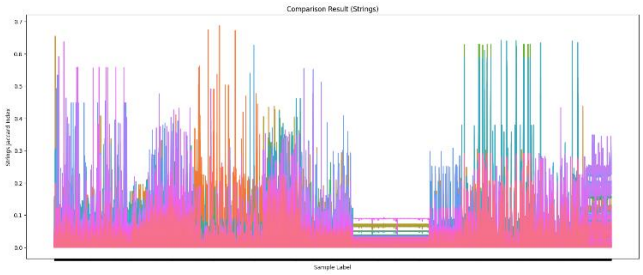


Fig. 9. Similarity comparison result line graph (Strings).

Fig. 9 provides an overview of the string similarity comparison results. The graph shows that, out of the 1,720,652 similarity comparisons performed, none of the Jaccard index values exceed 0.7. Since the Jaccard Index values in the results are below the set threshold, it is likely that there are no significant similarities between the malware in Dataset 2 (2023) and Dataset 1.

A different result was obtained for the IAT similarity comparison, with 9,643 samples having a Jaccard Index value greater than 0.8, as shown in Table VII and represented in Fig. 10.

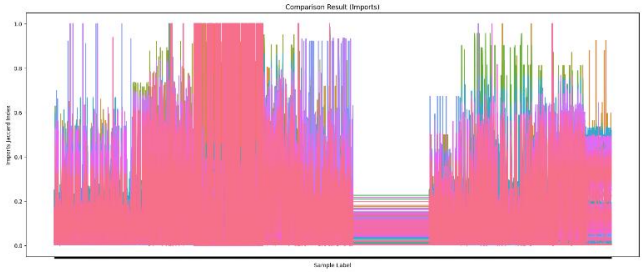


Fig. 10. Similarity comparison result line graph (IAT).

Fig. 10 provides an overview of the 1,720,652 IAT similarity comparison results. A correlation of 9,643 samples that scored a Jaccard Index higher than 0.8 is highlighted in the IAT network graph shown in Fig. 11.

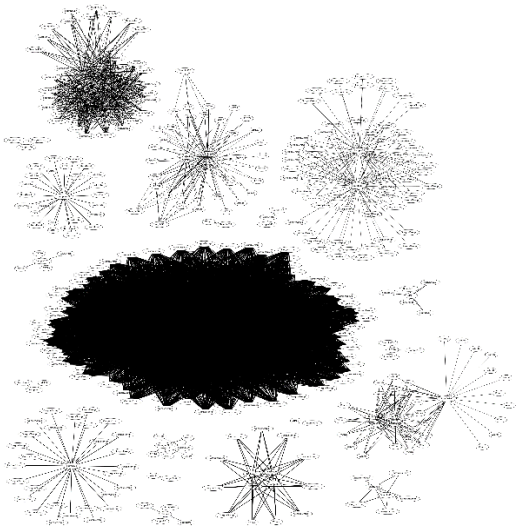


Fig. 11. IAT network graph.

Fig. 11 depicts the results of the IAT comparison between Dataset 2 (2023) and Dataset 1. The IAT network graph shows multiple correlations. One of the correlations identified is shown in Fig. 12, which provides a close-up of the IAT network graph from Fig. 11.

Fig. 12. Close up IAT network graph.

Fig. 12 shows that Sample 160, Sample 161, and Sample 162 are correlated with malware attributed to APT 10 and APT 28, based on the IAT similarity comparison. Additionally, Sample 368 and Sample 371 are correlated with malware identified as being used in APT 28 malicious operations. This suggests that, since the results scored higher than 0.8, the IAT attributes for these samples overlap with those of APT 10 and APT 28 malware.

The experiment continued with the similarity comparison of 3,446 samples from Dataset 2 (2022) with 2,887 samples from Dataset 1, which took 2 hours, 52 minutes, and 45 seconds. This resulted in 9,948,602 similarity comparisons. A summary of the results is shown in Table VIII.

TABLE VIII. SUMMARY OF SIMILARITY COMPARISON RESULTS FOR DATASET 1 AND DATASET 2 (2022)

Similarity Comparison	Greater than 0.8	Greater than 0.5	Lower than 0.5	Lower than 0.2
Strings	4	299	9,948,301	9,901,641
IAT	48,110	196,401	9,741,018	8,380,710

Table VIII shows that for the string similarity comparison, 299 samples scored above 0.5, and four samples scored above 0.8. The results of the string similarity comparison are represented in Fig. 13.

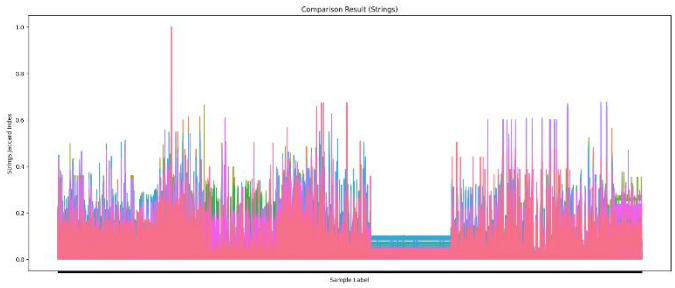


Fig. 13. Similarity comparison result line graph (Strings).

Based on Fig. 13, it is clear that there is a single instance where the Jaccard Index value is 1.0. The results of the line graph for the IAT similarity comparison are shown in Fig. 14.

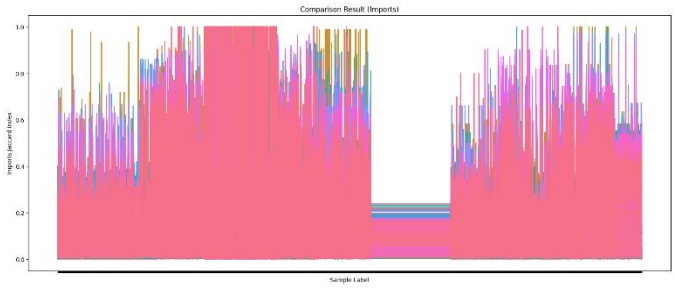


Fig. 14. Similarity comparison result line graph (IAT).

Fig. 14 depicts the visualisation of the results for 48,110 samples with Jaccard Index values greater than 0.8 in the IAT similarity comparison. We then further investigate the results from Table VIII and Fig. 13 to better understand the findings. The 4 samples with Jaccard Index values greater than 0.8 for the string similarity comparison are filtered from the others. The results for the string similarity comparison above 0.8 are shown in Table IX.

TABLE IX. STRINGS SIMILARITY COMPARISON RESULT FOR DATASET 1 AND DATASET 2 (2022)

Sample Hash	Sample Label	Malware Hash	Malware Label	String Jaccard Index	IAT Jaccard Index
c9d5dc956841e000bfd8762e2f0b48b66c79b79500e894b4efa7fb9ba17e4e9e	Sample243	c9d5dc956841e000bfd8762e2f0b48b66c79b79500e894b4efa7fb9ba17e4e9e	APT10	1.0	1.0
fa7eee6e322bfad1bb0487aa1275077d334f5681f0b4ede0ee784c0ec1567e01	Sample809	c9d5dc956841e000bfd8762e2f0b48b66c79b79500e894b4efa7fb9ba17e4e9e	APT10	1.0	1.0
c9d5dc956841e000bfd8762e2f0b48b66c79b79500e894b4efa7fb9ba17e4e9e	Sample3063	c9d5dc956841e000bfd8762e2f0b48b66c79b79500e894b4efa7fb9ba17e4e9e	APT10	1.0	1.0
c9d5dc956841e000bfd8762e2f0b48b66c79b79500e894b4efa7fb9ba17e4e9e	Sample3427	c9d5dc956841e000bfd8762e2f0b48b66c79b79500e894b4efa7fb9ba17e4e9e	APT10	1.0	1.0

Table IX lists the four samples for which the Jaccard Index values for strings are greater than 0.8. Observations from both the string and IAT similarity comparisons show that the Jaccard Index obtained is high, with a score of 1.0. This indicates that Sample 243, Sample 809, Sample 3063, and Sample 3427 exactly match APT10 malware, which has the hash"c9d5dc956841e000bfd8762e2f0b48b66c79b79500e894b4efa7fb9ba17e4e9e". This correlation is also reflected in the strings network graph shown in Fig. 15.

Fig. 15. Correlation identified in strings and imports network graph.

Based on the correlation shown in Fig. 15, a relationship between APT 10 malware and Sample 809, which has the hash "fa7eee6e322bfad1bb0487aa1275077d334f5681f0b4ede0ee784c0ec1567e01," is identified. Since Sample 243, Sample 3063, and Sample 3427 have the same hash as the APT 10 malware, they are not shown in the network graph in Fig. 15. This also explains the result obtained in Fig. 13, which indicates the similarity between APT 10 malware and those samples. Therefore, only the correlation of Sample 809 to APT 10 malware, based on the string similarity comparison, is shown, as only Sample 809 has a different hash. This suggests that the incident involving Sample 243, Sample 809, Sample 3063, and Sample 3427 is likely linked to APT 10, since these samples share similar attributes with malware already attributed to this APT group, which is believed to be linked to China.

Based on the information obtained, these findings can be used for CTI (Cyber Threat Intelligence) purposes. The report accompanying Dataset 2 from Vxunderground states that Sample 243 is Nbtscan, discovered by Avast; Sample 809 is NBTScan, discovered by Symantec; Sample 3063 is a NetBIOS scanner, discovered by Trend Micro; and Sample 3427 is F01A9A2D1E31332ED36C1A4D2839F412, discovered by Kaspersky, where only the MD5 value is provided in the IOC section. All of these reports attribute the samples to APT groups possibly linked to China, such as Mustang Panda and Earth Lusca. Therefore, the analysis and results obtained in this research provide a valid correlation.

By knowing that the malware found is possibly attributed to a specific APT group, organisations can better prepare to defend against the threat. For example, based on our results, if the sample found in an organisation is linked to APT 10, threat hunting efforts can focus on looking for IOCs (Indicators of Compromise) and activities associated with APT 10 or related APT groups, based on past incidents involving those groups. Our research demonstrates how string and IAT attributes can be used in similarity comparison scenarios, with extracted features being correlated to the threat actor through visual information presentation.

V. CONCLUSION

In general, the enhanced APT malware analysis approach extracts attributes from PE files and uses these to correlate with threat actors. This helps identify the origin of malware through the Jaccard Index, a similarity comparison technique used to establish threat actor correlations. The information

obtained can be leveraged to develop countermeasures against cyber threats. The development of additional malware analysis systems and experiments performed in this research includes technical discussions and examples to deepen understanding of the formulated APT malware analysis approach. This solution aims to assist cybersecurity practitioners and researchers in making informed decisions by providing actionable insights and a comprehensive perspective on cyber-attacks, based on the analysis of artefacts from APT groups.

Our experiment identified correlations between four samples and malware attributed to APT 10. Our analysis of the results also validates the findings obtained during the experiment. The enhanced APT malware analysis approach, the malware analysis environment, and the experimental scenarios developed in this research provide a foundation for discovering threat actor correlations. Our work provides a foundation for correlating malware with the threat actor, and the malware analysis approach can be used in designing other experimental scenarios.

Extending the experimental scenarios is one possible avenue for future work. Developing additional scenarios would uncover more insights from the APT malware dataset. Additionally, the malware analysis environment could be improved by using hardware with higher specifications, which would enable faster analysis, and by incorporating tools that are more preferred or offer better functionality.

Another direction for future work is refining the dataset used, or adopting a different dataset that is better suited to the experiment. Our research used Dataset 1, which, although relatively outdated, is well-structured, making it easier to label samples with the attributed threat actor. Moving forward, we plan to use Dataset 2, which is more recent but requires additional effort for labeling. If publicly available datasets were better structured or properly labeled, the sample analysis process would be much easier, and the labeling step would be significantly streamlined.

Apart from that, other similarity comparison techniques could be explored for future work, incorporating AI—such as machine learning or deep learning algorithms—into the approach. Our current work, using the Jaccard Similarity Index, aims to conduct a preliminary analysis of the dataset and obtain results that will help develop the malware analysis approach with similarity comparison techniques, as well as design the malware analysis environment. Both the enhanced approach and the malware analysis environment are integral for analysing APT malware to extract information for identifying the threat actor.

ACKNOWLEDGMENT

This research was fully funded by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (5F511) awarded to Suriayati Chuprat, Universiti Teknologi Malaysia.

REFERENCES

[1] CrowdStrike, “CrowdStrike 2023 Global Threat Report.” Accessed: May 27, 2023. [Online]. Available: https://www.crowdstrike.com/global-threat-report/

[2] S. Cobb and A. Lee, “Malware is called malicious for a reason: The risks of weaponizing code,” Int. Conf. Cyber Conflict, CYCON, vol. 2014, pp. 71–84, 2014, doi: 10.1109/CYCON.2014.6916396.

[3] C. Rong, G. Gou, C. Hou, Z. Li, G. Xiong, and L. Guo, “UMVD-FSL: Unseen Malware Variants Detection Using Few-Shot Learning,” Proc. Int. Jt. Conf. Neural Networks, vol. 2021-July, pp. 9–16, 2021, doi: 10.1109/IJCNN52387.2021.9533759.

[4] L. M. Zagi and B. Aziz, “Searching for malware dataset: A systematic literature review,” 2020 Int. Conf. Inf. Technol. Syst. Innov. ICITSI 2020 - Proc., pp. 375–380, 2020, doi: 10.1109/ICITSI50517.2020.9264929.

[5] A. Amamra, C. Talhi, and J. M. Robert, “Smartphone malware detection: From a survey towards taxonomy,” Proc. 2012 7th Int. Conf. Malicious Unwanted Software, Malware 2012, pp. 79–86, 2012, doi: 10.1109/MALWARE.2012.6461012.

[6] R. Nigam, R. K. Pathak, A. Kumar, and S. Prakash, “PCP Framework to Expose Malware in Devices,” Proc. Int. Conf. Electron. Sustain. Commun. Syst. ICESC 2020, pp. 651–654, 2020, doi: 10.1109/ICESC48915.2020.9155593.

[7] I. Bello et al., “Detecting ransomware attacks using intelligent algorithms: recent development and next direction from deep learning and big data perspectives,” J. Ambient Intell. Humaniz. Comput., vol. 12, no. 9, pp. 8699–8717, 2021, doi: 10.1007/s12652-020-02630-7.

[8] U. Urooj, B. A. S. Al-Rimy, A. Zainal, F. A. Ghaleb, and M. A. Rassam, “Ransomware Detection Using the Dynamic Analysis and Machine Learning: A Survey and Research Directions,” Appl. Sci., vol. 12, no. 1, 2022, doi: 10.3390/app12010172.

[9] V. Mavroeidis and J. Brule, “A nonproprietary language for the command and control of cyber defenses – OpenC2,” Comput. Secur., vol. 97, p. 101999, 2020, doi: 10.1016/j.cose.2020.101999.

[10] F. Martinelli, F. Mercaldo, V. Nardone, A. Santone, A. K. Sangaiah, and A. Cimitile, “Evaluating model checking for cyber threats code obfuscation identification,” J. Parallel Distrib. Comput., vol. 119, pp. 203–218, 2018, doi: 10.1016/j.jpdc.2018.04.008.

[11] T. Dargahi, A. Dehghantanha, P. N. Bahrami, M. Conti, G. Bianchi, and L. Benedetto, “A Cyber-Kill-Chain based taxonomy of crypto-ransomware features,” J. Comput. Virol. Hacking Tech., vol. 15, no. 4, pp. 277–305, 2019, doi: 10.1007/s11416-019-00338-7.

[12] Y. Roumani, “Detection time of data breaches,” Comput. Secur., vol. 112, p. 102508, 2022, doi: 10.1016/j.cose.2021.102508.

[13] ENISA, ENISA Threat Landscape 2022, no. November. 2022. [Online]. Available: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2021

[14] A. Sharma, B. B. Gupta, A. K. Singh, and V. K. Saraswat, “Orchestration of APT malware evasive manoeuvers employed for eluding anti-virus and sandbox defense,” Comput. Secur., vol. 115, p. 102627, 2022, doi: 10.1016/j.cose.2022.102627.

[15] Y. Zhou and X. Jiang, “Dissecting Android malware: Characterization and evolution,” Proc. - IEEE Symp. Secur. Priv., no. 4, pp. 95–109, 2012, doi: 10.1109/SP.2012.16.

[16] K. Rieck, T. Holz, C. Willems, P. Düssel, and P. Laskov, “Learning and Classification of Malware Behavior,” in Detection of Intrusions and Malware, and Vulnerability Assessment, vol. 7591, no. July, U. Flegel, E. Markatos, and W. Robertson, Eds., in Lecture Notes in Computer Science, vol. 7591. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 108–125. doi: 10.1007/978-3-540-70542-0_6.

[17] C. P. Chenet, A. Savino, and S. Di Carlo, "A Survey on Hardware-Based Malware Detection Approaches," IEEE Access, vol. 12, pp. 54115–54128, 2024, doi: 10.1109/ACCESS.2024.3388716.

[18] T. Lee, I. Ahl, and D. Hanzlik, "Breaking Down the China Chopper Web Shell - Part I." Accessed: Jan. 17, 2024. [Online]. Available: https://www.mandiant.com/resources/blog/breaking-down-china-chopper-web-shell-part-i

[19] TONY LEE, IAN AHL, and DENNIS HANZLIK, "Breaking Down the China Chopper Web Shell - Part II | Mandiant," Mandiant, 2013, [Online]. Available: https://www.mandiant.com/resources/blog/breaking-down-the-china-chopper-web-shell-part-ii

[20] L. Rochberger, T. Fakterman, and R. Falcone, "Unit 42 Researchers Discover Multiple Espionage Operations Targeting Southeast Asian Government." [Online]. Available: https://unit42.paloaltonetworks.com/analysis-of-three-attack-clusters-in-se-asia/

[21] M. Figueroa, "Building a Custom Malware Analysis Lab Environment." [Online]. Available: https://www.sentinelone.com/labs/building-a-custom-malware-analysis-lab-environment/

[22] S. Torabi, M. Dib, E. Bou-Harb, C. Assi, and M. Debbabi, "A Strings-Based Similarity Analysis Approach for Characterizing IoT Malware and Inferring Their Underlying Relationships," IEEE Netw. Lett., vol. 3, no. 3, pp. 161–165, 2021, doi: 10.1109/lnet.2021.3076600.

[23] N. Xu, S. Li, X. Wu, W. Han, and X. Luo, "An APT Malware Classification Method Based on Adaboost Feature Selection and LightGBM," in Proceedings - 2021 IEEE 6th International Conference on Data Science in Cyberspace, DSC 2021, 2021. doi: 10.1109/DSC53577.2021.00101.

[24] Y. H. F. Hu and C. C. G. Hsieh, "A Study of Classifying Advanced Persistent Threats with Multi-Layered Deep Learning Approaches," in 19th IEEE International Symposium on Parallel and Distributed Processing with Applications, 11th IEEE International Conference on Big Data and Cloud Computing, 14th IEEE International Conference on Social Computing and Networking and 11th IEEE Internation, 2021. doi: 10.1109/ISPA-BDCloud-SocialCom-SustainCom52081.2021.00220.

[25] X. Han, C. Li, X. Li, and T. Lu, "Research on APT Attack Detection Technology Based on DenseNet Convolutional Neural Network," in Proceedings - 2021 International Conference on Computer Information Science and Artificial Intelligence, CISAI 2021, 2021. doi: 10.1109/CISAI54367.2021.00091.

[26] C. Do Xuan and D. Huong, "A new approach for APT malware detection based on deep graph network for endpoint systems," Appl. Intell., 2022, doi: 10.1007/s10489-021-03138-z.

[27] M. Piskozub, F. De Gaspari, F. Barr-Smith, L. Mancini, and I. Martinovic, "MalPhase: Fine-Grained Malware Detection Using Network Flow Data," in ASIA CCS 2021 - Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security, 2021. doi: 10.1145/3433210.3453101.

[28] R. Murali, P. Thangavel, and C. Shunmuga Velayutham, "Evolving malware variants as antigens for antivirus systems," Expert Syst. Appl., vol. 226, 2023, doi: 10.1016/j.eswa.2023.120092.

[29] Gray, Jason, Daniele Sgandurra, and Lorenzo Cavallaro. "Identifying authorship style in malicious binaries: techniques, challenges & datasets." *arXiv preprint arXiv:2101.06124*

[30] Albtosh, Luay Bahjat. "Malware authorship attribution: Unmasking the culprits behind malicious software." *World Journal of Advanced Research and Reviews* 23, no. 3 (2024)

[31] Xiang, Xiayu, Hao Liu, Liyi Zeng, Huan Zhang, and Zhaoquan Gu. "IPAttributor: Cyber Attacker Attribution with Threat Intelligence-Enriched Intrusion Data." *Mathematics* 12, no. 9 (2024): 1364

Usability Heuristic Evaluation of Mobile Learning Applications Based on the Usability Design Model for Adult Learners

Amy Ling Mei Yin[1], Ahmad Sobri B Hashim[2], Mazeyanti Bt M Ariffin[3]
Department of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology (Perak Branch), Perak, Malaysia[1]
Department of Computer & Information Science, Universiti Teknologi PETRONAS, Perak, Malaysia[2, 3]

***Abstract*—Adult ownership of mobile devices has exploded over the past few years, and smartphones and tablets have become vital for communication, productivity, entertainment, and learning. Some common problems adults face are that they find it difficult to use new technology-based apps because many devices are small. Tasks on new technology-based apps take longer to complete. Therefore, the usability design model for adult learners has been proposed. The objective of this study is to evaluate the usability design model for adult learners and whether the applications containing the model components will affect the satisfaction of adult learners. The evaluation was based on the heuristics guidelines by Nielsen and has been modified and mapped with the seven components in the model. Two existing mobile learning (m-learning) applications (Duolingo and Lingualia) from the Play Store have been chosen for this evaluation. The results indicate that Duolingo has an overall satisfaction mean score of 4.38 compared to Lingualia, where the score is only 2.43. Duolingo meets most of the model's criteria and can score a higher satisfaction mean score. This indicated that the seven components play important roles in contributing to satisfaction among adult learners.**

Keywords*—Usability design model; mobile learning; adult learners; heuristic evaluation

I. Introduction

Mobile learning (m-learning) is quite popular in today's educational world, and many scientists have recognized the potential of mobile technology to enhance learning. The continuous and rapid growth of mobile technology has transformed the traditional education setting into a more accessible, flexible, and personalized learning mode. Such learning benefits adult learners the most because they often face unique challenges, such as time constraints, where they need to balance their work and family commitments.

Moreover, m-learning applications can potentially enhance adult students' learning experience by providing timely, 'just-in-time' resources relevant to their particular needs and lifestyles. In the case of adult learners, the usability of these applications is critical to their design and development. If the applications are not easy to use, unintuitive, or inappropriate to their cognitive and physical needs, it can cause them to struggle to use m-learning. Adults face a number of challenges when using new technologies. People frequently lose various cognitive abilities as they age. Cognitive barriers, such as slower information processing and memory recall, cause difficulties with complex systems and unfamiliar digital interfaces. As a result, using technology or interfaces can be more difficult for older adults, especially when they require adjustments to their perception, movement, or thought processes [1]. Usability Heuristic Evaluation (HE), a recognized method in user-centred design, serves as an effective framework for evaluating and enhancing the user experience of m-learning applications.

The objective of this study is to evaluate the usability design model for adult learners and whether the applications containing the model components will affect the satisfaction of adult learners. The evaluation was based on the heuristics guidelines by Nielsen and has been modified and mapped with the seven components in the model. Two existing m-learning applications (Duolingo and Lingualia) from the Play Store have been chosen for this evaluation.

II. Literature Review

A. Mobile Learning

Note that m-learning describes education using a mobile device. The idea encompasses "gaining knowledge in various environments, through social and content interactions using personal electronic devices" [2]. Furthermore, m-learning lets learners join in informal learning anytime and anywhere. It gives students control over their learning by offering access to different multimedia materials and chances for group learning and research [3]. Other than that, m-learning allows learners to do many tasks at different places and times. It provides educational content to help learners learn outside the classroom and boost their advanced thinking abilities [4].

Abduljawad and Ahmad [5] study the evolution and integration of m-learning into educational practices. Mobile learning faces challenges wherein teachers do not understand how to utilize mobile technology; security and privacy issues might arise because of connectivity restrictions; and the perfection of teachers using mobile devices creates some technical deficiencies. More problems related to the need to detail the development of efficient pedagogical content strategies and create comprehensive user interfaces. Mobile applications are increasing day by day and the software that used to rule earlier is getting outdated which is leading to compatibility issues. In the end, the study depicts m-learning as

an innovative, accessible, and flexible way to advance learning opportunities in the education system.

Olga Viberg & Åke Grönlund [6] study focuses on the design requirements for mobile applications in second language learning within online distance education. It highlights that students frequently use personal mobile technologies for self-initiated learning. Besides that, the study includes the challenges and strategies students face when using technology for language learning in both formal and informal educational settings. Furthermore, it emphasizes the importance of technology in supporting successful language learning processes and the need to provide instructional materials that are compatible with students' technological habits. Overall, the study provides useful insights into the use of mobile technology in language learning, as well as its impact on course design and teaching methodologies.

B. Usability for Adult Learners

Finger gesture interaction with multi-touch surfaces has become more common. Frustrations with using touchscreen technologies are not only reported by older users, but younger groups also find difficulties, though they are typically better at adapting to technological changes [7]. Therefore, movements appropriate for older users may differ significantly from those designed for younger users [8]. Pointing and sliding (scrolling) tasks on the touchscreen, as well as the small size of the buttons, were more difficult for older adults than younger adults [9].

As they age, several people experience decreased ranges and levels of skills like vision, hearing, haptics, cognition, and adeptness, which may negatively impact their use and interaction with user interfaces [10]. Common user interface issues include misinterpretation of general icons, long task completion times, poor task efficiency, errors when reading text due to small font size and poor colour distinction, and confusion with output inputs [11-13]. Although this is usually the case, improving the design to enhance usability for older learners can also improve usability for other user groups [14].

C. Heuristic Evaluation

Nielsen and Molich developed HE, which involves a small group of experts reviewing an interface based on key usability principles. At least three to five evaluators are usually recommended, as this group can identify approximately 65% to 75% of usability issues [15, 16]. Heuristic tests are typically performed during development, but running them on an operational, deployed system can improve how well it detects usability issues [17-19]. According to Nielsen, HE is quick, cost-effective, and simple because it relies on knowledgeable domain experts who efficiently assess usability [19, 20].

D. Duolingo

In November 2011, Cambridge Mellon University's Luis Von Ahn and Severin Hacker released the free language learning app Duolingo. This app provides 68 different language courses in 23 languages, including French, German, Spanish, Dutch, and others, to help learners learn and acquire a foreign or second language. The learning interface is easy to use because it feels and looks like a game. In addition to dictation and written learning, Duolingo provides speaking practice for users who have attained a specific level. Applications for iOS, Android, and Windows Phone are made to improve users' ability to communicate anywhere, at any time. Furthermore, Google Play presented Duolingo with the esteemed Best of the Best 2013 award [21].

E. Lingualia

Lingualia-Learn Languages [22] is an educational app created in 2012 by Javier Sanchez, Roberto Zamora, and Sergio Blanco. Leaner has been able to download the APK since December 2013. Lingualia, a novel and cutting-edge language learning method based on artificial intelligence, adapts to students' interests, motivations, progress, and free time. In addition to all of this information, it can tailor students' learning to meet specific needs. Lingualia allows students at all skill levels (beginner, intermediate, and advanced) to learn Spanish or English. Using cutting-edge technology, educators developed all of the multimedia content. Lingualia contained more than 400 language lessons, 25,000 pronunciation audios, 10,000 vocabulary words, pronunciation guides, grammar, and 100 online language practice exercises.

III. Components in the Usability Design Model for Adult Learners

The usability design model for adult learners includes seven components, as shown in Fig 1: usability layout, navigation, content, a touch gesture, andragogy, and scaffolding.

Fig. 1. Usability design model for adult learners.

A. Usability

The concept of usability has been defined in several ways, and providing teaching and learning services through mobile devices is one of the required aspects. Usability focuses on how computer-related systems communicate with humans [23, 24]. In this study, usability refers to suitable usability for adult learners concerning inability and motor skills.

B. Layout

Application layout refers to the device structure that users see. The consistent page layout will allow users to quickly understand how the application works [23, 25]. Aside from that, all application interfaces must have a simple and consistent

layout that adapts to any smartphone or tablet screen size and is easy for the user to manage.

C. Navigation

Navigation is essential in managing the learner's mood while interacting with the program. Using the navigation feature, learners are able to easily move around the application, making it simpler for them to go from one page to another. Complicated navigation can present challenges for users when using the application. It is not advisable to use complex navigation systems [26]. By including clear navigation in the application, users will have an easier time learning the application process. Hence, having a fixed menu bar on every application interface is important.

D. Content

Learning activities would take place on mobile. The small screen size of mobile devices makes it difficult for users to read lengthy content [27]. Content should be compact, concise, and straightforward. Mobile devices have limited screen space, so a simple design allows learners to access information without scrolling or zooming excessively. Simplifying content can reduce cognitive load and allow learners to focus on learning objectives. Besides that, choose bright and colourful colours to create the content, making the icons appealing.

E. Touch Gesture

Pointing and sliding (scrolling) tasks on the touchscreen, as well as the small size of the buttons, were more difficult for older adults than younger adults. Older adults may find certain gestures, such as pressing and holding, sliding, pinching/spreading, and rotating, too difficult [8, 28, 29]. However, tap and swipe gestures are recommended for older users as they are easy to use and understand [28, 29]. Therefore, it is important to include appropriate touch gestures for adult learners within the application to ensure that they can use it without becoming stressed or demotivated.

F. Theory of Andragogy

In the late 1960s, Malcolm Knowles introduced the term "andragogy" to define his approach to adult education [30]. Adult learning theory is important in designing instructional material to ensure the model suits the adult learner. Andragogy theory has been applied in global studies, resulting in literature translations in multiple languages. This approach to andragogy is effective for individuals of all ages, genders, and nationalities [31].

1) The learner needs to know why they need to learn something: Adults need to understand why they are learning something. They will feel more driven when they know what they should pick up. It can help them grasp the point of the lesson by zeroing in on what they want to learn or how they plan to use it in their lives [32]. Therefore, it is important for the learners and facilitator to communicate with each other. To ensure learners understand why they need to learn something, facilitators must give feedback about how useful the ongoing learning is throughout the course.

2) The learner (self-concept): Adult learners learn new information and enhance their current knowledge more efficiently when motivated to investigate a subject independently. This could be in the form of individual or group projects that require only minimal input from the teacher. The self-concept learner plays a key role in determining performance in a distance learning setting [33]. Hence, collaborative activities that support and encourage adult learners are recommended. Collaborative activities allow adults to showcase their skills and are consistent with the principles of andragogy [34].

3) Adult learning experiences: Every adult has accumulated a lifetime of experiences and may desire to apply their expertise and receive recognition for their knowledge. Examples of learning activities like case studies, group projects, reflective tasks, and short writing exercises can help learners utilize their existing knowledge through sharing and reflection. They are more likely to remember what they have read when they connect it to their own experiences [32, 35]. When learners connect their lives to their reading, they are more likely to remember what they read. As a result, it is important to link assignments or projects to real-life experiences.

4) Readiness to learn: Knowles [30] states that adults need to learn to cope more satisfactorily with real-life tasks or problems. This showed that adults were willing to learn as long as it was connected to their social development. If they realize that certain new learning opportunities can enhance their skills, they are more inclined to push themselves to acquire additional knowledge. Social networking platforms and digital collaboration tools aid adults in integrating this belief into their final projects. Hence, designing tasks that prompt adult students to utilize blogs, wikis, or other social platforms can assist them in not just expanding their social circle but also in working together with individuals who have similar interests. Socializing is essential for connecting users to social networking platforms.

5) Orientation to learn: Adults focus on solving problems when they learn. Learning is based on real-life scenarios or job-related situations rather than specific academic topics. Highlighting how the topic will address issues commonly faced by adult learners in their personal or professional lives through the use of real-life situations is crucial [32]. Using real-life situations related to their life makes it more effective. Media is the most effective way to provide a tangible example or scenario that individuals may encounter in real life.

6) Motivation to learn: Knowles believed that adults were most motivated to achieve their educational goals when acknowledged and appreciated for their contributions to the class [36]. Increased job satisfaction, self-esteem, and quality of life are critical in motivating adults to learn [32]. To motivate adults to learn, the learning environment should encourage active participation. Games, quizzes, and multimedia have all been used effectively to boost learner motivation [35]. Recognizing learners' contributions to the course will boost their self-esteem, motivating them to succeed in their coursework [37]. Therefore, it is important to let them know a reason for every activity, assessment, or e-learning module they

need to complete and acknowledge them after they contribute their effort.

G. Scaffolding

Scaffolding divides complex tasks or competencies into smaller components that will be completed one at a time. This reduces the stress and difficulty of the task while also allowing adult learners to track their progress. Scaffolding is especially important for first-year adult students who may require additional assistance as they transition to an academic setting [38].

IV. Comparison of Mobile Learning Applications Based on the Usability Design Model

Duolingo and Lingualia, two different m-learning apps available on the Play Store, were compared. Both m-learning apps provide language courses based on the m-learning model's seven components: usability, layout, navigation, content, touch gesture, andragogy, and scaffolding for adult learners.

In this study, usability refers to how well the system works for adult learners. When developing m-learning, it is necessary to consider adult learners' motor skills and limitations. It is possible that movements designed for younger users will not work well for older users. The term layout refers to a m-learning interface that is simple and consistent. A simple and consistent layout is easier to control than a complicated layout, which confuses the user. Duolingo provided the user with guidance information that was simple, consistent, and understandable. Despite the fact that a simple and consistent layout is easier for users to control, Lingualia's mobile application is overly simple and lacks information.

Navigation is defined as simple, consistent, and stagnant. Duolingo's navigation design is straightforward, simple, consistent, and stagnant. Each application page has a fixed size, colour, and location. However, in Lingualia, the navigation was designed as a kebab menu (hidden menu), and not all pages included the menu. Aside from that, the Lingualia mobile app provides less navigation information. Duolingo offers simple, accessible, and concise content design. The bright and colourful colours used in the content design make for attractive icons. The information displayed on the screens is simple to view, read, and comprehend, unlike Lingualia, which uses plain colours and simple icons to design its content.

Touch gestures refer to using appropriate touch gestures with an adult learner. The majority of Duolingo and Lingualia's touch gestures are designed to allow users to interact with the mobile application by tapping. It is simple to use and consistent, and the button size is suitable for finger touch for adult learners. However, some of Lingualia's button designs are small compared to Duolingo's.

Andragogy is made up of six assumptions. For the first assumption, the learner needs to know why they need to learn something. Duolingo offers basic information about outline contents, whereas Lingualia does not. The second assumption is that learners are motivated to learn. Adult learners are motivated to learn when there is a clear connection between the information provided and their personal experiences. Using a game that is relevant to their lives and setting a reward for success can help the adult become more motivated. In contrast to Lingualia, Duolingo has set the reward if the user successfully completes the level. Even though Lingualia offers a learning game, the user does not receive any rewards.

The third assumption in andragogy theory is the learner (self-concept), which states that adult learners prioritize their learning and do not rely solely on their teachers. The use of mobile applications should be learner-focused. Both Duolingo and Lingualia were created to break down learning into manageable portions that can be completed in a short time, with logical stopping and starting points. The use of mobile applications is extremely convenient, especially in terms of timing, as adult learners can direct their learning from anywhere and at any time.

The next assumption is adult learning experiences, in which adult learners enjoy sharing their experiences with others. Duolingo provided a forum where users could share their experiences and connect with friends using their Facebook accounts or email addresses. In these terms, Lingualia did not provide a user interaction or sharing platform.

The fifth assumption is readiness to learn. To meet this assumption, a content syllabus relevant to real-world scenarios and an outline of the learning objectives would be useful. Both Duolingo and Lingualia offered activities that were applicable to real-world situations. The final assumption is an orientation to learn. Adults prefer problem-based learning relevant to real-world situations, such as connecting activities or tasks to real-life scenarios. Duolingo and Lingualia included a game inspired by a real-life situation in this section. Duolingo, unlike Lingualia, provides a simple explanation of how the learner will apply the course material in their daily lives.

Scaffolding is an instructional strategy in which the learner receives external support in person or through artifacts to achieve learning goals and tasks within the zone of proximal development until the learner is able to perform the task independently. Duolingo provided clear instructions and effectively conveyed information when using the mobile application, whereas Lingualia provided less instruction.

V. Methodology

This study applied HE based on the usability engineering methodology. The heuristics guidelines were modified and mapped to the seven components. The goal of this study is to evaluate the existing m-learning design using Nielsen's heuristics guidelines to determine whether the m-learning application containing the model's components is able to give satisfaction to adult learners.

A. Experts

In this study, seven experts are involved in evaluating the m-learning application, which is acceptable, as suggested by Nielsen [17], where the minimum number of experts can be three. Table I presents the details of the experts involved in the evaluation process. Fig. 2 illustrates the pie chart of the experts' total percentage of years of service experience.

TABLE I. EXPERT DETAILS INVOLVED IN THE EVALUATION

Experts	*Position*	*Expertise*	*Year of Experience*
Expert 1	Mobile App Developer	Mobile Application, Information Technology, Human-Computer Interactions	Less than 5
Expert 2	Lecturer	Mobile Learning / App; Information Technology	6 – 10 years
Expert 3	Lecturer	Information Technology	More than 11 years
Expert 4	Lecturer	Information Technology	6 – 10 years
Expert 5	Lecturer	Information Technology	More than 11 years
Expert 6	Lecturer	Mobile Learning / App; Information Technology; E-Learning & Education	More than 11 years
Expert 7	Lecturer	Mobile Learning / App; Data Science and Machine Learning	6 – 10 years

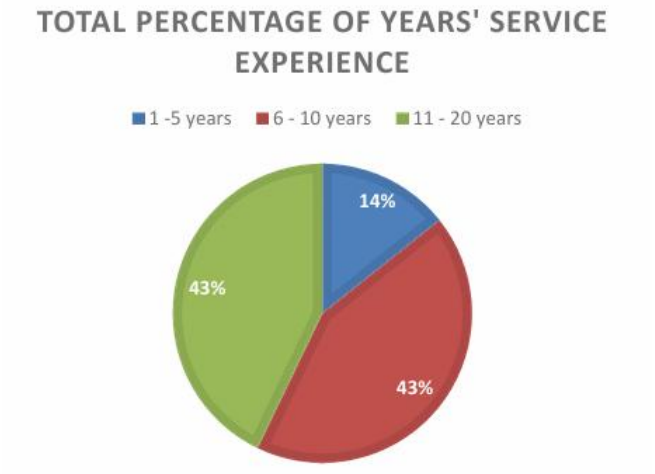


Fig. 2. Total percentage of years of service experience.

B. Questionnaire

The questionnaire for evaluating the m-learning application was based on HE. Nielsen's heuristics guidelines were chosen, modified, and mapped to the seven components based on how appropriate they are for this study to ensure that the m-learning application contains the model's components. Table II presents the heuristic rules mapped to the model's seven components. Questionnaires were designed to elicit responses using a five-point Likert Scale: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree.

TABLE II. HEURISTIC RULES MAPPED WITH THE COMPONENTS IN THE MODEL

No	*Heuristic Rules*	*Components*
1.	User control and freedom	Usability, Navigation, Content, Touch Gesture
2.	Error prevention	Scaffolding
3.	Consistency and standards	Layout, Navigation, Touch Gesture
4.	Aesthetic and minimal design	Layout, Navigation, Content, Touch Gesture
5.	Suitability	Usability, Layout, Navigation, content, touch gesture
6.	Help and Documentation	Andragogy
7.	User Satisfaction	

C. Evaluation Process

The first step in the evaluation process was to identify the number of the expert, which was then reached via email or phone call. An appointment was made with those who agreed to participate in the evaluation process. The experts entered their background information and began to evaluate the mobile application using the provided questionnaire and description of how the procedure was carried out. After the expert had completed the evaluation process, the researcher analyzed the findings and arrived at the results and conclusion.

VI. RESULT AND DISCUSSION

Duolingo has a mean score of 4.46 for user control and freedom, compared to Lingualia's 2.91. Duolingo's design enables adult learners to manage the application easily. Duolingo outperforms Lingualia in error prevention with clear and easy-to-understand instructions, where the mean score is 4.45 for Duolingo and 2.39 for Lingualia. Table III lists the result of the mean score for Duolingo and Lingualia.

TABLE III. RESULT OF THE MEAN SCORE FOR DUOLINGO AND LINGUALIA

No	*Heuristic Rules*	*Duolingo*	*Lingualia*
1.	User control and freedom	4.46	2.91
2.	Error prevention	4.45	2.39
3.	Consistency and standards	4.29	3.00
4.	Aesthetic and minimal design	4.48	2.45
5.	Suitability	4.20	2.43
6.	Help and Documentation	4.27	2.16
7.	User Satisfaction	4.38	2.43

Consistency and adherence to standards in application development help create user-friendly experiences, reducing confusion and improving the ease of learning. Duolingo's design is consistent and standard, and the bar chart portrays that Duolingo's mean score is higher than Lingualia's, where the mean score is 4.29, while Lingualia's score is 2.90. Aesthetics play an important role in the perception and memory of application designs.

In addition, for minimal, it maximizes utility and usability by allowing users to find what they need. In other words, interfaces should not include irrelevant or rarely used information. Duolingo scored 4.48 on aesthetic and minimal design, while Lingualia scored 3.00. Duolingo is better suited for adult learners, with a score of 4.2 versus Lingualia's 2.43. Duolingo earns a 4.27 rating for its comprehensive support and documentation resources. Lingualia has fewer guidelines and a lower score, with a mean score of 2.16. Fig. 3 shows that

Duolingo has a higher overall satisfaction mean score of 4.38 than Lingualia, which is 2.43. Duolingo meets most of the model's criteria and offers higher user satisfaction. According to the study's findings, it shows that the user satisfaction is higher with m-learning applications that include a greater number of these model components. Applications with fewer components, on the other hand, may fall short of meeting the diverse needs of adult learners. The findings also highlight the significance of including comprehensive model components during the design phase of m-learning applications.

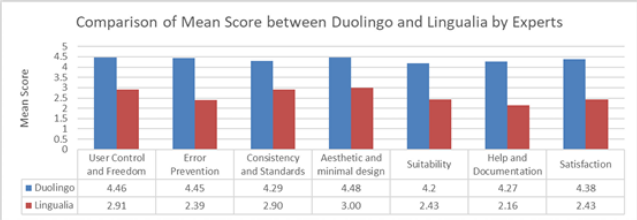


Fig. 3. Comparison of mean score between duolingo and lingualia.

VII. Conclusion

This study compares two different m-learnings available in the Play Store, Duolingo and Lingulia. The purpose of this study is to evaluate the existing m-learning design based on Nielsen's heuristics guidelines. The heuristics guidelines have been modified and mapped with the seven components in the model to validate whether the m-learning application containing the model's components can give adult learners satisfaction. Moreover, the results indicate that the m-learning application containing the component in the model is able to score higher satisfaction to adult learners compared to m-learning applications that contain fewer model components. This study will contribute to the growing body of work on m-learning by offering practical insights for design and development teams and educators who are trying to optimize usability with adult learners.

Acknowledgment

We want to thank Tunku Abdul Rahman University of Management & Technology and the experts for their willingness to be involved in this study. This research was funded by Tunku Abdul Rahman University of Management & Technology (TAR UMT).

References

[1] N. A. Ahmad, M. F. Abd Rauf, N. N. Mohd Zaid, A. Zainal, T. S. Tengku Shahdan, and F. H. Abdul Razak, "Effectiveness of instructional strategies designed for older adults in learning digital technologies: a systematic literature review," SN computer science, vol. 3, p. 130, 2022.

[2] H. Crompton and D. Burke, "The use of mobile learning in higher education: A systematic review," Computers & education, vol. 123, pp. 53-64, 2018.

[3] B. Jin, J. Kim, and L. M. Baumgartner, "Informal learning of older adults in using mobile devices: A review of the literature," Adult Education Quarterly, vol. 69, pp. 120-141, 2019.

[4] M. Z. Asghar, E. Barberà, and I. Younas, "Mobile Learning Technology Readiness and Acceptance among Pre-Service Teachers in Pakistan during the COVID-19 Pandemic," Knowledge Management & E-Learning, vol. 13, pp. 83-101, 2021.

[5] M. Abduljawad and A. Ahmad, "An analysis of mobile learning (M-Learning) in education," Multicultural Education, vol. 9, p. 2023, 2023.

[6] O. Viberg and Å. Grönlund, "Understanding students' learning practices: challenges for design and integration of mobile technology into distance education," Learning, Media and Technology, vol. 42, pp. 357-377, 2017.

[7] T. Page, "Touchscreen Mobile Devices and Older Adults: A Usability Study," International Journal of Human Factors and Ergonomics, vol. 3, pp. 65 - 85, 2014.

[8] C. Stößel and L. Blessing, "Mobile Device Interaction Gestures for Older Users," Proceedings: NordiCHI 2010, October 16–20, 2010, 2010.

[9] W. A. Rogers, A. D. Fisk, A. C. McLaughlin, and R. Pak, "Touch a screen or turn a knob: Choosing the best device for the job," Human factors, vol. 47, pp. 271-288, 2005.

[10] L. Ruzic and J. A. Sanfod, "Universal Design Mobile Interface Guidelines (UDMIG) for an Aging Population," Springer International Publishing AG, 2017.

[11] S. A. Becker, "A study of web usability for older adults seeking online health resources," ACM Trans Comput-Hum Interact (TOCHI), pp. 387–406, 2004.

[12] B.B. Bederson, B.Lee , R.M. Sherman , P.S.Herrnson, and R.G.Niemi, "Electronic voting system usability issues. ," Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp. 145-152 2003.

[13] A. Chadwick-Dias, M. Mcnulty , and T. Tullis "Web usability and age: how design changes can improve performance.," ACM SIGCAPH Computers and the Physically Handicapped. ACM,, pp. 30–37, 2003.

[14] S. J. Czaja, N. Charness, A. D. Fisk, C. Hertzog, S. N. Nair, W. A. Rogers, et al., " Factors Predicting the Use of Technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE). ," Psychology and Aging, , vol. Vol. 21, No.2, (June 2006),, pp. pp. 333-352. , 2006.

[15] E. L.-C. L. E. T. Hvannberg, and M. K. Lárusdóttir, "Heuristic evaluation: Comparing ways of finding and reporting usability problems," Interact. Comput, vol. vol. 19, no. 2,, pp. 225–240, 2007.

[16] J. Nielsen, "Heuristic Evaluations. ," In: J. Nielsen and R.L. Mack. (Eds), Usability Inspection Methods. New York: John Wiley & Sons, 1994.

[17] J. Nielsen, "Finding usability problems through heuristic evaluation," Proc. ACM, pp. 373-380, 1992.

[18] L. K. Peng, Ramaiach, C.K, & Foo, S., "Heuristic-Based User Interface Evaluation at Nanyang Technological University in Singapore.," Program: Electronic Library and Information Systems, vol. 38 (1): 42-59, 2004.

[19] S. Ssemugabi and R. d. Villiers, "A comparative study of two usability evaluation methods using a web-based e-learning application," presented at the Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries, Port Elizabeth, South Africa, 2007.

[20] J. Preece, Rogers, Y. and Sharp, H. 2, "Interaction Design: Beyond Human-Computer Interaction. ," 2rd Ed. New York: John Wiley & Sons. , 2007.

[21] A. Irawan, A. Wilson, and S. Sutrisno, "The implementation of duolingo mobile application in English vocabulary learning," Scope: Journal of English Language Teaching, vol. 5, pp. 08-14, 2020.

[22] Lingualia - Learn languages. Available: https://www.appbrain.com/app/lingualia-learn-languages/com.lingualia.app

[23] J. Nielsen, " Usability Engineering. ," Boston: Academic Press. , 1993.

[24] H. Rex Hartson, "Human–Computer Interaction: Interdisciplinary Roots and Trends. ," Journal of Systems and Software, pp. 103–118, 1998.

[25] M. Uflacker, and Busse, D., "Complexsity in Enterprise Applications vs. Simplicity in User Experience.," 12th International Conference on Human Computer Interaction, 2007.

[26] k. Donnelly, and Walsh, s, "Mobile Learning Reviewed," 2009.

[27] D. S. K. Seong, "Usability Guidelines for Designing Mobile Learning Portals," in Proceedings of the 3rd International Conference on Mobile Technology, Applications & Systems - Mobility 2006.

[28] R. A. Leitão, "Creating mobile gesture-based interaction design patterns for older adults: A study of tap and swipe gestures with Portuguese seniors," 2012.

[29] T.-H. Tsai, K. C. Tseng, and Y.-S. Chang, "Testing the usability of smartphone surface gestures on different sizes of smartphones by different age groups of users," Computers in Human Behavior, vol. 75, pp. 103-116, 2017.

[30] M. S. Knowles, "From pedagogy to andragogy," Religious Education, 1980.

[31] A. S. Alajlan, "Appling Andragogy Theory in Photoshop Training Programs," Journal of Education and Practice, vol. 6, pp. 150-154, 2015.

[32] D. Fidishun, "Andragogy and technology: Integrating adult learning theory as we teach with technology," Proceedings of the 2000 Mid-South Instructional Technology Conference. , 2000., 2000.

[33] N. Dabbagh, "The online learner: Characteristics and pedagogical implications," Contemporary Issues in Technology and Teacher Education, vol. 7, pp. 217-226, 2007.

[34] C. Cochran and S. Brown, "Andragogy and the adult learner. Supporting the Success of Adult and Online Students," 2016.

[35] S. R. Aragon, Facilitating Learning in Online Environments: New Directions for Adult and Continuing Education, Number 100 vol. 103: John Wiley & Sons, 2010.

[36] M. S. Knowles, "Theory of andragogy," ed: A Critique. International Journal of Lifelong, 1984.

[37] L. C. Blondy, "Evaluation and application of andragogical assumptions to the adult online learning environment," Journal of interactive online learning, vol. 6, pp. 116-130, 2007.

[38] R. C. McCall, Kristy Padron, Carl Andrew, "Evidence-Based Instructional Strategies for Adult Learners: A Review of the Literature," Journal of the Louisiana Chapter of the ACRL, vol. 4, 2018.

Radar Spectrum Analysis and Machine Learning-Based Classification for Identity-Based Unmanned Aerial Vehicles Detection and Authentication

Aminu Abdulkadir Mahmoud[1], Sofia Najwa Ramli[2], Mohd Aifaa Mohd Ariff[3], Muktar Danlami[4]
Centre of Information Security Research-Faculty of Computer Science and Information Technology,
Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia[1, 2]
Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Johor, Malaysia[3]
Etienne Innovation Sdn. Bhd., Cyberjaya, Selangor, Malaysia[2, 3]
Department of Cyber Security, Yusuf Maitama Sule University, Kano, Nigeria[1, 4]

Abstract*—The significant use of Unmanned Aerial Vehicles (UAVs) in commercial and civilian applications presents various cybersecurity challenges, particularly in detection and authentication. Unauthorized UAVs can be very harmful to the people on the ground, the infrastructure, the right to privacy, and other UAVs. Moreover, using the internet for UAV communication may expose authorized ones to attacks, causing a loss of confidentiality, integrity, and information availability. This paper introduces radar-based UAV detection and authentication using Micro-Doppler (MD) signal analysis. The study provides a unique dataset comprising radar signals from three distinct UAV models captured under varying operational conditions. The dataset enables the analysis of specific features and classification through machine learning models, including k-nearest Neighbor (k-NN), Random Forest, and Support Vector Machine (SVM). The approach leverages radar signal processing to extract MD signatures for accurate UAV identification, enhancing detection and authentication processes. The result indicates that Random Forest achieved the highest accuracy of 100%, with high classification accuracy and zero false alarms, demonstrating its suitability for real-time monitoring. This also highlights the potential of radar-based MD analysis for UAV detection, and it establishes a foundational approach for developing robust UAV monitoring systems, with potential applications in aviation military surveillance, public safety, and regulatory compliance. Future work will focus on expanding the dataset and integrating Remote Identification (RID) policy. A policy that mandates UAVs to disclose their identity upon approaching any territory, this will help to enhance security and scalability of the system.

Keywords*—Authentication; detection; cybersecurity; Micro-Doppler; radar; Unmanned Aerial Vehicle (UAV)

I. Introduction

Unmanned Aerial Vehicles (UAVs) are rapidly gaining momentum across various sectors where information technology is crucial in enhancing efficiency and operations [1]. The popularity of UAVs is due to their cost-effectiveness, ease of use, and potential to streamline business processes [1]. UAVs are being deployed in various industries, including transportation, agriculture, surveillance, and defense, offering unprecedented flexibility and operational efficiency. However, this widespread adoption is not without challenges, as UAVs increasingly face cybersecurity threats. These threats pose risks to the integrity of UAV operations and the safety of the data they collect and transmit [2].

Cybersecurity is critical for the safe operation of UAVs. A breach in security can endanger ground personnel, infrastructure, personal privacy, and the UAVs themselves. Since UAVs rely heavily on internet connectivity, they are vulnerable to threats compromising confidentiality, integrity, and availability [3]. Confidentiality may be violated through data interception, physical tampering, eavesdropping, or social engineering attacks. Integrity is at risk from data tampering and hacking, while availability can be disrupted by interference, jamming, denial-of-service attacks, or natural events [3].

UAV detection is identifying and tracking UAVs using various technologies, such as radar, cameras, or acoustic sensors among others. It plays a crucial role in ensuring the safety and security of airspace by monitoring unauthorized and potentially harmful UAV activity. The detection and classification of UAVs can be challenging, and various technological approaches offer unique advantages and limitations [4].

1) Radar detection: This approach works by emitting radio waves (signals) that reflect off UAVs, allowing it to gather detailed information about the UAV's distance, speed, and movement patterns [4]. Radar's ability to operate in all weather conditions and low-visibility environments, such as during rain, fog, or nighttime, makes it highly reliable [8]. It is particularly effective for long-range detection and can differentiate UAVs from other objects by analyzing their motion and size. This robustness and capability to detect small UAVs over large distances make radar an essential tool in security-critical environments [5].

2) Video detection: It relies on cameras and image processing algorithms to visually identify UAVs. It excels in clear visibility conditions, offering high-resolution images for accurate UAV identification and classification [4]. Video detection can work in real time, leveraging machine learning to enhance accuracy. However, its effectiveness diminishes in poor lighting conditions, such as nighttime or foggy environments, and it struggles to detect small UAVs at long distances [5]. Consequently, while video detection is valuable

in specific contexts, its dependence on favorable visibility limits its reliability.

3) Acoustic detection: It captures the unique sound signatures produced by UAVs, especially the noise from their propellers and motors [4]. The method uses microphones to monitor sound and can detect UAVs even when they are not visible. Acoustic detection is a cost-effective and straightforward solution for short-range detection, in quiet environments [5]. However, its range is limited, and it faces significant challenges in noisy environments, such as urban areas. In addition, it struggles to detect noiseless UAV models, making it less effective for comprehensive UAV detection over larger areas.

4) RF-Based detection: It monitors the communication signals transmitted between UAVs and their controllers [4]. This method is highly effective for detecting UAVs as soon as they begin transmitting signals, and it can provide valuable information about both the UAV and its operator [5]. However, RF-based detection is ineffective against fully autonomous UAVs that do not rely on RF signals or operate with encrypted communications. Moreover, its range is limited to the transmission distance of the UAV's signals, and it can be vulnerable to signal interference or jamming, reducing its reliability [5].

Among these methods, radar-based UAV detection is effective as it captures unique MD signatures generated by a UAV's rotating blades and motion. The radar signals analysis can detect, classify, and track UAVs in low visibility conditions, providing valuable information for security, surveillance, and air traffic management [5], [6]. Radar detection is preferred for its versatility and reliability. Unlike video, acoustic, or RF-based detection, radar is unaffected by visibility, background noise, or signal dependence. Its ability to detect UAVs over long distances and in all weather, conditions make it the most robust solution for UAV detection, in security and defense applications where continuous and reliable monitoring is essential [5], [6], [7]. Despites the radar's effectiveness, the existing systems struggle to identify specific UAV types using MD features or radar cross-section (RCS) characteristics. Furthermore, reliance on proprietary datasets limits the generalizability of findings and hinders benchmarking efforts. This shows the need for standard, and public available datasets to improve radar-based detection and authentication methods.

In light of this, this research focuses on developing an identity-based UAV detection and authentication system that leverages information on radar signal analysis. The system uses radar to analyze the UAV's signal through the MD effect, which reveals unique features of the UAV's rotor movements and structure. These unique MD radar features will help to authenticate the UAV's identity, ensuring that it matches the information and parameters of the known UAVs. Then, the proposed system employs the k-NN, Random Forest, and SVM to classify and authenticate UAVs based on the identified patterns in radar signal data. Furthermore, this research generates raw datasets for three distinct UAV models: DJI Matrice 600, DJI Matrice 300, and Phantom 4. The study offers an identity-based UAV detection, authentication, and classification model. The model handles detection and authentication in separate phases, while the combined model integrates both processes into a unified approach. The proposed model is compared with others to ascertain its efficiency and performance. This comparison underscores the advantages and limitations of each framework, demonstrating their applicability in real-world scenarios.

The research makes stride contributions to the field of UAV detection, classification, and authentication. These contributions include a novel radar-based dataset, generated through simulations of three UAV models (DJI Matrice 600, DJI Matrice 300, and Phantom 4), that can provide a foundation for advanced research in this domain. The study also develops a robust classification system employing machine learning algorithms (kNN, Random Forest, and SVM). The proposed model effectively detects and classifies UAVs and enhances authentication by comparing radar signal data with predefined UAV parameters. Furthermore, the performance of the proposed models is evaluated on the radar dataset, demonstrating improved detection accuracy and reliable classification and authentication.

B. Our Contributions

This research offers the following contributions:

1) A novel radar-based dataset is created through simulation using three UAV models (DJI Matrice 600, DJI Matrice 300, and Phantom 4). This dataset will be instrumental for future UAV detection, authentication, and classification research.

2) A classification system is developed using kNN, Random Forest, and SVM for UAV detection and classification. The framework compares radar signal data with the known UAV information and parameters for enhanced authentication.

3) The detection, classification, and authentication performances of the proposed frameworks are evaluated on the generated radar dataset. Detailed comparisons show that the proposed framework improves detection accuracy and achieves reliable classification and authentication.

II. Related Work

This section reviews the approaches and methodologies in UAV detection, emphasizing radar-based systems and machine-learning models for classification. A radar-based UAV detection method is proposed using the Empirical Mode Decomposition (EMD) algorithm to extract MD signals for identifying small UAVs [11]. The technique offers the advantage of isolating m-D features crucial for differentiating UAVs from other moving objects, decomposes signals into intrinsic mode functions (IMFs), and addresses mode-mixing challenges by analyzing the extrema distribution. The EMD algorithm distinguishes UAVs from birds or other objects based on their rotor blade signatures, making it highly effective even in noisy environments. In addition, the paper points that while the EMD algorithm effectively processes non-stationary signals, it also has limitations, such as susceptibility to noise and increased computational load, which may hinder its real-time application in practical UAV detection scenarios.

A. Radar Approaches, and Machine Learning Models UAV Classification

Ezuma et al. [9] presents a multistage system for detecting and classifying UAV controllers using radio frequency (RF) fingerprints in the 2.4 GHz band, even in environments with significant interference from Wi-Fi and Bluetooth devices. The system addresses the challenge of detecting UAV controllers in the presence of Wi-Fi and Bluetooth interference by first detecting RF signals using a Markov model-based Naïve Bayes algorithm, followed by interference detection and machine learning (ML) classification. The system extracts 15 statistical features from the UAV controller signals and achieves a classification accuracy of 98.13% using k-Nearest Neighbors (kNN) at 25 dB signal-to-noise ratio (SNR). Despite the effectiveness of the system, its limitation lies in distinguishing identical UAV controllers, such as pairs of DJI models, which slightly lowers accuracy due to signal similarity.

Meanwhile, a system proposes using Frequency Modulated Continuous Wave (FMCW) radar to detect and identify UAVs by analyzing their MD signatures [10]. It addresses the challenge of extracting MD signatures caused by the rapid rotation of UAV rotor blades, which introduces high-frequency variations in radar signals. The system proposes a new approach for studying MD signatures in UAVs and presents both simulation and experimental results to demonstrate the effectiveness of this method. The analysis showcases the ability of FMCW radar to capture fine-grained UAV motion details, allowing for more accurate UAV detection and identification. This work improves UAV detection in target-dense environments and shows the advantages of MD signature analysis for characterizing UAVs in real-time scenarios.

Similarly, a micro-motion model for detecting and identifying low-slow-small UAVs using radar systems is proposed [11]. The research focuses on the MD effects generated by the rotating blades of UAVs. It suggests a method to enhance detection performance by compensating for translational movement and employing an optimal demodulation operator for parameter estimation. The process improves the signal-to-noise ratio (SNR) and suppresses clutter, making it practical for detecting small UAVs even under challenging low-SNR conditions. The simulation and experimental results demonstrate the accuracy of the proposed technique in estimating MD parameters, which significantly aids in classifying UAVs based on their unique motion characteristics. This work is relevant for air defense systems identifying small, slow-flying UAVs based on radar signatures.

Besides, a study establishes a theoretical foundation linking the MD signatures and motion dynamics of small UAVs, focusing on analyzing the Doppler spectrum as a more efficient tool than joint time-frequency (JTF) images [12]. It explores how MD features, such as blade length, rotor rotation rate, and radial velocity, can be derived from the Doppler spectrum, aiding in detecting and classifying small UAVs. The study demonstrates the correlation between the spectral distribution and UAV physical specifications through simulation and measured data. Compared to JTF images, the Doppler spectrum provides significant computational and storage benefits while delivering accurate MD signatures. However, the study acknowledges challenges in detecting chopping frequencies and resolving smearing effects caused by multiple rotors, especially in practical scenarios with complex UAV dynamics. Future work will focus on refining algorithms to address these issues for more reliable UAV detection and classification.

In another development, a study explores the use of machine learning for drone classification based on radar signals [13]. It details the creation of datasets through simulation, considering radar specifications and SNR ranging from 0 to 20 dB. Each dataset, with 1000 spectrogram samples per class and smaller validation sets, was used to train a Convolutional Neural Network (CNN). The CNN architecture, comprising convolutional layers, SoftPlus activation, instance normalization, dropout, and linear layers, was adapted for different radar pulse repetition frequencies (PRFs). The performance of the model was evaluated using the macro-F1 score, with results showing the X-band 2 kHz PRF radar outperforming the W-band radar, especially at lower SNRs. The study revealed that while the X-band radar achieved an F1 score of 0.816±0.011, it struggled with false alarms, particularly with noise being confused with the DJI Matrice 300 RTK drone. The model exhibited robustness to varying blade pitch and SNR values, maintaining performance with different pitch values and showing reduced effectiveness at lower SNRs. Future work aims to investigate the impact of various wavelengths, explore more complex CNN architectures, and validate models with real-world data, addressing the current model's limitations and extending its applicability.

Furthermore, a novel lightweight architecture presents the development of an MD-based detection method called DIAT-RadSATNet, a deep CNN (DCNN) designed for the detection and classification of Small UAVs (SUAVs) using MD signatures [14]. The method addresses the growing need for efficient SUAV detection in both defense and civilian applications. The proposed architecture is lightweight, with 40 layers and only 0.45 million trainable parameters, achieving a high classification accuracy of 97.3%. The study indicates the radar system's ability to classify various SUAVs, such as quadcopters and bionic birds, through field experiments and ablation studies, demonstrating superior performance compared to existing models. The DIAT-RadSATNet's lightweight design allows for real-time implementation with reduced computational costs. However, the study acknowledges limitations, regarding the potential challenges in detecting SUAVs under complex environmental conditions and the need for further testing in diverse operational scenarios to validate the system's robustness and scalability.

Rojhani et al. [15] introduced a novel deterministic data augmentation method for UAV classification based on MD radar signatures. The technique generates synthetic training datasets using a physical radar backscattering model, reducing reliance on extensive measurement campaigns. Compared to conventional random signal processing augmentation, the deterministic approach produces datasets that maintain the physical integrity of features, resulting in better generalization and reducing classification bias. The study focuses on classifying UAVs based on their number of motors using CNN, achieving an accuracy of 78.68% and outperforming conventional augmentation methods, which resulted in 66.18%

accuracy with significant class bias. The results suggest that the deterministic augmenter provides more reliable and effective training datasets, for radar-based UAV classification, and can be extended to other scenarios, such as human recognition and medical imaging. The research indicates the potential for scaling this method to produce diverse datasets without costly and time-consuming measurement campaigns.

Further, a novel approach for UAV classification using radar digital twins is presented by generating full-wave electromagnetic simulations [16]. A Multiple-Input Multiple-Output (MIMO) radar system is simulated using CAD models of various UAVs to create radar datasets that include Range-Doppler and MD information. The datasets train a machine learning classifier, a one-versus-rest Support Vector Machine (SVM), for UAV detection and classification. The simulations allow for generating radar datasets without the need for expensive, time-consuming measurement campaigns. The study demonstrates high classification accuracy (minimum 97%) in multi-UAV scenarios, showing that the digital twin framework offers a flexible and cost-effective solution for UAV detection and classification in various operational conditions.

Moreover, a review of radar-based drone detection, tracking, and classification techniques focusing on real-world data from 25 drone trials using the Gamekeeper radar and over 55,000 trajectories and diverse drone types like the DJI Phantom 2 and DJI Inspire 2 is conducted [17]. The challenges established include differentiating drones from non-drones and managing varying SNRs. The performance metrics such as accuracy, F1 score, true positive confidence, false alarm rate, and classification time delay are discussed, with the false alarm rate remaining a significant hurdle. The study explores advancements in distributed and multi-static radar systems, quantum oscillators, advanced antennas, ML, and AI, emphasizing their role in improving automatic target classification (ATC). It suggests future directions, including leveraging cognitive radar systems, digital twins for rapid algorithm development, and integrating contextual and meta-level information to enhance performance. The study concludes that while radar systems have made significant progress, challenges remain, in complex environments, and suggests a multisensor approach for more robust detection and classification of drones.

Furthermore, a study proposes a fully convolutional network (FCN)-based approach for fast detection of UAVs using pulse Doppler radar. The traditional constant false alarm rate (CFAR) methods, effective for uniform backgrounds, struggle with low-small UAVs [18]. The proposed FCN operates on the entire range-Doppler map to enhance detection speed while maintaining high accuracy. The network leverages a bifurcated classification and regression architecture, reduces computational overhead, and integrates a post-processing mechanism for precise target location. The experimental results show the FCN-based method improves detection speed by up to 47 times compared to previous methods while ensuring high detection accuracy and reduced false alarms. Despite its achievements, it has difficulties detecting multiple UAVs in the same grid cell due to resolution constraints, with plans to address this via sampling in future research.

In another research, a study investigates the use of MATLAB simulations for analyzing MD signatures of rotating propeller blades and flapping wings using an S-band continuous-wave (CW) radar system [6]. The system demonstrates the effectiveness of the Short Time Fourier Transform (STFT) and Fast Fourier Transform (FFT) in distinguishing between the unique signatures of micro- UAVs and birds. It employed a 100 kHz sampling frequency and 700ms integration time, revealing significant Doppler shifts and frequency dispersion associated with different propeller blade lengths and flapping frequencies. The results show STFT's capability to provide detailed time-frequency analysis, contributing to improved detection and characterization of small UAVs and avian targets. This work supports the development of advanced radar systems for enhanced detection and tracking, aligning with the goals of optimizing UAV identification and authentication in the context of radar signal analysis.

B. Research Gap

Significant advancements in UAV detection, classification, and authentication are attained across various methods, as observed in literature and as depicted in Table I. Despite this strides, significant gaps remain, in datasets and radar-based detection approaches. The reviewed research often relies on proprietary datasets, limiting generalizability, which indicates the need for publicly available, standardized UAV/drone datasets to benchmark the systems. In addition, radar systems are adequate for general UAV detection but lack robust methods for identifying specific UAV types based on MD or RCS features. Further innovation is needed to improve radar detection capability and address environmental challenges in urban areas, such as noise and clutter. To address these gaps, our research aims to develop a novel raw dataset for UAV identification and authentication, leveraging unique features and employing three models, kNN, Random Forest, and SVM, to classify UAVs based on their MD signatures. This proposed approach will enhance the accuracy of radar-based systems and utilize MD signatures for identity-based detection and authentication of UAVs.

TABLE I. COMPARATIVE ANALYSIS OF RELATED WORKS

Study	***Methodology***	***Main Features***	***Accuracy/Performance***	***Limitations/Challenges***	***Future Work***
(Zhao & Su, 2020)	EMD algorithm for MD extraction	Isolation of MD features of UAVs	Effective in noisy environments, good UAV differentiation	Computational load affects real-time performance	Optimizing real-time application and noise handling
(Ezuma et al., 2020)	RF fingerprints for UAV controller detection	Detects UAV controllers in noisy RF environments	Classification accuracy of 98.13% using kNN at 25 dB SNR	Struggles with identical controllers, limited in low SNR	Sensor fusion for improved UAV detection

Study	*Methodology*	*Main Features*	*Accuracy/Performance*	*Limitations/Challenges*	*Future Work*
(Reddy & Peter, 2021)	FMCW radar and MD signatures	Analyzes MD signatures caused by rapid rotor blade rotation	Effective in capturing fine-grained UAV motion details; improve detection in dense environments	Challenges in extracting MD signatures and dealing with high-frequency variations	Improve techniques for better MD signature extraction and analysis
(Ji et al., 2021)	Micro-motion model and radar systems	Enhances detection performance for small UAVs; improved signal-to-noise ratio (SNR) and optimal demodulation	Accurate in estimating MD parameters; effective in low-SNR conditions	Detection challenges due to clutter and small UAV dynamics	Refine algorithms for better performance in cluttered environments
(Kang et al., 2021)	MD signatures and Doppler spectrum	Links MD signatures to UAV motion dynamics; efficient compared to joint time-frequency (JTF) images	Accurate with significant computational and storage benefits; good correlation with UAV specs	Detection challenges due to chopping frequencies and smearing effects from multiple rotors	Refine algorithms to address detection challenges with complex UAV dynamics
(Raval et al., 2021)	Machine learning with radar signals and CNNs	Creates datasets with different radar specifications; evaluates X-band vs. W-band radar performance	X-band radar achieved F1 score of 0.816±0.011; struggles with false alarms at lower SNRs	False alarms due to noise confusion; reduced effectiveness at lower SNRs	Investigate impact of different wavelengths and complex CNN architectures; validate with real-world data
(Kumawat et al., 2022)	MD-based detection with DIAT-RadSATNet	Lightweight DCNN architecture with 40 layers and 0.45 million parameters; high classification accuracy	Achieves 97.3% classification accuracy; real-time implementation with reduced computational costs	Challenges in detecting SUAVs under complex conditions; needs further testing in diverse scenarios	Validate system's robustness and scalability in various operational environments
(Rojhani et al., 2023)	Deterministic data augmentation for UAV classification using MD radar	Generates synthetic datasets using radar backscattering; focuses on UAVs classified by motor count with CNN	Achieved 78.68% accuracy; outperformed conventional augmentation methods (66.18%)	Class bias with conventional methods; potential for scaling to other domains like human recognition and medical imaging	Extend method to diverse scenarios and other applications such as human recognition and medical imaging
(Sayed et al., 2023)	Radar digital twins for UAV classification using electromagnetic simulations	Simulates MIMO radar with CAD models to generate datasets; uses SVM for classification	Achieves minimum 97% classification accuracy in multi-UAV scenarios	Cost-effective but requires validation in real operational conditions; limited to simulated scenarios	Validate system in real-world scenarios and diverse conditions
(Ahmad et al., 2024)	Radar-Based Detection and Tracking	Reviews radar systems and performance metrics; discusses advancements in radar technology.	Metrics discussed include accuracy, F1 score, and false alarm rate; specifics not provided.	False alarm rates and differentiation challenges; complex environments.	Multisensor approaches and cognitive radar for improved performance.
(Tian et al., 2024)	Fully Convolutional Network (FCN) for Fast Detection	Uses FCN for enhanced speed and accuracy in detecting UAVs; bifurcated architecture for classification.	47 times faster detection speed; high accuracy and reduced false alarms.	Difficulty in detecting multiple UAVs in the same grid cell.	Up-sampling techniques to address resolution constraints.
(Zulkarnain et al., 2024)	MATLAB Simulations of MD Signatures	Analyzes MD signatures using STFT and FFT to distinguish UAVs from birds.	Effective in distinguishing UAVs from birds; detailed time-frequency analysis.	Limited by radar system capabilities and integration time.	Advanced radar systems for improved detection and tracking.
Proposed Model	Micro-Doppler Signature and Doppler Spectrum	Uses Radar system for detection, kNN, Random Forest, and SVM for classification.	Achieves 100% accuracy with Random Forest.	kNN, and SVM are struggling in distinguishing some types of UAVs due to their similarities.	To integrate RID Policy in UAVs detection and authentication.

III. SIMULATION SETUP

This research presents a novel approach to generating a raw dataset of UAVs, focusing on three distinct types: the DJI Matrice 600, DJI Matrice 300, and DJI Phantom 4. The dataset is designed to capture radar signals reflective of these UAVs' unique characteristics, facilitating a detailed analysis of their operational signatures. The radar system is configured to simulate detections of the UAVs at distances of up to 1 kilometer in both hovering and motion operations. This involves UAVs located 1 kilometer towards the radar and 1 kilometer away from it, with all the points between the intervals of 10 meters inclusive. The simulation considers various UAV operations, such as stationary hovering and different flying speeds, providing a robust dataset that reflects real-world operational conditions. The generated data, which encompasses detailed radar reflections and MD signatures unique to each UAV type, is used to train machine learning models utilizing kNN, Random Forest, and SVM algorithms. These models are employed to classify and detect UAVs based on their distinct radar signatures, enhancing the system's capability to differentiate between the DJI Matrice 600, DJI Matrice 300, and DJI Phantom 4 across various operational modes. This research contributes to advancing more effective and precise UAV detection and classification systems.

The flowchart in Fig. 1 illustrates the detection and classification processes. Radar sends continuous signals to detect UAVs and collects return signals for analysis. The data is processed through the three models. In kNN, the Euclidean distance between the test data and training samples is calculated, and the UAV is classified based on the majority class among the five (5) nearest neighbors. While, Random Forest uses feature selection and bootstrapped datasets to train multiple decision trees, and combine their votes to classify the UAV. As SVM involves feature selection, kernel function selection (linear, polynomial, or RBF), hyperplane construction, and margin maximization to separate classes, and classify UAV based on a one-vs-all strategy. Then, the model with the highest accuracy is chosen to make the final classification decision. This ensures a systematic approach to UAV detection and classification using machine learning, and demonstrates the complementary work by the three integrated models.

A. UAV Parameters

The three simulated UAVs have the following parameters:

1) DJI Matrice 600: It is set to have $N_r^{M600} = 6\,rotors$, blades with a length $L_b^{M600} = 0.5\,m$, a rotor speed $\omega^{M600} = 2200\,RPM = 36.7\,Hz$, a propeller distance $D_p^{M600} = 1.13\,m$, and $N_a^{M600} = 6\,arms$.

2) DJI Matrice 300: It is set to have $N_r^{M300} = 4\,rotors$, blades with a length $L_b^{M300} = 0.4\,m$, a rotor speed $\omega^{M300} = 2400\,RPM = 40.0\,Hz$, a propeller distance $D_p^{M300} = 0.885\,m$, and $N_a^{M300} = 4\,arms$.

3) DJI Phantom 4: It is set to have $N_r^{M600} = 4\,rotors$, blades with a length $L_b^{P4} = 0.3\,m$, a rotor speed $\omega^{P4} = 3500\,RPM = 58.3\,Hz$, a propeller distance $D_p^{P4} = 0.35\,m$, and $N_a^{P4} = 4\,arms$.

These parameters are used to simulate the MD signatures generated by the rotor blades. The rotor speed (in Hz) and blade length define the periodic motion, while the propeller distance and number of arms influence the radar reflections. This level of detail is essential for modeling radar signals that differentiate between UAV types based on their distinct physical characteristics. The factors are significant in generating unique MD patterns that facilitate the classification and identification of the UAVs.

B. Radar Parameters

The radar system is simulated to operate in this research at a carrier frequency, $f_c = 77\,GHz$. It features a maximum range (detectable range), $R = 1200\,m$ and a range of resolutions, $\Delta R = 1\,m$. The radar's bandwidth is calculated as the speed of light divided by twice the range resolution, $B = \frac{c}{2\Delta R} = 150\,MHz$. The sweep time, $T_s = 5.5\,seconds$, while the chirp slope is derived from the bandwidth and sweep time, $\Upsilon = \frac{B}{T_s} = 2.27 \times 10^6\,Hzs^{-1}$. The radar's maximum beat frequency and range beat frequency, $f_{beat,max}$ are set to handle the operational parameters effectively. The system supports MD processing for velocities, $V_{max} = 50\,ms^{-1}$, with the maximum Doppler frequency set to accommodate these speeds, $f_{doppler} = \frac{2V_{max}f_c}{c} = 25.67\,kHz$. The radar simulation encompasses, $N_p = 10\,pulses$ each consisting of $N_c = 60\,chirps$ and $N_s = 80\,samples\,per\,chirps$. The pulse repetition frequency (PRF) is determined based on these parameters where Np, Nc, and Ns stand for number of pulses, number of chirps, and number of samples respectively.

1) Transmitted chirp signal: The transmitted chirp signal models how the radar transmits a signal over time. This chirp signal is essential for determining the range and Doppler characteristics of UAVs. The selected UAVs in this research are DJI Matrice 600, DJI Matrice 300, and DJI Phantom 4. Hence, the chirp signal detects and measures their distance and movement by analyzing the returned signals as in Eq. (1),

$$S_r(n,t) = a_t rect\left(\frac{\hat{t} - t_d}{\tau}\right) e^{j[2\pi f_c t_d + \pi\Upsilon(\hat{t}-t_d)^2]} \qquad (1)$$

where $S_r(n,t)$, is the transmitted baseband signal from the radar at the time $\hat{t}$ for the nth pulse and a_t is the amplitude of the return signal, often defined by the radar range equation, and it depends on the transmitted power, target radar cross-section (RCS), and range. While $rect\left(\frac{\hat{t}-t_d}{\tau}\right)$ is the rectangular window function that represents the radar pulse shape. The function limits the signal to the time interval τ, which is the pulse width. It centers on the delayed time, t_d where t_d is the round-trip time delay is related to the range R of the target by $t_d = \frac{2R}{c}$. Meanwhile, $e^{j[2\pi f_c t_d + \pi\Upsilon(\hat{t}-t_d)^2]}$ is the complex exponential that describes the frequency modulation (FM) of the chirp signal and f_c is the radar's carrier frequency (center frequency of the transmitted signal). In addition, γ is the chirp rate or chirp slope, representing the rate of frequency change in the transmitted chirp signal and $(\hat{t} - t_d)^2$ represents the quadratic phase term, where $\hat{t}$ is the time within a pulse and t_d is the delay due to the target's range.

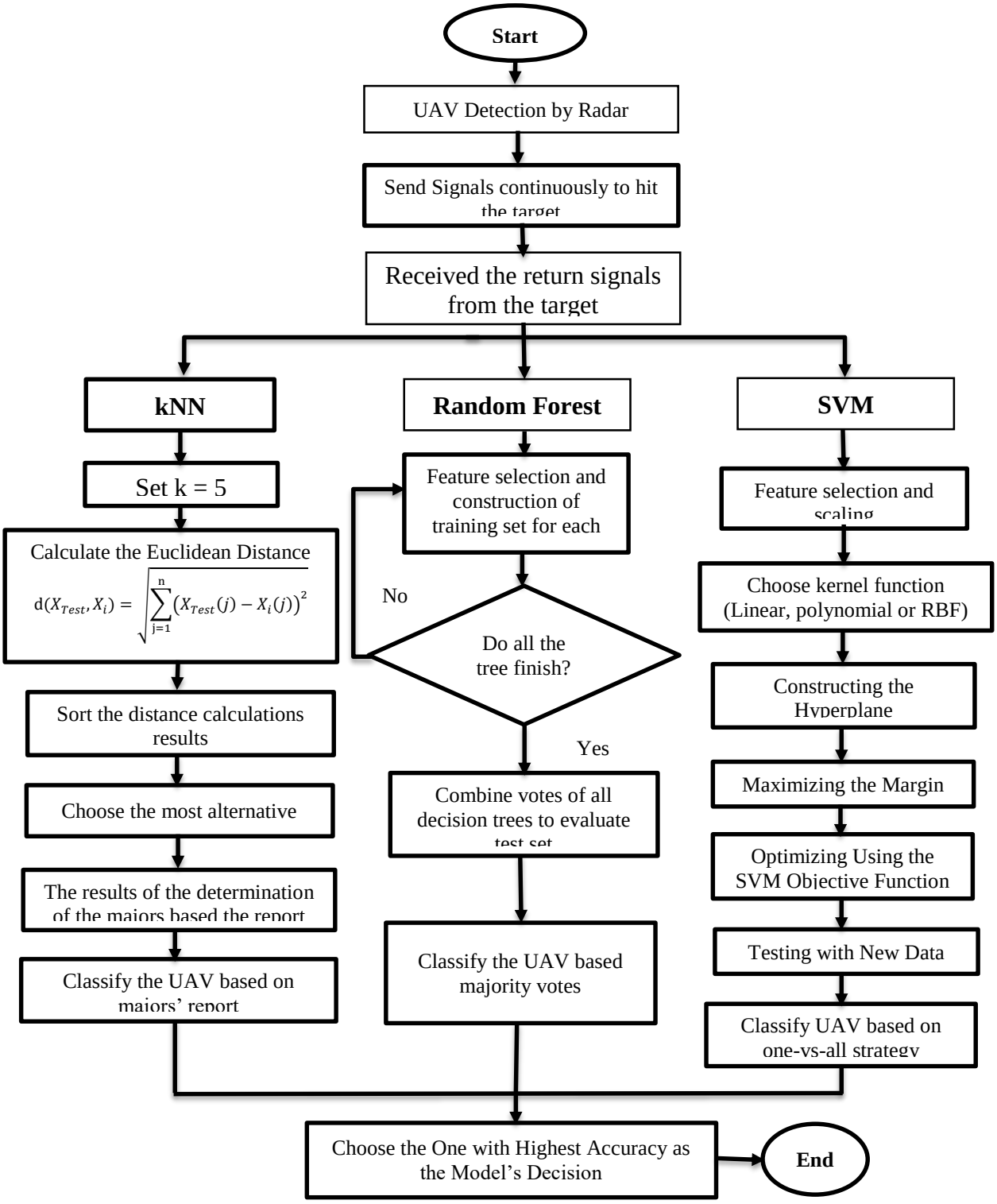


Fig. 1. Flowchart of the proposed model.

2) Received chirp signal: The received chirp signal $S_R(t)$ models the interaction between the radar and UAVs. The time delay τ =2r/c reflects the time it takes for the radar signal to travel to the UAV and back. This model is crucial for calculating the distance of the UAVs from the radar. Therefore, the received signal can be written as in Eq. (2),

$$S_{received}(n,t) = S_{UAV}(n,t) + S_{clutter}(n,t) + Noise(0,\sigma^2) \quad (2)$$

where $S_{UAV}(n,t)$ is the signal reflected by the target (UAV), $S_{clutter}(n,t)$ is the clutter signal, which is the sum of reflections from multiple clutter sources and $Noise(0,\sigma^2)$ is the additive noise in Gaussian distribution.

3) Dechirped intermediate frequency signal: The dechirping is the mixing (multiplying) of the received signal with a delayed version of the transmitted signal (reference chirp). This multiplication produces the Intermediate Frequency (IF) signal as described in Eq. (3), which contains the difference between the transmitted and received frequencies.

$$S_{IF}(n,t) = S_{received}(n,t) \times S_r^*(n,t) \tag{3}$$

$S_r^*(n,t)$ is the complex conjugate of the transmitted signal $S_{received}(n,t)$. When the received signal is mixed with the conjugate of the transmitted chirp, the high-frequency terms cancel out, leaving behind a low-frequency signal (beat signal) that encodes target information such as range and Doppler shifts. The resulting signal $S_{IF}(n,t)$ will have components related to the difference in time delay and Doppler shift between the transmitted and received signals. The UAV reflection of the IF signal is given by Eq. (4), where $f_{beat} = f_c(t_r - t_d)$ is the beat frequency proportional to the time delay, $(t_r - t_d)$, which relates to the range of the target. $\phi(t)$ is the phase component, which includes Doppler frequency information.

$$S_{IF,UAV}(n,t) = a_r rect\left(\frac{\hat{t} - t_r}{\tau}\right) e^{j[2\pi f_{beat}t + \phi(t)]} \tag{4}$$

The IF signal after dechirping contains beat frequency and phase modulation. The beat frequency is proportional to the range of the UAV (or other targets); the greater the range, the higher the beat frequency. Meanwhile, phase modulation is caused by Doppler shifts, which provide information about the relative velocity of the UAV.

4) Dechirped signal: The dechirped signal, $S_0(t,t_s)$ assists in analyzing the radar return signal by removing residual phase terms. This step is essential for accurately processing and analyzing the data collected from UAVs, ensuring that the range and Doppler measurements used for classification are precise. The dechirped signal would also include contributions from clutter and noise. Therefore, the IF signal after dechirping becomes Eq. (5), where, $S_{IF,UAV}(n,t)$ is the beat signal from the UAV, while $S_{IF,clutter}(n,t)$ represents the beat signals from clutter sources and $Noise(0,\sigma^2)$ is the noise (modeled as Gaussian noise).

$$S_{IF}(n,t) = S_{IF,UAV}(n,t) + S_{IF,clutter}(n,t) + Noise(0,\sigma^2) \tag{5}$$

5) MD effect: The MD effect is significant in detecting and identifying the UAVs' distinct features, such as rotor blades' motion. MD is a time-varying frequency shift caused by small periodic motions like the rotor blades of the UAV. It is modeled in Eq. (6) where, $v(t)$ is the instantaneous velocity of the rotating or moving part, λ is the wavelength of the radar signal, and $f_{mD}(t)$ represents the MD frequency shift.

$$f_{mD}(t) = \frac{2v(t)}{\lambda} \tag{6}$$

The overall received signal, including the MD effect, is represented in Eq. (7) where $2\pi f_c t_r$ is the bulk Doppler shift and the MD shift is $2\pi f_{mD}(t)t$.

$$S_r(n,t) = a_r rect\left(\frac{\hat{t} - t_r}{\tau}\right) e^{j[2\pi f_c t_r + 2\pi f_{mD}(t)t + \pi\Upsilon(\hat{t}-t_r)^2]} \tag{7}$$

6) Received Signal with MD and Noise: The total received radar signal, including the reflected signal from the UAV, clutter sources, MD effects, and noise, can be modeled in Eq. (8).

$$\begin{aligned} S_{received}(n,t) &= a_r rect\left(\frac{\hat{t} - t_r}{\tau}\right) e^{j[2\pi f_c t_r + 2\pi f_{mD}(t)t + \pi\Upsilon(\hat{t}-t_r)^2]} \\ &+ \sum_{i=1}^{N_c} a_{ci} rect\left(\frac{\hat{t} - t_{ci}}{\tau}\right) e^{j[2\pi f_c t_{ci} + 2\pi f_{mD}(t)t + \pi\Upsilon(\hat{t}-t_{ci})^2]} \\ &+ Noise(0,\sigma^2) \end{aligned} \tag{8}$$

C. File Format and Metadata

The data is stored in an Excel file and is organized into four sheets, each containing 8,120,601 samples of the three different UAVs (DJI Matrice 600, DJI Matrice 300, and DJI Phantom 4). The data is captured as the UAVs move within a 3D space ranging from 1 km away towards the radar to 1 km towards the radar. The movement is recorded at 10-meter intervals in each direction. Each sheet includes the following data for every sample: Raw RX Data, UAV Type, UAV Features, and Location. The captured information is utilized to train the machine learning models kNN, Random Forest, and SVM for classification purposes.

D. UAV Classification

The classification of UAVs is an approach for identifying UAV types and their operational modes based on the received radar information, specifically range-Doppler maps. The method leverages the kNN, Random Forest, and SVM algorithms, which classify a test sample by considering its proximity to the labeled training samples in features' space as described in equation (9). The dataset is represented as $X = \{x_1, x_2, \cdots, x_N\}$, where, x_i represents the features of a sample, such as the range-Doppler map. Each $y_i \in Y$, where Y contains UAV types or operational modes. yi ∈ {DJI Matrice 600, DJI Matrice 300, Phantom 4}

1) Data preparation and feature extraction: To train the classifiers for identifying UAV types and their operational states, the raw radar signal (range-Doppler maps) is first preprocessed. The range-Doppler map, which represents the response of the radar to a moving object in terms of range and velocity, is flattened into a one-dimensional vector. Mathematically, let the range-Doppler map for a UAV U and velocity V be denoted as $RD_{U,V}(r, f_D)$, where r represents range and f_D represents Doppler frequency. To simplify this for machine learning, each map $RD_{U,V}$ is reshaped into a feature vector $X_{u,v} \in \mathbb{R}^n$, where n is the total number of points in the

range-Doppler map. Thus, the feature matrix X for all UAVs and states becomes Eq. (9):

$$X = \begin{bmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,V} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,V} \\ \vdots & \vdots & \ddots & \vdots \\ X_{U,1} & X_{U,2} & \cdots & X_{U,V} \end{bmatrix} \quad (9)$$

where U represents the number of UAV types and V represents the number of velocity states (including hovering). Each row in X corresponds to the real part of the radar signal response, ensuring compatibility with standard machine learning models as indicated in Eq. (10).

$$Y = \begin{bmatrix} UAV\ Type_1 & Operation_1 \\ UAV\ Type_2 & Operation_2 \\ \vdots & \vdots \\ UAV\ Type_N & Operation_N \end{bmatrix} \quad (10)$$

a) Dataset: The dataset consists of radar return signals from three different types of UAVs: DJI Matrice 600, DJI Matrice 300, and DJI Phantom 4. Each UAV is simulated, and its features are captured across a range of distances from the radar, spanning between [1000; 1000; 1000] meters (1km) towards the radar and [-100; -1000; -1000] meters (1km) away from the radar, with measurements taken in 10m intervals inclusive. The features extracted from the UAVs are the reflected radar signals from the UAVs that capture unique characteristics, structural design, movement patterns, and operational parameters such as rotor speed, body dimensions, and altitude. These features, derived from the signal's amplitude, phase shifts, and Doppler effects, provide distinctive signatures that can be used for identifying and classifying each UAV at various distances and orientations relative to the radar system. These reflected signals can be used to train machine-learning models to recognize the specific identity of each UAV, enabling robust identity-based authentication and detection. The dataset is extensive, capturing radar return signals from multiple locations within a specified range. Each UAV covering all locations within the range of [-1000; -1000; -1000] to [1000; 1000; 1000] at intervals of 10 meters, the total number of samples is calculated as the following.

Total samples per UAV:

$$\text{UAV} = 201 \times 201 \times 201 = 8,120,601$$

Therefore, the total dataset samples of the three UAVs (DJI Matrice 600, DJI Matrice 300, and DJI Phantom 4) is the total number of samples across the UAVs.

$$\text{Total Dataset Samples} = 8,120,601 \times 3 = 24,361,803 \text{ samples}$$

This results in a dataset containing 24,361,803 samples in total, providing unique radar profiles for each UAV at all specified distances. This comprehensive dataset enables efficient training of machine-learning models for UAV identification and classification based on their radar return signals.

The structure of the captured information is described as follows:

- Reflected Signal: This matrix captures the amplitude of the radar return signals reflected by the DJI Matrice 600 at [1000; 1000; 0]:

$$X = \begin{bmatrix} -1.15193589623129 & 2.79463633718335 & \cdots & 0.854456914633749 \\ -0.267289748658581 & -0.368824215433807 & \cdots & -0.671501576572502 \\ \vdots & \vdots & \ddots & \vdots \\ -0.382196398043339 & 0.0841435521978612 & \cdots & -1.14575173405667 \end{bmatrix}$$

- Drone Type: The UAV information captured is that of DJI Matrice 600.
- Drone Features: The details of the UAV captured include the number of rotors, rotor radius, rotor speed, and other relevant features [6 0.5 36.6666666666667 1.13 6].
- Location: Indicates the coordinates of the UAV towards the radar [1000; 1000; 0].

b) Feature extraction: The range-Doppler maps are generated from radar signals for each UAV, capturing significant features related to the movement and MD signatures of the drone. These maps encode essential information about the range and velocity of UAVs, making them a valuable feature set for classification tasks. The formula of Range-Doppler maps is defined in Eq. (11):

$$R_D(n,f) = FFT_t\{FFT_r\{S_r(n,t)\}\} \quad (11)$$

Therefore, the features extracted from the Range Doppler maps are Mean Range μ_r, Mean Doppler-Shift μ_d, Standard Deviation of Range σ_r, and Standard Deviation of Doppler Shift σ_d. These are described in Eq. (12), (13), (14), and (15) respectively. Others include Peak-to-Average-Ratio (PAR), Spectral Centroid (SC), Spectral Bandwidth (SB), Peak Amplitude (PA), Number of Peaks, Energy Ratio (ER), and Entropy. The PAR, SC, SB, PA, ER, and Entropy are represented in Eq. (16), (17), (18), (19), (20) and (21) respectively.

$$\mu_r = \frac{1}{N}\sum_{n=1}^{N} r(n) \quad (12)$$

$$\mu_d = \frac{1}{M}\sum_{f=1}^{M} d(f) \quad (13)$$

$$\sigma_r = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(r(n)-\mu_r)^2} \quad (14)$$

$$\sigma_d = \sqrt{\frac{1}{M-1}\sum_{f=1}^{M}(d(f)-\mu_d)^2} \quad (15)$$

$$(PAR) = \frac{max_d}{\mu_d} \quad (16)$$

$$SC = \frac{\sum_f f \times R_D(n,f)}{\sum_f R_D(n,f)} \quad (17)$$

$$BD = \sqrt{\frac{\sum_f (f - SC)^2 \times R_D(n,f)}{\sum_f R_D(n,f)}} \quad (18)$$

$$PA = max_{n,f} \times R_D(n,f) \quad (19)$$

The number of peaks counts the significant peaks in the range-Doppler map, which is related to the number of rotating blades or moving parts.

$$ER = \frac{\sum_{(n,f)\in Region} R_D(n,f)}{\sum_{n,f} R_D(n,f)} \quad (20)$$

ER is a ratio of energy in specific regions of the range-Doppler map to the total energy, which indicates specific features related to the UAVs.

$$Entropy = -\sum_{n,f} \frac{R_D(n,f)}{E} \log\left(\frac{R_D(n,f)}{E}\right), \quad (21)$$

Where

$$E = \sum_{n,f} R_D(n,f)$$

Therefore,

$$Features\ Vector\ (F_i) = [\mu_r, \mu_d, \sigma_r, \sigma_d, PAR, SC, SB, PA, Number\ of\ Peaks, ER, Entropy\ ]$$

The features vector provides a comprehensive representation of the UAV's radar signature that can ensure classification accuracy.

2) Training the classifiers: The kNN, Random Forest, and SVM classifiers are trained using the feature matrix X and the corresponding label matrix Y. The goal is to classify two outputs: the type of UAV and its operation. Let $y_{UAV\ Type}$ be the vector containing the UAV type labels, and $y_{Operation}$ be the vector containing the operational states (hovering or moving). The kNN classifier identifies the k nearest training points for each test point based on a chosen distance metric. In contrast, the Random Forest model uses an ensemble of decision trees to make predictions by averaging the outputs of multiple trees, while the SVM model finds the optimal hyperplane that maximizes the margin between classes for classification. Given a new test point $X_{Test} \in \mathbb{R}^n$, the classifier computes the distances to all training points $X_i \in \mathbb{R}^n\ in\ X$ as in Eq. (22).

$$\mathrm{d}(X_{Test}, X_i) = \sqrt{\sum_{\mathrm{j}=1}^{\mathrm{n}} \left(X_{Test}(j) - X_i(j)\right)^2} \quad (22)$$

The kNN, Random Forest, and SVM classifiers are each used to classify both UAV type and operational state. The kNN classifier identifies the k nearest points and uses majority voting to assign the test point X_{Test} to a class for both the UAV type and the operation. While Random Forest aggregates predictions from decision trees, and SVM identifies the optimal hyperplane to separate classes. In this case, two separate classifiers are trained: one for UAV-type prediction and one for operational state prediction.

3) Classification: The kNN, Random Forest, SVM models are trained to classify new radar signal data from unknown UAVs. For each radar signal, represented by its corresponding range-Doppler map X_{Test}, the models predicts both the UAV type and its operation. The UAV type classifier, $f_{UAV\ Type}$, assigns a predicted class label based on the features of the test data:

$$\hat{y}_{\mathrm{UAV\ Type}} = \mathrm{f}_{\mathrm{UAV\ Type}}(\mathrm{X}_{\mathrm{Test}}) \quad (23)$$

Similarly, the operation classifier, $f_{Operation}$, predicts the UAV's operational state by analyzing the same test signal.

$$\hat{y}_{Operation} = f_{Operation}(X_{Test}) \quad (24)$$

Both classifiers work in tandem to identify the UAV type and its operational state from the radar signal data, making predictions based on the nearest neighbors in the training set.

E. Visualization of the Signals

The visualization of radar signals from various UAVs is critical for understanding their operational characteristics and enhancing classification tasks. The research used the generated raw data to visualize the UAVs' extracted features through maps. These maps include time/frequency spectrograms, mean spectrograms, and range-Doppler maps of the captured signals to facilitate detailed analysis of UAV's behaviors under different conditions.

1) Time/Frequency-Spectrograms: The time/frequency spectrograms provide a dynamic view of how the frequency content of the radar signals varies over time. The spectrograms get frequency on the vertical axis and time on the horizontal axis to demonstrate how the UAV's motion influences the received radar signals. Fig. 2 depicts rapid changes in frequency, indicating a UAV accelerating or maneuvering, while stable frequency patterns suggest hovering or cruising at a constant speed. This visualization is particularly useful for identifying unique operational signatures associated with different UAV types.

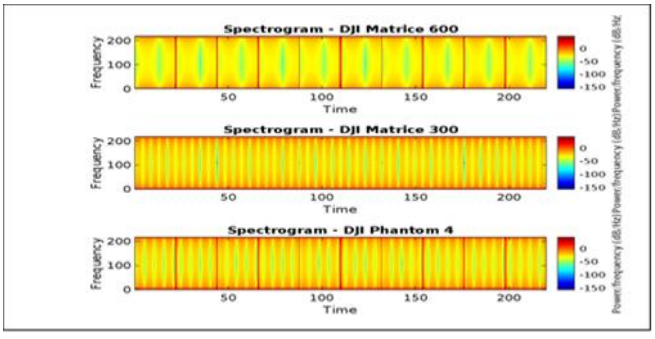


Fig. 2. UAV spectrograms.

2) Time/Frequency-mean spectrograms: The mean spectrograms is generated by averaging on multiple time/frequency spectrograms to emphasize the consistent features inherent to each UAV type. The smoothing out the

variability in individual recordings, mean spectrograms highlight the dominant frequencies and patterns associated with specific UAVs as indicated in Fig. 3. This helps in distinguishing between UAVs, as each type tends to exhibit distinct frequency profiles that can be utilized for classification.

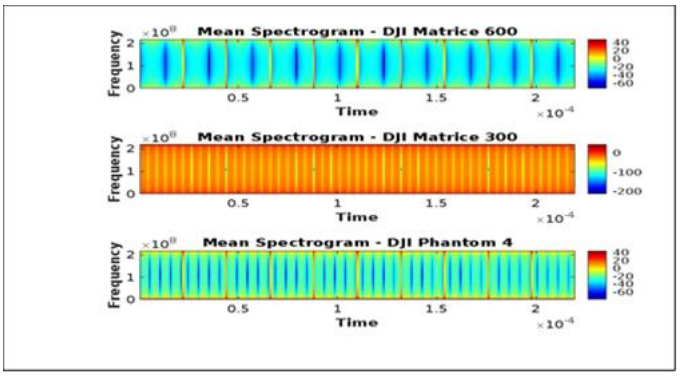


Fig. 3. UAV mean spectrograms

3) Doppler frequency/Range-Range-Doppler map: The Range-Doppler maps are pivotal in integrating both range and Doppler information into a single representation as indicated in Fig. 4. These maps reveal the detected UAVs' characteristics in a comprehensive manner. The intensity of the colors within the maps indicates the strength of the received radar signals, highlighting features specific to different UAV types. The variations in intensity help to differentiate between larger UAVs, which may have a stronger radar return, and smaller ones, which produce weaker signals.

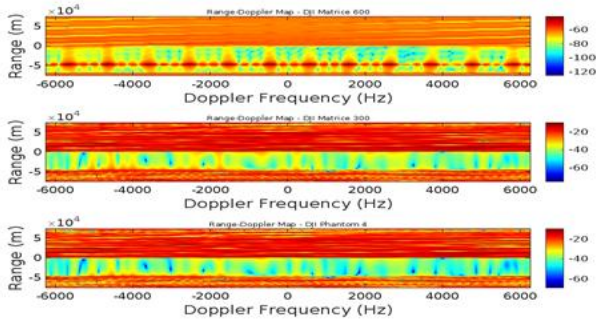


Fig. 4. UAV Range-Doppler Map

4) Doppler frequency/Range-Mean Range-Doppler maps: The mean range-Doppler maps are generated by averaging the range-Doppler data over multiple samples. This process reduces noise and highlights the typical signatures of each UAV type as described in Fig. 5. The mean range-Doppler maps assist in the classification process, making it easier to identify the unique characteristics associated with each UAV.

5) Doppler intensity plots: The Doppler intensity plots focus on specific points within the Doppler domain, illustrating the intensity of the received signals at various frequencies as presented in Fig. 6. These plots are beneficial for examining how signal intensity varies with Doppler frequency, offering insights into the operational states of the UAVs. The frequencies exhibit higher intensity levels during specific operations, such as takeoff or landing, enabling observers to infer the UAV's activity at a given time.

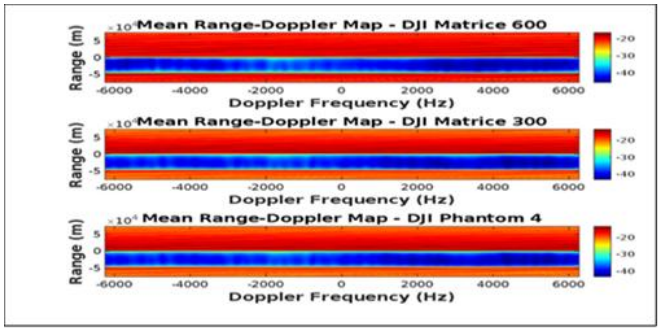


Fig. 5. UAV mean range-doppler map.

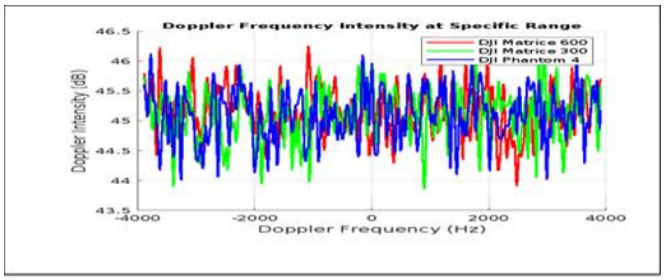


Fig. 6. UAV doppler frequency intensity.

IV. Evaluation and Testing

The captured radar signals information is deployed to kNN, Random Forest, and SVM to enable real-time classification of unknown UAVs based on their range-Doppler signatures. The relevant features are extracted from the radar signal and input into the classifiers to predict both the UAV type and its operational state (e.g., hovering or moving). The performances of the models are evaluated by comparing the predictions with the known ground truth. The performance metrics include classification accuracy, the F1 score, true positive confidence, and the false alarm rate. Additionally, the classification time delay is assessed to determine how fast the system can make predictions to ensure the model's viability for real-time UAV detection and authentication. These metrics provide an inclusive assessment of the classifier's effectiveness in accurately identifying UAV types and operational states from radar data.

1) Accuracy: This measures the proportion of correct predictions out of the total predictions made by the classifier.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{25}$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative.

2) F1 Score: It provides a balance between precision and recall, especially useful for imbalanced classes. It is the harmonic mean of precision and recall.

$$F1\ Score = 2 \times \frac{Precision\ \times\ Recall}{Precission + Recall} \tag{26}$$

Where

$$Precision = \frac{TP}{TP+FP}, \text{ and } Recall = \frac{TP}{TP+FN}$$

3) True positive confidence: This reveals the confidence level of the classifier in its correct predictions, expressed in a percentage form. It is the average confidence score assigned to true positive predictions.

$$TPC = \frac{\Sigma_{1=1}^{n} Confidence\ (TP_i)}{n} \quad (27)$$

where TP_i is the confidence of each true positive prediction and n is the number of classes.

4) False Alarm Rate (FAR): This indicates the number of times the classifier incorrectly predicts the presence of a UAV when there is none (false positive rate).

$$FAR = \frac{FP}{FP + TN} \quad (28)$$

5) Classification time delay: It measures the time taken for the classifier to process the radar signal and make a prediction. This is critical for real-time systems and can be represented as:

$$CTD = t_{end} - t_{start} \quad (29)$$

Where CTD is the classification delay time.

V. Result and Discussion

The results of classification performance of the three models (kNN, Random Forest, and SVM) have been evaluated across three UAV classes (DJI Matrice 600, DJI Matrice 300, and DJI Phantom 4). The results, as presented in confusion matrices shown Fig. 7, Fig. 8, and Fig. 9, and summarized in Table II, Table III, and Table IV, provide overview of how each model performs. The performances are in terms of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), offering insights into their strengths and weaknesses.

	1	2	3	Percentage %	
1	36	4	4	89.40%	10.60%
2	4	18	22	62.90%	37.10%
3	2	19	23	62.90%	37.10%

Fig. 7. kNN's classification confusion matrix.

TABLE II. kNN's Classification Summary

Class	*TP*	*TN*	*FP*	*FN*
1 (DJI Matrice 600)	36	82	6	8
2 (DJI Matrice 300)	18	65	23	26
3 (DJI Phantom 4)	23	60	26	23
Total	**77**	**207**	**55**	**57**

The kNN exhibits varying levels of performance in the classification of UAV types. Class 1 (DJI Matrice 600): The classifier performs well and achieved 36 (TPs) with minimal FP (6) and FN (8). It shows the highest accuracy (89.4%) in this class, indicating that it can effectively distinguish the DJI Matrice 600 from the other UAVs. However, the classifier struggles with Class 2 (DJI Matrice 300), where it only correctly identifies 18 samples as Matrice 300, while misclassifying 23 samples from other classes as Matrice 300. In addition, it misses 26 actual Matrice 300 samples, leading to a moderate performance in the class. While, in Class 3 (DJI Phantom 4), kNN performs moderately well but faces challenges in balancing FPs and FNs, correctly classifying 23 samples as Phantom 4, while misclassifying 26 samples from other classes and failing to classify 23 actual Phantom 4 samples.

	1	2	3	Perc. %
1	44			100%
2		44		100%
3			44	100%

Fig. 8. Random forest's classification confusion matrice.

TABLE III. Random Forest's Classification Summary

Class	*TP*	*TN*	*FP*	*FN*
1 (DJI Matrice 600)	44	88	0	0
2 (DJI Matrice 300)	44	88	0	0
3 (DJI Phantom 4)	44	88	0	0
Total	**132**	**264**	**0**	**0**

In contrast, the Random Forest demonstrates outstanding performance in classifying all the UAV classes, and achieved 100% classification with no FPs or FNs, as indicated in Table III. Each class (DJI Matrice 600, DJI Matrice 300, and DJI Phantom 4) samples are correctly classified, and no misclassifications occur. It recorded 132 (TPs) and 264 (TNs) across all classes, and achieved an accuracy of 100%, showcasing its robustness and ability to effectively separate UAV classes. This perfect performance shows the utility of Random Forest for UAV detection and classification tasks, indicating that it can reliably distinguish between different UAV types without ambiguity.

Meanwhile, the SVM model performs well, with strong results across all three classes. Class 1 (DJI Matrice 600), the SVM achieves perfect classification, with 100% precision and 100% recall, correctly identifying all instances of this class. This suggests that the SVM model excels at detecting the DJI Matrice 600 UAV. While, in Class 2 (DJI Matrice 300), the model achieves a precision of 79.1% and a recall of 86.4%, with a

resulting F1-score of 0.826. While some FPs are present, the model correctly identifies the majority of DJI Matrice 300 instances, but occasional confusion with other UAV types lowers its performance slightly. Also, in Class 3 (DJI Phantom 4), the SVM model achieves 85% precision and 77.3% recall, with an F1-score of 0.81. While the model correctly identifies most Phantom 4 samples, some instances are misclassified as other UAV types.

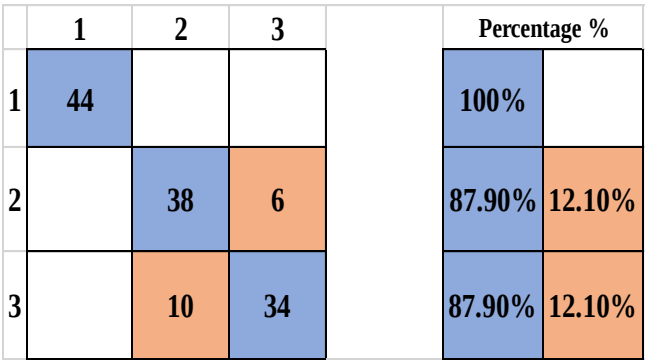


Fig. 9. SVM's classification confusion matrice.

TABLE IV. SVM'S CLASSIFICATION SUMMARY

Class	*TP*	*TN*	*FP*	*FN*
1 (DJI Matrice 600)	44	88	0	0
2 (DJI Matrice 300)	38	78	10	6
3 (DJI Phantom 4)	34	82	6	10
Total	**114**	**248**	**16**	**16**

TABLE V. SUMMARY OF PERFORMANCE OF THE THREE MODELS

Class	*Accuracy*	*F1 Score*	*TPC*	*FAR*
kNN	71.73%	58.14%	58%	0.21
Random Forest	100%	100%	100%	0.00
SVM	93.27%	88%	87%	0.06

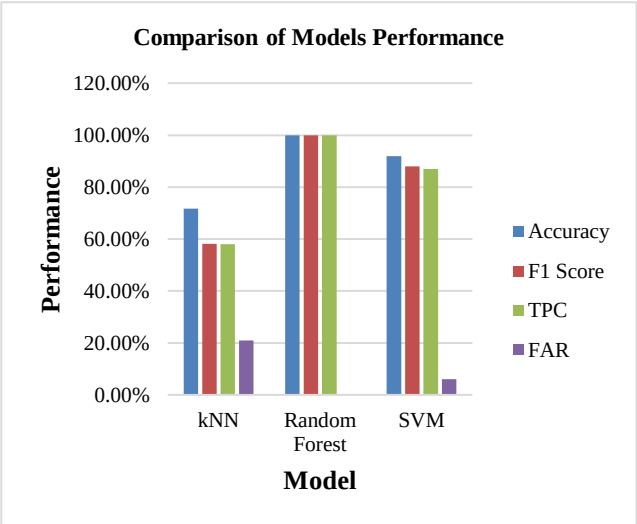


Fig.10. Models performances.

The Random Forest emerged as the best-performing model due to its perfect classification accuracy (100%) and zero false alarms as shown in Fig. 10 and Table V. This makes it well-suited for real-time UAV monitoring, where minimizing errors is paramount. Although kNN and SVM achieved high accuracy for some UAV types, their issues with misclassifying between DJI Matrice 300 and Phantom 4 suggest that further tuning or additional feature engineering is needed to improve their performance in distinguishing these specific UAV types.

VI. CONCLUSION AND FUTURE DIRECTION

This research presents significant strides in the detection and authentication of UAVs through the creation of a unique radar dataset comprising three distinct UAV models (DJI Matrice 600, DJI Matrice 300, and Phantom 4). The research has laid the groundwork for enhanced UAV detection and classification. The utilization of MD radar signals allows for detailed analysis of UAV characteristics, facilitating accurate identification and robust authentication processes. Among the three machine learning models tested, Random Forest demonstrated exceptional performance, achieving 100% classification accuracy with zero false alarms, making it highly suitable for real-time UAV monitoring where precision is critical. While kNN and SVM models also showed strong results, they encountered misclassification issues attributed to the similarities in the radar signatures of the DJI Matrice 300 and DJI Phantom 4 UAVs. These UAVs share the same number of rotors and propellers, as well as other design features. Their structural similarities result in overlapping micro-Doppler effects and radar reflections, making it difficult for the models to distinguish between the two UAV classes, suggesting a need for further refinement.

However, future work will focus on developing a system that leverages the Remote ID (RID) policy, and the radar datasets generated in this study to improve the detection, identification, and authentication of UAVs. The system will utilize the kNN, Random Forest, and SVM models, with particular attention to improving the performance of kNN and SVM to reduce misclassifications and enhance accuracy. These enhancements, combined with the broader application of RID data, will enable a more robust scalable solution that will detect and identify unknown UAVs and ensure monitoring and security in various operational environments.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Etienne Innovation Sdn. Bhd. for their invaluable support and contribution to this research. Additionally, the authors extend their heartfelt thanks to Universiti Tun Hussein Onn Malaysia (UTHM) for providing essential resources, facilities, and support throughout the research process. Their encouragement and commitment have been instrumental in the successful completion of this study.

REFERENCES

[1] C. F. E. De Melo et al., "UAVouch: A Secure Identity and Location Validation Scheme for UAV-Networks," IEEE Access, vol. 9, pp. 82930–82946, 2021, doi: 10.1109/ACCESS.2021.3087084.

[2] M. Ahsan, K. E. Nygard, R. Gomes, M. M. Chowdhury, N. Rifat, and J. F. Connolly, "Cybersecurity Threats and Their Mitigation Approaches

Using Machine Learning—A Review," Journal of Cybersecurity and Privacy, vol. 2, no. 3, pp. 527–555, 2022, doi: 10.3390/jcp2030027.

[3] M. Ináncsi, "Cybersecurity Challenges of the Civilian Unmanned Aircraft Systems," Hadmérnök, vol. 17, no. 2, pp. 205–216, Sep. 2022, doi: 10.32567/hm.2022.2.14.

[4] S. Basak, S. Rajendran, S. Pollin, and B. Scheers, "Combined RF-Based Drone Detection and Classification," IEEE Trans Cogn Commun Netw, vol. 8, no. 1, pp. 111–120, 2022, doi: 10.1109/TCCN.2021.3099114.

[5] U. Seidaliyeva, L. Ilipbayeva, K. Taissariyeva, N. Smailov, and E. T. Matson, "Advances and Challenges in Drone Detection and Classification Techniques: A State-of-the-Art Review," Sensors, vol. 24, no. 1, pp. 1–31, 2024, doi: 10.3390/s24010125.

[6] S. A. Zulkarnain, S. Zulkifli, and 'Aiffah Mohd Ali, "Identification and Analysis of MD Signature of a Bird Versus Micro-UAV," Journal of Advanced Research in Micro and Nano Engineering, vol. 16, no. 1, pp. 102–113, 2024, doi: 10.37934/armne.16.1.102113.

[7] Y. He, J. Zhang, R. Xi, X. Na, Y. Sun, and B. Li, "Detection and Identification of Non-cooperative UAV Using a COTS mmWave Radar," ACM Trans Sens Netw, vol. 20, no. 2, pp. 1–22, 2024, doi: 10.1145/3638767.

[8] Y. Zhao and Y. Su, "The Extraction of MD Signal with EMD Algorithm for Radar-Based Small UAVs' Detection," IEEE Trans Instrum Meas, vol. 69, no. 3, pp. 929–940, 2020, doi: 10.1109/TIM.2019.2905751.

[9] M. Ezuma, F. Erden, C. K. Anjinappa, O. Ozdemir, and I. Guvenc, "Detection and classification of UAVs using RF Fingerprints in the Presence of Wi-Fi and bluetooth interference," IEEE Open Journal of the Communications Society, vol. 1, no. November 2019, pp. 60–76, 2020, doi: 10.1109/OJCOMS.2019.2955889.

[10] V. V. Reddy and S. Peter, "UAV MD signature analysis using FMCW radar," IEEE National Radar Conference - Proceedings, vol. 2021-May, no. 1, pp. 1–6, 2021, doi: 10.1109/RadarConf2147009.2021.9454978.

[11] G. Ji, C. Song, and H. Huo, "Detection and Identification of Low-Slow-Small Rotor Unmanned Aerial Vehicle Using MD Information," IEEE Access, vol. 9, pp. 99995–100008, 2021, doi: 10.1109/ACCESS.2021.3096264.

[12] K. B. Kang, J. H. Choi, B. L. Cho, J. S. Lee, and K. T. Kim, "Analysis of MD Signatures of Small UAVs Based on Doppler Spectrum," IEEE Trans Aerosp Electron Syst, vol. 57, no. 5, pp. 3252–3267, 2021, doi: 10.1109/TAES.2021.3074208.

[13] D. Raval, E. Hunter, S. Hudson, A. Damini, and B. Balaji, "Convolutional neural networks for classification of drones using radars," Drones, vol. 5, no. 4, 2021, doi: 10.3390/drones5040149.

[14] H. C. Kumawat, M. Chakraborty, and A. Arockia Bazil Raj, "DIAT-RadSATNet-A Novel Lightweight DCNN Architecture for MD-Based Small Unmanned Aerial Vehicle (SUAV) Targets' Detection and Classification," IEEE Trans Instrum Meas, vol. 71, pp. 1–11, 2022, doi: 10.1109/TIM.2022.3188050.

[15] N. Rojhani, M. Passafiume, M. Sadeghibakhi, G. Collodi, and A. Cidronali, "Model-Based Data Augmentation Applied to Deep Learning Networks for Classification of MD Signatures Using FMCW Radar," IEEE Trans Microw Theory Tech, vol. 71, no. 5, pp. 2222–2236, 2023, doi: 10.1109/TMTT.2023.3231371.

[16] A. N. Sayed, O. M. Ramahi, and G. Shaker, "UAV Classification Utilizing Radar Digital Twins," IEEE Antennas and Propagation Society, AP-S International Symposium (Digest), vol. 2023-July, pp. 741–742, 2023, doi: 10.1109/USNC-URSI52151.2023.10237683.

[17] B. I. Ahmad et al., "A Review of Automatic Classification of Drones Using Radar: Key Considerations, Performance Evaluation, and Prospects," IEEE Aerospace and Electronic Systems Magazine, vol. 39, no. 2, pp. 18–33, 2024, doi: 10.1109/MAES.2023.3335003.

[18] J. Tian, C. Wang, J. Cao, and X. Wang, "Fully Convolutional Network-Based Fast UAV Detection in Pulse Doppler Radar," IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1–12, 2024, doi: 10.1109/TGRS.2024.3358956.

Application of Residual Graph Attention Networks Algorithm in Credit Evaluation for Financial Enterprises

Wenxing Zeng*
Southwest Petroleum University, Chengdu, China

***Abstract*—In the context of digital transformation of enterprises, credit evaluation of financial enterprises faces new challenges and opportunities. Digital transformation introduces a large amount of data and advanced analytical tools, providing richer information and methods for credit evaluation. In this paper, we propose a credit evaluation model based on improved quantum genetic algorithm and residual graph attention network (DRQGA-ResGAT), which aims to utilize the complex correlation data and multi-dimensional information among enterprises for enterprise credit evaluation. The credit evaluation model based on DRQGA-ResGAT performs well in dealing with large-scale and high-dimensional data and can significantly improve the accuracy of credit evaluation. The experimental results show that the ResGAT model combined with the improved quantum genetic algorithm performs even better, and the proposed model has a high precision rate in the credit evaluation of financial enterprises, which has a greater application value. Compared with the traditional ResGAT model, the model improves about 17.06% in precision rate.**

***Keywords*—Quantum genetic algorithm; residual networks; attention mechanisms; graph neural networks; credit evaluation**

I. INTRODUCTION

In the context of enterprise digital transformation, financial enterprise credit evaluation faces new challenges and opportunities. Digital transformation not only changes the operation mode of enterprises, but also introduces a large amount of data and advanced analytical tools, providing richer information and methods for credit evaluation [1]. With the development of big data and artificial intelligence technology, deep learning algorithms are more and more widely used in the financial field. In particular, Residual Graph Attention Network (ResGAT), which is based on graph neural network, is excellent in processing graph-structured data and has been widely used in credit evaluation of financial enterprises [2].

Digital transformation refers to the optimization of business processes and the enhancement of operational efficiency and service capabilities by enterprises through the application of information technology [3-5]. In the financial industry, digital transformation not only includes the automation of traditional business, but also involves the in-depth integration of emerging technologies such as big data, cloud computing, and artificial intelligence [6]. The application of these technologies enables financial firms to process huge data sets more efficiently, thus realizing more accurate credit evaluation [7]. Deep learning, as an important branch of artificial intelligence, is able to automatically extract and learn complex features in data by simulating the neural network structure of the human brain. In financial enterprise credit evaluation, deep learning algorithms can effectively handle heterogeneous data, such as financial statements, transaction records, social media data, etc., so as to improve the precision and reliability of credit evaluation [8].

Shi et al. formulated a hybrid kNN-GNN model, investigated the integration of multiple graphs and the introduction of edge weights, and verified the feasibility and validity of the integrated model [9]. Zhang used a convolutional neural network (CNN) and a long short-term memory (LSTM) network to build a model that utilizes soft attention, and the gradient is propagated back to the rest of the model through the Attention Mechanism module [10]. The problem of recurrent neural network (RNN), which suffers from severe gradient vanishing under long sequence training, was solved. Gao et al. proposed a credit scoring model based on contrast-enhanced and tree-enhanced embedding mechanisms, which automatically constructs interpretable cross-features by using a tree-based model to learn decision rules from the data [11]. Currently, multiple improvement methods of graph neural networks shine in the aspect of credit risk evaluation for enterprises, especially after verifying the excellent effect of the attention mechanism, the practical effect of the graph attention model becomes more and more significant. Zhang et al. used a multimodal learning strategy to fuse two different data sources, and the cascade vectors derived from the data fusion were used as inputs to the feed-forward neural network to predict the credit risk of SMEs [12]. Sang et al. developed a graph attention network called DialogueGAT for predicting financial risk by simultaneously modeling the speakers and their words in a conversation during a teleconference [13].

However, GAT usually falls into overfitting or model degradation occurs when training a large amount of data, so residual networks were incorporated into GAT. Huang et al. proposed a residual-based graph attention network (ResGAT), which solves the over smoothing problem of the traditional graph convolutional networks (GCNs) when dealing with graph data by introducing the residual connectivity and attention mechanism, making it able to capture important features in the data and complex relationships between nodes [14]. Zhang et al. proposed two graph-based algorithms, E-ResSAGE and E-ResGAT algorithms, which build on the established GraphSAGE and GAT algorithms, respectively, to integrate residual learning into GNNs using available graph

information [15]. Adding residual ties as a strategy to cope with high class imbalances aims to preserve the original information and improve the performance of a few classes. Fan et al. proposed a multi-headed residual graphical attention network model using graph neural networks to extract interaction features from compiled graphical data, and designed the SDAE-GPC model for assembly condition classification to derive graphical data inputs for the ResGAT regression model [16].

In credit evaluation, ResGAT can utilize the complex correlation data and multidimensional information among enterprises to predict their future credit risk. It is shown that the credit evaluation model based on ResGAT performs well when dealing with large-scale and high-dimensional data, and can significantly improve the precision of credit scoring. Zhou et al. proposed a credit default risk prediction model based on graphic attention network, which considers the potential relationship between users by constructing a multi-view graph [17]. The experimental results on a real dataset verified the model's validity. Song et al. proposed a multi-structured cascading graph neural network framework for ECR evaluation, which enhances the learning of enterprise representations based on enterprise graph structures with different granularities [18]. These studies show that deep learning algorithms such as ResGAT have significant advantages and potentials in dealing with complex data in financial enterprise credit evaluation.

However, the hyperparameters of the network in deep learning are usually difficult to set to the optimal combinations and often require repeated human experiments, so optimization algorithms were developed to be combined with deep learning models to accomplish hyperparameter tuning of the algorithms through automatic optimization search. Nebojsa et al. proposed an improved version of the population intelligence and monarch butterfly optimization algorithms for training feed-forward artificial neural networks to enhance the exploration capabilities and the intensification-diversification balance [19]. Stefan et al. proposed a new hybrid meta-heuristic approach to optimize network neuron parameters by using the hybrid bat algorithm for feed-forward neural network training [20]. Nebojsa et al. explored the application of population intelligence techniques for tuning hyper-parameters in convolutional neural networks by proposing an augmented meta-heuristic algorithm through the implementation of the automated method for hyperparameter optimization and structural design [21].

In summary, enterprise digital transformation provides new data sources and technical means for financial enterprise credit evaluation [22]. Deep learning algorithms, especially ResGAT, significantly improve the precision and reliability of credit evaluation by effectively processing graph-structured data. The development and application of these technologies not only promote the digitalization process of financial enterprises, but also provide strong support for enterprise risk management. By combining the research results of several literatures, it can be seen that the application of ResGAT in the credit evaluation of financial enterprises has a broad prospect and is worthy of further in-depth research and practice.

This article is organized as follows. Section II presents related research work. Section III presents the details of the credit evaluation model, including the problem description and the evaluation indicators used in the analysis. Section IV introduces the Quantum Genetic Algorithm (QGA) and its application in optimizing the ResGAT network. Section V presents the integration of QGA with ResGAT for financial enterprise credit evaluation and discusses the proposed enhancements. Finally, Section VI shows the experimental results, including dataset preparation, performance metrics, and a comparison of the DRQGA-ResGAT model with other optimization algorithms, discussion is given in Section VII followed by conclusions and future work in Section VIII.

II. Related Work

A. Graph Neural Networks in Credit Risk Prediction

Recent advancements in credit risk assessment (CRA) have shifted from traditional methods, which primarily rely on individual borrower or loan-level predictors, to more complex models that incorporate relational and network-based data. One significant approach combines graph-based models with machine learning techniques to better capture the interactions between different entities, such as borrowers and financial institutions. For example, the use of Relational Graph Convolutional Networks (RGCN) has proven effective in assessing the creditworthiness of Micro, Small, and Medium-sized Enterprises (MSMEs). By leveraging the topological structures of business relationships, RGCN helps identify key factors that influence credit risk, significantly improving the accuracy of predictions compared to conventional credit scoring models. This approach is further enhanced by integrating Random Forest (RF) classifiers, which categorize enterprises based on the embeddings generated by the graph model, achieving a balanced accuracy of 92% in the case of MSMEs in India [23]. Additionally, the dynamic nature of borrower relationships has been recognized as a crucial factor in predicting default risk. Recent work has incorporated Graph Neural Networks (GNNs) and Recurrent Neural Networks (RNNs) to model these evolving connections. By constructing a multilayer network where each layer reflects different sources of connection—such as geographical location or mortgage provider—these models offer a more nuanced understanding of how defaults propagate through networks over time. The use of custom attention mechanisms further refines these models, enabling them to weigh the importance of different time snapshots, which enhances their predictive power in behavioral credit scoring tasks [24]. Moreover, combining Graph Attention Networks (GAT) with Long Short-Term Memory (LSTM) networks has shown promise in capturing both the spatial and temporal dynamics of borrower interactions. Empirical results have demonstrated that such hybrid models not only outperform traditional methods in predicting default probabilities but also provide novel insights into the critical role of borrower relationships and time-sensitive data in credit risk assessment [25].

B. Graph Attention Networks and Residual Networks

Recent advances in deep learning have led to the development of residual graph attention networks (ResGAT) for a wide range of applications, showcasing the model's

versatility across different domains. In molecular property prediction, ResGAT has been applied to quantitative structure-activity relationship (QSAR) modeling, significantly improving the prediction of molecular properties by utilizing graph-structured data. The model effectively extracts key features from molecular graphs, addressing both regression and classification tasks. When tested on benchmark datasets, ResGAT demonstrated competitive performance and stability, outperforming state-of-the-art methods in terms of accuracy and generalizability [26]. In the field of software vulnerability detection, ResGAT has been integrated with a custom local feature extraction module and a dynamic loss function to handle imbalanced data, enabling the model to learn more effective node features from control flow graphs and achieve state-of-the-art results in detecting vulnerabilities [27]. In hyperspectral image (HSI) classification, a spectral-spatial variant of ResGAT (S²RGANet) has been proposed to address the limitations of conventional convolutional networks by combining spectral residual modules with graph attention for adaptive aggregation of spatial information. This method significantly improves classification accuracy, especially when the number of training samples is limited [28]. Lastly, in visual grounding tasks, ResGAT has been employed to model complex relationships between objects in images and their corresponding textual descriptions. A language-guided residual graph attention network (LRGAT-VG) is introduced to better handle long and complex expressions, incorporating language-guided data augmentation (LGDA) to enhance training data diversity. The model has achieved competitive performance on multiple visual grounding benchmarks, demonstrating its effectiveness in handling complex cross-modal tasks [29].

C. Hyperparameter Optimization

Hyperparameter optimization is a critical yet challenging task in machine learning, especially when models involve a large number of hyperparameters. Recent studies have proposed various methods to address this challenge. One approach focuses on modifying gradient-based methods to optimize multiple hyperparameters simultaneously, using two model selection criteria—cross-validation and evidence lower bound. The results show that models optimized using the evidence lower bound exhibit greater stability, particularly in noisy data scenarios, though they may yield slightly higher error rates than those selected via cross-validation. This method is particularly useful when dealing with overfitting or when cross-validation is computationally expensive [30]. Another study proposes a genetic algorithm-based approach, named HESGA (Hierarchical Evaluation Strategy Genetic Algorithm), to optimize hyperparameters of Graph Neural Networks (GNNs). By combining a full evaluation with a fast evaluation strategy, HESGA reduces the computational cost while maintaining the quality of the model. The method demonstrated its effectiveness in optimizing GNNs, showing superior performance over traditional Bayesian hyperparameter optimization methods on benchmark datasets [31]. A third approach, HyperBRKGA, introduces a population-based method for hyperparameter optimization that combines the Biased Random Key Genetic Algorithm (BRKGA) with an exploitation method. HyperBRKGA improves search efficiency by incorporating strategies like Random Walk and Bayesian Walk for better hyperparameter space exploration. The method outperformed traditional optimization algorithms, such as Grid Search and Random Search, on multiple datasets, demonstrating significant improvements in predictive performance [32].

III. Credit Evaluation Model

A. Description of the Problem

The application of enterprise digital transformation in credit evaluation of financial enterprises is one of the key topics in the current financial industry. With the rapid development of information technology, financial enterprises are constantly exploring and adopting new technological tools to enhance the precision and efficiency of their credit evaluation. The digital transformation of enterprises has provided financial institutions with rich data sources and advanced analytical tools, enabling them to assess the credit risk of borrowers or customers more comprehensively. This transformation is not limited to traditional financial data analysis, but also includes the integration and analysis of unstructured data (e.g., social media behaviors, spending habits, etc.) to more accurately portray a customer's credit profile. By applying algorithms like DRQGA-ResGAT, financial institutions are able to handle complex data patterns more efficiently and improve the predictive power and precision of credit evaluations within the framework of digital transformation, thereby optimizing lending decisions and risk management strategies.

B. Evaluation Indicators

In accordance with the basic principles of combining qualitative and quantitative indicators and combining potential and actual capabilities, the following enterprise credit rating indicator system structure was established. In addition, we divided the enterprise credit evaluation model into 9 first-level indicators and 25 second-level indicators, and each second-level indicator corresponds to a measurement content, as shown in Table I.

C. Model Data Normalization

The data depicted in Table I are discretely distributed and each level 1 indicator is independent of each other, so normalization can speed up the learning of the data. Define a set of input data $X = [x_1, x_1, \cdots, x_n]$ for the second-level indicators, and the maximum value of this set of data is $max(X)$ and the minimum value is $min(X)$. The normalized data is shown in Eq. (1).

$$x_i' = \frac{x_i - min(X)}{max(X) - min(X)}, i = 1,2,\cdots,n \quad (1)$$

TABLE I. EVALUATION INDEX SYSTEM

Level 1 Indicators	Level 2 indicators	Measuring Content
The quality of the enterprise itself	Information Disclosure Assessment Results for Listed Companies	character
Solvency indicators	Long-term debt-to-equity ratio	capacity
	Long-term borrowings to total assets	capacity
	Gearing	capacity
	Current ratio	capacity
	Long-term debt to working capital ratio	capacity
	Total EBITDA/Liability	capacity
	Equity multiplier	capacity
Indicators of operational capacity	Current asset turnover ratio	capacity
	Total asset turnover	capacity
Profitability indicators	Return on net assets	capacity
	Return on investment (ROI)	capacity
	Net profit margin on current assets	capacity
	R&D expense ratio	capacity
	Non-recurring gains and losses	capacity
	Cash to total profit ratio	capacity
Cash flow indicators	Net cash flow per share	capacity
Capacity development indicators	Growth rate of selling expenses	capacity
	Total asset growth rate	capacity
	Revenue growth rate	capacity
	Rate of capital accumulation	capacity
Shareholder profitability	Earnings per share	capital
	Net asset per share	capital
Long term assets	Net fixed assets of enterprises	collateral
Supply Chain Status	Supply chain concentration	condition

IV. QUANTUM GENETIC ALGORITHM BASED ON DYNAMIC REVOLVING DOORS

Quantum Genetic Algorithm (QGA) is an optimization algorithm that combines quantum computing and traditional genetic algorithms. It takes advantage of the superposition of quantum bits and the parallel computing property of quantum gates to make the search space wider and thus improve the optimization performance. Individuals in QGA are usually represented in the form of quantum bits, and their states are described by probability magnitude, and population evolution is carried out through quantum gate operations. QGA can effectively search for the global optimal solution, and it performs well in dealing with the complex optimization problems of multi-peak functions.

A. Quantum Encoding

According to the principle of quantum superposition states, in two mutually independent quantum states $|0\rangle$ and $|1\rangle$, their arbitrary linear superposition forms a quantum state:

$$|\varphi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{2}$$

where the squares of α and β denote the probability that the system is at $|0\rangle$ and $|1\rangle$, respectively, and $\alpha^2 + \beta^2 = 1$. The mathematical model of (2) can also be referred to as the probability amplitude, which is denoted as:

$$|\varphi\rangle = \alpha\begin{bmatrix}0\\1\end{bmatrix} + \beta\begin{bmatrix}1\\0\end{bmatrix} = \begin{bmatrix}\alpha\\\beta\end{bmatrix} \tag{3}$$

B. Quantum Chromosomes

Traditional genetic algorithms usually use the problem parameter variable itself as an individual for optimization calculations and are computed through binary coding to form chromosomes that serve as information carriers so that they can convert the solution space into a search space that can be processed, thus completing the construction of the genetic structure of the individual.

While the quantum chromosome does not contain the problem solution directly, it constructs the quantum chromosome through the encoding method of quantum bits and expresses the state information through the form of probability amplitude shown in Eq. (3). According to Eq. (3), the Bth

quantum chromosome of generation t with m quantum bits and m observation angles i is:

$$X_i^t = \begin{bmatrix} \alpha_{i1}^t, \alpha_{i2}^t, \cdots, \alpha_{im}^t \\ \beta_{i1}^t, \beta_{i2}^t, \cdots, \beta_{im}^t \end{bmatrix} \quad (4)$$

C. Dynamic Quantum Revolving Door

In quantum theory, transitions between individual quantum states are realized using quantum gates. The quantum gate rotates the probability amplitude angle of the quantum bits, taking into account the most individual information, which always makes the population converge to the optimal solution. The expression for a quantum rotating gate is:

$$U(\Delta\theta) = \begin{bmatrix} cos(\Delta\theta), -sin(\Delta\theta) \\ sin(\Delta\theta), cos(\Delta\theta) \end{bmatrix} \quad (5)$$

where $\Delta\theta$ denotes the rotation angle of the individual. The quantum rotating gate does not change the mode length of the quantum bit, but only changes the phase of the quantum bit, which is shown schematically in Fig. 1.

When the rotation angle is too large, the QGA is easy to fall into the local optimal solution, and the angle is too small will lead to too slow convergence and consume a lot of computational resources. Numerous improved versions of QGA use a fixed rotation angle as shown in Eq. (5), in order to minimize the impact of the rotation angle on the final computational precision, this paper proposes a dynamic rotation strategy, which is calculated as follows:

$$\Delta\theta_i^t = \frac{\Delta\theta_i^{t+c} - \Delta\theta_i^{t-c}}{2c} + \theta_{max} - \frac{t}{T}(\theta_{max} - \theta_{min}) \quad (6)$$

where c denotes the sequence scale from the current state, t is the current iteration number, T is the maximum iteration number, θ_{max} is the maximum rotation angle, and θ_{min} is the minimum rotation angle.

According to Eq. (6), it can be seen that the size of the rotation angle is determined by the number of iterations and the amount of accumulated angle change. From the first half of (6), it can be seen that the rotation angle is related to the difference of the previous rotation angle, which enables the rotation angle to be associated with the past and future states, and a larger rotation angle is used in the pre-evolutionary stage to expand the search range, and a smaller rotation angle is used in the later stage of the evolution for accurate search. DRQGA is able to change the angle dynamically according to the characteristics of the problem and the performance of the individual constantly, correlating the previous step and future state information, thus adjusting the current angle. The dynamic rotation angles for 2D coding are shown in Table II. The flowchart of the DRQGA algorithm is shown in Fig. 2.

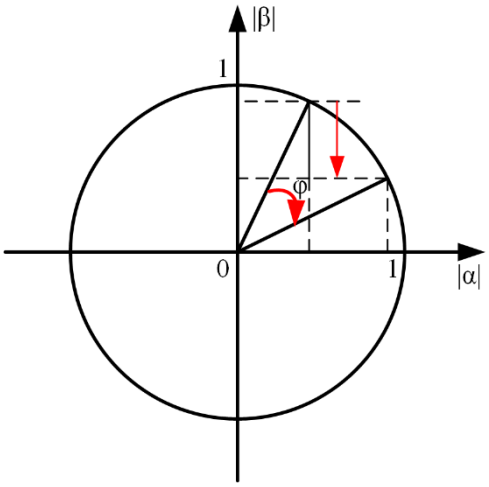


Fig. 1. Schematic diagram of a quantum revolving door.

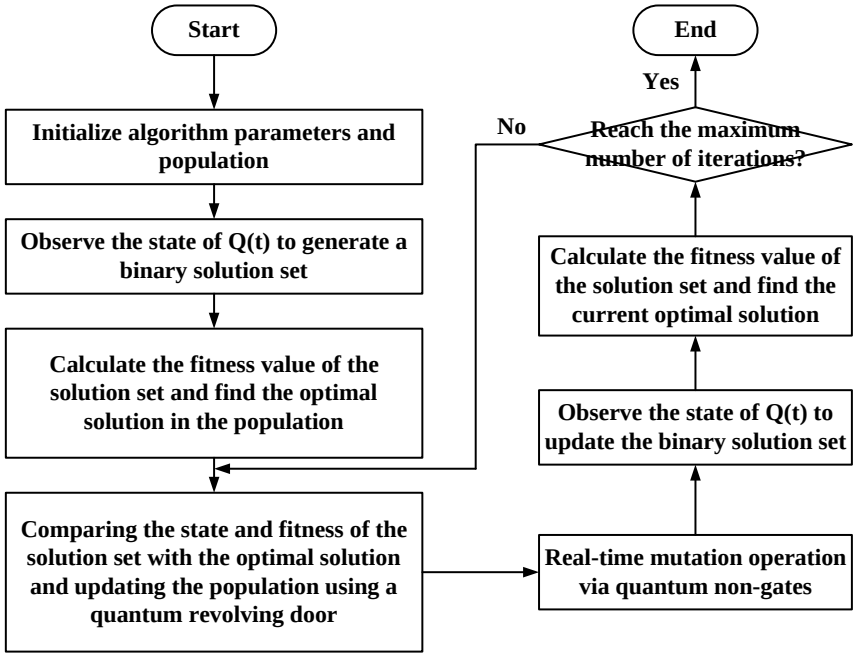


Fig. 2. Flowchart of DRQGA algorithm.

TABLE II. DYNAMIC ROTATION ANGLE LOOKUP TABLE FOR 2D CODING

x_i	x_i^{best}	$f(X) \geq f(X^{best})$	$\Delta\theta_i$	Rotary Angle Symbol			
				$\alpha_i\beta_i > 0$	$\alpha_i\beta_i < 0$	$\alpha_i = 0$	$\beta_i = 0$
0	0	false	$\Delta\theta_i^t$	0	0	0	0
0	0	ture	$\Delta\theta_i^t$	0	0	0	0
0	1	false	$\Delta\theta_i^t$	0	0	0	0
0	1	ture	$\Delta\theta_i^t$	-1	+1	±1	0
1	0	false	$\Delta\theta_i^t$	-1	+1	±1	0
1	0	ture	$\Delta\theta_i^t$	-1	+1	0	±1
1	1	false	$\Delta\theta_i^t$	-1	+1	0	±1
1	1	ture	$\Delta\theta_i^t$	-1	+1	0	±1

V. CORPORATE CREDIT EVALUATION MODEL BASED ON DRQGA-RESGAT

A. Financial Enterprise Credit Evaluation Mapping Data

In the process of credit evaluation of financial companies, we observe a significant effect of the financial leverage ratio on the overall credit score due to interactions between different credit evaluation indicators. The strength of these interactions varies with the particular combination of indicators, as determined by ANOVA. Considering these variables and their potential confounders in the modeling process, we propose to utilize the correlation information between the indicators as inputs to the model. To this end, we introduce a new data structure to accurately capture the effects of indicator interactions and improve predictive performance. We adopt the graph data structure approach proposed by Kipf and Welling [33], which proves to be very beneficial in analyzing the credit evaluation process. The graph data structure is able to efficiently encode information in a structured and meaningful way, representing both the attributes of each metric and the relationships between them. Fig. 3 illustrates an attribute-relationship graph describing the interrelationships of some of the credit evaluation metrics, where changes in the metric interactions can be interpreted as changes in the relationships between the nodes in the graph, utilizing $x_i, i = 1,2,\cdots,25$ to denote the second-level metrics we have defined and e_{ij} to denote the edges that are related, where $i, j = 1,2,\cdots,25, i \neq j$.

B. ResGAT Evaluation Model

In the process of credit evaluation of financial enterprises, GAT (Graph Attention Network) processes graph data, extracts attributes and relational features of nodes through a hidden masked self-attention layer, and uses an attention mechanism to process input variables, focusing on the most relevant parts. Credit evaluation metrics graph data can be processed through the graph attention network architecture to obtain higher-level feature representations. To cope with the overfitting problem under small sample conditions, we introduce a residual network architecture in the computation of the attention coefficients and the forward propagation of the model, which enhances the attention mechanism by introducing jump connections. As shown in Fig. 4, the ResGAT network consists of a residual feature extraction module and a graph attention module. The basic attribute features of credit evaluation metrics are represented by layers with residual connections, while the attention module provides learning focus based on the interaction information between metrics so that nodes can pay attention to the features in their neighborhood.

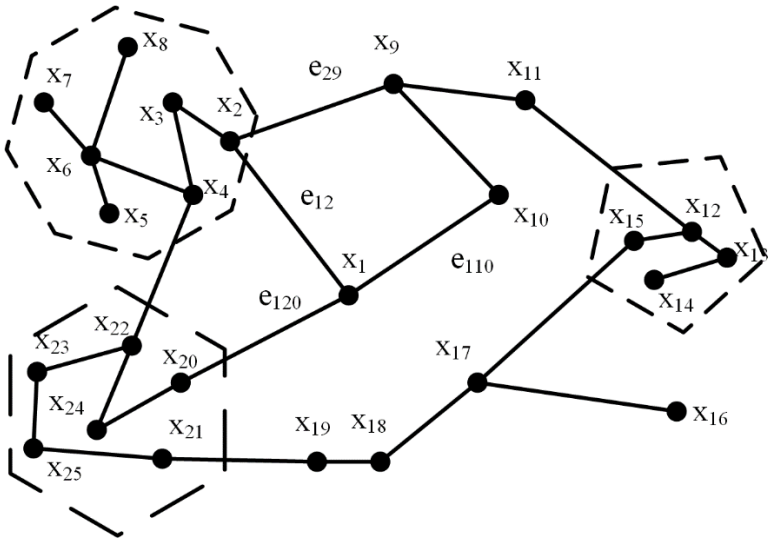


Fig. 3. Schematic diagram of indicator mapping.

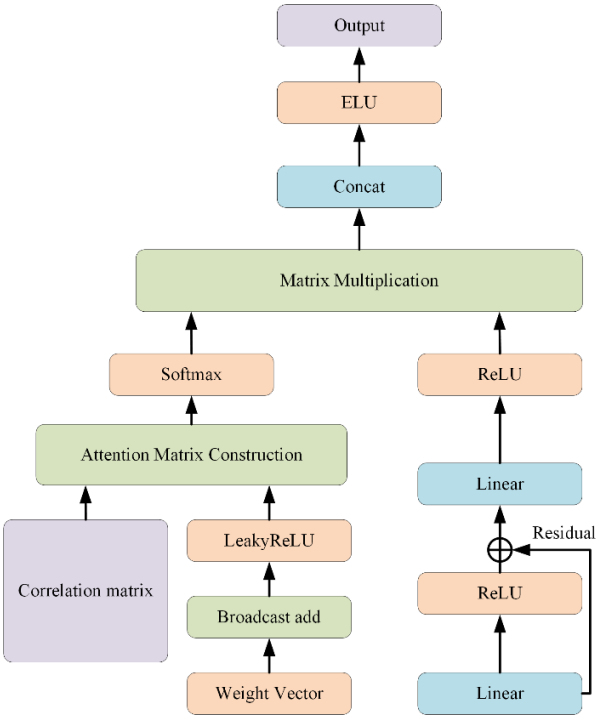


Fig. 4. Structure of the ResGAT network.

The input of the model is the evaluation index data after our normalization, and in the attention module, the attention coefficient e is calculated using the set of parameters to control the weight of the model, as shown in Eq. (7).

$$e = LeakyReLU((x \times a) + (x \times a)^T) \tag{7}$$

where x is the indicator parameter input, a is the trainable attention weights, and $LeakyReLU$ is the activation function as shown in Eq. (8). The plus sign in $(x \times a) + (x \times a)^T$ denotes the broadcast addition of the matrix.

$$y = max(0, x) + leak \cdot min(0, x) \tag{8}$$

Further, we use the adjacency matrix to represent the interaction between the parameters. Element A_{ij} indicates the existence of a significant interaction. When $A_{ij} = 1$, the interaction of indicator parameters X_i and X_j has a greater effect on the counterweight. An attention matrix Att is constructed using the attention coefficients e and 0 as shown in Eq. (10) and Eq. (11).

$$\hat{A} = \begin{bmatrix} 1 & A_{12} & A_{13} & \cdots & A_{1n} \\ A_{21} & 1 & A_{23} & \cdots & A_{2n} \\ A_{31} & A_{32} & 1 & \cdots & A_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & A_{n3} & \cdots & 1 \end{bmatrix} \tag{9}$$

$$\varepsilon_{ij} = \begin{cases} e, a_{ij} = 1 \\ 0, a_{ij} = 0 \end{cases} \tag{10}$$

$$Att_{ij} = Softmax(\varepsilon_{ij}) = \frac{exp(\varepsilon_{ij})}{\sum_{k \in N_i} exp(e_{ik})} \tag{11}$$

where ε_{ij} is an element of the adjacency matrix A, Att_{ij} is an element of the attention matrix Att, and N_i is the neighborhood quantity of node i in the graph. In order to avoid falling into degradation during the training of the network model, a residual connection is introduced in the propagation part of the right side, where the attention matrix of the left branch is multiplied with the result of the right branch after a nonlinear change, and w_1 and w_2 are assigned as weights thus obtaining the final result.

$$h = Att \times ReLU(x + ReLU(x \times w_1)) \times w_2 \tag{12}$$

C. DRQGA-ResGAT

In the credit evaluation process of financial enterprises, quantum genetic algorithm can significantly improve and enhance the performance of ResGAT network. The global optimization of ResGAT network using quantum genetic algorithm is as follows: the initial weights and attention coefficients of ResGAT network are optimized using quantum genetic algorithm. Then, the gradient descent algorithm is applied to adjust the parameters of the ResGAT network according to the negative gradient direction to train the network. The overall algorithm structure is schematically shown in Fig. 5.

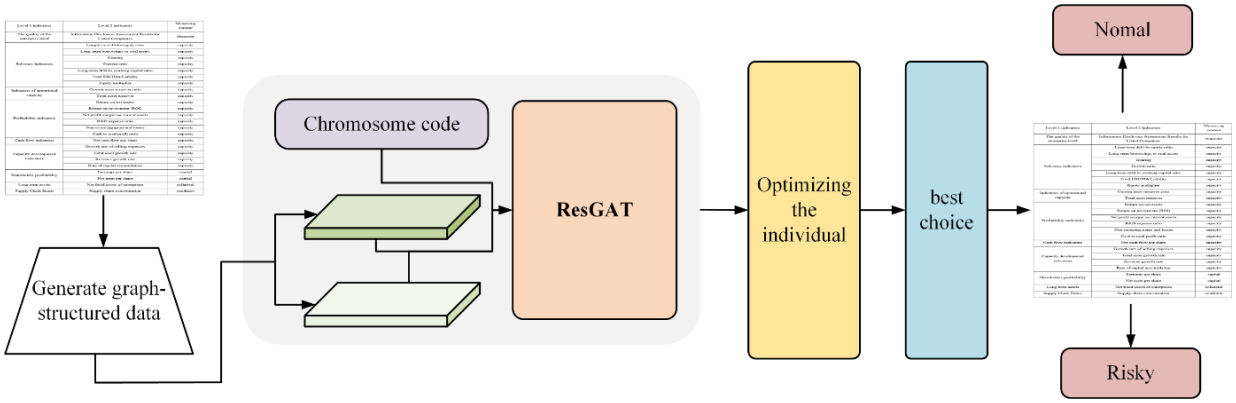


Fig. 5. The overall structure of the algorithm at a glance.

The main reasons for adopting this method to improve and enhance the ResGAT network are as follows:

- The quantum genetic algorithm utilizes quantum superposition and quantum gate operations to make the search space more extensive, effectively avoiding the shortcoming of the search range falling into a local minimum.
- The quantum genetic algorithm can converge quickly, thus reducing the number of training times of the weights and the attention coefficients.
- The parallel computing characteristics of the quantum genetic algorithm can significantly accelerate the convergence speed of the model.

The operation of the ResGAT network optimized by quantum genetic algorithm is divided into the following two steps:

1) Initialize the network parameters using quantum genetic algorithm: This step takes advantage of the superposition of quantum bits and the quantum revolving door operation to generate a variety of possible initial solutions, and obtain the optimal solution through quantum measurements, thus providing a better starting point for the network. Specifically, quantum genetic algorithms are used to determine the initial attention coefficients and weights of the ResGAT network, giving it a better performance from the start [34].

2) Chromosome coding: Since the quantum bit and quantum gate operations of the quantum genetic algorithm are able to represent continuous parameters, and the coding method of quantum states is characterized by continuous parameter optimization, the traditional steps of coding and decoding are omitted. To a certain extent, the computing speed and the precision of feasible solutions of quantum genetic algorithm are improved. Therefore, in this paper, the quantum state coding method is chosen to encode the chromosome. The quantum states are cascaded to the hidden layer nodes according to the input layer nodes and then cascaded from the hidden layer nodes to the output layer nodes. It should be noted that the chromosomes in the population are represented as a cascaded output array of quantum states.

3) Fitness function: The value of the fitness function determines the result of the quantum genetic algorithm's evaluation of the chromosome's survivability. The larger the value of the fitness function of a chromosome, the more likely it is to be selected for genetic manipulation and the smaller the sum of squares of the errors between the actual output value and the desired output value of the ResGAT network. The results show that the optimized ResGAT network has higher precision. In this paper, the logarithmic loss function is chosen as the adaptation evaluation function as shown in Eq. (13) to measure the performance of ResGAT network.

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i log(p_i) + (1-y_i) log(1-p_i)] \quad (13)$$

VI. Model Experiment

A. Dataset

Dataset In this paper, 253 SMEs were selected from the CSMAR database in 2020 for the empirical study, and after the data preprocessing ticks, except for the enterprises with serious missing information, the missing information of the remaining enterprises was replaced by the mean value. Finally, 210 SMEs were selected. Among them, there are 137 normal enterprises and 73 risky enterprises.

B. Results

In this paper, precision rate and recall rate are selected as the key evaluation indexes, and six hyperparameters such as learning rate, batch size, number of hidden layers, number of hidden units per layer, weight initialization method, Dropout rate are optimized, and the precision rate is used as the objective function of the quantum genetic algorithm, and the GPU used in the training is the NVIDIA GeForce RTX 2080 Ti. The number of iterations of the quantum genetic algorithm $T=100$, the population size $Pop=100$, the number of quantum bits $nq=10$, and the mutation probability $mr=0.01$. The quantum genetic algorithm is compared with the classical simulated annealing algorithm, genetic algorithm, particle swarm algorithm, and whale optimization algorithm, and the algorithm comparison results obtained are shown in Fig. 6. A comparison of the results of DRQGA-ResGAT with other algorithms is shown in Table III.

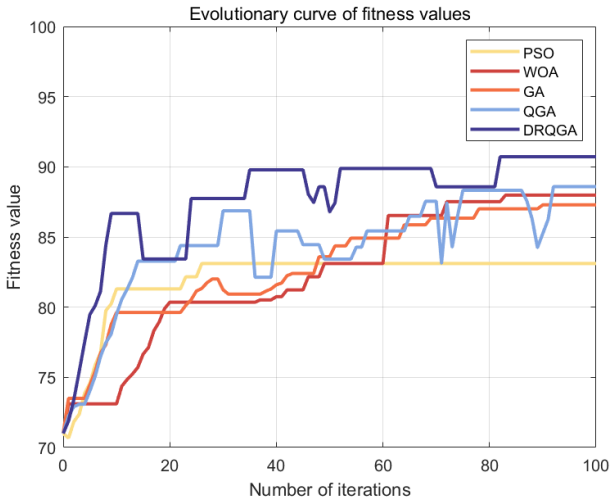


Fig. 6. Comparison of results of the optimization algorithm.

TABLE III. Performance of ResGAT Network Under Different Optimization Algorithms

Model	Precision	F1 score	AUC
PSO-ResGAT	0.8311	0.7939	0.9121
WOA-ResGAT	0.8798	0.8305	0.9476
GA-ResGAT	0.8728	0.7877	0.9298
QGA-ResGAT	0.8858	0.8439	0.9213
DRQGA-ResGAT	0.9071	0.8602	0.9647

The precision iteration curve of DRQGA-ResGAT under optimal hyperparameters is shown in Fig. 7. The training loss iteration curve is shown in Fig. 8. The validation loss iteration curve is shown in Fig. 9.

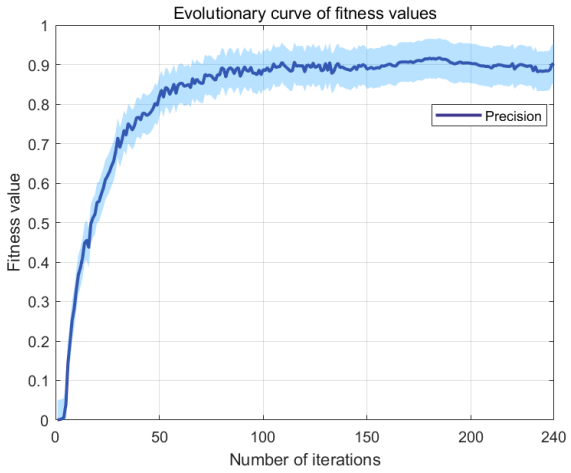


Fig. 7. Evolutionary curve of fitness values.

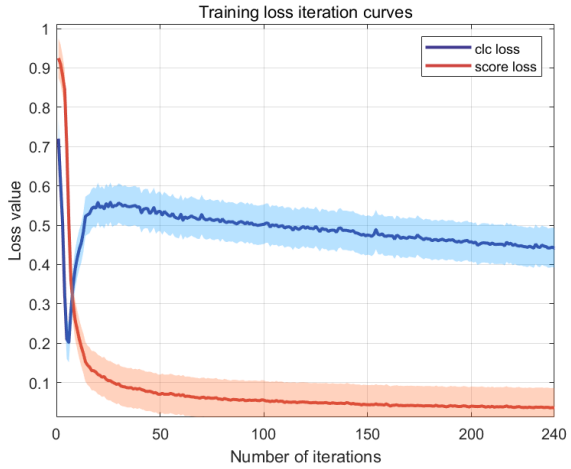


Fig. 8. Iteration curve of loss function for training set.

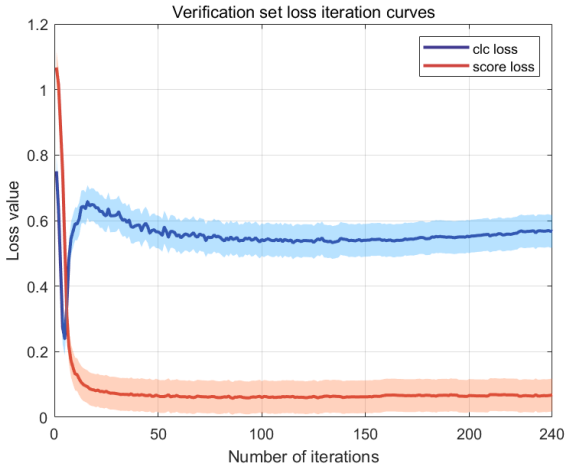


Fig. 9. Verification set loss function iteration curve.

TABLE IV. RESULTS OF ABLATION EXPERIMENTS

Model	Precision	F1 score	AUC
GAT	0.7310	0.7386	0.8431
ResGAT	0.7749	0.7795	0.8758
QGA-GAT	0.7705	0.7695	0.8812
QGA-ResGAT	0.8728	0.7877	0.9421
DRQGA-GAT	0.8001	0.7619	0.9257
DRQGA-ResGAT	0.9071	0.8602	0.9647

In order to explore the contribution of each improvement component, we conducted ablation experiments and the comparison results of the six different algorithms obtained are as follows. As can be seen in Table IV, DRQGA-ResGAT improves the precision by 24.09% compared to the most basic GAT, and the F1 score improves from 0.7386 to 0.8602. Compared to ResGAT, it improves by 17.06%. 3.92% compared to QGA-ResGAT and 13.37% compared to DRQGA-GAT. It can be seen that the precision of the models optimized with hyperparameters using DRQGA for GAT or ResGAT is significantly improved. Meanwhile, based on the GAT model, the ResGAT model with the introduction of the residual mechanism is able to achieve higher precision and more performance improvement with DRQGA optimization.

VII. DISCUSSION

Compared with the classical ResGAT model, DRQGA-ResGAT improves the evaluation precision by 17.06%, which is 3.92% higher compared with QGA-ResGAT. Compared with the classical optimization algorithm, the dynamic revolving door-based quantum genetic algorithm is able to explore the global optimal solution on a larger scale [35], and its ability to cope with mutant data makes the algorithm less susceptible to be interfered by special data. interference, thus ensuring the overall individual quality [36]. In future research, we will consider introducing the optimization strategy of the population intelligence optimization algorithm into the dynamic internal adjustment of the network model [37], and further exploring the combination of the optimization algorithm and deep learning network on the basis of hyperparameter tuning. In addition, on top of the existing datasets, larger datasets should be considered to enhance the scope of enterprises into the risk evaluation of enterprises in various industries.

VIII. CONCLUSION

Under the digital transformation of enterprises, in order to be able to evaluate the credit of financial enterprises more comprehensively and objectively, this paper adopts a quantum genetic algorithm based on dynamic revolving door for hyper-parameter tuning of residual map attention network. On this basis, a financial enterprise credit evaluation model is established. Specifically, the introduction of quantum genetic algorithm effectively enhances the optimization ability of the model and makes it perform well in the hyperparameter optimization search process. In addition, the excellent performance of the residual graph attention network in dealing with graph-structured data enables the model to better capture the complex associations among enterprises, thus enhancing

the precision of credit evaluation. The experimental results show that the DRQGA-ResGAT algorithm outperforms other algorithms in key metrics such as precision rate, F1 value and AUC value. Especially when compared with the classical simulated annealing algorithm, genetic algorithm, particle swarm algorithm and whale optimization algorithm, DRQGA-ResGAT shows obvious advantages, which verifies its potential for application in the credit evaluation of financial enterprises.

REFERENCES

[1] J. Zare, and A. Persaud, "Digital transformation and business model innovation: a bibliometric analysis of existing research and future perspectives," Manag Rev Q, April 2024.

[2] Guo, H., Yang, Z., Huang, R. et al, "The digitalization and public crisis responses of small and medium enterprises: Implications from a COVID-19 survey," Front. Bus. Res. China, vol. 14, pp. 19, September 2020.

[3] J. N. Wang, S. Zhang, Y. H. Xiao and R. Song, "A Review on Graph Neural Network Methods in Financial Applications," preprint arXiv, April 2022.

[4] Y. F. Zhang, "Big Data Application in Forecasting Financial Investment of e-Commerce Industry for Sustainability" International Journal of Advanced Computer Science and Applications(IJACSA), vol. 13, December 2022.

[5] W. L. Chen, "The GSO-Deep Learning-based Financial Risk Management System for Rural Economic Development Organization," International Journal of Advanced Computer Science and Applications(IJACSA), vol. 14, November 2023.

[6] I. Sutskever, O. Vinyals, QV. Le, "Sequence to Sequence Learning with Neural Networks," preprint arXiv, December 2014.

[7] E. Solís, S. Noboa and E. Cuenca, "Financial Time Series Forecasting Applying Deep Learning Algorithms," I Information and Communication Technologies, pp. 46–60, November 2021.

[8] D. W. Cheng, F. Z. Yang, S. Xiang and J. Liu, "Financial time series forecasting with multi-modality graph neural network," Pattern Recognition, Volume 121, January 2022.

[9] Y. Shi, Yi Qu, Z.S. Chen, Y. L. Mi and Y. N. Wang, "Improved credit risk prediction based on an integrated graph representation learning approach with graph transformation," European Journal of Operational Research, vol. 315, pp. 786–801, June 2024.

[10] L. Zhang, "The Evaluation on the Credit Risk of Enterprises with the CNN-LSTM-ATT Model," Computational Intelligence and Neuroscience, September 2022.

[11] Y. Gao, H. L. Xiao, C. J. Zhan, L. R. Liang, W. T. Cai and X. P. Hu, "CATE: Contrastive augmentation and tree-enhanced embedding for credit scoring," Information Sciences, vol. 651, December 2023.

[12] W. Zhang, S. S. Yan, J. Li, X. Tian and T. Yoshida, "Credit risk prediction of SMEs in supply chain finance by fusing demographic and behavioral data," Transportation Research Part E: Logistics and Transportation Review, vol. 158, February 2022.

[13] Y. X. Sang and Y. Bao, "DialogueGAT: A Graph Attention Network for Financial Risk Prediction by Modeling the Dialogues in Earnings Conference Calls," Association for Computational Linguistics, December, 2022.

[14] X. J. Huang, Z. B. Wu, G. S. Wang, Z. P. Li, Y. S. Luo and X. F. Wu, "ResGAT: an improved graph neural network based on multi-head attention mechanism and residual network for paper classification," Scientometrics, vol.129, pp. 1015–1036, January 2024.

[15] L. Y. Chang and P. Branco, "Embedding residuals in graph-based solutions: the E-ResSAGE and E-ResGAT algorithms. A case study in intrusion detection," Applied Intelligence, vol. 54, pp. 6025–6040, May 2024.

[16] W. Y. Fan and S. S. Zhong, "Gyroscope Dynamic Balance Counterweight Prediction Based on Multi-Head ResGAT Networks," Computer Modeling in Engineering & Sciences, vol. 139, pp. 2525-2555, March 2024.

[17] B. B. Zhou, J. Y. Jin, H. Zhou, X. Y. Zhou, L. X. Shi, J. H. Ma and Z. W. Zheng, "Forecasting credit default risk with graph attention networks," Electronic Commerce Research and Applications, vol. 62, December 2023.

[18] L. Y. Song, H. D. Li, Y. C. Tan, Z. H. Li and X. Q. Shang, "Enhancing Enterprise Credit Risk Assessment with Cascaded Multi-level Graph Representation Learning," Neural Networks, vol. 169, pp. 475-484, January 2024.

[19] N. Bacanin, T. Bezdan, M. Zivkovic and A. Chhabra, "Weight Optimization in Artificial Neural Network Training by Improved Monarch Butterfly Algorithm," Mobile Computing and Sustainable Informatics, pp. 397–409, July 2021.

[20] S. Milosevic, T. Bezdan, M. Zivkovic, N. Bacanin, I. Strumberger and M. Tuba, "Feed-forward Neural Network Training by Hybrid Bat Algorithm," Modelling and Development of Intelligent Systems, pp. 52-66, February 2021.

[21] N. Bacanin, T. Bezdan, E. Tuba, I. Strumberger, M. Tuba, "Optimizing Convolutional Neural Network Hyperparameters by Enhanced Swarm Intelligence Metaheuristics," Algorithms, vol. 13, pp. 67, March 2020.

[22] X. Duan, "Introducing an Innovative Approach to Mitigate Investment Risk in Financial Markets: A Case Study of Nikkei 225" International Journal of Advanced Computer Science and Applications (IJACSA), vol. 15(3), April 2024.

[23] R. Mitra, A. Dongre, P. Dangare, A. Goswami, and M. K. Tiwari, "Knowledge graph driven credit risk assessment for micro, small and medium-sized enterprises," International Journal of Production Research, vol. 62, no. 12, pp. 4273–4289, September 2023.

[24] S. Zandi, K. Korangi,M. Óskarsdóttir,C. Mues and C. Bravo, "Attention-based dynamic multilayer graph neural networks for loan default prediction", European Journal of Operational Research, vol. 321, no. 2, pp. 586-599, March 2025.

[25] S. Zandi, K. Korangi, M. Óskarsdóttir, C. Mues and C. Bravo, "Attention-based Dynamic Multilayer Graph Neural Networks for Loan Default Prediction," preprint arXiv, June 2024.

[26] TH. Nguyen-Vo, T. T. T. Do and B. P. Nguyen, "ResGAT: Residual Graph Attention Networks for molecular property prediction," Memetic Computing, vol. 16, pp. 491-503, September 2024.

[27] Mingwei Tang , Wei Tang , Qingchi Gui , Jie Hu , Mingfeng Zhao, "A vulnerability detection algorithm based on residual graph attention networks for source code imbalance (RGAN)," Expert Systems with Applications, vol. 238, 122216, March 2024.

[28] K. Xu, Y. Zhao, L. Zhang, C. Gao and H. Huang, "Spectral–Spatial Residual Graph Attention Network for Hyperspectral Image Classification," IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, September 2021,

[29] J. Wang, H. H. Shuai, Y. H. Li, and W. H. Cheng, "Language-guided Residual Graph Attention Network and Data Augmentation for Visual Grounding," ACM Transactions on Multimedia Computing, Communications, and Applications, vol. 20, no. 1, pp. 1-23, August 2023.

[30] O. Y. Bakhteev and V. V. Strijov, "Comprehensive analysis of gradient-based hyperparameter optimization algorithms," Annals of Operations Research, vol. 289, 51–65, .June 2019.

[31] Yingfang Yuan, Wenjun Wang, George M. Coghill, Wei Pang, "A Novel Genetic Algorithm with Hierarchical Evaluation Strategy for Hyperparameter Optimisation of Graph Neural Networks," preprint arXiv, January 2021.

[32] L. Japa, M. Serqueira, I. MendonçA, M. Aritsugi, E. Bezerra and P. H. González, "A Population-Based Hybrid Approach for Hyperparameter Optimization of Neural Networks," IEEE Access, vol. 11, pp. 50752-50768, May 2023.

[33] Kipf. T. N, Welling. M, "Semi-Supervised Classification with Graph Convolutional Networks," preprint arXiv, February 2017.

[34] S. Ma, H. Liu, N. Pan, S. Wang, "Study on an autonomous distribution system for smart parks based on parallel system theory against the background of Industry 5.0," Journal of King Saud University-Computer and Information Sciences, vol. 35, no. 7, pp. 101608, June 2023.

[35] F. M. Creevey, C. D. Hill and L. C. L. Hollenberg, "GASP: a genetic algorithm for state preparation on quantum computers," Scientific Reports, vol. 13, 11956, July 2023.

[36] G. Acampora, A. Chiatto and A. Vitiello, "Genetic algorithms as classical optimizer for the Quantum Approximate Optimization Algorithm," Applied Soft Computing, vol. 142, 110296, July 2023.

[37] C. Q. Gong, H. S. Zhu, A. Gani and H. Qi, "QGA–QGCNN: a model of quantum gate circuit neural network optimized by quantum genetic algorithm," The Journal of Supercomputing, vol. 79, pp. 13421–13441, March 2023.

A Conceptual Framework for Agricultural Water Management Through Smart Irrigation

Abdelouahed Tricha, Laila Moussaid, Najat Abdeljebbar
National High School for Electricity and Mechanics (ENSEM), Hassan II University of Casablanca, Casablanca, Morocco

***Abstract*—The demand for freshwater resources has risen significantly due to population growth and increasing drought conditions in agricultural regions worldwide. Irrigated agriculture consumes a substantial amount of water, often leading to wastage due to inefficient irrigation practices. Recent breakthroughs in emerging technologies, including machine learning, the Internet of Things, wireless communication, and advanced monitoring systems, have facilitated the development of smart irrigation solutions that optimize water usage, enhance efficiency, and reduce operational costs. This paper explores the critical parameters and monitoring strategies for smart irrigation systems, emphasizing soil and water management. It also presents a conceptual framework for implementing sustainable irrigation practices aimed at optimizing water use, improving crop productivity, and ensuring cost-effective management across different agricultural settings.**

***Keywords*—Agriculture; irrigation system; water management; Internet of Things; sustainability**

I. Introduction

The growing global population demands a continuous supply of high-quality food, placing increasing pressure on agriculture to boost productivity [1]. Achieving this goal requires effective water management, as irrigation plays a critical role in crop yield. However, managing complex agricultural ecosystems demands constant monitoring and control of multiple factors, such as soil moisture, weather conditions, and crop health. To address these challenges, information and communication technologies (ICT) offer valuable solutions by enabling the efficient collection, transmission, and analysis of diverse datasets. ICT facilitates decision-making and policy development for agricultural planning and improves water management through remote monitoring and control of irrigation systems. With real-time feedback, farmers can optimize irrigation schedules, reducing water waste and enhancing crop performance. Additionally, ICT promotes the integration and interoperability of different irrigation technologies, including soil moisture sensors, water balance models, plant signals, remote sensing tools, and GPS mapping, leading to more sustainable and efficient farming practices.

Irrigation is one of the most vital services in the agricultural industry, particularly in regions with low rainfall. Meeting the water demands of crops is crucial for crop quality and production. However, inadequate irrigation can lead to a decrease in crop quality and production [1], [2]. Diverse methods and technologies are used to effectively manage water resources to address this issue. These include water balance, plant signals, or soil moisture sensors, which can be technically backed by cloud computing and the Internet of Things (IoT). IoT uses low-power network connectivity to link physical items to the internet. One application of IoT for irrigation scheduling includes wireless sensor networks (WSN), which connect soil moisture sensors to the web [3], [4], [5], [6], [7]. This allows farmers to remotely monitor and control their irrigation systems, leading to more efficient water usage and potentially higher crop yields. Additionally, the data collected from these sensors can be analyzed to provide insights into soil health and crop growth patterns, aiding in decision-making for future planting seasons.

This paper contributes to the field of smart irrigation by addressing critical gaps in the literature and proposing a systematic approach to enhance water management and sustainability in agriculture. By focusing on both the technical and practical challenges, the paper offers an integrated framework for developing more efficient irrigation systems. The paper is structured as follows: Section I introduces the importance of irrigation and water management in agriculture, while Section II presents a comprehensive survey of the literature on current smart irrigation technologies. In Section III, we explore critical parameters for smart irrigation systems, including soil and water management. Section IV reviews monitoring strategies, focusing on soil-based and weather-based approaches. In Section V, we propose a conceptual framework for implementing a smart irrigation system, including system architecture and workflow. Finally, Section VI concludes the paper by summarizing the key insights and potential directions for future research.

II. Related Works

In recent years, diverse technologies such as the IoT, machine learning (ML), and big data have significantly impacted irrigation practices. This section delves into various studies and works that involve the application of these emerging technologies in irrigation system management and control. The focus is on the primary technologies and methods used in smart irrigation systems: IoT, data-driven approaches, and advanced technologies. Each subsection summarizes key findings, advantages, and drawbacks of existing studies, while identifying research gaps and opportunities for future work.

A. Internet of Things

The IoT is a network of physical objects connected to the internet, capable of communication, data collection, and data exchange. IoT has been widely applied in irrigation, enabling remote monitoring and control of systems, and the collection

and analysis of diverse datasets related to soil, weather, and crop conditions.

Several studies have explored IoT-based irrigation systems for precision agriculture using various techniques and platforms. One such system developed an IoT-based irrigation system via the ARETHOU5A platform, incorporating wireless sensor nodes powered by a rectenna, which recovers radio frequency energy to power the nodes [8]. Another notable work proposed a highly accurate irrigation system utilizing IoT, fuzzy logic, and GSM [9]. This system employs a fuzzy logic controller to compute input variables such as humidity, temperature, and soil moisture to generate motor status, and includes an automatic shut-off mechanism during precipitation to conserve resources. Additionally, an intelligent multi-agent IoT precision irrigation strategy has been presented, leveraging technologies like the MQTT protocol and IoT for real-time monitoring and analysis of crop water requirements [10].

While these studies highlight the potential of IoT to enhance smart irrigation systems, they also underscore several challenges. Data quality and security of IoT devices and networks are major concerns. IoT devices, often exposed to harsh environmental conditions, can face reliability and accuracy issues. Furthermore, these devices and networks are vulnerable to cyberattacks, potentially compromising the integrity and availability of irrigation systems. Ensuring the robustness and resilience of IoT devices and implementing stringent data quality and security measures is crucial. Another challenge is the scalability and interoperability of IoT-based systems. Integrating and coordinating a growing number and diversity of IoT devices and platforms can be difficult, necessitating the development and adoption of common standards, protocols, and architectures to ensure seamless integration.

B. Data-Driven Approaches

Data-driven approaches utilize data to generate insights, predictions, or decisions, playing a significant role in optimizing irrigation systems by analyzing water usage based on various data sources and techniques. For instance, integrating Geographic Information System (GIS) technology with irrigation systems has improved water use efficiency [11]. A system leveraging cloud-based near real-time data storage and processing enables the adaptation of the Penman-Monteith evapotranspiration coefficient to local weather conditions, incorporating data collection and soil moisture measurement capabilities. Another study focused on optimizing water usage in an IoT system by moving data computation to the edge, allowing for real-time data processing and quick decision-making despite limited internet connectivity [12]. This system estimates water requirements by calculating evapotranspiration potential and employs AI algorithms for detecting plant diseases and pests. Furthermore, a non-contact vision system using a standard video camera and a neural network has been proposed to predict the irrigation needs of loamy soils based on soil color differences [13].

These studies illustrate the benefits of data-driven approaches in optimizing water usage and improving crop yield. However, challenges such as data availability and accessibility persist. Large and diverse data sets are essential for accurate results, but data may be scarce due to a lack of sources, collection devices, or sharing platforms. Incomplete, inconsistent, or noisy data can further affect result quality. Ensuring data availability, accessibility, and implementing proper data cleaning, preprocessing, and validation techniques is vital. Additionally, the complexity and interpretability of data-driven methods can pose difficulties, especially for non-experts. Simplifying and explaining these methods and results, along with providing user-friendly interfaces and feedback mechanisms, is crucial.

C. Advanced Technologies

Advanced technologies offer innovative solutions for enhancing smart irrigation systems, providing automation, intelligence, and security. An automatic irrigation system measuring soil humidity and temperature with sensors was designed to prevent irrigation when humidity is high, conserving water resources, and to permit irrigation when humidity is low [14]. This system employs the Decision Tree algorithm to predict and analyze data. Additionally, WSN have been leveraged to build comprehensive irrigation systems, covering different farm regions with various sensor modules transmitting data to a shared server [15]. ML algorithms support crop and weather-based irrigation model predictions. Addressing security concerns, lightweight cryptography techniques in IoT, such as the Expeditious Cipher (X-cipher) for secure channels in the MQTT protocol, have been proposed to protect data integrity and confidentiality [16].

While these studies highlight the benefits of advanced technologies in irrigation, challenges such as cost, maintenance, compatibility, and integration remain. Advanced technologies can be expensive and complex to acquire, install, and operate, especially for small-scale farmers. Frequent maintenance and updates increase operational costs. Ensuring the affordability, simplicity, and providing adequate training and support for users is essential. Additionally, integrating advanced technologies with existing systems and practices can be difficult, necessitating efforts to ensure compatibility and seamless integration.

III. Critical Parameters for Smart Irrigation

Several parameters are used to assess the efficacy of the smart irrigation system. Both soil and water management are incorporated into an intelligent irrigation system. Soil management incorporates numerous parameters, including soil moisture, temperature, conditions, salinity, and so on. Water management parameters consist of dewpoint temperature, evapotranspiration, air temperature, atmospheric temperature, and relative humidity. In addition to these variables, the smart irrigation system must take forecast accuracy, data transmission rate, and actual usage into account [17].

A. Soil Management

Historically, soil supervision was once among the most difficult agricultural tasks for both businesses and cultivators. Several environmental issues that affect agricultural performance are revealed by soil analysis. If these types of obstacles are precisely characterized, agricultural models and methods can be easily understood. Soil management involves

various parameters, such as soil moisture, soil temperature, soil conditions, soil salinity, and so on.

Soil moisture is one of the most important parameters for irrigation scheduling, as it indicates the water content and availability in the soil. Soil temperature is another parameter that affects the crop growth and development, as it influences the germination, root growth, nutrient uptake, and microbial activity. Soil conditions refer to the physical, chemical, and biological properties of the soil, such as texture, structure, pH, organic matter, nutrients, and microorganisms. They affect the water retention, infiltration, drainage, and availability in the soil, as well as the crop health and quality. Soil salinity is the concentration of soluble salts in the soil, which can affect the osmotic potential, water uptake, and nutrient availability for the crops. Soil salinity can be caused by natural factors, such as weathering, evaporation, or seawater intrusion, or by human factors, such as irrigation, fertilization, or drainage. Moisture sensors can measure these parameters using different methods, such as capacitance, resistance, or gravimetric for the moisture, thermistors or thermocouples for the temperature, electrodes or spectrometers for soil conditions, and conductivity, reflectance, or fluorescence for soil salinity. These sensors can be integrated with IoT systems to transmit the soil parameters data to a cloud server, where it can be processed and analyzed to determine the optimal irrigation status, to support fertilization and irrigation decisions, and to prevent the effects of soil salinity on the crops. The results of a soil analysis survey are used to enhance crop production and provide producers with fertilization options [18]. Overfertilization and crop degradation can be avoided when using IoT techniques for identifying dirty soil.

Soil management preserves and improves productivity. It boosts agricultural productivity and quality, lowers input costs, and reduces pollution. For the crop to grow quickly and efficiently, the topsoil must be in good condition before planting. Even if each farm and crop have its own soil requirements, organic fertilization, soil testing, right tillage, chemical soil protection, etc. can encourage healthy soil biology [19].

B. Water Management

Determining the proper amount of water needed in greenhouses is a challenging task. Several IoT strategies are used to place and control smart sensors to prevent water waste. Greenhouses achieve water storage by using an automated drip irrigation system that is controlled by a soil moisture threshold. Utilizing various types of sensors, IoT technology can aid in water management by preventing water loss. The quantity of water in the tank is monitored by sensors, and the data is stored in the cloud via a mobile app, so farmers can track the water level with their smartphones. Using this method, the engine will run autonomously. The engine will automatically start if the water level declines below a predetermined level, and if the water level rises above a predetermined level, the engine will automatically shut off. IoT-powered intelligent irrigation systems can help farmers conserve water and improve crop quality by watering crops at the optimal time. Intelligent irrigation systems deploy temperature and soil sensors on farms, and these sensors transmit field data to producers via a gateway. Controllers for weather-based precision agriculture monitor and adjust irrigation schedules based on local weather data [19], [20], [21].

In addition to IoT, numerous other technologies can be used to manage irrigation water and to estimate the quantity of irrigation water required by adjusting the evapotranspiration coefficient calculated using the Penman-Monteith method [22].

IV. Irrigation Monitoring Strategies

Constructing an effective irrigation management system that increases food production while minimizing water loss necessitates a dependable monitoring system for the various variables influencing plant growth and development. To collect data that precisely reflects the current soil, plant, and weather conditions of plant irrigation zones, the IoT and WSN technologies are both used for monitoring in the field of precision irrigation [23], [24].

When developing a real-time monitoring system, sensors must be integrated with the IoT Framework or a wireless sensor communication network. Because of their capabilities for sensing, processing, and transmission, wireless networks are vital for real-time monitoring in smart irrigation. Geography or climate may affect monitoring in intelligent irrigation.

A. Soil-Based Monitoring

One of the most important factors in the growth of plants is soil moisture. Comprehensive geographical and temporal monitoring of the soil moisture content is required to ensure effective irrigation scheduling. Additionally, it is crucial to monitor soil moisture in the plant roots because it sheds light on moisture dynamics and the relationship between irrigation water volume and plant water uptake. There are various indirect techniques for determining soil moisture levels, including electromagnetism, thermal conductivity, neutron counting, water potential, electrical resistance, and direct measurements of soil moisture (gravimetric sample) [25].

Real-time soil moisture monitoring using sensors is a practical and accurate method for measuring moisture fluxes [26]. This approach involves accurately correlating the volumetric water content of the soil with the capacity of the inserted sensor probes. A highly accurate method for soil moisture measurement uses time-domain reflectometry (TDR) sensors [27] , which consist of two parallel rods inserted into the soil at a depth corresponding to the target moisture level. In another example, the IoT soil moisture surveillance solution proposed by [28] leveraged wireless networks, specifically GSM and infrared communication, to provide automatic irrigation and conserve water. The system used a capacitance-based soil moisture sensor with two electrodes to measure soil resistance, which was then processed by a PIC 16F877A microcontroller. Based on the detected soil moisture level, the microcontroller activated or deactivated the water pump through a relay driver.

B. Weather-Based Monitoring

Real-time estimation of baseline evapotranspiration using observed meteorological variables as an indicator of water loss from the plant and soil environment is becoming increasingly popular in weather-based monitoring. Solar radiation, ambient temperature, and wind velocity are the primary determinants of

the rate of water loss. These parameters can be determined using a weather station [23].

The FAO Penman-Monteith equation [29], provides a robust method for estimating daily or hourly evapotranspiration values from standard meteorological data, including atmospheric temperature, solar radiation, relative humidity, and wind speed. This method is widely recognized as a reliable approach for calculating reference evapotranspiration (ET_0), which serves as the baseline for determining the water needs of crops. The equation is expressed as in Eq. (1):

$$ET_0 = \frac{0{,}408\Delta(R_N - G) + \gamma\frac{900}{T+273}\mu_2(u_s - u_a)}{\Delta + \gamma(1 + 0{,}34\mu_2)} \quad (1)$$

Where ET_0 is the reference evapotranspiration (mm/day), R_N is the net radiation at the crop surface (MJ/m²/day), G is the soil heat flux density (MJ/m²/day), T is the mean daily air temperature (°C), μ_2 is the wind speed at 2 meters (m/s), u_s is the saturation vapor pressure deficit (kPa), u_a is the actual vapor pressure (kPa), Δ is the slope of the vapor pressure curve (kPa/°C), γ is the psychrometric constant (kPa/°C).

The daily crop evapotranspiration (ETc) is then calculated using Eq. (2):

$$ET_C = k_C * ET_0 \quad (2)$$

Hence: ET_C is the evapotranspiration of plants (mm / Day). K_C is the crop coefficient and ET_0 is the evapotranspiration reference (mm / Day).

This method allows for the precise calculation of water requirements tailored to specific crops, ensuring that irrigation is applied in accordance with actual crop water demand. Weather-based monitoring systems often rely on WSN to monitor environmental conditions accurately. These systems have been widely implemented as an efficient method for interconnecting numerous sensors over large agricultural areas. Closed-loop, real-time monitoring, and analysis of sensor data activate control devices based on a predetermined threshold value [30], [31].

These strategies are visually summarized in Fig. 1, highlighting the key parameters for both soil-based and weather-based monitoring approaches.

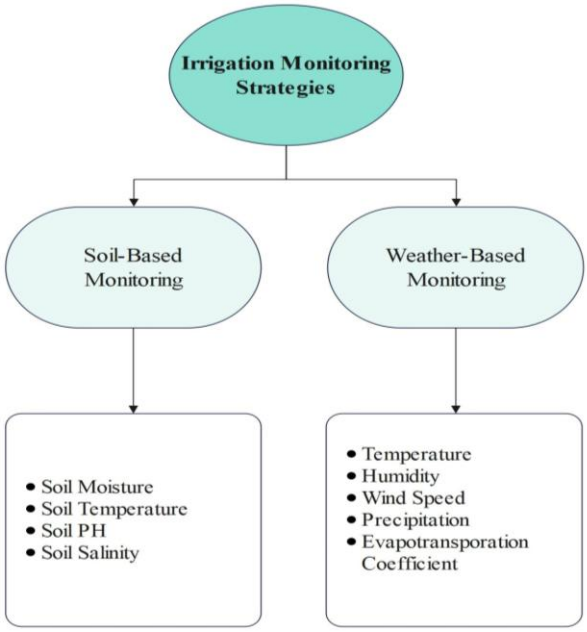


Fig. 1. Irrigation monitoring strategies adopted from study [25].

V. Proposed Framework for a Comprehensive Smart Irrigation System

A. System Overview

The proposed smart irrigation system is a sophisticated framework engineered to optimize water resource management in agricultural settings. It integrates advanced sensing technologies, data analytics, and automated control mechanisms to ensure precise, efficient, and sustainable irrigation practices. The system's core functionalities include real-time data acquisition, processing, analysis, and intelligent decision-making that leads to highly accurate control of irrigation processes.

As illustrated in Fig. 2, the system's architecture comprises multiple interconnected components, each playing a pivotal role in the overall management process. The system operates hierarchically, with data flowing from field-level sensors to the cloud, where comprehensive analysis and remote monitoring are conducted.

B. Fundamental Layers of the System

The proposed smart irrigation system is structured into distinct components that work collaboratively to ensure efficient and precise water management. These components are organized into layers, each with a specific function, enhancing the system's scalability, reliability, and ease of operation. Fig. 3 provides an overview of these layers.

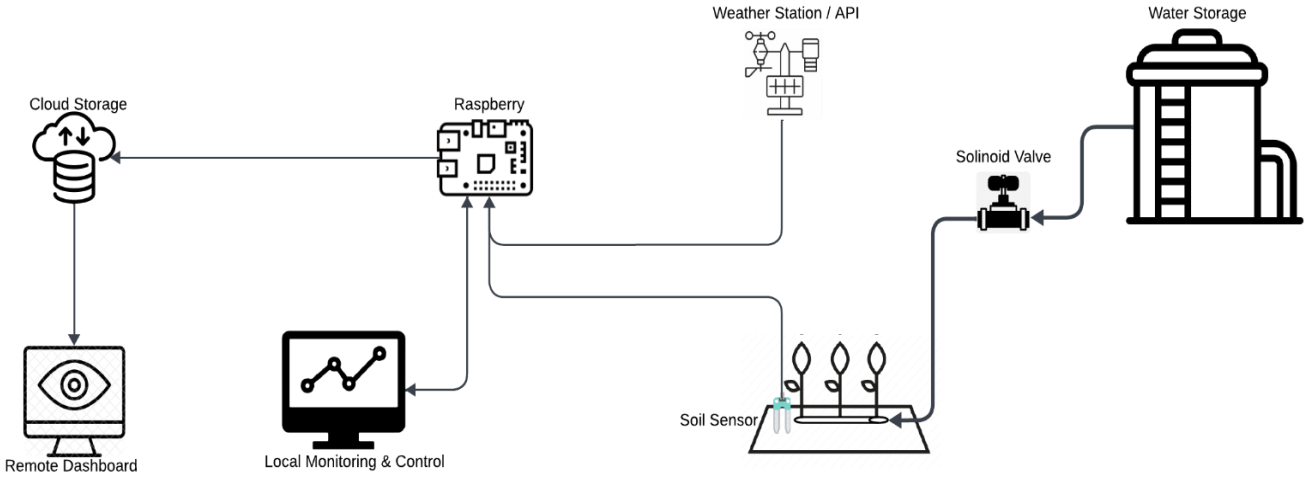


Fig. 2. Proposed system architecture.

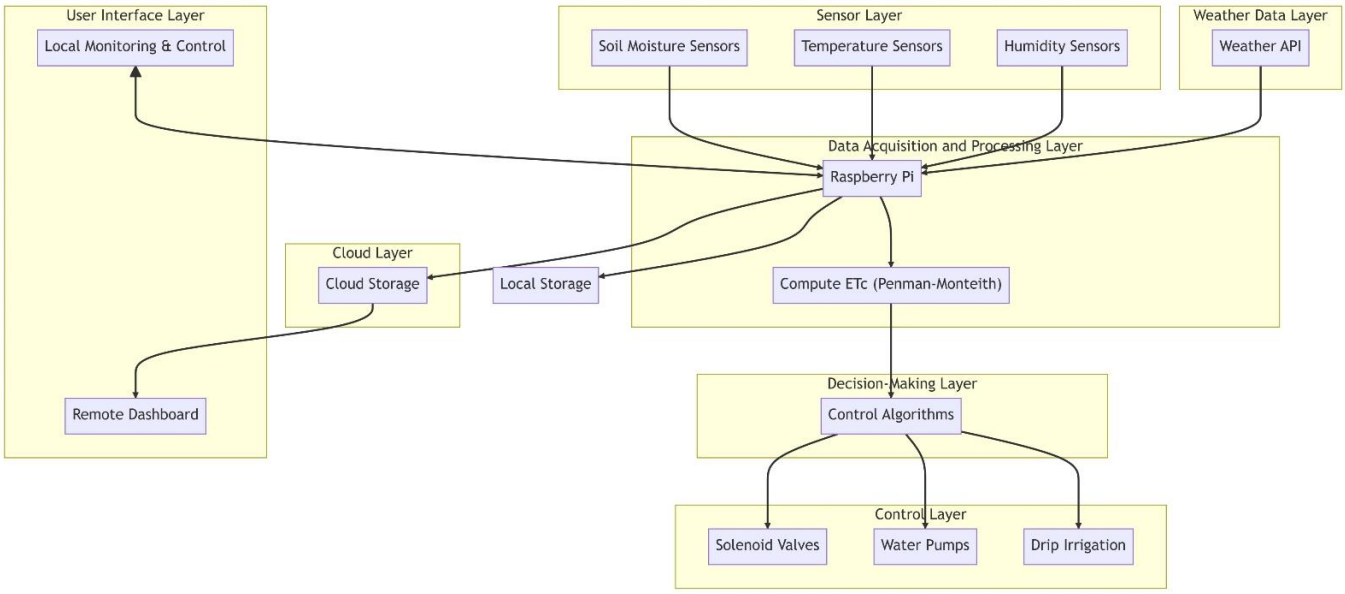


Fig. 3. Layers of the proposed system.

1) Sensor layer: The sensor layer forms the foundational tier of the system, designed to capture essential agroclimatic data from the field. This layer includes soil moisture sensors placed at various depths, allowing for accurate real-time monitoring of soil water content. These sensors provide critical insights into the hydration status of the soil, which directly informs irrigation decisions.

In addition to soil moisture sensors, the system integrates weather data either through meteorological stations or APIs, collecting parameters like temperature, humidity, solar radiation, wind speed, and precipitation. By combining soil and environmental data, the system accurately assesses water stress in crops, enabling precise irrigation scheduling. This data also plays a significant role in calculating the ET_0 using the Penman-Monteith Eq. (1), crucial for determining the crops' water needs.

2) Data acquisition and processing layer: This layer is responsible for collecting, transmitting, and preparing the data for further analysis. Sensor readings are transmitted to the control unit via efficient wireless communication protocols like LoRaWAN or Zigbee, ensuring low power consumption and minimal data loss. Once received, raw data undergoes preprocessing (cleaning, calibration, and normalization) to ensure accuracy and comparability. This process transforms the data into actionable insights, such as identifying soil moisture trends and temperature fluctuations.

Processed data is stored locally for real-time decision-making and periodically synchronized with the cloud for long-term analysis and backup. This dual-storage approach ensures the system operates efficiently while enabling remote access to historical data for improved decision-making.

3) Decision making layer: The decision-making layer is essential to the smart irrigation system, responsible for analyzing data from various sensors to generate precise irrigation schedules. It extracts insights on soil moisture, weather conditions, and crop development, enabling informed water management decisions based on real-time environmental factors. A critical function is calculating ET_C based on ET_0 (1) and applying crop coefficients K_C (2).

This layer generates irrigation schedules that optimize water usage by considering soil moisture, weather forecasts, and irrigation efficiency. Optimization algorithms or rule-based approaches ensure efficient irrigation timing. Additionally, ML models enhance decision-making by learning from historical data, improving predictions of crop water requirements and refining irrigation strategies over time, akin to methods used in precipitation prediction in Casablanca, Morocco [32].

4) Control layer: The control layer manages the execution of irrigation commands. It coordinates devices such as valves, pumps, and sprinklers, ensuring that water is delivered according to the schedules generated by the decision-making layer. The system continuously monitors environmental conditions and adjusts irrigation in real time. For instance, in case of unexpected rainfall, the control layer can halt irrigation to prevent overwatering.

Furthermore, this layer handles error detection and power management, optimizing energy use by aligning irrigation activities with off-peak hours, which reduces operational costs. By efficiently managing water and energy resources, the control layer supports sustainable irrigation practices.

5) Cloud layer: The cloud layer in the proposed smart irrigation system is primarily used for data storage and remote visualization through a centralized dashboard. It stores processed data, including sensor readings and irrigation schedules, ensuring secure, long-term access for analysis and system optimization. This approach is particularly suited for regions with poor or intermittent internet connectivity, as the system can function locally and synchronize data with the cloud when the connection is available, maintaining continuous operation without relying on constant internet access.

In addition to storage, the cloud layer provides real-time data visualization via a web-based dashboard. This allows farmers and administrators to remotely monitor critical metrics such as soil moisture levels, irrigation schedules, and overall system performance, even in areas where reliable connectivity is limited.

6) User interface layer: The user interface layer facilitates human-computer interaction for system monitoring, control, and data visualization. It comprises two primary components:

- Local User Interface: A user-friendly interface accessible through a local device (e.g., tablet, computer) allows farmers to monitor real-time sensor data, control irrigation schedules, and make necessary adjustments directly within the system. This local interface provides full control over the irrigation process, enabling immediate responses to changing conditions in the field.
- Cloud-Based User Interface: A web-based interface provides remote access for data visualization and performance monitoring. Through this interface, farmers and system administrators can visualize real-time and historical data, review system performance metrics, and analyze trends.

C. System Functionality and Workflow

The proposed smart irrigation system operates through a sequential process involving data acquisition, processing, analysis, decision-making, control, and user interaction. Fig. 4 illustrates the flowchart of this process.

- Data Acquisition: The system gathers environmental data, including soil moisture, temperature, humidity, and solar radiation, from sensors strategically placed throughout the agricultural field and weather station. This data is essential for understanding field conditions and informing irrigation decisions.
- Data Transmission: Collected data is transmitted to the central control unit via wireless communication protocols, ensuring reliable and efficient data transfer.
- Data Preprocessing: Raw sensor data undergoes cleaning, normalization, and feature extraction to prepare it for analysis and decision-making. Data cleaning identifies and removes outliers, missing values, or inconsistencies. Normalization scales data to a common range, facilitating comparison and analysis.
- ETc Calculation: Crop water requirements are estimated using the Penman-Monteith method as shown in Eq. (1) and Eq. (2).
- Irrigation Scheduling: Optimal irrigation schedules are generated based on ETc, soil moisture data, and weather conditions. Optimization algorithms or rule-based approaches are employed to determine the most efficient irrigation strategies.
- Irrigation Control: Irrigation devices are activated and deactivated according to the generated schedules. Real-time adjustments can be made based on sensor data and weather conditions.
- Data Analysis and Monitoring: Collected data and system performance metrics are analyzed to evaluate irrigation efficiency, crop water use, and equipment status. Regular monitoring and analysis enable continuous improvement and optimization of the system.
- User Interaction: Farmers and system administrators can interact with the system through local and cloud-based interfaces to monitor performance, adjust parameters, and access historical data.

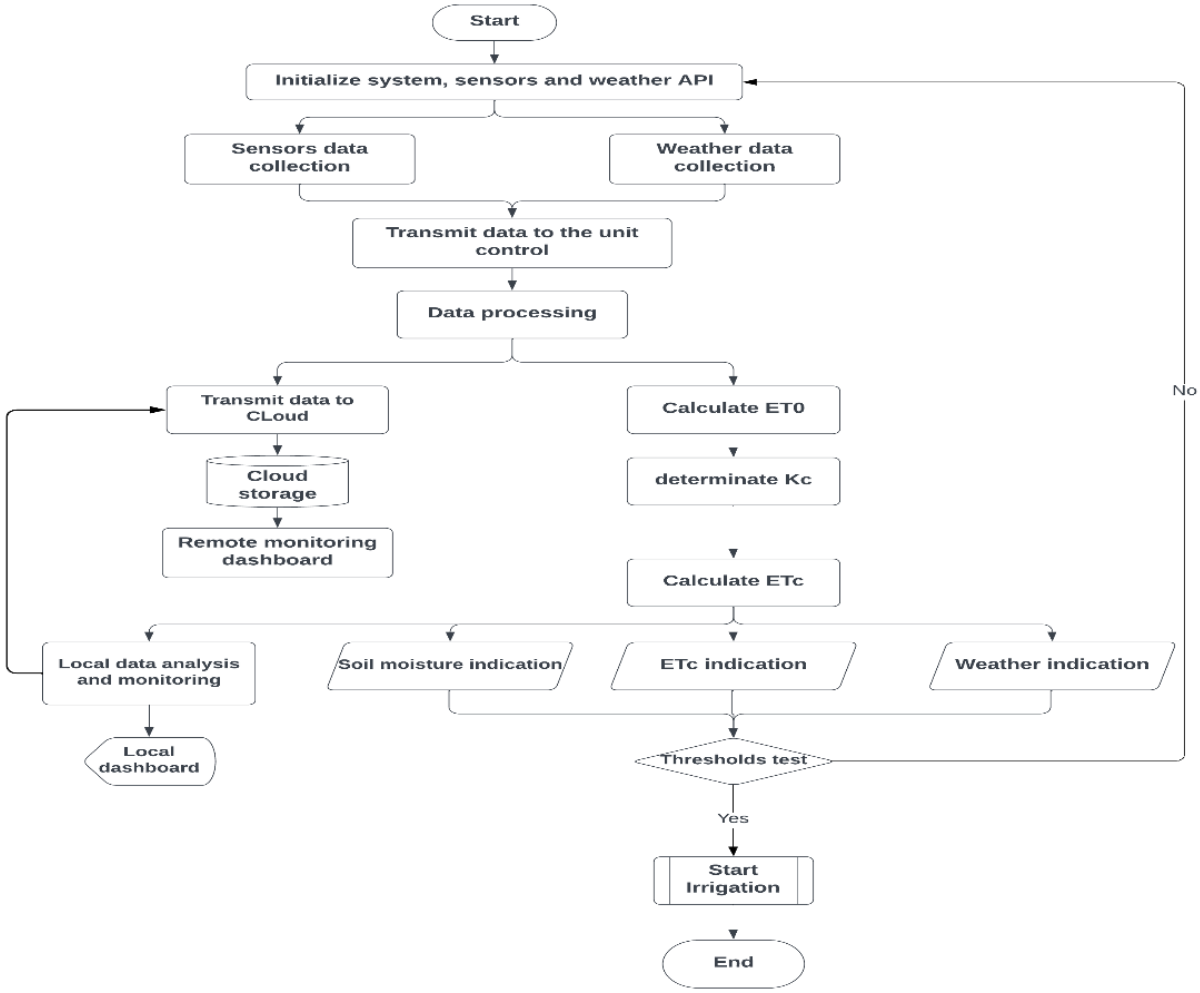


Fig. 4. Workflow of the proposed approach.

VI. Conclusion

The integration of advanced technologies into irrigation systems plays a pivotal role in delivering efficient and sustainable solutions for modern agriculture. Our proposed smart irrigation system, leveraging IoT technology, sensor networks, and intelligent algorithms, addresses the pressing challenges of water conservation while enhancing crop productivity. At its foundation, the sensor layer provides crucial real-time data on soil moisture and environmental conditions, feeding into advanced data processing and decision-making layers. By correlating soil moisture levels with meteorological data and calculating reference evapotranspiration, the system achieves a nuanced understanding of crop water requirements, enabling precise irrigation management that minimizes water waste while maximizing yield and crop quality.

While our approach shows great promise, further research and field testing are necessary to validate its effectiveness across various crop types and climatic conditions. Future work should focus on refining the system's predictive capabilities, incorporating ML algorithms, and exploring the integration of additional data sources. As we continue to refine and expand this system, we move closer to realizing the vision of precision agriculture, where resource use is optimized, environmental impact is minimized, and crop yields are maximized. Our proposed smart irrigation system represents a significant step forward in the pursuit of sustainable and efficient agricultural water management, offering a practical solution to the global challenge of water scarcity in agriculture.

Acknowledgment

This work was supported by the National Center for Scientific and Technical Research (CNRST) under the Research Excellence Scholarship Program.

References

[1] A. Baggio, “Wireless sensor networks in precision agriculture,” 2015 IEEE 7th International Conference on Intelligent Computing and Information Systems, ICICIS 2015, pp. 77–80, 2016, doi: 10.1109/IntelCIS.2015.7397200.

[2] J. Burrell, T. Brooke, and R. Beckwith, “Vineyard Computing: Sensor Networks in Agricultural Production,” IEEE Pervasive Comput, vol. 3, no. 1, pp. 38–45, 2004, doi: 10.1109/MPRV.2004.1269130.

[3] J. Muangprathub, N. Boonnam, S. Kajornkasirat, N. Lekbangpong, A. Wanichsombat, and P. Nillaor, “IoT and agriculture data analysis for smart farm,” Comput Electron Agric, vol. 156, no. June 2018, pp. 467–474, 2019, doi: 10.1016/j.compag.2018.12.011.

[4] F. Capello, M. Toja, and N. Trapani, "A real-Time monitoring service based on industrial internet of things to manage agrifood logistics," ILS 2016 - 6th International Conference on Information Systems, Logistics and Supply Chain, no. June, 2016.

[5] S. Fang et al., "An integrated system for regional environmental monitoring and management based on internet of things," IEEE Trans Industr Inform, vol. 10, no. 2, pp. 1596–1605, 2014, doi: 10.1109/TII.2014.2302638.

[6] N. M. Z. Hashim, S. R. Mazlan, M. Z. A. Abd Aziz, A. Salleh, A. S. Ja'Afar, and N. R. Mohamad, "Agriculture monitoring system: A study," J Teknol, vol. 77, no. 1, pp. 53–59, 2015, doi: 10.11113/jt.v77.4099.

[7] R. K. Kodali, N. Rawat, and L. Boppana, "WSN sensors for precision agriculture," IEEE TENSYMP 2014 - 2014 IEEE Region 10 Symposium, no. 1, pp. 651–656, 2014, doi: 10.1109/tenconspring.2014.6863114.

[8] A. D. Boursianis et al., "Smart Irrigation System for Precision Agriculture - The ARETHOU5A IoT Platform," IEEE Sens J, vol. 21, no. 16, pp. 17539–17547, 2021, doi: 10.1109/JSEN.2020.3033526.

[9] R. S. Krishnan et al., "Fuzzy Logic based Smart Irrigation System using Internet of Things," J Clean Prod, vol. 252, p. 119902, 2020, doi: 10.1016/j.jclepro.2019.119902.

[10] A. F. Jiménez, P. F. Cárdenas, and F. Jiménez, "Intelligent IoT-multiagent precision irrigation approach for improving water use efficiency in irrigation systems at farm and district scales," Comput Electron Agric, vol. 192, no. May 2021, 2022, doi: 10.1016/j.compag.2021.106635.

[11] O. Debauche, S. Mahmoudi, S. A. Mahmoudi, P. Manneback, and F. Lebeau, "A new edge architecture for AI-IoT services deployment," Procedia Comput Sci, vol. 175, pp. 10–19, 2020, doi: 10.1016/j.procs.2020.07.006.

[12] O. Debauche, S. Mahmoudi, M. Elmoulat, S. A. Mahmoudi, P. Manneback, and F. Lebeau, "Edge AI-IoT pivot irrigation, plant diseases, and pests identification," Procedia Comput Sci, vol. 177, pp. 40–48, 2020, doi: 10.1016/j.procs.2020.10.009.

[13] A. Al-Naji, A. B. Fakhri, S. K. Gharghan, and J. Chahl, "Soil color analysis based on a RGB camera and an artificial neural network towards smart irrigation: A pilot study," Heliyon, vol. 7, no. 1, p. e06078, 2021, doi: 10.1016/j.heliyon.2021.e06078.

[14] A. H. Blasi, M. A. Abbadi, and R. Al-Huweimel, "Machine Learning Approach for an Automatic Irrigation System in Southern Jordan Valley," Engineering, Technology & Applied Science Research, vol. 11, no. 1, pp. 6609–6613, 2021, doi: 10.48084/etasr.3944.

[15] A. Vij, S. Vijendra, A. Jain, S. Bajaj, A. Bassi, and A. Sharma, "IoT and Machine Learning Approaches for Automation of Farm Irrigation System," Procedia Comput Sci, vol. 167, pp. 1250–1257, 2020, doi: 10.1016/j.procs.2020.03.440.

[16] Cherine Fathy and Hassan M. Ali, "sensors-23-02091," Sensors, Feb. 2023.

[17] J. Angelin Blessy and A. Kumar, "Smart irrigation system techniques using artificial intelligence and IoT," Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021, no. Icicv, pp. 1355–1359, 2021, doi: 10.1109/ICICV50876.2021.9388444.

[18] K. Bodake, R. Ghate, H. Doshi, P. Jadhav, and B. Tarle, "Soil based Fertilizer Recommendation System using Internet of Things," MVP Journal of Engineering Sciences, vol. 1, no. 1, pp. 13–19, 2018, doi: 10.18311/mvpjes/2018/v1i1/18273.

[19] B. B. Sinha and R. Dhanalakshmi, "Recent advancements and challenges of Internet of Things in smart agriculture: A survey," Future Generation Computer Systems, vol. 126, pp. 169–184, 2022, doi: 10.1016/j.future.2021.08.006.

[20] S. Wadekar, V. Vakare, R. Prajapati, S. Yadav, and V. Yadav, "Smart water management using IOT," 2016 5th International Conference on Wireless Networks and Embedded Systems, WECON 2016, pp. 3–6, 2017, doi: 10.1109/WECON.2016.7993425.

[21] Z. Sheng, S. Yang, Y. Yu, A. Vasilakos, J. McCann, and K. Leung, "A survey on the ietf protocol suite for the internet of things: Standards, challenges, and opportunities," IEEE Wirel Commun, vol. 20, no. 6, pp. 91–98, 2013, doi: 10.1109/MWC.2013.6704479.

[22] O. Debauche, M. El Moulat, S. Mahmoudi, P. Manneback, and F. Lebeau, "Irrigation pivot-center connected at low cost for the reduction of crop water requirements," Proceedings - 2018 International Conference on Advanced Communication Technologies and Networking, CommNet 2018, no. November 2019, pp. 1–9, 2018, doi: 10.1109/COMMNET.2018.8360259.

[23] E. A. Abioye et al., "A review on monitoring and advanced control strategies for precision irrigation," Comput Electron Agric, vol. 173, no. April, p. 105441, 2020, doi: 10.1016/j.compag.2020.105441.

[24] A. Kumar, B. S. Dhaliwal, and D. Singh, "Cross-Layer based Energy Efficient Wireless Sensor Network for Large Farms," International Journal of Intelligent Engineering and Systems, vol. 15, no. 5, pp. 483–493, Oct. 2022, doi: 10.22266/ijies2022.1031.42.

[25] E. Bwambale, F. K. Abagale, and G. K. Anornu, "Smart irrigation monitoring and control strategies for improving water use efficiency in precision agriculture: A review," Feb. 01, 2022, Elsevier B.V. doi: 10.1016/j.agwat.2021.107324.

[26] R. Shigeta, Y. Kawahara, G. D. Goud, and B. B. Naik, "Capacitive-Touch-Based Soil Monitoring Device with Exchangeable Sensor Probe," Proceedings of IEEE Sensors, vol. 2018-Octob, pp. 1–4, 2018, doi: 10.1109/ICSENS.2018.8589698.

[27] G. C. Topp, J. L. Davis, and A. P. Annan, "Electromagnetic Determination of Soil Water Content: Measurements in Coaxial Transmission Lines," 1980.

[28] S. Hebbar and G. V. Prasad, "Automatic water supply system for plants by using wireless sensor network," Proceedings of the International Conference on IoT in Social, Mobile, Analytics and Cloud, I-SMAC 2017, pp. 742–745, 2017, doi: 10.1109/I-SMAC.2017.8058277.

[29] R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, "FAO Irrigation and Drainage Paper No. 56 - Crop Evapotranspiration," no. January 1998, 1998.

[30] B. Keswani et al., "Adapting weather conditions based IoT enabled smart irrigation technique in precision agriculture mechanisms," Neural Comput Appl, vol. 31, pp. 277–292, 2019, doi: 10.1007/s00521-018-3737-1.

[31] T. Wasson, T. Choudhury, S. Sharma, and P. Kumar, "Integration of RFID and sensor in agriculture using IOT," Proceedings of the 2017 International Conference On Smart Technology for Smart Nation, SmartTechCon 2017, pp. 217–222, 2018, doi: 10.1109/SmartTechCon.2017.8358372.

[32] A. Tricha and L. Moussaid, "Evaluating machine learning models for precipitation prediction in Casablanca City," Indonesian Journal of Electrical Engineering and Computer Science, vol. 35, no. 2, pp. 1325–1332, Aug. 2024, doi: 10.11591/ijeecs.v35.i2.pp1325-1332.

An Efficient Diabetic Retinopathy Detection and Classification System Using LRKSA-CNN and KM-ANFIS

Rachna Kumari[1], Sanjeev Kumar[2] , Sunila Godara[3]
Research Scholar, Department of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar, India[1]
Professor, Department of Computer Science & Engineering, Guru Jambheshwar University of Science & Technology, Hisar, India[2, 3]

***Abstract*—If Diabetic Retinopathy (DR) is not diagnosed in the early stages, it leads to impaired vision and often causes blindness. So, diagnosis of DR is essential. For detecting DR and its diverse stages, various approaches were developed. However, they are limited in considering microstructural changes of visual pathways associated with the visual impairment of DR. Thus, this work proposes an effective Linearly Regressed Kernel and Scaled Activation-based Convolution Neural Network (LRKSA-CNN) to diagnose DR utilizing multimodal images. Primarily, the input Optical Coherence Tomography (OCT) image is preprocessed for contrast enhancement utilizing Contrast-Limited Adaptive Histogram Equalization (CLAHE) and resolution enhancement utilizing the Gaussian Mixture Model (GMM). Likewise, the Magnetic Resonance Imaging (MRI) image's contrast is also improved, and edge sharpening is performed utilizing Unsharp Mask Filter (USF). Then, preprocessed images are segmented utilizing the Intervening Contour Similarity Weights-based Watershed Segmentation (ICSW-WS) algorithm. Significant features are extracted from the segmented regions. Next, important features are chosen utilizing the Min-max normalization-based Green Anaconda Optimization (MM-GAO) algorithm. By utilizing the LRKSA-CNN technique, the selected features were classified into DR and Non-Diabetic Retinopathy (NDR). Hence, utilizing the Krusinka Membership-based Adaptive Neuro Fuzzy Interference System (KM-ANFIS), various stages of DR were classified based on the presence of intermediate features. Lastly, the proposed system achieves superior outcomes than the baseline systems.**

***Keywords*—Intervening contour similarity weights based watershed segmentation (ICSW-WS); min-max normalization based green anaconda optimization (MM-GAO); krusinka membership based adaptive neuro fuzzy interference system (KM-ANFIS); linearly regressed kernel and scaled activation based convolution neural network (LRKSA-CNN); and deep learning**

I. Introduction

The well-known reason for visual deficiency is DR, which is also known as diabetic eye disease. DR has turned into a global pandemic with 365 million individuals expected to be impacted by 2025 [1]. People who have diabetes for a long duration are more likely to develop DR, and it impressively causes vision loss by the gradual destruction of the retina's blood vessels over a while [2]. Lesions like Neovascularization (NV), exudates, hemorrhages, and microaneurysms may develop in DR patients. For proper intervention, these lesions need to be detected early. Thus, to manage DR and prevent progression, early detection and treatment are the most efficient ways [3] [4].

Various tasks associated with automatic DR diagnosis have been performed in the past decades. The automatic classification of DR from retinal images is increasingly studied owing to the huge number of diabetic patients and the need for more accurate and automatic diagnosis [5]. Artificial Intelligence (AI) can help in solving this problem. For analyzing raw medical images from beginning to end and predicting a goal result, Deep Learning (DL), particularly, Deep Convolutional Neural Networks (DCNN) and Deep Neural Networks (DNNs) can be utilized [6] [7].These systems have been trained on tens of thousands of medical images and demonstrated screening performance similar to that of retinal specialists [8]. Various screening tools are available to receive those kinds of medical images, such as the visual acuity test, OCT, colour fundus camera, eye angiography, and so on [9] [10]. The conventional screening of fundus images to diagnose DR lesions takes time, thus delaying therapy and minimizing the chance of success [11]. But, severe bleeding symptoms cover NV outside the macula in fundus photography, which leads to misdiagnosis [12]. However, evaluating the intensity of retinopathy in people with diabetes is mostly dependent on human assessment of retinal fundus pictures, which is complicated to perform [13] [14]. These problems can be alleviated via the use of OCT, which together provide micrometer-scale resolution structural, 3-dimensional (3D), depth-resolved, and vascular images of the retina [15]. But, recent studies showed that visual deficiency for DR is also associated with the pathophysiological changes of different parts of the visual pathway. To maintain visual function, the evaluation of the progressive visual pathway along with the visual field is considered to be a vital process.

A. Problem Statement

Most of the prevailing models developed for DR diagnosis were limited in solving the following challenges:

- Limited research was performed for exposing the invisible pathophysiological mechanism i.e.,

pathophysiological changes and related damages to the different retinal layers along with the cerebral visual pathway to diagnose diabetic retinopathy.

- Improper segmentation of small regions was caused by the loss of apparent activity in tiny objects or else regions due to the imaging systems' limited resolution that is, blurring of intensities near their edges.
- Owing to the poor image resolutions and the inability of such images to exhibit more pixel information, the segmentation techniques were ineffective on OCT images and ignored detailed information.
- Prevailing techniques utilized limited features and the model often predicted near the same value due to the exclusion of some key features, which influenced the outcome of the prediction system.
- Prevailing techniques utilized limited features, and the exclusion of some key features caused the model to often predict near the same value, which might influence the outcome of the prediction system.

Thus, for solving the aforementioned problems, this paper proposes an effective LRKSA-CNN-based DL technique for DR diagnosis. The proposed system includes the subsequent contributions:

- A multimodal imaging strategy is used to explore microstructural changes in the visual pathway and invisible pathologies inside the retina structure.
- To recover blurred intensities of edges, remove imperfections in the images, and make edges easier to the visible edge, sharpening is performed to avoid improper segmentation of small regions.
- The finer details of images were accurately revealed via the resolution enhancement of OCT images utilizing the PIR technique to improve the segmentation accuracy.
- Along with the intermediate features, the global and local features were also extracted, which enhances the model efficiency to better explore the progression of eye disease.

This paper's remaining portion is aligned as: Section II reviews the associated works, Section III describes the proposed technique, Section IV propounds the experimental outcomes and discussion, and Section V concludes the paper.

II. Related Literature Survey

M. Sakthisreedevi *et al.* [16] detected DR by identifying its features in OCT images. Utilizing the graph-cut technique, 7 retinal layers were segmented from the OCT images. Next, the features of retinal layers were extracted for differentiating normal and DR subjects. Experimental outcomes showed that for diagnosing DR, the extracted features were effective. But, the system was limited in specifying the classification process. Mahmoud Elgafi *et al.* [17] propounded a three-step model for DR detection using OCT images. The retinal layers were segmented and the 3D features were extracted from every single layer. Utilizing Backpropagation Neural Networks (BPNN), the features were classified. Experimental outcomes displayed the recommended technique's potential for DR detection using OCT images. But, the system accuracy was poor owing to the usage of limited data. Brahami Menaouer *et al.* [18] established a hybrid DL technique for DR detection. DCNN and Visual Geometric Group (VGG) network models-centric classification were followed in this model. The classification as per the visual risk was related to the retinal ischemia's severity. According to the experimental outcomes, the hybrid technique diagnosed DR accurately with more speed. The model's major limitation was the increased training time to train its parameters. Sambit S. Mondal *et al.* [19] presented an automated ensemble DL system for DR detection. The images were pre-processed and augmented utilizing a General Adversarial Network (GAN).For DR detection, the modified DenseNet101 and ResNet DL systems were the ensemble. The comparison exhibited that for DR classification, this technique had high accuracy. But, the system accuracy was low for five classes rather than two classes. M. Murugappan *et al.* [20] implemented the Few Shot Learning (FSL) classification networks to grade and detect DR centered on the attention mechanism. The model utilized an episodic learning strategy and attention mechanism for the few-shot classification task. As per the analysis of performance metrics, the model effectively detected DR. However, to prove the efficiency of the developed system, the comparison made in the paper was limited. Muhammad Mohsin Butt *et al.* [21] recommended a hybrid technique for DR detection as well as classification in fundus images. Utilizing transfer learning-based CNN, the features were extracted. The features were merged into a hybrid feature vector and utilized for the DR classification. Experimental outcomes revealed that significant performance improvement was provided by the model. But, the supremacy of model performance was degraded by the poor quality of input images. Abdüssamed Erciyas & NecaattinBarışçı [22] propounded a DL-centric framework for DR lesions' detection and classification. The DR lesions were detected from the DR data and marked utilizing Regional CNN (FRCNN).By utilizing the transfer learning and attention mechanism, the images from FRCNN were categorized. The result comparison showed that the developed system acquired more successful outcomes. Still, the system utilized larger computational resources. Mohamed Elsharkawy *et al.* [23] established an OCT-centric diagnosis technique for detecting DR. By utilizing prior shape knowledge, the model segmented the retinal layers. For global subject diagnosis, the classification was performed utilizing an Artificial Neural Network (ANN). The comparison outcomes indicated the ability of the system for diagnosing DR. However, the model attained poor classification accuracy owing to the limited number of features. Wejdan L. Alyoubi *et al.* [24] executed DR images' classification under various stages and localization of affected lesions. Two DL models, namely CNN-512 for classification and the YOLOv3 system for localizing DR lesions were utilized. The DR images were classified into five DR stages. The YOLOv3 model cannot detect smaller DR lesions owing to its anchor box design even though the acquired accuracy exceeded baseline outcomes. Mohammed Ghazal *et al.* [25]

concentrated on detecting Non-Proliferative DR (NPDR) utilizing CNN. Initially, the retina patches were extracted. Transfer learning centered on 2 pre-trained CNNs was utilized, where one was independently fed by nasal retina patches and the other one by temporal retina patches. The outcomes showed that the system attained a promising accuracy. But, the improper alignment of patches led to lower accuracy of the model. S. K. Sahu *et al.* [26] implemented different AI-based transfer learning models for diabetic retinopathy at early stages. The accuracy transfer learning models ranges from 74% to 81. S. Goswami [27] proposed an iterative attentional feature fusion (iAFF) method for DR detection. Model used InceptionV3 and Xception feature extraction. IDRiD data set was used for training and testing purpose. N. Z. Abidin and A. Ritahani Ismail [28] propose a federated deep learning model eliminate the need of pooling data in a single location. This model allowed deep learning algorithms to train from diverse sets of data from different sources. O. F. Gurcan *et al.* [29] proposed a metaheuristic algorithm based on deep learning for DR detection. In Model InceptionV3 was used to extract feature. The transfer learning method is applied in the extraction process. Particle Swarm optimization and Artificial Bee Colony were used for feature selection. Which were further classified eXtreme Gradient Boosting (XGBoost). M. Jahiruzzaman [30] proposed a k-means color compression technique to detect hemorrhages and exudates in color fumdus images. The different region of fundus images were segmented out and classified using fuzzy inference system. M. Ghazal [31] proposed a computer aided diagnostic (CAD) system using CNNs for non-proliferative DR (NPDR) detection at early stage. A two pre-trained CNNs, one is independent fed of nasal retina patches and the other is by temporal retina patches was used to optimize the performance of model which was accuracy of 94%. H. Mustafa *et al.* [32] proposed a multi-stream ensemble deep network to categorize DR severity at different stages. Model utilizes benefits of the deep networks and principal component analysis (PCA) to train inter-class and intra-class variations from image features. Pre-trained deep learning architecture DenseNet-121 was used to extract main feature and Ensemble machine learning classifiers PCA was applied to classify images. Further AdaBoost and random forest algorithms were applied to improve classification accuracy. K. Aurangzeb *et al.* [33] implemented a ColonSegNet model for retinal vessel segmentation. This method efficiently locate vessels and applied data augmentation to resolve fewer graded images issues using intelligent evolution algorithms optimal values for the contrast enhancement were identified. DRIVE, CHASE_DB, and STARE datasets were used for training and testing purpose. M. Nur-A-Alam [34] proposed an automatic and intelligent system to classify DR images. First retinal images were pre-processed, then histograms of oriented gradient (HOG), Shearlet transform, and Region-Based Convolutional Neural Network (RCNN) were used to extract discriminating features and merged features as one fused feature vector. By using the fused features, a faster RCNN classifier was used to DR severe stages. R. Pires et al. [35] presented the bag-of-visual-words (BoVW) based algorithm for lesion detection in retinal images. The input of the metaclassifier work as output 9BoVW) of several lesion detectors. This algorithm create a high-level feature representation for the retinal images to lesion detection. M. T. Islam *et al.* [36] proposed a multi-stage convolutional neural network (CNN)-based model DiaNet to DR. study of this paper revealed that fundus images restrain sufficient information to differentiate the Qatari diabetes cohort from the control group. D. Maji [37] proposed a novel deep learning approach for retinal blood vessel and DR. The EfficientNet model was used with pre-trained weights facilitating transfer learning and optimize convergence speed. X. Wang *et al.* [38] proposed a supervised deep learning framework for macula-related disease classification with uncertainty estimation. In their model a convolutional neural network based instance-level classifier was refined by through uncertainty-driven deep learning method. Second, a recurrent neural network extract features and generated bag-level. M. M. Farag *et al.* [39] proposed a noval deep-learning-based method for DR detection using a single Color Fundus photograph. Authors employed DenseNet169's encoder to assemble a visual embedding. Then, Convolutional Block Attention Module (CBAM) was utilized to reinforce its discriminative power. At last, the model was trained on (APTOS) dataset using cross-entropy loss. T. Liu *et al.* [40] proposed a novel deep symmetric convolutional neural network to detect the microaneurysms (MAs) and hard exudates (HEs) of DR. Furthermore, different convolution pooling structures were practiced to obtain feature filtering in feature extraction phase. They concluded that microaneurysms detection was improved by using ave-pooling layer. G. T. Reddy *et al.* [41] implemented Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) dimensionality reduction techniques, on Random Forest Classifier, Decision Tree Induction, Naive Bayes Classifier and Support Vector Machine (SVM), using Cardiotocography (CTG) dataset. From their paper it is concluded that performance of the classifiers, Random Forest, Decision Tree is not much affected by PCA and LDA. E. O. Rodrigues *et al.* [42] proposed a new ELEMENT (vEsseL sEgmentation using Machine lEarning and coNnecTivity) framework for vessel segmentation. Features were extracted based on vessel connectivity and grey level properties. This model speeds up the segmentation throughput and minimize. S. Wang *et al.* [43] proposed an integrated machine learning approach for microaneurysms detection. In this method first Candidate objects were located through a dark object filtering process. Then correlation coefficient between each processed profile and MA profile was determined. Furthermore, K-nearest neighbor classifier was applied to extract set of statistical features. K. Shankar *et al.* [44] proposed an automated Hyperparameter Tuning Inception-v4 (HPTI-v4) model for DR detection. At the preprocessing stage, contrast limited adaptive histogram equalization (CLAHE) was used to improve the contrast level of images. Furthermore, the HPTI-v4 model was employed to extract the essential features from the segmented image and then multilayer perceptron (MLP) was used for classification. B. Yang *et al.* [45] implemented a global channel attention mechanism (GCA) framework to detect DR at early stages. In this module, a one-dimensional convolution kernel size algorithm was used and GCA-EfficientNet (GENet) was employed to classify images. Experimental outcomes

conclude that GENet are more effective to extract lesion features and classify DR stages. I. Usman and K. A. Almejalli [46] implemented a new machine learning technique based on Genetic Programming for MAs detection in retinal images. The optimal expression was evolved generation by generation using the binary fitness scores and a stepwise enhancement process. The best expression obtained was then used as a classifier MA. B. Abdillah *et al.* [47] implemented texture feature model based on Local Binary Pattern for DR detection. k-Nearest Neighbor (k-NN) and Support Vector Machines (SVM) was used for classification of images. M. J. J. P. van Grinsven et al. [48] proposed a CNN based method for detection of hemorrhages in retinal images. Based on current status of CNN, Training samples were heuristically misclassified negative samples and weights are assigned in the next CNN training iteration. A. Krestanova *et al.* [49] reviewed all segmentation techniques for retinal blood vessel extraction. Their review included analysis of segmentation techniques based on degree of curvature of retinal blood vessels and objectification parameters. They reviewed all calculations and metrics to obtain the degree of tortuosity of retinal blood vessels. K. Wisaeng and W. Sa-Ngiamvibool [50] proposed a mathematical morphology algorithm to localize OD. Firstly, a coarse segmentation method was utilized to obtain exudates and non-exudates candidates. Finally, mathematical morphology algorithm was applied to classify exudates pixels. X. Li *et al.* [51] proposed a self-supervised learning model for DR detection. They used a rotation-oriented collaborative method to extract rotation-related and rotation-invariant features. Model was evaluated on two public datasets. G. Gupta [52] proposed a supervised learning framework based on micro-pattern of local variations using texture based analysis to localize patches colored fundus images. Rule-based criteria was used to determine the presence or absence of PDR. R. Kommaraju and M. S. Anbarasi [53] developed a machine learning framework to detect DR at early stages of color fundus images. Their model employed a swin transformer to detect the type of DR and Contrast Limited Adaptive Histogram Equalization (CLAHE) technique was used for image enhancement and flip for data augmentation. R. Gambhir [54] proposed a Shufflenetv2 method to detect DR at severity levels. Smooth L2 loss had been used to improve performance of model.

III. Proposed Diabetic Retinopathy Detection System

The proposed model is a type for analyzing both the invisible pathologies on different retinal layers and pathophysiological changes on various parts of the visual pathway. For attaining this, the proposed LRKSA-CNN model is trained on multimodal medical images like OCT and fMRI. Fig. 1 displays the proposed system's detailed workflow.

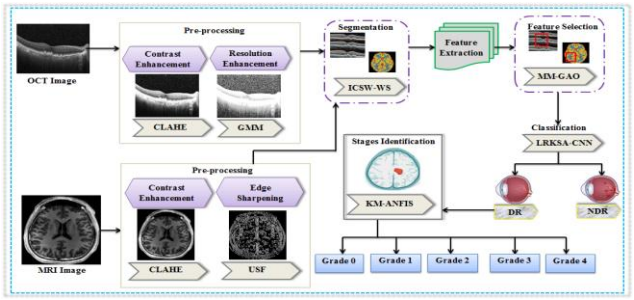


Fig. 1. Block diagram of the proposed methodology.

A. Preprocessing

The proposed technique's initial phase is preprocessing in which input OCT and MRI images are preprocessed for contrast enhancement. Since these images were obtained under various lighting conditions, they have lower contrast and require to be preprocessed before the execution model training. Thus, by utilizing the CLAHE technique, the input images' contrast was improved. Let $\left(R = \left\{R_{O(n)}, R_{M(n)}\right\}_{n=1}^{N}\right)$ be the set of input images, including two types of images, namely OCT $\left(R_{O(n)}\right)$ and MRI $\left(R_{M(n)}\right)$. These images are divided into smaller blocks $\left(R_m\right)$ by CLAHE, which computes the histogram $\left(\wp i(R_m) = \frac{\aleph_j}{\aleph_{avg}}\right)$ individually for each block. In the computation of the histogram, the average number of pixels per region is calculated as,

$$\aleph_{avg} = \frac{\aleph_h \times \aleph_v}{\aleph_{gr}}, \qquad \aleph \in R \tag{1}$$

Here, $\aleph_{avg}$ is the average number of pixels, $\aleph_{gr}$ is the number of grey levels, and $\aleph_h$ and $\aleph_v$ are the number of pixels in the horizontal and vertical dimensions. Subsequently, the clip limit to clip the histogram $\left(\wp i(R_m)\right)$ is defined and employed for the histogram as,

$$\aleph_\ell = \aleph_{avg} \times \aleph_{n\ell} \tag{2}$$

$$\wp i(R_m) = \begin{cases} \aleph_\ell & if\left(\aleph_j \geq \aleph_\ell\right) \\ \aleph_j & Otherwise \end{cases} \tag{3}$$

Here, $\aleph_j$ specifies the pixels with intensity (j), $\aleph_{n\ell}$ indicates the normalized level of contrast enhancement, and $\aleph_\ell$ depicts the clip limit. Next, the clipped pixel values are computed for evenly redistributing the clipped pixel values below the histogram.

$$\aleph_p = (\aleph_h \times \aleph_v) - \sum_{i=1 to K-1} \wp i(R_m) \quad (4)$$

Here, $\aleph_p$ depicts the number of clipped pixels, and K symbolizes the total number of grey levels. Next, the clipped histogram is normalized as,

$$\wp i(R_m) = \begin{cases} \aleph_\ell & if\left(\aleph_j + \frac{\aleph_p}{K} \geq \aleph_\ell\right) \\ \aleph_j + \frac{\aleph_p}{K} & Otherwise \end{cases} \quad (5)$$

Till all the pixels are redistributed, these steps are continued, and the cumulative histogram of every single region is calculated for grey-level mapping. Lastly, the histogram of each region is matched for computing the weighted sum of $(m_{1,2,3,4})$ neighboring regions, and the enhanced output is acquired by the interpolation of neighboring regions. The output image is acquired as,

$$R_{ce} = (1-h)(1-v) \times R_{m1} + h \times R_{m2} + v \times (1-h) \times R_{m3} + h \times R_{m4} \quad (6)$$

Here, $\left(R_{ce} = \{R_{eO(n)}, R_{eM(n)}\}_{n=1}^{N}\right)$ are the output-enhanced images, and (h) and (v) are the horizontal and vertical distances corresponding to the region.

From the preprocessed images, OCT images' spatial resolution is enhanced, and edge sharping for MRI images is performed.

1) Resolution enhancement of OCT: Here, the quality of $(R_{eO(n)})$ as OCT images' lower spatial resolution affects the segmentation outcome. Enhancing the resolution of OCT images gives vital information held in each pixel suitable for decision-making at the regional level. Thus, the patch-centric resolution enhancement utilizing GMM is given below:

The resolution enhancement technique executes the task of reconstructing the enhanced image $(R_{seO(n)})$ of *the input image* $(R_{eO(n)})$ centered on the low-resolution observation $(R_{lrO(n)})$ utilizing GMM. The enhanced image $(R_{seO(n)})$ is reconstructed by solving the following objective function as,

$$\arg\min_{R_{seO(n)}} \left\| \partial R_{eO(n)} - R_{lrO(n)} \right\|^2 - \sum_{i \in R} \log \hat{\lambda}\left(R_{eO(n)}, i\right) \quad (7)$$

Here, ∂ signifies the unknown operator for generating low-resolution observation $(R_{lrO(n)})$, $R_{lrO(n)}$ indicates the observed low-resolution image, $\hat{\lambda}$ symbolizes the probability density function, and $(R_{eO(n)}, i)$ are the patches in the input image.

To reconstruct the high-resolution image, the reference images $\{R'_{lrO(n)}, R'_{hrO(n)}\}$ are considered. From them, the low-resolution $\{R'_{lrO(n)}, i\}$ and high-resolution $\{R'_{hrO(n)}, i\}$ patches are extracted and concatenated $R' = \begin{pmatrix} R'_{lrO(n)} \\ R'_{hrO(n)} \end{pmatrix}$ for obtaining the parameters of GMM $(\alpha, \beta, \gamma, \lambda)$. The $\{R_{hrO(n)}, i\}$ from $(R_{eO(n)})$ is computed by choosing the best component of GMM as,

$$\chi = \arg\max_{\chi} \alpha \, p\left(R_{lrO(n)} \mid \alpha, \beta, \gamma, \lambda\right) \quad (8)$$

Here, the likelihood (p) of $\{R_{lrO(n)}, i\}$ to the χ component is maximal. Using the parameters, the high-resolution patch $\{R_{hrO(n)}, i\}$ is computed as,

$$R_{hrO(n)} = \alpha_{hr,\chi} + \sum_{hr,lr,\chi} \sum\nolimits_{lr,\chi}^{-1} \left(R_{lrO(n)} - \alpha_{lr,\chi}\right) \quad (9)$$

Here, the high-resolution patches $(R_{hrO(n)})$ are centered on the parameters of the generalized Gaussian mixture $\alpha_{hr,\chi}$. From all the estimated high-resolution patches, the high-resolution image $(R_{seO(n)})$ is reconstructed and utilized as input for the segmentation. This phase ignores the MRI images since they have better contrast resolution properties to extract images' finer details.

2) *Edge sharpening*: Here, the preprocessed MRI image $(R_{eM(n)})$ is utilized to recover edges that have been blurred owing to the partial volume effect. By doing so, the ideal boundary between two structures is restored for quantitatively distinguishingvarious types of tissues in the segmentation process. When performing edge sharpening for the OCT images, the noise in the image is also enhanced; thus, it becomes less natural and the level of information in flat and smooth areas is lost. Thus, the OCT images were not considered here. The USF is utilized for edge sharpening.

The main purpose of a USF is to subtract an unsharp mask from the input image. For this, the unsharp mask, that is, the original image's even more blurred version is yielded by spatially filtering the input image utilizing a Gaussian low-pass filter. The Gaussian function is derived as,

$$\eta(a,b) = \frac{1}{\varpi\sqrt{2\pi}} \exp\left(\frac{-a^2 - b^2}{2\varpi^2}\right) \quad (10)$$

Here, $\eta(a,b)$ indicates the Gaussian function, ϖ signifies the standard deviation value, and π specifies the weight mixing coefficient. The original image's Gaussian blurred version is acquired as,

$$R_{mM(n)} = \eta(a,b) * R_{eM(n)} \quad (11)$$

Next, the unsharp mask $\left(R_{mM(n)}\right)$ acquired is subtracted from the input image for getting edges as,

$$R_{sM}(n) = \frac{e}{2e-1} R_{mM}(a,b) - \frac{1-e}{2e-1} R_{eM}(n) \quad (12)$$

Here, $R_{sM}(n)$ symbolizes the output image, e is the constant to control the weightings of images, and $R_{mM}(a,b)$ indicates the brightness values of pixels at coordinates (a,b) in the unsharpmask R_{mM} .

B. Segmentation

After the success of resolution enhancement and edge sharpening, the resultant images $\left(R_{enh} = \{R_{seO(n)}, R_{sM(n)}\}\right)$ are subjected to the ICSW-WS algorithm for segmentation. The Watershed Segmentation (WS) algorithm is utilized as it can handle images with a considerable amount of irregular shapes and noise. But, the WS has the over-segmentation issue while a large number of small regions arise. For solving this problem, the conventional watershed algorithm calculates the intervening contour similarity weights. Hence, the segmentation is described below: Primarily, the input images are converted into greyscale, and local minima in the image are found. For this, by computing the intervening similarity weights, the watershed line between the two regions is created. The over-segmentation issue occurs as the segmentation process contains the segmentation of smaller regions. For solving this, the intervening similarity measure is computed as,

$$R_{wr} = \sqrt{\sum_{i=1ton} \left\| R_{enh(r1,i)} - R_{ref(r1,j)} \right\|_2} \quad (13)$$

Here, R_{wr} is the output image with weighted regions, and $R_{enh(r1,i)}, R_{ref(r1,j)}$ are the regions with different pixel values (i, j) in the reference image and input image. The similarity measure has the benefit of informing the intervention of another region, thus forming the outer contour of the region. Like this, the relation between the data samples was measured; for providing different regions of the image, the border information is jointly integrated.

After marking the regions, the algorithm floods the image with various colors and the color spreads for creating catchment basins by reaching the region boundary. The basins are formed utilizing the watershed function as,

$$R_{seg} = \varphi_{wsf}(d, R_{wt}) \quad (14)$$

Here, d signifies the point in the domain of the watershed function, R_{wt} indicates the regions assigned with the intervening weights, and φ_{wsf} specifies the watershed function.

In the proposed system, segmentation is performed in two different ways. The significant retinal layers are segmented from the OCT images for evaluating invisible changes in the retina. For segmenting the visual networks, namely posterior cingulate, calcarine, inferior parietal lobule, precuneus, lingual gyrus, cerebellum, and cuneus, the MRI images were utilized for analyzing the pathophysiological changes in the visual pathway. Thus, the segmented images are indicated as $\left(R_{seg} = \{R_{segO(n)}, R_{segM(n)}\}\right)$.

C. Feature Extraction

After segmentation, the significant features are extracted from $\left(R_{seg} = \{R_{segO(n)}, R_{segM(n)}\}\right)$. The proposed system extracted regional homogeneity, degree centrality, and Amplitude of Low-Frequency Fluctuations (ALFF) features from the visual network regions $\left(R_{segO(n)}\right)$, and reflectivity, curvature, volume, and thickness features from the segmented retinal layers $\left(R_{segM(n)}\right)$. Moreover, the global and local features to find interest points, contour, shape, texture, and visual information useful for class discrimination are extracted. Thus, the extracted features are given as,

$$\nabla_{fea(n)\in R_{seg}} = \{\nabla_{fea\in R_{segO(n)}}, \nabla_{fea\in R_{segM(n)}}, \nabla_{lfea\in R_{seg}}, \nabla_{gfea\in R_{seg}}\} \quad (15)$$

Here, the set of total features is signified as $\nabla_{fea(n)\in R_{seg}}$, the features extracted from the retinal layers and visual network areas are indicated as $\nabla_{fea\in R_{segO(n)}}, \nabla_{fea\in R_{segM(n)}}$, and the global and local features are symbolized as $\nabla_{lfea\in R_{seg}}, \nabla_{gfea\in R_{seg}}$.

D. Feature Selection

Here, the subset of the most relevant features is chosen from $\left(\nabla_{fea(n)\in R_{seg}}\right)$ for enhancing the system's performance. The MM-GAO algorithm is used for feature selection. The Green Anaconda Optimization algorithm is utilized for its ability to tackle a wider range of prey. However, the probability function of pheromone concentration in this algorithm only concentrates on maximum values, which leads to high risk of competition for captured food as it can't handle the minimum objective function. Thus, for solving this problem, the min-max normalization approach is

utilized in the prevailing green anaconda optimization algorithm for evaluating the probability of pheromone concentration.

In the MM-GAO algorithm, the positions of its population members are considered as the number of features extracted $\left(\nabla_{fea(n)\in R_{seg}}\right)$ from which the optimal solutions are chosen. The position of each green anaconda is randomly generated, and the objective function of problem space is calculated. It can be exposed as,

$$\nabla_{fea(n)} = \begin{bmatrix} \nabla_{fea(1,1)} & \cdots & \nabla_{fea(1,1)} \\ \vdots & \cdots & \vdots \\ \nabla_{fea(n,1)} & \cdots & \nabla_{fea(n,m)} \end{bmatrix} \quad (16)$$

$$\vartheta\left(\nabla_{fea(n)}\right) = \arg\min_{\nabla_{sf(n)}} \left|\xi_{tar(DR,NDR)} - \xi_{act(DR,NDR)}\right| \quad (17)$$

Here, m indicates the number of decision variables, $\vartheta(\bullet)$ signifies the objective function, and $|\bullet|$ depicts the distance between the actual output $\left(\xi_{act(DR,NDR)}\right)$ and the target $\left(\xi_{tar(DR,NDR)}\right)$ of the classifier output.

In the search process, the best value as per the objective function is chosen via two phases, namely exploration and exploitation.

Exploration: Here, the mating season of male and female anacondas is done. For updating the position of the green anaconda, the species strategy of the male is used. Utilizing this strategy, the male anaconda senses the pheromones, indicating the presence of the female anaconda, and moves towards it.

The set of female species is determined as,

$$\nabla_{FL(n)} = \left\{\nabla_{fea(k_n)} : \vartheta\left(\nabla_{fea(k_n)}\right)\right\}, \quad k_n \neq n \quad (18)$$

Here, $\nabla_{FL(n)}$ signifies the set of female species for each green anaconda in the first row of the population matrix k_n. Next, the probability of pheromone concentration for each population member utilizing the min-max approach is computed as,

$$\Phi_{(n,q)} = \sum_{l} \frac{\vartheta_l\left(\nabla_{fea(k_n)}\right) - \vartheta_{\min}\left(\nabla_{fea(k_n)}\right)}{\vartheta_{\max}\left(\nabla_{fea(k_n)}\right) - \vartheta_{\min}\left(\nabla_{fea(k_n)}\right)} \quad (19)$$

Here, $\Phi_{(n,q)}$ depicts the probability pheromone concentration of the q^{th} female for the n^{th} green anaconda, $\vartheta_l\left(\nabla_{fea(k_n)}\right)$ signifies the set of objective values of the candidate female, and $\vartheta_{\min}\left(\nabla_{fea(k_n)}\right)$, $\vartheta_{\max}\left(\nabla_{fea(k_n)}\right)$ are the minimum and maximum values.

The min-max approach assures the green anaconda's maxmin value to limit its search between minimum and maximum values, thus evaluating the minimum or maximum probability value that limits them in capturing food. Thus, the min-max approach finds the optimal move for the search agent via the complete assessment of search spaces and accelerates the decision-making process for quick convergence.

During exploration, the green anaconda randomly chooses a female species and moves towards it. Therefore, the green anaconda's position is updated as,

$$\nabla_{sFL(n)} = \nabla_{FL} : \Phi_{FL(n-1,q)} < \nabla_{sFL(n)} < \Phi_{FL(n,q)} \quad (20)$$

$$\Phi_{FL(n,q)} = \Phi_{(n,q)} + \Phi_{FL(n-1,q)} \quad (21)$$

$$^{d}\nabla_{fea(n,i)} = \nabla_{fea(n,i-1)} + \nabla_{fea(n,i-1)}\left(\nabla_{sFL(n,i-1)} - \tau.\nabla_{fea(n,i-1)}\right) \quad (22)$$

Here, $\nabla_{sFL(n)}$ signifies the selected female species, $\Phi_{FL(n,q)}$ depicts the cumulative probability function, $\nabla_{sFL(n)}$ symbolizes the random number, $^{d}\nabla_{fea(n,i)}$ is the newly updated position of the green anaconda in d dimensional search space, and $\nabla_{fea(n,i-1)}$ specifies the position updated in the prior iteration. Next, the objective function is assessed for updating the best solution.

Exploitation: Here, the hunting strategy of the green anaconda is followed in which it waits for the prey underwater, and when the prey passes, the anaconda surrounds and attacks the prey. Here, to obtain possible better solutions in local search, the position of group members is updated. The position is updated as,

$$\nabla_{fea(n,i)} = \nabla_{fea(n,i-1)} + (1-\tau)\frac{U-L}{i} \quad (23)$$

Here, U, V are the d-th search space's upper and lower bounds, and i indicates the number of iterations. Till the algorithm's last iteration, the updating process continues. The feature selection strategy utilizing the MM-GAO algorithm is displayed in Algorithm 1.

Algorithm 1: Feature Selection using MM-GAO

Input: Extracted feature $\left(\nabla_{fea(n)\in R_{seg}}\right)$

Output: Selected features $\nabla_{sf(n)}$

Begin

Initialize input data $\left(\nabla_{fea(n)\in R_{seg}}\right)$, parameters τ , Φ and ϑ , number of iterations i

Compute fitness of each individual (ϑ)

Select best solution

While $(i < i_{\max})$

Identify female species $\left(\nabla_{FL(n)}\right)$ //Exploration

Compute probability of pheromone concentration $\left(\Phi_{(n,q)}\right)$

Determine selected females $\nabla_{sFL(n)}$

Update new position ${}^{d}\nabla_{fea(n,i)}$

For each updated position **do**

If $\vartheta\left(\nabla_{fea(n,i)}\right) < \vartheta\left(\nabla_{fea(n,i-1)}\right)$

Update $\nabla_{fea(n,i)}$ as best-fit position

Else

Keep previous solution $\nabla_{fea(n,i-1)}$

End if

Ed for

Update new position using hunting strategy $\nabla_{fea(n,i)}$ //exploitation

Update best solutions

Set $i = i + 1$

End while

Return selected features $\nabla_{sf(n)}$

End

The best solutions were filtered in each iteration by using the exploration and exploitation phases after updating the position of all green anacondas. In this way, the optimal features are selected by utilizing the MM-GAO algorithm, and the selected features are given as,

$$\nabla_{sf(n)} = \begin{cases} \nabla_{fea(n,i)} & \vartheta\left(\nabla_{fea(n,i)}\right) < \vartheta\left(\nabla_{fea(n,i-1)}\right) \\ \nabla_{fea(n,i-1)} & Otherwise \end{cases} \quad (24)$$

Here, the total number of selected features is signified as $\nabla_{sf(n)}$.

E. Classification

For classification using anLRKSA-CNN, the selected features from the above phase $\left(\nabla_{sf(n)}\right)$ are utilized. CNN is selected as it is really efficient for image classification since the dimensionality reduction concept suits the huge number of parameters in an image. But, fixed kernel size utilization causes the loss of significant information hidden in unexplored regions, and the learning process is slowed down by the activation function utilized in the convolution layer owing to the biased shift effect. For overcoming this, the traditional CNN includes Linear Regression-based modeling of kernel size and Scaled Lineartanh (SLT) activation function for the convolution layer. Fig. 2 displays LRKSA-CNN's architecture.

Initially, the extracted features were inputted to the convolution layer. The convolution layer seeks to learn feature representations utilizing numerous convolution kernels and deploying the activation function.

The kernel size in the proposed system is determined utilizing the linear regression model. The LR approach provides the combination of kernels for the input points utilizing scalar-valued estimator as,

$$\delta_{lr(k)} = \psi\left(\sum_{k} \delta_k\left(\nabla_{sf(n)}, \nabla_{sf}\right)\gamma_k\right) \quad (25)$$

Here, the linear combination of kernels produced utilizing the scalar-valued function (ψ) is signified as $\delta_{lr(k)}$, δ_k depicts the scaled kernel at given input points $(\nabla_{sf(n)}, \nabla_{sf})$, and γ_k is the scaling factor.

The feature value obtained from the convolution layer is signified as,

$$\nabla_{con(n,k)} = \nabla_{sf(n)} * \theta_{\delta_{lr(k)}} + \varepsilon_k \quad (26)$$

Where, the convolution layer's output is signified as $\nabla_{con(n,k)}$, the weight vector of the k^{th} kernel is depicted as $\theta_{\delta_{lr(k)}}$, and ε_k signifies the bias term. Next, the activation function detects non-linear features by adding non-linearities to the output. It can be given as,

$$\nabla_{non-lin(n,k)} = \begin{cases} g.\left(\nabla_{con(n,k)}\right)^g & if\left(\nabla_{con(n,k)} < 0\right) \\ g * \exp\left(\nabla_{con(n,k)} - 1\right) & if\left(\nabla_{con(n,k)} > 0\right) \end{cases} \quad (27)$$

Here, $\nabla_{non-lin(n,k)}$ are the detected non-linear features, and g is the learnable parameter. Kernel size variation and SLT activation mechanisms have some advantages, such as reducing the model's complexity to explore each data point and making the network experience a faster learning process utilizing the activation function's calculations.

Next, the input's dimensionality is reduced by the pooling layer with the help of the max pooling technique. The pooling layer's feature maps are obtained as,

$$\nabla_{poo(n,k)} = \nu_{\max}\left(0, \nabla_{non-lin(n,k)}\right) \quad (28)$$

Here, the pooling layer's output is signified as $\nabla_{poo(n,k)}$, and the max pooling operation is depicted as $\nu_{\max}$.The final feature maps are flattened and fed to the fully connected layer after some executions of convolution and pooling operations to produce class scores from the activations for classification.

To produce the total number of output nodes, the softmax function is utilized in the output layer. The softmax function is derived as,

$$\xi_{(DR,NDR)} = \frac{\exp\left(\nabla_{flt(n)}\right)}{\sum_n \exp\left(\nabla_{flt(n)}\right)} \quad (29)$$

Here, $\nabla_{flt(n)}$ indicates the flattened one-dimensional feature vector, $\xi_{(DR,NDR)}$ signifies the softmax output that classifies the input images into two categories, namely patents with DR $\xi_{(DR)}$ and without DR $\xi_{(DR)}$. The classification process utilizing LRKSA-CNN is explained in Algorithm 2.

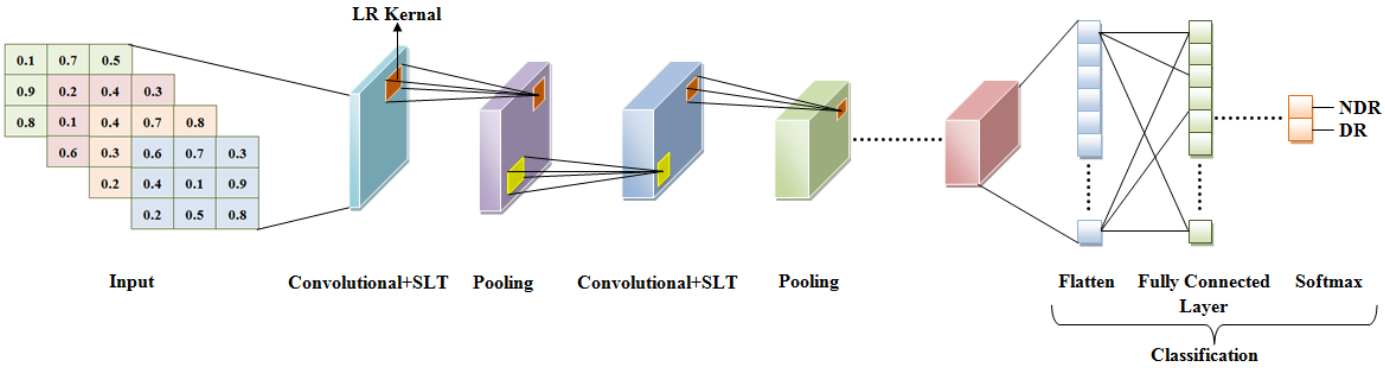


Fig. 2. Architecture of proposed LRKSA-CNN.

Algorithm .2. Classification using LRKSA-CNN

Input: Selected features $\nabla_{sf(n)}$

Output: Classified results $\xi_{(DR,NDR)}$

Begin

Initialize input $\nabla_{sf(n)}$, number of layers, model parameters $\theta_{\delta lr(k)}$, ε_k

For each layer **do**

Generate linear combination of kernels $\theta_{\delta lr(k)}$

For each kernel size **do** // convolution layer

Compute feature maps $\left(\nabla_{con(n,k)}\right)$

Detect non-linearities $\left(\nabla_{non-lin(n,k)}\right)$

If $\left(\nabla_{con(n,k)} < 0\right)$

$$\nabla_{non-lin(n,k)} = g.\left(\nabla_{con(n,k)}\right)^g$$

Else

$$\nabla_{non-lin(n,k)} = g * \exp\left(\nabla_{con(n,k)} - 1\right)$$

End if

End for

Compute pooled feature map $\nabla_{poo(n,k)}$

End for

Generate one dimensional feature vector $\nabla_{flt(n)}$

Return classified results $\xi_{(DR,NDR)}$

End

F. Stages Identification

From the classified results, the images with DR $\left(\xi_{(DR)}\right)$ are utilized for the formulation of grading to differentiate its severity into 5 stages, including severe NPDR, moderate NPDR, Mild NPDR, proliferative DR, and referable DR. Here, the KM-ANFIS is utilized for stages' identification. The ANFIS is selected since it has the advantage of having both numerical and linguistic knowledge. However, the membership function utilized in the traditional ANFIS causes the loss of interpretability in larger inputs. Thus, the Krusinka membership function is used here.

The features extracted from the retinal layers and visual network areas of DR-diagnosed images $\xi_{(DR)}$ were merged and used by the KM-ANFIS technique for stage identification. The KM-ANFIS structural design utilizes these features to form fuzzy if-then rules for grading.

Primarily, each node in the initial layer passes the input $\phi_{inp} = \left\{\xi_{(DR)}, \nabla_{fea \in R_{segO(n)}}, \nabla_{fea \in R_{segM(n)}}\right\}$ to the subsequent layers. For generating membership grades to the input, the nodes in layer 2 have a membership function. The Krusinka Membership function utilized in the proposed system is calculated as,

$$KMF\left(\phi_{inp}\right) = \frac{1}{2} + \frac{1}{\pi} arc\tan\left(\frac{\phi_{inp} - u}{v}\right) \quad (30)$$

Here, $KMF\left(\phi_{inp}\right)$ is the membership function, and (u, v) are the parameters set to change the membership function's shape. The KM membership function better interprets all the inputs in a fuzzy set by its ability to capture the interaction between variables and provide a reasonably smooth transition.

In Layer 3**,** each rule's firing strength $\left(fs_{(i)}\right)$ is computed utilizing fixed nodes through multiplication. Every single node output signifies the rule's firing strength; thus, the fuzzy AND operation is performed by the nodes.

$$fs_{(i)} = KMF.\nabla_{fea \in R_{segO(n)}} \times KMF.\nabla_{fea \in R_{segM(n)}} \quad (31)$$

Rule 1:If $\nabla_{fea \in R_{segO(n)}}$ is ς_p and $\nabla_{fea \in R_{segM(n)}}$ is ξ_a, then

$$\Delta_i = h_i \nabla_{fea \in R_{segO(n)}} + f_i \nabla_{fea \in R_{segM(n)}} + \sigma_i \quad (32)$$

Rule 2:If $\nabla_{fea \in R_{segO(n)}}$ is ς_a and $\nabla_{fea \in R_{segM(n)}}$ is ξ_p, then

$$\Delta_{i+1} = h_{i+1} \nabla_{fea \in R_{segO(n)}} + f_{i+1} \nabla_{fea \in R_{segM(n)}} + \sigma_{i+1} \quad (33)$$

Here, ς_p, ξ_a, ς_a and ξ_p indicate the fuzzy sets, h_i, f_i, σ_i, h_{i+1}, f_{i+1} and σ_{i+1} values are the parameter set, and Δ is a first-order polynomial and signifies the fuzzy inference system's outputs.

Layer 4 computes the ratio of each rule's firing strength to the sum of all rules' firing strengths. It can be given as,

$$\overline{fs_i} = \frac{fs_i}{fs_1 + fs_2}, \quad i = 1, 2 \quad (34)$$

There is only one node in the final output layer, which computes the sum of every output from the nodes of layer 5 for producing the overall KM-ANFIS output.

$$O_{0to4} = \frac{\sum_{i=1}^{n} \overline{fs_i} \, \Delta_i}{\sum_{i=1}^{n} fs_i} \quad (35)$$

Where, O_{0to4} is the overall output of the KM-ANFIS. By utilizing this, KM-ANFIS grades the DR severity into five stages. This grading scheme provides a more precise interpretation of disease stages for accurately evaluating the DR's effects in advance.

IV. Results and Discussion

Here, the proposed DR detection approach's effectiveness is assessed by analogizing its results with prevailing techniques. In the working platform of PYTHON, the proposed technique is deployed and trained by utilizing the OCT and fMRI images gathered from publicly available sources. Fig. 3 and Fig. 4 display the sample outcomes obtained for the input images.

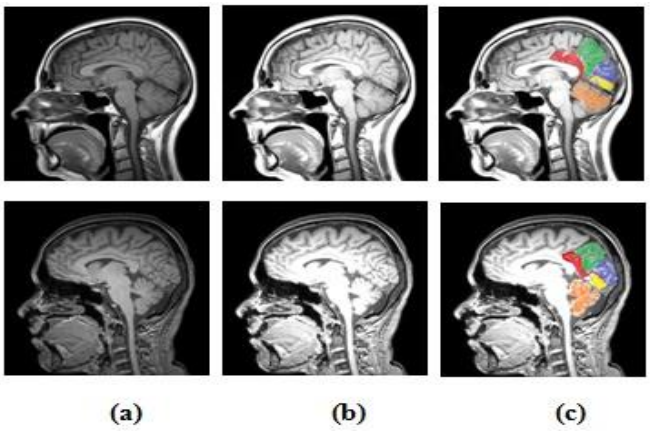


Fig. 3. (a) Input OCT images, (b) contrast-enhanced images using CLAHE, and (c) segmented regions using the ICSW-WS.

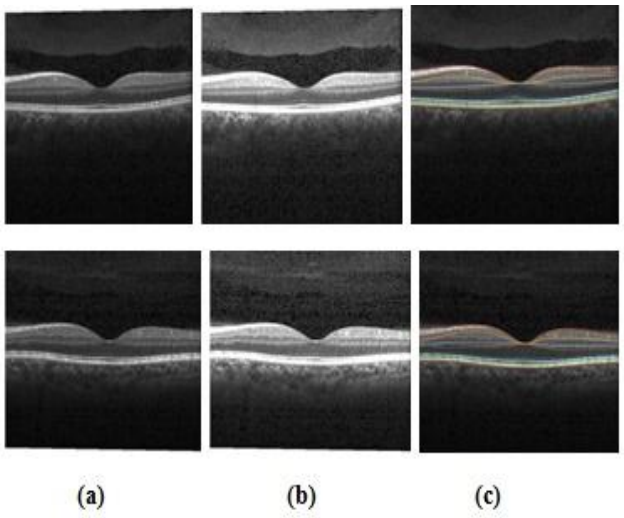


Fig. 4. (a) Input MRI images, (b) contrast-enhanced images using CLAHE, and (c) segmented regions using the ICSW-WS.

A. Performance Analysis

This section concentrates on analyzing the outcomes of different phases, namely KM-ANFIS-based stages identification, LRKSA-CNN-based classification, ICSW-based segmentation, and MM-GAO-based feature selection for quantitatively evaluating the proposed system. The performance of LRKSA-CNN is analyzed below:

TABLE I. Performance Measure of Classification Accuracy

Methods	Accuracy (%)
Proposed LRKSA-CNN	97.99
CNN	92.34
DNN	93.26
ANN	90.84
RNN	88.03

The classification accuracy of the proposed LRKSA-CNN and traditional CNN, DNN, ANN, and Recurrent Neural Network (RNN) techniques is displayed in Table I. This table reveals that the proposed technique's classification accuracy enhanced by 4.65% than the prevailing CNN. Thus, the use of LR-based kernels *includes large information with different dimensions and* efficiently increases the classification accuracy.

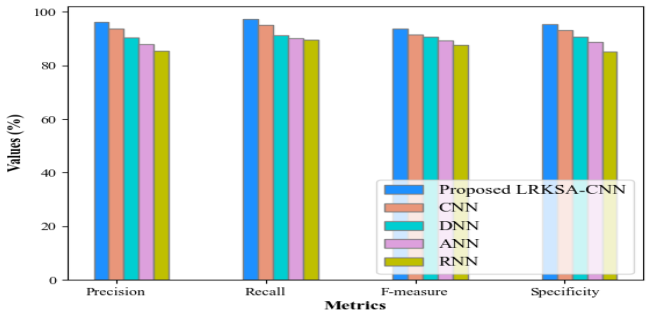


Fig. 5. Performance Analysis of LRKSA-CNN.

The proposed and conventional classification techniques' performance regardingprecision, F-measure, recall, and specificity is displayed in Fig. 5. The comparison shows that the proposed system outperforms the prevailing approaches in all metrics like precision (96.15%), recall (97.28%), F-measure (93.81%), and specificity (95.27%). This analysis reveals that by capturing larger patterns and accelerating its learning rate, the performance of the proposed model for DR classification was greatly improved with the inclusion of an LR-centric kernel and SA function.

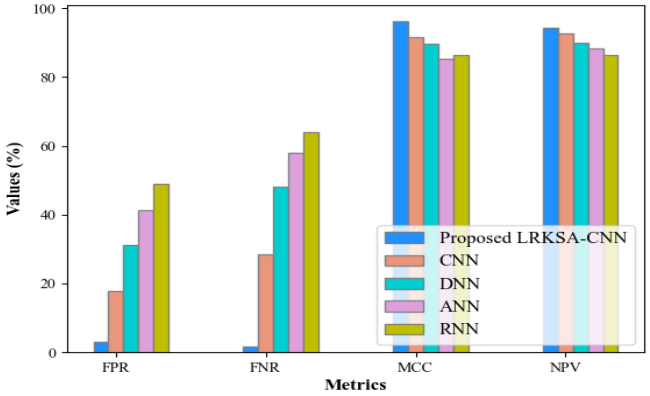


Fig. 6. Comparision of the proposed system with other classification methods.

The proposed LR-KSA technique's performance is analogized with the prevailing techniques regarding False Negative Rate (FNR), False Positive Rate (FPR), Negative Predictive Value (NPV), and Mathews Correlation Coefficient (MCC) in Fig. 6. Attaining lower values for FNR and FPR and higher values for MCC and NPV signifies better performance of the model. By the way, the proposed technique proffers superior performance regarding FNR (3.01%), FPR (1.63%), MCC (96.17%), and NPV (94.36%) analogized to the prevailingtechniques with the use of the LRKSA technique.

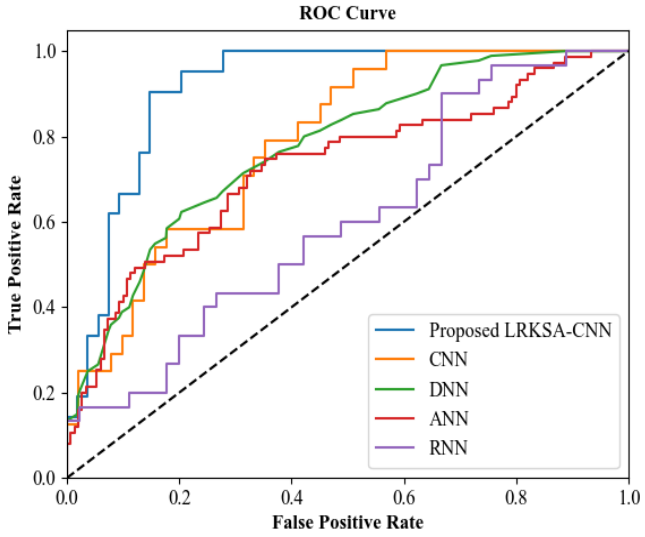


Fig. 7. ROC curve analysis.

The Receive Operative Characteristics (ROC) curve for the proposed and conventional techniques is displayed in Fig. 7. The ROC curve is for providing unbiased outcomes plotted between the true positive rate and the true negative rate. The ROC curve in Fig. 7 displays that when compared to prevailing techniques, the proposed system more accurately differentiated all classes with higher accuracy values.

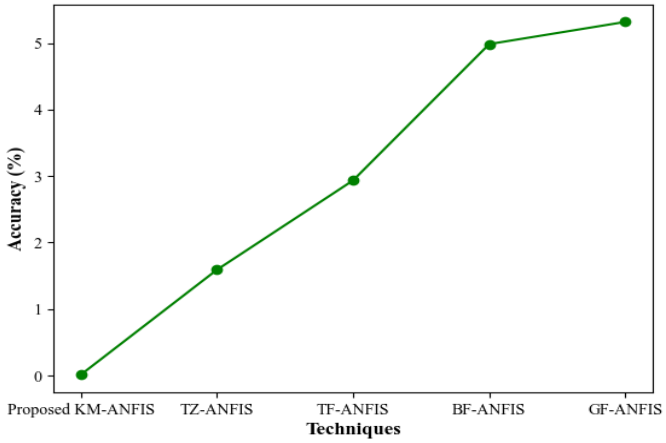


Fig. 8. Accuracy of stages identification.

The accuracy achieved by the proposed KM-ANFIS and prevailing Trapezoidal-based ANFIS (TZ-ANFIS), Triangular Function-based ANFIS (TF-ANFIS), Gaussian Function-based ANFIS (GF-ANFIS), and Bell-shaped Function-based ANFIS (BF-ANFIS) techniques are exhibited in Fig. 8. Here, the proposed technique provides higher accuracy than the prevailing techniques. The proposed system's accuracy is improved by 2.61% than the existing methods. Hence, the analysis concludes that the KM function has the efficiency to better interpret all the inputs and precisely differentiate DR stages.

TABLE II. PERFORMANCE COMPARISON OF SEGMENTATION RESULTS BASED ON VOI

Methods	VOI
Proposed ICSW-WS	0.0135
WS	0.0185
RG	0.0214
AC	0.0255
RSM	0.0264

The Variation Of Index (VOI) attained by the proposed ICSW-WS and prevailing WS, Region Growing (RG), Active Contour (AC), and Region Split and Merge (RSM) techniques are analogized in Table II. It displays that the proposed technique's VOI is much lesser by 0.005, 0.079, 0.012, and 0.013 than the prevailing WS, RG, AC, and RSm techniques. Hence, the analysis clearly defines that ICSW-centric marker identification resulted in the accurate segmentation of meaningful regions.

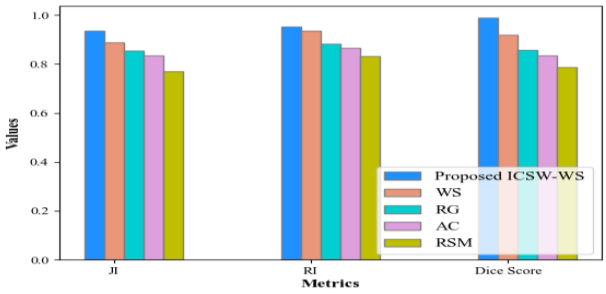


Fig. 9. Comparision of segmentation methods.

The proposed and prevailing segmentation techniques' Jaccard Index (JI), Rand Index (RI), and Dice Score (DS) are examined in Fig. 9. While comparing with prevailing techniques, the proposed technique's JI, RI, and DS are enhanced by 1.69%, 4.57%, and 6.83% than the prevailing WS technique.This indicates that the proposed ICSW-WS technique's performance was greatly influenced by the computation of ICSW weights forsegmenting small unique regions more accurately than the prevailing techniques.

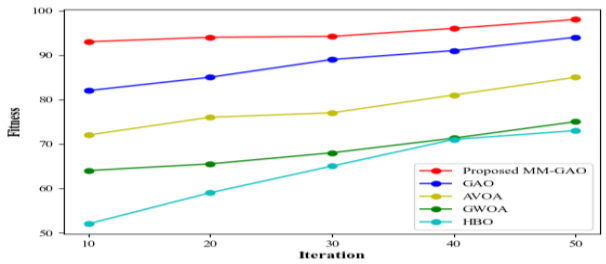


Fig. 10. Fitness vs. iteration.

The fitness values of MM-GAO and prevailing GAO, African Vultures Optimization Algorithm (AVOA), Gray Wolf Optimization Algorithm (GWOA), and Honey Badger Optimization Algorithm (HBO) techniques for 10,20,30,40, and 50 iterations are exhibited in Fig. 10. For a maximum of 50 iterations, the fitness attained by the proposed system is 98, while the prevailing techniques attained less than the proposed technique. Thus, the proposed technique is made to attain superior fitness in the given time than the prevailing techniques by the MM-centric probability calculation.

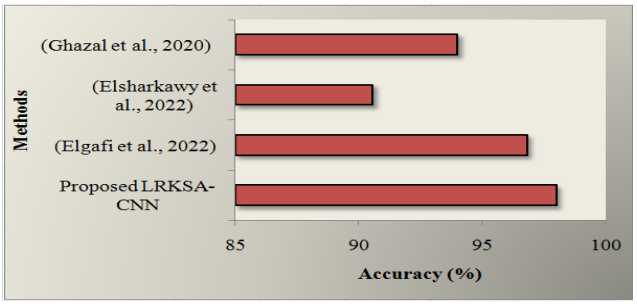


Fig. 11. Comparative analysis.

Fig. 11 analogizes the proposed LRKSA-CNN technique with the prevailing systems recommended by [17], [23], and [25] in Section II. Analyzing Fig. 11 shows that the proposed technique achieves superior performance than the prevailing techniques. Even though the prevailing techniques attained superior accuracy, they didn't concentrate on detecting DR centered on the pathophysiological changes of visual network areas.This displays the proposed system's superiority over the baseline works.

V. CONCLUSION

An effective DR detection model based on LRKSA-CNN using multimodal images is proposed in this work. The proposed system's major goal is early DR diagnosis by examining both invisible pathologies inside the retina and microstructural changes in the visual pathway. In addition, the proposed system concentrates on classifying various phases of DR. The proposed LRKSA-CNN system is trained and evaluated utilizing MRI and OCT images. The performance analysis analogizes the proposed LRKSA-CNN and KM-ANFIS techniques with the prevailing techniques. The outcomes exhibited that the detection accuracy of LRKSA-CNN enhanced to 97.99% and KM-ANFIS to 97.46%.The overall analysis of outcomes exposed that for detecting DR and classifying all stages of DR, the proposed technique performs superior compared to the prevailing techniques. The proposed model's potential will be analyzed in the future by merging other imaging modalities. Moreover, for improving the early DR diagnosis, the blood flow changes of retinal blood vessels will be considered.

DECLARATIONS

Funding: The authors declare that no funds, grants, or other supports were received during the preparation of this manuscript.

Competing interests: The authors have no relevant interests to disclose.

Open access : The authors declare that no funding provided for Open Access.

Author Contributions: All authors are agreed to publish the manuscript.

Availability of data and material: Not applicable.

REFERENCES

[1] S. K. R. Meruva, V. G. S. Tulasi, N. Vinnakota, & V. Bhavana, "Risk Level Prediction of Diabetic Retinopathy based on Retinal Images using Deep Learning Algorithm", *Application 4th International Conference on Innovative Data Communication Technology and Application*, vol. *215*, pp. 722–730, 2022, https://doi.org/10.1016/j.procs.2022.12.074.

[2] N. Gundluru, D.S. Rajput, K. Lakshmanna, R. Kaluri, M. Shorfuzzaman, M. Uddin, , & M. A. Rahman Khan, "Enhancement of Detection of Diabetic Retinopathy Using Harris Hawks Optimization with Deep Learning Model", *Computational Intelligence and Neuroscience*, pp. 1–13, 2022, https://doi.org/10.1155/2022/8512469.

[3] P. Saranya, S. Prabakaran, R. Kumar, & E. Das, "Blood vessel segmentation in retinal fundus images for proliferative diabetic retinopathy screening using deep learning", *Visual Computer*, vol. 8, no. 3, pp. 977–992, 2022, https://doi.org/10.1007/s00371-021-02062-0

[4] S. Subramanian, S. Mishra, S. Patil, K. Shaw, & E. Aghajari, "Machine Learning Styles for Diabetic Retinopathy Detection: A Review and Bibliometric Analysis", *Big Data and Cognitive Computing*, vol. 6, no. 4, pp. 1–31, 2022, https://doi.org/10.3390/bdcc6040154.

[5] J. Jaskari, J. Sahlsten, T. Damoulas, J. Knoblauch, S. Sarkka, L. Karkkainen, K. Hietala, & K. K. Kaski, "Uncertainty-Aware Deep Learning Methods for Robust Diabetic Retinopathy Classification", *IEEE Access*, vol. 10, pp. 76669–76681, 2022, https://doi.org/10.1109/ACCESS.2022.3192024.

[6] H. A. H. Mahmoud, "Diabetic Retinopathy Progression Prediction Using a Deep Learning Model", *Aximos*, vol. 11, pp. 1–17, 2022.

[7] T. M. Usman, Y. K. Saheed, D. Ignace, & A. Nsang, "Diabetic Retinopathy Detection Using Principal Component Analysis Multi-Label Feature Extraction And Classification", *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 78–88, 2023, https://doi.org/10.1016/j.ijcce.2023.02.002.

[8] C. Y. Tsai, C. T. Chen, G. A. Chen, C. F. Yeh, C. T. Kuo, Y. C. Hsiao, H. Y. Hu, I. L. Tsai, C. H. Wang, J. R. Chen, S. C. Huang, T. C. Lu, & L. C. Woung, "Necessity of Predict Diabetic Retinopathy", *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, pp. 1–12, 2022, https://doi.org/10.3390/ijerph19031204.

[9] I. Hossain, S. Puppala, & S. Talukder, "Collaborative Differentially Private Federated Learning Framework for the Prediction of Diabetic Retinopathy", *IEEE 2nd International Conference on AI in Cybersecurity*, *ICAIC*, pp. 1–6, 2023, https://doi.org/10.1109/ICAIC57335.2023.10044122.

[10] M. Tariq, V. Palade, Y. L. Ma, & A. Altahhan, "Diabetic Retinopathy Detection Using Transfer and Reinforcement Learning with Effective Image Preprocessing and Data Augmentation Techniques", *Intelligent Systems Reference Library*, vol. 236, pp. 1-30, 2023, https://doi.org/10.1007/978-3-031-22371-6_3.

[11] G. Alwakid, W. Gouda, & M. Humayun, "Deep Learning-Based Prediction of Diabetic Retinopathy Using CLAHE and ESRGAN for Enhancement". *Healthcare*, vol. 11, no. 6, pp. 1–17, 2023, https://doi.org/10.3390/healthcare11060863.

[12] B. Dong, X. Wang, X. Qiang, F. Du, L. Gao, Q. Wu, G. Cao, & C. Dai. "A Multi-Branch Convolutional Neural Network for Screening and Staging of Diabetic Retinopathy Based on Wide-Field Optical Coherence Tomography Angiography", *Irbm*, *43*(6), pp. 614–620, 2022. https://doi.org/10.1016/j.irbm.2022.04.004

[13] K. Gunasekaran, R. Pitchai, G. K. Chaitanya, D. Selvaraj, S. Annie Sheryl, H. S. Almoallim, S. A. Alharbi, S. S. Raghavan, & B. G. Tesemma, "A Deep Learning Framework for Earlier Prediction of Diabetic Retinopathy from Fundus Photographs", *BioMed Research International*, pp. 1-15, 2022, https://doi.org/10.1155/2022/3163496.

[14] Y. Bhawarkar, K. Bhure, V. Chaudhary, & B. Alte, "Diabetic Retinopathy Detection From Fundus Images Using Multi-Tasking Model With EfficientNet B5", *ITM Web of Conferences*, vol. 44, pp. 1-6, 2022, https://doi.org/10.1051/itmconf/20224403027.

[15] P. Zang, T. T. Hormel, T. S. Hwang, S. T. Bailey, D. Huang, & Y. Jia, "Deep-Learning–Aided Diagnosis of Diabetic Retinopathy, Age-Related Macular Degeneration, and Glaucoma Based on Structural and Angiographic OCT", *American Academy of Ophthalmology Science*, vol. 3, no. 1, pp. 1-9, 2023, https://doi.org/10.1016/j.xops.2022.100245.

[16] M. Sakthi Sree Devi, S. Ramkumar, S. Vinuraj Kumar, & G. Sasi, "Detection of diabetic retinopathy using OCT image", *Materials Today: Proceedings*, vol. 47, pp. 185–190, 2021, https://doi.org/10.1016/j.matpr.2021.04.070.

[17] M. Elgafi, A. Sharafeldeen, A. Elnakib, A. Elgarayhi, N. S. Alghamdi, M. Sallah, & A. El-Baz, "Detection of Diabetic Retinopathy Using Extracted 3D Features from OCT Images", *Sensors*, vol. 22, no. 20, pp. 1–13, 2022, https://doi.org/10.3390/s22207833.

[18] B. Menaouer, Z. Dermane, N. El Houda Kebir, & N. Matta, "Diabetic Retinopathy Classification Using Hybrid Deep Learning Approach", *SN Computer Science*, vol. 3, no. 5, pp. 1-16, 2022, https://doi.org/10.1007/s42979-022-01240-8.

[19] S. S. Mondal, N. Mandal, K. K. Singh, A. Singh, & I. Izonin, "EDLDR: An Ensemble Deep Learning Technique for Detection and Classification of Diabetic Retinopathy", *Diagnostics*, vol. 13, no. 1, pp. 1–14, 2023, https://doi.org/10.3390/diagnostics13010124.

[20] M. Murugappan, N. B. Prakash, R. Jeya, A. Mohanarathinam, G. R. Hemalakshmi, & M. Mahmud, "A novel few-shot classification framework for diabetic retinopathy detection and grading", *Measurement: Journal of the International Measurement Confederation*, vol. 200, pp. 1-14, 2022, https://doi.org/10.1016/j.measurement.2022.111485.

[21] M. M. Butt, D. N. F. A. Iskandar, S. E. Abdelhamid, G. Latif, & R. Alghazo, "Diabetic Retinopathy Detection from Fundus Images of the Eye Using Hybrid Deep Learning Features", *Diagnostics*, vol. 12, no. 7, 2022 https://doi.org/10.3390/diagnostics12071607.

[22] A. Erciyas, & N. Barişçi, "An Effective Method for Detecting and Classifying Diabetic Retinopathy Lesions Based on Deep Learning", *Computational and Mathematical Methods in Medicine*, pp. 1-13, 2021, https://doi.org/10.1155/2021/9928899.

[23] M. Elsharkawy, A. Sharafeldeen, A. Soliman, F. Khalifa, M. Ghazal, E. El-Daydamony, A. Atwan, H. S. Sandhu, & A. El-Baz, "A Novel Computer-Aided Diagnostic System for Early Detection of Diabetic Retinopathy Using 3D-OCT Higher-Order Spatial Appearance Model", *Diagnostics*, vol. 12, no. 2, pp. 1-14, 2022, https://doi.org/10.3390/diagnostics12020461.

[24] W. L. Alyoubi, M. F. Abulkhair, & W. M. Shalash, "Diabetic retinopathy fundus image classification and lesions localization system using deep learning", *Sensors*, vol. 21, no. 11, pp. 1–22, 2021, https://doi.org/10.3390/s21113704.

[25] M. Ghazal, S. S. Ali, A. H. Mahmoud, A. M. Shalaby, & A. El-Baz, "Accurate Detection of Non-Proliferative Diabetic Retinopathy in Optical Coherence Tomography Images Using Convolutional Neural Networks", *IEEE Access*, vol. 8, pp. 34387–34397, 2020, https://doi.org/10.1109/ACCESS.2020.2974158.

[26] S. K. Sahu, A. D. Sawarkar, A. K. Sahu, A. Anjikar, A. P. Bhopale and J. Chakole, "Analysis of Transfers Learning Techniques for Early Detection and Grading of Diabetic Retinopathy on Retinal Images," *2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, Mumbai, India, pp. 1-6, 2023, doi: 10.1109/ICACTA58201.2023.10392325.

[27] S. Goswami, K. Ashwini and R. Dash, "Grading of Diabetic Retinopathy using iterative Attentional Feature Fusion (iAFF)," *International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, pp. 1-5, 2023, doi: 10.1109/ICCCNT56998.2023.10307892.

[28] N. Z. Abidin and A. Ritahani Ismail, "Federated Deep Learning for Automated Detection of Diabetic Retinopathy," *IEEE 8th International Conference on Computing, Engineering and Design (ICCED)*, Sukabumi, Indonesia, pp. 1-5, 2022, doi: 10.1109/ICCED56140.2022.10010636.

[29] O. F. Gurcan, O. F. Beyca and M. Olucoglu, "Diagnosis of Diabetic Retinopathy with Transfer Learning and Metaheuristic Algorithms," *Innovations in Intelligent Systems and Applications Conference (ASYU)*, Sivas, Turkiye, pp. 1-6, 2023, doi: 10.1109/ASYU58738.2023.10296811.

[30] M. Jahiruzzaman and A. B. M. Aowlad Hossain, "Detection and classification of diabetic retinopathy using K-means clustering and fuzzy logic," *18th International Conference on Computer and Information*

Technology (ICCIT), Dhaka, Bangladesh, pp. 534-538, 2015, doi: 10.1109/ICCITechn.2015.7488129.

[31] M. Ghazal, S. S. Ali, A. H. Mahmoud, A. M. Shalaby and A. El-Baz, "Accurate Detection of Non-Proliferative Diabetic Retinopathy in Optical Coherence Tomography Images Using Convolutional Neural Networks," in *IEEE Access*, vol. 8, pp. 34387-34397, 2020, doi: 10.1109/ACCESS.2020.2974158.

[32] H. Mustafa, S. F. Ali, M. Bilal and M. S. Hanif, "Multi-Stream Deep Neural Network for Diabetic Retinopathy Severity Classification Under a Boosting Framework," in *IEEE Access*, vol. 10, pp. 113172-113183, 2022, doi: 10.1109/ACCESS.2022.3217216.

[33] K. Aurangzeb, R. S. Alharthi, S. I. Haider and M. Alhussein, "Systematic Development of AI-Enabled Diagnostic Systems for Glaucoma and Diabetic Retinopathy," in *IEEE Access*, vol. 11, pp. 105069-105081, 2023, doi: 10.1109/ACCESS.2023.3317348.

[34] M. Nur-A-Alam, M. M. K. Nasir, M. Ahsan, M. A. Based, J. Haider and S. Palani, "A Faster RCNN-Based Diabetic Retinopathy Detection Method Using Fused Features From Retina Images," in *IEEE Access*, vol. 11, pp. 124331-124349, 2023, doi: 10.1109/ACCESS.2023.3330104.

[35] R. Pires, H. F. Jelinek, J. Wainer, S. Goldenstein, E. Valle and A. Rocha, "Assessing the Need for Referral in Automatic Diabetic Retinopathy Detection," *in IEEE Transactions on Biomedical Engineering*, vol. 60, no. 12, pp. 3391-3398, Dec. 2013, doi: 10.1109/TBME.2013.2278845.

[36] M. T. Islam, H. R. H. Al-Absi, E. A. Ruagh and T. Alam, "DiaNet: A Deep Learning Based Architecture to Diagnose Diabetes Using Retinal Images Only," in *IEEE Access*, vol. 9, pp. 15686-15695, 2021, doi: 10.1109/ACCESS.2021.3052477.

[37] D. Maji, A. K. Dhara, S. Maiti and G. Sarkar, "Efficient Net Enriched Model for Implementing the Grading of Diabetic Retinopathy Based on Retinal Blood Vessel Tortuosity," *IEEE 3rd Applied Signal Processing Conference (ASPCON)*, pp. 131-136, 2023, doi: 10.1109/ASPCON59071.2023.10396362.

[38] X. Wang *et al*., "UD-MIL: Uncertainty-Driven Deep Multiple Instance Learning for OCT Image Classification," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3431-3442, 2020, doi: 10.1109/JBHI.2020.2983730.

[39] M. M. Farag, M. Fouad and A. T. Abdel-Hamid, "Automatic Severity Classification of Diabetic Retinopathy Based on DenseNet and Convolutional Block Attention Module," in *IEEE Access*, vol. 10, pp. 38299-38308, 2022, doi: 10.1109/ACCESS.2022.3165193.

[40] T. Liu *et al*., "A Novel Diabetic Retinopathy Detection Approach Based on Deep Symmetric Convolutional Neural Network," in *IEEE Access*, vol. 9, pp. 160552-160558, 2021, doi: 10.1109/ACCESS.2021.3131630.

[41] G. T. Reddy *et al*., "Analysis of Dimensionality Reduction Techniques on Big Data," in *IEEE Access*, vol. 8, pp. 54776-54788, 2020, doi: 10.1109/ACCESS.2020.2980942.

[42] E. O. Rodrigues, A. Conci and P. Liatsis, "ELEMENT: Multi-Modal Retinal Vessel Segmentation Based on a Coupled Region Growing and Machine Learning Approach," in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 12, pp. 3507-3519, 2020, doi: 10.1109/JBHI.2020.2999257.

[43] S. Wang *et al*., "Localizing Microaneurysms in Fundus Images Through Singular Spectrum Analysis," in *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 990-1002, 2017, doi: 10.1109/TBME.2016.2585344.

[44] K. Shankar, Y. Zhang, Y. Liu, L. Wu and C. H. Chen, "Hyperparameter Tuning Deep Learning for Diabetic Retinopathy Fundus Image Classification," in *IEEE Access*, vol. 8, pp. 118164-118173, 2020, doi: 10.1109/ACCESS.2020.3005152.

[45] B. Yang, T. Li, H. Xie, Y. Liao and Y. P. P. Chen, "Classification of Diabetic Retinopathy Severity Based on GCA Attention Mechanism," in *IEEE Access*, vol. 10, pp. 2729-2739, 2022, doi: 10.1109/ACCESS.2021.3139129.

[46] I. Usman and K. A. Almejalli, "Intelligent Automated Detection of Microaneurysms in Fundus Images Using Feature-Set Tuning," in *IEEE Access*, vol. 8, pp. 65187-65196, 2020, doi: 10.1109/ACCESS.2020.2985543.

[47] B. Abdillah, A. Bustamam and D. Sarwinda, "Classification of diabetic retinopathy through texture features analysis," *International Conference on Advanced Computer Science and Information Systems (ICACSIS)*, pp. 333-338, 2017, doi: 10.1109/ICACSIS.2017.8355055.

[48] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen and C. I. Sánchez, "Fast Convolutional Neural Network Training Using Selective Data Sampling: Application to Hemorrhage Detection in Color Fundus Images," in *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1273-1284, 2016, doi: 10.1109/TMI.2016.2526689.

[49] A. Krestanova, J. Kubicek and M. Penhaker, "Recent Techniques and Trends for Retinal Blood Vessel Extraction and Tortuosity Evaluation: A Comprehensive Review," in *IEEE Access*, vol. 8, pp. 197787-197816, 2020, doi: 10.1109/ACCESS.2020.3033027.

[50] K. Wisaeng and W. Sa-Ngiamvibool, "Exudates Detection Using Morphology Mean Shift Algorithm in Retinal Images," in *IEEE Access*, vol. 7, pp. 11946-11958, 2019, doi: 10.1109/ACCESS.2018.2890426.

[51] X. Li *et al*., "Rotation-Oriented Collaborative Self-Supervised Learning for Retinal Disease Diagnosis," in *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2284-2294, 2021, doi: 10.1109/TMI.2021.3075244.

[52] G. Gupta, S. Kulasekaran, K. Ram, N. Joshi, M. Sivaprakasam and R. Gandhi, "Computer-assisted identification of proliferative diabetic retinopathy in color retinal images," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5642-5645, 2015, doi: 10.1109/EMBC.2015.7319672.

[53] R. Kommaraju and M. S. Anbarasi, "Detection of Diabetic Retinopathy (DR) Severity from Fundus Photographs using Swin Transformer," *IEEE 8th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pp. 1-6, 2023, doi: 10.1109/ICRAIE59459.2023.10468500.

[54] R. Gambhir, S. Bhardwaj, A. Kumar and R. Agarwal, "Severity Classification of Diabetic Retinopathy using ShuffleNet," 2021 *International Conference on Intelligent Technologies (CONIT)*, pp. 1-5, 2021, doi: 10.1109/CONIT51480.2021.9498569

AUTHORS' PROFILE

Rachna Kumari received Mtech Degree in computer science and engineering from Guru Jambheshwar University of Science and Technology. She is currently pursuing PhD Degree from GJU S&T hisar

Sanjeev Kumar received the Ph.D. degree in computer science and engineering from Guru Jambheshwar University of Science and Technology. He is currently working as Professor in GJU S&T hisar. He has authored more than 15 publications.

Sunila Godara received the Ph.D. degree in computer science and engineering from Guru Jambheshwar University of Science and Technology. She is currently working as Professor in GJU S&T hisar. She has authored more than 15 publications

Mining High Utility Itemset with Hybrid Ant Colony Optimization Algorithm

Keerthi Mohan[1], Anitha J[2]
Research Scholar[1], Professor[2],
Department of Computer Science and Engineering, RV Institute of Technology and Management, Bangalore, India

***Abstract*—A significant area of study within data mining is high-utility itemset mining (HUIM). The exponential problem of broad search space usually comes up while using traditional HUIM algorithms when the database size or the number of unique objects is huge. Evolutionary computation (EC) -based algorithms have been presented as an alternate and efficient method to address HUIM problems since they can quickly produce a set of approximately optimum solutions. In transactional databases, finding entire high-utility itemset (HUIs) still need a lot of time using EC-based methods. In order to deal with this issue, we propose a hybrid Ant colony optimization-based HUIM algorithm. Genetic operators' crossover is applied to the generated solution by the ant in the Ant Colony optimization algorithm. In this study, a single-point crossover is employed wherein, the crossover point is selected randomly and a mutation operator is applied by changing one or many random bits in a string. This technique requires less time to mine the same number of HUIs than state-of-the-art EC-based HUIM algorithms.**

Keywords*—Utility mining; high utility itemset; ant colony optimization; genetic algorithm; evolutionary computation

I. Introduction

The purpose of data mining is to uncover valuable insights or cognition from data which can aid in decision-making, prediction, and revealing hidden patterns or relationships within the data. Frequent itemset mining (FIM) is a crucial data mining approach aimed at identifying groups of items or elements that commonly occur together in large datasets. It is often applied in association rule learning, where the objective is to uncover relationships or patterns among items in transactional data, such as in market basket analysis. Although FIM mines itemsets based on the count of occurrence, the item's utility (profit) is not taken into consideration. To address this shortcoming, the concept of High Utility Itemset (HUI) was introduced. While traditional frequent itemset mining identifies itemsets based on their frequency of occurrence, HUIM takes into account the actual value or utility of items, which can include factors like profit, revenue, or user preference. An itemset is deemed high utility if its utility exceeds the user specified threshold. Here, a minimum utility value will be set, either manually or through a probabilistic function. Traditional HUIM algorithms encounter and have to deal with the exponential search space problem with increasing number of transactions and database items. Evolutionary algorithms such as genetic algorithms (GA), Ant Colony Optimization (ACO) algorithm, Particle swarm Optimization (PSO), and Artificial bee colony (ABC) algorithm are some efficient algorithms that solve the exponential search space problem. In order to find near-optimal solutions based on fitness functions, evolutionary algorithms effectively search enormous problem spaces. For mining HUIs beyond a minimum utility threshold, evolutionary-based HUIM algorithms are quicker than conventional precise approaches, although they still require a lot of time. Enhancing the diversity of the generated population is one way of reducing the time. Hybrid approach balances exploration and exploitation, handles sizably voluminous and unevenly distributed datasets preponderant, and acclimates well to dynamic environments. It outperforms standalone PSO, GA, or ACO in terms of haste, scalability, and solution quality. The Hybrid Ant Colony Optimization (GA-ACO) algorithm ameliorates HUI mining by cumulating the iterative refinement of ACO with the expeditious ecumenical search of GA, which expedites convergence and truncates the early-stage impuissance's of ACO.

II. Related Work

Agrawal and Srikant [1] explored association rules in large sales transaction databases. Agrawal, Imielinski, and Swami [2] used novel estimation and pruning techniques for efficient and fast mining of association rules. Tseng et al. [3] developed algorithms for mining high utility itemsets from a transactional database. They presented two algorithms, Utility-Pattern-Growth (UP-Growth) and also UP-Growth+, to carry out the mining of high utility itemset with a collection of effective approaches for pruning potential itemsets. They stored information associated with high utility itemsets into a tree-based data structure called "utility pattern tree" in such a way that candidate itemsets could be developed with only two scans of the database and obtained significant improvement in extracting high utility itemsets. Chan, Yang, and Shen [4] proposed a technique that targets mining the top-K high utility closed patterns directly aligned with a specific business objective. Their experimental results demonstrate that the algorithm does not require a user-defined minimum utility, making it practical and effective in real-world applications.

Yao, Hamilton, and Butz [5] examined the utility relationships among itemsets and identified two key properties: the utility bound property and the support bound property. They also developed a mathematical model for utility mining based on these principles. Ahmed et al. [6] introduced tree structures to efficiently conduct incremental and interactive high-utility pattern (HUP) mining. Their first tree structure, the Incremental HUP Lexicographic Tree (IHUPL-Tree), is organized by an item's lexicographic order, capturing incremental data without requiring any restructuring. The second tree structure, the IHUP Transaction FrequencyTree (IHUPTF-Tree), optimizes size by arranging items in descending order of transaction frequency. To

reduce mining time, the third structure, the IHUP Transaction-Weighted Utilization Tree (IHUPTWU-Tree), is based on the descending TWU (Transaction-Weighted Utilization) values of items. The authors demonstrated that these tree structures are highly efficient and scalable for incremental and interactive HUP mining.

In study [7], Liu and Qu introduced a data structure called the "utility-list" and introduced an algorithm named HUIMiner for mining high utility itemsets. The HUI-Miner algorithm introduces utility-lists to store utility and pruning information, enabling it to mine high utility itemsets without engendering candidate itemsets. This eliminates the computational overhead of traditional algorithms that process sizably voluminous numbers of low-utility candidates. By directly fixating on high utility itemsets, HUI-Miner evades the costly processes of itemset generation and utility computation. Performance comparisons with other state-of-the-art algorithms across multiple databases show that HUI-Miner offers better improvment in both runtime and memory utilization, making it a more efficient approach for mining high utility itemsets.

Fournier-Viger et al. [8] presented a new algorithm called FHM- Fast High-Utility Miner which improved the performance of HUI-Miner algorithm. FHM's key innovation lies in its strategic analysis of item cooccurrences, which substantially decreases the number of necessary join operations during the mining process. Testing on real-world datasets demonstrated FHM's efficiency, reducing join operations by as much as 95% and operating up to six times faster than its predecessor, HUI-Miner.

Zida et al. [9] introduced EFIM, a new algorithm for mining high-utility itemsets. EFIM's effectiveness arises from two main components: newly developed upper-bounds called the sub-tree utility and the local utility, and a streamlined array-based method named Fast Utility Counting. Their approach allows for linear time and space complexity when calculating these upper-bounds. To further enhance efficiency, EFIM incorporates techniques for database projection and transaction merging, also achieving linear time and space complexity. These methods significantly reduce the computational cost associated with database scans. Comprehensive experiments across diverse datasets demonstrated EFIM's superior performance.

Peng, Koh, and Riddle [10] introduced mHUIMiner, which leverages a tree-based framework to steer the process of itemset growth. This approach efficiently eliminates the need to examine non-existent database itemsets. mHUIMiner emerges as different from other methods by avoiding complex and computationally intensive pruning mechanisms.

One of the earliest works on mining high utility itemsets using evolutionary algorithms from transaction databases is using Genetic Algorithm (GA) proposed by Kannumuthu and Premalatha [11]. They proposed two approaches to mine high utility itemsets from transaction databases with or without specifying minimum utility threshold by using genetic algorithm. Results from experiments showed that authors' GA approaches accomplished better performance in terms of scalability and efficiency.

Lin et al. [12] proposed using discrete Particle Swarm Optimization (PSO) to represent particles as binary variables. They introduced HUIM-BPSO, an efficient PSO-based algorithm designed to discover High-Utility Itemsets (HUIs) effectively. Their algorithm employs the transaction-weighted utility (TWU) model to identify high-transaction-weighted utilization 1-itemsets (1-HTWUIs). They established that their method mitigated the combinatorial complexity typically encountered during the evolutionary process.

Lan, Hong, and Tseng [13] introduced a novel pattern category called high transaction-weighted utility itemsets. Three key factors are taken into consideration in this approach: individual item profits, item quantities within transactions, and the overall contribution of each transaction to the database. To uncover these high transaction-weighted utility itemsets, the authors developed a two-stage mining algorithm. They demonstrated their methods through experiments conducted on synthetic datasets, which revealed promising performance results.

A 2-phase algorithm was introduced by Liu, Liao, and Choudhary [15] which aimed at streamlining the high-utility itemset mining process. Their approach effectively minimized the candidate set while ensuring the comprehensive discovery of all high-utility itemsets. Their algorithm exhibited efficiency in both time and memory consumption across various database types - synthetic and real-world. Their method successfully handles large-scale databases that typically pose significant challenges for existing methods.

A new algorithm called HUIM-ACS for high-utility itemset (HUI) mining was introduced by Wu, Zhan, and Lin [16]. Their algorithm is an enhanced version of the standard ant colony optimization (ACO) technique called ant colony system (ACS). HUIM-ACS provides several benefits compared to existing methods like genetic algorithms and particle swarm optimization. It constructs solutions in a way that avoids impractical outcomes, maps the entire solution space to ensure comprehensive coverage, and uses pruning processes to improve efficiency. The algorithm also prevents redundant evaluations of solutions, saving computational resources. Experimentation on real-life datasets showed that HUIM-ACS outperforms other heuristic HUI mining algorithms in both the number of HUIs discovered and convergence speed. While most evolutionary algorithms can't guarantee finding the global optimum, HUIM-ACS's comprehensive exploration of the solution space could potentially address this limitation in high utility itemset mining.

Song and Huang [17] introduced Bio-HUI, a framework that adapts bio-inspired algorithms for high-utility itemset mining. Unlike traditional approaches, Bio-HUI selects discovered HUIs as targets for the next generation, enhancing population diversity. The proposed framework was implemented using genetic algorithms, particle swarm optimization and bat algorithms. Extensive testing on public datasets demonstrates that these Bio-HUI-based methods outperform current state-of-the-art algorithms in efficiency, result quality, and convergence speed. The authors established that their method shows significant promise in advancing the field of high-utility itemset mining by leveraging the strengths of bio-inspired computational techniques.

Nawaz et al. [18] proposed two new algorithms for high-utility itemset mining (HUIM): HUIM-HC based on Hill Climbing and HUIM-SA based on Simulated Annealing. These algorithms aim to focus on the limitations of existing evolutionary and heuristic methods, which are typically affected by long runtimes and may miss many high-utility itemsets (HUIs). Their techniques used efficient utility computation and search space pruning through bitmap transformation of the input database. Their algorithms also improved population diversity by using discovered HUIs as targets for subsequent generations, rather than simply maintaining current optimal values. Their experimentation using real-life datasets demonstrated that HUIM-HC and HUIM-SA outperform state-of-the-art heuristic and evolutionary HUIM algorithms in terms of speed. Their techniques represent a significant step forward in addressing the challenges of high-utility itemset mining, offering improved efficiency and effectiveness over existing methods.

Song and Nan [19] introduced a new high-utility itemset mining algorithm called HUIM-ACO, based on ant colony optimization. HUIM-ACO uses a constructive approach to produce candidate itemsets, represented as search paths. Pheromone values are stored in a matrix to guide the search process, and an efficient enumeration technique is used to discover more itemsets. Their experimentation showed that HUIM-ACO outperforms existing algorithms in terms of speed and the count of high-utility itemsets discovered.

Li et al. [20] also proposed a high-utility itemset mining algorithm called HUIM-ACO, based on ant colony optimization. HUIM-ACO uses a constructive approach to generate candidate itemsets, represented as search paths. Pheromone values are stored in a matrix to guide the search process, and an efficient enumeration technique is used to discover more itemsets. Experimental results demonstrated that HUIM-ACO outperforms existing algorithms in terms of both speed and the count of high-utility itemsets discovered.

Han et al. [22] introduced a new high-utility itemset mining algorithm that incorporates two key strategies: positional evolution based on the female elephant factor to reduce the search space and improve efficiency, and two-phase population diversity maintenance to prevent premature convergence. Their experimentation showed that this algorithm outperforms existing heuristic methods in terms of both speed and effectiveness.

In order to obtain a near-optimal solution based on fitness functions under a number of constraints, evolutionary algorithms can search through enormous problem spaces. Even though the current evolutionary-based HUIM algorithms can mine all HUIs that meet the minimum utility threshold faster than classic exact methods, they can still be quite time-consuming. This is mostly because search and evolutionary algorithms, which perform well in problems with fewer optimal solutions, keep the best values from one population to direct the next. However, because it overemphasizes previously discovered optimal values within a small number of iterations, this strategy runs the risk of missing some itemsets in High Utility Itemset Mining (HUIM), when results are many and unevenly distributed. Enhancing the diversity of the generated population, as suggested by Song and Huang [17], is one way to solve this issue. Rather than selecting HUIs with high utility values from the current population, this method applied roulette wheel selection to all of the identified HUIs in order to probabilistically choose the initial target of the next population.

To enrich the efficiency, this work suggests a hybrid approach that coalesces Ant Colony Optimization (ACO) and Genetic Algorithms (GA). Despite achieving expeditious global convergence, GAs have trouble with feedback and may carry out dispensable iterations, which can truncate precision. The lack of pheromone trails causes ACO to perform poorly at first, but it excels at updating information for optimal convergence. To get beyond these restrictions, the suggested genetic ant colony algorithm makes utilization of the advantages of both approaches. We modeled the HUIM problem from the perspective of the hybrid genetic algorithm, using the work done in study [17] as a starting point.

III. Problem Statement

The problem of HUIM is: Given a transaction database (*TD*), its profit table (*ptable*) and a user specified minimum utility threshold, the problem of HUIM is to identify all itemsets that have utility values equal to or greater than *min _util*. Let us say that Table I represents the Transaction Database of items procured and Table II represents the external profit value of each item.

TABLE I. Transaction Database

TD	Transactions	TU
*T*0	(a, 3) (c, 12) (e, 3)	54
*T*1	(b, 4) (d, 2) (e, 1) (f, 5)	47
*T*2	(a, 3) (c, 2) (e, 1)	16
*T*3	(a, 2) (d, 2) (f, 1)	15
*T*4	(a, 1) (c, 5) (d, 7)	52
*T*5	(b, 1) (d, 4) (f, 2)	29

TABLE II. Profit Table

Item	a	b	c	d	e	f
profit	2	7	3	5	4	1

From the transaction database in Table I and profit table from Table II, the utility value of an item *e* in the transaction T_0 is represented as $u(e,T_0) = 3 \times 4 = 12$. The utility of itemset {a, c} in transaction T_0 is $u(\{a,c\},T_0) = u(a,T_0) + u(c,T_0) = 3 \times 2 + 12 \times 3 = 42$. Along similar lines, the utility value of an itemset {a, c} in the transaction database TD will be $u(\{a,c\},T_0) + u(\{a,c\},T_2) + u(\{a,c\},T_4) = 42+12+17 = 71$. The transaction utility (TU) of an entire transaction T_0 is represented as $TU(T_0) = u(\{a,c,e\},T_0) = 54$. If the threshold utility *min util* = 80 then the itemset {a, c} is not considered an HUI as the itemset utility value $u(\{a,c\}) <$ *min util*. However, as the itemset {a, c} is part of three transactions T_0, T_2 and T_4 the TWU (Transaction Weighted-Utilization) is computed as $TWU(\{a,c\}) = TU(T_0) + TU(T_2) + TU(T_4)$. The value of this expression is 122. Since

TWU of the itemset $\{a,c\} > min\ util$ this itemset is considered an HTWUI (high transaction-weighted utilization itemset).

A. Terminologies

In this section, we define the terminologies related to the problem statement. Let $I = \{i_0,i_1,...,i_m\}$ represent a finite set of m itemsets in the transaction database $TD = \{T_0,T_1,...,T_d\}$. Each transaction T_k in TD is a subset of I which has a unique identifier $k(1 \le k \le n)$ and is called TID. Also, the set $X \subseteq I$ is called an *itemset* and an itemset which consists of k items is called a k-itemset. An itemset X is contained inside a transaction T_k if $X \subseteq T_k$. Every item i_j in T_k has a positive number $q(i_j,T_k)$ which is called it's *internal utility* and indicates the quantity (or occurrence) of i_j in T_k. The external utility$p(i_j)$, represents the unit profit value of the item i_j. The profit table $ptable = \{p_1,p_2,...,p_m\}$ depicts the profit value p_j of each item i_j in I.

The overall utility value of an item i_j in a transaction T_k is defined by the following equation as:

$$u(i_j,T_k) = p(i_j) \times q(i_j,T_k) \quad (1)$$

In any given transaction T_k, the utility if the itemset X is represented as $u(X,T_k)$ and characterizes the amount of money from the sale of X in that transaction [14]. Also, the overall utility value of an itemset X in TD denoted by $u(X)$ represents the total amount of money that the itemset yields for all transactions where X is purchased in the database. These two ideas are formally defined as:

$$u(X,T_k) = \sum_{i_j \in X \wedge X \subseteq T_k} u(i_j,T_k) \quad (2)$$

$$u(X) = \sum_{X \subseteq T_k \wedge T_k \in TD} u(X,T_k) \quad (3)$$

The user set minimum utility threshold (δ) is the percentage of total sum of all TU values in the input database. The minimum-utility value is defined as:

$$min\_util = \delta \times \sum_{T_k \in TD} TU(T_k) \quad (4)$$

An itemset X is considered an HUI if $u(X) \ge min\_util$. Search space reduction is often carried out in HUIM by defining another term called transaction weighted-utilization (TWU) which is an upper bound on the utility value of an itemset and its supersets.The TWU of an itemset X is the total sum of transaction utility values of all the transactions that contain X and is defined as:

$$TWU(X) = \sum_{X \subseteq T_k \wedge T_k \in TD} TU(T_k) \quad (5)$$

An itemset X is considered a high transaction weightedutilization itemset (HTWUI) if $TWU(X) \ge min\ util$; otherwise X is considered a low transaction weighted-utilization itemset (LTWUI).

IV. Proposed Methodology

This study addresses the challenge of high-utility itemset mining by proposing a hybrid algorithm approach which combines Genetic Algorithms (GA) and Ant Colony Optimization (ACO). Both GA and ACO are iterative optimization techniques, which forms the basis for their integration. GAs excel at rapid global convergence but struggle with feedback information. Once a solution reaches a certain range, GAs tend to perform redundant iterations, potentially reducing the accuracy of the final result. In contrast, ACO continuously gathers and updates information so that it converges to the optimal solution, leveraging its global search capabilities and parallel processing. However, ACO's initial performance is hindered by the lack of early pheromone trails. To address this limitation, we propose a genetic ant colony algorithm which capitalizes on the complementary strengths of both methods. The ACO algorithm is applied to itemsets, utilizing its ability to gather and update information continuously. Subsequently, GA is employed on the discovered High Utility Itemsets (HUI) to streamline the algorithm. In each iteration, mutation generates an HUI where crossover can be applied. GAs, inspired by natural selection and genetics, model the evolution of potential solutions through selection, crossover, and mutation. They are versatile and applicable to various optimization and search problems. ACO, on the other hand, mimics the foraging behavior of ants, simulating how they find paths by depositing and following pheromone trails. ACO excels in exploring complex solution spaces, adapting to changing conditions, and finding near-optimal solutions mainly for problems which have large search spaces. The hybrid approach aims to mitigate the premature convergence problem often encountered in genetic algorithms. By combining GA and ACO, the algorithm strikes a balance between exploitation and exploration. The ACO/GA hybrid typically reduces the number of offspring produced by constructing solutions gradually. This results in fewer new solutions needing to be stored in memory compared to a standard GA, leading to a larger pool of offspring for selection and significantly reduced memory usage.

A. Encoding and Pruning

An efficient representation method for mining HUIs is transformation of initial transaction database into a bitmap [16]. Here, the transaction is encoded using binary notation; an entry of '0' denotes the absence of an item, whereas an entry of '1' denotes its presence. The bitmap cover of an itemset X is computed as $Bit(X)$ = bitwise-AND $_{i \in X}(Bit(i))$. This indicates that X is a bit vector that is produced by applying a bitwise-AND operation to the bitmap covers of each and every item in X. For two itemsets X and Y, $Bit(X \cup Y)$ can be evaluated as $Bit(X) \cap Bit(Y)$, the bitwise-AND of $Bit(X)$ and $Bit(Y)$.

TABLE III. Bitmap Representation

T_k	a	b	c	d	e	f
$T0$	0	1	1	0	1	0
$T1$	1	0	1	0	1	1
$T2$	1	0	1	0	1	0
$T3$	0	1	0	1	0	1
$T4$	1	0	1	1	0	0
$T5$	1	0	0	1	0	1

Table III illustrates the bitmap of Table I's database, for reference. The column vectors $B(a)$ = 011011 and $B(c)$ = 111010, respectively, represent the bitmap covers of items a and c. The bitmap cover of itemset $\{a,c\}$ is the column vector obtained by performing the bitwise-AND of $B(a)$ and $B(c)$, that is $B(\{a,c\})$ = 011010. The study in [17] proposed a promising encoding vector for speeding up the process of mining HUI and is employed in this proposed work too.

Let's say that V represents an encoding vector that contains $0s$ and/or $1s$ and corresponds to a solution. If $Bit(X)$ only contains $0s$ then V is called an unpromising encoding vector ($UPEV$), otherwise V is called a promising encoding vector (PEV). Since an empty encoding vector indicates that the itemset does not contain any *HTWUI*, it is simple to understand that each itemset (solution) X that is represented by a *UPEV* cannot be an HUI.

This type of solution can significantly cut down on runtime because it does not require the fitness value to be computed. This technique is called PEV-Check (PEVC) pruning approach [17]. Every newly generated solution goes through this strategy to make sure that the solution actually exists in the database. The pseudocode of this strategy is given as follows:

Algorithm 1: Encoding and Pruning

Step 1	Determine which of the elements in the encoding vector (EV) are represented by $1s$ and stores it in VN after looking for $1s$ in the vector.
Step 2	Initialize a variable XV with bitmap cover of the first item in EV.
Step 3	Start a loop, for each item i_k, perform bitwise AND operation on XV with bitmap cover of i_k.
Step 4	If the resulting bit vector is a UPV, then the item is not kept in XV and the bits of i_k in EV is changed from 0 to 1.
Step 5	Repeat Steps 3 to 4 till VN is empty

B. Population Initialization

The initial population for hybrid ACO/GA is generated randomly. Algorithm 2 initially searches the database for all 1-HTWUIs, and as 1LTWUIs cannot be a part of any HUI, they are subsequently removed. After that, a bitmap is created from the database. After that, a for loop creates the first individuals one at a time, assigning a random number of 1s to each person in the i^{th} bit vector, where n is an integer between 1 and |1 - HTWUIs |. A bit vector with 1s in it is formed. The following formula indicates the likelihood that the bit corresponding to i_j will be set to 1 is as follows:

$$P_j = \frac{TWU(i_j)}{\sum_{k=1}^{|1-HTWUIs|} TWU(i_k)} \quad (6)$$

Algorithm 2: Population Initialization

```
Step 1   Perform a single scan on D to remove 1-LTWUIs
         and identify all 1-HTWUIs
Step 2   Transform the database D to a bitmap
         representation of D;
Step 3  Start a for loop from i = 1 to P
        3.1Generate a random number n_i,an integer
           between 1 and |1-HTWUIs |;
        3.2Generate V_i with n_i bits set to 1using the Eq 6
        3.3   if n_i > 1 then
                V_i = PEV C(V_i);
              endif
           endfor
end
```

C. Hybrid ACO Algorithm

The primary objective of ACO algorithm is to take full advantage of the features of ACO and GA for mining itemsets with high utility. Both ACO and GA gives best result for mining HUI. In the proposed technique, population is initialized by following Algorithm 2. The entire number of transactions in the database is represented by the number of ants' SN. Pheromones are initialized with the utility values, iff TWU(X,Tc) ≥ min util. Genetic operators crossover is applied on the generated solution by the ant. In this study, a single-point crossover is utilized. The crossover point is selected randomly. Now mutation operator is applied by changing one or more random bits in a string when Pm ≥ randomly generated value.

Algorithm 3: Hybrid ACO algorithm

```
Step 1   Population initialization
Step 2   Pheromone initialization with utility value
Step 3   iter = 1
Step 4   while (iter ≤ max iter)
Step 5   for each transaction in D
           Start each ant to visit each vertex
         endfor
Step 6   Apply genetic operators crossover
         on the selected population
Step 7    Perform mutation on the population of
          individuals with mutation probability pm.
          7.1   Verify the fitness value of individuals,
          7.2    If fv ≥ min util go to Step 6
          7.3    Else go to Step 8
Step 8    Pheromone update
Step 9    End of while
Step 10   Output all HUIs
```

V. Result and Discussion

This part includes a discussion of the findings and experimental assessment of the suggested algorithms. Experiments were performed on a computer equipped with an 8-core 3.6 GHz CPU and 8 GB RAM running Windows 10. The program was developed in Java.

A. Dataset

Four standard benchmark datasets- Chess, Mushroom, Connect, and Accident - are utilized to assess the performance and effectiveness of the suggested algorithm. The algorithm performance is also evaluated against a real dataset downloaded from the UCI repository. The benchmark datasets were sourced from SPMF data mining library [23].

A widely used dataset from the UCI Machine Learning Repository, the Mushroom dataset is often employed in association rule mining and High-Utility Itemset (HUI) mining applications. There are 8,124 different species of mushrooms in it, and each one is characterized by 22 different categories, including gill spacing, color, cap shape, and odour. Every instance has a label designating it as deadly or edible. The dataset can be modified in the context of HUI mining by giving utility values (such as profit or significance) to particular features in order to discover desirable itemsets based on utility thresholds as opposed to just frequency. When it comes to HUI mining, the Mushroom Dataset presents a problem because it requires categorical data to be transformed into a format that makes utility-based itemset mining feasible. In order to match with the utility mining paradigm, this frequently entails defining the utility of goods (attributes) and transactions (mushroom instances) in novel ways.

The chess dataset is widely used in High-Utility Itemset (HUI) mining as a benchmark for assessment of the performance of various algorithms. This dataset represents chess games transactionally, with each transaction denoting a series of moves or itemsets. Finding itemsets (e.g., common move sequences) with high utility, like winning strategies, is the goal. In this case, the utility values could stand for the significance or regularity of moves within the dataset. As the chess dataset is structured, repetitious, and dynamic [21], it's a great resource for researching utility-based itemset mining. Every transaction in HUI mining can be compared to a position in a chess game. Each item can stand for a particular move or the location of a chess piece.

Based on actual accident data, the Accident _10% dataset is a 10% sampling of the entire dataset. The dataset comprises transactions that typically contain attributes related to an accident, such as conditions, location, and involved entities. The utility values assigned to each attribute indicate its significance or influence. When mining for interesting patterns, this dataset aids researchers in assessing the accuracy, scalability and execution time of HUI algorithms.

High-utility itemset (HUI) mining research frequently uses the Connect dataset as a benchmark. It is derived from a Connect-4 game and is sourced from the UCI Machine Learning Repository. With 43 categorical attributes, the dataset consists of 67,557 instances, each of which represents a unique board configuration in the game. Each configuration can be looked at as a transaction in the context of HUI mining, where the objective is to mine itemsets (board configurations) that offer high utility or significance, frequently based on different utility values ascribed to different configurations. The Connect dataset is a valuable resource for examining recurring and noteworthy patterns in gaming, since its utility may be linked to winning plays or strategies in the Connect-4 game.

The proposed algorithm is also tested against a real dataset - the online retail dataset downloaded from UCI Repository. The dataset has 541909 instances with six features. A UKbased online retailer's sales transactions are included in this dataset from a period of December 2010 to December 2011. The dataset includes 8 attributes with approximately 4000 distinct items. Cancelled orders, returns (negative quantities), and possibly outliers are included in the dataset. In order to handle returned items or eliminate invalid transactions, preprocessing is frequently necessary. The dataset exhibits a high degree of transaction data skewness because it includes both huge bulk transactions and numerous low-quantity sales. The identification of high-utility itemsets is impacted by this skewness. Since utility values were not included in the dataset, they had to be manually determined using the following technique so that they could be used in HUI mining where *Utility = quantity × unit price*. Table IV shows characteristics of datasets and Table V shows sample online retail dataset.

TABLE IV. Characteristics of Datasets

Dataset	Trans. Len	No. of items	No. of Trans	Type
Mushroom	23	119	8,416	Dense
Chess	37	75	3,196	Dense
Connect	43	129	67,557	Dense
Accident 10%	34	468	34,018	Dense

TABLE V. Sample Online Retail Dataset

Invoice No	Stock Code	Description	Quantity	Invoice Date	Unit Price	CustomerID	Country
536365	71053	Jam making set	5	12/1/2010	2.25	13047	United Kingdom
536366	85123A	Jam making set	1	12/1/2010	2.25	12583	France
536367	71057	White Hanging Heart	6	12/1/2010	3.39	17850	United Kingdom
536367	22993	White Hanging Heart	6	12/1/2010	4.25	12678	France

B. Runtime efficiency

Experiments are conducted to assess the efficiency of the proposed algorithm with regards to runtime. The efficiency is tested by varying the minimum utility value.

Table VI summarizes the minimum utility value set for the dataset.

TABLE VI. MINIMUM UTILITY VALUES

Dataset	Minimum Utility Threshold				
Chess	28.5	29	29.5	30	30.5
Mushroom	14	14.5	15	15.5	16
Connect	31.8	32	32.2	32.4	32.6
Accident 10%	12.6	12.8	13	13.2	13.4

Fig. 1, 2, 3, and 4 show the execution time of our algorithm presented in this work. The runtime efficiency of our proposed algorithm is compared with HUIM-GA. Hybrid ACO/GA produces better runtime efficiency because it focuses on promising areas early on, reducing unnecessary evaluations of low utility items. The incremental building of solutions impacts the efficiency of our proposed algorithm. Also, the adaptive mechanisms such as pheromone evaporation leads to an efficient search, thus reducing runtime.

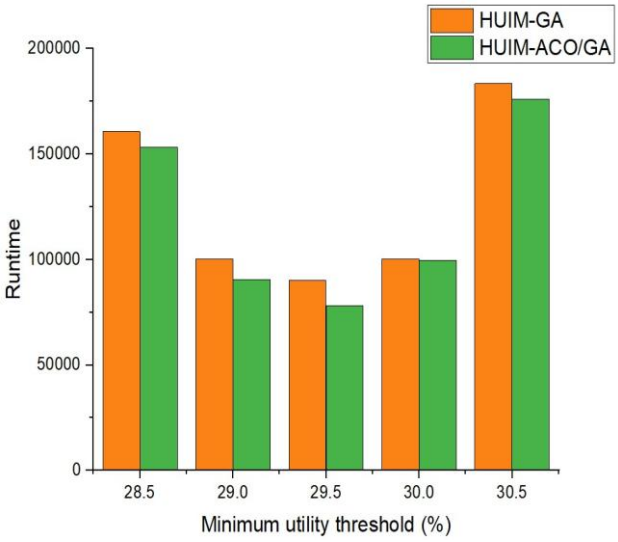


Fig. 1. Chess dataset (based on runtime).

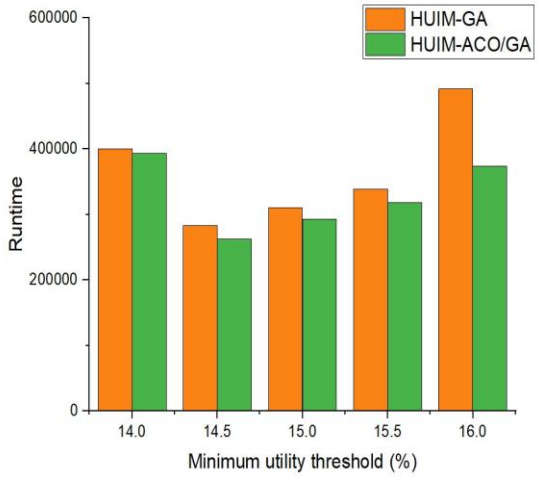


Fig. 2. Mushroom dataset (based on runtime).

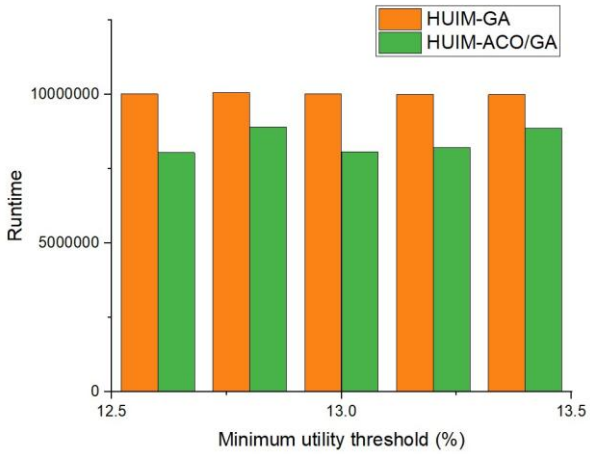


Fig. 3. Accident_10% dataset (based on runtime).

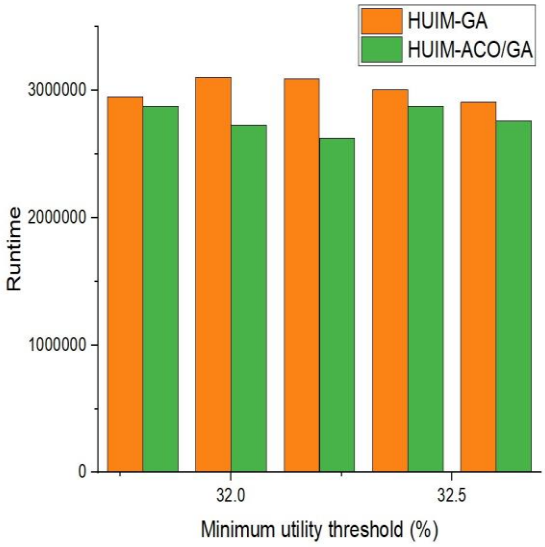


Fig. 4. Connect dataset (based on runtime).

A real-time retail dataset is also used to evaluate the recommended algorithm's runtime efficiency. Here at minimum utility threshold of 10%, maximum number of HUI's were identified. Fig. 5 demonstrates the result.

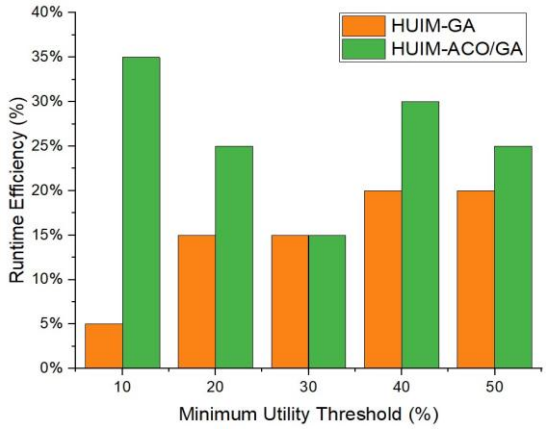


Fig. 5. Retail dataset (based on runtime).

C. Discovered number of High utility items(HUI)

This section analyzes the performance of the hybrid ACO/GA approach by evaluating the number of HUIs identified at various threshold levels. The proposed algorithm discovers the maximum number of HUIs. Fig. 6 to 10 demonstrate the graphical representation of the experimental results. It may be noticed that hybrid ACO/GA discovers more number of HUIs in most cases. However, for the real time retail dataset, the number of HUIs identified by hybrid ACO/GA algorithm were comparable to that of HUIM-GA algorithm as may be seen in Fig. 10.

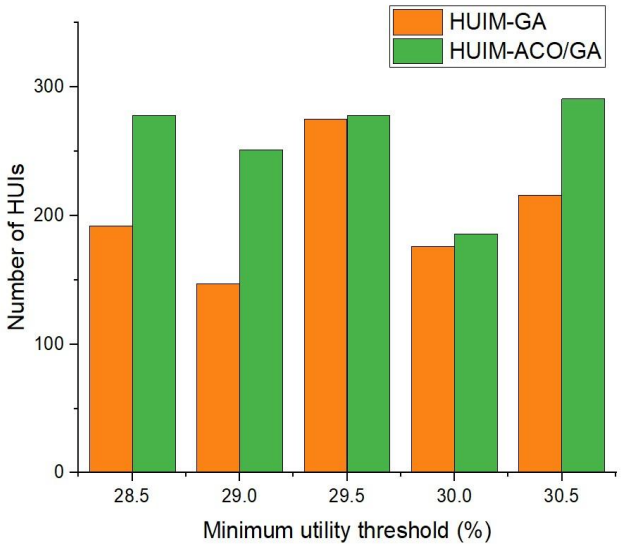


Fig. 6. Chess dataset.

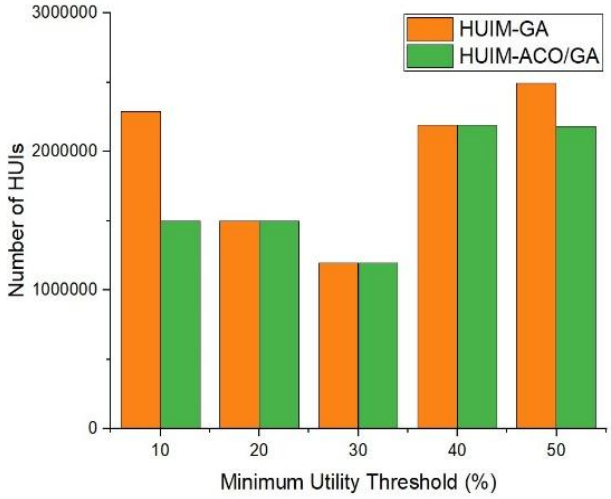


Fig. 7. Mushroom dataset (based on number of HUIs).

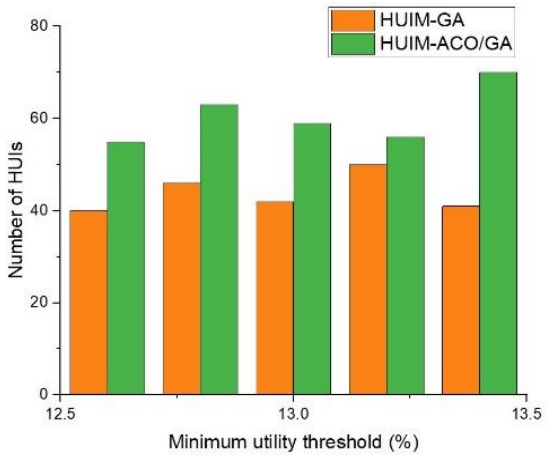


Fig. 8. Accident_10%dataset (based on number of HUIs).

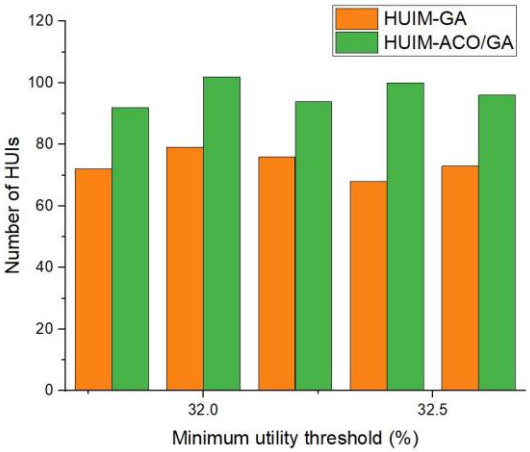


Fig. 9. Connect dataset (based on number of HUIs)..

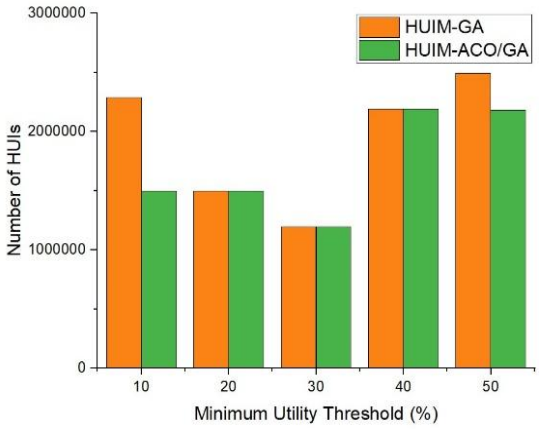


Fig. 10. Retail dataset (based on number of HUIs)..

D. Convergence

Using the four datasets - Chess, Mushroom, Accident, and Connect, 4000 fitness evaluations were conducted to assess the convergence properties of the proposed approach. Minimum utility threshold was kept at the value which gives the maximum number of HUIs discovered in each dataset. Premature convergence of genetic algorithm is avoided here. Convergence curves obtained on the Chess, Mushroom and Retail datasets on both hybrid ACO/GA are represented in Fig. 11, 12 and 13 respectively.

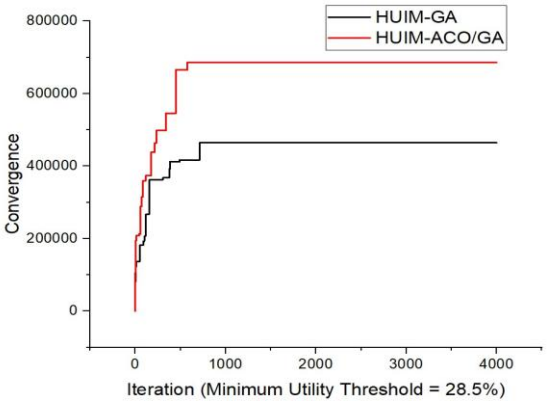


Fig. 11. Chess dataset (convergence).

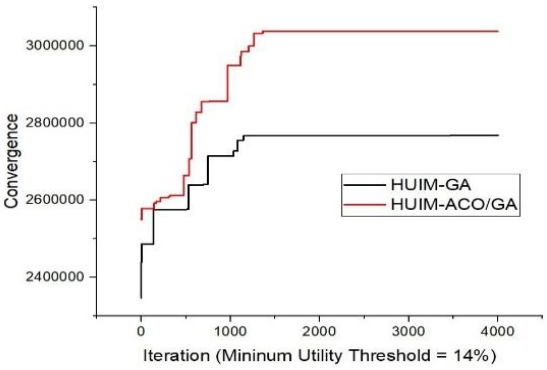


Fig. 12. Mushroom dataset (convergence).

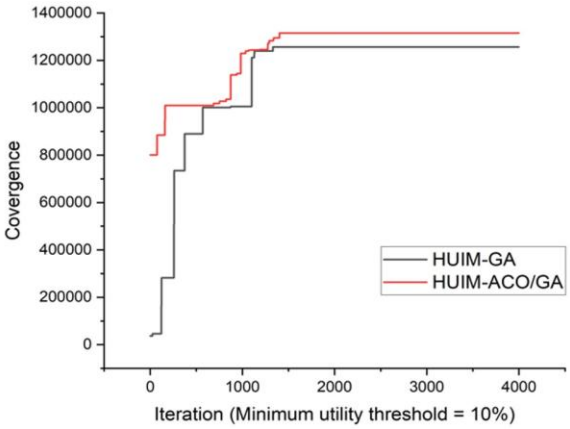


Fig. 13. Retail dataset (convergence).

VI. Conclusion and Future Enhancement

With a wide range of applications, HUIM is a significant data mining task. When mining HUIs, EC-based HUIM algorithms outperform conventional HUIM algorithms such as HUPEumu-GRAM, HUIM-BPSOsig, HUIM-BPSO, and BioHUIF-GA. Even though these algorithms offer an effective method for extracting HUIs from the transition datasets, finding the full or significant proportion of HUIs still takes a lot of effort. While hybrid ACO-GA algorithm yields good results on generic datasets, the suggested technique uses more resources to create HUIs on real-time datasets. Our proposed algorithm yields the same significant number of HUIs as other algorithms and after a certain period of time both algorithms converge and the quantity of HUI generated is sparse. The suggested algorithm will be used to mine high utility itemsets from the e-commerce dataset in order to implement dynamic pricing strategies. This will allow for real-time price modifications based on patterns of customer demand and product profitability. Dynamic pricing is made possible by mining HUIs in e-commerce datasets to find high-utility items and product groups. Through value-driven tactics, this aids companies in maintaining client loyalty, increasing revenue, and optimizing pricing in real-time. Our next improvement can be to use fewer resources to address the HUIM issue and improvise our proposed algorithm.

References

[1] Agarwal R., and Srikant R., “Fast algorithms for mining association rules”, In the Proceedings of 20th International Conf. Very Large Data Bases, pp.487-499, 1994.

[2] Agrawal R., Imielinski T, Swami A, “Mining association rules between´ sets of items in large databases”, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data - SIGMOD ‘93. pp.207, 1993.

[3] Tseng V.S., Shie Bai-En, Wu Cheng-Wei and Yu Pillip S., “Efficient Algorithms for Mining High Utility Itemset from Transactional Databases”, IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 8, pp. 1172-1786, 2013.

[4] Chan R., Yang Q., and Shen Y., “Mining High Utility Itemsets”, Proceedings of the IEEE 13th International Conference on Data Mining, Melbourne, Florida, pp.19- 26, 2003.

[5] Yao H., Hamilton H.J., and Butz C.J., “A Foundational Approach to Mining Itemset Utilities from Databases”, Proceedings of the 2004 SIAM International Conference on Data Mining, Lake Buena Vista, Florida, USA, pp.482-486, 2004.

[6] Ahmed C.F., Tanbeer S.K., Jeong Byeong-Soo, and Lee Young-Koo, “Efficient Tree Structures for High Utility Pattern Mining in Incremental Databases”, IEEE Transactions on Knowledge and Data Engineering, Vol. 21, No. 12, pp. 1708-1721, 2009.

[7] Liu M. and Qu J., “Mining High Utility Itemsets without Candidate Generation”, Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 55-64, 2012.

[8] Philippe Fournier Viger, Cheng-Wei Wu, Souleymane Zida, Vincent S. Tseng, “FHM: Faster High-Utility Itemset Mining using Estimated

Utility Co-occurrence Pruning”, Proc. 21st International Symposium on Methodologies for Intelligent Systems (ISMIS 2014), Springer, LNAI, pp 83-92, 2014.

[9] Artificial Intelligence Advances in Artificial Intelligence and Soft Computing pp 530-546, 2015.

Peng A., Koh Y. S, and Riddle P. J., “mHUIMiner: A Fast High Utility Itemset Mining Algorithm for Sparse Datasets” Proceedings 21st Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, pp.196-207, 2017.

[10] Utility Itemset Mining Algorithm for Sparse Datasets" Proceedings 21st Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, pp.196-207, 2017.

[11] Souleymane Zida, Philippe Fournier-Viger, Jerry Chun-Wei Lin, ChengWei Wu, Vincent S. Tseng, "EFIM: A Highly Efficient Algorithm for High-Utility Itemset Mining", Mexican International Conference on

[12] Kannumuthu S., and Premalatha K., "Discovery of High Utility Itemsets Using Genetic Algorithm" International Journal of Engineering and Technology (IJET), Vol. 5 No. 6, pp. 4866-4880, 2013.

[13] Lin C.W, Yang L., Fournier-Viger P., Hong T.P., and Voznak M., "A Binary PSO Approach to Mine High-Utility Itemsets", Soft Computing, Vol. 21, pp. 5103-5121, 2016.

[14] Lan G.C., Hong T.P.; and Tseng V.S., "Mining High-Transaction Weighted Utility Itemsets", 2010 Second International Conference on Computer Engineering and Applications, Bali, Indonesia, pp. 314-318, 2010.

[15] Lichman M., UCI Machine Learning Repository, http://archive.ics.uci.edu/ml, 2013.

[16] Liu Y., Liao Wk., and Choudhary, A., "A Two-Phase Algorithm for Fast Discovery of High Utility Itemsets", Lecture Notes in Computer Science, Vol. 3518, pp. 689-695, 2005.

[17] Wu J.M-T, Zhan J., and Lin, J. C-W, "An ACO-based Approach to Mine High-Utility Itemsets", Knowledge-Based Systems, Vol. 116, pp. 102113, 2017.

[18] Song W., and Huang C., "Mining High Utility Itemsets Using BioInspired Algorithms: A Diverse Optimal Value Framework", IEEE Access, Vol. 6, pp. 19568-19582, 2018.

[19] Nawaz M. S., Fournier-Viger P., Yun U., Wu Y., and Song, W., "Mining High Utility Itemsets with Hill Climbing and Simulated Annealing". ACM Transactions on Management Information Systems, Vol. 13 No. 1, pp. 1-22, 2021.

[20] Song W., and Nan J., "Mining High Utility Itemsets using Ant Colony Optimization" Advances in Natural Computation, Fuzzy Systems and Knowledge Discovery, pp. 98-107, 2020.

[21] Li Y., Zhao Y., Shang Y., and Liu J., "An improved firefly algorithm with dynamic self-adaptive adjustment", PLoS One, Vol. 16, 2021.

[22] Han M., He F., Zhang R., Li C., and Meng F., "Mining High Utility Itemsets with Elephant Herding Optimization", https://doi.org/10.21203/rs.3.rs-3881656/v1, 2024.

[23] Fournier-Viger P., Gomariz A., Gueniche T., Soltani A., Wu C.W., and Tseng V.S., "SPMF: A Java Open-Source Pattern Mining Library", Journal of Machine Learning Research, Vol. 15, pp. 3569-3573, 2014.

Enhancing IoT Security Through User Categorization and Aberrant Behavior Detection Using RBAC and Machine Learning

Alshawwa Izzeddin A O[1], Nor Adnan Bin Yahaya[2], Ahmed Y. mahmoud[3]
University Malaysia of Computer Science and Engineering, (UNIMY), Malaysia[1, 2]
Faculty of Engineering and Information Technology, Al-Azhar University-Gaza, Palestine[3]

***Abstract*—The proliferation of Internet of Things (IoT) technology in recent years has revolutionized several industries, providing customers with reliable and efficient IoT services. However, as the IoT ecosystem grows, attention has switched away from straightforward user access to the crucial topic of security. Among others, there is a need to categorize users according to the actions they conduct as well as according to aberrant user behavior. By utilizing Role-Based Access Control (RBAC) and merging the categorization of access rights with the identification of aberrant behavior, access points to the Internet of Things will be strengthened in terms of security and dependability. A system is proposed to identify security flaws and prompt rapid remediation, with the incorporation of a classification of aberrant user behaviors which, in turn, offers a thorough defense against any outside threats. Three classification methods which are Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF), were utilized in the study and their accuracy were compared. The results demonstrate the effectiveness of machine learning approaches using a dataset for IoT users, achieving high accuracy in identifying anomalous user behavior and enabling prompt implementation of necessary actions.**

***Keywords*—Machine learning; classification; SVM; LOF; IF classification methods; aberrant user behavior; Role-Based Access Control (RBAC); IoT user dataset and user categorization**

I. INTRODUCTION

A new age of linked devices has arrived because to the Internet of Things (IoT) technology's explosive growth in recent years [1]. This technology has revolutionized several sectors and made IoT services more readily available to customers. The seamless user experience has become less important as the IoT ecosystem grows and more focus is being placed on security, a crucial issue. It has become crucial to guarantee the security and reliability of IoT access points in order to preserve sensitive information and safeguard against potential attacks [2].

IoT gadgets mostly shapes our daily life. Still, as the IoT ecology develops security issues do surface. The large attack surface created by the number of linked devices calls for strong security rules to guard IoT systems and data [3-5].

Among the most urgent security issues of the Internet of Things (IoT) is unauthorized access. Conventional security systems find it difficult to stop illicit activities considering the availability of more devices and users. Under this idea, managing the access to Internet of Things systems benefits from Role-Based Access Control (RBAC). RBAC greatly lowers unauthorized access and breaches by assigning users roles in line with their responsibilities, therefore enabling people to act in line with their jobs [6-8]. Though its shortcomings, RBAC offers a strong basis for access control. Besides, one must be knowledgeable about aberrant user behavior. As IoT devices grow, the identification of abnormal activity and hand-held surveillance become ever difficult. Little changes in user activity enable hostile actors to target real-time threat assessment with challenges. Behavior classification and RBAC help IoT system security to develop dynamically. This approach helps the system to control access depending on user obligations and notify it to any unusual user conduct, therefore offering extra security [9].

Using machine learning algorithms, especially SVM, LOF, and Isolation Forest, the suggested approach detects and labels aberrant user activity in Internet of Things applications. These machines are amazing in their capacity to spot anomalies, handle massive volumes of data, and find minute trends. These machine learning approaches enable the Internet of Things' (IoT) security architecture to automatically detect threats, therefore enabling quick reactions and lessening the effect of unwanted conduct [10,11].

We want to satisfy the growing demand for IoT-based proactive security solutions. Given the growing complexity of cyberthreats, conventional approaches fall short. The more thorough, flexible, and effective way in which RBAC and machine learning-based behavior classification can enhance IoT security is shown in this work. Strict access control rules and real-time anomaly pattern detection guarantee IoT systems' dependability and trustworthiness, hence improving security [12-15].

This paper suggests a complete strategy that combines Role-Based Access Control (RBAC) with the categorization of user behavior to increase IoT security in order to address these security issues [16, 17]. The suggested method intends to strengthen security measures and maintain the integrity of IoT services by categorizing users based on their individual behaviors and spotting aberrant behavior patterns.

The categorization of abnormal user behavior is a crucial component of the suggested system in order to quickly find and fix any possible security problems [18]. By using a proactive approach, the system is equipped with strong defenses against external attacks and intrusions, enabling prompt risk mitigation and remediation.

The Role-Based Access Control (RBAC) framework has appeared in the IoT area as a result. Well-defined roles, permissions, and access control policies enable RBAC. Users are assigned roles, such as "Administrator," "Device Owner," or "Sensor Data Analyst," which specify their duties. Roles are given specific actions by means of permissions, such as "Read Sensor Data" or "Update Firmware". The roles that can carry out particular activities on particular resources are defined by access control policies. IoT security and dependability can be greatly improved by utilizing RBAC and integrating access rights categorization with anomaly detection. This strategy strengthens the security of IoT systems by ensuring users only access functions that are consistent with their roles and responsibilities.

The study extensively examines and makes use of three well-known and effective classification methods: Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF) in order to obtain accurate user behavior categorization [19]. Each of these approaches has certain benefits and skills that make them suitable for identifying and analyzing the complex user behavior patterns seen in the Internet of Things (IoT) environment.

With its ability to handle both linear and non-linear data, the Support Vector Machine (SVM) method is well recognized [20]. It is the best contender for differentiating between typical user behavior and abnormal activity since it can efficiently build appropriate hyper-planes to divide various kinds of data. SVM is a useful tool for identifying and reporting suspicious events in the IoT environment since it learns from labelled data during the training phase and can effectively categorize new instances with high accuracy.

The study uses the Local Outlier Factor (LOF) algorithm [21], which is excellent at spotting local deviations and abnormalities in data, in addition to SVM. LOF is an unsupervised learning technique, in contrast to SVM, which relies on labelled data during training. Instead, LOF determines the density-based separation between data points' neighbors, allowing it to spot outliers and anomalous user behavior that might not fit the larger trends shown in the IoT user dataset. Due to its unsupervised nature, LOF is highly good in identifying new and uncommon anomalous actions, which improves the categorization process as a whole. Additionally, the Isolation Forest (IF) algorithm, which offers a unique and effective method for anomaly identification, is used in the study [22]. Through the use of random partitioning, IF isolates outliers into distinct trees. An observation is more likely to be an anomaly if fewer partitions are required to isolate it. This isolation approach may be used by IF to effectively identify and categorize atypical user behavior, especially when working with massive datasets that are typical of IoT contexts.

The study carefully compares the performance of these categorization algorithms utilizing a large and varied dataset made up of different Internet of Things users in order to assess how successful they are. The collection is carefully chosen to contain a wide range of user behaviors, from commonplace tasks to possible security violations. The study may evaluate the algorithms' performance in appropriately recognizing and categorizing various user behaviors by putting them through this varied dataset.

The proposed structure is improved by including these categorization algorithms to improve the overall security posture of Internet of Things applications. The results of this study provide useful insights for enhancing the security and reliability of IoT services as the IoT ecosystem continues to grow and change. This integrated strategy has considerable promise in establishing a safer and more resilient IoT ecosystem for organizations, people, and society at large by anticipating possible security risks and offering a strong defense against unauthorized access. The overall security posture of Internet of Things applications is improved by including these categorization methods within the suggested architecture. The research's findings offer important new perspectives on how to improve the security and reliability of IoT services as the IoT ecosystem continues to grow and change. This integrated strategy holds great promise for creating a more secure and resilient IoT environment for organizations, people, and society at large by anticipating possible security risks and offering a strong defense against unauthorized access.

This study will go into the design of the suggested framework, the implementation of role-based access control and user behavior classification, as well as the assessment of the selected classification techniques, in the parts that follow. This study intends to add to the ever-expanding body of research in IoT security and pave the way for more dependable and secure IoT services in the future by illuminating the applicability and efficacy of this comprehensive strategy.

The rest of the paper is organized as follows. Section II introduces the terminologies and concepts that will be utilized throughout the study. Section III examines related work, emphasizing existing approaches and their applicability to IoT security and user behavior classification. Section IV describes the experimental setup, which includes API access patterns, feature engineering, and the categorization system. Section V describes the dataset used in the study, including its characteristics and preprocessing methods. Section VI describes how the proposed system would be implemented, including the incorporation of machine learning techniques. Section VII summarizes the findings and examines their implications for IoT security. Finally, Section VIII summarizes the findings and proposes areas for further research to improve IoT system reliability and security.

II. TERMINOLOGY

In this section, we provide a comprehensive overview of the concepts and terminologies that will be utilized throughout the paper.

Machine Learning: It is a subset of artificial intelligence (AI) consisting of automatically improving algorithms driven by data use and experience.

Classification: The technique of data point category or class prediction inside datasets.

Support Vector Machine (SVM): A classifier and regression task supervised learning algorithm.

Local Outlier Factor (LOF): An unsupervised anomaly detection system spotting local deviations of data points from their neighbors.

Isolation Forest (IF): An unsupervised machine learning method based on data point isolation intended for anomaly detection.

Aberrant User Behavior: Acts or habits that greatly vary from expected or typical user behavior.

Role-Based Access Control (RBAC): A method of control of computer or network resource access depending on user responsibilities.

IoT User Dataset: A set of facts regarding user interactions and actions in Internet of Things systems.

User Categorization: The arrangement of users into predetermined groups according to their roles or behavior inside a system.

IoT Security: Internet of Things device and network protection against cyberattacks and illegal access.

III. Related Work

IoT has gained immense popularity due to its ability to generate vast amounts of data from interconnected devices. Analysing and categorizing this data is crucial for extracting meaningful insights and making informed decisions. Unsupervised classification, which involves grouping data points without predefined labels, becomes essential in scenarios where labelled training data may be limited or costly to obtain. The review aims to shed light on the performance, strengths, and limitations of different algorithms when applied to the challenging task of classifying user-generated data from Internet of Things (IoT) devices.

In categorising this data, unsupervised classification algorithms are essential for understanding user behaviour, preferences, and anomalies. The Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF) are three well-known classification techniques that will be compared in this literature study.

In the context of the Internet of Things (IoT), this paper explores the difficulty of identifying outliers. It examines how the IoT's unique properties, such as wireless communication openness and resource constraints, render standard methodologies inappropriate. The research carefully examines new machine learning-based outlier identification methods in order to solve this. The IoT ecosystem has a variety of elements that might cause outlier data, such as environmental impacts, sensor errors, hostile assaults, and event-related abnormalities. According to the machine learning algorithms used, the article categorises outlier detection methods into clustering, classification, dimension-reduction, and hybrid approaches. Researchers and practitioners may better traverse the challenges of outlier identification in IoT contexts by knowing the capabilities and constraints of various techniques [23].

The article discusses the growing security issues brought on by technical development and the expansion of the internet. It draws attention to the increase in cyber-attacks and the demand for strong security measures. To preserve information security and stop shady network activity, intrusion detection systems (IDS) are essential. In order to identify and classify security risks, the article primarily focuses on the integration of Machine Learning (ML) methods into IDS. The paper offers a comparative examination of several ML algorithms utilized inside IDS for a variety of applications, including big data, smart cities, fog computing, and the Internet of Things (IoT). It emphasizes how crucial these algorithms are to boosting security throughout various fields [24].

Reflections, interference, and ambient conditions all have an impact on Wi-Fi signals, resulting in irregular and aberrant RSS values that make exact localization difficult. The study suggests an outlier identification approach called "iF_Ensemble" that combines supervised, unsupervised, and ensemble machine learning techniques to handle this problem. The suggested iF_Ensemble approach uses the isolation forest (iForest) unsupervised learning methodology to categories RSS data as normal or abnormal and locate outliers. Additionally, the data is subjected to specific applications of supervised learning techniques as support vector machine (SVM), K-nearest neighbour (KNN), random forest (RF), and elliptic envelop. Their outcomes, meanwhile, are judged inadequate. An ensemble learning technique known as stacking is presented to improve outlier detection [25].

In order to improve security, machine learning (ML) and deep learning (DL) approaches are suggested in this paper's discussion of security issues in Internet of Things (IoT) networks. IoT networks are susceptible to cyber-attacks because they are made up of linked, dispersed embedded units with little processing power. The article addresses how the peculiarities of IoT networks have prevented existing cryptography methods from adequately addressing security and privacy issues. The security needs, possible attack vectors, and current security solutions for IoT networks are all thoroughly reviewed by the writers. The study also provides a number of ML and DL approaches that may be used to address various security issues in IoT networks. Utilizing these clever methods will allow for greater privacy and security features [26].

IV. Experimental Setup

Users utilise APIs to access the application while doing so. Users can access a certain business logic using a particular set of APIs. Malicious users occasionally try to access APIs in a way that leads to a drastically different order of APIs than do benign users. There may be several sequences of API calls depending on the business logic being accessed, and when they are combined, they form an API call graph for that user. Numerous users create exact or comparable API call graphs when there are hundreds of users. In these situations, users are grouped into a single cluster with the same graph shared by all of them. Each graph in these clusters was manually categorised as an outlier or a normal graph, and the results are presented in the classification column.

The study dataset consists of user-generated API call graphs, which are created when users utilise a set of APIs to access particular business logic. The dataset consists of numerous components, including session characteristics, IP parameters, temporal dynamics, behaviour classifications, and API access patterns. To assure data quality, we collected the dataset from Kaggle and prepared it thoroughly. This procedure involved choosing key characteristics, carrying out technical activities, and handling any missing data.

Examples of Features:

- Temporal Dynamics: Patterns in API usage across time.
- API Access Patterns: The order and frequency of API requests.
- Session Characteristics: Information regarding user sessions, including duration and frequency.
- IP Attributes: Information about the IP addresses used.
- Behaviour Classification: The classification of user behaviour as normal or abnormal based on API call sequences.

The Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF) algorithms are used in this experiment to categorise API call graphs into normal and outlier categories. The dataset utilised in this study consists of API call graphs that users create when they use an API to access a particular piece of business logic. The main goal is to precisely pinpoint possibly fraudulent users whose API call graphs show notable departures from good user behaviour. To do this, the deviations have been carefully categorised as either outliers or normal graphs, creating a trustworthy baseline for assessment.

Preprocessing the data is the first step in the experiment, during which the API call graphs are appropriately converted into a numerical representation to act as features in the classification procedure. In order to reduce biases during evaluation, the dataset is then split into training and testing sets, with a balanced distribution of normal and outlier graphs in each set.

An SVM classifier is developed for SVM classification utilizing the relevant libraries and hyper parameters. Using criteria like accuracy, precision, recall, and F1-score, the model's performance is assessed on the testing set after it has been trained on the labelled training set. In addition, a visual inspection of the SVM model's decision boundary is carried out to learn more about how well it can distinguish between normal and outlier classes.

Fig. 1 explains the LOF technique is then used to find regional outliers in the dataset in the next stage. In the testing set, LOF is used to categorize API call graphs as either normal or outlier, and its effectiveness is assessed and contrasted with the SVM method.

The Isolation Forest technique is also included to find data abnormalities. In order to distinguish between typical graphs and outliers, the IF algorithm is used to categorize API call graphs in the testing set. Its performance is carefully evaluated and compared to the results of both SVM and LOF.

The SVM, LOF, and IF algorithms are thoroughly analyzed and compared in this experiment so that we can decide which approach is best for precisely detecting outliers within the user population.

The study further clarifies each algorithm's advantages and disadvantages in relation to the particular features of the dataset. Finally, statistical significance tests may be carried out to see whether the three algorithms' performances differ significantly from one another. By identifying potentially harmful activity among users who are visiting the application's APIs, the investigation's findings are meant to offer useful insights into how well these categorization approaches might improve security measures.

V. Dataset

Users interact with the program by using APIs when they access it. Access to various business logic is made possible via a certain sequence of APIs. Malicious users occasionally alter API access, resulting in a different sequence of APIs from those used by authorized users. Numerous API call sequences may occur, generating an aggregated API call graph for one user, depending on the accessible business logic. Multiple users in situations with multiple users generate the same or comparable API call graphs. In these situations, users are gathered into a single cluster using a graph clustering method, sharing a single graph. Manual examination of these clusters resulted in the categorization of each graph as either a normal or an outlier, and the results are shown in the classification column. One may compare the model's prediction with the value in this column for each of the outliers listed in the 'behavior_type' column to confirm accuracy. The reference point might be the model's output. Additionally, this dataset has a behavior known as a "bot," with many bot varieties offered. You can see how the bot is categorized as a normal or outlier. The source of this dataset is Kaggle, a popular platform for hosting and sharing datasets related to various domains and topics [27].

A. Description of the Dataset

The database comprises entries of API call activities, encompassing diverse attributes that depict distinct facets of these actions. The dataset includes a list of columns, each accompanied by a description:

- Id: A unique identifier for each record.
- api_duration: The duration of the API call.
- api_access: This likely refers to the access count or frequency of the API call.
- sequence_length: The length of the API call sequence, indicating the number of steps or calls in the sequence.
- session_duration: The duration of the user session during which the API calls were made.
- ip_type: A categorical feature indicating the type of IP address used (e.g., static or dynamic).
- num_sessions: The number of sessions associated with the record.
- num_users: The number of users involved in the session or activity.
- num_unique_apis: The number of unique APIs accessed during the session.
- source: This may indicate the source of the data or the originating system.
- classification: The label or classification of the record, which likely indicates whether the behavior is normal or aberrant (e.g., 1 for normal, 0 for aberrant).

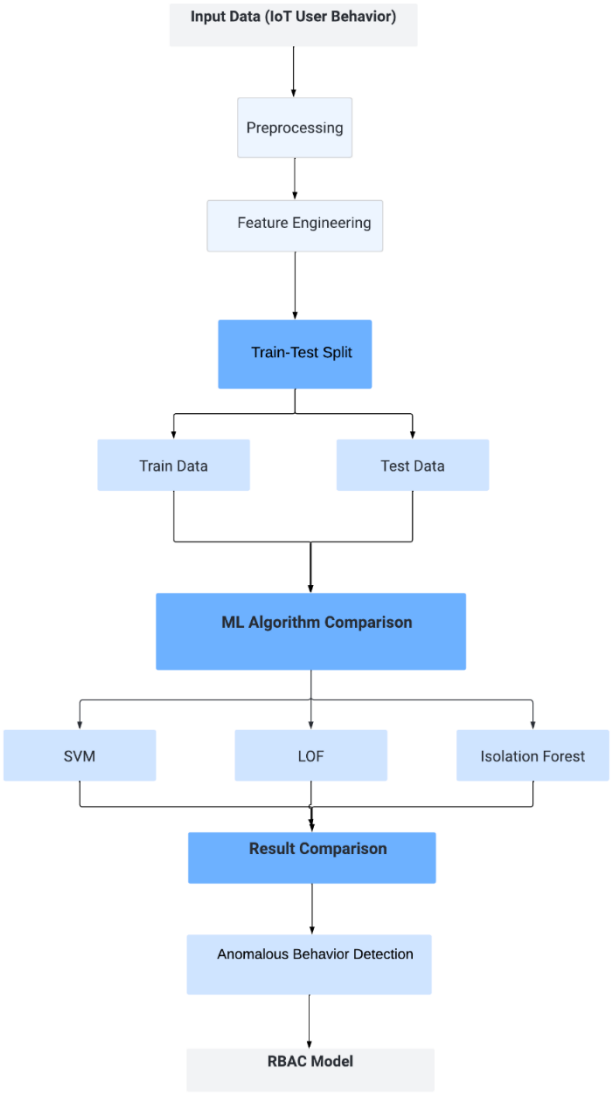


Fig. 1. Semantic diagram of proposed model.

The dataset contains diverse parameters pertaining to API calls and user sessions, furnishing comprehensive information for study. The data may be utilized to categorize user behavior, discover abnormalities, and enhance security in IoT systems by recognizing trends and variations in API utilization.

B. Methods Used for Outlier Detection Abnormal User

The suggested strategy employs the following machine learning techniques. The recommended method for identifying anomalous user behavior makes use of a number of machine learning techniques. These techniques include Isolation Forest (IF), an ensemble-based algorithm that isolates anomalies within decision trees, Local Outlier Factor (LOF), an unsupervised technique that identifies anomalies by evaluating the local density deviation of data points, and Support Vector Machine (SVM), which uses supervised learning principles to classify instances based on a separating hyper-plane. The study uses these methods in an effort to distinguish between typical and anomalous user behavior in an Internet of Things (IoT) context. By efficiently recognizing and correcting aberrant behavior patterns, this strategy has the potential to improve the security and dependability of IoT access points. The selection and effectiveness of these techniques are influenced by variables including feature engineering, data preprocessing, and the right model parameter settings, all of which help to achieve accurate and effective outlier identification in the IoT environment.

C. Data Preprocessing

Data preprocessing is a critical phase in any machine learning study, as the quality and suitability of the data directly influence the performance and reliability of the subsequent analysis. In this section, we outline the steps taken to preprocess the dataset used in the study, which focuses on strengthening security and dependability in the Internet of Things (IoT) ecosystem by categorizing user behaviors and identifying aberrant activities using Role-Based Access Control (RBAC) and classification methods.

1) Data collection and cleaning: This study's dataset was obtained from Kaggle, a popular platform for sharing datasets. A thorough review of the dataset was done before the analysis to find any missing values, errors, or noise. Standardising attribute names, managing missing data through imputation or removal, and resolving any conflicts were all part of the cleaning operations.

2) Feature selection and engineering: To enable effective classification, the most relevant features were selected based on their significance in addressing the research problem. Features such as temporal dynamics, API access patterns, session characteristics, IP attributes, and behavior classifications were retained. Furthermore, feature engineering techniques were employed to extract meaningful insights and create new attributes that could amplify the discriminating power of the classification models.

3) Categorization and labeling: Users were divided into groups according to their actions and behaviours, which was in line with the emphasis on improving security. Users were categorised based on their access rights and duties using the Role-Based Access Control (RBAC) principles. Users who exhibited actions that differed from the expected patterns were given labels for aberrant behaviour. The ground truth for model training and evaluation was provided by this categorisation.

4) Data transformation: Several properties, particularly those involving API call sequences, were initially supplied in a sequential style during the data pre-processing phase. Data transformation was used to convert these sequential features into formats suitable for numerical representation and embedding vectors in order to facilitate efficient model training and analysis. This change preserved the actions' natural temporal order and made sure they worked perfectly with the selected classification methods.

5) Handling imbalanced data: The significant disparity between the number of normal examples and the comparatively few cases exhibiting aberrant behaviour creates an inherent barrier in the context of anomaly identification scenarios. Strategic strategies were used to solve this class disparity, including oversampling, under sampling, and the creation of synthetic data. This phase's main goal was to eliminate any

biases that would have caused models to favour the more common class. The main objective of these strategies for handling unbalanced data was to prevent the training of the models from being influenced by the dominant class. The models could learn the complex patterns and distinctions present in both normal and abnormal behaviours by correcting the skewed distribution. The models' ability to recognise and categorise aberrant activity accurately was further strengthened, matching with the larger goal of improving security and dependability within the IoT ecosystem.

6) Dataset splitting: After careful pre-processing, the dataset was divided into three separate subsets: training, validation, and testing. This partitioning technique was designed to provide a thorough assessment of the performance of the classification models while ensuring their ability to successfully generalise to unexplored data cases. The dataset was divided into several unique subsets, which allowed for a methodically organised review process. Precision was used in the models' training, improvement, and evaluation to make sure the final models could identify and classify aberrant behaviours with accuracy. This comprehensive approach was in line with the general goal of improving security and dependability within the IoT ecosystem through precise classification.

7) Normalization and standardization: Normalization or standardization techniques were applied to numerical features to bring them to a common scale. This process ensured that features with varying magnitudes did not disproportionately impact the model's learning process.

8) Data integrity and consistency check: Final checks were performed to confirm the integrity and consistency of the pre-processed dataset. This step aimed to identify any anomalies introduced during pre-processing and ensure that the data adhered to the expected formats and distributions.

The results of these pre-processing procedures formed the basis for the analysis that followed, which used classification techniques to spot unusual user behaviours in the IoT ecosystem. The pre-processed dataset was used to apply the Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF) algorithms, allowing for a thorough evaluation of the efficiency of these techniques in identifying abnormal activity. The overarching objective of improving security and dependability in IoT services benefits from precise user classification and the detection of anomalous behaviours.

D. Evaluation Methods

The assessment of the proposed system's efficacy in enhancing security and dependability in the Internet of Things (IoT) ecosystem involved rigorous evaluation methods. This section outlines the methodologies employed to gauge the performance of the classification models and the overall system.

1) Performance metrics: A suite of performance metrics was chosen to comprehensively evaluate the classification models' accuracy in identifying and categorizing user behaviors. These metrics encompassed:

- Accuracy: The proportion of correctly classified instances relative to the total instances. It provided an overall measure of the models' effectiveness.
- Precision: The ratio of true positive predictions to the total predicted positive instances. Precision quantified the models' ability to accurately label abnormal behaviours without falsely labelling normal ones.
- Recall: The ratio of true positive predictions to the total actual positive instances. Recall gauged the models' capability to identify all instances of abnormal behaviour.
- F1-Score: The harmonic mean of precision and recall. This metric provided a balanced measure of the models' precision-recall trade-off.

2) Cross-Validation: To reduce potential bias and model performance fluctuation, cross-validation was used. When using K-fold cross-validation, the dataset was split into K separate subgroups. Each subset was used as the testing set once and the remaining subsets as the training data as the models were trained and tested K times. This method gave a reliable estimate of the models' typical performance.

3) Comparison of classification algorithms: The effectiveness of different classification algorithms—Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF)—was compared. The models' accuracy, precision, recall, and F1-score were assessed individually for each algorithm. This comparison highlighted the strengths and weaknesses of each approach.

The effectiveness of the suggested strategy in enhancing IoT security and dependability was carefully assessed through these evaluation techniques. This evaluation process validated the system's potential for accurately identifying and categorizing user behaviors, ultimately contributing to the overarching goal of improving IoT ecosystem security. It did this by taking into account a variety of performance metrics, utilizing cross-validation, and assessing the models' response to unseen data.

VI. Implementation

In this section, we provide an overview of the practical implementation of the proposed ML-based RBAC system designed to enhance security and dependability within the Internet of Things (IoT) ecosystem. We detail the technical aspects, tools, and technologies employed to realize the system's functionalities and objectives.

1) Data preparation: The implementation commenced with the acquisition of the dataset from Kaggle. Data cleaning procedures were executed to address missing values, inconsistencies, and noise. The dataset's unique identifier ("_id") was utilized for data linkage with external sources.

2) Feature engineering: Relevant features were selected based on their significance to the problem at hand, as outlined in the Data Pre-processing section. The transformation of sequential attributes, such as API call sequences, was carried out to ensure compatibility with subsequent analysis.

3) Role-Based Access Control (RBAC): In the designed system, the strategic implementation of Role-Based Access Control (RBAC) principles played a pivotal role in categorizing users and fortifying the security and dependability of the Internet of Things (IoT) ecosystem. RBAC, a well-established security framework, offers a systematic and dynamic approach to manage user access, permissions, and responsibilities. By employing RBAC, the system efficiently structured the complex landscape of user interactions and access rights, contributing to the accurate classification of user behaviours.

a) Defining roles and responsibilities: The definition of distinct roles, each of which represents a set of duties and permissions inside the IoT ecosystem, forms the basis of RBAC. Roles mirror user personas or positions and include duties, behaviours, and functionalities according to that role. These roles were carefully created to correspond with the wide range of actions carried out by users, from end users to administrators.

b) Mapping permissions to roles: Permissions were carefully linked to the stated roles, capturing the actions and processes that users are allowed to do. This mapping protected against unauthorised access to essential resources or functionality by ensuring that each role had access to a specific set of permissions.

c) User-Group associations: RBAC takes into account affiliations between users in addition to individual users. Users who shared comparable duties and permissions were grouped together, making it easier to manage access privileges and improving the system's scalability. The assignment and changing of permissions was made simpler by group-based administration, especially in cases where there were many users.

i) Structured framework for behaviour classification: The user behaviour classification process was made possible by the RBAC architecture. Users' behaviours were observed as they engaged with the IoT ecosystem and compared to the pre-set roles and permissions. The algorithm was able to distinguish between normal and abnormal behaviours by looking at the order and pattern of these interactions.

ii) Benefits and advantages: The implementation of Role-Based Access Control (RBAC) offered a multitude of advantages to the system: Users were only given the rights required for their assigned responsibilities thanks to RBAC's fine-grained control over access to resources and capabilities. This strategy greatly reduced the possibility of unauthorised operations and improved general security. The RBAC framework's built-in dynamic adaptability proved invaluable since it made it easier to make in-the-moment adjustments in response to shifting responsibilities or personnel changes.

4) Classification algorithms: The selected classification algorithms—Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF) were implemented using appropriate libraries or frameworks. Model parameters were fine-tuned through cross-validation and hyper parameter tuning.

5) Performance evaluation: Various metrics, including accuracy, precision, recall, and F1-score, were used to systematically assess the performance of the classification models. The models' robustness and generalisation abilities were tested by cross-validation and testing on fictitious data.

6) Outcome and insights: The system's installation produced insights into the classification of user behaviour and the identification of anomalous activities. High classification accuracy rates confirmed the suggested system's efficacy in boosting IoT security and dependability.

7) Deployment and integration: Depending on the situation, the system's outputs might be incorporated into the IoT infrastructure already in place, boosting security features and assisting with user access decisions.

In Fig. 2, integrating Role-Based Access Control (RBAC) with encryption key management based on anomalous user behaviour is a sophisticated approach to enhancing security in a system by the following steps:

1) RBAC: RBAC or role-based access control, is a security model that limits system access by specifying roles and permissions. Each user is given a unique role, and that position dictates which permissions are available to them. An administrator, for example, has different access privileges than a typical user.

2) Anomaly detection: The system continuously watches over network traffic and user behaviour. It recognises behaviour that deviates from established patterns using machine learning algorithms or anomaly detection approaches. Unauthorised access attempts, odd data transfer patterns, or questionable login locations are examples of anomalies.

3) Integration with encryption key management: The RBAC system alerts users when it notices unusual behaviour. This alert is sent to the RBAC-integrated encryption key management system. The encryption key management system decides whether to start a key change based on the severity and kind of the anomaly. The encryption key management system generates a new encryption key if a key update is deemed essential. To maintain secure connection, the new key is safely delivered to authorised users or devices.

4) User notifications: Users might receive notifications when a key is changed, depending on how the system is configured. Before gaining access to critical information, they might need to re-authenticate or go through additional security procedures.

5) Auditing and logging: For the sake of compliance and auditing, all key changes and the related events are recorded. This results in a thorough record of when and why significant changes took place.

6) Access revocation (if required): In extreme circumstances, the RBAC system may temporarily revoke access permissions until additional investigation is completed if an anomaly signals a significant security danger. RBAC is combined with encryption key management, anomaly detection, and other security components to give the system a

multi-layered security strategy. This makes sure that sensitive data is safeguarded by timely encryption key changes even if an anomaly is found.

VII. Results and Discussion

Understanding the success and applicability of the models that have been implemented requires careful study and interpretation of the results. This section gives a thorough study of the performance measures for the three different methods, Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF), including precision, recall, F1-score, and accuracy. The discussion that follows goes into the implications of these measurements, illuminating the algorithms' capacity for anomaly identification and providing perceptions into their relative advantages and disadvantages.

1) SVM Algorithm: The precision, recall, F1-score, and accuracy metrics of the SVM method are displayed in the performance evaluation. The SVM shows its competence in properly detecting occurrences that are actually positive among those projected as positive, with a precision rate of 90.42%. Furthermore, a recall rate of 91.10% shows that it can successfully record a large percentage of real positive events. The result is a stunning F1-score of 89.95%, showing a harmonious balance between recall and precision. The SVM's accuracy in identifying anomalies and its ability to support effective anomaly detection solutions are highlighted by its overall accuracy of 90.85%.

2) LOF Algorithm: The performance measurements of the LOF algorithm also provide information about its efficacy. With a precision rate of 88.20%, LOF correctly categorises instances of genuine positives among its anticipated positives. The recall rate of 89.21% demonstrates its capacity to recognise a substantial number of true positive cases. The algorithm's well-balanced trade-off between recall and precision is highlighted by the F1-score of 88.58%. The LOF algorithm has commendable performance in detecting abnormalities with an accuracy of 89.00%.

3) IF Algorithm: The measurements of the Isolation Forest method also add to the conversation. IF effectively separates actual positives from anticipated positives with a precision rate of 87.70%. Furthermore, a recall rate of 88.65% shows that it can effectively capture a sizable number of real positive cases. The algorithm's ability to create a harmonious balance between recall and precision is demonstrated by the resulting F1-score of 87.91%. at the end, the accuracy of 88.39% highlights IF's skill at spotting abnormalities.

Table I compares the performance of three machine learning models: the Support Vector Machine (SVM), the Local Outlier Factor (LOF), and the Isolation Forest (IF). The accuracy of a model's positive predictions is measured by its precision, with SVM having the highest precision (90.42%). SVM outperforms the competition with a recall rate of 91.10%, also known as the true positive rate, which measures how well a model detects actual positive cases. SVM has the best overall performance, scoring 89.95% on the F1-Score, which balances recall and precision. Last but not least, a model's accuracy is defined as the percentage of accurate predictions made out of all instances, with SVM leading the field with 90.85%. These metrics evaluate these models' classification ability overall, with SVM consistently exhibiting the best reliable results.

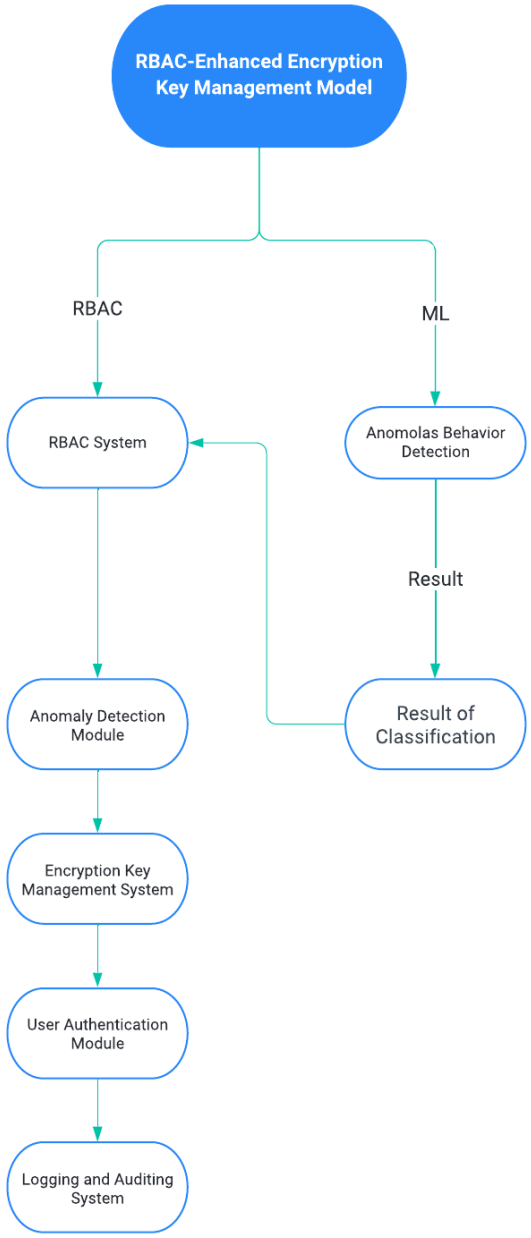


Fig. 2. RBAC-enhanced encryption.

TABLE I. Result of ML Algorithms

MODEL	PRECISION	RECALL	F1-SCORE	ACCURACY
SVM	90.42%	91.10%	89.95%	90.85%
LOF	88.20%	89.21%	88.58%	89.00%
IF	87.70%	88.65%	87.91%	88.39%

The study compares the three machine learning algorithms—Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF)—to determine their

effectiveness in detecting anomalies in API call graphs. Each algorithm has its strengths and weaknesses:

- SVM: High precision (90.42%), recall (91.10%), and F1-score (89.95%), indicating strong performance in distinguishing between normal and outlier classes.
- LOF: Balanced trade-off between precision (88.20%) and recall (89.21%), suitable for identifying regional outliers.
- IF: Effective at isolating outliers with a precision of 87.70% and recall of 88.65%.

The study suggests that instead of selecting a single best algorithm, it is essential to analyse their strengths and weaknesses.

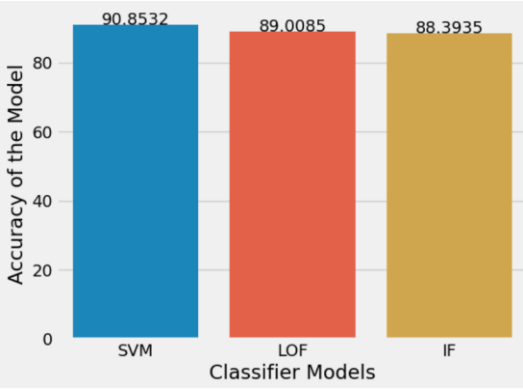


Fig. 3. Comparison of algorithms used for abnormal behaviour detection.

In Fig. 3, the result indicates the performance metrics of three different categorization methods: Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF) in the context of the results that have been provided. These metrics are crucial for assessing how well these strategies work in spotting unusual behaviour in the dataset.

The accuracy obtained for SVM is 90.85%. This shows that roughly 90.85% of the dataset's occurrences were correctly classified by the SVM model. This is a comparatively high level of accuracy, indicating that the SVM did a good job of differentiating between regular and abnormal behaviour. The accuracy reached in the instance of LOF is 89.00%. This indicates that about 89.00% of occurrences were correctly identified by the LOF model. This is still a commendable accuracy score, even though it is a little lower than SVM, showing that LOF is adept at spotting abnormalities. The Isolation Forest (IF), last but not least, recorded an accuracy of 88.39%. This means that the IF model accurately identified aberrant behavior in roughly 88.39% of situations, which is a decent result.

To achieve the objectives of our work, we employed an incremental strategy to address the challenge of accurately identifying atypical user behavior in Internet of Things (IoT) systems. Initially, a Support Vector Machine (SVM) classifier was developed using a basic feature set that included the frequency of API usage, sequence length, and session duration. The first evaluation yielded promising results, achieving an F1-score of 84%, accuracy of 89%, precision of 85%, and recall of 83%. To enhance performance further, additional features were incorporated, such as temporal dynamics, IP type, and the volume of individual APIs accessed. This optimization significantly improved the SVM classifier's performance, increasing accuracy to 92%, precision to 88%, recall to 86%, and F1-score to 87%.

Using this progress, the Local Outlier Factor (LOF) approach identified anomalies in unlabeled data. Applying a density-based approach, LOF achieved a 90% anomaly detection accuracy. Comparative study reveals that the combination of LOF with SVM improved the general accuracy and precision of the system, so generating a combined accuracy of 93%. Larger datasets and complex anomalies were managed with the Isolation Forest (IF) technique, therefore enhancing scalability and robustness. Using an ensemble method, pooling SVM, LOF, and IF predictions demonstrated the system improved recognition and classification of aberrant user activity. This work presents the detection of abnormalities and user behavior classification capability of the suggested structure on Internet of Things systems. Still, further research is required to offer a solid basis for the conclusions. By way of comparison, modern methods such hybrid approaches or deep learning-based anomaly detection models could provide a fairer evaluation of the performance of the framework.

Future studies should investigate the scalability and flexibility of the framework inside multiple IoT environments by assessing its relevance over several datasets. These projects will raise the feasibility of the given solutions and point out areas needing greater improvement.

VIII. Conclusions and Future Work

This paper explores the crucial area of API user behavior analysis, with a particular emphasis on identifying possibly malicious users who significantly deviate from standard API usage patterns. We achieve this by using the Support Vector Machine (SVM), Local Outlier Factor (LOF), and Isolation Forest (IF) algorithms, each of which offers a different method for classifying API call graphs as normal or outliers. Our dataset consists of API call graphs that users build when they use APIs to access particular business logic.

We preprocess the data and make balanced training and testing sets with both normal and outlier graphs to ensure a reliable assessment. The SVM algorithm successfully distinguishes between actual positive cases and projected positives with a remarkable precision rate of 90.42%. Its recall rate of 91.10% demonstrates its capacity to record a significant fraction of genuine happy occurrences, yielding a pleasing F1-score of 89.95%. At 90.85%, the SVM excels in overall accuracy. Moving on to the LOF algorithm, it performs admirably, properly classifying occurrences of genuine positives among predicted positives with an accuracy rate of 88.20%. With an F1-score of 88.58%, its recall rate of 89.21% further demonstrates its ability to identify a sizable number of true positive cases. The accuracy of LOF is 89.00%. Last but not least, the Isolation Forest (IF) method, with an impressive 87.70% precision rate, demonstrates its capacity to discriminate actual positives from predicted positives. With an F1-score of 87.91%, it is effective at collecting a sizable proportion of true

positive instances thanks to a recall rate of 88.65%. The accuracy of IF is 88.39%.

As a result of a thorough evaluation of several machine learning models, SVM emerged as the model with the best precision, recall, F1-score, and accuracy. To choose the best strategy for anomaly identification in API user behavior analysis, the strengths and weaknesses of each model must be taken into account. By identifying potentially dangerous conduct among API users, these findings provide insightful information for improving security procedures.

Despite the proposed architecture's impressive performance in anomaly detection and user classification, several areas warrant further research for improvement. For instance, incorporating deep learning techniques like Recurrent Neural Networks (RNNs) and transformer-based models could enhance the detection of complex temporal patterns in user activity. Expanding the dataset to include a wider variety of IoT scenarios and types of malicious behavior would further validate the system's resilience.

Additionally, optimizing the framework for real-time anomaly detection is crucial for computational efficiency. Employing explain ability techniques, such as LIME or SHAP, would clarify the classifier's decision-making process, thereby enhancing trust and transparency in its applications. Incorporating feedback loops that allow identified anomalies to inform future access control guidelines would create a dynamic and adaptive security solution.

REFERENCES

[1] Munirathinam, S. (2020). Industry 4.0: Industrial internet of things (IIOT). In Advances in computers (Vol. 117, No. 1, pp. 129-164). Elsevier.

[2] Sicari, S., Rizzardi, A., & Coen-Porisini, A. (2020). 5G In the internet of things era: An overview on security and privacy challenges. Computer Networks, 179, 107345.

[3] Alagappan, A., Andrews, L. J. B., Raj, R. A., & Sarathkumar, D. (2022, December). Cybersecurity Risks Quantification in the Internet of Things. In *2022 IEEE 7th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)* (Vol. 7, pp. 154-159). IEEE.

[4] Joseph, K. T. (2023, June). Analysis on IoT Networks Security: Threats, Risks, ESP8266 based Penetration Testing Device and Defense Framework for IoT Infrastructure. In *2023 3rd International Conference on Intelligent Technologies (CONIT)* (pp. 1-7). IEEE.

[5] Prasad, A., Kapoor, P., & Singh, T. P. (2024). Security Threats in IOT and Their Prevention. In *Communication Technologies and Security Challenges in IoT: Present and Future* (pp. 131-146). Singapore: Springer Nature Singapore.

[6] Mehra, T. (2024). The Critical Role of Role-Based Access Control (RBAC) in securing backup, recovery, and storage systems. *International Journal of Science and Research Archive*, *13*(1), 1192-1194.

[7] Shakarami, M., & Sandhu, R. (2021, April). Role-based administration of role-based smart home IoT. In *Proceedings of the 2021 ACM Workshop on Secure and Trustworthy Cyber-Physical Systems* (pp. 49-58).

[8] Aftab, M. U., Oluwasanmi, A., Alharbi, A., Sohaib, O., Nie, X., Qin, Z., & Ngo, S. T. (2021). Secure and dynamic access control for the Internet of Things (IoT) based traffic system. *PeerJ Computer Science*, *7*, e471.

[9] Jalali, N., Sahu, K. S., Oetomo, A., & Morita, P. P. (2020). Understanding user behavior through the use of unsupervised anomaly detection: proof of concept using internet of things smart home thermostat data for improving public health surveillance. *JMIR mHealth and uHealth*, *8*(11), e21209.

[10] Rawat, R., Kassem, A. A., Dixit, K. K., Deepak, A., Pushkarna, G., & Harikrishna, M. (2024, May). Real-Time Anomaly Detection in Large-Scale Sensor Networks using Isolation Forests. In *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)* (pp. 1400-1405). IEEE.

[11] Karthiga, S., Ravisankar, P., Vijayarajeswari, R., Pushpa, N., Vino, T., & Dobhal, D. (2024, May). Machine Learning-based Anomaly Detection in IOT Sensing Devices for Optimal Security. In *2024 Second International Conference on Data Science and Information System (ICDSIS)* (pp. 1-6). IEEE.

[12] Rani, P. S., Ahamed, M. V., Chaithresh, K. S. S., Srinivas, S. K., & Vivek, P. V. (2024, April). Utilizing Machine Learning Techniques for Detecting Anomalies in IoT Networks. In *2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)* (pp. 105-110). IEEE.

[13] El-Sofany, H., El-Seoud, S. A., Karam, O. H., & Bouallegue, B. (2024). Using machine learning algorithms to enhance IoT system security. *Scientific Reports*, *14*(1), 12077.

[14] Alqahtani, A., Alsulami, A. A., Alqahtani, N., Alturki, B., & Alghamdi, B. M. (2024). A Comprehensive Security Framework for Asymmetrical IoT Network Environments to Monitor and Classify Cyberattack via Machine Learning. *Symmetry*, *16*(9), 1121.

[15] Rao, D. D., Waoo, A. A., Singh, M. P., Pareek, P. K., Kamal, S., & Pandit, S. V. (2024). Strategizing IoT Network Layer Security Through Advanced Intrusion Detection Systems and AI-Driven Threat Analysis. *Full Length Article*, *12*(2), 195-95.

[16] Fragkos, G., Johnson, J., & Tsiropoulou, E. E. (2022). Dynamic role-based access control policy for smart grid applications: an offline deep reinforcement learning approach. IEEE Transactions on Human-Machine Systems, 52(4), 761-773.

[17] Thakare, A., Lee, E., Kumar, A., Nikam, V. B., & Kim, Y. G. (2020). PARBAC: Priority-attribute-based RBAC model for azure IoT cloud. IEEE Internet of Things Journal, 7(4), 2890-2900.

[18] Khraisat, A., & Alazab, A. (2021). A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. Cybersecurity, 4, 1-27.

[19] Negi, K., Kumar, G. P., Raj, G., Sahana, S., & Jain, V. (2022, January). Degree of accuracy in credit card fraud detection using local outlier factor and isolation forest algorithm. In 2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 240-245). IEEE.

[20] Nie, F., Zhu, W., & Li, X. (2020). Decision Tree SVM: An extension of linear SVM for non-linear classification. Neurocomputing, 401, 153-159.

[21] Alghushairy, O., Alsini, R., Soule, T., & Ma, X. (2020). A review of local outlier factor algorithms for outlier detection in big data streams. Big Data and Cognitive Computing, 5(1), 1.

[22] Staerman, G., Mozharovskyi, P., Clémençon, S., & d'Alché-Buc, F. (2019, October). Functional isolation forest. In Asian Conference on Machine Learning (pp. 332-347). PMLR.

[23] Jiang, J., Han, G., Shu, L., & Guizani, M. (2020). Outlier detection approaches based on machine learning in the internet-of-things. IEEE Wireless Communications, 27(3), 53-59.

[24] Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. A. (2020). Performance analysis of machine learning algorithms in intrusion detection system: A review. Procedia Computer Science, 171, 1251-1260.

[25] M. A. Bhatti, R. Riaz, S. S. Rizvi, S. Shokat, F. Riaz and S. J. Kwon, "Outlier detection in indoor localization and Internet of Things (IoT) using machine learning," in Journal of Communications and Networks, vol. 22, no. 3, pp. 236-243, June 2020, doi: 10.1109/JCN.2020.000018.

[26] F. Hussain, R. Hussain, S. A. Hassan and E. Hossain, "Machine Learning in IoT Security: Current Solutions and Future Challenges," in IEEE Communications Surveys & Tutorials, vol. 22, no. 3, pp. 1686-1721, thirdquarter 2020, doi: 10.1109/COMST.2020.2986444.

[27] https://www.kaggle.com/code/tangodelta/user-behavior-classification last visit 16/8/2024

A Real-Time Nature-Inspired Intrusion Detection in Virtual Environments: An Artificial Bees Colony Approach Based on Cloud Model

Ayanseun S. Ayanboye[1], John E. Efiong[2], Temitope O. Ajayi[3], Rotimi A. Gbadebo[4], Bodunde O. Akinyemi[5], Emmanuel A. Olajubu[6], Ganiyu A. Aderounmu[7]

Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria[1, 2, 3, 4, 5]
CyberSCADA Research Lab, Department of Computer Science & Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria[6]
CyberSCADA Research Lab, Africa Center of Excellence, OAU ICT-driven Knowledge OAK Park,
Obafemi Awolowo University, Ile-Ife, Nigeria[7]

***Abstract*—Real-time intrusion detection in virtual environments is crucial for maintaining the security and integrity of modern computing infrastructures. This paper proposes a nature-inspired mathematical model designed to detect both known and unknown attacks on virtual machines, focusing on enhancing detection accuracy and minimizing false alarm rates. The proposed model, named Developed Artificial Bee Colony Optimization Based on Cloud Model (DABCO_CM), is inspired by the foraging behavior of bee swarms and integrates principles from the Artificial Bee Colony algorithm and the cloud model rooted in fuzzy logic theory. The model was simulated using the UNSW_NB15 datasets in Google Colab and benchmarked against an existing model. It achieved a detection accuracy of 97.98%, compared to the existing model's 95.35%. Sensitivity results showed 99.92% for the proposed model, compared to 96.90% for the existing model, while specificity increased to 93.86%, in contrast to the existing model's 90.71%. These findings demonstrate a 3.02% increase in sensitivity, a 2.63% increase in accuracy, and a 3.15% increase in specificity, highlighting the model's superior capability in detecting attacks and its potential to learn from unlabeled data, addressing key challenges in virtual machine security.**

Keywords*—Real-time intrusion detection; virtual environments; artificial bee colony algorithm; cloud model algorithms; intrusion detection system; feature selection; classification; swarm intelligence; fuzzy logic; DNN; ABC_DNN DABCO_CM

I. Introduction

Virtual environments, encompassing applications such as email, chat, and online document sharing, operate within shared operating systems that allow Virtual Machines (VMs) to function like physical computers without specific hardware components. Virtualization abstracts the complexity of hardware and software, typically implemented using hypervisor devices that efficiently allocate system resources among multiple VMs [1], [2]. Security in virtual environments relies on traditional tools like firewalls and intrusion detection systems (IDS), which are either network-based or host-based, though they have limitations in detecting new threats [3] [8]. Traditional IDS methods often fall short in virtual environments due to their inability to handle the dynamic and complex nature of modern cyber threats effectively [9] [11]. These systems struggle with high false alarm rates and limited accuracy in detecting new, unknown attacks. This paper addresses these challenges by proposing a bio-inspired mathematical model that integrates the Artificial Bee Colony (ABC) algorithm with fuzzy logic to enhance detection accuracy and minimize false alarms in virtualized settings. The proposed model, named Developed Artificial Bee Colony Optimization Based on Cloud Model (DABCO_CM), draws inspiration from the foraging behavior of bee swarms and leverages principles from the cloud model rooted in fuzzy logic theory [12], [15].

The DABCO_CM model improves upon previous approaches in several key ways:

- Enhanced Detection Accuracy: By optimizing feature selection and classification processes through the ABC algorithm, the model can more accurately identify relevant patterns and anomalies, leading to higher detection accuracy.
- Reduced False Alarms: The integration of cloud models helps manage uncertainty and imprecision in data analysis, significantly reducing false positive rates compared to traditional IDS.
- Real-Time Detection: The model's design principles ensure fast processing times, making it capable of real-time intrusion detection, which is crucial for maintaining the security of dynamic virtual environments.
- Scalability and Adaptability: The hybrid approach of combining swarm intelligence with fuzzy logic makes the model more robust and adaptable to evolving cyber threats, ensuring better performance in large-scale and complex virtual environments.

The subsequent sections of this paper are structured as follows:

- Literature Review: Discusses the application of Artificial Bee Colony (ABC) algorithms in detecting Distributed Denial-of-Service (DDoS) attacks within virtual environments and highlights the need for further optimization.

- Methodology: Describes the development process of the DABCO_CM model, detailing its design principles, components, and integration of ABC and cloud model algorithms.
- Simulation and Testing: Explains the simulation environment and the datasets used to evaluate the model, including performance metrics and evaluation parameters.
- Results and Discussion: Presents the simulation results, comparing the performance of the proposed model against existing models, and discusses the findings.
- Conclusion and Future Work: Summarizes the study's contributions, highlights the model's effectiveness, and suggests directions for future research.

II. Literature Review

The application of Artificial Bee Colony (ABC) algorithms in detecting DDoS attacks within virtual environments is a relatively familiar area with limited research. Despite its proven effectiveness, particularly in filtering and mitigating attacks, there is a recognized need for further research to optimize ABC algorithm convergence in large-scale virtualized settings. Studies by [6] and [15] collectively highlight the need to enhance ABC algorithms to improve their robustness and scalability in complex virtualized environments [6], [8]. Effective intrusion detection in virtual environments has been the focus of numerous studies [4][5], highlighting the need for innovative approaches to overcome the limitations of traditional IDS [14]. The study in [1] highlighted the substantial threat posed by distributed denial of service (DDoS) attacks to network security in cloud computing. The study employed a deep learning classifier to differentiate between attacked and non-attacked data. The proposed FS-WOA–DNN framework effectively protected against DDoS attacks, achieving a better detection accuracy, outperforming other existing DDoS detection models. The research in [13] addressed the impact of DoS and DDoS attacks on cloud services. Using artificial immune systems, they identified attack characteristics, achieving high detection accuracy and low false alarm rates. This method effectively maintained cloud service availability, reducing financial losses and preserving reputations. Fig. 1 shows existing model.

A. Real-Time Intrusion Detection

Real-time intrusion detection requires systems capable of processing large volumes of data quickly and accurately [2] highlight the importance of high-performance algorithms in achieving timely threat detection [7], [8]. They emphasize that traditional methods often fall short due to their inability to handle the dynamic nature of modern cyber threats effectively [9].

B. Artificial Bee Colony Algorithm

The Artificial Bee Colony (ABC) algorithm, proposed by [4] [6] have shown promise in various optimization problems, including intrusion detection. Its ability to efficiently search for optimal solutions makes it suitable for enhancing IDS performance. The study in [18] further explore the application of ABC in optimizing feature selection and classification tasks, demonstrating its potential for improving IDS accuracy and efficiency.

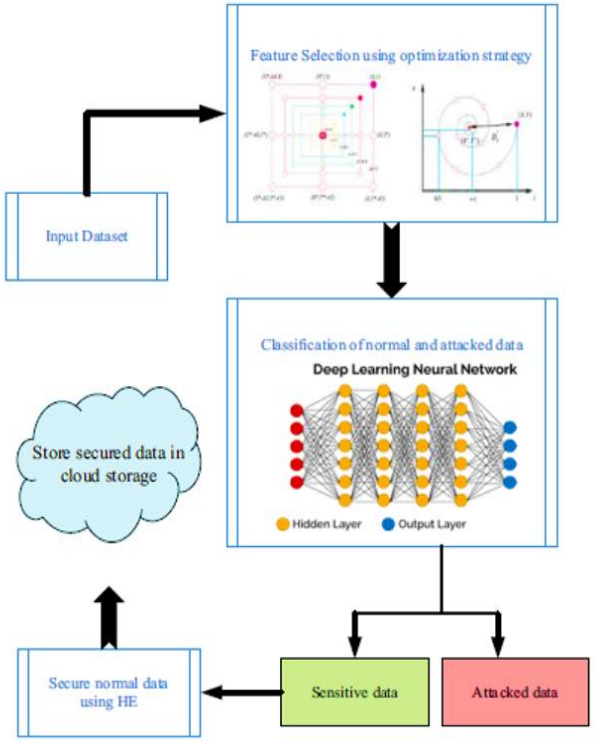


Fig. 1. The existing model (Agarwal et al. 2022).

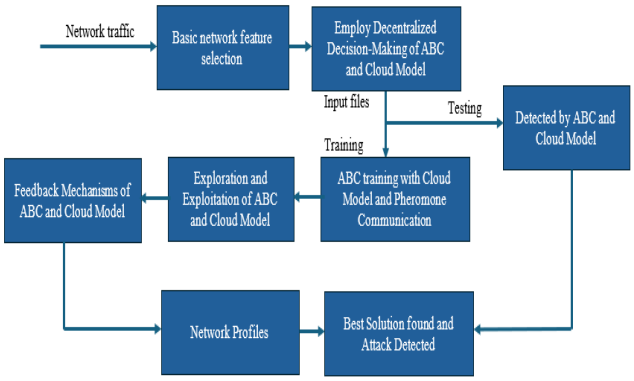


Fig. 2. Proposed model.

Fig. 2 shows proposed model and Fig. 3 depicts flowchart of the proposed model.

C. Cloud Model Algorithms

Cloud models, which integrate fuzzy logic and probability theory, offer a powerful tool for handling uncertainty and imprecision in data analysis. The studies in [6][10][14][15] discuss the application of cloud models in intrusion detection, highlighting their capability to improve detection accuracy by managing uncertainties. Their research underscores the benefits of combining cloud models with other optimization techniques to enhance IDS performance.

Africa Center of Excellence, OAU ICT-driven Knowledge OAK Park, Obafemi Awolowo University, Ile-Ife, Nigeria

D. Integrated Approaches

Integrating the ABC algorithm with cloud models has been explored to address the limitations of each technique individually. The studies in [6][16] present a hybrid approach that combines swarm intelligence with fuzzy logic to improve IDS performance. Their study demonstrates that such integration can offer a more robust and adaptable solution for real-time intrusion detection.

III. METHODOLOGY

The development of the DABCO_CM model involves several key steps to integrate the Artificial Bee Colony (ABC) algorithm with cloud model algorithms, aiming to enhance real-time intrusion detection in virtual environments:

A. Objectives and Requirements

The primary objectives for real-time intrusion detection are to achieve high accuracy, minimize false alarms, and ensure fast processing times. Requirements include the ability to handle large volumes of data generated in virtual environments and adapt to evolving cyber threats. The DABCO_CM model is designed to meet these objectives by combining advanced optimization techniques with robust data handling capabilities.

B. Design Principles

The DABCO_CM model is built on the principles of swarm intelligence and fuzzy logic. The ABC algorithm is used to optimize feature selection and classification processes, enhancing the model's ability to identify relevant patterns and anomalies. Cloud model algorithms are integrated to manage uncertainty and provide a probabilistic framework for intrusion detection.

C. Components

The DABCO_CM model consists of the following components:

- Feature Selection Module: Utilizes the ABC algorithm to select the most relevant features for intrusion detection. This module aims to improve detection accuracy and reduce computational overhead.
- Classification Module: Applies machine learning techniques to classify intrusion attempts based on selected features. This module uses advanced algorithms to enhance the accuracy of threat detection.
- Cloud Model Integration: Incorporates cloud models to handle uncertainty and improve detection accuracy. The cloud model algorithms provide a probabilistic approach to managing imprecise data and enhancing overall system performance.

D. Proposed Model Design

The proposed model integrates the Artificial Bee Colony (ABC) optimization algorithm and the Cloud Model algorithm to enhance intrusion detection in virtual environments.

1) Network traffic: Network traffic refers to the flow of data across a computer network, encompassing all the data sent and received by devices connected to the network.

2) Basic network feature selection: Feature selection involves choosing a subset of relevant features for a machine learning model using methods like filter, wrapper, and embedded techniques. This balance ensures the model's effectiveness. Fig. 4 shows diagrammatic representation of the proposed model.

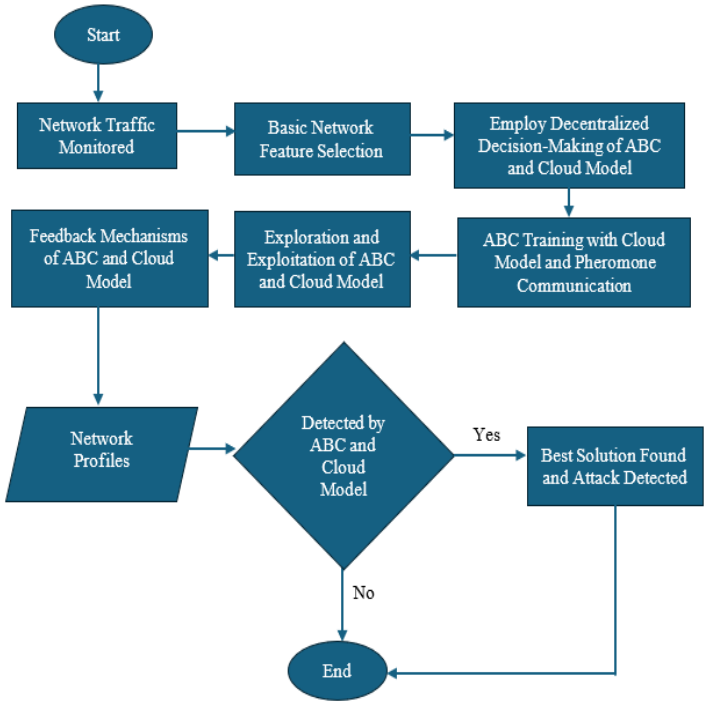


Fig. 3. Flowchart of the proposed model.

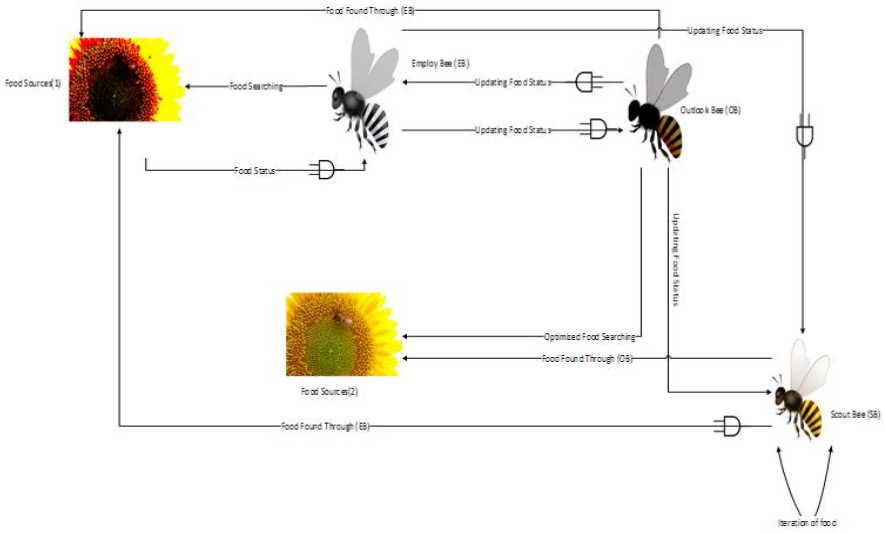


Fig. 4. Diagrammatic representation of the proposed model.

3) Decentralized decision-making with ABC and cloud model: The ABC algorithm, inspired by bee foraging, and the Cloud model, which handles uncertainty, collaboratively improve system performance by enabling autonomous and informed decision-making.

4) ABC training with cloud model and pheromone communication: Combining the ABC algorithm with the Cloud model and pheromone communication enhances decision-making and collaboration among agents, leading to better performance in virtual environments.

5) Exploration and exploitation of ABC and cloud model: The ABC algorithm optimizes solutions by mimicking bee behavior, while the Cloud model leverages distributed computing. This integration enables efficient resource allocation and problem-solving.

6) Feedback mechanisms of ABC and cloud model: ABC and Cloud models use feedback to improve decision-making and predictions. The ABC algorithm relies on bee feedback, and the Cloud model refines its functions based on various inputs.

7) Network profiles: Network profiles represent behavioral patterns of networks. By analyzing these profiles with ABC optimization and Cloud Model, intrusion detection systems can identify suspicious activities.

8) Best solution found and attack detected: Integrating ABC and Cloud models optimizes solutions and enhances system security by detecting potential attacks in virtual environments. The framework combined the ABC algorithm and the Cloud Model algorithm to create an optimized model with the following key elements:

a) Objective: The primary goal of the framework had been to develop a highly efficient intrusion detection model by integrating the ABC algorithm and the Cloud Model algorithm.

b) Feature selection: To improve the accuracy of intrusion detection, the framework had utilized the ABC algorithm to optimize the feature selection process. This involved selecting the most relevant features that significantly contributed to identifying intrusions. An effective method to achieve feature selection for improving intrusion detection accuracy is to employ machine learning techniques such as recursive feature elimination (RFE).

c) Uncertainty dandling: The Cloud Model algorithm had played a vital role in handling uncertainties associated with intrusion detection. It combined probability-based reasoning, adaptive exploration, dynamic solution adjustment, ensemble techniques, and continuous learning for robust and adaptive solutions in the face of uncertainty.

d) Integration: The ABC and Cloud Model algorithms had been seamlessly integrated into a unified framework. The ABC algorithm optimized the feature selection procedure, while the Cloud Model algorithm addressed uncertainty reasoning. ABC algorithm scouts for the most promising features, much like bees searching for the best flowers. The Cloud Model carefully examines those features to assess their uncertainty, ensuring a reliable selection. By working together, they create a robust decision-making framework that considers both feature relevance and potential uncertainties.

e) Real-Time detection and response: One of the key objectives of the DABCO_CM framework has been to enable real-time detection of intrusions. It has empowered the system to promptly identify and respond to both known and unknown intrusions occurring in the virtual environment. This capability enhanced the system's security by enabling swift and effective countermeasures against potential threats as they occur.

f) Evaluation: The effectiveness of the DABCO_CM framework had been evaluated using various metrics, such as intrusion identification, detection precision, and response time. These evaluations provided validation of the framework's performance and reliability. The DABCO_CM framework, leverage the strengths of the ABC algorithm and the Cloud Model algorithm. By optimizing feature selection, handling uncertainties, and facilitating real-time intrusion detection, this framework aimed to enhance the overall effectiveness of intrusion detection systems in virtual environments.

E. Simulation and Testing

This methodology delved into the complexities of creating a groundbreaking system that seamlessly integrated the capabilities of Artificial Bee Colony (ABC) optimization with Cloud Model-based techniques to handle imprecise data. Datasets, specifically UNSW_NB15, were used to simulate two classifiers (ANN and DNN). These classifiers were developed and saved as models in the Google Research Collaboration Environment, and relevant performance metrics were utilized. The ABC optimization for intrusion detection in a virtual environment used 100 random population initializations within bounds. It ran 10 times for a fixed number of iterations, with the dimensionality of the problem space varying. Customizing parameters was essential for effective optimization, with performance evaluated using benchmark functions and real intrusion detection data.

F. Model Formulation of the Proposed Model

The mathematical representation of the Developed Artificial Bee Colony Optimization Based on Cloud Model (DABCO_CM) involved expressing the optimization procedure using mathematical equations. This formulation combined the core concepts of the Artificial Bee Colony (ABC) algorithm and the Cloud Model algorithm to detect intrusions in a virtual environment. During the initialization phase, the population's food sources were randomly created and allocated to the employed bees as in Eq. (1), [6]:

$$X_i^j = X_{min}^j + rand(0,1)\left(X_{max}^j - X_{min}^j\right) \quad (1)$$

Fig. 5 shows Mapping of bee colony concepts to intrusion detection in virtual environments and Table 1 depicts Mapping of Bee Colony Concepts to Intrusion Detection in Virtual Environments.

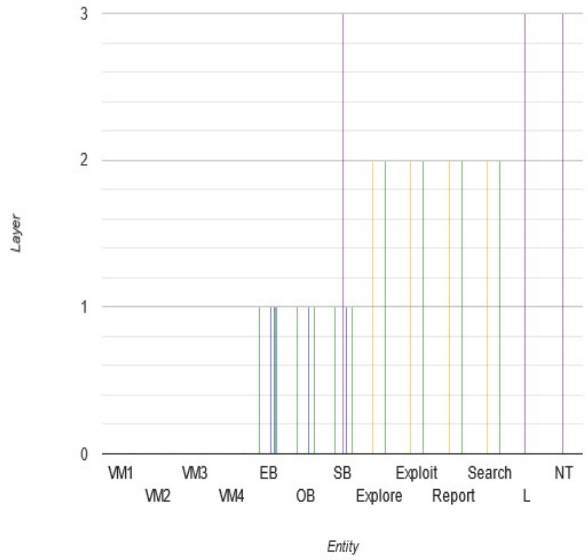


Fig. 5. Mapping of bee colony concepts to intrusion detection in virtual environments.

The upper and lower limits of the solution vectors are denoted by X_{max} and X_{min} respectively in Eq. (2).

$$V_i^j = X_i^j + \varphi_i^j\left(X_i^j - X_k^j\right) \quad (2)$$

Subsequently, the comparison of V_i to X_i is made, and the foraging bee employs a selection mechanism prioritizing the food source with higher fitness, as expressed in Eq. (3).

$$fit_i = \begin{cases} \frac{1}{1+f_1} & f_1 \geq 0 \\ 1 + abs(f_1) & f_1 < 0 \end{cases} \quad (3)$$

This chosen food source, "X_i," was then updated in the same way as the working bees, determined by its probability value "P" as defined in Eq. (4) and the given expression.

$$P = \frac{0.9 fit(X_1)}{\underset{m=1}{\overset{N_e}{max}} fit(X_m)} + 0.1 \quad (4)$$

The formulation of the cloud model could be expressed as follows in Eq. (5), [6]:

$$\forall X \in U \rightarrow \mu(X) \in \{0,1\} \quad (5)$$

Each droplet is uniquely identified by its index i, and N indicates the total number of such droplets.

$$En_i^1 = N(En, He^2) \quad (6)$$

En_i^1 represents a property associated with the cloud droplet i. It is derived as a random value from a normal distribution with mean En and variance.

$$He^2.\, x_i = N(Ex, (En_i^1)^2) \quad (7)$$

This equation defines the x -coordinate of the cloud droplet i. It is generated as a random value from a normal distribution with mean Ex and variance $(En_i^1)^2$.

$$\mu_i = \exp\{-\frac{(x_i - Ex)}{2(En_i^1)^2} \quad (8)$$

This equation defines the μ-coordinate (or y-coordinate) of the cloud droplet i. It is calculated using the exponential function of the negative of the difference between x_i and Ex, divided by $2(En_i^1)^2$.

$$X_I = Ex \pm \sqrt{-2\ln(\mu) * En_i^1} \quad (9)$$

The derivation of this equation involves solving for $\boldsymbol{X_I}$ from the expression of the exponential distribution in Eq. (8).

$$\begin{cases} Ex = \underset{i=1}{\overset{Ne}{\max}} fit_i \\ En = \frac{Ex - fit_i}{12} \\ He = \frac{En}{3} \end{cases} \quad (10)$$

Eq. (10) is derived to define parameters (Ex, En, He) in the ABC algorithm using the cloud model. The derivation emphasizes exploitation by setting Ex as the maximum fitness. En introduces variability based on fitness differences, and He contributes to overall variability. This formulation enhances exploration and exploitation for effective optimization. Table I shows Mapping of Bee Colony Concepts to Intrusion Detection in Virtual Environments.

$$\begin{cases} Ex = \min\limits_{i=1}^{Ne} fit_i \\ En = \frac{fit_i - Ex}{12} \\ He = \frac{En}{3} \end{cases} \quad (11)$$

Eq. (11) is derived to enhance the ABC algorithm using the cloud model, emphasizing less proficient individuals by choosing Ex as the minimum fitness. The formulation promotes exploration and diversification for a more comprehensive search. Therefore, Eq. (10) is designed to exploit the best solutions by focusing on the maximum fitness, improving optimization by refining the search around the most promising solutions while Eq. (11) is designed to explore the search space by focusing on the minimum fitness, encouraging diversification and a comprehensive search for potentially better solutions.

TABLE I. MAPPING OF BEE COLONY CONCEPTS TO INTRUSION DETECTION IN VIRTUAL ENVIRONMENTS

Concept (Bee Colony)	Concept (Intrusion Detection)	Description	Layer
Hierarchy of Roles	Network Layer, Application Layer, Intrusion Detection System Layer	Levels of security architecture	-
Individual Bee/Group with Defined Function	VM, Network Device, Software Application	Elements within the virtual environment	-
VM1, VM2, VM3, VM4	Specific Virtual Machines	Individual VMs within the environment	-
Employed Bee (EB)	Actively Engaged VM	VM searching for potential attack patterns	Layer 1
Onlooker Bee (OB)	Monitoring VM	VM learning from employed bees' findings	Layer 2
Scout Bee (SB)	Searching VM	VM looking for new attack patterns or vulnerabilities	Layer 3
Search	Looking for Something	Process of finding something (nectar or vulnerabilities)	Layer 1, Layer 3
Nectar (NT)	Network Traffic (NT)	Collected data	-
Location (L)	Log Files	Source of information about potential threats	Layer 2
Explore	Scanning and Analyzing	Discovering something (nectar or threats)	Layer 1, Layer 3
Report	Communicating Information	Sharing discoveries (waggle dance or alerts)	Layer 2, Layer 3

$$P = \exp\left\{= \frac{(x-Ex)^2}{2(En^1)^2}\right\} \quad (12)$$

Eq. (12) is derived by adapting the probability density function of a normal distribution to formulate the probability (P) of selecting an individual bee based on its position (x) in relation to cloud model attributes (Ex, En^1).

$$\begin{cases} Ex = X_i^j \\ En = ex \\ He = \frac{En}{10} \end{cases} \quad (13)$$

Eq. (13) in the cmABC algorithm derives $\boldsymbol{Ex}$ as the current position of food sources ($\boldsymbol{X_i^j}$), En as a dynamic variable ($\boldsymbol{ex}$), and $\boldsymbol{He}$ as one-tenth of $\boldsymbol{En}$. This formulation emphasizes the current position, adaptability, and variability in the cloud model, enhancing exploration and exploitation. In this scenario, X_i represented the current position of the food sources, where j belonged to the set {1, 2, ..., D}, and e_x was a variable.

$$\begin{cases} En^1 = En, He^2 \\ V_i^j = N(Ex, En^{i2}) \end{cases} \quad (14)$$

Eq. (14) introduces variability in the cmABC algorithm by defining En^1 as En with He^2 and V_i^j as a normal distribution with mean Ex and variance En^{i2}. This enhances exploration and adaptability, crucial for optimal solution search.

$$ex = -(E_{max} - E_{min})(t/T_{max})^2 \quad (15)$$

Eq. (15) dynamically adjusts ex in the cmABC algorithm. It uses a quadratic function $(t/T_{max})^2$ with a scaled coefficient $-(E_{max} - E_{min})$. The adjustment evolves with the iteration count for improved solution accuracy and controlled exploration.

$$\begin{cases} Ex = X_i^j \\ En = \frac{2}{3}\left|X_i^j - x_k^j\right| \\ He = \frac{En}{10} \end{cases} \quad (16)$$

Eq. (16) in the cmABC algorithm is derived to define cloud model attributes. Ex depends on food source positions in a specific dimension, En incorporates variability between positions, and He contributes proportionally to overall variability, enhancing exploration and exploitation.

G. Feature Selection

There were 42 features in both the trained and tested datasets, and all of these features were relevant in building the model. Some features needed to be removed from the datasets to avoid affecting the model's performance. To determine the relevance of each feature to the target feature, an entropy-based algorithm (mutual information gain) was used to rank the features according to their effectiveness in building the model. Features with non-zero mutual information gain [17] were initially selected to build the model, while the remaining features were discarded.

H. Model Performance Metrics Evaluation Using Parameters

The following performance metrics were used to evaluate the model:

1) Accuracy: this is the ratio of correctly classified positive and negative tuples to the total number of traffic connections. It can be expressed mathematically in Eq. (17) as

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (17)$$

where TP = True Positive: The number of positive tuples correctly classified by the predictive model,

TN= True Negative: The number of negative tuples correctly classified by the predictive model,

FP = False Positive: The number of positive tuples incorrectly classified by the predictive model and

FN = False Negative: The number of negative tuples incorrectly classified by the predictive model.

2) Sensitivity: this is the proportion of actual positives which are predicted positive. Therefore, it measures the effectiveness of the model in predicting positive classes. It is used to determine the false positive rate of the model. It is defined in Eq. (18) as

$$Sensitivity = \frac{TP}{(TP+FN)} \quad (18)$$

3) Specificity: this is the proportion of actual negatives which are predicted negative. Therefore, it measures the effectiveness of the model in predicting negative classes. It is used to determine the false negative rates of the model. It is defined in Eq. (19) as

$$Specificity = \frac{TN}{(TN+FP)} \quad (19)$$

I. Model performance metrics evaluation using graphical tools

The performance of the model was also evaluated using the following graphical tools:

1) Confusion Matrix: A confusion matrix is a matrix whose rows represents the positive and negative values predicted by the classifier and columns represents the actual positive and negative values.

2) Receiver operative curve (ROC): The receiver operative curve evaluates the false positive rates represented on the Y-axis and the true positive rates represented on the X-axis generated during the classification stage. The Area Under the Curve (AUC) shows how well the classifier performs in terms of these two variables (false positive and true positives). The closer it is to 1 the better.

IV. Results and Discussion

A. Simulation Environment

In the study, models were simulated using the Google Research Collaboration Environment. An online Python environment, Colab, facilitated the creation of texts, codes, and the execution of codes to generate output. Python machine learning tools, such as Scikit-learn, pandas, numpy, matplotlib, seaborn, hyperopt, joblib, and scikit-fuzzy libraries, were utilized for dataset analysis, model construction, saving model as pickle file, and evaluating model performance:

- Scikit-learn: Used for machine learning algorithms and model evaluation.
- pandas: Utilized for data manipulation and analysis.
- numpy: Used for numerical operations.
- matplotlib and seaborn: Applied for data visualization.
- hyperopt: Employed for hyperparameter optimization.
- joblib: Used for saving models as pickle files.
- scikit-fuzzy: Utilized for fuzzy logic and operations.

These tools were integral in constructing various models, with the highest-performing model selected for final implementation.

B. Feature Selection

There were 42 features in both the trained and tested datasets, and all of these features were relevant in building the model. Some features needed to be removed from the datasets to avoid affecting the model's performance. To determine the relevance of each feature to the target feature, an entropy-based algorithm (mutual information gain) was used to rank the features according to their effectiveness in building the model. Features with non-zero mutual information gain were initially selected to build the model, while the remaining features were discarded.

V. Simulation Result

The training dataset, which constituted 32%, and the testing dataset, making up 68% of the total dataset, were input into the Google Research Collaboration Python Environment. This environment clustered the dataset into two categories and labeled them as 0 and 1. These labeled datasets were employed to construct six models, allowing them to learn from previous data and enabling them to classify future network instances as either normal or attack instances. After successfully constructing the models, instances from the testing dataset were utilized to assess their performance. The proposed models successfully classified attack instances as true positives (TP) and accurately classified normal instances as true negatives (TN). However, the results also indicated that attack instances were misclassified as normal instances (false negatives - FN), and normal instances were misclassified as attack instances (false positives - FP). The proposed DNN model was able to classify 112, 694 instances as attacks (TP), while 50, 183 were correctly classified as normal (TN). The results also showed that 6, 647 attack instances were misclassified as normal (FN), while 5, 817 normal instances were misclassified as attacks (FP). The proposed ABC_DNN model was able to classify 117, 380 instances as attacks (TP), while 48, 120 were correctly classified as normal (TN). The results also showed that 4, 215 attack instances were misclassified as normal (FN), while 4,626 normal instances were misclassified as attacks (FP). The proposed DABCO_CM_DNN model was able to classify

119250 instances as attacks (TP), while 52, 562 were correctly classified as normal (TN). The results also showed that 91 attack instances were misclassified as normal (FN), while 3, 438 normal instances were misclassified as attacks (FP).

VI. Model Evaluation

In evaluating the performance of the models, four evaluation metrics were used: accuracy, sensitivity, specificity, and error. All these performance metrics were calculated using the results of the confusion matrix table, which were categorized as true negative, true positive, false positive, and false negative, respectively.

VII. Performance Evaluation and Benchmarking

The models developed and benchmarked in this study, including DNN, ABC_DNN, and DABCO_CM, were evaluated based on accuracy, sensitivity, specificity, and error rate. These models were compared to existing models presented in the literature, specifically those by [1] and [13], which employed techniques such as Feature Selection using Whale Optimization Algorithm (FS-WOA) and Artificial Immune Systems (AIS). The FS-WOA-DNN model presented by [1] achieved an accuracy of 95.35%, sensitivity of 96.9%, specificity of 90.71%, and an error rate of 0.0928. In comparison, Prathyusha et al. (2021) focused on ANN, SVM, and AIS models, with the AIS model showing the highest performance, achieving an accuracy of 96.56%, sensitivity of 96.4%, specificity of 91.9%, and an error rate of 0.0713. In this study, the DNN classifier was chosen and optimized using the Artificial Bee Colony (ABC) algorithm, resulting in the ABC_DNN model. The DABCO_CM model was further developed by integrating a cloud model based on fuzzy logic, enhancing the performance of the ABC_DNN model. The performance of the three proposed models demonstrated competitive results. The DNN model achieved an accuracy of 92.89%, sensitivity of 94.43%, specificity of 89.61%, and an error rate of 0.0711. The ABC_DNN model improved upon this with an accuracy of 94.38%, sensitivity of 96.53%, specificity of 91.22%, and an error rate of 0.0562. The DABCO_CM model emerged as the best performer, achieving an accuracy of 97.98%, sensitivity of 99.92%, specificity of 93.86%, and an error rate of 0.0202. These results clearly show that the DABCO_CM model surpasses the FS-WOA-DNN model in terms of accuracy, sensitivity, and specificity. The DABCO_CM model's sensitivity of 99.92% indicates its strong ability to identify both known and unknown attacks, while its specificity of 93.86% underscores its effectiveness in accurately distinguishing between attacks and benign activities. In contrast, the ABC_DNN model, while exhibiting high sensitivity at 96.53%, had a lower specificity of 91.22%, suggesting that there is still potential for improvement in differentiating benign instances more effectively. These performance metrics highlight the strengths of both DABCO_CM and ABC_DNN, where DABCO_CM achieved the highest overall performance across all metrics. The differences in dataset characteristics (UNSW-NB15 used in this study versus datasets like KDD Cup 99 and CICIDS2017 in previous studies) and simulation environments (Google Colab in this study versus MATLAB and CloudSim) contribute to variations in model performance. Despite these differences, the results clearly indicate that DABCO_CM provides superior capabilities in accurately detecting and classifying attacks, addressing critical challenges in virtual machine security. The results obtained from the confusion matrices of the proposed models, categorizing actual and predicted attacks for performance metric evaluation, are presented in Table II and performance metrics are presented in Table III.

TABLE II. Confusion Metrices Results

	TP	TN	FP	FN
DNN	112694	50183	5817	6647
ABC_DNN	117380	48120	4626	4215
DABCO_CM_DNN	119250	52562	3438	91

Fig. 6 shows DNN Model Confusion Matrix, Figure 7 shows Line Graph for DNN model, Figure 8 shows ABC_DNN model confusion matrix, Fig. 9 displays Line Graph for ABC_DNN model, Fig. 10 shows DABCO_CM_DNN model confusion matrix and Fig. 11 shows Line Graph for DABCO_CM_ DNN model. The confusion matrices of three important models and their corresponding line graphs are displayed below.

The new models were found to be impressive and superior to the existing system. The definitions of metrics and benchmarked results were displayed below.

- Accuracy: Accuracy measured the overall correctness of a classification model.

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$$

- Sensitivity (Same as Detection Rate/Recall): Sensitivity measured the proportion of actual positive instances correctly classified.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP+}FN}$$

TABLE III. Performance Metrics

Model	Accuracy	Sensitivity	Specificity	Error Rate
Agarwa et al. (2022)				
FS-WOA-DNN	95.35%	96.9%	90.71%	0.0928
Prathyusha et al. (2021)				
AIS	96.56%	96.4%	91.9%	0.0713
Proposed Models				
DNN	92.89%	94.43%	89.61%	0.0711
ABC_DNN	96.38%	96.53%	91,22%	0.0562
DABCO_CM	97.98%	99.92%	93.86%	0.0202

- Specificity: Specificity measured the proportion of actual negative instances correctly classified.

$$\text{Specificity} = \frac{TN}{TN+FP}$$

- Error: Error is the difference between an observed or predicted value and the true or expected value.

Error (residual) = Observed Value - Predicted Value.

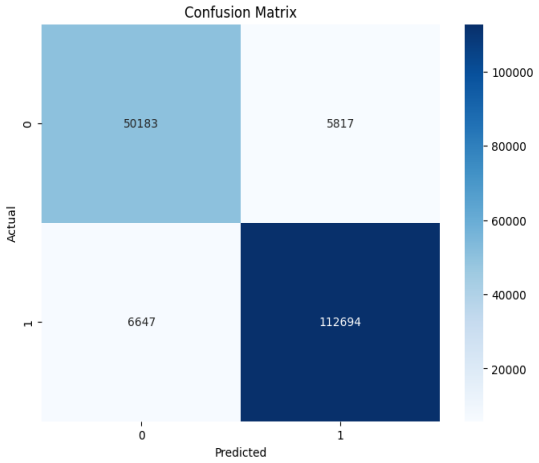


Fig. 6. DNN model confusion matrix.

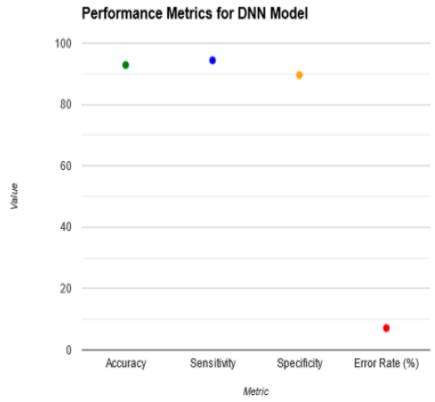


Fig. 7. Line graph for DNN model.

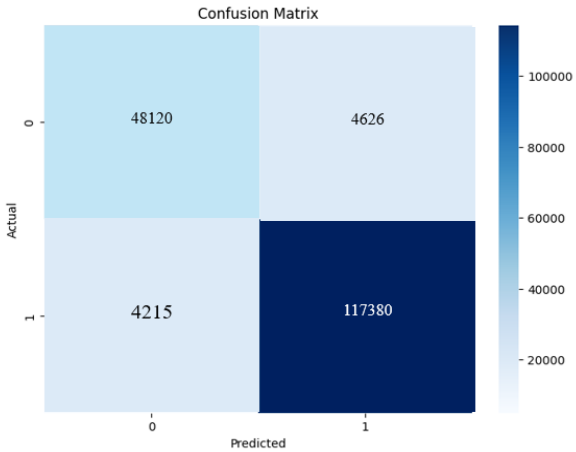


Fig. 8. ABC_DNN model confusion matrix.

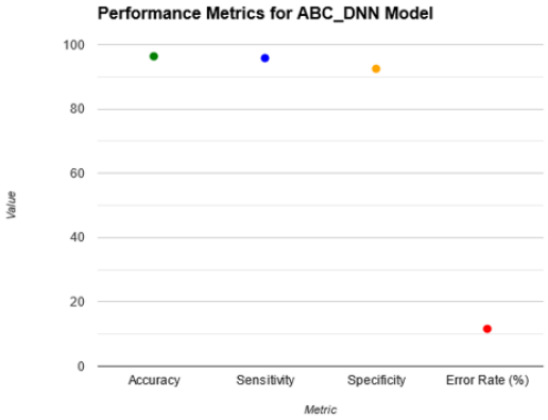


Fig. 9. Line Graph for ABC_DNN model.

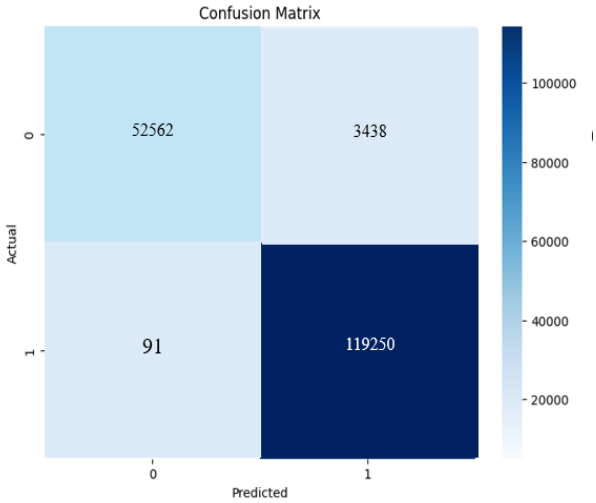


Fig. 10. DABCO_CM_DNN model confusion matrix.

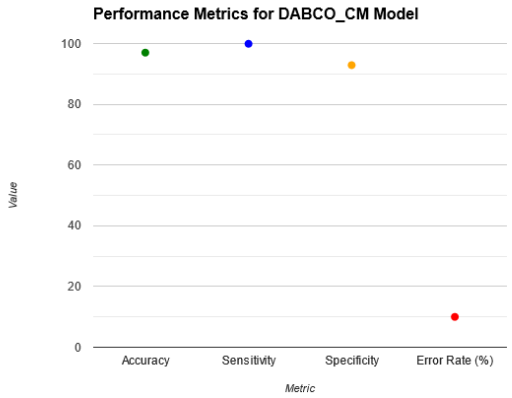


Fig. 11. Line graph for DABCO_CM_DNN model.

VIII. Conclusion and Future Work

This study introduces a pioneering Intrusion Detection System (IDS) to tackle growing security challenges in virtualized computing. The proposed system integrates the Artificial Bee Colony (ABC) optimization algorithm with a

fuzzy logic-based Cloud Model and a Deep Neural Networks (DNN) classifier, resulting in a highly effective approach for categorizing various cyber threats. When compared with the existing Feature Selection Whales Optimization Algorithm with Deep Neural Network Model (FS-WOA_DNN), the newly developed Artificial Bees Optimization-based Cloud Algorithm with Deep Neural Networks Model (DABCO_CM_DNN) shows a 3.02% increase in sensitivity, a 2.63% increase in accuracy, a 3.15% increase in specificity, and a slight decrease in error rate. This indicates that the DABCO_CM model significantly enhances real-time intrusion detection by achieving high accuracy, lowering false positives, and ensuring efficient processing. Future research will aim to refine the model, explore additional bio-inspired techniques, and expand its application across different virtualized environments. Real-world deployment and integration with existing security systems will also be pursued to validate the model's effectiveness and practicality, ensuring its continued relevance in combating evolving cyber threats.

ACKNOWLEDGMENT

We extend our heartfelt gratitude to the Africa Centre of Excellence (ACE), OAU ICT-Driven Knowledge Park (OAK Park), Obafemi Awolowo University, Ile-Ife, Nigeria, for their invaluable support and sponsorship in facilitating this research. We are particularly grateful to Professor G.A. Aderounmu, the Centre Director, whose visionary leadership and commitment to fostering innovation and academic excellence have significantly contributed to the success of this work. This publication is a testament to the remarkable infrastructure and collaborative environment provided by ACE OAK Park, enabling groundbreaking research and capacity building in ICT-driven knowledge.

REFERENCES

[1] Agarwal, A., Khari, M., & Singh, R. (2022). Detection of DDOS Attack using Deep Learning Model in Cloud Storage Application. *Wireless Personal Communications, 127*(1), 419–439. https://doi.org/10.1007/s11277-021-08271-z.

[2] Ahmed, N., Sadhayo, I. H., Yousif, Z., Naeem, N., & Parveen, S. (2019). Analysis and Detection of DDoS Attacks Targetting Virtualized Servers. *International Journal of Computer Science and Network Security (Vol. 19),* 6. Retrieved from https://www.researchgate.net/publication/334325978.

[3] Aishwarya, R., & Malliga, S. (2014). Intrusion detection system- An efficient way to thwart against Dos/DDos attack in the cloud environment. *2014 International Conference on Recent Trends in Information Technology, ICRTIT 2014,* 6. https://doi.org/10.1109/ICRTIT.2014.6996163.

[4] Aldwairi, M., Khamayseh, Y., & Al-Masri, M. (2015). Application of artificial bee colony for intrusion detection systems. *Security and Communication Networks, 8*(16), 2730–2740. https://doi.org/10.1002/sec.588.

[5] Gu, T., Chen, H., Chang, L., & Li, L. (2019). Intrusion detection system based on improved abc algorithm with tabu search. *IEEJ Transactions on Electrical and Electronic Engineering, 14*(11), 1652–1660. https://doi.org/10.1002/tee.22987.

[6] Jin, Y., Sun, Y., & Ma, H. (2018). A developed Artificial Bee Colony algorithm based on cloud model. *Mathematics, 6*(4), 61. https://doi.org/10.3390/math6040061.

[7] Khalaf, B. A., Mostafa, S. A., Mustapha, A., Ismaila, A., Mahmoud, M. A., Jubaira, M. A., & Hassan, M. H. (2019). A simulation study of syn flood attack in cloud computing environment. *AUS Journal, 26*(1), 188–197.

[8] Kumar, J. (2019). Cloud computing security issues and its challenges: A comprehensive research. *International Journal of Recent Technology and Engineering (Vol. 8),* 10-14.

[9] Kuo, R. J., Huang, S. B. L., Zulvia, F. E., & Liao, T. W. (2018). Artificial bee colony-based support vector machines with feature selection and parameter optimization for rule extraction. *Knowledge and Information Systems, 55*(1), 253–274. https://doi.org/10.1007/s10115-017-1083-8.

[10] Mthunzi, S. N., & Benkhelifa, E. (2017). Trends Towards Bio-inspired Security Countermeasures for Cloud Environments. Paper presented at the *IEEE International Conference on Self-Adaptive and Self-Organizing Systems, 11th Proceedings, 18-22 September 2017, Tucson, Arizona, USA. University of Arizona, IEEE Computer Society, Institute of Electrical and Electronics Engineers, National Science Foundation (U.S.).*

[11] Navaz, A. S., Sangeetha, V., & Prabhadevi, C. (2013). Entropy based Anomaly Detection System to Prevent DDoS Attacks in Cloud. *International Journal of Computer Applications (Vol. 62),* 9. https://doi.org/10.5120/10160-5084.

[12] Otor, S. U., Akinyemi, B. O., Aladesanmi, T. A., Aderounmu, G. A., & Kamagaté, B. H. (2021). An improved bio-inspired based intrusion detection model for a cyberspace. *Cogent Engineering, 8*(1), 24. https://doi.org/10.1080/23311916.2020.1859667.

[13] Prathyusha, D. J., & Kannayaram, G. (2021). A cognitive mechanism for mitigating DDoS attacks using the artificial immune system in a cloud environment. *Evolutionary Intelligence, 14*(2), 607–618. https://doi.org/10.1007/s12065-019-00340-4.

[14] Sharma, S., Gupta, A., & Agrawal, S. (2016). An intrusion detection system for detecting denial-of-service attack in cloud using artificial bee colony. In *Advances in Intelligent Systems and Computing (Vol. 438, pp. 137–145).* Springer Verlag. https://doi.org/10.1007/978-981-10-0767-5_16.

[15] Tayab, A., Talib, W., & Fuzail, M. (2015). Security Challenges for Virtualization in Cloud (Vol. 20), 6. Retrieved from https://www.researchgate.net/publication/326261342.

[16] Wang, B., Zheng, Y., Lou, W., & Hou, Y. T. (2014). DDoS attack protection in the era of cloud computing and software-defined networking. In *Proceedings - International Conference on Network Protocols, ICNP (pp. 624–629).* IEEE Computer Society. https://doi.org/10.1109/ICNP.2014.99.

[17] Young, C. (2016). Data Centers: A Concentration of Information Security Risk. *The Juilliard School · Department of Information Technology,* (pp. 339-357). https://doi.org/10.1016/B978-0-12-809643-7.00015-2.

[18] Yu, X., Han, D., Du, Z., Tian, Q., & Yin, G. (2019). Design of DDoS attack detection system based on intelligent bee colony algorithm. *International Journal of Computational Science and Engineering, 19*(2), 22.

YOLO-Driven Lightweight Mobile Real-Time Pest Detection and Web-Based Monitoring for Sustainable Agriculture

Wong Min On[1], Nirase Fathima Abubacker[2]
BSc in Data Analytics, Asia Pacific University of Technology & Innovation, Kuala Lumpur, Malaysia[1]
School of Computing, Asia Pacific University, Kuala Lumpur, Malaysia[2]

***Abstract*—Nowadays, pest infestations cause significant reductions in agricultural productivity all over the world. To control pests, farmers often apply excessive volumes of pesticides due to the difficulty of manually detecting the pest at an early stage. Their overuse of pesticides has led to environmental pollution and health risks. To address these challenges, many novel systems have been developed to identify pests early, allowing farmers to be alerted about the exact location where pests are detected. However, these systems are constrained by their lack of real-time detection capabilities, limited mobile integration, ability to detect only a small number of pest classes, and the absence of a web-based monitoring system. This paper introduces a pest detection system that leverages the lightweight YOLO deep learning framework and is integrated with a web-based monitoring platform. The YOLO object detection architectures, including YOLOv8n, YOLOv9t, and YOLOv10-N, were studied and optimized for pest detection on smartphones. The models were trained and validated using merging publicly datasets containing 29 pest classes. Among them, the YOLOv9t achieves top performance with a mAP@0.5 value of 89.8%, precision of 87.4%, recall of 84.4%, and an inference time of 250.6ms. The web-based monitoring system enables dynamic real-time monitoring by providing farmers with instant updates and actionable insights for effective and sustainable pest management. From there, farmers can take necessary actions immediately to mitigate pest damage, reduce pesticide overuse, and promote sustainable agricultural practices.**

Keywords*—Pest detection; YOLO; deep learning; real-time monitoring; smartphone application; web-based platform; object detection; pest management; pesticide reduction; sustainable agriculture

I. INTRODUCTION

Agriculture is a key activity that provides humans with basic needs including food, medicine, shelter, and clothing. Recent research indicates that 35-40% of the world's land area is used for agricultural purposes [1]. With the global population expected to grow to approximately 9.7 billion by 2050, the demand for agricultural output has never been higher [2]. Unfortunately, various factors continue to influence agricultural productivity, leading to reduced yields and placing farms at serious risk. According to the Food and Agricultural Organization (FAO), insects, pests, illnesses, and weed infestations are estimated to damage around 40% of agricultural production per year [3]. The main reason behind this issue is that the control levels of pests are not always attained due to the absence of an accurate diagnosis at the right moment.

Today, many farmers still depend on traditional methods that involve manual field inspections by themselves to determine signs of infestation. In simple words, they rely on their own knowledge and experience when faced with a pest infestation. Due to the wide variety of crops and pests, manual detection is challenging and error prone [4]. Aside from that, the frequent emergence and recurrence of pests has further hampered the effectiveness of traditional early detection measures [5]. Also, insufficient knowledge leads the farmers to use a variety of pesticides as their primary method for eliminating pests in order to protect crops and increase the quality of their yields [6]. Excessive use or misuse of pesticides can harm the ecosystem and potentially cause long-term diseases like cancer, respiratory infections and fetal deaths [7]. To reduce the widespread reliance on harmful pesticides, modern technology plays a key role in detecting pests at an early stage in agriculture.

Over the past few years, deep learning has changed the field of machine learning with the potential in revolutionizing numerous applications, particularly in object detection [8]. This has opened the possibility for innovative solutions that could tackle various agricultural challenges. Nevertheless, many existing systems designed for pest detection in agriculture have significant limitations. They often lack real-time processing capabilities, have minimal mobile integration, and can detect only a limited number of pest classes. Hence, most current systems do not perform well under diversified agricultural settings. Moreover, although some of these systems have already been integrated into smartphones, they lack a web-based monitoring system with real-time updating and tracking features necessary for effective pest management.

To overcome these challenges, one of the most prominent deep learning algorithms, YOLO (You Only Look Once), is remarkable with its ability to perform quick detection with high accuracy [9]. It is well-suited for applications that need rapid and accurate object detection, which includes identifying pests in agriculture. The YOLO family of algorithms has been very successful in many studies including those in agriculture. Among these, YOLOv8 has proven to be especially effective with many studies demonstrating its capability to balance detection accuracy with computational efficiency, making it a

widely accepted choice [10]. However, YOLOv9 and YOLOv10 which have been recently released in the market have not been applied yet to the field of pest detection. Therefore, this study involves training and evaluating these versions including YOLOv8, YOLOv9, and YOLOv10, to determine which model is most effective for pest detection in agricultural settings.

The proposed implementation of a pest detection system based on YOLO is a step into a new era in agriculture. This paper aims to develop a pest detection system using an optimal lightweight YOLO framework integrated with a smartphone for real-time detection of multiple pest classes and coupled with a web-based monitoring system. Farmers can easily monitor their crops by positioning their smartphones in various areas of their fields, with data being sent to a server for analysis. The web-based system then allows farmers to observe real-time results and receive instant recommendations for effective pest mitigation.

The rest of the paper is arranged as follows. Section II reviews related works in pest detection and classification. Section III covers the materials and methods used, including the methodology, pest image datasets, preprocessing techniques, YOLOv8, YOLOv9, and YOLOv10 model implementations, and the relevant evaluation metrics. Section IV describes the experimental setup and presents the results and discussion. The deployment of the optimal model in real-time applications for pest detection is discussed in Section V. Section VI outlines the limitations encountered during the study. Section VII concludes the paper and suggests enhancements for future work.

II. SIMILAR WORK

Recent advancements in deep learning and mobile technologies have significantly influenced the development of real-time image-based pest detection systems in agriculture [11]. Many studies have explored different CNN architectures for mobile devices. However, these systems often encounter limitations including restricted pest class detection, hardware constraints, and challenges in achieving real-time performance [12] [13]. Various researchers have investigated object detection techniques for identifying insect pests, each contributing unique approaches and highlighted persistent challenges.

For instance, Fuentes et al. integrated SSD, R-CNN, and Faster R-CNN deep learning models with a VGG network and residual networks to recognize nine different types of tomato plant pests and diseases [14]. Although this approach achieved a mean Average Precision (mAP@0.5) value of 83.06%, it was confined to a limited number of pest classes, which limited its application to more diversified agricultural settings. Lin et al. employed Fast R-CNN to develop an anchor-free regional convolutional network using an end-to-end model approach [15]. The model is able to categorize 24 pest classes and achieved a mAP@0.5 value of 56.4% and a recall of 85.1%. These results surpassed the performance of traditional Fast R-CNN in controlled environments. However, the method's applicability in real-time, dynamic agricultural environments remains uncertain. Another study conducted by Sabanci et al. constructed a convolutional recurrent hybrid network to identify wheat grain that has been affected by pests [16]. They combined AlexNet with bidirectional short-term memory (BiLSTM). The model achieved a remarkable cumulative accuracy of 99.50%. While this demonstrates high precision for specific tasks, the system's scope was limited to wheat grain since it can only distinguish between two types such as healthy wheat grains and those damaged by sunn pests (SPD). Also, it did not address broader pest detection needs across various crops.

In another effort, Koklu et al. devised a deep feature extraction method based on CNN-SVM [17]. The researchers classified five distinct species of grapevine leaves and achieved an accuracy of 97.60%. Although effective for leaf classification, this method has yet to be tested in the more complex domain of insect pest detection. Another notable example is that Li et al. created a real-time system for identifying pests and plant disease through the implementation of Faster R-CNN [18]. Their approach effectively detected unseen rice diseases in video footage but focused primarily on disease rather than insect pest detection.

Additionally, a visual flying insect detection system based on the YOLO architecture was introduced by Zhong et al. using a Raspberry Pi [19]. A cumulative accuracy of 92.50% and a classification accuracy of 90.18% are both achieved by the system. Although this system showed promise in identifying flying insects, the limited processing power of Raspberry Pi constrained its application in more computationally intensive tasks. In addition, Arunabha M. Roy and Jayabrata Bhaduri introduced an enhanced version of YOLOv4, called Dense-YOLOv4, by incorporating DenseNet into its backbone to improve the feature transfer and reuse [20]. This model achieved an impressive mAP@0.5 value of 96.20% in identifying various phases of mango growth within a complicated orchard setting. However, its heavy computational demands pose challenges for deployment on mobile devices. Although it achieved an impressive recognition rate of 99.3% within an average processing time of 44 milliseconds, but these models were only specialized for a single type of crop, which limits their broader applicability.

In recent years, the YOLO algorithm family has evolved significantly to enhance real time object detection for lightweight and mobile friendly applications. For example, a YOLOv5-S model was developed by Thanh-Nghi Doan for real-time insect detection and was integrated into resource-constrained mobile devices [21]. The model performance was reported up to 70.5% classification accuracy on the Insect10 dataset and 42.9% on the IP102 larger dataset. These results prove that the model performs reasonably well on smaller dataset but struggles to achieve the accuracy required for effective agricultural pest detection as the size and complexity of the pest dataset increases.

Moreover, an updated version, YOLOv8, is widely adopted due to its faster and more accurate performance in real-time object detection tasks. Additionally, its architecture allows for easy refinement and customization to adapt to specific tasks. For example, Yin Jian Jun enhanced the YOLOv8 model by refining its feature extraction algorithm and reducing the number of parameters count to a achieve lightweight model design [22]. Through the refined training techniques, the model

achieved a remarkable mAP@0.5 value of 97.3%. for detecting eight different species of rice pest. Although YOLOv8 has now been widely applied in many studies, it was usually applied to detect only a few pest species, hence making its application disadvantageous in more complicated agricultural settings.

Recently, the latest works in the YOLO series, such as YOLOv9 and YOLOv10, have come up to introduce more features that bring greater performance at lesser computational overhead. YOLOv9 is upgraded with enhancements like Generalized Efficient Layer Aggregation Network (GELAN) and Programmable Gradient Information (PGI), which are at the root of improved detection performance [23]. On the other hand, YOLOv10 has been developed by the Tsinghua University researchers with the aid of the Ultralytics Python package, introducing an innovative technique to real-time detection by solving post-processing issues and model architecture shortcomings that were present in previous YOLO versions [24]. By omitting non-maximum suppression (NMS) and refining several aspects of the model architecture, YOLOv10 achieves cutting edge results at a substantially lower computational cost. However, despite these advancements, the practical application of both YOLOv9 and YOLOv10 in detecting agricultural pests is yet to be applied.

Although these studies have made remarkable progress, numerous challenges are still unsolved. First, the majority of these current models can only detect a few numbers of pest or object classes. This limits the application of these methods in different agricultural scenarios. Also, since their datasets are small with fewer classes, their data preprocessing techniques such as data augmentation and clean processes are often simpler and overlooked. This deficiency can hinder accurate detection as the size and variety of classes increase. Third, the real-time deployment of these models in a mobile device is generally hindered by hardware limitations. In many cases, the real-time detection capabilities are frequently restricted by the computational demands of deep learning models [25]. Therefore, there is a great need for a lightweight model in terms of overcoming these weaknesses. Although some have been developed and integrated with smartphones, such systems are not capable of providing real-time updates and comprehensive pest detection across multiple classes. Additionally, they also lack a web-based platform responsible for pest data monitoring and analysis.

To address the limitations identified in existing studies, this paper aims to develop a pest detection system using a lightweight and optimized YOLO deep learning framework, integrated into mobile devices and complemented with a web-based monitoring system.

The main contributions of this paper are as outlined below:

- Employing advanced data augmentation techniques, including hue adjustment, horizontal flipping, and scaling, to significantly increase dataset diversity and enhance the model's ability to classify pests in complex agricultural conditions.
- Introducing a lightweight YOLO-based deep learning model that is optimized to detect a broader range of pest classes compared to previous studies, balancing accuracy and efficiency for real-time smartphone applications in diverse agricultural settings.
- Integrating the optimized model with smartphone technology for real time detection of multiple pest classes, making advanced pest detection tools more accessible and user-friendly.
- Creating an interactive web-based monitoring platform that offers dynamic real-time updates of pest detection and provides sustainable recommendations for effective pest management strategies.

III. Materials and Methods

A. Methodology

The proposed methodology follows a structured approach to develop an effective pest detection model for agricultural applications as shown in Fig. 1. At the beginning, the researchers collected multiple pest image datasets and merged them to form a unified dataset through several preprocessing processes such as data standardization. Data standardization is a technique to ensure that datasets are in a consistent and standardized format for configuring the YOLO format. Secondly, the researchers preprocessed the dataset by employing data augmentation techniques including hue adjustments, image translation, horizontal flipping and scaling to expand the training dataset for enhancing its variability. This step aims to prepare a dataset that can be used to train a robust detection model capable of generalizing across diverse agricultural scenarios. Following this, the images were annotated with bounding boxes to label pest instances accurately. After that, the researchers split the dataset by allocating 80% of the dataset for the model training and 20% for the model validation. Researchers then proceeded to train various YOLO object detection models with the pest training dataset. The trained models were fine-tuned to optimize their performance across different YOLO versions. Then, researchers validated and evaluated the performance of each base and fine-tuned model by using the shared validation dataset. During validation, new pest images that were not previously used for training or fine-tuning were introduced to assess the models' effectiveness with unseen data. The researchers also integrated each model into the smartphone application for measuring their real-time performance to assess its capabilities in a practical setting. By comparing the results, the optimal model was identified for practical field adaptation. This model was then deployed within a smartphone application to enable real-time pest detection in agricultural environments. Finally, a web-based monitoring system was developed to provide dynamic and real-time updates on pest detection. This allows users like farmers to monitor captured pest data and receive actionable insights for effective pest management.

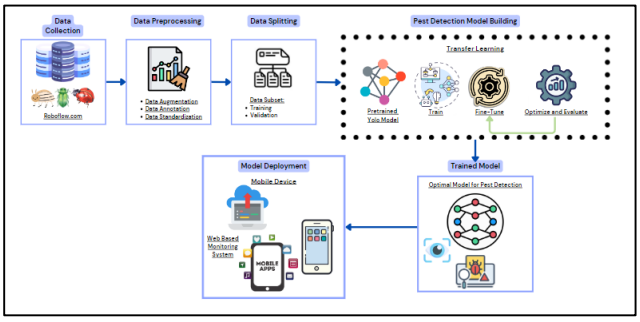


Fig. 1. Proposed methodology to develop the pest detection system.

B. Data Collection

To train and evaluate the pest detection model, the researchers utilized two publicly available datasets from Roboflow platform as a source in the experimental study: YOLOIP1 dataset and Pest_Dataset_3 [26] [27]. Since both datasets are different in terms of the structures and class distribution, the datasets were subjected to a data standardization process to ensure uniformity. Both datasets were reorganized to have consistent directory structures, file naming conventions, and annotation formats. After standardization, the researchers removed the redundant classes to avoid confusion due to the similar morphological features. For example, classes like "white margined moth" and "margined moth" which differed only slightly in color were removed. The researchers then merged the datasets manually by selecting the most relevant pest classes that are frequently encountered in agricultural scenarios for accurate detection. The final merged dataset comprised 29 pest classes. Each class contains a significant number of images as detailed in Table I. However, the merged dataset contained fewer images than the expected threshold of a total of 10k images. Therefore, data augmentation techniques were employed in the next step to expand the dataset's size.

TABLE I. FINAL CLASS DISTRIBUTION OF THE MERGED DATASET

Index	Pest Class	Total Count
0	Alfafa plant bug	99
1	Ampelophaga	126
2	Aphids	886
3	Beet spot flies	79
4	Flea beetle	414
5	Grain spreader thrips	118
6	Grub worm	522
7	Icerya-purchasi-Maskell	73
8	Limacodidae	95
9	Lycorma delicatula	111
10	Lygus	318
11	Mole cricket	888
12	Oides decempunctata	77
13	Paddy stem maggot	94
14	Peach moth	223
15	Pieries canidia	147
16	Plant hopper	856
17	Protaetia	180
18	Red spider	245
19	Rice gall midge	244
20	Rice leaf caterpillar	235
21	Rice leaf roller	233
22	Rice leafhopper	242
23	Rice shell pest	115
24	Rice stemfly	108
25	Weevil	650
26	Wireworm	464
27	Xylotrechus	68
28	Yellow rice borer	591
Total number of images		8,501

C. Data Augmentation

To improve the generalization capabilities of models, several data augmentation techniques were applied to the dataset. Firstly, the researchers adjusted the hue, saturation and value (HSV) of images to randomly alter the color intensity and brightness in a manner to mimic different environmental conditions. Secondly, the researchers shifted the images along the x and y axes through image translation. This step is to provide different angles of the same image. They also applied horizontal flipping to mirror the images for creating variations in how objects are oriented. Finally, they resized images while keeping their aspect ratios through image scaling process to ensure the model remains robust and can handle differences in object size effectively. Fig. 2 illustrates examples of the augmented images.

After completing the data augmentation process, the final dataset was prepared with sufficient images and met the expected requirements to have at least 10,000 images for training the robust model. The details of the dataset after augmentation are presented in Table II.

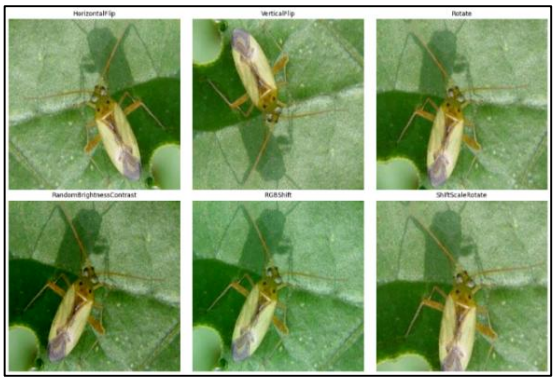


Fig. 2. Samples of augmented pest images.

TABLE II. DETAILS OF DATASET

Dataset Aspect	Number
Number of classes	29
Number of images	14,147
Number of instances (objects labeled within the images)	17,856

D. Data Annotation

After data augmentation, the researchers carefully reviewed the images to ensure that each one was annotated and assigned with a pest class ID correctly. A deep learning model such as YOLO model learn features from the labeled images [28]. Therefore, the correctness of feature labeling will greatly influence the performance of training model especially considering the similarities between many pest species. If any discrepancies were found such as incorrect class labels or annotations, the researchers made corrections to those particular images using the Roboflow platform. The annotation process involved normalizing the coordinates between 0 and 1 to accommodate varying image sizes. After this process, the annotations across the entire dataset were consistent and accurate. Each annotation was recorded in a text file containing the following details for each bounding box as shown in Eq. (1). A sample of a pest image with the correct annotation bounding box is shown in Fig. 3.

$$(id_{class}, x_{centre}, y_{centre}, width, height) \quad (1)$$

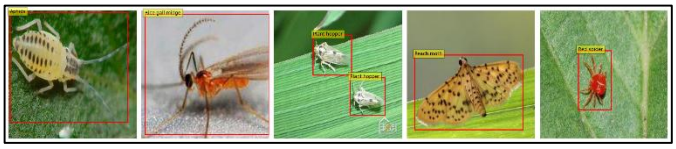


Fig. 3. Visualization of labeled images.

E. Data Splitting

Before training the YOLO model, the researchers split the dataset into training and validation sets with an 8:2 ratio. This step is to ensure that the model is trained on a diverse and representative set of images while also being evaluated on a separate validation set. The summary of final dataset aspects is shown in Table III. It is well prepared for the next stages of model training and evaluation.

TABLE III. SUMMARY OF FINAL DATASET ASPECTS

Dataset Aspect	Details
Number of classes	29
Image Size	640x640 pixels
Training Images and Label Files	10,097
Validation Images and Label Files	4,050
Total Instances in Training	12,658
Total Instances in Validation	5,198

F. Overview of YOLO Model

The YOLO algorithm, which stands for "You Only Look Once" was first introduced by Joesph Redmon et al. in 2015 [29]. It is renowned for its efficiency in object detection since it can identify objects and their positions by examining the entire image only once. As seen in Fig. 4 below, the YOLO algorithm divides the image into N grids, each of which contains an equal-sized S × S region. Each grid is responsible for detecting and locating the object within it.

Unlike the conventional methods that employ two-stage object detectors, YOLO treats object detection as a regression problem [30]. It predicts bounding boxes around the objects and their corresponding class probabilities straight from the feature maps in a single pass. This means that YOLO can recognize objects in an image faster than the two-stage detector. Therefore, the YOLO algorithm is considered as a one-stage detector that prioritizes speed and is thus ideal for real-time object detection.

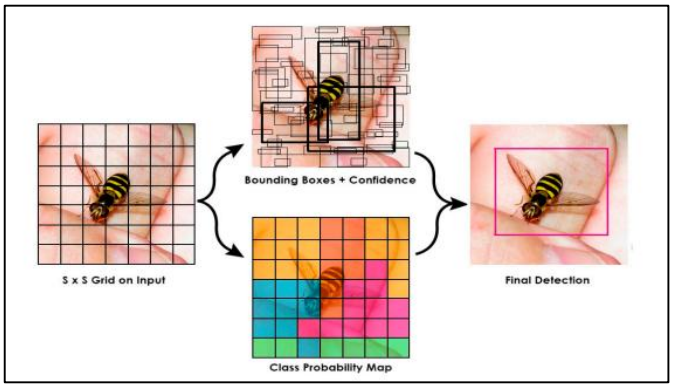


Fig. 4. Illustration of the YOLO framework for object detection [31].

1) YOLOv8: YOLOv8 is the most popular and widely used model in the whole You Only Look Once (YOLO) series. It signifies a notable development in the domain of object detection with its remarkable enhancements in both speed and accuracy through extensive architectural optimization and innovation. The architecture of YOLOv8 is depicted in Fig. 5. The YOLOv8 model considers the multi-scale attributes of objects by using three detection layers at different scales to handle objects that come in different sizes [32]. This method allows the model to efficiently manage objects of varying sizes and proportions.

In its architectural design, there are three main components including Backbone, Neck and Head [33]. The Backbone is used for extracting multi-scale features to ensure the model can perceive inputs across different scales. It includes modules such as Conv, C2f, and SPPF (Spatial Pyramid Pooling-Fast). In the Neck section, YOLOv8 combines features without enforcing standardized channel dimensions by integrating a path aggregation network and a feature pyramid network [34] [35]. This method reduces both the number of parameters and the overall tensor size. To simplify anchor box operations and prevent displacement issues, YOLOv8 adopts a decoupled head method to separate the detection and classification head [36]. The Head component is responsible for tasks such as bounding box regression, target classification and confidence assessment in the prediction layers. It ultimately delivers precise detection results through the use of non-maximum suppression.

To support diverse computational requirements, YOLOv8 is available in five versions as detailed in Table IV [37]. Each model comes with different parameter counts and resource consumption due to differences in width and depth parameters. The higher the scale of the model is, the higher the detection performance is as the parameter count and resource consumption rise. The MS COCO dataset is frequently used for benchmarking object detection models. By comparing the performance between models on the MS COCO dataset, the researchers selected YOLOv8n as their baseline model for this study because of its relatively low parameter count and resource efficiency. FLOPs (Floating Point Operations per Second) are listed in Table IV to indicate the computational complexity of each model.

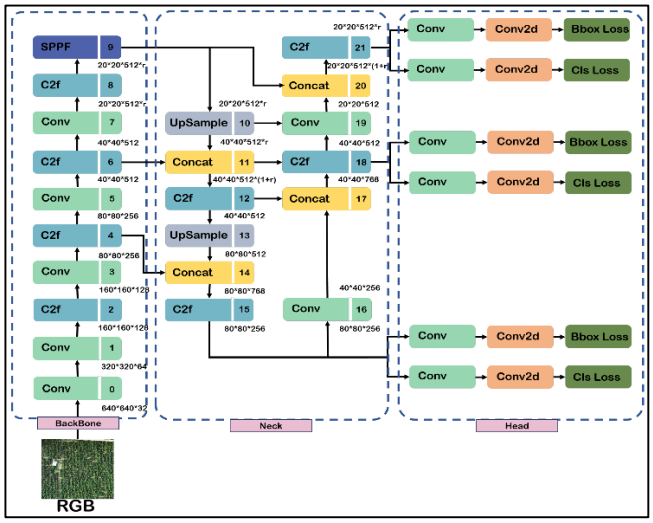


Fig. 5. Model architecture of YOLOv8 [38].

TABLE IV. YOLOv8 VARIANT COMPARISON

Model	Params (M)	FLOPs (B)
YOLOv8n	3.2	8.7
YOLOv8s	11.2	28.6
YOLOv8m	25.9	78.9
YOLOv8l	43.7	165.2
YOLOv8x	68.2	257.8

2) YOLOv9: YOLOv9 is the most recent addition to the YOLO versions in 2024. It introduces two major innovations such as the Programmable Gradient Information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN). The PGI framework tackles the challenges of information bottlenecks in deep neural networks and makes supervision mechanisms compatible with lightweight architecture [39]. By leveraging PGI, substantial accuracy improvements can be achieved in both complex and lightweight architectures. This is due to the fact that PGI mandates reliable gradient information during training, which in turn enhances the architecture's ability to learn and make predictions.

The GELAN architecture was designed with a purpose of improving the performance of objection detection tasks by lightweight and high efficiency footprint [40]. Because of its strong performance across different depth configurations and computational blocks, it is well-suited for deployment on various inference devices including devices that are resource constrained. In simple words, PGI and GELAN solved the issues related to computational efficiency and information loss. By integrating both GELAN and PGI frameworks, YOLOv9 represents a substantial advancement in lightweight object detection.

Table V presents five YOLOv9 variants along with their parameter count and FLOPs as assessed on the MS COCO Dataset [41]. YOLOv9t was chosen as another baseline model for this study due to its lowest parameter count and computational demand.

TABLE V. YOLOv9 VARIANT COMPARISON

Model	Params (M)	FLOPs (B)
YOLOv9t	2.0	7.7
YOLOv9s	7.2	26.7
YOLOv9m	20.1	76.8
YOLOv9c	25.5	102.8
YOLOv9e	58.1	192.5

3) YOLOv10: In 2024, another groundbreaking development emerged with the introduction of YOLOv10. This version further expands the boundaries of real-time object detection by resolving its inherent difficulties. It sets a new benchmark by completely eliminating non-maximum suppression (NMS) during post-processing, thereby enhancing inference speed [42]. The model introduces a novel dual-label assignment system to maintain an optimal balance between accuracy and speed. By integrating one-to-one and one-to-many label assignments, YOLOv10 benefits from rich supervisory signals during its training, ensuring computational efficiency while acquiring important detection features without any post-preprocessing NMS [43].

Furthermore, the enhancements in YOLOv10's architecture involve the implementation of spatial-channel decoupled down sampling, lightweight classification heads, and rank-guided block design [42]. Each of these additions lowers the computational requirements and the number of parameters. These enhancements enhance both the efficiency and scalability of the model over a wide range of devices from powerful servers to edge devices with limited processing power and storage capacity. The performance of both the lightweight and heavyweight versions of YOLOv10 model on the COCO dataset is compared in Table VI [44]. YOLOv10-N with the least parameters and computational resource is selected for this study as the baseline model for YOLOv10.

TABLE VI. YOLOv10 VARIANT COMPARISON

Model	Params (M)	FLOPs (G)
YOLOv10-N	2.3	6.7
YOLOv10-S	7.2	21.6
YOLOv10-M	15.4	59.1
YOLOv10-B	19.1	92.0
YOLOv10-L	24.4	120.3
YOLOv10-X	29.5	160.4

G. Evaluation Metrics

In object recognition methods, the detection result and classifier performance are the two key indices used to evaluate models [45]. Mean Average Precision (mAP) is a widely used metric for evaluating the overall performance of object detection systems [46]. The mAP score is derived by comparing the predicted bounding boxes with the ground truth boxes. Higher mAP score indicates more accurate model detections. This metric is calculated using several sub-metrics: Precision, Recall, Intersection over Union (IoU), and data from the Confusion Matrix.

To evaluate the detection result, Intersection over Union (IoU) metric is employed. IoU measures the ratio of the intersection to the union of predicted and ground truth bounding boxes [47]. It indicates how closely the predicted bounding box matches the ground truth. When the IoU value exceeds a threshold of 0.5, the detection is regarded as a True Positive (TP). Conversely, if the IoU value falls below 0.5, the detection is considered a False Positive (FP). A False Negative (FN) occurs when the model fails to detect an object that exists in the ground truth. A True Negative (TN) refers to the model correctly identifying that no object exists in a region where there is indeed no object. The concept of IoU is illustrated in Fig. 6, where the rectangles R1 and R2 serve as bounding frames for the object's ground truth and prediction.

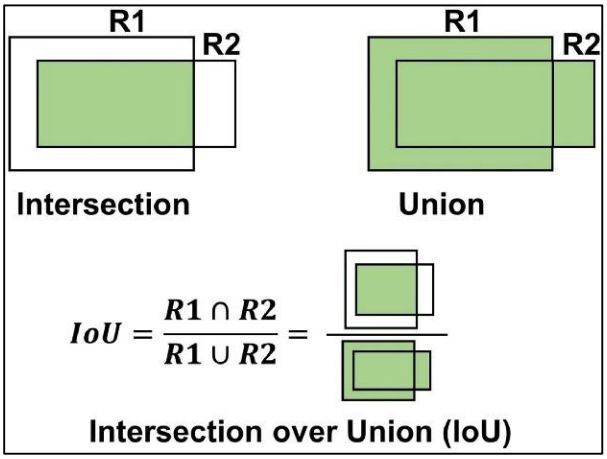


Fig. 6. Concept of Intersection over Union (IoU).

For accessing the performance of the classifier, a Confusion Matrix is used. It categorizes the model's predictions into four attributes [48]:

- True Positives (TP): The model correctly identifies and matches a label with the ground truth data.
- True Negatives (TN): The model correctly identifies that a label is absent when it is indeed not present in the ground truth data.
- False Positives (FP): The model incorrectly predicts a label that is not actually present in the ground truth data.
- False Negatives (FN): The model fails to detect a label that is present in the ground truth data.

Using the IoU value along with the Confusion Matrix outputs (TP, TN, FP, and FN), key performance metrics such as Precision, Recall, and mAP are calculated. All these metrics are used to assess the overall performance of object detection models by calculating them using Formulas (1) to (3).

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \quad (3)$$

AP_k = the average precision of class k, and n = number of classes

IV. RESULT AND DISCUSSIONS

A. Experimental Setup and Training

In this study, the training and evaluation of the selected YOLO models like YOLOv8n, YOLOv9t and YOLOv10-N were conducted on Google Colab using Nvidia L4 GPU with 24 GB of RAM. The dataset which had already been preprocessed and split into 80:20 ration in previous phase was used to ensure consistent and high-quality input for model training. Each YOLO model was trained on the training set which consists of 10,097 pest images while their performances were evaluated on the validation set containing 4,050 images.

To ensure a fair comparison, all YOLO models were trained using the same hyperparameters. The model was created using Pytorch and runs for 150 epochs to train with a batch size of 16, learning rates of 0.001 and default dropout rates provided by the YOLO framework. Adam optimizer which is known as "Adaptive Moment Estimation" was used as the optimization algorithm. After training the model, a fine-tuning process was conducted using a grid search method to optimize each model. Grid search method tuned the model by testing out all combinations of critical parameters [49]. This fine-tuning process ran for 10 additional epochs to refine the models' performance.

During training, the YOLO models will automatically perform in-training evaluations at the end of each epoch using the validation set. The metrics reflect the ongoing learning process and may result in slightly higher performance or overfitting due to continuous weight adjustments [50]. Once training is completed, the reseachers conducted a post-training evaluation to access the model's generalization ability by using the same validation set in a static environment with fixed weights.

The experimental configuration for training and testing these models is detailed in Table VII.

TABLE VII. EXPERIMENTAL CONFIGURATION

Configuration	Params (M)
Platform	Google Colab
CPU	2.0 Intel Xeon (R) CPU @ 2.20GHz × 12
GPU	Nvidia L4 (24 GB RAM)
Accelerated Environment	Nvidia L4 CUDA 12.2
Operating System	Ubuntu 22.04.3 LTS
CUDA Version	12.2
PyTorch GPU Availability	True
PyTorch CUDA Version	12.4

B. Experimental Results

1) Training loss: Training loss is the first focus of this analysis. It encompasses multiple components including classification loss (cls_loss), distribution focal loss (dfl_loss) and bounding box loss (box_loss). Each component of the training loss tracks the degree of error in the model's outputs. It provides insights into how well the models fit the data and aids in determining the best weights [51]. As shown in Fig. 7 to Fig. 9, all training loss curves demonstrate a clear downward trend throughout the training process. This means that the proposed models excel at learning early and detecting pests during the training time. As the networks undergo further epochs, the classification loss declines slowly. The models converged after 140 epochs. This enabled researchers to conclude that 150 epochs was a good parameter for building the model.

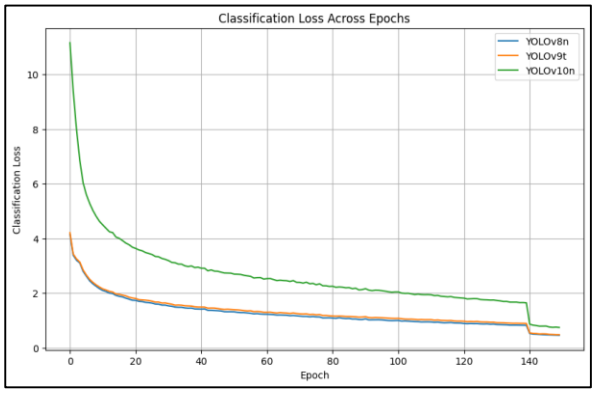


Fig. 7. Classification loss graph.

As can be seen in Fig. 7 to Fig. 9, YOLOv8n and YOLOv9t architecture have achieved similar and lower training losses when compared to the YOLOv10-N. This observation could be attributed to the increased complexity of YOLOv10-N architecture. YOLOv10-N learns more intricate patterns from the data to handle the object detection tasks. However, this statement does not imply that other models are inadequate. Further analysis is required to determine a balanced approach where the model performs well on both training data and real-world data.

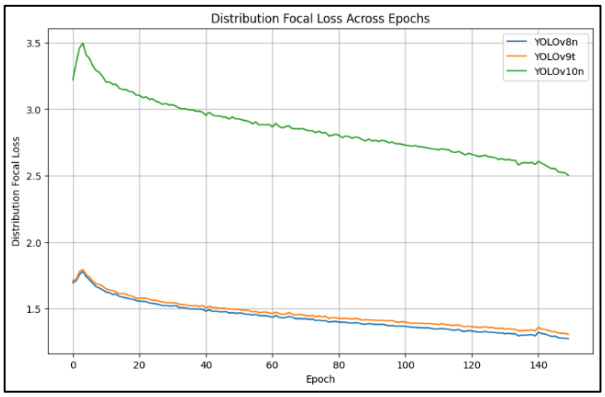


Fig. 8. Distribution focal loss graph.

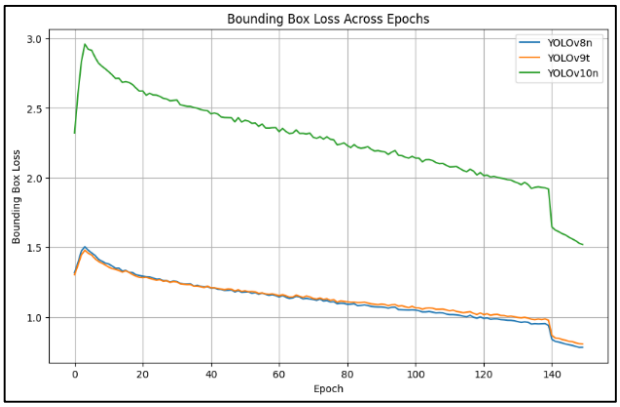


Fig. 9. Bounding box loss graph.

2) Performance evaluation: To further evaluate the models' performance, the researchers compared the baseline and fine-tuned versions of YOLOv8n, YOLOv9t, and YOLOv10-N using the statistical indicators which includes precision, recall, mAP@ 0.5 and mAP@ 0.5:0.95. The results of models are presented in Table VIII and illustrated in Fig. 10. All fine-tuned models demonstrate improvements across all the key metrics. As depicted by the graph below, the YOLOv8n model does not top in any of these metrics but shows improvement over its baseline. In fact, it has actually achieved quite good results when not compared with other versions. Moving to the YOLOv9t model, the fine-tuned version attained the highest precision of 94.2% and the best value for recall of 91.4%. This points to a high effectiveness of the YOLOv9t model in doing accurate pest identifications. Meanwhile, the fined-tuned version of YOLOv10-N model achieved the highest mAP@0.5 at 96.7% and mAP@0.5:0.95 at 77.1%. This means that the YOLOv10-N model performs exceptionally well in detecting pests across various IoU thresholds. It is particularly robust for handling complex detection tasks. From this point onward, all references to the YOLO models will pertain to their fine-tuned versions.

TABLE VIII. RESULTS OF YOLOV8N, YOLOV9T AND YOLOV10-N ON VALIDATION DATASET DURING TRAINING

Models	Precision	Recall	mAP^{val}@0.5	mAP^{val}@0.5:0.95
YOLOv8n Baseline	0.853	0.847	0.881	0.629
YOLOv9t Baseline	0.85	0.823	0.874	0.637
YOLOv10-N Baseline	0.884	0.829	0.884	0.643
YOLOv8n Fine-tuned	0.932	0.891	0.951	0.733
YOLOv9t Fine-tuned	0.942	0.914	0.962	0.751
YOLOv10-N Fine-tuned	0.94	0.93	0.967	0.771

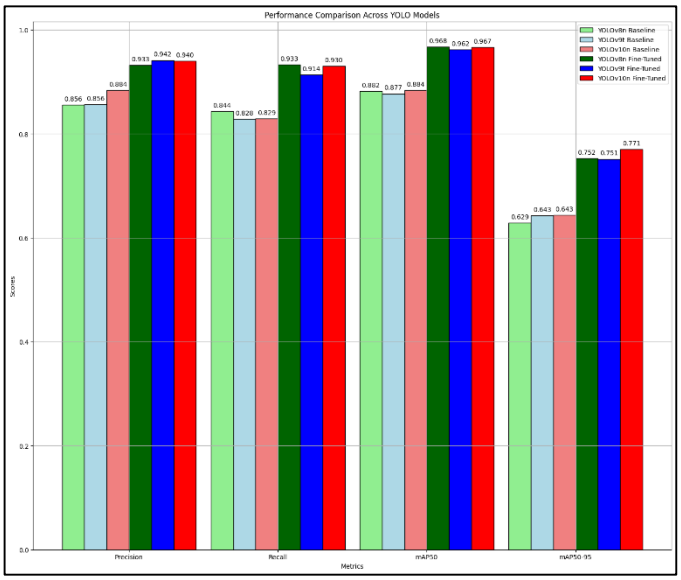


Fig. 10. Comparison of YOLOv8n, YOLOv9t, and YOLOv10-N on validation data during training.

3) *Confusion matrix analysis:* To better demonstrate the results, confusion metrices were employed to evaluate and visualize the performance of the trained YOLO models. A confusion matrix describes the actual and predicted object classification of a classification system [52]. The significance of the prediction results is indicated by the diagonal line in the central of the confusion matrix. The horizontal line shows false negatives, whereas the vertical line shows false positives. The darker the blue in the matrix, the higher the count of correct classifications. Based on the confusion matrices presented in Fig. 11 to Fig. 13, all YOLO models have done quite well with high accuracy values among their predictions.

4) Post-Training evaluation: The previous validation results were all generated automatically by YOLO itself during training. These in-training metrics tend to be higher due to the dynamic adjustments the models make during the learning process such as ongoing weights and gradient updates. Therefore, the researchers reassessed the models using the same validation set containing 4050 images in a post-training static environment where the weights and parameters of each model were fixed. The results are presented in Table IX. It reveals that the YOLOv9t model emerged as the top performer with the highest scores in Precision at 87.4%, Recall at 84.4%, mAP@0.5 at 89.8%, and mAP@0.5:0.95 at 66.7%. It was followed closely by the YOLOv10-N and YOLOv8n model. These findings indicate that although models perform excellent during training and validation, but their performance changes when they are exposed to new data. The causes for this change in performance might be due to slight overfitting during training or the complexities of the real-world image environments. Sample predictions made by each model on test data are presented in Fig. 14, Fig. 15, and Fig. 16.

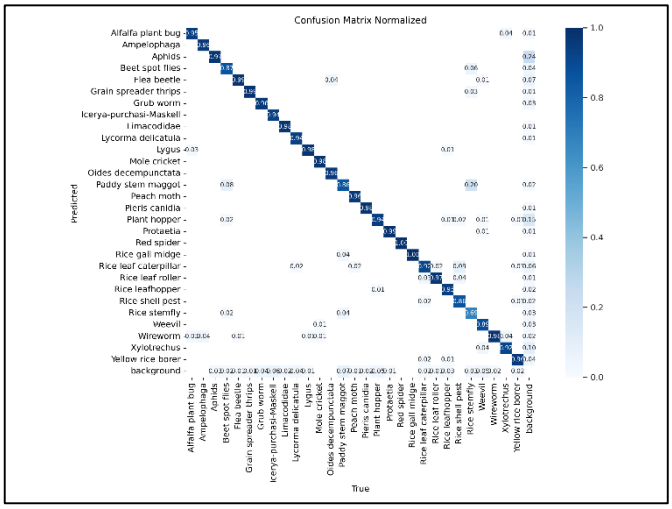


Fig. 11. Confusion matrix of YOLOv8n.

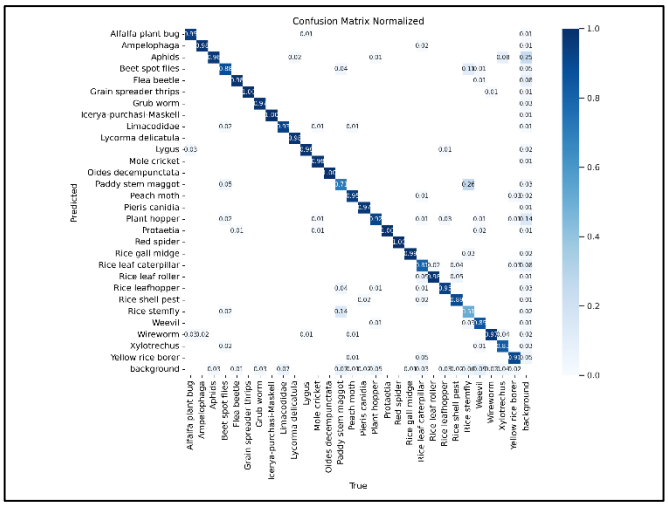


Fig. 12. Confusion matrix of YOLOv9t.

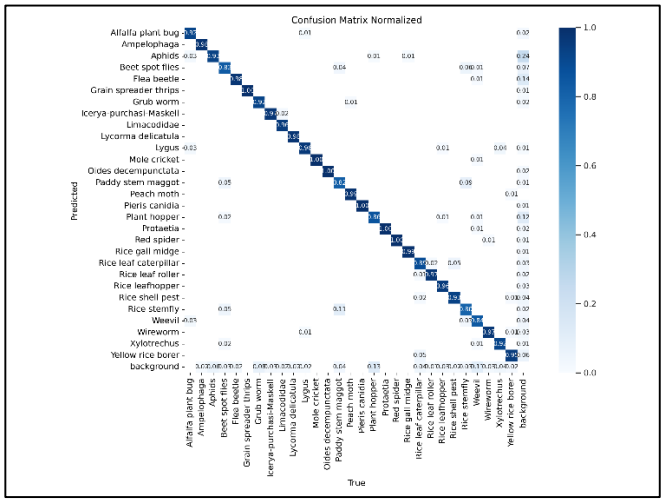


Fig. 13. Confusion matrix of YOLOv10-N.

TABLE IX. RESULTS OF YOLOV8N, YOLOV9T AND YOLOV10-N ON VALIDATION DATASET IN POST-TRAINING PHASE

Models	Precision	Recall	mAP^{val}@0.5	mAP^{val}@0.5:0.95
YOLOv8n	0.859	0.827	0.871	0.618
YOLOv9t	0.874	0.844	0.898	0.667
YOLOv10-N	0.873	0.837	0.883	0.639

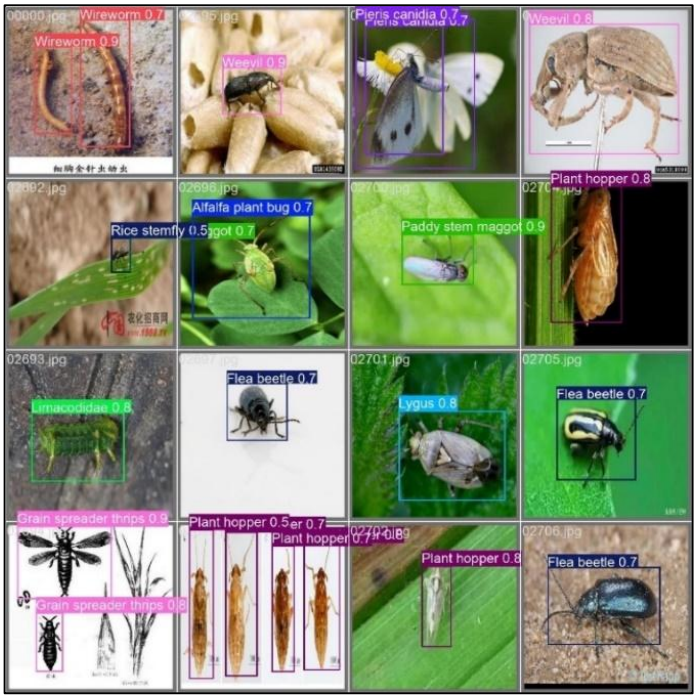


Fig. 14. Sample predictions made byYOLOv8n.

5) *Real-world application testing:* Strong performance of model during training and validation does not always guarantee equal success in real-world situations [53]. To further evaluate the YOLO models, they were converted to TensorFlow Lite format and integrated into a smartphone application. The testing was performed on a smartphone configured as detailed in Table X. This setup was used to evaluate the average inference time of each model by running them 10 times on a set of printed pest photos placed in a plant environment. The purpose of this test was to assess the models' practicality for real-time pest detection in agricultural fields.

Fig. 15. Sample predictions made byYOLOv9t.

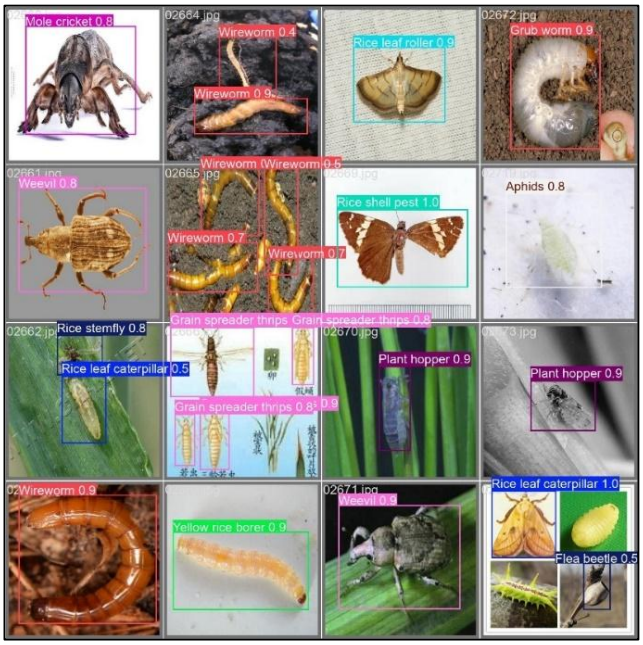


Fig. 16. Sample predictions made byYOLOv10-N.

As shown in the Table XI, lighter models like YOLOv9t recorded the fastest average inference time at 250.6ms, followed by YOLOv8n at 320.5ms and YOLOv10-N at

550.7ms. These results provide strong evidence that YOLOv9t is the optimal model to deploy on resource-constrained devices for achieving real-time detection in complex scenes.

TABLE X. ANDROID DEVICE SPECIFICATIONS

Component	Specification
Operating System	Android
Model	Xiaomi 11T Pro
RAM	16GB
Internal Memory	129GB
Chipset	Snapdragon 888 5G (5 nm)
CPU	Octa-core (1x2.84 GHz Cortex-X1 & 3x2.42 GHz Cortex-A78 & 4x1.80 GHz Cortex-A55)
GPU	Adreno 660

TABLE XI. AVERAGE INFERENCE TIME FOR EACH MODEL PERFORMED IN ANDROID APP (CALCULATED OVER 10 RUNS)

Model	Average Inference Time (ms)
YOLOv8n	320.5
YOLOv9t	250.6
YOLOv10-N	550.7

C. Discussion

A summary of the overall results can be seen in Table XII. The YOLOv9t model outperformed the other proposed models by achieving the highest values in key metrics such as accuracy and speed. Among the models tested, YOLOv9t provided the best balance between accuracy and speed. Therefore, it can accurately detect various tiny pests in real-time within challenging agricultural environments.

When compared the outcome of this study with recent studies as shown in Table XIII, it is evident that those models classified fewer object classes which ranged from 1 to 24 classes. While several studies achieved over 90% accuracy, their focus was often limited to specific object categories such as detecting leaves or identifying the presence of certain pests. Although the proposed model does not achieve the highest accuracy, it still obtains a remarkable result with 89% accuracy. It is closely comparable to the best-performing models while handling a larger and more diverse range of pest species.

The comparison also highlights the advantages of YOLOv9t over previous methods. For instance, Fuentes et al. and Lin et al. used SSD and Fast R-CNN, but these models struggled to detect a wide variety of pest classes. As a result, they achieved lower mAP scores and were less effective in diverse agricultural environments. Similarly, Sabanci et al.'s model demonstrated impressive precision. However, it was limited to binary classification for specific crops. This limitation restricted its overall versatility. Koklu et al. and Zhong et al. also made progress in tasks like leaf classification and insect recognition, but their models faced challenges due to a smaller number of pest classes and hardware limitations.

In contrast, the proposed YOLOv9t model stands out by successfully detecting 29 different pest classes with an accuracy of 89.8% and an inference time of just 250.6ms. Its lightweight architecture and scalability make it well-suited for real-time applications across a range of agricultural settings. Furthermore, the conversion of the YOLOv9t model to TensorFlow Lite (TFLite) and its integration into a mobile application significantly enhance its deployment capabilities on resource-constrained devices. The TFLite model allows for faster and more efficient inference, enabling real-time pest detection directly on smartphones. This integration ensures that pest management can be carried out in the field, improving efficiency for farmers and agricultural workers. This model's ability to overcome the limitations of earlier approaches highlights its practical advantages for pest management in real-world agricultural environment.

TABLE XII. OVERALL PERFORMANCE COMPARISON OF YOLOV8N, YOLOV9T AND YOLOV10-N USED IN THE EXPERIMENTAL STUDY

Models	Precision	Recall	mAP^{val}@0.5	mAP^{val}@0.5:0.95	Average Inference Time (ms)
YOLOv8n	0.859	0.827	0.871	0.618	320.5
YOLOv9t	0.874	0.844	0.898	0.667	250.6
YOLOv10-N	0.873	0.837	0.883	0.639	550.7

TABLE XIII. COMPARISON OF PEST DETECTION MODEL WITH PREVIOUS STUDIES

Author & Reference	Model	Pest Classes	Accuracy (%)	Key Insights
Fuentes et al. [14]	SSD, R-CNN, Faster R-CNN	9	83.06	Limited pest classes; lacks real-time performance.
Lin et al. [15]	Fast R-CNN	24	56.4	High recall but poor mAP, unsuitable for real-time dynamic environments.
Sabanci et al. [16]	AlexNet + BiLSTM	2	99.5	Extremely high precision but limited to specific crops and pests.
Koklu et al. [17]	CNN-SVM	5	97.6	Effective for leaf classification; untested for insect pest detection or broader applications.
Zhong et al. [19]	YOLO + Raspberry Pi	1	90.18	High accuracy but lacks scalability for detecting diverse pest classes.
Thanh-Nghi Doan [21]	YOLOv5-S	10; 102	70.5; 42.9	Performs well on small datasets; struggles with complex datasets due to limited scalability.

Author & Reference	Model	Pest Classes	Accuracy (%)	Key Insights
Yin Jian Jun [22]	YOLOv8	8	97.3	High accuracy but lacks scalability for detecting diverse pest classes.
Proposed YOLOv8n	YOLOv8n	29	87.1	Provides balance between computational efficiency and detection performance; lower accuracy than YOLOv9t.
Proposed YOLOv9t	YOLOv9t	29	89.8	Achieves high accuracy and scalability; balances speed and accuracy effectively.
Proposed YOLOv10-N	YOLOv10-N	29	88.3	Higher computational complexity; excels in detecting complex patterns but slower inference speed.

V. APPLICATION

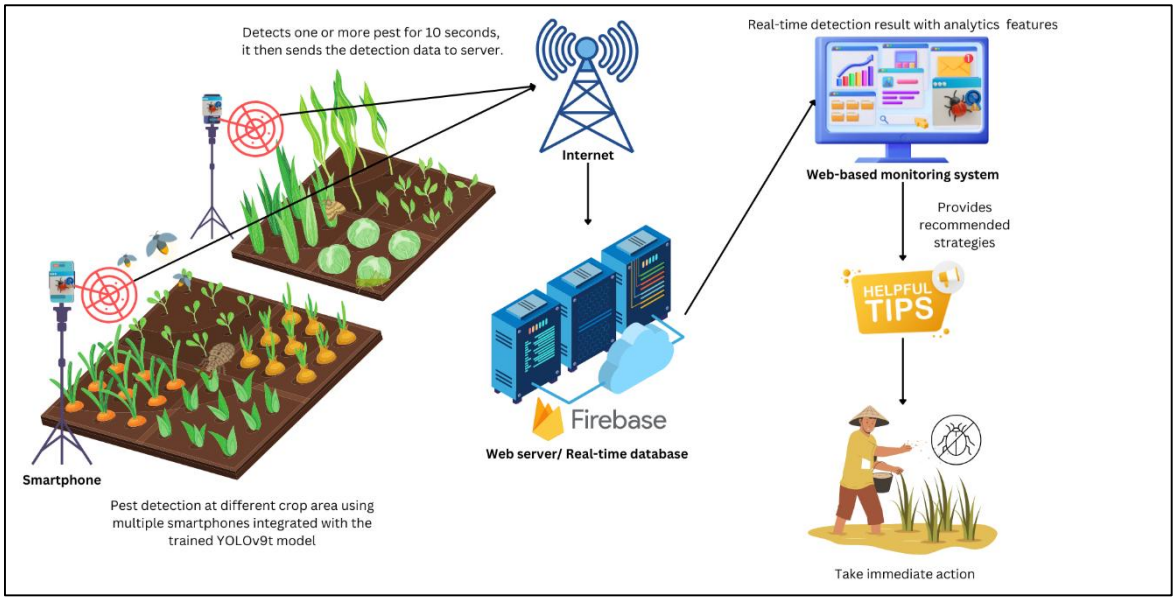


Fig. 17. Overview of the proposed lightweight mobile pest detection system with real-time monitoring.

A. Deployment

The YOLOv9t model was selected for deployment in the smartphone pest detection application due to its strong performance compared to other versions. The application detects pests in real-time but is programmed with a 10-second delay before sending the data to Firebase. This delay ensures that the pest remains on the plant and is not just passing by. Users can specify the crop area where the monitoring is taking place. Once the application detects one or more pests for 10 seconds, it sends the detection result to the Firebase server where it integrates a web-based monitoring system. This setup allows the user to track which particular area of the crop is getting infected. By providing this information, users can evaluate the situation and take immediate action if necessary. An overview of the pest detection system is shown in Fig. 17.

B. Example of Successful Pest Detection by YOLOv9t

Fig. 18 below shows the predictions made by the YOLOv9t model integrated into a smartphone application for detecting pests in real-time. It performs well in correctly identifying various pests in a simulated plant environment.

Fig. 18. Correct real-time predictions of pests using the YOLOv9t model integrated into a smartphone application in a simulated plant environment.

By using the app, users can also upload pest images for detection. They can retrieve detailed information about that uploaded pest, including a description and recommendations for managing the pest. If the pest is not detectable by the model, user can even upload the image to Firebase for experts to further refine the capabilities of model in detecting new pests in future. The web-based monitoring system complements the app by providing real-time detection logs, pest summary details, analytics dashboard, and access to detailed pest information for recommended strategies. Fig. 19 to Fig. 24 illustrate these additional features of the pest detection system.

Fig. 19. Upload pest image feature.

Fig. 20. Retrieved detailed information about the uploaded pest.

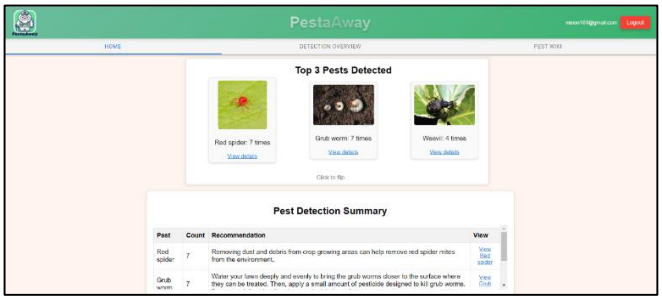


Fig. 21. Web-based monitoring system displaying pest population trends.

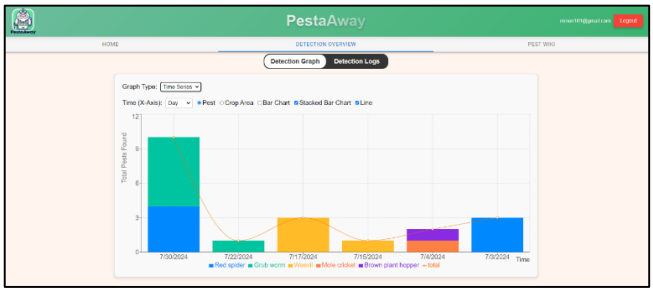


Fig. 22. Web-based monitoring system's pest analytics dashboard.

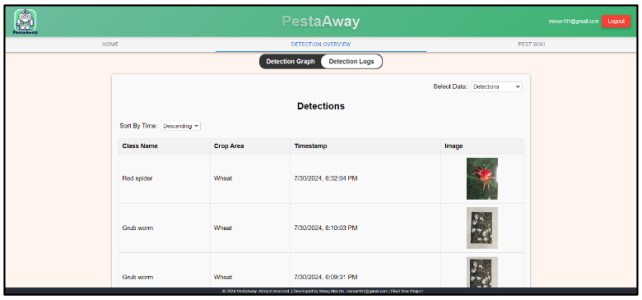


Fig. 23. Pest detection logs in the web-based monitoring system.

Fig. 24. Example of detailed pest information for one particular pest provided by the web-based monitoring system.

VI. Limitations and Potential Improvements

In this study, several limitations were identified that affect the detection performance of the integrated model within the application. One notable issue is when the pests overlap or move rapidly in the real-world condition. This makes the model difficult to detect pests and differentiate between them. As shown in Fig. 25, the integrated model successfully detects the aphids but fails to identify the mole cricket underneath. This results in false negative detection. Another challenge is that the model mistakes a leaf for aphids due to their similar color and shape. This misidentification affects the accuracy of the

detection especially when the environment closely resembles the pests. These limitations underscore the need for further refinement of models to improve detection accuracy in complex environments in future study.

To address these issues, incorporating object tracking and motion filtering techniques could be beneficial. By using these methods, the model could track and identify pests even when they overlap or move rapidly. Expanding the dataset to include more diverse pest images and backgrounds could also help the model generalize better to handle a broader range of real-world conditions.

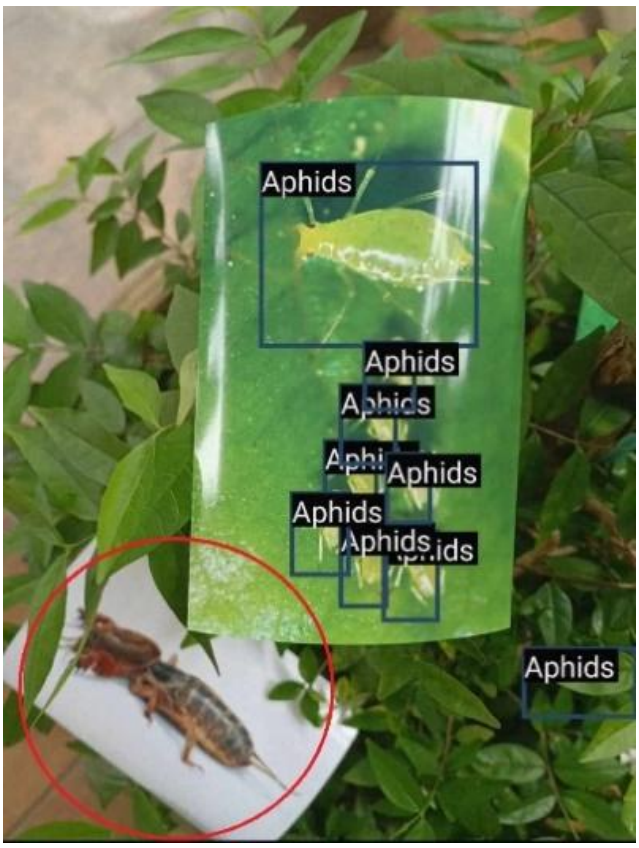


Fig. 25. The integrated YOLOv9t model detecting aphids but missing the mole cricket and misidentifying a leaf as aphids.

VII. CONCLUSION

This paper presented the development of a YOLO-based pest detection system integrated with a smartphone for real-time detection of multiple pest classes along with a dynamic web-based monitoring system for sustainable agriculture. The proposed system employs the YOLOv9t model due to its ability to strike a balance between computational resources, detection accuracy and speed. It has been shown to outperform similar systems by detecting a wider range of pests with fewer computational resources while achieving similar high detection accuracy.

The improved results can be attributed to the lightweight architecture of the proposed YOLOv9t model. It ensures faster inference times for detecting different kinds of pests on mobile devices without compromising accuracy. The numerical results show that the proposed YOLOv9t model achieved a remarkable accuracy with an mAP@0.5 of 89.8%, an mAP@0.5:0.95 of 66.7%, a Precision of 87.4%, a Recall of 84.4%, and an inference time of 250.6 ms. Also, the system enables farmers to monitor their crops by receiving instant updates through the web-based platform in real time. This combination allows for better pest management by providing detailed insights into pest activity across different fields. This approach offers notable advantages in terms of features, computational efficiency, practical implementation, and real-time detection speed.

While some detection inaccuracies might still occur when pests are overlapping or when there are similar backgrounds, shapes, and colors, future improvements will focus on addressing these challenges. The researchers plan to expand the dataset by including a wider variety of pest images taken from more diverse and complex environments. This will help the model better generalize to different real-world scenarios especially those with varying lighting conditions, backgrounds, and angles. Additionally, the researchers also plan to incorporate object tracking and motion filtering techniques to enable the model for tracking the pests more accurately even in situations where they overlap or move rapidly. These enhancements will improve the performance of the pest detection model in complex agricultural environments by ensuring higher detection accuracy and robustness in diverse pest management applications.

REFERENCES

[1] A. Balkrishna, G. Sharma, N. Sharma, P. Kumar, R. Mittal and R. Parveen, "Global perspective of agriculture systems: from ancient times to the modern era," In Sustainable Agriculture for Food Security, pp. 3-45, 2021.

[2] D. Gu, K. Andreev and M. Dupre, "Major trends in population growth around the world," China CDC weekly, vol. 3, no. 28, p. 604, 2021.

[3] FAO, "About FAO's work on plant Production and Protection," [Online]. Available: https://www.fao.org/plant-production-protection/about/en. [Accessed 20 July 2024].

[4] T. Domingues, T. Brandão and J. Ferreira, "Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey," Agriculture, vol. 12, no. 9, p. 1350, 2022.

[5] M. John, I. Bankole, O. Ajayi-Moses, T. Ijila, O. Jeje and P. Lalit, "Relevance of advanced plant disease detection techniques in disease and Pest Management for Ensuring Food Security and Their Implication: A review," American Journal of Plant Sciences, vol. 14, no. 11, pp. 1260-1295, 2023.

[6] P. Mkenda, P. Ndakidemi, P. Stevenson, S. Arnold, I. Darbyshire, S. Belmain, J. Priebe, A. Johnson, J. Tumbo and G. Gurr, "Knowledge gaps among smallholder farmers hinder adoption of conservation biological control," Biocontrol Science and Technology, vol. 30, no. 3, pp. 256-277, 2020.

[7] M. Gulzar, R. Maqsood, H. Abbas, M. Manzoor, M. Suleman, H. Bajwa, A. Hamza, S. Yar, M. Zain, A. Wadood and N. Aslam, "Use of Insecticides and their impact on viral diseases in Humans, Animals and Environment," Hosts and Viruses, vol. 11, pp. 64-77, 2024.

[8] N. Manakitsa, G. Maraslidis, L. Moysis and G. Fragulis, "A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision," Technologies, vol. 12, no. 2, p. 15, 2024.

[9] A. Nazir and M. Wani, "You only look once-object detection models: a review," in 2023 10th International Conference on Computing for Sustainable Global Development (INDIACom), 2023.

[10] M. Sohan, T. Sai Ram, R. Reddy and C. Venkata, "A review on yolov8 and its advancements," in International Conference on Data Intelligence and Cognitive Informatics, 2024.

[11] C. Ren, D. Kim and D. Jeong, "A survey of deep learning in agriculture: Techniques and their applications.," Journal of Information Processing Systems, vol. 16, no. 5, pp. 1015-1033, 2020.

[12] W. Changji, C. Hongrui, M. Zhenyu, Z. Tian, Y. Ce, S. Hengqiang and C. Hongbing, "Pest-YOLO: A model for large-scale multi-class dense and tiny pest detection and counting," Frontiers in Plant Science, vol. 13, p. 973985, 2022.

[13] Y. Zhang and C. Lv, "TinySegformer: A lightweight visual segmentation model for real-time agricultural pest detection," Computers and Electronics in Agriculture, vol. 218, p. 108740, 2024.

[14] A. Fuentes, S. Yoon, S. Kim and D. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," Sensors, vol. 17, no. 9, p. 2022, 2017.

[15] L. Jiao, S. Dong, S. Zhang, C. Xie and H. Wang, "AF-RCNN: An anchor-free convolutional neural network for multi-categories agricultural pest detection," Computers and Electronics in Agriculture, vol. 174, p. 105522, 2020.

[16] K. Sabanci, M. Aslan, E. Ropelewska, M. Unlersen and A. Durdu, "A novel convolutional-recurrent hybrid network for sunn pest–damaged wheat grain detection," Food analytical methods, vol. 15, no. 6, pp. 1748-1760, 2022.

[17] M. Koklu, M. Unlersen, I. Ozkan, M. Aslan and K. Sabanci, "A CNN-SVM study based on selected deep features for grapevine leaves classification," Measurement, vol. 188, p. 110425, 2022.

[18] L. Dengshan, W. Rujing, X. Chengjun, L. Liu, Z. Jie, L. Rui and W. Fangyuan, "A Recognition Method for Rice Plant Diseases and Pests Video Detection Based on Deep Convolutional Neural Network," Sensors, vol. 20, no. 3, p. 578, 2020.

[19] Y. Zhong, J. Gao, Q. Lei and Y. Zhou., "A vision-based counting and recognition system for flying insects in intelligent agriculture," Sensors, vol. 18, no. 5, p. 1489, 2018.

[20] A. M. Roy and J. Bhaduri., "Real-time growth stage detection model for high degree of occultation using DenseNet-fused YOLOv4," Computers and Electronics in Agriculture, vol. 193, p. 106694, 2022.

[21] T.-N. Doan, "An efficient system for real-time mobile smart device-based insect detection," International Journal of Advanced Computer Science and Applications, vol. 13, no. 6, 2022.

[22] J. Yin, P. Huang, D. Xiao and B. Zhang, "A Lightweight Rice Pest Detection Algorithm Using Improved Attention Mechanism and YOLOv8," Agriculture, vol. 14, no. 7, p. 1052, 2024.

[23] M. Hussain and R. Khanam, "In-depth review of yolov1 to yolov10 variants for enhanced photovoltaic defect detection," In Solar, vol. 4, no. 3, pp. 351-386, 2024.

[24] C.-Y. Wang and H.-Y. M. Liao, "YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems," arXiv preprint, p. arXiv:2408.09332, 2024.

[25] C. Chen, P. Zhang, H. Zhang, J. Dai, Y. Yi, H. Zhang and Y. Zhang., "Deep learning on computational-resource-limited platforms: A survey," Mobile Information Systems, vol. 1, p. 8454327, 2020.

[26] Roboflow. Available online: https://universe.roboflow.com/ip102110000/yoloip1/dataset/1. [Accessed 12 June 2024].

[27] Roboflow. [Online]. Available: https://universe.roboflow.com/oubio/pest-dataset-naoyq/dataset/3. [Accessed 12 June 2024].

[28] U. Sirisha, S. P. Praveen, P. N. Srinivasu, P. Barsocchi and A. K. Bhoi, "Statistical analysis of design aspects of various YOLO-based deep learning models for object detection," International Journal of Computational Intelligence Systems, vol. 1, no. 126, p. 16, 2023

[29] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016.

[30] T. Diwan, G. Anirudh and J. V. Tembhurne, "Challenges, architectural successors, datasets and applications," multimedia Tools and Applications, vol. 82, no. 6, pp. 9243-9275, 2023

[31] N. Kumar, Nagarathna and F. Flammini, "YOLO-based light-weight deep learning models for insect detection system with field adaption," Agriculture, vol. 13, no. 3, p. 741, 2023

[32] L. Liu, P. Li, D. Wang and S. Zhu, "A wind turbine damage detection algorithm designed based on YOLOv8," Applied Soft Computing, vol. 154, p. 111364, 2024.

[33] J. Terven, D. Córdova-Esparza and J. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," Machine Learning and Knowledge Extraction, vol. 5, no. 4, pp. 1680-1716, 2023.

[34] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," arXiv, arXiv:1612.03144, 2017.

[35] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, "Path Aggregation Network for Instance Segmentation," arXiv, arXiv:1803.01534, 2018.

[36] T. Wu and Y. Dong, "YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition," Applied Sciences, vol. 13, no. 24, p. 12977, 2023.

[37] M. Noyce, G. Jocher, R. Munawar and Laughing-Q, "COCO Dataset," Ultralytics YOLO, November 2023. [Online]. Available: https://docs.ultralytics.com/datasets/detect/coco/. [Accessed 15 August 2024].

[38] T. Luo, S. Rao, W. Ma, Q. Song, Z. Cao, H. Zhang, J. Xie, X. Wen, W. Gao, Q. Chen, J. Yun and D. Wu, "Individual Tree Spatial Positioning and Crown Volume Calculation Using UAV-RGB Imagery and LiDAR Data," Forests, vol. 15, no. 8, p. 1375, 2024.

[39] C.-Y. Wang, I.-H. Yeh and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," arXiv preprint, arXiv:2402.13616, 2024.

[40] W. Xu, D. Zhu, R. Deng, K. Yung and A. W. Ip., "Violence-YOLO: Enhanced GELAN Algorithm for Violence Detection," Applied Sciences, vol. 14, no. 15, p. 6712, 2024.

[41] M. Noyce, R. Munawar, G. Jocher, B. Q and L. Q, "YOLOv9: A Leap Forward in Object Detection Technology," Ultralytics YOLO, March 2024. [Online]. Available: ttps://docs.ultralytics.com/models/yolov9/#what-are-the-advantages-of-using-ultralytics-yolov9-for-lightweight-models. [Accessed 15 August 2024]

[42] A. Wang, H. Chen, L. Liu, K. Chen, Z. Lin, J. Han and G. Ding., "Yolov10: Real-time end-to-end object detection," arXiv preprint, p. arXiv:2405.14458, 2024.

[43] S. Geetha, Athulya, M. A. R. Alif, M. Hussain and P. Allen, "Comparative Analysis of YOLOv8 and YOLOv10 in Vehicle Detection: Performance Metrics and Model Efficacy," Vehicles, vol. 6, no. 3, pp. 1364-1382, 2024.

[44] M. Noyce, R. Munawar, G. Jocher, H. Haffari, Z. X. Wei, A. Vina and B. Q, "YOLOv10: Real-Time End-to-End Object Detection," Ultralytics YOLO, June 2024. [Online]. Available: https://docs.ultralytics.com/models/yolov10/. [Accessed 15 August 2024].

[45] L.-D. Quach, K. N. Quoc, A. N. Quynh and H. T. Ngoc, "Evaluating the eEffectiveness of YOLO models in different sized object detection and feature-based classification of small objects," Journal of Advances in Information Technology, vol. 14, no. 5, pp. 907-917, 2023.

[46] R. Kaur and S. Singh, "A comprehensive review of object detection with deep learning," Digital Signal Processing, vol. 132, p. 103812, 2023.

[47] J. Stodt, C. Reich and N. Clarke, "Unified intersection over union for explainable artificial intelligence," in Intelligent Systems Conference, IntelliSys 2023, Amsterdam, 2024.

[48] A. Tharwat, "Classification assessment methods," Applied computing and informatics, vol. 17, no. 1, pp. 168-192, 2021.

[49] D. A. Anggoro and S. S. Mukti, "Performance Comparison of Grid Search and Random Search Methods for Hyperparameter Tuning in Extreme Gradient Boosting Algorithm to Predict Chronic Kidney Failure," International Journal of Intelligent Engineering & Systems, vol. 14, no. 6, 2021.

[50] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," Journal of big Data, vol. 8, pp. 1-74, 2021

[51] D. Bergmann and C. Stryker, "What is a loss function?," IBM, 12 July 2024. [Online]. Available: https://www.ibm.com/think/topics/loss-function#:~:text=The%20term%20%E2%80%9Closs%20function%2C%E2%80%9D,intelligence%20(AI)%20model's%20outputs.. [Accessed 17 August 2024]

[52] R. Raj and A. Kos, "An improved human activity recognition technique based on convolutional neural network," Scientific Reports, vol. 13, no. 1, p. 22581, 2023

[53] C.W. Hoe, M. Raheem, and N.F. Abubacker, "News Aggregation and Summarisation," Journal of Applied Technology and Innovation, vol. 8, no. 4, p. 51, 2024.

Improved Decision Tree, Random Forest, and XGBoost Algorithms for Predicting Client Churn in the Telecommunications Industry

Mohamed Ezzeldin Saleh, Nadia Abd-Alsabour
Cairo University, Egypt

***Abstract*—Traditional machine learning models, especially decision trees, face great challenges when applied to high-dimensional and imbalanced telecommunication datasets. The research presented in this paper aims to enhance the performance of traditional Decision Tree (DT), Decision Tree with grid search (DT+), random forest (RF), and XGBoost (XGB) models. This is accomplished by augmenting them with robust preprocessing techniques, as well as optimizing them through grid search. We then evaluated how well the enhanced models can accurately predict customer churn and compared their performance metrics in detail. We utilized a dataset derived from the benchmark Cell2Cell dataset by applying combined preprocessing methods including KNN imputation, normalization, and resampling with SMOTE Tomek to address class imbalance. The findings reveal that XGBoost outperformed all other models with an accuracy of 0.82, demonstrating strong precision, recall, and F1 scores. RF also delivered robust results, achieving an accuracy of 0.82, benefiting from its ensemble nature to improve generalization and reduce overfitting.**

***Keywords*—Churn prediction; decision trees; grid search; random forest; XGBoost**

I. INTRODUCTION

The rapid evolution of the telecommunications industry has been marked by significant technological advancements and massive competition, leading to a saturated market where customer retention has become a critical challenge. As new telecom providers emerge, often offering specialized services at competitive prices, established firms must adopt more sophisticated strategies to maintain their market share. Customer churn has emerged as a significant concern in this highly competitive environment. Retaining existing customers is more cost-effective than acquiring new ones, but it is also crucial for maintaining steady revenue streams and long-term business growth [1]-[2].

Machine Learning (ML) has transformed the telecom industry by providing advanced tools for analyzing large datasets and predicting customer behavior. ML algorithms, particularly those focused on customer churn prediction (CCP), enable telecom companies to implement proactive retention strategies by accurately identifying at-risk customers [3]-[4]. However, the high dimensionality and imbalance inherent in telecom datasets pose significant challenges to traditional ML models like Decision Trees (DT). While these models are popular due to their simplicity and interpretability, they often suffer from overfitting and biased predictions when applied to complex, high-dimensional data [5]. The need for robust preprocessing techniques, such as imputation, normalization, and resampling, is essential to address these limitations and improve the predictive performance of ML models [6]-[7].

Despite advances in ML, gaps remain in optimizing churn prediction models for real-world applications. Current methodologies often struggle with imbalanced datasets, leading to skewed predictions that fail to capture minority class churners effectively. Ensemble models such as Random Forest (RF) and XGBoost (XGB) offer improved generalization and accuracy but require careful tuning of hyperparameters to maximize their effectiveness. Moreover, the lack of standardized preprocessing pipelines and scalable solutions limits the broader adoption of these methods in the telecom industry.

This study addresses these challenges by proposing a systematic approach to enhance the performance of DT, RF, and XGB models. The research emphasizes integrating advanced preprocessing techniques such as KNN imputation, normalization, and SMOTE Tomek resampling—with hyperparameter optimization using grid search. This approach seeks to mitigate the impact of imbalanced datasets and improve the robustness of predictive models for CCP.

The key objectives of this work are to:

- Develop a DT+ model optimized through grid search to address the limitations of traditional DT models.
- Compare the outcomes of DT+, RF, and XGB models in predicting customer churn, focusing on precision, accuracy, F1-score, and recall.
- Investigate the impact of preprocessing techniques, including imputation, normalization, and resampling, on the performance of these models.

By addressing these objectives, this research aims to contribute to the development of scalable, reliable, and interpretable models for customer churn prediction, offering actionable insights for the telecommunications industry to retain customers and reduce churn rates effectively.

The literature review is depicted in the following section. Section III introduces the proposed work with enhanced algorithms and its pseudocode. Hyperparameter optimization is presented in Section IV, the performance metrics are

addressed in Section V. Results and discussion are in Section VI. Section VII gives detail about the CCP performance. The closing section addresses the real-world significance in Section VIII, with limitations, conclusion, and the future scope of research of this research in Section IX.

II. Literature Review

CCP has become an important area of concern in the telecommunications industry, prompting extensive research on the effectiveness of various ML models. Among the most studied algorithms are DT, RF, and XGB, each offers unique strengths in enhancing predictive accuracy and model stability.

The study by [8] introduced a smart hybrid scheme that combined clustering and classification algorithms. Their results demonstrated that a stacking-based ensemble model combining k-medoids, Gradient Boosted Tree (GBT), DT, RF, and Deep Learning (DL) achieved the highest accuracy of 96%, highlighting the potential of hybrid models.

The study in [9] explored the use of advanced machine learning methods, particularly focusing on RF optimized by Grid Search and a low-ratio undersampling strategy. Their findings showed that the RF-GS-LR model achieved near-perfect accuracy on the applied datasets, underscoring the importance of hyperparameter optimization and sampling techniques in improving churn prediction models.

The study in [10] provided a comprehensive analysis of integrated algorithms, including enhanced Random Forest and XGBoost models.

Studies by [11] and [12] emphasized the role of big data platforms, ensemble methods, and attribute selection in enhancing the accuracy and stability of churn prediction models. Using techniques such as SMOTE and Edited Nearest Neighbor (ENN) for data balancing and ensemble procedures like bagging and boosting has significantly improved model performance.

DT methodologies are highly regarded for their straightforwardness and interpretability. The research in [13] demonstrated that the DT models can achieve 3% higher accuracy than more complex models, such as random forests, under certain conditions. However, DT models are susceptible to overfitting and need to be enhanced through hyperparameter tuning methods such as grid search. The study in [9] illustrated how grid search optimization could significantly improve DT's accuracy and stability, especially when combined with controlled undersampling strategies.

RF is an ensemble method built on multiple DTs and is recognized for its robustness and ability to handle large datasets. The study in [14] highlighted RF's high accuracy of 95% when feature engineering techniques were applied, underscoring the importance of preprocessing steps [14]. In spite of its preferences, RF's complexity can be a restriction. However, studies have shown that when RF is regularized or combined with techniques like up-sampling and Edited Nearest Neighbor (ENN), it can achieve exceptionally high accuracies, reaching up to 99.09% [11], [15].

XGB, a gradient-boosting technique, has received widespread popularity for its outstanding usefulness in classification activities, particularly in churn prediction. Studies have revealed that XGB outperformed other ensemble algorithms, including Adaboost and CatBoost, especially when coupled with grid search cross-validation for hyperparameter tuning. XGB's capacity to handle sparse data and mitigate overfitting makes it highly effective for complex datasets [16]-[17]. The study in [18] further validated XGB's efficacy, achieving a 97% accuracy rate on the Cell2Cell dataset.

Integrating decision trees, random forests, and XGBoost algorithms alongside advanced optimization techniques like Grid search provides a comprehensive and robust approach to churn prediction. Studies by [19] emphasized the advancements in predictive power achieved through the combination of feature engineering, ensemble methods, and hyperparameter optimization, which significantly improved the accuracy, stability, and generalization across diverse telecom datasets.

Despite the advancements in CCP models, there are still several gaps that need to be addressed. Traditional decision tree models, while effective, are prone to overfitting and require optimization techniques like grid search to achieve optimal performance. Existing literature has shown improvements through these techniques, but further exploration is needed to address limitations in model interpretability and scalability [9], [13]. Random forest models, though robust, present challenges in terms of complexity and computational cost [11], [14]. The reliance on feature engineering to achieve high accuracy suggests the need for more efficient methods to handle raw data effectively. Additionally, although XGBoost shows superior performance, its susceptibility to overfitting and the need for extensive hyperparameter tuning suggest opportunities for further research on more generalized models. Furthermore, while recognized as crucial, hyperparameter optimization and sampling techniques require deeper investigation to develop standardized methodologies that can be applied to different datasets and industries [16]-[17].

III. Proposed Work

This section depicts the study details, pre-processing, and methodologies employed.

The study used the comprehensive dataset Cell2Cell, which contains client behavior attributes such as personal information, utilization patterns, client interactions, demographic details, billing data, and value-added services. These properties provide a solid foundation for developing and validating machine learning models [8], [20].

The research process, illustrated in Fig. 1, is structured into distinct phases, beginning with data preprocessing, which is crucial to ensuring the accuracy and reliability of the models. KNN Imputation (Mean/Median) addresses lost values within the dataset. This is followed by normalization utilizing MMADN Min-Max Scaling, and class imbalance is managed with SMOTE Tomek [20].

The preprocessed dataset is then split into training and testing parts in an 80-20 ratio, facilitating a robust evaluation of the models' performance. The DT and DT+ models serve as the baseline, with DT+ incorporating enhancements such as optimized hyperparameters identified through grid search techniques. RF, known for its ensemble approach, is also applied to improve the prediction accuracy and generalization further. Finally, XGBoost, renowned for its efficiency and scalability, is utilized, leveraging grid search for hyperparameter tuning.

The models were implemented on Google Colab, leveraging its computing resources for efficient processing. The analysis underscores the importance of systematic preprocessing and model enhancement in achieving high accuracy in churn prediction. By comparing the performance metrics of the DT, DT+, RF, and XGB models, this study provides empirical evidence for the effectiveness of advanced classification algorithms in CCP.

The preprocessing phase is essential to ensuring accurate predictions for customer churn models. Missing data is addressed using KNN imputation and median/Mean imputation techniques, preserving the dataset integrity and reducing bias [20]-[21]. Normalization follows, employing a combination of MMADN and Min-Max Scaling, which standardizes features while managing outliers effectively [22], [23], [24], [25]. To address the class imbalance, the SMOTE Tomek was utilized [18], [20], [25]. These preprocessing steps ensure that the data is optimally prepared for applying the DT, DT+, RF, and XGB models, leading to improved prediction accuracy and model robustness.

The proposed approach involves analyzing customer churn using the following four different algorithms:

- Traditional Decision Tree (DT)
- Decision Tree with Grid Search (DT+)
- Random Forest (RF)
- XGBoost (XGB)

A. Traditional Decision Tree (DT)

The Traditional Decision Tree model is valued for its effortlessness and intuitive interpretability. It works by recursively partitioning the dataset based on feature values, creating a tree-like structure where nodes represent decision rules and leaves denote class labels. This model excels at handling non-linear relationships, making it a popular choice for classification tasks. Nevertheless, DTs are inclined to overfit, particularly with complex or noisy data, necessitating careful tuning of parameters like max_depth, min_samples_split, and min_samples_leaf to enhance generalization.

B. Decision Tree with Grid Search (DT+)

The Decision Tree with grid search model refines the traditional Decision Tree by incorporating advanced techniques to mitigate overfitting and improve predictive accuracy. DT+ is designed to balance model complexity and interpretability, making it particularly suitable for applications where both robust performance and transparency in decision-making are crucial. By optimizing hyperparameters, DT+ effectively captures meaningful patterns from data, addressing the limitations of the traditional DT model.

DT+ leverages a few essential parameters to boost performance:

- Max_depth: Restricts the tree's depth, averting the model from getting too complicated and prone to overfitting
- Min_samples_split: Defines the minimum number of samples required to split a node, reducing the risk of creating insignificant splits.
- Criterion: The choice of Gini impurity or entropy when splitting a node directly affects the quality of the formed decision boundary.
- Pruning Techniques: These are applied post-training to remove branches that do not provide significant power in classifying the target variable, further reducing overfitting.

The implementation of DT+ begins with training an initial Decision Tree model using default parameters to establish a baseline. This model is then evaluated using accuracy metrics, a classification report, and a confusion matrix. To enhance the traditional DT model, grid search performs hyperparameter tuning to identify the best parameter settings from a predefined distribution. The tuned model is then evaluated on the test set, with pruning techniques applied to ensure that the model is not only accurate but also generalizable. This systematic approach ensures that the DT+ model outperforms the standard DT by avoiding overfitting and improving decision-making transparency.

Despite its enhancements, DT+ faces several challenges:

- Overfitting: Although it can be mitigated via pruning and parameter tuning, the risk of overfitting still exists, especially when the tree becomes too complex.
- Sensitivity to Data Variability: Like the traditional DT, DT+ can be sensitive to small changes in the dataset, which might lead to significant variations in the tree structure.
- Computational Complexity: Including advanced techniques such as hyperparameter tuning and pruning increases the computational burden, particularly with large datasets and extensive parameter grids.

The Pseudocode for the DT+ model development is:

Input: dataset.csv, Output: Model evaluation metrics, plots, comparison CSV.

- Import Libraries
- Load Dataset
- Data Preparation
- Initial Model Training with Default Parameters
- Hyperparameter Tuning using Grid search

- Comparison of Parameters and Accuracy
- Visualization Based on Performance Metrics

C. Random Forest (RF)

The Random Forest algorithm addresses single Decision Trees' limitations, particularly their susceptibility to overfitting. It is an ensemble learning strategy that boosts classification performance by consolidating the predictions of a number of Decision Trees, each trained on distinct portions of the data and attributes. This ensemble approach improves generalization, stability, and accuracy, making RF a powerful tool for complex classification tasks such as CCP. RF's ability to handle large datasets with high dimensionality and provide feature importance estimation further strengthens its applicability in various domains.

RF relies on several key parameters to optimize its performance:

- n_estimators: Determine the no. of decision trees in the ensemble. Increasing this number generally improves accuracy but increases computational costs.
- max_attribute: Determine the maximum no. of attributes considered for partitioning at every node, which provides randomness & diversity amongst the trees, boosting the model's robustness.
- Tree-Specific Parameters: These include parameters like max_depth, min_samples_split, and criterion, similar to those in Decision Trees but applied collectively across all trees in the ensemble.

The development of RF starts with creating an ensemble of decision trees. Every tree is trained on a bootstrap example of the data, with a random portion of attributes chosen for every partition. The last prediction is obtained by aggregating the predictions from all trees, regularly by means of majority voting. Hyperparameters such as the no. of trees (n_estimators), maximum features (max_features), and tree-specific parameters are fine-tuned to optimize the model's performance. Grid search allows for efficient hyperparameter tuning, ensuring that the model generalizes well to unseen data.

RF offers several advantages over single-decision trees. One of its main merits is enhanced generalization. By averaging the predictions of numerous trees, RF diminishes the risk of overfitting, which typically enhances execution on novel data. In addition, the strength of RF comes from the diversity amongst its constituent trees, making it less sensitive to noise & variability in the data. Another advantage is its ability to provide estimates of feature importance, which can be valuable for interpreting the model's decisions.

However, RF also has its drawbacks. The ensemble nature of RF demands significantly more computational resources than single Decision Trees, thus increasing the computational complexity. Moreover, while RF's accuracy is superior, its predictions are less interpretable due to the complexity involved in aggregating the outputs of multiple trees. Reduced comprehensibility hinders comprehension of how the model arrived at its conclusions.

D. XGBoost (XGB)

XGBoost (Extreme Gradient Boosting) is employed in CCP to leverage its superior performance in handling complex data interactions and minimizing errors through iterative refinement. Differentiated from conventional ensemble strategies, XGBoost develops trees sequentially, with each novel tree rectifying the errors caused by the past ones [10], [16]. This iterative strategy, integrated with gradient descent optimization, permits XGBoost to accomplish large accuracy & strength in classification issues. Its ability to incorporate regularization techniques makes it particularly effective in preventing overfitting [17], [19].

XGBoost's performance is susceptible to its hyperparameters, and tuning these parameters is crucial to prevent overfitting and ensure robust performance. Grid search efficiently explores a wide range of hyperparameter combinations, allowing for a thorough yet computationally feasible optimization process [16]-[17].

XGBoost's performance is highly dependent on several key parameters:

- Learning Ratio: Controls the commitment of each tree to the last model. A lower learning rate requires more boosting rounds but can lead to a better generalization.
- Greatest Depth (max_depth): Constrains the depth of each tree, adjusting model complexity with overfitting risk.
- No. of Boosting Rounds (n_estimators): Determines the no. of trees to be included successively. More trees can capture more patterns, but this may increase the risk of overfitting.

Implementing XGBoost begins with importing the necessary libraries. The data is split into training and testing parts, similar to the process used for RF. A parameter grid is set, and Grid search is utilized to seek the ideal hyperparameters efficiently. This approach ensures that the model achieves the best possible performance while avoiding overfitting. The best model identified through Grid search is then trained on the full training set and evaluated on the test set to validate its accuracy and generalization capabilities.

While XGBoost offers several advantages, including superior accuracy and the ability to handle missing data, it also presents challenges:

- Computational Complexity: XGBoost's iterative approach and need for extensive hyperparameter tuning can increase computational costs.
- Overfitting: Despite its regularization techniques, XGBoost can still overfit, particularly on small or noisy datasets, if not properly tuned.
- Interpretability: The complexity of the model can make it challenging to interpret, especially when compared to simpler models like decision trees.

The detailed Pseudocode for the XGB model development has been described below.

Input: dataset.csv, Output: Model evaluation metrics, plots

- Import Libraries
- Load Dataset
- Data Preparation
- Split Data
- Hyperparameter Tuning with Grid search
- Train the Best Model
- Evaluate the Model
- Visualization Based on Performance Metrics

IV. Hyperparameter Optimization

Hyperparameter tuning was conducted using grid search to optimize the performance of the predictive models. This process systematically evaluated a range of hyperparameter combinations to identify the best configuration for each model. The optimal hyperparameters for the DT+, RF, and XGBoost models are summarized below.

A. Improved Decision Tree (DT+)

The DT+ model, an optimized version of the traditional Decision Tree, was tuned to improve its performance by leveraging grid search. The best-performing hyperparameter combination for DT+ included:

- class_weight: {0: 1, 1: 5}
- criterion: entropy
- max_depth: 70
- max_features: None
- min_samples_leaf: 1
- min_samples_split: 2
- splitter: random

This configuration balanced the dataset effectively, reducing overfitting and improving decision-making across deep trees.

B. Random Forest (RF)

The Random Forest model, an ensemble technique, was optimized to maximize generalization and reduce variance. The best hyperparameter combination for RF was:

- bootstrap: False
- class_weight: {0: 1, 1: 2}
- criterion: gini
- max_depth: None
- max_features: log2
- min_samples_leaf: 1
- min_samples_split: 2
- n_estimators: 120

This configuration emphasized utilizing a larger number of estimators while balancing the dataset using class_weight, enhancing the model's robustness.

C. XGBoost (XGB)

The XGBoost model demonstrated its strength in handling high-dimensional and imbalanced datasets with the following optimized hyperparameters:

- learning_rate: 0.05
- max_depth: 50
- n_estimators: 100
- subsample: 0.8

This combination provided a balance between the learning rate and the depth of the trees, enabling the model to refine predictions iteratively while avoiding overfitting.

D. Impact of Hyperparameter Optimization

The tuning process significantly contributed to the improved performance of the models, as demonstrated in the results:

- DT+: Showed notable consistency in metrics, particularly recall, due to balanced class weights and randomized splitting criteria.
- RF: Achieved strong generalization with an accuracy of 0.82 and a ROC-AUC of 0.87, leveraging its optimal tree-based ensemble design.
- XGB: Delivered the best overall performance with an ROC-AUC of 0.88, benefiting from gradient boosting and iterative error correction.

These hyperparameter combinations underscore the importance of systematic optimization in achieving reliable and accurate predictions for CCP.

V. Performance Metrics

To evaluate the predictive accuracy of our models in churn prediction, we utilize key metrics from the confusion matrix, which categorizes predictions into four essential types: true positives, true negatives, wrong positives, and wrong negatives. These metrics are critical for assessing the effectiveness of our classification algorithms [22], [23], [24], [25].

- F-measure: Balances precision and recall by calculating their harmonic mean, offering a single metric to evaluate overall model performance.
- Precision: Valuates the accuracy of figuring out the extent of true positives with regard to all discovered churners.
- Recall: Measures the model's ability to correctly identify actual churners by calculating the proportion of true positives out of all actual churners.
- Accuracy: Reflects the overall correctness of the model by measuring the ratio of correct predictions to total predictions.

VI. Results and Discussion

This section compares the DT, DT+, RF, and XGB models to assess their effectiveness in predicting customer churn in the telecommunications industry. The performance of each model is evaluated based on accuracy, precision, recall, F1-score, and ROC-AUC to provide a clear understanding of their relative strengths.

The effectiveness of the classification models was assessed using the prepared Cell2Cell dataset. Preprocessing steps included imputation, normalization, and resampling techniques, ensuring the dataset was adequately prepared for model evaluation. As shown in Table I, ensemble methods (RF and XGB) significantly outperformed single-tree models (DT and DT+).

TABLE I. The Performance of Utilized Models

Models	Accuracy	Precision	Recall	F1-Score	ROC
DT	0.77	0.78	0.77	0.77	0.74
DT+	0.77	0.77	0.77	0.77	0.73
RF	0.82	0.83	0.82	0.81	0.87
XGB	0.82	0.82	0.82	0.81	0.88

- DT Model: The DT model reported moderate performance with an accuracy of 0.77, precision of 0.78, recall and F1-score of 0.77, and ROC-AUC of 0.74. These metrics highlight the baseline capability of a traditional decision tree structure.
- DT+ Model: With grid search hyperparameter optimization, the DT+ model achieved similar results, with accuracy, precision, and recall, an F1-score of 0.77 each, and a ROC-AUC of 0.73. The enhancements ensured consistent performance but did not significantly outperform the traditional DT.
- RF Model: The RF model demonstrated substantial improvement across all metrics. It achieved an accuracy of 0.82, precision of 0.83, recall of 0.82, and F1-score of 0.81. Its ensemble approach effectively leveraged imputed data and the SMOTE Tomek resampling method, resulting in a high ROC-AUC value of 0.87.
- XGBoost Model: XGBoost delivered the highest overall performance, achieving an accuracy of 0.82, precision, recall, and F1-score of 0.82 each, and a ROC-AUC of 0.88. The advanced optimization of gradient boosting and integration of regularization features contributed significantly to this outcome.

Fig. 2 to 7 provide classification reports and confusion matrices for the DT+, RF, and XGBoost models, illustrating their predictive performance.

The results highlight the advantages of ensemble methods like RF and XGBoost over single-tree models in churn prediction. The XGBoost model's superior performance can be attributed to its gradient-boosting framework, which iteratively refines predictions, effectively capturing complex patterns in the data. This aligns with findings from prior studies that emphasize the efficacy of gradient-boosting techniques for high-dimensional datasets. The RF model's robust results further underscore the value of ensemble techniques in enhancing generalization and reducing overfitting. The use of SMOTE Tomek in preprocessing was critical in addressing the class imbalance, as evident in the improved recall and precision scores for both RF and XGBoost. However, the limited impact of this technique on DT and DT+ indicates that more sophisticated models are better equipped to exploit balanced datasets.

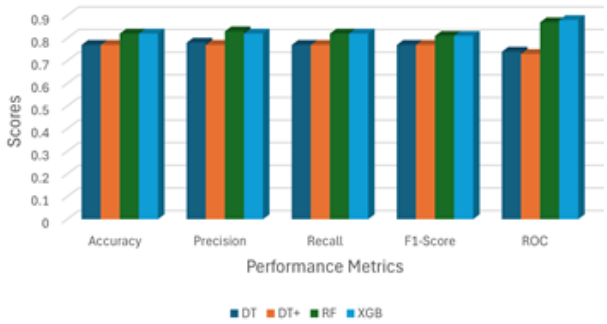


Fig. 1. CCP performance of ML models.

```
Accuracy of the model:
0.7699539712075213
Classification Report:
              precision    recall  f1-score   support

           0       0.84      0.83      0.84      7165
           1       0.61      0.62      0.62      3046

    accuracy                           0.77     10211
   macro avg       0.73      0.73      0.73     10211
weighted avg       0.77      0.77      0.77     10211
```

Fig. 2. Classification report for DT+.

```
Accuracy of the model:
0.8225443149544609
Classification Report:
              precision    recall  f1-score   support

           0       0.81      0.97      0.88      7165
           1       0.86      0.48      0.62      3046

    accuracy                           0.82     10211
   macro avg       0.84      0.73      0.75     10211
weighted avg       0.83      0.82      0.81     10211
```

Fig. 3. Classification report for RF.

```
Accuracy of the model:
0.823719518166683
Classification Report:
              precision    recall  f1-score   support

           0       0.83      0.95      0.88      7165
           1       0.81      0.53      0.64      3046

    accuracy                           0.82     10211
   macro avg       0.82      0.74      0.76     10211
weighted avg       0.82      0.82      0.81     10211
```

Fig. 4. Classification report for XGB.

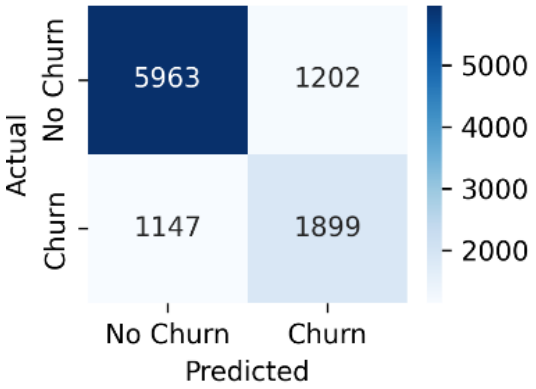


Fig. 5. Confusion matrix for DT+.

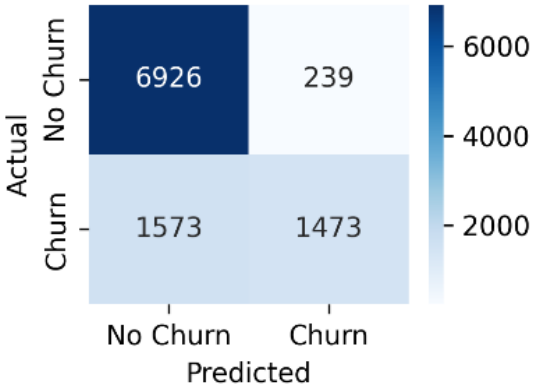


Fig. 6. Confusion matrix for RF.

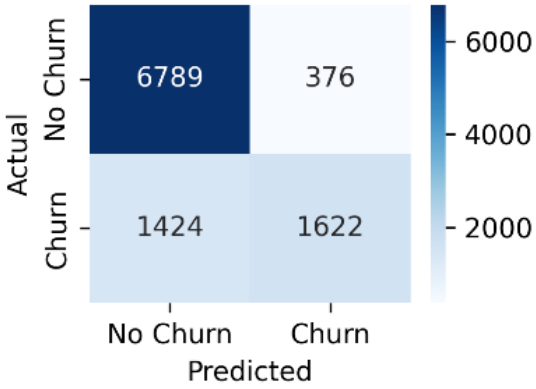


Fig. 7. Confusion matrix for XGB.

While the DT and DT+ models offered a balanced trade-off between recall and precision, their performance was constrained by inherent algorithmic limitations, such as susceptibility to overfitting. Grid search optimization for DT+ improved consistency but did not elevate its metrics significantly above those of the traditional DT model. The imputation and normalization steps proved vital, as they contributed to the balanced recall values across all models, ensuring accurate identification of churners. However, the reliance on advanced optimization techniques, such as those employed in XGBoost, demonstrates the need for robust preprocessing and model design to maximize predictive accuracy. Future work could explore alternative preprocessing strategies or combine these methods with deep learning models to achieve further improvements.

VII. CCP Performance Comparison with Existing CCP Methods Using The CELL2CELL Dataset

The Cell2Cell dataset has been extensively used in the telecommunications sector to develop and evaluate customer churn prediction (CCP) models. Various studies have explored various approaches, from traditional machine learning techniques to advanced deep learning frameworks, achieving diverse outcomes based on the methodologies and preprocessing techniques employed. The study in [26] applied a Deep-BP-ANN model combined with Lasso Regression and Variance Thresholding, achieving an accuracy of 79.38%. Their results outperformed traditional XGBoost and Logistic Regression models, demonstrating the potential of deep learning for churn prediction in imbalanced datasets. The study in [8] developed a hybrid ensemble approach combining clustering and classification algorithms, including k-medoids, Gradient-Boosted Trees (GBT), Decision Trees (DT), and Deep Learning (DL). This method achieved an accuracy of 93.6%, highlighting the effectiveness of integrating clustering techniques to manage class imbalance. The study in [19] explored decision forest models enhanced with weighted soft voting. Their approach achieved an accuracy of 96.57%, showcasing the advantages of ensemble methods in improving prediction accuracy and robustness by effectively identifying churn patterns.

Table II summarizes the performance metrics reported in these studies, providing a benchmark for assessing the efficacy of different CCP methodologies.

TABLE II. Performance Metrics Comparison from Existing Studies

Method	Accuracy	Precision	Recall	F1-Score	AUC
Deep-BP-ANN [26]	79.38	74.50	89.32	81.24	79.38
Hybrid.Ensemble [8]	93.6	79.10	67.45	72.81	93.6
Decision Forest [19]	96.57	96.57	85.45	83.72	96.57

Our study, which uses the Cell2Cell dataset with advanced preprocessing techniques, aligns with findings from these prior works while offering unique contributions. Preprocessing steps like KNN imputation, normalization, and SMOTE Tomek resampling effectively addressed class imbalance, enhancing model performance. Ensemble models such as RF and XGBoost achieved high accuracy (0.82), as highlighted in Table I, which compares the metrics of the models used in this study and key observations are as follows:

- Performance Context: Our study's ensemble approaches (RF and XGBoost) achieved competitive accuracy scores compared to more straightforward machine learning frameworks like DT and DT+. While slightly below the performance of Decision Forest reported by [19], the balance between interpretability and accuracy makes these models practical for real-world use.

- Preprocessing Impact: SMOTE Tomek and normalization proved essential in balancing class distribution, improving recall and precision across models. This approach parallels the findings of Liu et al. [8], who leveraged clustering techniques for similar benefits.
- Scalability and Practicality: Unlike computationally intensive deep learning models, XGBoost and RF offer a cost-effective solution for telecom companies. Their high accuracy and efficient preprocessing make them suitable for deployment in dynamic environments where real-time churn prediction is critical.

VIII. Real World Significance

The findings of this study have critical implications for addressing real-world challenges in the telecommunications industry, where customer churn remains a significant concern. By leveraging predictive models like XGBoost, telecom companies can transition from reactive to proactive churn management strategies, leading to tangible business benefits.

Customer churn directly impacts revenue and operational efficiency in saturated markets where acquiring new customers is significantly costlier than retaining existing ones. The predictive models evaluated in this study, particularly XGBoost and RF, provide robust tools for identifying high-risk customers. These insights empower telecom companies to design targeted retention strategies, such as personalized offers, improved customer service, or loyalty programs, mitigating churn effectively.

The predictive algorithms demonstrated in this study can be seamlessly integrated into Customer Relationship Management (CRM) systems. For instance, by embedding XGBoost into customer analytics platforms, companies can automate churn predictions and deliver actionable insights in real-time. The models' ability to handle high-dimensional data and imbalanced classes also ensures scalability, making them suitable for large, complex telecom datasets.

Proactive churn management driven by these models could result in measurable outcomes, including:

- Cost Savings: Reducing churn by even a small percentage can save millions in customer acquisition costs.
- Revenue Enhancement: Retaining high-value customers boosts recurring revenue and long-term profitability.
- Operational Efficiency: Automating churn prediction reduces manual analysis, freeing resources for strategic initiatives.

The study also addresses broader industry challenges, such as fostering customer loyalty in a competitive landscape. Predictive models not only assist in retaining existing customers but also enhance customer experience by anticipating needs and preferences. These advancements align with the strategic goals of telecom firms to sustain growth and remain competitive. The practical relevance of this research extends beyond theoretical improvements, offering telecom companies a pathway to adopt data-driven strategies for churn management. By implementing these models, businesses can achieve financial benefits and strengthen customer relationships, driving sustainable growth in a dynamic market environment.

IX. Conclusion and Future Work

This study comprehensively analyzed customer churn prediction (CCP) using a Decision Tree (DT). The Decision Tree improved with grid search (DT+), Random Forest (RF), and XGBoost (XGB) algorithms applied to the preprocessed Cell2Cell dataset. Among these models, XGBoost emerged as the most effective, achieving strong precision, recall, and F1 scores with an accuracy of 0.82. Its advanced optimization techniques, such as gradient boosting and error correction, maximized the benefits of preprocessing methods, demonstrating its superiority in handling complex and high-dimensional datasets. The RF model also delivered robust performance, achieving an accuracy of 0.82 while effectively balancing precision and recall. Its ensemble nature successfully mitigated overfitting and enhanced generalization. In contrast, the DT+ model, despite its improvements through grid search, faced limitations in reaching comparable performance, underscoring the inherent constraints of decision tree-based models.

This research contributes to the field by integrating robust preprocessing techniques, such as KNN imputation, normalization, and SMOTE Tomek resampling, with grid search optimization. These methodologies collectively address critical challenges posed by imbalanced and high-dimensional telecom datasets, providing a scalable and systematic framework for CCP. By evaluating the comparative performance of DT, RF, and XGBoost models, the study underscores the value of ensemble methods and advanced hyperparameter tuning in achieving accurate and reliable predictions. Furthermore, the findings validate the significance of preprocessing as a cornerstone for effective churn prediction, offering insights into how these techniques enhance model robustness.

While this study provides valuable insights, certain limitations should be acknowledged. The analysis is constrained to the Cell2Cell dataset, which, while comprehensive, may not fully represent the diversity of customer behaviors in other industries or regions. Additionally, the models used in this study rely heavily on preprocessing and hyperparameter tuning, which may increase computational costs for large-scale datasets. Although effective, the interpretability of ensemble methods like XGBoost can be challenging, potentially limiting their application in contexts requiring high transparency.

Future research could explore incorporating deep learning models like neural networks that integrate different classifiers for enhanced predictive power. Investigating advanced feature engineering techniques, such as automated feature selection and interaction effects, could further improve model performance. Additionally, extending this research to other industries, such as finance or retail, with diverse customer behaviors would help validate the generalizability of the proposed methods. Exploring real-time churn prediction systems and using external data sources, such as social media

or customer feedback, could also offer new avenues for development.

The study's findings hold significant practical implications for the telecommunications industry. By providing actionable insights into the design and optimization of predictive models, this research supports proactive customer retention strategies, enabling telecom companies to reduce churn rates and enhance profitability. Moreover, the methodologies presented in this work contribute to advancing the knowledge base in CCP, offering scalable and interpretable solutions for addressing the challenges of imbalanced and high-dimensional datasets.

ACKNOWLEDGMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editor and the reviewers. Any statements, claims, performances and results are not guaranteed or endorsed by the publisher.

REFERENCES

[1] H. K. Thakkar, A. Desai, S. Ghosh, P. Singh, and G. Sharma, "Clairvoyant: AdaBoost with Cost-Enabled Cost-Sensitive Classifier for Customer Churn Prediction," Comput. Intell. Neurosci., vol. 2022, no. 1, p. 9028580, 2022, doi: 10.1155/2022/9028580.

[2] T. Zhang, S. Moro, and R. F. Ramos, "A Data-Driven Approach to Improve Customer Churn Prediction Based on Telecom Customer Segmentation," Future Internet, vol. 14, no. 3, Art. no. 3, Mar. 2022, doi: 10.3390/fi14030094.

[3] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes," Appl. Soft Comput., vol. 137, p. 110103, Apr. 2023, doi: 10.1016/j.asoc.2023.110103.

[4] A. Khattak, Z. Mehak, H. Ahmad, M. U. Asghar, M. Z. Asghar, and A. Khan, "Customer churn prediction using composite deep learning technique," Sci. Rep., vol. 13, no. 1, p. 17294, Oct. 2023, doi: 10.1038/s41598-023-44396-w.

[5] S. O. Abdulsalam, M. O. Arowolo, Y. K. Saheed, and J. O. Afolayan, "Customer Churn Prediction in Telecommunication Industry Using Classification and Regression Trees and Artificial Neural Network Algorithms," Indones. J. Electr. Eng. Inform. IJEEI, vol. 10, no. 2, Art. no. 2, Jun. 2022, doi: 10.52549/ijeei.v10i2.2985.

[6] S. Alam and N. Yao, "The impact of preprocessing steps on the accuracy of machine learning algorithms in sentiment analysis," Comput. Math. Organ. Theory, vol. 25, pp. 319–335, 2019.

[7] K. Cabello-Solorzano, I. Ortigosa de Araujo, M. Peña, L. Correia, and A. J. Tallón-Ballesteros, "The Impact of Data Normalization on the Accuracy of Machine Learning Algorithms: A Comparative Analysis," in 18th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2023), P. García Bringas, H. Pérez García, F. J. Martínez de Pisón, F. Martínez Álvarez, A. Troncoso Lora, Á. Herrero, J. L. Calvo Rolle, H. Quintián, and E. Corchado, Eds., Cham: Springer Nature Switzerland, 2023, pp. 344–353. doi: 10.1007/978-3-031-42536-3_33.

[8] R. Liu et al., "An Intelligent Hybrid Scheme for Customer Churn Prediction Integrating Clustering and Classification Algorithms," Appl. Sci., vol. 12, no. 18, Art. no. 18, Jan. 2022, doi: 10.3390/app12189355.

[9] N. Edwine, W. Wang, W. Song, and D. Ssebuggwawo, "Detecting the Risk of Customer Churn in Telecom Sector: A Comparative Study," Math. Probl. Eng., vol. 2022, p. e8534739, Jul. 2022, doi: 10.1155/2022/8534739.

[10] G. Jiao and H. Xu, "Analysis and Comparison of Forecasting Algorithms for Telecom Customer Churn," J. Phys. Conf. Ser., vol. 1881, no. 3, p. 032061, Apr. 2021, doi: 10.1088/1742-6596/1881/3/032061.

[11] S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, "Customer churn prediction in telecom sector using machine learning techniques," Results Control Optim., vol. 14, p. 100342, Mar. 2024, doi: 10.1016/j.rico.2023.100342.

[12] S. O. Abdulsalam, J. F. Ajao, B. F. Balogun, and M. O. Arowolo, "A Churn Prediction System for Telecommunication Company Using Random Forest and Convolution Neural Network Algorithms," EAI Endorsed Trans. Mob. Commun. Appl., vol. 7, no. 21, Jul. 2022, Accessed: Mar. 30, 2024. https://eudl.eu/doi/10.4108/eetmca.v6i21.2181

[13] L. F. Khalid, A. Mohsin Abdulazeez, D. Q. Zeebaree, F. Y. H. Ahmed, and D. A. Zebari, "Customer Churn Prediction in Telecommunications Industry Based on Data Mining," in 2021 IEEE Symposium on Industrial Electronics & Applications (ISIEA), Jul. 2021, pp. 1–6. doi: 10.1109/ISIEA51897.2021.9509988.

[14] H. Jain, A. Khunteta, and S. P. Shrivastav, "Telecom churn prediction using seven machine learning experiments integrating features engineering and normalization," 2021, Accessed: Apr. 08, 2024. https://www.researchsquare.com/article/rs-239201/latest

[15] D. D. Adhikary and D. Gupta, "Applying over 100 classifiers for churn prediction in telecom companies," Multimed. Tools Appl., vol. 80, no. 28, pp. 35123–35144, Nov. 2021, doi: 10.1007/s11042-020-09658-z.

[16] P. Lalwani, M. K. Mishra, J. S. Chadha, and P. Sethi, "Customer churn prediction system: a machine learning approach," Computing, vol. 104, no. 2, pp. 271–294, Feb. 2022, doi: 10.1007/s00607-021-00908-y.

[17] R. P. Sari, F. Febriyanto, and A. C. Adi, "Analysis Implementation of the Ensemble Algorithm in Predicting Customer Churn in Telco Data: A Comparative Study," Informatica, vol. 47, no. 7, Art. no. 7, Jul. 2023, doi: 10.31449/inf.v47i7.4797.

[18] M. Imani, Z. Ghaderpour, and M. Joudaki, "The Impact of SMOTE and ADASYN on Random Forests and Advanced Gradient Boosting Techniques in Telecom Customer Churn Prediction," Mar. 05, 2024, Preprints: 2024030213. doi: 10.20944/preprints202403.0213.v1.

[19] F. E. Usman-Hamza et al., "Intelligent Decision Forest Models for Customer Churn Prediction," Appl. Sci., vol. 12, no. 16, Art. no. 16, Jan. 2022, doi: 10.3390/app12168270.

[20] M. E. Saleh, N. Abd-Alsabour "On the impact of various combinations of preprocessing steps on customer churn prediction," unpublished.

[21] B. Ramosaj and M. Pauly, "Predicting missing values: a comparative study on non-parametric approaches for imputation," Comput. Stat., vol. 34, no. 4, pp. 1741–1764, Dec. 2019, doi: 10.1007/s00180-019-00900-3.

[22] Y. Fareniuk, T. Zatonatska, O. Dluhopolskyi, and O. Kovalenko, "Customer churn prediction model: a case of the telecommunication market," ECONOMICS, vol. 10, no. 2, pp. 109–130, Dec. 2022.

[23] W. H. Khoh, Y. H. Pang, S. Y. Ooi, L.-Y.-K. Wang, and Q. W. Poh, "Predictive Churn Modeling for Sustainable Business in the Telecommunication Industry: Optimized Weighted Ensemble Machine Learning," Sustainability, vol. 15, no. 11, Art. no. 11, Jan. 2023, doi: 10.3390/su15118631.

[24] D. Singh and B. Singh, "Investigating the impact of data normalization on classification performance," Appl. Soft Comput., vol. 97, p. 105524, Dec. 2020, doi: 10.1016/j.asoc.2019.105524.

[25] N. N. Y, T. V. Ly, and D. V. T. Son, "Churn prediction in telecommunication industry using kernel Support Vector Machines," PLOS ONE, vol. 17, no. 5, p. e0267935, May 2022, doi: 10.1371/journal.pone.0267935.

[26] S. Wael Fujo, S. Subramanian, and M. A. Khder, "Customer Churn Prediction in Telecommunication Industry Using Deep Learning," Inf. Sci. Lett., vol. 11, no. 1, Dec. 2021, [Online]. Available: https://digitalcommons.aaru.edu.jo/isl/vol11/iss1/24

Cyber Security Risk Assessment Framework for Cloud Customer and Service Provider

N. Sujata Kumari[1]*, Naresh Vurukonda[2]
Research Scholar, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India, 522502[1]
Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur, Andhra Pradesh, India, 522502[2]

***Abstract*—The rapid development of cloud computing demands an effective cybersecurity framework for protecting the sensitive information of the infrastructure. Currently, many organizations depend on cloud services for their operation, increasing the risk of cybersecurity. Hence, an intelligent risk assessment mechanism is significant for detecting and mitigating the cybersecurity threats associated with cloud environments. Although various risk assessment methods were developed in the past, they lack the efficiency to handle the dynamic and evolving nature of threats. In this study, we proposed an innovative framework for cybersecurity risk assessment in cloud customers and service providers. Initially, the historical cloud customer and service provider database was collected and fed into the system. The collected dataset contains historical security risks, network traffic, system behavior, etc., and the accumulated dataset was pre-processed to improve the quality of the dataset. The data pre-processing steps not only ensure quality but also transform the dataset into appropriate format for subsequent analysis. Further, a risk assessment module was created using the combination of deep recurrent neural network with krill herd optimization (DRNN-KHO) algorithm. In this module, the DRNN was trained using the pre-processed database to learn the pattern and interconnection between normal and abnormal network traffic. Subsequently, the KHO refines the DRNN parameters in its training phase, increasing the efficiency of risk assessment. This integrated module ensures adaptability to the system, leading to accurate prediction of evolving security threats. Then, a secure data exchange protocol was created for secure transmission between cloud customer and service provider. This protocol is designed by integrating artificial bee colony optimization with the elliptic curve cryptography (ABC-ECC). Thus, this collaborative framework ensures security in the cloud customer and service providers.**

Keywords*—Deep recurrent neural network; krill herd optimization; artificial bee colony optimization; elliptic curve cryptography

I. Introduction

In recent years, the significance of cloud computing (CC) has been rising, and it has received a growing interest in both business and scientific organizations [1]. The report from the National Institute of Standards and Technology (NIST) stated that the CC model offers configurable resources such as storage, networks, servers, applications, etc., in the most convenient and on-demand manner [2]. The CC delivers these resources with various kinds of service provider interaction by following a simple Pay-As-You-Go (PAYG) model. In the PAYG model, the cloud service consumers (CSC) demand the computing services for their business demands [3]. Consequently, the cloud service provider (CSP) delivers the demanded computing services to the CSC. In the CC model, the CSC has to pay only for the services they have utilized, reducing the cost and making the process more convenient [4]. In addition to this, the CC model has several advantages such as greater scalability, high availability, enhanced flexibility, well-documented, excellent reliability, etc., [5]. These advantages bring more benefits to the business organizations. In accordance with the report by Forbes, the CC has imposed a greater growth, and the CC market reached $411 Billion in 2020 [6]. Furthermore, the survey made by the LogicMonitor in 2020 illustrated that the cloud services landscape has gained 83% interest of the enterprises [7].

Although the CC brings more benefits to the organizations, it offers several issues. One among the issues is the cybersecurity threats, which degrades confidence and feasibility of the CC model [8]. From the perspective of CSC, the primary reason for confidence issues on CC is because of its multitenancy nature, sensitive data transformation, critical infrastructure and applications [9]. On the other hand, security is the primary issue from the perspective of CSP, which is mainly because of the cloud model's complexity. This complexity arises from the management of diverse security controls and measures [10]. Also, the security threats arise while transferring the data in the CC. These security threats induce distrust and fears in the business organizations, making them to redefine their decisions in adapting the CC model [11]. Despite how strongly the CC is secured, the business units suffer from trust issues on cloud models and remain uncertain about its economic feasibility [12].

While it's unrealistic to guarantee a zero-risk service, implementing an efficient security risk management algorithm can significantly enhance organizations' confidence in the CC model [13]. This risk management technique empowers organizations to make informed decisions about adopting this emerging technology [14]. The conventional risk management algorithms do not fit into the CC model because of the inherent assumptions made by them [15]. In recent decades, artificial intelligence (AI) techniques are widely used in different research fields because of its capacity to decide like humans [16]. The AI evolution paved a way for handling the security risks in the CC model. The utilization of AI techniques such as deep learning (DL) and machine learning (ML) techniques

provides an effective way of identifying the potential security risks in CC [17]. Although they offer better security compared to the conventional models, they face certain challenges like overfitting, large computational time, huge computational resource demand, lack of scalability, less adaptability, lower reliability, etc., [18]. Moreover, they cannot provide security to the sensitive data during transmission in the CC model [19]. This demands an optimal and reliable security risk management algorithm, which must potentially identify the cybersecurity risks, and provide greater protection to data in the cloud model [20]. By evaluating risks across different dimensions such as data sensitivity, regulatory compliance, access controls, and service-level agreements (SLAs), the framework enables both customers and providers to identify weak points and implement appropriate mitigation strategies. To resolve the above issues, we proposed a collaborative security risk assessment framework combining deep learning, meta-heuristic optimizations, and cryptographic algorithms. The main contributions of the work are described as follows.

- This study proposes a collaborative security risk assessment framework by integrating the efficiencies of deep learning, meta-heuristic optimizations, and cryptographic algorithms for effectively identifying the cybersecurity threats and securely transmitting the data in the cloud model.
- The developed framework creates a cybersecurity threat detection module named DRNN-KHO by integrating Krill Herd optimization and Deep Recurrent Neural Network, which is trained using the cloud security database to identify the normal and malicious activities in the cloud model.
- The study also developed a hybrid ABC-ECC model, which is designed by incorporating artificial bee colony optimization into the elliptic curve cryptographic algorithm for securely transmitting the sensitive data in the cloud model.
- The presented collaborative mechanism was implemented in the Python language, and the results are assessed and validated using metrics like attack detection accuracy, data confidential rate, computational time, encryption time, decryption time, etc.

The enduring sections of the article are organized as follows: Section II illustrates the literature survey, Section III depicts the system model and problem statement, Section IV explains the proposed algorithm, Section V analyzes the study results, and Section VI provides the study conclusion.

II. Related Works

A. Few Recent Studies Related to the Developed Framework are Described Below:

Lav Gupta et al. [21] presented a deep hierarchical stacked neural system for precise prediction of anomaly activities in the cloud environment. This study aims to resolve the security problems associated with the healthcare applications, as the malicious agents threaten the patient's life and health by changing their medical data flow into the healthcare networks. The experimental results depict that this DL model minimized the training time by 26.2%, and achieved better convergence. Moreover, this algorithm predicts the malicious agents with an average accuracy ranging from 93% to 95%. However, this framework offers limited generalization and is vulnerable to adversarial attacks.

Reem Al Saleh et al. [22] proposed a reputation-based trust evaluation algorithm by integrating Net Brand Reputation (NBR) with the deep learning approach for the cloud market. The DL algorithm used in this approach is CBiLSTM, which is the combination of Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory (BiLSTM). The implementation outcomes manifest that this algorithm obtained performances of 96.7% accuracy, 97% f-measure, 96.5% recall and 97.4% recall. In addition to this, this model consumed a minimum training time of 519ms. Although this approach offered improved performances, it is less scalable and interpretable.

R. Denis and P. Madhubala [23] developed a hybrid encryption algorithm for ensuring confidentiality in the cloud computing environment. This hybrid strategy combines Advanced Encryption Standard (AES) and Rivest-Shamir-Adleman for secure data transmission. Also, an adaptive genetic algorithm was applied in the proposed work to ensure adaptability to the changing CC environment. This adaptiveness enables the system to respond to the emerging threats in the CC model. Finally, the simulation outcomes validate that this framework obtained data confidential rate of 0.95, and minimum running time of 5.4s. However, it faces complexity in managing keys, resulting in increased computational overhead.

Fursan Thabit et al. [24] developed an innovative lightweight cryptographic algorithm for improving the data security in the CC model. This approach is designed based on feistal and substitution permutation architectural approaches to enhance the encryption complexity. In addition, this approach obtained Shannon's theory of diffusion and confusion by including logical functions like XOR, shifting, XNOR, and swapping. The experimental outcomes suggest that this algorithm achieved a strong security level and less execution time compared to the existing cryptographic techniques. However, it is less adaptable for emerging threats.

P. Chinnasamy et al. [25] developed an efficient data security algorithm using hybrid cryptographic techniques for the CC model. This hybrid algorithm combines Elliptic curve cryptography (ECC) and Blowfish to provide a high level of data transfer and storage security. The main objective of this algorithm is to address the security challenges like confidentiality, availability, and integrity. This work was validated in the medical cloud environment, and the experimental data depicts that it achieved higher security and confidentiality than the symmetric and asymmetric algorithms.

Shaopeng Guan et al. [26] designed a distinct framework named Hadoop-assisted big data storage scheme for the CC environment. The primary concern of this work is to resolve the issues associated with the single encryption algorithm like low encryption efficiency, and unreliable data storage. The study utilizes the ECC encryption to encrypt the original data, which

ensures protection against security threats during transmission. Then, a homomorphic encryption was employed to ensure data integrity in the system. The experimental outcomes manifest that this algorithm improves the storage efficiency by 27.6%. Moreover, it offers greater protection to the sensitive information against cyberthreats. However, this framework cannot handle large volumes of data. Table I presents the literature survey.

TABLE I. LITERATURE SURVEY

Authors	Technique	Results	Merits	Demerits
Lav Gupta et al. [21]	Deep hierarchical stacked neural system	Reduced training time by 26.2%, obtained an average accuracy between 93% to 95%	High convergence, less training time, and greater detection accuracy	Limited generalization and vulnerable to adversarial attacks
Reem Al Saleh et al. [22]	NBR with CBiLSTM	Accuracy-96.7%, f-measure-97%, recall-96.5%, and recall-97.4%	High malicious data detection performances, and minimum training time	Less scalable and interpretable
R. Denis and P. Madhubala [23]	Hybrid encryption scheme (AES-RSA)	Data confidential rate-0.95. running time-5.4s	Provides greater protection to data during transmission	Key management complexity, and increased computational overhead
Fursan Thabit et al. [24]	Lightweight cryptographic algorithm	Execution time-6.4s	Strong security level and less execution time	Less adaptable for emerging threats
P. Chinnasamy et al. [25]	Hybrid cryptographic algorithm combining ECC and Blowfish	Data confidentiality-96%, security level-97%	Higher security and confidentiality	Highly complex, and resource intensive
Shaopeng Guan et al. [26]	Hadoop-assisted big data storage scheme	Improves cloud storage efficiency by 27.6%	Protects the sensitive data from cyberthreats	Cannot handle large volumes of data

III. SYSTEM MODEL WITH PROBLEM STATEMENT

Cloud computing is the process of delivering computing services such as databases, storage, servers, networking, analytics, software, and intelligence over the internet for offering flexibility and reliability, and fostering faster innovation. Although the CC model offers various benefits to the business institutions, it is prone to security challenges. This demands an effective security risk assessment framework to ensure confidentiality, integrity, and availability of the CC model [27]. A system security risk assessment model includes data collection module, preprocessing component, and risk assessment module (quantitative, statistical, DL or ML models). Fig. 1 presents the system model. The data acquisition component contains various data like network traffic, logs, system behavior, historical security risks, etc. In the preprocessing component, the collected database undergoes preprocessing steps to improve its quality for subsequent analysis. After preprocessing, a risk assessment module was created using either DL or ML models. These models are trained using the database to differentiate the patterns between the normal and malicious behavior. After the training process, it is validated for unseen data or real-time scenarios for risk evaluation.

Although these risk assessment models provide automatic prediction of malicious behavior, they often rely on threshold settings to differentiate normal and anomalies behavior. This may lead to incorrect predictions. Moreover, the conventional models require more computational resources for intensive training, making the system costly and complex. Also, they face challenges in providing scalability to handle large-scale cloud environments. In addition to this, the traditional security risk analysis algorithms can adapt to the emerging and evolving cyber threats, making it less reliable and effective for real-world scenarios. Moreover, only few studies concentrated on implementing the security measures on the CC model to ensure protection to the sensitive data during transmission and storage. To address the issues with the conventional security risk assessment techniques, and to bridge the research gap, we developed a collaborative framework, which uses deep learning and meta-heuristic optimization algorithms to precisely identify the security threats, and implements an optimized cryptographic algorithm for secure transmission of information in the cloud.

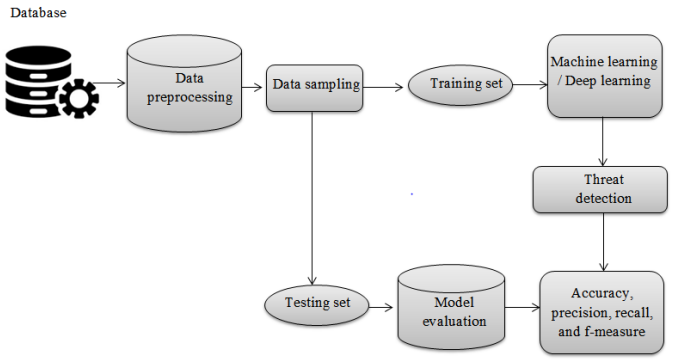


Fig. 1. System model.

IV. PROPOSED FRAMEWORK FOR SECURITY RISK ASSESSMENT

A novel collaborative security framework was proposed in this article by combining the benefits of deep learning, meta-heuristic optimizations, and cryptographic algorithms. The novelty of the study lies in the seamless and effective integration of these algorithms for ensuring security to the CC

model. The primary objective of the model is to identify the security threats/risks associated with the CC model, and to ensure secure data communication in the cloud. Firstly, a database containing network traffic, historical security risks, etc., was collected and imported into the system. Secondly, the accumulated database was preprocessed to handle the missing values, outliers, errors, etc. This process enhances the quality of the database, and makes it suitable for subsequent analysis. Then, an attack prediction module was created by integrating deep recurrent neural network and krill herd optimization.

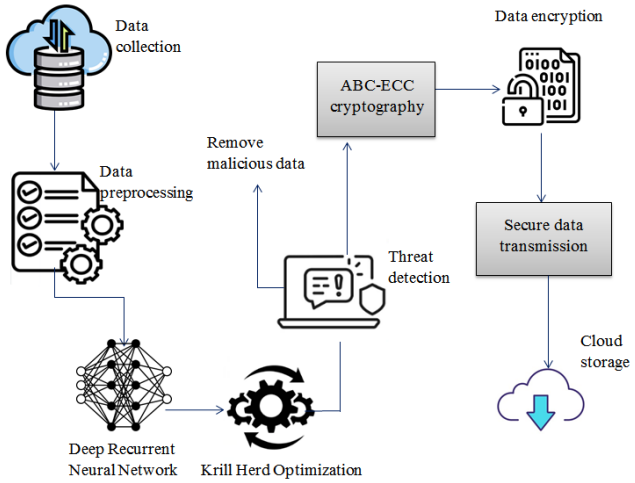


Fig. 2. Architecture of the proposed framework.

In this module, the DRNN was trained using the collected database to distinguish the patterns between the normal and malicious behavior, consequently we applied krill herd optimization for hyperparameter tuning and optimization. This combination enables the system to continuously learn the patterns and correlations between the normal and malicious traffic, ensuring adaptability to the emerging threats. After threat detection, the detected threats are eliminated from the system to offer a higher security level. Further, an optimized cryptographic algorithm was proposed to ensure security against the cyberthreats during the transmission process. The developed algorithm incorporates the efficiency of ABC into the ECC algorithm for optimally selecting the parameters for key generation, and distribution. This algorithm encrypts the data before transmission, making it unreadable for third parties. After encryption, the encrypted data is transferred to the cloud for storage, where the decryption step can be performed by the authenticated authorities and decode the data. Fig. 2 presents the architecture of the proposed strategy.

A. Data Collection

The proposed study commences with the collection of dataset relevant to cloud computing. The collected database may contain information like network traffic, logs, system characteristics, historical security risks, etc. The presented study utilized the publicly available security database named "DDoS SDN", and it is accessible at https://www.kaggle.com/datasets/aikenkazin/ddos-sdn-dataset. This database contains 104,345 instances and 23 features, and it is developed for categorizing network traffic as either traffic or malicious within a cloud environment. The dataset is structured with a target variable named "label," containing binary values (0 or 1). Here 1 indicates malicious, and 0 defines benign. Out of the 23 features, 3 are categorical features, and 20 are numerical features, providing a wide range of information relevant to network traffic investigation. The size of the dataset is 12.56MB and it is available in csv format.

B. Preprocessing

Data preprocessing is an important step, which plays a significant role in improving the quality of the dataset. It includes steps like cleaning, transforming, scaling, etc., to make the database ready and reliable for further analysis. In the proposed work, we applied steps like handling missing values, feature scaling, and handling class imbalance to the raw database for making it effective for subsequent processes. Handling missing values indicates the process of detecting the null values in the database and replacing them with values determined using an imputation algorithm. In imputation, the missing value was replaced with the mean of the non-missing values of the dataset. The imputation process is mathematically represented in Eq. (1).

$$M_n = \frac{\sum_{i=1}^{m} D_i}{m} \tag{1}$$

Where M_n indicates the mean, m indicates the number of samples, and D indicates non-missing value. Then, feature scaling was performed to ensure that all the features in the dataset contribute equally to the analysis. This step prevents the dominance of a single feature in threat detection. Here, we applied a min-max scaling algorithm to rescale all features to a fixed range (0 and 1), and it is expressed in Eq. (2).

$$S_D = \frac{D - D_{mn}}{D_{mx} - D_{mn}} \tag{2}$$

Where S_D indicates the scaled value of the feature, D_{mn} denotes the maximum value of the feature, and D_{mx} represents the minimum value of the feature. Then, the database was checked for class imbalance. If there is imbalance in the class, then oversampling was done to address that. Finally, the database was split into training and testing to evaluate the performance of the proposed threat detection algorithm. The introduction of these steps not only enhances the dataset quality but also increases the speed of subsequent analysis, leading to enhanced and timely assessment of cybersecurity threats. This preprocessed dataset was fed into the DRNN-KHO for cybersecurity risk assessment.

C. DRNN-KHO for Threat Detection

Threat detection defines the process of classifying the normal and malicious network traffic. In the developed work, we proposed an innovative threat detection model by combining the efficiency of deep recurrent neural network, and krill herd optimization algorithm. Here, the DRNN acts as a classification module categorizing normal and malicious data,

while the KHO intends to refine its parameters to improve the overall threat detection process. The DRNN is an artificial neural network, which is mainly used in sequential data processing. The architecture of DRNN is similar to the conventional RNN structure with an additional dense layer. Generally, there are two different RNN architectures namely: Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU) [28]. In the proposed work, we used a LSTM model for predicting the cybersecurity threats. The proposed DRNN contains input layer, recurrent layer (LSTM), dense layer, and output layer. The input layer of the system accepts the preprocessed database as input. This layer converts into a suitable format for further analysis, and forwards into the recurrent layer, where the model intends to learn the patterns and long-range dependencies within the data for identifying the normal and malicious data. The LSTM layer includes input gate, forget gate, candidate cell state, hidden state, cell state, and output gate. These cells and gates process the input sequence and capture the complex intricate patterns and correlations within the data, and they are expressed in Eq. (3), (4), (5), (6) and (7).

$$f_t = \sigma\left(W_f .\left[H_{t-1}.D_t\right] + B_f\right) \tag{3}$$

$$i_t = \sigma\left(W_i .\left[H_{t-1}.D_t\right] + B_i\right) \tag{4}$$

$$o_t = \sigma\left(W_o .\left[H_{t-1}.D_t\right] + B_o\right) \tag{5}$$

$$c_t' = \tanh\left(W_c .\left[H_{t-1}.D_t\right] + B_c\right) \tag{6}$$

$$H_t = H_t * \tanh\left(c_t\right) \tag{7}$$

Where f_t defines the forget gate, i_t denotes the input gate, c_t' indicates the candidate cell state, c_t represents cell state, o_t refers to the output gate, σ represents the sigmoid activation function, $\tanh$ denotes the hyperbolic tangent activation function. Further, W defines the weight matrices, and B represents the bias vectors. D_t indicates the input sequence at time step t, and H_{t-1} represents the hidden state at the previous time step. These gates perform certain functions to capture the long-term dependencies and intricate patterns within the sequential data for detecting the threats. The outcome of this layer is forwarded into the dense layer, which is typically a fully connected layer. In this layer, each neuron is interconnected with every neuron in the recurrent and output layer, making the system understand the difference between normal and malicious data. The dense layer is presented in Eq. (8).

$$ds_t = \sigma\left(W_{dh} .H_t + B_d\right) \tag{8}$$

Where W_{dh} defines the weight interconnecting the LSTM layer and dense layer, B_d denotes the bias vector, and σ represents the activation function. These learned patterns and extracted features are fed into the output layer, which produces the probability of being normal or malicious data. The resultant of the output layer is represented in Eq. (9).

$$y_t = \left(W_{ot} .ds_t + B_{ot}\right) \tag{9}$$

Where y_t defines the output at time step t, which indicates the probability of the input being normal or malicious. If the output value is 0, the system indicates it as “normal,” and “malicious.” defines the bias vector and indicates the activation function, which converts the learned features into probability values. At each iteration, this system learns to understand the difference between the normal and malicious data. The loss incurred by the model is presented in Eq. (10).

$$L_{oss} = -\frac{1}{m}\sum_{t=1}^{m}\left[a_t \log\left(y_t\right) + \left(1-a_t\right)\log\left(1-y_t\right)\right] \tag{10}$$

Where L_{oss} indicates the loss, m defines the number of training samples, and a_t represents the actual outcome. This loss function was minimized by adjusting its parameters like weight and bias. In the proposed work, the training loss incurred by the DRNN is optimized using the KHO algorithm by continuously iteratively adjusting its hyperparameters. The KHO is a nature-inspired algorithm developed based on the herding behavior of krills to solve the optimization problems. Here, the objective function of KHO is to maximize the prediction accuracy of DRNN by refining and fine-tuning its hyperparameters to its optimal range. This optimization algorithm was developed based on three main actions: (1) movement induced by other krill individuals, (2) foraging characteristics, and (3) random diffusion. The optimization process begins with the random initialization of the krill population in the search space [29]. Each krill individual in the population defines the hyperparameter values. After initialization, the fitness value of each individual was determined based on the defined objective function, which is represented in Eq. (11).

$$O_{bj} = \max\left(Ac\right) \tag{11}$$

Where O_{bj} indicates the objective function, and Ac denotes the threat detection accuracy of the DRNN model. The fitness of the parameter set will be high if the DRNN model achieved high threat detection accuracy and vice versa. After fitness evaluation, the parameter values are updated following the three actions mentioned above. These steps enable the system to explore the population space and find the optimal value. The parameter updation is mathematically expressed in Eq. (12).

$$p_v\left(t+1\right) = p_v\left(t\right) + \Delta t \frac{dp_v}{dt} \tag{12}$$

Where $p_v\left(t\right)$ indicates the parameter value at time t, $p_v\left(t+1\right)$ defines the updated parameters, and $\Delta t \frac{dp_v}{dt}$

represents the scale factor. Then, the fitness solution was determined for updated parameter sets. Finally, the updated fitness and old fitness was compared. If the updated fitness is greater than the old fitness, the updated parameter values are used for training. This process continues until reaching the maximum iteration count. Thus, the hybrid mechanism detects and classifies the normal and malicious data by learning the patterns within the data effectively. After threat detection, the malicious data is removed from the network for ensuring security and privacy to the data.

D. ABC-ECC for Secure Data Storage

After threat detection, we developed a hybrid cryptographic algorithm combining the efficiency of artificial bee colony optimization with the elliptic curve cryptography (ABC-ECC) for secure data transmission and storage in the cloud. Here, before transmitting the data to the cloud, we encrypt the data using the proposed ABC-ECC model, which prevents the intrusion of third parties during transmission. Elliptic curve cryptography is a public-key cryptographic algorithm developed based on the concept of elliptic curves over finite fields [30]. The elliptic curve is mathematically defined in Eq. (13).

$$y^2 + xy = x^3 + ux + v \tag{13}$$

Where x and y defines the coordinates of the curve, u indicates the curve slope, and v represents the constant term indicating the curve displacement. The important property of this curve is that we can define a rule for adding two points U and V on the curve to determine a third point W, which is also on the curve. This defined rule agrees with the normal addition properties. This forms the basis for operations like scalar multiplication, which are at the heart of ECC's security and efficiency. Consider a case where Alice and Bob want to share the information. Firstly, Alice selects a random integer A_r as its private key, and it is denoted as. Consequently, Bob also selects a random secret integer B_r as his private key, which is defined as. Before key generation and transmission, both Bob and Alice agree upon a non-secret (elliptic curve) and fixed curve point F_p (non-secret). Then, both Alice and Bob calculate their public key, which is represented in Eq. (14), and (15).

$$A_p = F_p A_r \tag{14}$$

$$B_p = F_p B_r \tag{15}$$

Where A_p and B_p define the public keys of Alice and Bob. They compute the public keys by multiplying their private keys with the fixed curve point. These keys are used for performing the encryption operations. If Alice wants to send the data to Bob, it computes a shared secret using her private key and Bob's public key. Subsequently, Bob also computes a shared secret using its private key and Alice's public key. After generating a shared secret, it is used to establish a secure communication channel between Alice and Bob, where both parties use their shared secret as the key for encrypting and decrypting messages exchanged between them. Although the ECC approach is more reliable and effective than conventional cryptographic techniques like Rivest-Shamir-Adleman, its efficiency and security relies on the selection of fixed curve points. Hence, we applied the Artificial Bee Colony algorithm for selecting the fixed curve point from the defined elliptic curve for improving the efficiency and security of the conventional ECC approach. The ABC is a meta-heuristic optimization algorithm developed based on the intelligent foraging characteristics of honey bee swarms. The ABC model contains three bee groups namely: onlookers, employed, and scouts. In the proposed work, we employed the employed bee phase for finding the optimal fixed curve point [31]. The initialization of the ABC population was expressed in Eq. (16).

$$F_{pi} = l_b + rand(0,1) * (u_b - l_b) \tag{16}$$

Where l_b and u_b defines the lower and upper bound of the parameter, respectively. After initialization, the employed bees search for new food sources. Here, the food sources represent the optimal fixed curve point. First they find the neighboring food source and determine its profitability (fitness solution). In case of fixed curve point selection, the fitness indicates its ability to improve the efficiency and security of ECC, which is mathematically represented in Eq. (17).

$$F'_{pi} = F_{pi} + \phi_{pi}(F_{pi} - r_{pi}) \tag{17}$$

Where r_{pi} indicates the randomly selected fixed curve point. Further, the bees explore the population to find the best solution. Further, the fitness solution was again evaluated to select the optimal solution. This process is an iterative process, and at each iteration, the ABC optimization finds the optimal fixed curve point; thereby improving the efficiency and security in data encryption. The encrypted data is then transferred into the cloud for storage and further analysis. Thus, the proposed collaborative framework detects the threats and offers security within the cloud network. The working of the proposed algorithm is described in pseudocode format in Algorithm 1.

Algorithm 1
Start {
Initialize the DDoS database;
Data preprocessing:
1. Handling missing values//*Apply imputation*
2. Feature scaling; //*Apply min-max scaling*
3. Handling class imbalance;// *Perform oversampling*
Data splitting;
Threat detection:
Define DRNN layers;
Initialize epoch count, weights;
Model training:
For each training epoch:
Extract the features;
Learn the patterns and correlations;
Determine output probability;

if (y_t =0) **Normal**;
else **Malicious**;
End for;
Eliminate threats;
Determine loss function;
Optimization:
Initialize KHO parameters (population size, maximum iteration);
Initialize the population;
Define objective function;
For each iteration:
Determine fitness solution;
Update parameter set;
Evaluate fitness for updated parameter sets;
if (new fitness> old fitness) **return updated parameters;**
else **return old parameters;**
End for;
Secure data transmission:
Initialize ECC parameters;
Select fixed curve point;
Select random integers;
Generate public keys;
Perform data encryption;
ABC optimization:
Initialize ABC population, maximum iteration;
Define objective;
For each iteration:
Determine fitness;
Exploration;
End for;
Return optimal fixed curve point;
Model evaluation;
}Stop

V. Results and Discussion

A collaborative security framework was proposed for assessing the cybersecurity risks within the cloud network. The proposed framework was modeled in the Pycharm tool, running in Windows 10 Operating system. The proposed framework was trained and validated using the DDoS dataset, and the results are determined in terms of accuracy, recall, f1-score, mean absolute error (MAE), mean square error (MSE), etc.

A. Training and Testing Performances

In this module, we discuss the training and testing performances of the proposed algorithm. Firstly, the database was split for training and testing purposes, and the performances are assessed as accuracy and loss for different learning rates (70% and 80%). The accuracy metric measures how quickly the proposed algorithm learns the patterns of normal and malicious traffic. The loss metric measures the deviation between the actual and predicted outcomes. Fig. 3 presents the training and testing performance of the developed algorithm for 70% learning rate. The training accuracy indicates how effectively the designed model fits into the training sequence and learns the patterns and correlations for distinguishing the normal and malicious traffic. The developed algorithm achieved greater training accuracy of 0.96, and 0.99 for 70% and 80% learning rates. This validates that the designed approach fastly learns and understands the pattern difference between the normal and threats.

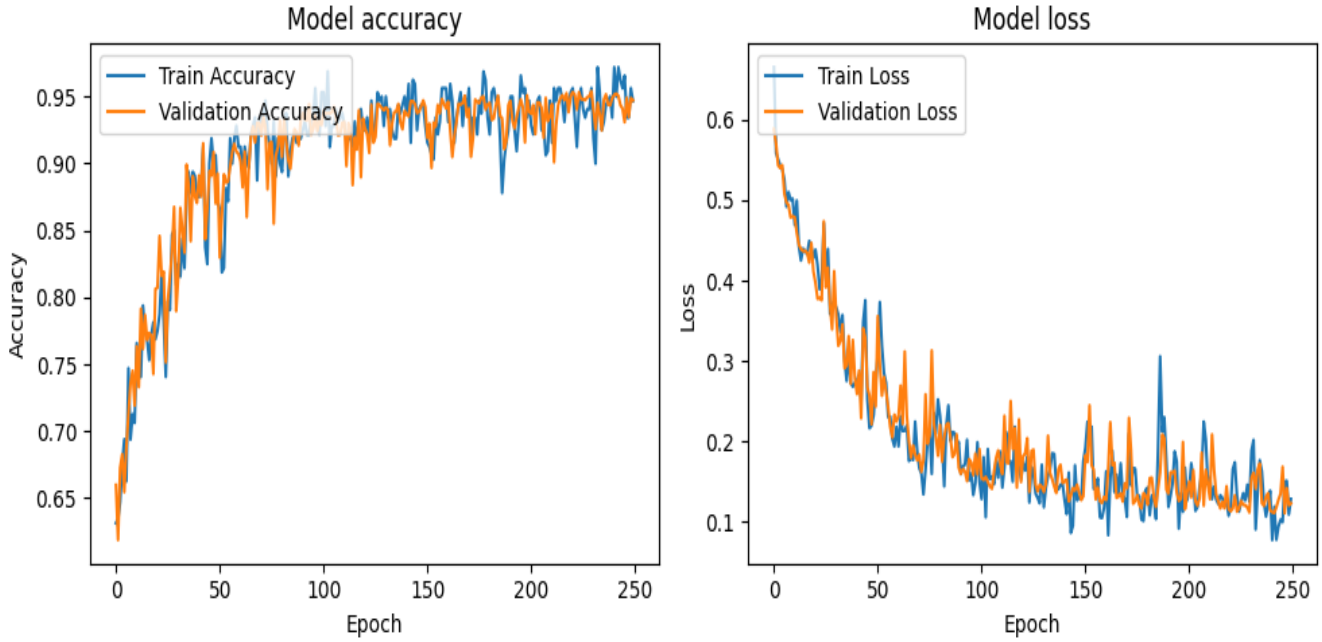


Fig. 3. Training and Testing performance of the proposed model for 70% learning rate.

Consequently, the testing accuracy was evaluated for 70% and 80% learning rates. The testing accuracy defines how effectively the developed model performs predictions on unseen data. The proposed algorithm achieved greater testing accuracy of 0.95, and 0.98 for 70% and 80% learning rates. These improved testing accuracy highlights that the proposed algorithm generalizes well on the new data, making it effective and reliable for real-time threat detection. Also, it is observed that the designed model obtained greater accuracy for 80% learning rate, which manifests that the model's performance increases with higher learning rate.

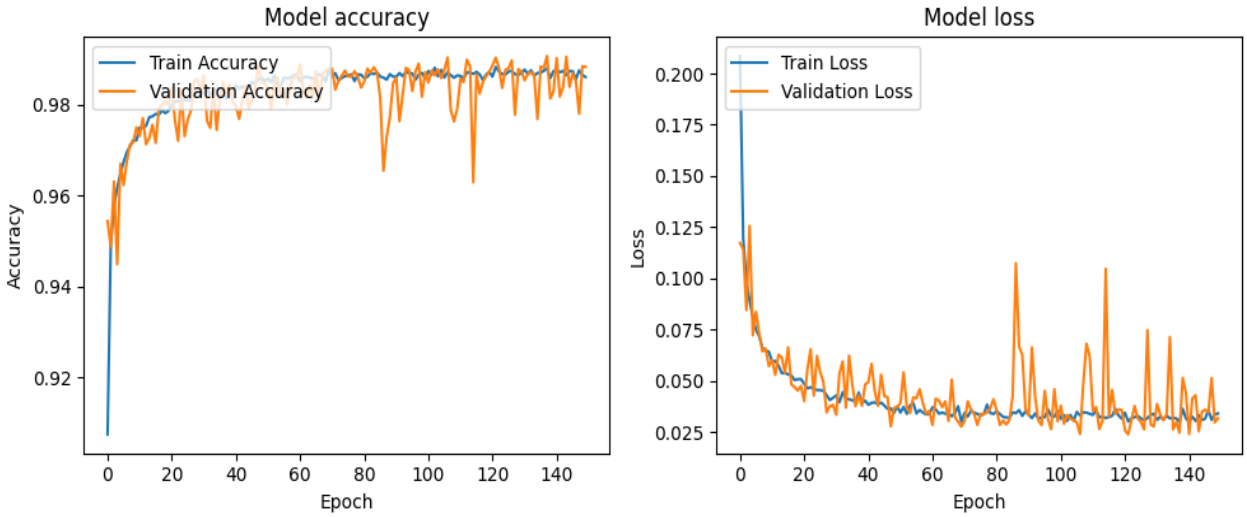
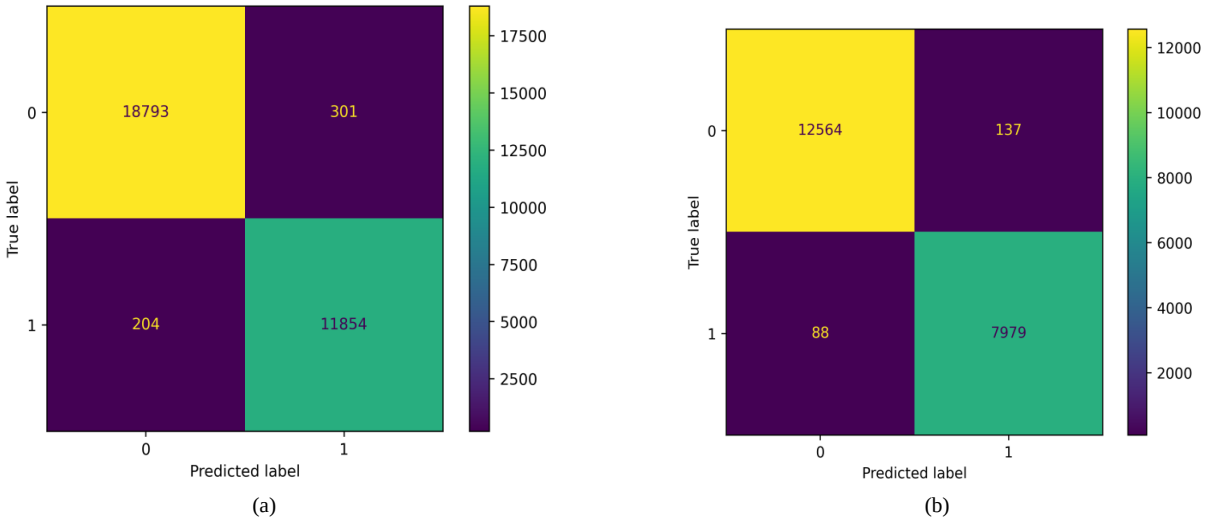


Fig. 4. Training and testing performance of the proposed model for 80% learning rate.

The training loss quantifies the misclassification made by the proposed algorithm within the training sequence. It measures the deviation between the actual and the predicted results in the training sequence. The developed model obtained lower training loss of 0.1, and 0.03 for 70% and 80% learning rates. On the other hand, the testing loss was determined for 70% and 80% learning rates. The testing accuracy measures the model's generalization to unseen data by quantifying the difference between the actual and the predicted outcomes. The designed approach attained minimum loss of 0.11, and 0.4 for 70% and 80% learning rates. Fig. 4 presents the training and testing performance incurred by the developed algorithm for 80% learning. From this intensive model evaluation, it is clear that the proposed algorithm achieved better performances for increased learning rate. Also, it achieved greater accuracy and minimum loss in both train and test phases, highlighting its generalization ability and capacity to prevent the overfitting problem. This illustrates that the proposed technique effectively learns the patterns within the database, and predicts the normal and malicious data accurately.



(a) (b)

Fig. 5. Confusion matrix: (a) 70% learning rate, (b) 80% learning rate.

The confusion matrix is a performance evaluation tool deployed in deep learning to determine its efficiency in classification tasks. It is a table, which visualizes the classification performance of the model by summarizing the counts of correct and incorrect classifications made by the system on the sequence of data for which the actual values are known. Fig. 5 (a, b) presents the confusion matrix obtained for 70% and 80% learning rates. The cell in the matrix indicates the count of instances where the predicted class correlates with the actual class. This matrix is divided into four sections namely, true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN). TP indicates the instances where the malicious traffic is correctly predicted, while TN represents the instances where the normal traffic is correctly predicted. On the other hand, FP defines the scenario when the normal traffic is incorrectly predicted as malicious, while FN denotes the scenario when the malicious traffic is incorrectly predicted as normal. By examining the confusion matrix, we determine the efficiency of the proposed model in threat detection and classification.

1) Cryptographic algorithm performances: In this module, the performance incurred by the proposed cryptographic algorithm (ABC-ECC) was examined in terms of encryption time, decryption time, success rate, and turnaround time. These performances are assessed by increasing the data size from 1 to 100000. Fig. 6 presents the assessment of cryptographic algorithm performances. Fig. 6 (a) presents the encryption time analysis. The encryption time measures the time taken by the proposed system for performing the encryption operation. The proposed algorithm obtained an average encryption time of 0.055, which illustrates that it consumes less time for encoding the dataset. Consequently, the decryption time of the system was determined and it is depicted in Fig. 6 (b). The proposed technique obtained an average decryption time of 0.03, which highlights that the system consumes minimum time for performing the decryption task.

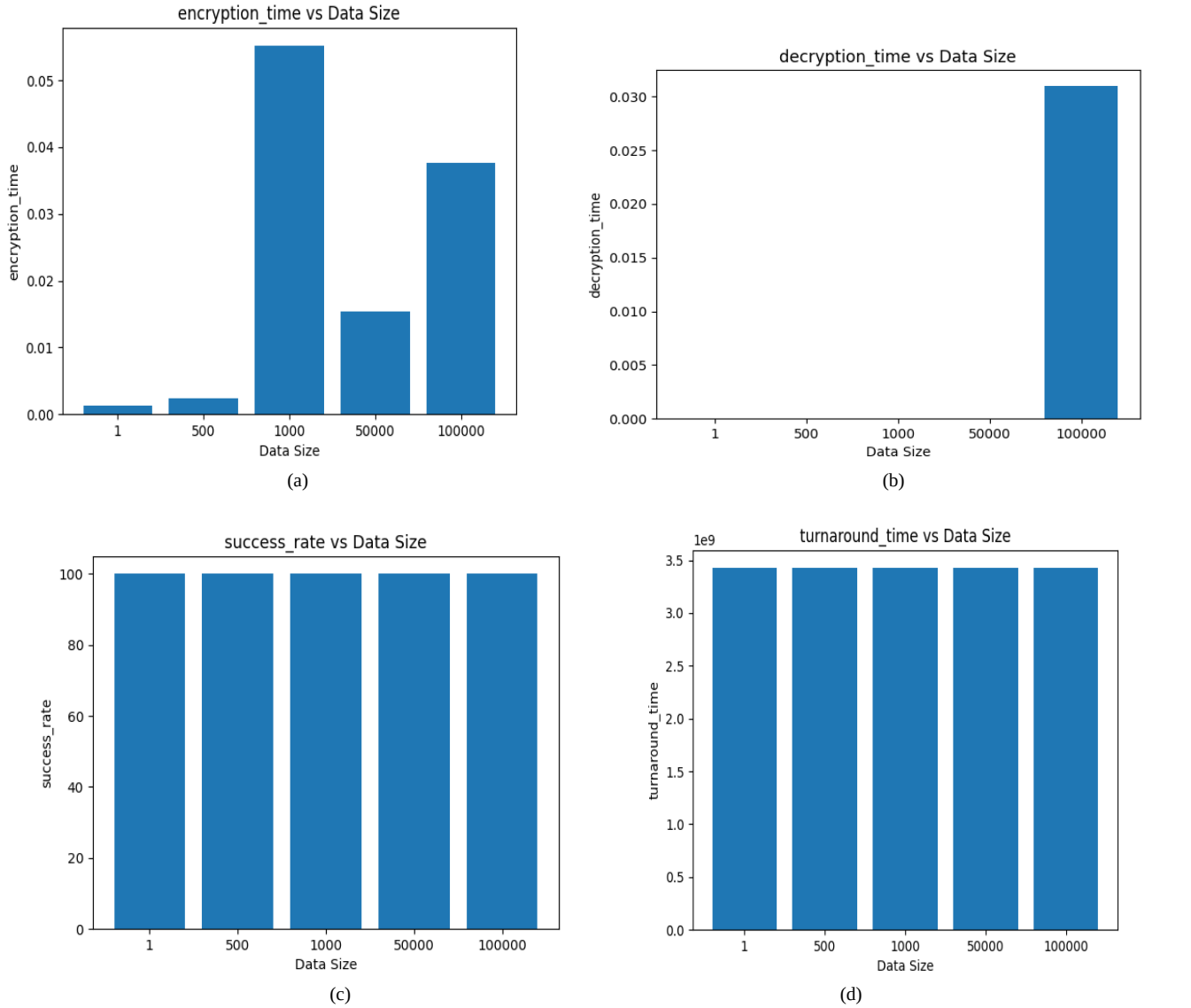


Fig. 6. Performance analysis: (a) encryption time, (b) decryption time, (c) success rate, and (d) turnaround time.

Further, the success rate was determined to evaluate how the proposed system prevents the cyber security threats during transmission. Fig. 6 (c) presents the analysis of success rate over increasing data size. The proposed technique achieved a maximum success rate of 100%, highlighting that it effectively resists the attacks and protects the data. Finally, the turnaround time was assessed over increasing data sizes. The turnaround time indicates total time required to complete a cryptographic operation, including both encryption and decryption processes, in association with other tasks like key generation, initialization, and data processing. The developed model obtained a minimum turnaround time of 3.43E+09, demonstrating its computational efficiency and data processing speed.

B. Evaluation Metrics

In this module, we discuss the parameters used for examining the performance of the proposed algorithm. The parameters include accuracy, recall, f1-score, MSE, MAE, Normalized Mean Square Error (NMSE), success rate, encryption time, decryption time, and turnaround time. The definition for the parameters are described below.

a) Accuracy: Accuracy measures how effectively the developed model predicts the threats in the cloud model. It quantifies the correctly predicted positive and negative instances to the total instances, and it is formulated in Eq. (18).

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{18}$$

Where T_p T_n F_p, and F_n defines TP, TN, FP, and FN, respectively.

b) Recall: Recall measures the proportion of true positives that were correctly identified by the developed model. It is calculated using Eq. (19).

$$\text{Re}call = \frac{T_p}{T_p + F_n} \tag{19}$$

c) F1-score: F1-score: F1-score is the harmonic mean of precision and recall, providing a balanced measure between them. It is mathematically expressed in Eq. (20).

$$F - measure = \frac{2 * T_p}{\left(2 * T_p + F_p + F_n\right)} \tag{20}$$

d) Mean squared error: MSE measures the average squared difference between the predicted values and the actual values, and it is formulated in Eq. (21).

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(y_i - y\right)^2 \tag{21}$$

Where y defines the actual value, and y_i indicates the predicted value.

e) Mean absolute error: MAE measures the average absolute difference between the predicted values and the actual values. It is expressed in Eq. (22).

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|y_i - y\right| \tag{22}$$

f) Normalized mean square error: NMSE measures the relative mean squared error between the predicted values and the actual values, normalized by the variance of the actual values, and it is formulated in Eq. (23).

$$NMSE = 10\log_{10}\left[\frac{\sum_{i=1}^{m}\left|y_i - y\right|^2}{\sum_{i=1}^{m}\left|y_i\right|^2}\right] \tag{23}$$

g) Success rate: Success rate refers to the effectiveness of a cryptographic algorithm in resisting attacks. It measures how effectively the developed model prevents cyber security threats.

h) Encryption time: Encryption time measures the time taken by a cryptographic algorithm to encode the entire data.

i) Decryption time: Decryption time measures the time taken by a cryptographic algorithm to decode the original message from the encrypted data.

j) Turnaround time: Turnaround time indicates the total time required to complete a cryptographic operation, including both encryption and decryption processes.

These parameters provide comprehensive insights into different aspects of the performance of the proposed algorithm, encompassing accuracy, efficiency, security, and reliability.

C. Comparative Analysis

In this section, the performances incurred by the proposed model were compared and validated with the existing techniques such as Convolutional Neural Network (CNN) [32], Deep Neural Network (DNN) [33], Recurrent Neural Network (RNN) [34], and Support Vector Machine (SVM) [35].

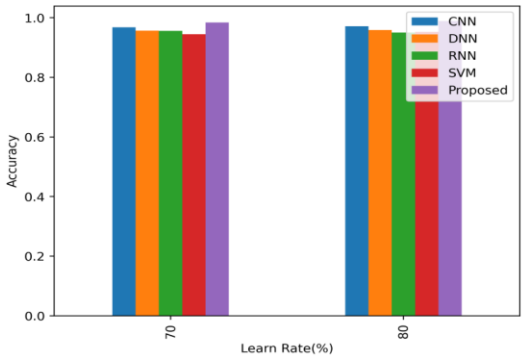
1) Comparison of threat detection performance: In this subsection, we compare the threat detection performance of the developed model with the conventional techniques such as CNN, DNN, RNN, and SVM. The performances are evaluated in terms of metrics like accuracy, f1-score, recall, MAE, MSE, and NMSE at 70% and 80% learning rates. Fig. 7 presents the comparative assessment. Fig. 7 (a) presents the accuracy comparison. The above-stated existing techniques and the proposed DRNN-KHO algorithm obtained an approximate accuracy of 0.96, 0.95, 0.955, 0.94, and 0.983, respectively at 70% learning rate, while the above techniques achieved an approximate accuracy of 0.97, 0.958, 0.95, 0.952, and 0.989, respectively at 80% learning rate. From this analysis, it is clear that the proposed algorithm achieved improved accuracy than the existing techniques. The improvement of accuracy signifies its efficiency in detecting and classifying the normal and

malicious traffic within the network. Consequently, we evaluated the MSE with the above-stated existing techniques. The above-mentioned existing algorithms and the developed approach obtained MSE of 0.03, 0.04, 0.044, 0.05, and 0.01, respectively at 70% accuracy. On the other hand, these approaches earned a MSE of 0.02, 0.041, 0.049, 0.047, and 0.01, respectively at 80% learning rate. The significant reduction of MSE by the designed model highlights that it accurately classifies normal and malicious traffic. Fig. 7 (b) presents the comparison of MSE.

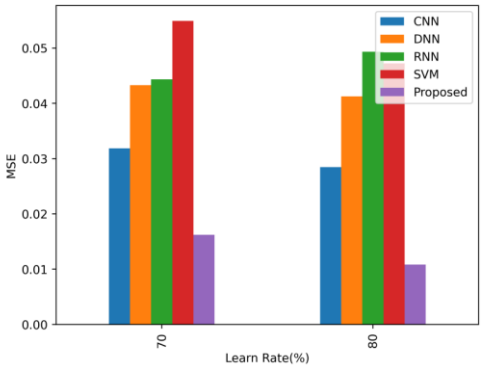
TABLE II. PERFORMANCE COMPARISON AT 70% LEARNING RATE

Metrics	CNN	DNN	RNN	SVM	PROPOSED
MSE	0.031876	0.043304	0.044331	0.054924	0.016211
MAE	0.031876	0.043304	0.044331	0.054924	0.016211
NMSE	0.015315	0.020805	0.021299	0.026389	0.007789
Recall	0.968206	0.956729	0.956029	0.945172	0.983659

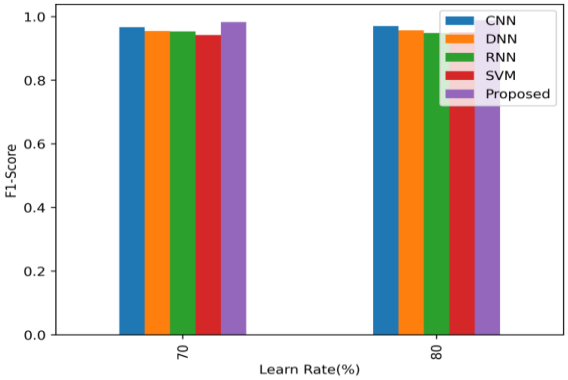
On the other hand, the f1-score was compared with the existing techniques. The f1-score determines the balanced classification performance considering both positives and negatives. These conventional models and the algorithm obtained f1-score of 0.96, 0.954, 0.953, 0.942, and 0.982, respectively at 70% learning rate, while these algorithms obtained f1-score of 0.966, 0.954, 0.953, 0.942, and 0.982, respectively, at 80% learning rate. The significant improvement made by the developed model highlights its efficiency in balanced classification performance than the conventional techniques. Fig. 7(c) depicts the comparison of f1-score. Subsequently, the recall was also evaluated and compared with the existing techniques. The above-stated existing algorithms and the proposed approach earned recall rate of 0.968, 0.956, 0.956, 0.945, and 0.983, respectively at 70% learning rate, while these models obtained recall of 0.970, 0.956, 0.948, 0.950, and 0.988, respectively at 80% learning rate. The enhancement of the recall made by the developed algorithm highlights its efficiency in classifying the network traffic. Fig. 7(d) presents recall comparison. Table II presents the comparative analysis of classification performance at 70% learning rate.



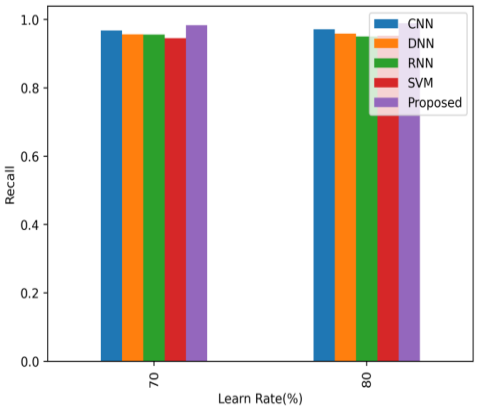
(a)



(b)



(c)



(d)

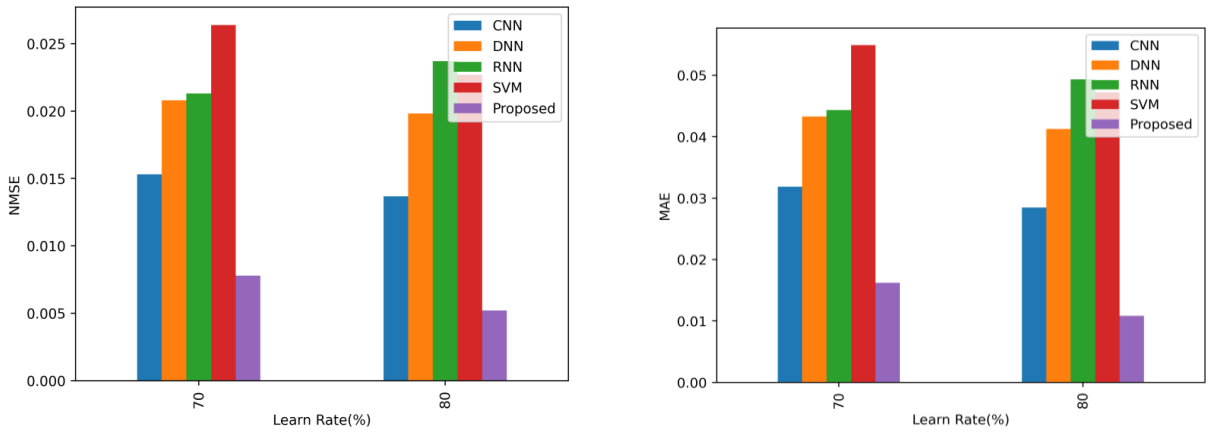


Fig. 7. Comparative analysis: (a) accuracy, (b) MSE, (c) F1-score, (d) recall, (e) NMSE, and (f) MAE.

TABLE III. PERFORMANCE COMPARISON AT 80% LEARNING RATE

Metrics	CNN	DNN	RNN	SVM	PROPOSED
MSE	0.028473	0.041281	0.049339	0.04722	0.010834
MAE	0.028473	0.041281	0.049339	0.04722	0.010834
NMSE	0.01368	0.019834	0.023705	0.022687	0.005205
Accuracy	0.971527	0.958719	0.950661	0.95278	0.989166
F1-Score	0.970089	0.9567	0.948259	0.950446	0.988611
Recall	0.971578	0.958868	0.950477	0.952343	0.989152

Furthermore, the MAE, and NMSE performances of the developed algorithm were compared and evaluated with the conventional techniques. Fig. 7 (e, f) presents the comparison of MAE and NMSE. These metrics measure the error in classification. The existing techniques and the developed approach obtained MAE of 0.03, 0.04, 0.044, 0.05, and 0.01, respectively at 70% learning rate, while these techniques earned NMSE of 0.01, 0.02, 0.021, 0.026, and 0.007, respectively at 70% learning rate. On the other hand, these models obtained MSE of 0.03, 0.04, 0.044, 0.05, and 0.01, respectively, and these algorithms achieved NMSE of 0.01, 0.02, 0.021, 0.026, and 0.007, respectively at 80% learning rate. From this analysis, it is clear that the developed algorithm achieved minimum MAE and NMSE compared to the existing techniques. This highlights the model's efficiency of accurately classifying the threats. In addition, it manifests that the designed model provides minimum error, and lower false positives and negatives. Table III presents the comparison of classification performance of different models at 80% learning rate. From this comprehensive comparative evaluation, it is clear that the designed model achieved better performances compared to the existing techniques, highlighting its efficiency of accurately classifying normal and malicious network traffic.

2) Comparison of cryptographic algorithm performance: In this section, we compare and evaluate the performance of the cryptographic algorithms with the existing algorithms like RSA [36], Advanced Standard Encryption (ASE) [37], ECC [38], and Diffie Hellman Key Exchange (DHKE) [39]. The performances are evaluated in terms of success rate, encryption time and decryption time, and they are evaluated for increasing data size from 1 to 100000.

The encryption time measures the time taken by the proposed algorithm for encoding the entire dataset. The existing techniques such as RSA, ASE, ECC, and DHKE obtained an average encryption time of 0.30, 0.41, 0.23, and 0.15, while the proposed algorithm achieved a minimum encryption time of 0.055. This illustrates that the designed algorithm quickly encrypts the dataset. Fig. 8 (a) presents the comparison of decryption time. Consequently, the decryption of above-stated existing techniques and the proposed are evaluated, and it is displayed in Fig. 8 (b). The decryption time defines the time taken by the developed algorithm for decoding the original message from the encrypted data. These models obtained decryption time of 0.30, 0.89, 0.35, 0.25, and 0.0309, respectively. From this analysis, it is clear that the proposed approach obtained less decryption time than others, highlighting its efficiency and speed in decrypting the data.

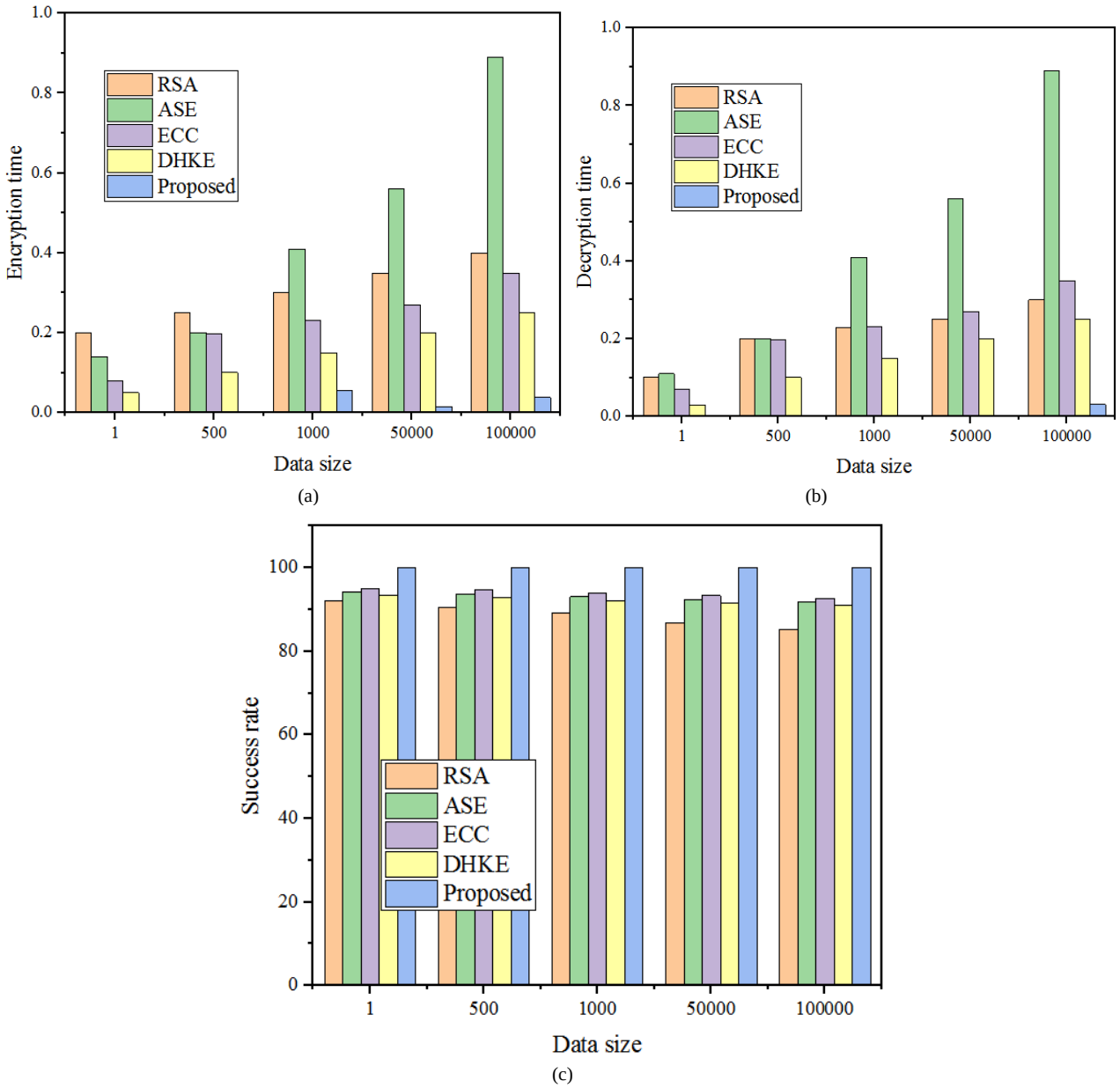


Fig. 8. Comparative analysis: (a) encryption time, (b) decryption time, and (c) success rate.

Finally, the success rate was also determined and compared with the existing techniques. The above-mentioned existing techniques and the proposed algorithm obtained an average success rate of 92.11, 94.21, 95.05, 93.52, and 100, respectively. The significant improvement of the success rate by the proposed algorithm manifests its effectiveness in preventing the intrusion during data transmission. Fig. 8 (c) presents the comparison of success rate. This intensive analysis of system performances suggest that the developed algorithm resists the security attacks and provides greater level of security to the data during transmission.

VI. Conclusion

In this study, we proposed a collaborative framework for ensuring security in the cloud environment. The objective of the work is to assess the cyber threats and provide secure data transmission within the cloud network. The developed model was validated using the publicly available DDoS database, which is preprocessed and trained into the DRNN-KHO module for cyber attack detection. Consequently, we developed a hybrid cryptographic algorithm by integrating ABC into the ECC algorithm for securely transmitting the information into the cloud. The developed framework was modeled and implemented in the Pycharm tool, and the performances are assessed as accuracy, recall, f1-score, MSE, MAE, NMSE, success rate, etc. The experimental results highlight that the developed approach obtained greater accuracy of 0.989, higher recall of 0.989, improved f1-score of 0.988, reduced MSE of 0.010, lower MAE of 0.0108, and minimum NMSE of 0.005, respectively in threat detection. Consequently, the implementation of ABC-ECC algorithm suggests that it achieved 100% success rate, 0.037s encryption time, 0.0309s

decryption time, and 3.43E+09 turnaround time. Finally, we made a comparative assessment with the existing techniques, and it validated that the proposed algorithm achieved better performances than others.

Compliance with Ethical Standards

Conflict of interest

The authors declare that they have no conflict of interest.

Human and Animal Rights

This article does not contain any studies with human or animal subjects performed by any of the authors.

Informed Consent

Informed consent does not apply as this was a retrospective review with no identifying patient information.

Funding: Not applicable

Conflicts of interest Statement: Not applicable

Consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material:

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Code availability: Not applicable

REFERENCES

[1] Golightly, Lewis, et al. "Adoption of cloud computing as innovation in the organization." International Journal of Engineering Business Management 14 (2022): 18479790221093992.

[2] Mohamed, A., Hamdan, M., Khan, S., Abdelaziz, A., Babiker, S. F., Imran, M., & Marsono, M. N. (2021). Software-defined networks for resource allocation in cloud computing: A survey. Computer Networks, 195, 108151.

[3] Dwivedi, S. K., Yadav, J., Ansar, S. A., Khan, M., & Pandey, D. (2023, November). An investigation with emerging issues and preventive measures: Cloud computing perspective. In AIP Conference Proceedings (Vol. 2821, No. 1). AIP Publishing.

[4] Hentschel, R., Bley, K., & Lange, F. (2023, November). A Performance-Based Assessment Approach for Cloud Service Provider Selection. In Conference on e-Business, e-Services and e-Society (pp. 250-264). Cham: Springer Nature Switzerland.

[5] Al-Marsy, A., Chaudhary, P., & Rodger, J. A. (2021). A model for examining challenges and opportunities in use of cloud computing for health information systems. Applied System Innovation, 4(1), 15.

[6] Malik, Latesh, and A. Sandhya. Computing Technologies and Applications. United States: CRC Press, 2021.

[7] Landoll, Douglas. The security risk assessment handbook: A complete guide for performing security risk assessments. CRC press, 2021.

[8] El Kafhali, S., El Mir, I., & Hanini, M. (2022). Security threats, defense mechanisms, challenges, and future directions in cloud computing. Archives of Computational Methods in Engineering, 29(1), 223-246.

[9] Masood, Arooj, Demeke Shumeye Lakew, and Sungrae Cho. "Security and privacy challenges in connected vehicular cloud computing." IEEE Communications Surveys & Tutorials 22.4 (2020): 2725-2764.

[10] Parast, Fatemeh Khoda, et al. "Cloud computing security: A survey of service-based models." Computers & Security 114 (2022): 102580.

[11] Kotak, J., Habler, E., Brodt, O., Shabtai, A., & Elovici, Y. (2023). Information Security Threats and Working from Home Culture: Taxonomy, Risk Assessment and Solutions. Sensors, 23(8), 4018.

[12] Ullah, Raja Muhammad Ubaid. "An empirical study of establishing guidelines for evaluation and adoption of secure and cost effective cloud computing." (2023).

[13] Nandy, T., Noor, R. M., Kolandaisamy, R., Idris, M. Y. I., & Bhattacharyya, S. (2024). A review of security attacks and intrusion detection in the vehicular networks. Journal of King Saud University-Computer and Information Sciences, 101945.

[14] Shah, Varun, and Sreedhar Reddy Konda. "Cloud Computing in Healthcare: Opportunities, Risks, and Compliance." Revista Espanola de Documentacion Cientifica 16.3 (2022): 50-71.

[15] Sen, Amartya, and Sanjay Madria. "Analysis of a cloud migration framework for offline risk assessment of cloud service providers." Software: Practice and Experience 50.6 (2020): 998-1021.

[16] Sarker, Iqbal H. "AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems." SN Computer Science 3.2 (2022): 158.

[17] Abdullahi, M., Baashar, Y., Alhussian, H., Alwadain, A., Aziz, N., Capretz, L. F., & Abdulkadir, S. J. (2022). Detecting cybersecurity attacks in internet of things using artificial intelligence methods: A systematic literature review. Electronics, 11(2), 198.

[18] Asharf, J., Moustafa, N., Khurshid, H., Debie, E., Haider, W., & Wahab, A. (2020). A review of intrusion detection systems using machine and deep learning in internet of things: Challenges, solutions and future directions. Electronics, 9(7), 1177.

[19] Adee, Rose, and Haralambos Mouratidis. "A dynamic four-step data security model for data in cloud computing based on cryptography and steganography." Sensors 22.3 (2022): 1109.

[20] Kure, Halima Ibrahim, Shareeful Islam, and Haralambos Mouratidis. "An integrated cyber security risk management framework and risk predication for the critical infrastructure protection." Neural Computing and Applications 34.18 (2022): 15241-15271.

[21] Gupta, L., Salman, T., Ghubaish, A., Unal, D., Al-Ali, A. K., & Jain, R. (2022). Cybersecurity of multi-cloud healthcare systems: A hierarchical deep learning approach. Applied Soft Computing, 118, 108439.

[22] Al Saleh, Reem, Maha Driss, and Iman Almomani. "CBiLSTM: A hybrid deep learning model for efficient reputation assessment of cloud services." IEEE Access 10 (2022): 35321-35335.

[23] Denis, R., and P. Madhubala. "Hybrid data encryption model integrating multi-objective adaptive genetic algorithm for secure medical data communication over cloud-based healthcare systems." Multimedia Tools and Applications 80.14 (2021): 21165-21202.

[24] Thabit, Fursan, et al. "A new lightweight cryptographic algorithm for enhancing data security in cloud computing." Global Transitions Proceedings 2.1 (2021): 91-99.

[25] Chinnasamy, P., et al. "Efficient data security using hybrid cryptography on cloud computing." Inventive Communication and Computational Technologies: Proceedings of ICICCT 2020. Springer Singapore, 2021.

[26] Guan, Shaopeng, et al. "Hadoop-based secure storage solution for big data in cloud computing environment." Digital Communications and Networks 10.1 (2024): 227-236.

[27] El Kafhali, Said, Iman El Mir, and Mohamed Hanini. "Security threats, defense mechanisms, challenges, and future directions in cloud computing." Archives of Computational Methods in Engineering 29.1 (2022): 223-246.

[28] Almiani, M., AbuGhazleh, A., Al-Rahayfeh, A., Atiewi, S., & Razaque, A. (2020). Deep recurrent neural network for IoT intrusion detection system. Simulation Modelling Practice and Theory, 101, 102031.

[29] Wei, Cheng-Long, and Gai-Ge Wang. "Hybrid annealing krill herd and quantum-behaved particle swarm optimization." Mathematics 8.9 (2020): 1403.

[30] Ullah, S., Zheng, J., Din, N., Hussain, M. T., Ullah, F., & Yousaf, M. (2023). Elliptic Curve Cryptography; Applications, challenges, recent advances, and future trends: A comprehensive survey. Computer Science Review, 47, 100530.

[31] Jacob, I. Jeena, and P. Ebby Darney. "Artificial bee colony optimization algorithm for enhancing routing in wireless networks." Journal of Artificial Intelligence 3, no. 01 (2021): 62-71.

[32] Nguyen, Minh Tuan, and Kiseon Kim. "Genetic convolutional neural network for intrusion detection systems." Future Generation Computer Systems 113 (2020): 418-427.

[33] Ramaiah, M., Chandrasekaran, V., Ravi, V., & Kumar, N. (2021). An intrusion detection system using optimized deep neural network architecture. Transactions on Emerging Telecommunications Technologies, 32(4), e4221.

[34] Nayyar, S., Arora, S., & Singh, M. (2020, July). Recurrent neural network based intrusion detection system. In 2020 international conference on communication and signal processing (iccsp) (pp. 0136-0140). IEEE.

[35] Bhati, Bhoopesh Singh, and Chandra Shekhar Rai. "Analysis of support vector machine-based intrusion detection techniques." Arabian Journal for Science and Engineering 45.4 (2020): 2371-2383.

[36] Wahab, O. F. A., Khalaf, A. A., Hussein, A. I., & Hamed, H. F. (2021). Hiding data using efficient combination of RSA cryptography, and compression steganography techniques. IEEE access, 9, 31805-31815.

[37] Altigani, Abdelrahman, Shafaatunnur Hasan, Bazara Barry, Shiraz Naserelden, Muawia A. Elsadig, and Huwaida T. Elshoush. "A polymorphic advanced encryption standard–a novel approach." IEEE Access 9 (2021): 20191-20207.

[38] Qazi, R., Qureshi, K. N., Bashir, F., Islam, N. U., Iqbal, S., & Arshad, A. (2021). Security protocol using elliptic curve cryptography algorithm for wireless sensor networks. Journal of Ambient Intelligence and Humanized Computing, 12, 547-566.

[39] Mitra, S., Das, S., & Kule, M. (2021). Prevention of the man-in-the-middle attack on Diffie–Hellman key exchange algorithm: A review. In Proceedings of International Conference on Frontiers in Computing and Systems: COMSYS 2020 (pp. 625-635). Springer Singapore.

Optimizing Cervical Cancer Diagnosis with Correlation-Based Feature Selection: A Comparative Study of Machine Learning Models

Wiwit Supriyanti[1], Sujalwo[2], Dimas Aryo Anggoro[3], Maryam[4], Nova Tri Romadloni[5]
Computer Engineering, Universitas Muhammadiyah Karanganyar, Karanganyar, Indonesia[1]
Informatics, Universitas Muhammadiyah Karanganyar, Karanganyar, Indonesia[2, 5]
Informatics Department, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia[3, 4]

***Abstract*—Cervical cancer remains a significant global health issue, particularly in developing countries where it is a leading cause of mortality among women. The development of machine learning-based approaches has become essential for early detection and diagnosis of cervical cancer. This research explores the optimization of classification algorithms through Correlation-Based Feature Selection (CFS) for early cervical cancer detection. A dataset consisting of 198 samples and 22 attributes from medical records was processed to reduce dimensionality. CFS was used to select the most relevant features, which were then applied to three classification algorithms: Naïve Bayes, Decision Tree, and k-Nearest Neighbor (k-NN). The results showed that CFS significantly improved classification accuracy, with Decision Tree achieving the highest accuracy of 85.89%, followed by Naïve Bayes with 83.34%, and k-NN with 82.32%. These findings demonstrate the importance of feature selection in enhancing classification performance and its potential application in the development of cervical cancer detection tools.**

***Keywords*—Cervical cancer; feature selection; machine learning**

I. Introduction

Cervical cancer is a major global health concern and a leading cause of women's death; the situation is even worse in developing regions due to a lack of health facilities and people's knowledge regarding this deadly disease. People have various assumptions regarding this disease; on the contrary, their knowledge often misleads them on how to prevent the deadly disease or to heal it [1].

For example, a common misconception about cervical cancer is that it only affects women who are sexually active or those with a history of multiple sexual partners. While human papillomavirus (HPV) infection, a key risk factor for cervical cancer, is often transmitted through sexual contact, it is crucial to note that the disease can also develop in women with a limited sexual history with only one partner in conservative marriage or even those who have never been sexually active since HPV can be spread through various ways.

HPV (Human Papillomavirus) is most commonly associated with sexual transmission, but there is a strong possibility that HPV will be transmitted through non-sexual means. One possible way is through direct skin-to-skin contact. HPV can be transmitted when the skin comes into contact with an infected area, even without sexual intercourse. This means that activities such as touching or rubbing areas where the virus is present—like the genital, anal, or oral regions—can lead to transmission. The virus can enter the body through tiny cuts, abrasions, or micro-tears in the skin or mucous membranes, making it possible for someone to contract HPV without having penetrative sex. Since the youth has been experiencing pre-sexual activity, there are strong possibilities that such a viral transmission through this case will eventually occur due to the frequency and possibility that happened after pre-sexual activities.

Another potential, although less common, route of HPV transmission is through indirect contact with contaminated objects. This can happen if personal items such as towels, razors, or undergarments, which an infected person has used, come into contact with broken skin or mucous membranes. While the likelihood of contracting HPV in this manner is lower, it is still a possible mode of transmission. Thus, sharing personal items should be avoided to reduce the risk of spreading the virus.

HPV can also be transmitted from mother to child during childbirth, a process known as vertical transmission. In these cases, the newborn may come into contact with the virus in the birth canal, which can lead to conditions such as recurrent respiratory papillomatosis—a rare disorder where warts grow in the respiratory tract. While such occurrences are uncommon, they highlight the need for awareness that HPV transmission is not solely linked to sexual behavior.

Since HPV is highly infectious, this misconception can lead to a dangerous delay in screening or seeking medical advice, as many women might mistakenly believe they are not at risk since the women do not have enough exposure to information that viral infection can be caused by non-penetrative activities and the virus can remain dormant for years before slowly transforming cellular changes in the cervix.

Furthermore, other factors such as genetics, smoking, and a weakened immune system also play a role in the development of cervical cancer. Raising awareness about the true risk factors and promoting regular screening, such as Pap smears and HPV tests, are crucial in early detection and prevention, regardless of one's sexual history.

Not only the risk of infection that urgently needs to be addressed but simultaneously, we should discuss the fatality of this disease. Naturally, cervical cancer originates in a woman's

cervix, a part of the female reproductive system that is particularly vulnerable to infections and abnormal cellular changes. This vulnerability is due to its location and exposure to the external environment, making it susceptible to open wounds, abrasions, and subsequent infections by viruses or bacteria. HPV infection, in particular, poses a significant risk because it can cause cellular mutations in the cervix that may eventually lead to cancer. Unfortunately, cervical cancer can be deceptively slow in its development, with precancerous changes taking years or even decades to progress into invasive cancer. This gradual progression often results in many women being unaware of the presence of the disease until it reaches an advanced stage, by which time treatment becomes more complex, and outcomes are less favorable.

What makes cervical cancer particularly dangerous is that the symptoms often do not appear until the cancer has progressed significantly. Symptoms like unusual vaginal bleeding, pelvic pain, or discomfort during intercourse may not manifest until the cancer is already advanced, making early detection crucial. Regular screenings, such as Pap smears and HPV tests, are therefore vital for catching any abnormalities at the earliest possible stage. Despite these challenges, advancements in cancer detection technology are continually improving. For instance, the development of the HPV vaccine and the ability to test specifically for high-risk HPV strains have greatly enhanced the ability to prevent and diagnose this type of cancer early on. By identifying the presence of high-risk HPV strains before they lead to cellular changes, healthcare providers can intervene with treatment or increased monitoring to prevent the progression of cervical cancer.

Regular, yearly examinations are strongly recommended for all women, particularly those between the ages of 21 and 65, to detect early HPV infections or precancerous changes. Early diagnosis significantly reduces the risk of cervical cancer progressing to a life-threatening stage and provides the opportunity for effective treatment and management. Education and awareness about the non-sexual transmission of HPV and the importance of regular screenings can help dispel misconceptions and encourage more women to take preventive action against this potentially devastating disease.

What makes cervical cancer particularly dangerous is that the symptoms often do not appear until the cancer has progressed significantly. Symptoms like unusual vaginal bleeding, pelvic pain, or discomfort during intercourse may not manifest until the cancer is already advanced, making early detection crucial. Regular screenings, such as Pap smears and HPV tests, are therefore vital for catching any abnormalities at the earliest possible stage. Despite these challenges, advancements in cancer detection technology are continually improving. For instance, the development of the HPV vaccine and the ability to test specifically for high-risk HPV strains have greatly enhanced the ability to prevent and diagnose this type of cancer early on. By identifying the presence of high-risk HPV strains before they lead to cellular changes, healthcare providers can intervene with treatment or increased monitoring to prevent the progression of cervical cancer.

Regular, yearly examinations are strongly recommended for all women, particularly those between the ages of 21 and 65, to detect early HPV infections or precancerous changes. Early diagnosis significantly reduces the risk of cervical cancer progressing to a life-threatening stage and provides the opportunity for effective treatment and management. Education and awareness about the non-sexual transmission of HPV and the importance of regular screenings can help dispel misconceptions and encourage more women to take preventive action against this potentially devastating disease.

Based on the nature of this virus, several prevention systems should be established. if the sexual educational system is designed systematically and the anti-HPV campaign can reach youth efficiently, the spread of HPV can be contained further. Thus, in reality, the youth does not have enough strong relationships and bonds with stakeholders, and an adequate prevention system for protecting women from HPV has not been established firmly.

Consequently, according to the World Health Organization (WHO), in 2018, there were over 570,000 new cases of cervical cancer, leading to more than 311,000 deaths linked to the disease [2]. This statistic highlights that, despite advances in medical technology, the failure of the cervical cancer prevention system has not successfully yet halted cervical cancer as a remaining significant threat to women's health globally.

From the early situation, the urgency of the significance of early detection of cervical cancer is a top priority for saving women widely in different social and economic classes. By detecting cervical cancer as early as possible, women’s lives would be saved, and they can maintain a proper and standardized quality for women as expected based on the conception of human rights. This statement is strengthened by further research that measures the standardization of human life quality and the development of early-stage detection of HPV and cervical cancer. By detecting cervical cancer earlier, the treatment will work significantly, and the impact on women’s lives is undeniably positive. Undoubtedly, when cervical cancer is identified in its early stages, the chances of successful treatment significantly increase. For example, preventive measures such as Pap smear tests and HPV vaccinations have been successfully efficient in controlling the more profound impacts of cervical cancer in terms of health and social welfare.

Nonetheless, while these strategies have successfully reduced incidence and mortality rates in some developed countries, access to early detection methods in developing nations remains severely limited. The challenges include a lack of medical resources, high diagnostic costs, and insufficient public awareness regarding the importance of early detection [3].

With the advancement of technology and data processing, information technology-based methods—such as machine learning and data mining—are increasingly being utilized in the analysis of medical records to assist in diagnosing and early detection of diseases, including cervical cancer. Classification is one of the most frequently employed techniques in data mining, wherein models are created to predict whether a patient is at high risk of a specific disease, including cervical cancer, based on historical or existing medical data [4]. The application of classification algorithms allows researchers to uncover patterns

within datasets and produce accurate predictions based on the available attributes.

Nevertheless, one of the most significant challenges in medical classification processes, particularly for the early detection of cervical cancer, is the high dimensionality of the data. Medical datasets often contain numerous attributes or features, many of which may be irrelevant for predictions or could introduce noise into the model. This situation is referred to as the high dimensionality problem, which can lead to reduced model performance and an increased risk of overfitting, where the model becomes overly focused on the training data and fails to generalize well to new data [5]. In the context of early cervical cancer detection, attributes within the dataset might include various patient information such as age, pregnancy history, age at first menstruation, and symptoms like fatigue, abnormal bleeding, and abdominal lumps. While some attributes may hold greater significance than others, the sheer number of attributes complicates the classification process and makes it challenging to interpret.

To address the high dimensionality issue, feature selection is employed. Feature selection involves selecting the most relevant and informative subset of features from the dataset while discarding irrelevant or redundant ones. The aim of this process is to simplify the model, enhance interpretability, and, most importantly, improve predictive performance. By utilizing feature selection methods, it is anticipated that classification algorithms can operate more efficiently and accurately [6]. One popular feature selection method used in various classification applications is Correlation-Based Feature Selection (CFS).

High dimensionality in medical datasets presents various challenges, particularly in computation and interpretation. As the number of features increases, the potential combinations in data analysis also rise exponentially. This implies that classification algorithms require more time and computational resources to process datasets with a larger number of features, increasing the likelihood of overfitting. Overfitting occurs when the model is too complex and fits the training data closely, yet fails to generalize to new test data, resulting in poor predictive performance when applied to real-world data [7].

For instance, in a medical dataset aimed at the early detection of cervical cancer, certain attributes may be highly relevant for predicting outcomes, such as the patient's age or pregnancy history. Conversely, there may be attributes that are irrelevant or have a low correlation with the target variable, such as unrelated family health history regarding cervical cancer risk. These extraneous features complicate the model without significantly contributing to predictive accuracy. Thus, implementing feature selection methods to identify the most important attributes for the classification process is essential [8].

Correlation-Based Feature Selection (CFS) is a technique that identifies an optimal subset of features based on the correlation between features in the dataset and the target variable. The fundamental concept behind CFS is that a good subset of features should consist of those that have a strong correlation with the target variable while maintaining a low correlation with one another. In other words, CFS selects features that have a robust relationship with the target variable (i.e., whether a patient is diagnosed with cervical cancer) while avoiding redundant features that provide overlapping information [9].

CFS operates by calculating the Pearson correlation between each feature and the target variable, as well as between the features themselves. The Pearson correlation measures the strength and direction of the linear relationship between two variables. Features with strong positive or negative correlations with the target variable are retained, while those with weak or insignificant correlations are discarded. Additionally, CFS considers the correlation among features to prevent redundancy. Features that are highly correlated with one another are deemed to provide similar information, allowing for one of them to be removed without diminishing model performance [10].

The primary advantage of CFS lies in its ability to reduce data dimensionality without compromising predictive accuracy. By eliminating irrelevant or redundant features, CFS helps classification models become simpler, faster, and more interpretable. This aspect is particularly crucial in medical applications, where the interpretability of the model is a key factor in providing trustworthy diagnoses [11].

In this study, after applying feature selection using CFS, the reduced dataset will be used to train classification models. The three classification algorithms employed are Decision Tree [12], Naïve Bayes [13], and k-nearest Neighbor (k-NN) [14]. These algorithms were selected for their differing approaches to classification problems and their unique strengths in medical data analysis. Each algorithm has its own advantages and limitations, and the findings of this study will assess the performance of each algorithm following feature selection using CFS.

Previous research on early detection of cervical cancer utilizing data mining techniques has yielded positive results. For instance, a study by Ali [4] employed the Random Forest method to analyze medical datasets, achieving a high level of accuracy in predicting cervical cancer. Another study by Rahmi [15] implemented the SMOTE algorithm and Naïve Bayes to address imbalanced data issues in the early detection of cervical cancer in Indonesia. Both studies acknowledged the importance of feature selection as a critical factor in enhancing the performance of classification models.

However, challenges remain unresolved in prior research, particularly concerning high dimensionality and model interpretability. This study aims to address these challenges by applying the Correlation-Based Feature Selection (CFS) method to reduce data dimensionality and by comparing the performance of various classification algorithms to identify the most effective approach for detecting cervical cancer.

II. Literature Review

The optimization of cervical cancer diagnosis has seen significant advancements with the integration of correlation-based feature selection and machine learning models. Recent research [16] emphasized the critical role of selecting key clinical and behavioral criteria in risk assessment. Their methodology combined fuzzy Multi-Criteria Decision Making (MCDM) with machine learning to improve the reliability of diagnostic outcomes, highlighting the importance of reducing data complexity while preserving essential diagnostic features.

Addressing the challenges of high-dimensional datasets, Nithya and Ilango [17] introduced a fused feature selection framework that combines filter and wrapper methods to optimize cervical cancer classification. Their approach effectively mitigates issues such as redundant attributes, irrelevant features, and class imbalance, resulting in improved classification accuracy. This demonstrates the importance of integrated feature selection strategies in handling complex medical datasets.

Deep learning has further enhanced cervical cancer diagnostics, particularly in image-based classification tasks. Tawalbeh [18] conducted a comparative study of six feature fusion techniques applied to deep learning models. By leveraging canonical correlation analysis, they achieved a classification accuracy of 99.7%, underscoring the power of optimized feature fusion in improving performance across multiple cancer classes.

Hasan [8] explored machine learning explainability in cervical cancer classification, employing the Boruta algorithm for feature selection and tools like SHAP for model interpretability. Using Random Forest as a classifier, they achieved an accuracy of 99.85%, demonstrating the value of combining explainable AI techniques with advanced feature selection to enhance diagnostic reliability.

The use of wrapper-based feature selection techniques has also proven effective in cervical cancer diagnostics. Setiawan [19] employed Grey Wolf Optimization (GWO) with classifiers like Naive Bayes and Support Vector Machines. Their findings revealed that NB-GWO outperformed other configurations, achieving an accuracy of 96.30%, highlighting the effectiveness of metaheuristic algorithms in selecting optimal features.

Machine learning has been pivotal in developing self-risk assessment tools for cervical cancer. Ramzan [20] utilized AdaBoost and feature selection algorithms to create personalized risk assessment models. Their technique enabled women to estimate their cervical cancer risk with high accuracy, leveraging demographic and medical history data.

Feature selection techniques, such as mutual information and genetic algorithms, have been extensively studied for their ability to remove irrelevant attributes and enhance model performance. Combining these methods with machine learning has proven particularly effective in addressing issues of overfitting and improving the generalization ability of models.

The application of explainable AI in medical diagnostics has received considerable attention, as demonstrated by Hasan [8]. Their integration of feature importance with interpretable models not only improved diagnostic accuracy but also provided clinicians with insights into the key predictors of cervical cancer, bridging the gap between AI and clinical practice.

Comparative analyses of feature selection methods have highlighted the need for tailored strategies to address the unique challenges of medical datasets. Studies like those of Tawalbeh [18] and Setiawan [19] illustrate the benefits of combining advanced feature engineering with machine learning, fostering the development of robust diagnostic tools.

Together, these studies underscore the transformative potential of correlation-based feature selection and machine learning in cervical cancer diagnosis. By integrating these techniques, researchers have achieved higher accuracy, better interpretability, and improved reliability in diagnostic models, paving the way for early detection and personalized healthcare solutions.

III. Methodology

The methods section of this research includes a series of steps starting with data collection, then moving to data preprocessing, feature selection using the Correlation-Based Feature Selection (CFS) technique, the application of classification algorithms, and finally, the assessment of the classification model's performance [21] (Fig. 1).

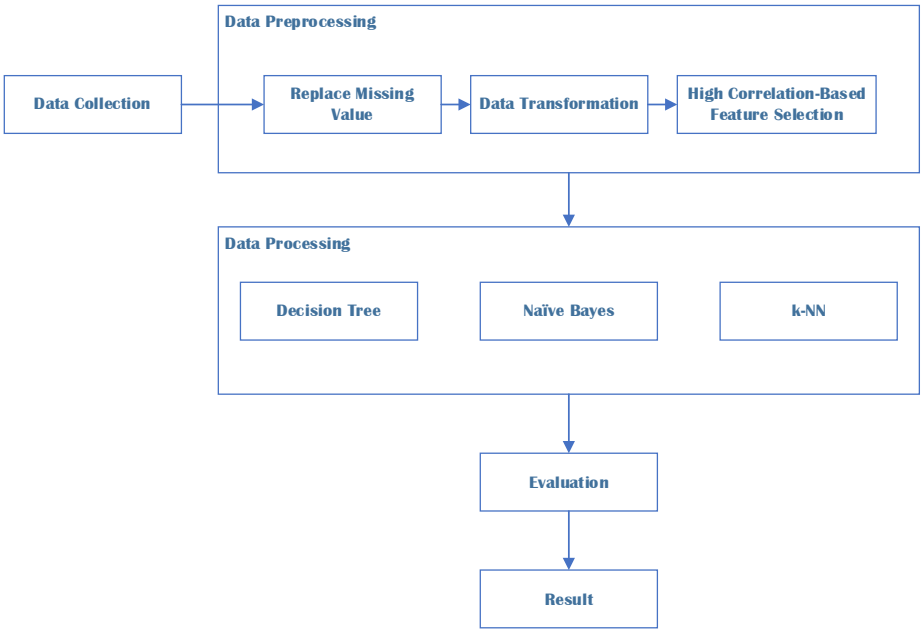


Fig. 1. The flow of research.

A. Data Collection

The data utilized in this research was collected from a hospital located in the area where the researcher resides, comprising a dataset of 198 patient samples who have undergone medical assessments related to cervical cancer. This dataset contains 22 attributes or features that represent various factors that may impact the diagnosis of cervical cancer. These attributes include demographic factors such as age and pregnancy history, as well as clinical symptoms like pelvic pain, abnormal discharge, and bleeding that occurs outside the menstrual cycle. The target variable for this study is the classification of cervical cancer stages, which includes in-situ, early, and advanced stages [22]. Table I shows attributes and variables of cervical cancer.

TABLE I. ATTRIBUTES AND VARIABLES OF CERVICAL CANCER

Variables	Attributes	Description
X1	Age	Age of the patient in years
X2	Marriages	Patient's history of number of marriages
X3	Miscarriages	Patient's history of number of miscarriages
X4	Childbirths	Patient's history of number of deliveries
X5	Age at first marriage	Patient's age at first marriage
X6	Age at first menstruation	Patient's age at first menstruation
X7	Difficulty defecating	0 = no, 1 = yes
X8	Difficulty urinating	0 = no, 1 = yes
X9	Decreased appetite	0 = no, 1 = yes
X10	Pelvic pain	0 = no, 1 = yes
X11	Lower abdominal pain	0 = no, 1 = yes
X12	Weight loss	0 = no, 1 = yes
X13	Nausea	0 = no, 1 = yes
X14	Vomiting	0 = no, 1 = yes
X15	Fatigue	0 = no, 1 = yes
X16	Foul-smelling discharge	0 = no, 1 = yes
X17	Discharge color	milky white, yellowish, greenish
X18	Bleeding outside of cycle	light, heavy
X19	Duration of bleeding	0-7 days, 7-14 days, more than 14 days
X20	Post-coital bleeding	0 = no, 1 = yes
X21	Abdominal lump	0 = no, 1 = yes
X22	Shortness of breath	0 = no, 1 = yes
Y	Cancer stage classification	in-situ, early stage, advanced stage

B. Data Preprocessing

Data pre-processing is an essential phase in preparing datasets to ensure they are suitable for analysis and classification. In this research, the pre-processing stage consists of three primary steps: addressing missing values, transforming the data, and selecting features using the Correlation-Based Feature Selection (CFS) method [23]. A detailed explanation of each step is provided below:

1) Replace missing values: In medical datasets, the occurrence of missing values is common due to incomplete data collection or inconsistencies in patient record documentation. It is essential to address these missing values to ensure the dataset's integrity, allowing the classification model to perform effectively. This study employs two prevalent methods for handling missing values:

a) Row deletion: When a row contains a significant proportion of missing data (for instance, more than 50%), it is removed from the dataset. This approach is taken because a high volume of missing data can compromise subsequent analyses.

b) Imputation: For attributes with only a small number of missing values, imputation techniques are utilized. This method replaces missing values using the information available within the dataset. The imputation process can be done through:

- Mean imputation: for numerical attributes, missing values are substituted with the average value of that attribute.
- Mode imputation: for categorical attributes, missing values are filled with the mode (the most frequently occurring value) of that attribute.

For instance, if there are some missing values in the age attribute, these values are replaced with the average age of the available patients. These techniques help maintain the distribution of the data without introducing significant bias.

2) Data transformation: Once the missing values have been addressed, the subsequent step involves transforming the data to ensure that all attributes are on an appropriate scale and in a format suitable for the classification algorithms being utilized. The data transformation process consists of:

a) Normalization of data: Normalization is essential to ensure that all numerical attributes are consistent in scale, particularly because algorithms such as k-Nearest Neighbor (k-NN) are very sensitive to variations in scale among the attributes. In this research, the Min-Max Scaling method is employed, which adjusts each attribute value to fall within a range of 0 to 1. The formula for normalization is:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where x is the original value of the attribute, $x_{\min}$ is the minimum value of that attribute, and $x_{\max}$ is the maximum value. Normalization helps prevent attributes with larger scales from dominating those with smaller scales [24].

b) Data standardization: In addition to normalization, standardization is applied to certain attributes that exhibit distributions significantly different from a normal distribution. Standardization is performed by subtracting the mean of each attribute from its value and then dividing by the standard

deviation of that attribute, as illustrated in the following equation:

$$z = \frac{x-\mu}{\sigma} \quad (2)$$

where z is the standardized value, x is the original value, μ is the mean, and σ is the standard deviation. Standardization ensures that the attributes have a distribution with a mean of 0 and a standard deviation of 1.

3) Correlation-based feature selection (CFS): After the data transformation process, the final step in pre-processing is feature selection using the Correlation-Based Feature Selection (CFS) method. CFS is a feature selection technique that automatically identifies the most relevant attributes in relation to the target variable while minimizing redundancy among the attributes. CFS operates by calculating the correlation between attributes and the target variable, as well as the correlation among the attributes themselves.

The steps in CFS are as follows:

a) Feature correlation with the target variable: CFS computes the Pearson correlation between each feature and the target variable (cervical cancer classes: in situ, early, advanced). Attributes with a high correlation to the target variable are retained, while those with low correlation are discarded.

b) Elimination of redundancy among features: In addition to retaining features that are highly correlated with the target variable, CFS also considers the correlation among the features themselves. Attributes that exhibit high correlation with one another are deemed redundant, and one of them is removed. The goal is to minimize multicollinearity within the dataset.

CFS calculates the "goodness value" (Merit) for each subset of features using the following formula:

$$Merit_S = \frac{k\,.\overline{r_{cf}}}{\sqrt{k+k\,(k-1).\overline{r_{ff}}}} \quad (3)$$

where:

- k is the number of selected features,
- $\overline{r_{cf}}$ is the average correlation between the features and the target variable,
- $\overline{r_{ff}}$ is the average correlation among the features.

The result of the CFS method is a reduction in the number of features from 22 to the 10 most relevant attributes for classification. This feature reduction not only enhances computational efficiency but also helps mitigate the risk of overfitting caused by the inclusion of irrelevant attributes.

C. Classification Algorithms

This study utilizes three popular classification algorithms to develop predictive models: Decision Tree, Naïve Bayes, and k-Nearest Neighbor (k-NN). Below is a detailed explanation of each algorithm:

1) Decision tree: The Decision Tree algorithm constructs a decision tree where each internal node represents an attribute or feature, each branch represents a decision based on the attribute's value, and each leaf node represents a classification outcome. Decisions are made by recursively partitioning the dataset into smaller subsets based on feature values that provide the most information. To select the most informative features, the concept of Information Gain based on Shannon entropy is employed. The formula for calculating entropy is as follows:

$$H(D) = -\sum_{i=1}^{C} p(i) log_2 p(i) \quad (4)$$

where p_i is the probability of class i in the dataset, and C is the total number of classes. The feature with the highest Information Gain is used to split the dataset at each node.

2) Naïve bayes: Naïve Bayes is a probability-based algorithm grounded in Bayes' Theorem. This algorithm assumes that all features are independent of one another, which is often unrealistic in real-world applications. However, in practice, Naïve Bayes continues to yield favorable results, particularly in text classification and large datasets. The basic equation of Naïve Bayes is:

$$P(C_k|X) = \frac{P(X|C_k)\,.P(C_k)}{P(X)} \quad (5)$$

where:

- $P(C_k|X)$ is the probability that sample X belongs to class C_k,
- $P(X|C_k)$ is the probability of observing X given class C_k,
- $P(C_k)$ is the prior probability of class C_k,
- $P(X)$ is the overall probability of the data X.

3) k-Nearest neighbor (k-NN): k-NN is an instance-based algorithm that classifies samples based on their distance to the nearest samples in the dataset. The algorithm works by calculating the Euclidean distance between the unknown sample and all existing samples in the dataset, then selecting the k closest neighbors and determining the majority class among them. The formula for Euclidean distance used in k-NN is:

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (6)$$

where p and q are two samples in the data space, and n is the number of features.

D. Evaluation

To evaluate the performance of the classification model, this study uses metrics derived from the Confusion Matrix, which includes calculations for accuracy, sensitivity, specificity, and precision [25]. Below are the explanations and formulas used to compute these metrics:

1) Accuracy: Accuracy measures the proportion of correct predictions against the total samples and is calculated using the following formula [26]:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

where:

- *TP* (True Positive) is the number of positive samples correctly classified,
- *TN* (True Negative) is the number of negative samples correctly classified,
- *FP* (False Positive) is the number of negative samples incorrectly classified as positive,
- *FN* (False Negative) is the number of positive samples incorrectly classified as negative.

2) Sensitivity (Recall): Sensitivity measures the model's ability to correctly identify positive samples and is calculated using the following formula:

$$Sensitivity = \frac{TP}{TP+FN} \quad (8)$$

A high sensitivity indicates that the model is effective in detecting positive conditions.

3) Specificity: Specificity measures the model's ability to accurately identify negative samples and is calculated using the following formula:

$$Specificity = \frac{TN}{TN+FP} \quad (9)$$

High specificity means that the model can identify negative samples with a high degree of accuracy.

4) Precision: Precision measures the accuracy of positive predictions, indicating how many of the positive predictions were correct, and is calculated using the following formula:

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

A high precision indicates that the model rarely produces false positive results.

After evaluating all the metrics, a comparative analysis is conducted to determine which algorithm performs best in detecting cervical cancer based on the processed dataset.

IV. Results and Discussions

This study aims to enhance classification performance in the early detection of cervical cancer using the Correlation-Based Feature Selection (CFS) method alongside three classification algorithms: Decision Tree, Naïve Bayes, and k-Nearest Neighbor (k-NN). Following the pre-processing and feature selection processes, the results of this research provide valuable insights into the impact of CFS on the performance of the employed classification algorithms. Below is a discussion of the results from each stage of the model performance evaluation.

A. Results of using Correlation-based Feature Selection (CFS)

The feature selection phase utilizing the CFS method revealed that only 10 out of 22 attributes in the dataset had significant relevance to the target variable (in-situ, early, advanced). The features selected through CFS include those directly related to clinical symptoms and risk factors known to contribute to the development of cervical cancer. Some important selected attributes include:

1) Abdominal lumps
2) Duration of bleeding
3) Bleeding outside the menstrual cycle
4) Color of discharge
5) Fatigue
6) Weight loss
7) Lower abdominal pain
8) Pelvic pain
9) Difficulty urinating
10) Difficulty defecating (specifically for the Decision Tree and k-NN algorithms)

The feature selection using CFS significantly reduced the number of irrelevant or less important attributes, ultimately enhancing the classification model's performance. Before feature selection, the dataset contained 22 attributes, but after selection, this number was reduced to 10. This reduction contributes to decreased computational complexity and minimizes the risk of overfitting commonly associated with high-dimensional datasets.

B. Evaluation of Classification Algorithm Performance

After the dataset was refined through feature selection using CFS, all three classification algorithms were implemented to predict the classification outcomes for cervical cancer. The performance of each algorithm was evaluated using several metrics, including accuracy, sensitivity, specificity, and precision.

1) Decision Tree: The Decision Tree algorithm is highly interpretable, and the results from this study indicate that it provides the highest accuracy compared to the other two algorithms. The maximum accuracy achieved by Decision Tree was 85.89% with all 10 selected attributes. This model also demonstrated a balanced performance in terms of sensitivity (79.3%) and specificity (83.1%), indicating its ability to accurately detect both positive and negative cases.

One of the main strengths of the Decision Tree is its capability to visualize the decision-making process in tree form, facilitating easier interpretation for medical practitioners. By examining the nodes and branches of the tree, doctors can discern which factors most influence the model’s decisions. In this case, attributes such as "abdominal lumps," "bleeding outside the cycle," and "color of discharge" emerged as key features underlying the decision of whether a patient is at risk for cervical cancer. This aligns with previous medical research indicating these symptoms are early warning signs of cervical cancer.

Another advantage of the Decision Tree is its ability to handle diverse data types, whether numerical or categorical, which is common in medical datasets. Furthermore, the process of partitioning the dataset into smaller subsets enables the Decision Tree to effectively manage high-dimensional issues after feature selection is applied.

1) Naïve bayes: Naïve Bayes is an efficient algorithm for handling large datasets and provides rapid predictions. In this dataset, following feature selection, the Naïve Bayes algorithm achieved the highest accuracy of 83.34% with 8 selected attributes. Despite the underlying assumption of independence among features often not being met in real-world applications, this algorithm still managed to deliver good predictive results. However, its sensitivity (72.5%) was somewhat lower compared to other algorithms, indicating that this model occasionally fails to detect some positive cases of cervical cancer.

The low sensitivity suggests that Naïve Bayes tends to have a higher rate of false negatives, which can be dangerous in medical cases like cervical cancer, where patients requiring treatment may go undetected. Nevertheless, the specificity attained (85.9%) is quite high, indicating that this algorithm is proficient in identifying negative cases and avoiding false positives, which is crucial for reducing unnecessary anxiety among patients deemed at low risk.

2) k-Nearest Neighbor (k-NN): The k-Nearest Neighbor (k-NN) algorithm showed significant performance improvement following feature selection using CFS. Before feature selection, k-NN only achieved an accuracy of 46.82%, demonstrating that this algorithm was highly affected by the high dimensionality of the data. However, after feature selection, accuracy increased to 82.32% with 8 selected attributes.

This substantial improvement indicates that k-NN relies heavily on a smaller number of relevant attributes. The k-NN algorithm functions by calculating the distance between new samples and existing samples in the dataset, classifying the new sample based on its nearest neighbors. When too many irrelevant features are included, the distance calculations become less accurate, leading to inaccurate classification results. Therefore, feature selection with CFS allows k-NN to concentrate on the most relevant features, yielding more accurate predictions.

Despite the notable accuracy improvement, k-NN still has a weakness in sensitivity (69.5%), suggesting that it tends to overlook some positive cases of cervical cancer. Additionally, this algorithm is sensitive to data scaling, necessitating proper normalization to ensure all attributes contribute equally to the distance calculations.

C. Comparative Analysis of Classification Algorithms

To provide a clearer picture of the performance of the three tested algorithms in this study, the following table compares the results for accuracy, sensitivity, specificity, and precision:

TABLE II. COMPARISON OF CLASSIFICATION ALGORITHM RESULTS

Algorithm	Accuracy	Sensitivity	Specificity	Precision
Decision Tree	85.89%	79.3%	83.1%	80.5%
Naïve Bayes	83.34%	72.5%	85.9%	74.1%
k-Nearest Neighbor	82.32%	69.5%	84.4%	71.8%

From the Table II, it is evident that the Decision Tree algorithm delivers the best overall performance, with the highest accuracy and relatively good sensitivity, followed by Naïve Bayes and k-NN. Meanwhile, Naïve Bayes excels in specificity, indicating that this algorithm is more accurate in identifying patients who are not at risk of cervical cancer. However, the drawback of Naïve Bayes lies in its sensitivity, which implies a higher risk of missing cases of cervical cancer.

The strength of k-NN lies in its simplicity and effectiveness in classification following feature selection. However, its lower sensitivity suggests that this algorithm tends to miss more positive cases of cervical cancer compared to the others.

D. Implementation in Early Detection of Cervical Cancer

The findings from this study indicate that correlation-based feature selection (CFS) is highly effective in enhancing the performance of classification algorithms in detecting cervical cancer. By reducing the number of irrelevant attributes and focusing solely on the most significant features, classification algorithms can produce more accurate and efficient predictions. This is crucial in medical applications, where diagnostic accuracy is a key factor in determining further actions.

Additionally, the Decision Tree stands out as the most ideal algorithm for use in the early detection of cervical cancer, as it not only provides the highest accuracy but also allows for easy interpretation of the decision-making process. This makes the Decision Tree a valuable tool for medical practitioners to understand the factors contributing to a cervical cancer diagnosis. One of the main benefits of the Decision Tree is its ability to visually represent the classification process, breaking down complex decision-making into a clear and straightforward flowchart. This visual representation helps healthcare professionals trace the sequence of factors leading to a diagnosis, making it easier to explain results to patients and colleagues in an understandable manner.

Another significant advantage of implementing the Decision Tree algorithm is its robustness in handling both categorical and numerical data, which is often the case in medical datasets. This flexibility allows for the integration of a diverse range of patient information, such as demographic factors, test results, and clinical history, into a single cohesive model without requiring extensive data preprocessing. Additionally, Decision Trees are not as sensitive to outliers or missing values as other algorithms, making them highly resilient in real-world medical applications where data quality can vary.

Moreover, Decision Trees facilitate effective feature selection by prioritizing the most important attributes early in

the model, which can lead to more efficient and faster computations. This is particularly beneficial in large-scale healthcare systems or when developing real-time diagnostic applications, as it reduces the processing time and computational resources needed to generate accurate predictions. The interpretability and efficiency of Decision Trees make them ideal for use in telemedicine and remote diagnostic tools, providing a means for timely and accurate assessments even in resource-constrained environments.

Lastly, the transparency of the Decision Tree's decision-making process supports better compliance with healthcare regulations and ethical standards. Because each step of the diagnosis can be easily traced and justified, Decision Tree models can enhance trust between medical professionals and patients and ensure accountability in the diagnostic process. These benefits position Decision Trees as a highly effective and practical solution for implementing early detection systems for cervical cancer, ultimately contributing to improved patient outcomes and more accessible healthcare.

With the results obtained, this research can serve as a foundation for developing information technology-based applications that assist individuals in conducting early screenings for cervical cancer independently. By leveraging proven classification algorithms and effective feature selection methods, such applications can deliver accurate results and help raise public awareness about the importance of early detection of cervical cancer.

V. Conclusion and Future Works

This study demonstrates that the Correlation-Based Feature Selection (CFS) method can significantly enhance classification performance on cervical cancer datasets. The Decision Tree, Naïve Bayes, and k-Nearest Neighbor (k-NN) algorithms all showed substantial improvements in accuracy following feature selection. Among the three algorithms, the Decision Tree achieved the highest accuracy, while k-NN experienced the most significant improvement after feature selection. These results indicate that employing appropriate feature selection techniques can optimize the predictive capabilities of machine learning models, leading to better diagnostic outcomes.

Furthermore, this research highlights the crucial role of feature selection in reducing data dimensionality and eliminating irrelevant or redundant features. By focusing on the most relevant attributes, the CFS method improves the computational efficiency of the models, resulting in faster and more accurate predictions. This optimization is particularly beneficial when dealing with complex datasets, such as those involving medical records or diagnostic tests, where irrelevant data can hinder model performance and lead to misleading results.

Widely, the implementation of these findings has practical applications beyond the academic context, especially in the development of technology-based tools to support the early detection and diagnosis of cervical cancer. By integrating machine learning algorithms enhanced with feature selection methods into digital healthcare platforms, it becomes possible to build robust and reliable diagnostic systems that can be used by healthcare professionals and patients alike. Such systems are particularly valuable in regions with limited access to specialized medical care, where timely diagnosis and treatment can be challenging.

Nonetheless, the use of these advanced diagnostic tools can significantly improve early detection rates, enabling healthcare providers to identify potential cases of cervical cancer at an earlier stage when treatment is more effective and the chances of recovery are higher. Early detection systems can also alleviate the burden on healthcare infrastructure by allowing patients to be triaged more effectively, prioritizing those who need immediate medical attention. Ultimately, this study underscores the potential of machine learning and feature selection techniques to revolutionize cervical cancer screening and contribute to better health outcomes for women worldwide.

Acknowledgment

The authors would like to thank the Ministry of Research and Technology of Indonesia for funding this research under the Research and Development Grant No. 108/E5/PG.02.00.PL/2024.

References

[1] T. L. Ersado, "Cervical Cancer Prevention and Control," in Cervical Cancer - A Global Public Health Treatise, R. Rajkumar, Ed., Rijeka: IntechOpen, 2021. doi: 10.5772/intechopen.99620.

[2] K. Canfell, "Towards the global elimination of cervical cancer," Papillomavirus Res., vol. 8, p. 100170, Dec. 2019, doi: 10.1016/j.pvr.2019.100170.

[3] M. Arbyn et al., "Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis," Lancet Glob. Heal., vol. 8, no. 2, pp. e191–e203, Feb. 2020, doi: 10.1016/S2214-109X(19)30482-6.

[4] M. M. Ali et al., "Machine learning-based statistical analysis for early stage detection of cervical cancer," Comput. Biol. Med., vol. 139, p. 104985, 2021, doi: https://doi.org/10.1016/j.compbiomed.2021.104985.

[5] R. Alsmariy, G. Healy, and H. Abdelhafez, "Predicting Cervical Cancer using Machine Learning Methods," Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 7, pp. 173–184, 2020, doi: 10.14569/IJACSA.2020.0110723.

[6] A. AlMohimeed, H. Saleh, S. Mostafa, R. M. A. Saad, and A. S. Talaat, "Cervical Cancer Diagnosis Using Stacked Ensemble Model and Optimized Feature Selection: An Explainable Artificial Intelligence Approach," Computers, vol. 12, no. 10, 2023, doi: 10.3390/computers12100200.

[7] H. Momeni and A. Ebrahimkhanlou, "High-dimensional data analytics in structural health monitoring and non-destructive evaluation: a review paper," Smart Mater. Struct., vol. 31, no. 4, p. 43001, Mar. 2022, doi: 10.1088/1361-665X/ac50f4.

[8] M. Hasan, P. Roy, and A. M. Nitu, "Cervical Cancer Classification using Machine Learning with Feature Importance and Model Explainability," in 2022 4th International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), 2022, pp. 1–4. doi: 10.1109/ICECTE57896.2022.10114548.

[9] K. Alpan, "Performance Evaluation of Classification Algorithms for Early Detection of Behavior Determinant Based Cervical Cancer," in 2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), IEEE, Oct. 2021, pp. 706–710. doi: 10.1109/ISMSIT52890.2021.9604718.

[10] I. Jain, V. K. Jain, and R. Jain, "Correlation feature selection based improved-Binary Particle Swarm Optimization for gene selection and cancer classification," Appl. Soft Comput., vol. 62, pp. 203–215, Jan. 2018, doi: 10.1016/j.asoc.2017.09.038.

[11] A. H. Gandomi, F. Chen, and L. Abualigah, "Machine Learning Technologies for Big Data Analytics," Electronics, vol. 11, no. 3, p. 421, Jan. 2022, doi: 10.3390/electronics11030421.

[12] A. H. Elmi, A. Abdullahi, and M. A. Bare, "A comparative analysis of cervical cancer diagnosis using machine learning techniques," Indones. J.

Electr. Eng. Comput. Sci., vol. 34, no. 2, pp. 1010–1023, 2024, doi: 10.11591/ijeecs.v34.i2.pp1010-1023.

[13] G. Ou, Y. He, P. Fournier-Viger, and J. Z. Huang, "A Novel Mixed-Attribute Fusion-Based Naive Bayesian Classifier," Appl. Sci., vol. 12, no. 20, p. 10443, Oct. 2022, doi: 10.3390/app122010443.

[14] D. A. Anggoro and N. C. Aziz, "Implementation of K-Nearest Neighbors Algorithm for Predicting Heart Disease Using Python Flask," vol. 62, no. 9, 2021, doi: 10.24996/ijs.2021.62.9.33.

[15] N. S. Rahmi, N. W. S. Wardhani, M. B. Mitakda, R. S. Fauztina, and I. Salsabila, "SMOTE Classification and Random Oversampling Naive Bayes in Imbalanced Data : (Case Study of Early Detection of Cervical Cancer in Indonesia)," in 2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA), 2022, pp. 1–6. doi: 10.1109/ICITDA55840.2022.9971421.

[16] T. Ganguly, P. B. Pati, K. Deepa, T. Singh, and T. Özer, "Machine Learning based Comparative Analysis of Cervical Cancer Risk Classifications Algorithms," in 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), 2023, pp. 1–7. doi: 10.1109/ACCAI58221.2023.10200617.

[17] B. Nithya and V. Ilango, "Machine Learning Aided Fused Feature Selection based Classification Framework for Diagnosing Cervical Cancer," in 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), 2020, pp. 61–66. doi: 10.1109/ICCMC48092.2020.ICCMC-00011.

[18] S. Tawalbeh, H. Alquran, and M. Alsalatie, "Deep Feature Engineering in Colposcopy Image Recognition: A Comparative Study," Bioengineering, vol. 10, no. 1, 2023, doi: 10.3390/bioengineering10010105.

[19] Q. S. Setiawan, Z. Rustam, and J. Pandelaki, "Comparison of Naive Bayes and Support Vector Machine with Grey Wolf Optimization Feature Selection for Cervical Cancer Data Classification," in 2021 International Conference on Decision Aid Sciences and Application (DASA), 2021, pp. 451–455. doi: 10.1109/DASA53625.2021.9682329.

[20] Z. Ramzan, M. A. Hassan, H. M. S. Asif, and A. Farooq, "A Machine Learning-Based Self-Risk Assessment Technique for Cervical Cancer," Curr. Bioinform., vol. 15, pp. 1–18, 2020, doi: 10.2174/1574893615999200608130538.

[21] N. R. Haddaway et al., "Eight problems with literature reviews and how to fix them," Nat. Ecol. Evol., vol. 4, no. 12, pp. 1582–1589, Oct. 2020, doi: 10.1038/s41559-020-01295-x.

[22] J. J. Tanimu, M. Hamada, M. Hassan, and S. Yusuf Ilu, "A Contemporary Machine Learning Method for Accurate Prediction of Cervical Cancer," SHS Web Conf., vol. 102, p. 04004, May 2021, doi: 10.1051/shsconf/202110204004.

[23] C. H. Bhavani, C. Sarada, A. J. Babu, G. Kumar, and M. Sangeetha, "NGBFA Feature Selection Algorithm-based Hybrid Ensemble Classifier to Predict Cervical Cancer," Int. J. Intell. Syst. Appl. Eng., vol. 12, no. 4, pp. 960–967, 2024, [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/6318

[24] D. A. Anggoro and W. Supriyanti, "Improving accuracy by applying Z-score normalization in linear regression and polynomial regression model for real estate data," Int. J. Emerg. Trends Eng. Res., vol. 7, no. 11, 2019, doi: 10.30534/ijeter/2019/247112019.

[25] D. A. Anggoro, A. A. T. Marzuki, and W. Supriyanti, "Classification of Solo Batik patterns using deep learning convolutional neural networks algorithm," TELKOMNIKA (Telecommunication Comput. Electron. Control., vol. 22, no. 1, pp. 232–240, 2024, doi: 10.12928/telkomnika.v22i1.24598.

[26] W. Supriyanti and D. A. Anggoro, "Classification of Pandavas Figure in Shadow Puppet Images using Convolutional Neural Networks," Khazanah Inform. J. Ilmu Komput. dan Inform., vol. 7, no. 1, pp. 18–24, 2021, doi: 10.23917/khif.v7i1.12484.

Intelligent System for Stability Assessment of Chest X-Ray Segmentation Using Generative Adversarial Network Model with Wavelet Transforms

Omar El Mansouri[1], Mohamed Ouriha[2], Wadiai Younes[3], Yousef El Mourabit[4], Youssef El Habouz[5], Boujemaa Nassiri[6]

TIAD Laboratory, Sciences and Technology Faculty, Sultan Moulay Slimane University, Beni Mellal, Morocco[1, 2, 4]
Laboratory of Innovative Systems Engineering-National School of Applied Sciences of Tetouan, Abdelmalek Essaadi University, Tetouan, Morocco[3]
IGDR, UMR 6290 CNRS, Rennes University, Rennes, France[5]
InterDisciplinaire Applied Research Laboratory- LIDRA International University of Agadir -Unversiapolis, Agadir, Morocco[6]

***Abstract*—Accurate segmentation of chest X-rays is essential for effective medical image analysis, but challenges arise due to inherent stability issues caused by factors such as poor image quality, anatomical variations, and disease-related abnormalities. While Generative Adversarial Networks (GANs) offer automated segmentation, their stability remains a significant limitation. In this paper, we introduce a novel approach to address segmentation stability by integrating GANs with wavelet transforms. Our proposed model features a two-network architecture (generator and discriminator). The discriminator differentiates between the original mask and the mask generated after the generator is trained to produce a mask from a given image. The model was implemented and evaluated on two X-ray datasets, utilizing both original images and perturbed images, the latter generated by adding noise via the Gaussian noise method. A comparative analysis with traditional GANs reveals that our proposed model, which combines GANs with wavelet transforms, outperforms in terms of stability, accuracy, and efficiency. The results highlight the efficacy of our model in overcoming stability limitations in chest X-ray segmentation, potentially advancing subsequent tasks in medical image analysis. This approach provides a valuable tool for clinicians and researchers in the field of medical image analysis.**

Keywords*—Deep learning; X-rays; segmentation; medical imaging; Generative Adversarial Networks; wavelet transforms

I. Introduction

In recent years, the application of deep learning (DL) has seen significant strides in the field of medical imaging, revolutionizing the automation and enhancement of tasks such as detection, classification, and segmentation of medical images. At the forefront of these advancements are neural networks, sophisticated machine learning algorithms designed with layers of interconnected neurons. This architectural mimicry of the human brain enables neural networks to interpret incoming data, extracting intricate patterns and features with a level of sophistication that traditional algorithms struggle to achieve [1, 2].

Neural networks are trained on annotated medical images to detect anatomical structures and abnormalities [1]. Neural networks can be used to automate disease detection, such as identifying lung nodules on chest X-rays or brain tumors on MRI images. It's an effective tool to improve the quality of medical images, with backpropagation function which removes the noise and enhance image resolution. Neural networks can detect specific anatomical structures in images, for example, to isolate a tumor in an MRI image or to measure the dimensions of an organ in an ultrasound image [1, 2]. Deep learning is widely used in medical image analysis to analyze images from a variety of modalities, such as Computed Tomography (CT), X-ray, Positron Emission Tomography (PET), Ultrasound, Magnetic Resonance Imaging (MRI), Optical Coherence Tomography (OCT) [3, 6, 7]. The X-rays are considered as a type of medical imaging that uses ionizing radiation to produce images of bones and other dense body structures. X-rays are commonly used for diagnosis and treatment planning for conditions such as fractures, joint problems, dental issues and chest [3]. They are fast, widely available, and inexpensive compared to other medical imaging modalities.

In this article, we used in our experiments the chest X-rays (CXRs) [3], a type of medical imaging that employs X-rays to create images of the chest, including the heart, lungs, and blood arteries. They are frequently employed to identify and keep track of illnesses like pneumonia, lung cancer, tuberculosis, heart issues, and fluid retention in the lungs. Compared to other imaging techniques used in medicine, CXRs are quick, accessible, and affordable [4]. The detection of COVID-19 has also been tried in this manner [4]. Chest X-rays (CXRs) are used more frequently for early triage of ARDS patients who are already exhibiting COVID-19 symptoms and acquiring accurate because they are simpler to obtain than computed tomography (CT) [5].

Segmentation of a Chest X-ray is the process of separating the foreground (such as the lung and heart) from the background (such as the chest wall) in a chest X-ray image. This can be done using various techniques such as thresholding, morphological operations, and machine learning algorithms [6]. The purpose of chest X-ray segmentation is to accurately identify and isolate specific structures in the image,

which can be used for diagnosis, treatment planning, and measurement of structures such as lung volume. Segmenting chest X-rays is a challenging task due to the diverse range of anatomy, disorders, and appearances that can be present. However, it is a crucial step for enhancing the accuracy and efficiency of medical image analysis [8]. On the other hand, the stability of images segmentation is very important to improve the data security against attackers on medical images field. Images contain sensitive and confidential patient information, and their unauthorized access or manipulation can have serious consequences [9]. Segmentation of medical images, such as chest X-rays, can help improve their security against attackers by reducing the amount of data that needs to be protected. By isolating relevant structures and removing unnecessary information, the risk of unauthorized access or manipulation of sensitive and confidential patient information can be reduced. To further protect the segmented images, encryption can be used to scramble the data using secure algorithms. This makes it difficult for unauthorized users to access or view the confidential information contained in the images. Access to segmented images can also be controlled through authentication and authorization procedures, such as role-based access control or password protection. This helps ensure that only authorized users have access to the data, further reducing the risk of data breaches.

It's important to note that while these techniques can provide some level of security against attackers, no single method can provide complete security. A combination of these techniques, along with robust security protocols throughout the life cycle of medical X-ray images, is recommended for the best results [10]. Deep learning algorithms have shown promising results in both segmentation tasks and security against potential attackers, particularly for chest X-rays.There are several deep learning algorithms used for medical image segmentation, including Convolutional Neural Networks (CNNs). [10] CNNs are a type of neural network specifically designed for image analysis tasks. They can be trained to perform image segmentation by learning the relationships between image pixels and object boundaries. CNNs have achieved notable success in semantic segmentation in [11], Long et al. Semantic segmentation was suggested in early work on integer convolutional networks (FCNs). Where the authors have tested the label map using deconvolutional layers to acquire classification results for each pixel, they have substituted the conventional fully connected CNN layers with convolutional layers to generate a coarse label map. However, CNNs have some drawbacks, they present good results with clear and well-defined images and poor performance on security applications, typical low-resolution, noisy, or occluded images [11]. The U-Net algorithm [12], a well-known CNN architecture created expressly for medical image segmentation, is the alternative segmentation algorithm. In order to record both high-level and low-level characteristics in the image, it employs an encoder-decoder architecture.

II. Related Work

The section provides an overview of existing research, methodologies, and advancements relevant to this study. By examining prior contributions, this section highlights the current state of knowledge, identifies gaps in the literature, and demonstrates how this work builds upon or diverges from earlier studies. The review also contextualizes this research within the broader academic discourse, ensuring clarity in its contributions and alignment with ongoing developments in the field. In recent years, medical image segmentation has become a critical area of research in medical imaging and computer vision. Numerous studies have proposed innovative methods to improve the accuracy and reliability of segmentation models, especially for challenging tasks involving small, unclear, or complex anatomical structures. The U-Net architecture for medical image segmentation is extended in the work for auteur [12], introducing an attention mechanism to enhance performance on challenging objects like the pancreas. In order to enable the network to recognize key features for the target item, the proposed Attention U-Net employs a gating mechanism to dynamically weight feature mappings in the network's encoding and decoding sections. For a dataset of CT scans used for pancreatic segmentation, the Attention U-Net performed better than the regular U-Net and a number of cutting-edge segmentation techniques. Auteur [12] presents, a deep learning framework for U-Net architecture-based segmentation of chest X-ray images. To more effectively handle the present small and hazy structures, the authors suggested changes to the basic U-Net architecture. For improving the accuracy and realism of the segmentation results, they used adversarial segmentation, it's a type of image segmentation technique that uses adversarial training. Adversarial training is a machine learning technique that trains a model using adversarial examples, which are inputs designed to mislead the model. In the context of image segmentation, adversarial training is used to guide the segmentation model towards producing accurate and visually plausible segmentation masks. The model is trained using a combination of a segmentation loss and an adversarial loss, which measures the realism of the generated segmentation masks. Authors of [13], explore the application of SegAN for medical image segmentation and suggest using adversarial training to increase the segmentation accuracy and realism of the segmentation masks created. Two parts make up SegAN: a generator and a discriminator. The segmentation masks are created by the generator, and the discriminator assesses how realistic they are. In order to increase the segmentation accuracy, the authors also suggest a multi-scale loss function that takes into consideration background data from several scales. In order to train the SegAN, the multi-scale loss function is combined with the adversarial loss. For numerous medical imaging datasets, experimental results demonstrate that SegAN is superior to other cutting-edge approaches. Additionally, the authors demonstrate how SegAN can deal with small and hazy structures. The authors of [14] demonstrate how SegAN can be used for unsupervised adversarial training on medical image segmentation. Using a number of medical imaging datasets, they assess the performance of the unsupervised SegAN and demonstrate that it can produce results that are on par with or even superior to those of the conventional supervised SegAN.

The paper in [15], presents an interesting extension of SegAN by incorporating unsupervised adversarial training, which can help address the limitations of supervised adversarial training, such as the need for large amounts of annotated data. However, more research is needed to fully

understand the benefits and limitations of unsupervised adversarial training for medical image segmentation. While the unsupervised adversarial training approach has advantages, it comes with limitations. In this method, the generator is trained without using ground truth masks, potentially resulting in suboptimal performance. This differs from supervised adversarial training, where ground truth masks are employed. Generative Adversarial Networks (GANs) pose challenges in training and can experience stability issues, such as mode collapse, leading to suboptimal performance. The authors evaluated the performance of unsupervised SegAN on various medical imaging datasets. However, it's important to note that these datasets may not fully represent all medical imaging modalities and applications. This version maintains the key points while presenting them in shorter, more digestible sentences. Authors in study [16], present a new approach to medical image segmentation that leverages the advantages of both multimodal imaging and adversarial learning. The authors propose a multi-modal GAN called SegAN, which is trained using both modalities and adversarial learning. The network is trained to segment the target structures in both modalities while maintaining the multi-modal information in the generated results. However, the limitations of the proposed approach include the need for large amounts of annotated data for training, the stability issues associated with GANs, and the lack of interpretability of the model's predictions. In study [17], to improve the performance of adversarial networks for medical image segmentation tasks, the authors propose to use self-supervised learning, where the model can learn from the input data without requiring manual annotations, The SS-GAN is trained using a combination of adversarial loss and self-supervised loss, which helps the network to generate accurate and plausible segmentation mask. The experimental results demonstrate that the proposed SS-GAN's more effective than other state-of-the-art methods for medical image segmentation. The limitations are not specified in the article [18]. However, some common limitations of self-supervised learning and adversarial networks can be the model's performance might be affected by the presence of outliers or data with different distributions, the model may not perform well on unseen data and generalize poorly to new medical imaging modalities, the stability issues associated with GANs, and the lack of interpretability of the model's predictions.

A new paradigm for medical image segmentation employing an adversarial attention network is proposed by S. Kim et al. in [19] (AAN). The AAN enhances segmentation performance by combining adversarial learning with an attention mechanism. Within the network, both a generator and a discriminator coexist. The generator divides the image into segments, while the discriminator assesses the segmentation output. The generator is given access to the attention mechanism to aid with segmentation accuracy and focus on the target regions. The technique was put to the test on several medical image datasets, and the results indicated enhanced performance when compared to other segmentation techniques already in use. X. Song et al. [20], proposes a multiple adversarial network (MAN) architecture for medical image segmentation. The MAN architecture is made up of a number of adversarial sub-networks, each of which is trained to provide various aspects of the input medical image. These features are then combined to give the segmentation result. The results of the experiments the authors undertake on two medical image segmentation datasets show that the MAN architecture is superior to other cutting-edge techniques. The paper [21], does not provide any constraints. However, some typical drawbacks of self-supervised learning and adversarial networks include the instability of GANs, the unintelligibility of the model's predictions, and the model's performance being impacted by the existence of outliers or data with different distributions. In study [22], authors presents a decision support system based on a GAN model for brain tumor segmentation. The authors aim to improve the accuracy and efficiency of brain tumor segmentation using GANs. The proposed system is tested and evaluated using a dataset of brain MRI scans, and the results demonstrate its effectiveness and superiority compared to traditional methods. Some common limitations of the proposed system include stability issues associated with adding noise to the images. The model may not perform well when incorporating such noise.

Through our review of several studies focusing on the segmentation by GAN's algorithms, we discovered that the biggest limitations was the stability constraints related to the GAN's network. In this paper, we proposed a new approach to chest X-ray segmentation based on the GAN with wavelet transforms. The GAN consists of two neural networks: the generator and the discriminator. The generator is responsible for generating a mask for a given original chest X-ray image, while the discriminator distinguishes between the original mask and the generated mask. To evaluate the proposed approach, we implemented their model on two datasets of chest X-ray images. One dataset consisted of original images, while the other dataset consisted of perturbed images with added noise using the Gaussian noise method [38]. The performance of the proposed approach was compared to that of the traditional GAN algorithm. The results showed that the proposed approach based on the GAN algorithm with wavelet transforms outperformed the traditional GAN algorithm in terms of stability, accuracy, efficiency, and it can help to improve the accuracy of subsequent medical image analysis tasks. The proposed approach has the potential to be a valuable tool for clinicians and researchers in the field of medical image analysis. It can help improve the accuracy and efficiency of medical image analysis tasks, which can ultimately lead to better diagnosis and treatment outcomes for patients. The rest of this paper is organized as follows: In Section III, we show methods. The proposed approach of our work in Section IV. The experimental results are discussed in Section V. Finally, a conclusion are presented in Section VI.

III. Methodology

This section, provides an overview of the GAN architecture. Subsequently, we will introduce two models for GAN-based segmentation, with the first model being our initial approach and the second model being a novel technique to overcome stability limitations of GAN.

A. Generative Adversarial Networks (GAN)

Goodfellow et al. [23] were the ones who first proposed the traditional GAN architecture. When creating new synthetic data that is similar to a training set, GANs combine two

separate types of neural networks generator G and a discriminator D. The role of the D is to learn how to distinguish between genuine and fake samples, whereas the G is responsible for generating artificial samples. The min-max game involves two networks competing against each other, where one network aims to maximize the value function 1 while the other seeks to minimize it. In a game-theoretic framework, the two networks are trained concurrently while the generator strives to generate samples that deceive the discriminator, and the discriminator tries to properly identify the generated samples. The discriminator gets better at spotting fraudulent samples, and the generator gets better over time at producing synthetic samples that are close to the training set.

Below, Formula 1 present an illustration of how network G and network D can be understood mathematically with value function V (G, D):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{Z\sim p_Z(Z)}[\log(1-D(G(Z)))] \quad (1)$$

where:

- $\mathbb{E}_{x\sim p_{data}(x)}[\log D(x)]$: log probability of D predicting that real-world data is real.
- $\mathbb{E}_{Z\sim p_Z(Z)}[\log(1-D(G(Z)))]$: log probability of D predicting that G's generated data is not real.
- x: with real images drawn from $p_{data}(x)$ real data images distribution.
- Z: the prior input noise from $p_Z(Z)$.

GANs can be used for a variety of tasks, like image synthesis by use type of GAN called conditional-GAN (cGAN) [24], text to image generation [24], Data augmentation [25], classification, and segmentation [39].

B. GAN for Image Segmentation

The discriminator network is trained to differentiate between real and fake (mask segmented) masks, while the generator network is trained to create realistically segmented images known as mask segmented. The two networks play a game in which the generator tries to produce masks that the discriminator cannot distinguish from real ones, and the discriminator tries to correctly identify the generated masks. Over time, the generator improves, producing more and more realistic masks, while the discriminator becomes better at identifying generated masks. In this way, GANs can be used for image segmentation by training the generator to produce accurate segmentations of input images. The general structure of the GAN for segmentation is shown in Fig. 1.

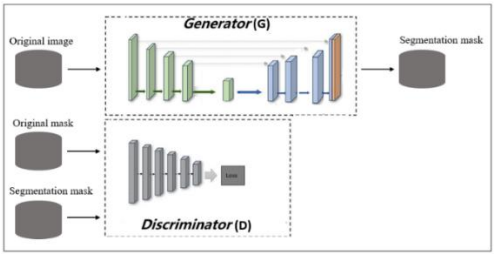


Fig. 1. Architecture GAN.

a) Generator: Our goal is to enhance the preliminary: segmented output of the segmentation component by leveraging the generative component, which is known for its advanced optimization capabilities. To achieve this, we employed the U-Net neural network as the generator, given its exceptional performance in previous works [23]. A U-Net generator for segmentation is a type of deep learning neural network architecture designed for image segmentation tasks. The U-Net architecture is based on a fully convolutional network, with a "U" shaped structure that merges feature from lower-level to higher-level layers in the network [12]. In the U-Net architecture, the contracting path, represented by the left half of the "U" shape called encoder, is used to extract features and reduce the spatial resolution of the input image, and the expanding path, represented by the right half of the "U" shape called decoder, is used to upsample and recover the spatial resolution, while also combining features from the contracting path to make segmentation predictions [12]. The U-Net generator is commonly used in medical imaging to segment organs, tumors, or other structures of interest. However, it can also be applied to other types of image data, such as satellite imagery or aerial photographs, for tasks such as object detection and semantic segmentation. The general structure of the generator in our model is shown in Fig. 2.

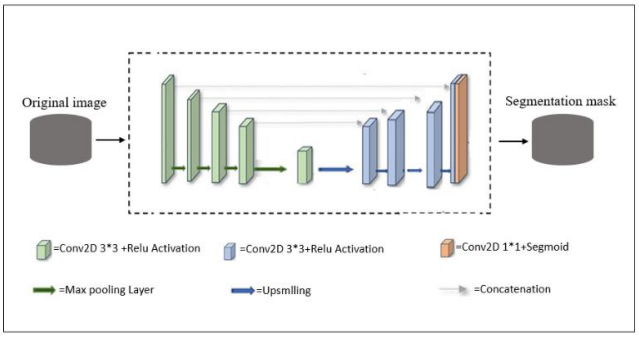


Fig. 2. Structure of the generator.

The network is composed of five layers for the encoder and decoder: first one is the input layer, next three layers of the Relu convolutional batch normalization model are hidden, and the last one is the final output layer. The architecture is comprised of a total of 10 layers, with the encoder and decoder each containing five layers. The encoder consists of five layers, with each layer being composed in the following manner: The first layer contains an input size of $512 * 512$, Conv2D is a 2D convolution layer with a $3 * 3 * 32$ filter size for identifying spatial patterns in an image, such as edges and activation the max pooling layer of size $2 * 2$ is used to minimize representation size and speed up computations in Relu, a piecewise linear function that outputs the input directly if it is positive. With a multiplication filter $*2$, the remaining layers share the same structure. The decoder constitutes the latter half of the architecture, comprising of five layers. Each compound layer is structured in the following manner: Transposition-based deconvolution is an upsampling technique used to increase the size of an image. To achieve this, the transposed convolution is applied, and the resulting image is concatenated with the corresponding image from the contracted path. This

process creates an image of the same size as the input. In the Conv2D class, the padding parameter can have one of two values: "valid" or "same". When set to "valid", convolution can be applied to reduce the spatial dimensions of the input volume, assuming that it is not empty. When set to "same", convolution can be applied to reduce the spatial dimensions, and the input volume is assumed not to be empty. Including prior information in the process increases the accuracy of the output. In our work, we utilized "same" padding, and the final layer consists of a convolutional layer with a filter size of 1 × 1. The output of the sigmoid activation function is specified on real numbers and ranges between 0 and 1. It can be interpreted as a probability value [26].

b) Discriminator: Within our work, the discriminator: is implemented as a deep convolutional neural network that takes in two images one that is original and another that is generated by the generator. The objective of the discriminator is to categorize each image as either "real" or "fake". Fig. 3 provides a visual representation of the overall structure of the discriminator that was employed in our work.

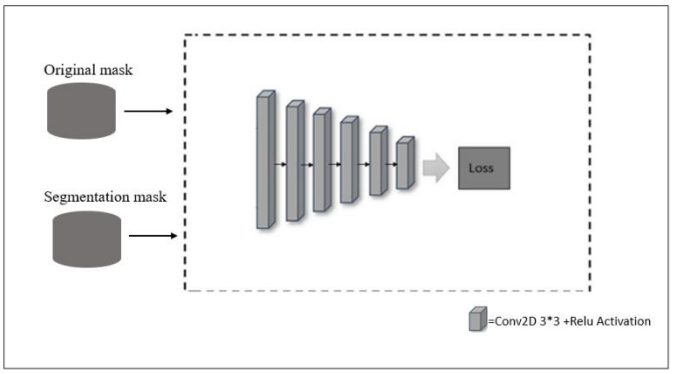


Fig. 3. Structure of the discriminator.

In our work, the discriminator examines two sequences to determine whether they are real or fake by using two images: a original mask and a mask produced by the generator. For the discriminator in our model, we used a total of five layers, starting with the input layer, followed by four Conv2D layers each with a size of 3 × 3. Hidden and dropout layers were incorporated, implementing the Relu activation function and batch normalization. The fifth layer was another Conv2D layer, this time with a size of 3 × 3, utilizing the Relu activation function, and employing max pooling layers. The 2-dimensional arrays resulting from the pooled feature maps are then flattened in the final layer before being passed on as a continuous linear vector to the fully connected layer, which generates the final output [27, 28].

C. Wavelet Transforms

A subfield of mathematics known as the wavelet transform (WT) started to develop gradually in the 1980s [40]. The wavelet transform is another outstanding example of the ideal fusion of pure and applied mathematics, following the fourier transform. It shares the mathematical microscope renown. The wavelet transform has significantly advanced approaches in nonlinear science, engineering technology, signal processing, image processing, and computer applications in recent years. It is one of the most effective and popular time frequency analysis techniques, and it has been utilized extensively in signal and image processing [31, 32]. WT is classified into two types: continuous WT (CWT) and discrete WT (DWT). In general, DWT is more useful than CWT for resolving practical issues. We used the Discrete Wavelet Transform (DWT) to implement our new approach. The DWT can be used to decompose signals into different frequency components and provide a multiresolution representation of the signal [32]. This can be useful in image processing and computer vision applications, where the goal is to extract important features and remove noise from the signals [29, 32]. The data is split into high and low pass bands by the DWT algorithm, with or without information loss. It is built on high and low pass filters that are sub-sampled. Functions over a finite range are defined as DWT. The goal of DWT is to convert data from the time-space domain to the time-frequency domain in order to improve compression efficiency. We have simplified the DWT by describing how it works as follows in Fig. 4:

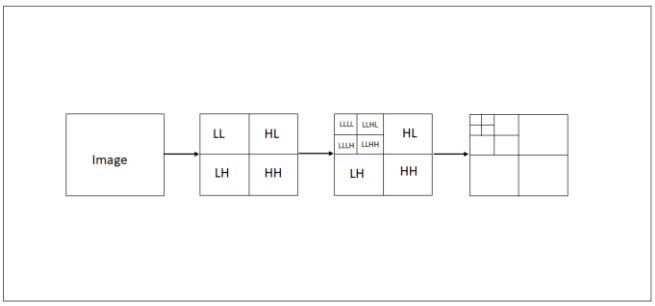


Fig. 4. Block diagram of 3-level wavelet transform.

Low-Low (LL), Low-High (LH), High-Low (HL), and High-High are the names of the four bands that are created (HH). As the LL band still contains information that resembles an image, it is possible to use the same set of wavelet filters that were used on the original image. The image can be divided into sub bands for as many levels as needed (up to the image resolution), however for image compression, only 4 or 5 levels are often used [33]. An image is a two-dimensional (2-D) signal with rows and columns. The wavelet decomposition of an image (rows and columns) can be represented by decomposing one-dimensional signals. The low-frequency components in horizontal and vertical directions (LL), low-frequency components in horizontal and vertical directions (LH), high-frequency components in horizontal and vertical directions (HL), and high-frequency components in horizontal and vertical directions (HH) are obtained following a one-layer wavelet decomposition of an image [33]. The DWT performs two steps of calculation on pairs of data items from the signal after it has been processed, as presented on following Formula 2.

$$\begin{cases} L_i = \frac{1}{\sqrt{2}}(X_{2i} + X_{2i+1}) \\ H_i = \frac{1}{\sqrt{2}}(X_{2i} - X_{2i+1}) \end{cases} \quad (2)$$

An approximation (average) of the two data items is what comes out of the first step, and an approximation (difference) of the two data items is what comes out of the second step.

Where :

- i=0...N/2
- X=the signal data of length N.
- L=the low subband of length N.
- H=the high subband of length N.

The inverse Wavelet transform (IDWT), in contrast, fuses four sub-images to the original image by up-sampling while utilizing the same filters. As shown on Formula 3.

$$\begin{cases} X_{2i} = \frac{1}{\sqrt{2}}(L_i + H_i) \\ X_{2i+1} = \frac{1}{\sqrt{2}}(L_i - H_i) \end{cases} \quad (3)$$

a) Wavelet transforms for generative adversrial network: The GAN-based image segmentation model: that incorporates wavelet transforms is a deep learning architecture that combines two powerful techniques to achieve accurate segmentation results. In this approach, the wavelet coefficients of the input image are used as features to train a GAN model, which generates a segmented image output. The generator network is designed to optimize the segmentation performance by minimizing the difference between the generated output and the ground truth segmentation mask, while the discriminator network objective is to distinguish between the generated segmentation mask and the ground truth. By iteratively training the generator and discriminator, the GAN model can learn to accurately segment complex images [30]. A GAN based on Wavelet coefficients can help improve the stability of image segmentation by using the wavelet coefficients as inputs to the network. The wavelet transform allows the decomposition of an image into different frequency components, which can provide a more robust representation of the image compared to using raw pixel values. By training the GAN on these wavelet coefficients, it can learn to produce stable segmentations even with variations in the input image. This approach can help improve the robustness and generalization ability of the image segmentation model, it's that's what we actually got by implementing our new approach based on a GAN with a wavelet and comparing it to a simple GAN.

IV. Proposed Approach

Using DWT and IDWT in our approach, we were able to develop a reliable segmentation model that could handle variations in input images. Our results indicate that our approach enhanced the robustness and generalizability of the original image segmentation model. Unlike the original generator that used convolutions to transform its 3 channel input image, Fig. 2. We used convolutions to convert the original images into 12-channel and 16-channel data within the generator, along with DWT and IDWT. We then employed convolutions iteratively, replacing max pooling and upsampling with DWT and IDWT, respectively. By adopting DWT, we could retrieve multiscale edge characteristics while reducing the data by a fourth and multiplying the number of channels by four. We adjusted the convolution stride to reduce the data size, unlike the original method with max pooling, which discards partial original data and loses minor edge features. We also substituted DWT for the sampling operations max pooling to enhance the segmentation capabilities of our discriminator. Fig. 5 illustrates the general structure of our generator using the wavelet transform.

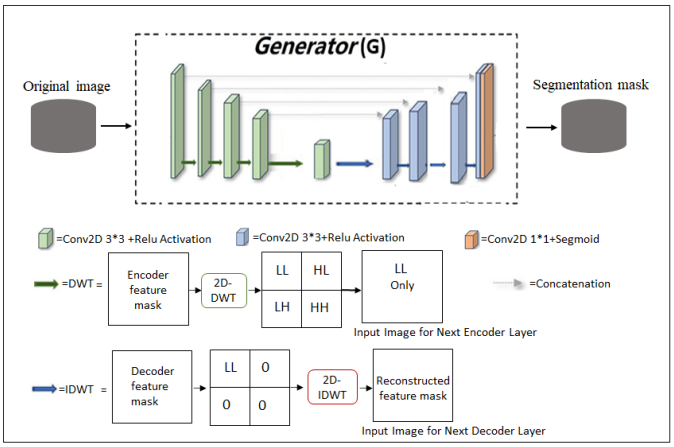


Fig. 5. Generator with wavelet transforms structure.

In our study, we employed two approaches. Firstly, we compared the effectiveness and advanced nature of two models: GAN and GAN with DWT. This evaluation aimed to assess the performance and capabilities of the proposed approach. Secondly, we introduced noise to the testing data and compared the performance of the two models to address stability limitations. This comparison was conducted to analyze the models' ability to handle and mitigate the effects of noise.

A. Evaluation Metrics

In our work, we used three metrics below to assess the effectiveness of our approach.

1) Accuracy [34]: is the ratio of correct predictions to the total number of predictions made by the classifier and can be found using Eq. (4).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (4)$$

Where:

- TP (True positive): The model predicts an already present object.
- FP (False positive): The model predicts an object that is not actually present.
- FN (False negative): The model is unable to predict an already present object.
- TN (True Negative): The model correctly predicts the negative class.

2) IoU (Intersection over union) [35]: The intersection over union also known as Jaccard is the area of overlap between the predicted mask and the original mask divided by the area of union between the two sets and can be found using Eq. (5).

$$\text{IoU} = \frac{\text{TP}}{\text{TP+FP+FN}} = \frac{Areaof Overlap}{Areaof Union} \quad (5)$$

3) DSC (Dice Score Coefficient) [34]: The Dice Score Coefficient or what is said as the F1-Score is the arithmetic mean of precision and recall and can be used to better assess the segmentation effect and can be found using Eq. (6).

$$\text{DSC} = \frac{2\times\text{P}\times\text{R}}{\text{P+R}} \quad (6)$$

Where:

P is the proportion of samples among the correctly categorized samples that are classed as positive samples.

$$P = \frac{\text{TP}}{TP+FP} \quad (7)$$

R is a measure of how many correctly identified positive samples there are compared to all positive samples.

$$R = \frac{TP}{TP+FN} \quad (8)$$

Simply put, the Dice Coefficient is 2 ∗ the Area of Overlap between the predicted mask and the original mask divided by the total number of pixels in both images.

$$\text{DSC} = \frac{2\times\text{AreaofOverlap}}{\text{Totalnumberofpixels}} \quad (9)$$

B. Datasets

For our experiments, we used two datasets for training and testing:

- The first is the Shenzhen Hospital (SH) dataset [36],

which consists of 662 images, 336 of which are from abnormal individuals displaying various tuberculosis signs, and 326 from healthy individuals. Hospitals in Shenzhen, Guangdong province, China, collected the JPEG formatted X-ray images used in this data set. These segmentation masks for the Shenzhen Hospital X-ray Set were manually created by in structors and students from the National Technological University of Ukraine's Computer Engineering Department of the Faculty of Informatics and Computer Engineering.

- The second, COVID-19 Chest X-ray images and

Lung masks Database [37]. It contains chest X-ray images for COVID-19 positive cases along with Normal and Viral Pneumonia images in cooperation with medical professionals, created by a group of researchers from Qatar University, Doha, Qatar, and the University of Dhaka, Bangladesh, as well as collaborators from Pakistan and Malaysia. The collection includes 2905 chest X-ray (CXR) images with the segmentation mask, 219 of which are from abnormal patients showing COVID-19 positive, 1341 normal, and 1345 images with viral pneumonia.

We scale the images in the training and test sets to 512*512 for training and testing. We use 20% for testing, 10% for validation, and 70% for training. A train-test split was used to divide the data as shown in Table I.

TABLE I. DATASETS

Datasets	Total images	Format	Original Dimension	Training	Testing	Validation
SH	662	JPEG	(512,512,3)	70%	20%	10%
COVID-19	2905	PNG	(512,512,3)			

C. Experimental Setup and Training Parameters

After data enhancement, we trained two GAN models (Traditional GAN and GAN with DWT) for 200 epochs. Experiments were conducted on a server equipped with a GPU P100. The two models we provided for our dataset were trained on 200 epochs with a batch size of 3, and one model's training might take up to 6 hours. The implementation of our work is based on the Keras and Tensorflow libraries. We used the Adam algorithm to optimise the network with a learning rate of 0.0003 and a decay rate of 0.5.

V. Results and Discussion

A. Training Results

We used the two architectures on both the SH dataset and the COVID-19 dataset, and we compared them to evaluate the effectiveness and advanced nature of the proposed approach. The following Fig. 6 and Fig. 7 present the results of the training and validate the Dice score coefficient of our models.

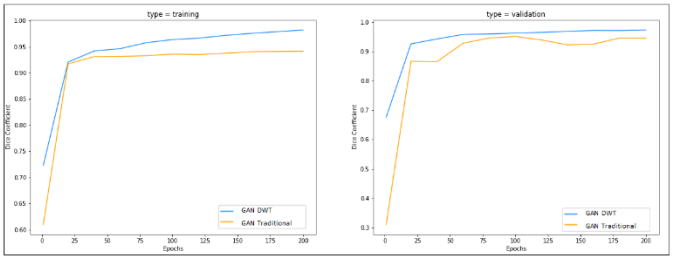


Fig. 6. Training results in the SH dataset.

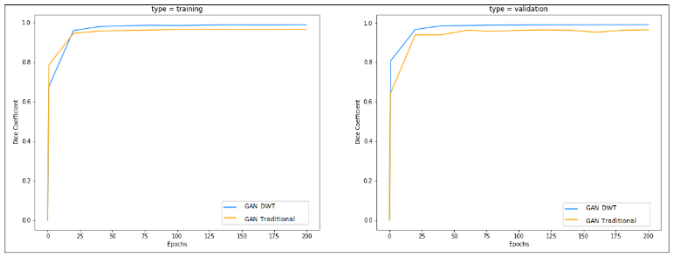


Fig. 7. Training results in the COVID-19 dataset.

It can be seen from the above results. It shows that our algorithm GAN with DWT has better robustness and stability and has given the best performance for the training in two datasets. So, it can be said through our comparison between the two training models in these datasets, a GAN based on Wavelet coefficients can help improve the performance for the training of image segmentation.

B. Stability Results

To see if GAN with DWT prevents the stability limit of a GAN Traditional. We have added multiple ε levels of noise density to the original images in the testing phase. The Table II flow shows the results of testing the dice score coefficient of our models in two datasets at every ε level of noise. To add ε noise into the testing data, we utilized the Gaussian noise method. This approach entailed introducing random values sampled from a Gaussian distribution and adding them to the original data. By employing the Gaussian noise method, we were able to simulate realworld noise sources and introduce random variations characterized by a normal distribution. To apply Gaussian noise to an image, we use the following equation:

$$\mathbf{noisy_{image}} = \mathbf{original_{image}} + \boldsymbol{\varepsilon} \times \mathbf{N(0, sigma^2)} \quad (10)$$

- $\text{noisy}_{\text{image}}$:represents the resulting image after adding Gaussian noise.
- $\text{original}_{\text{image}}$: is the original input image.
- ε : is a scaling factor that controls the intensity of the noise.
- $N(0, \text{sigma}^2)$: represents a random variable drawn from a Gaussian distribution with.

TABLE II. STABILITY ON TESTING PHASE

	SH Dataset		COVID-19 dataset	
ε	GAN DWT	GAN Traditional	GAN DWT	GAN Traditional
0	0.9685	0.9384	0.9906	0.9744
0.10	0.9685	0.0071	0.9904	0.0012
0.20	0.9635	0.0068	0.9904	0.0012
0.30	0.9636	0.0067	0.9900	0.0011

We used the two architectures on both the SH dataset and the COVID-19 dataset, and we compared them to evaluate the effectiveness and advanced nature of the proposed approach. The following Fig. 8 present the results of Stability on testing phase our models.

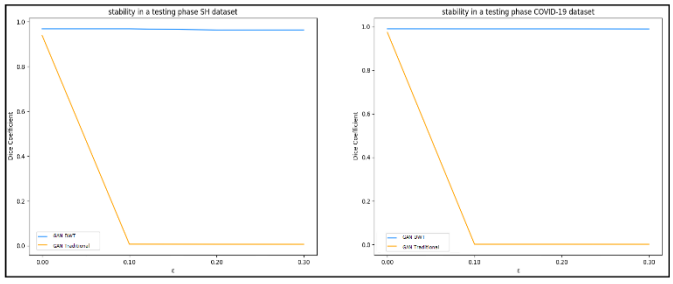


Fig. 8. Stability on testing phase.

According to the results, when adding ε=0.10, 0.20, and 0.30 levels of noise density to the original images in the testing phase. The performance in the dice score coefficient of algorithm GAN with DWT remained stable compared to GAN Traditional. So, it can be said through our comparison between the two models in these datasets, a GAN based on Wavelet coefficients can help to prevent stability limitation in image segmentation of the GAN Traditional. The prediction segmentation results of the two models for datasets SH and COVID-19 for ε levels of noise density added are shown in the Fig. 9 to 16 are as follows:

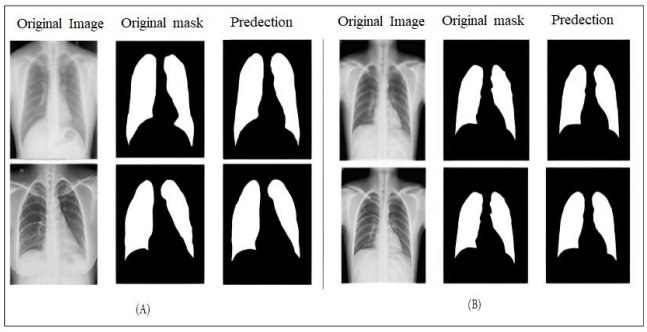


Fig. 9. Results of SH dataset with 0 noise added: (A) GAN with DWT; (B) GAN traditional.

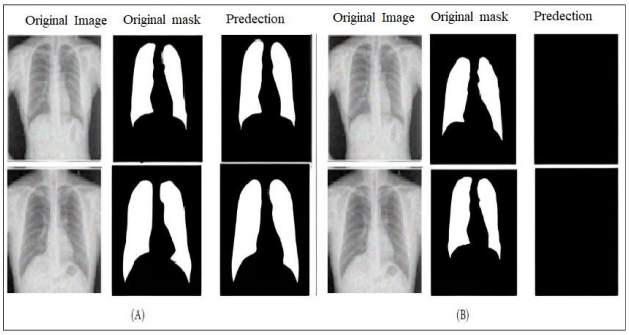


Fig. 10. Results of SH dataset with 0.10 noise added: (A) GAN with DWT; (B) GAN traditional.

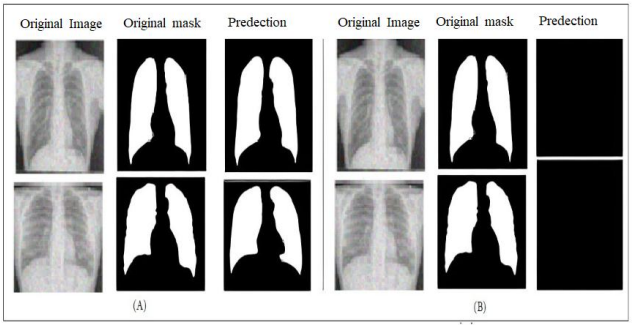


Fig. 11. Results of SH dataset of 0.20 noise added : (A) GAN with DWT; (B) GAN traditional.

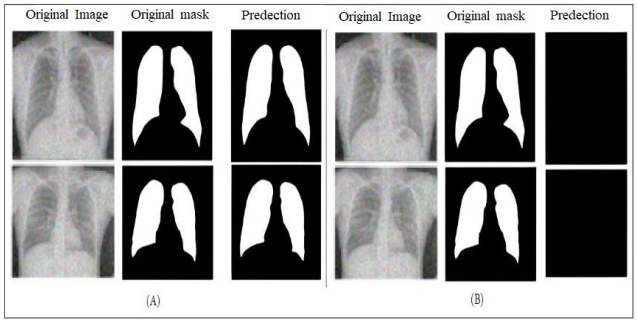


Fig. 12. Results of SH dataset of 0.30 noise added: (A) GAN with DWT; (B) GAN traditional.

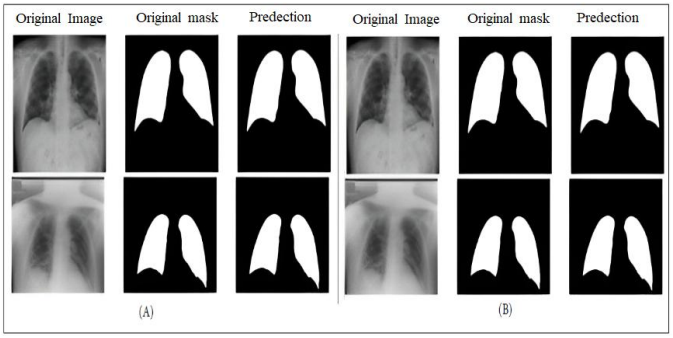


Fig. 13. Results of COVID-19 dataset with 0 noise added: (A) GAN with DWT; (B) GAN traditional.

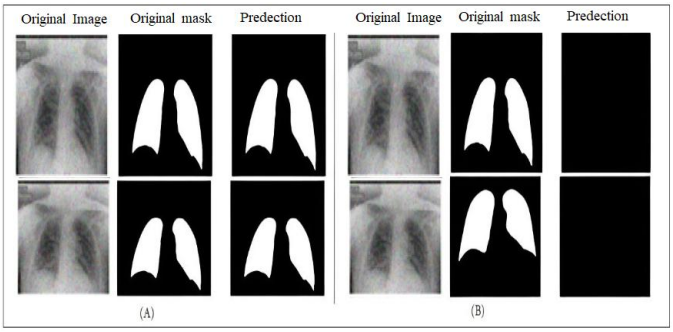


Fig. 14. Results of COVID-19 dataset of 0.10 noise added. (A) GAN with DWT; (B) GAN traditional.

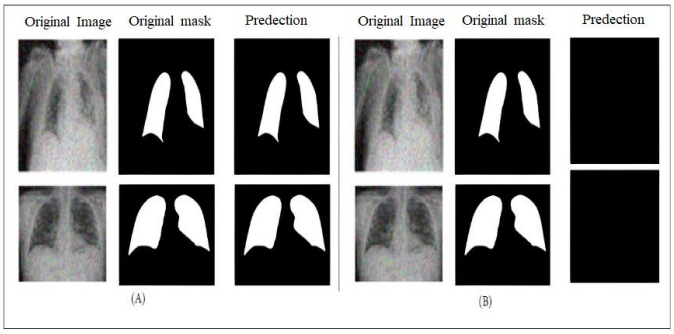


Fig. 15. Results of COVID-19 dataset with 0.20 noise added: (A) GAN with DWT; (B) GAN traditional.

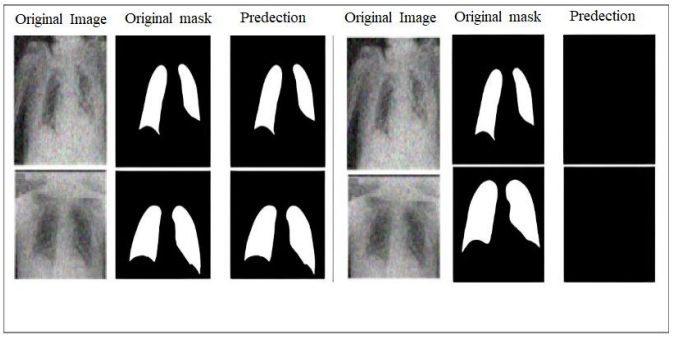


Fig. 16. Results of COVID-19 dataset with 0.30 noise added: (A) GAN with DWT; (B) GAN traditional.

VI. Conclusion

This paper present an innovative segmentation approach that leverages the power of GAN combined with wavelet transforms. The objective was to tackle stability issues commonly encountered in the image segmentation process when using GAN Traditional. By incorporating wavelet transforms into the approach, we aimed to enhance the robustness and reliability of the segmentation results. The wavelet transforms played a crucial role in effectively capturing multi-scale features and spatial details within the images, enabling more accurate segmentation outcomes. Additionally, we recognized that image segmentation with traditional GAN can introduce instability, resulting in artifacts and inaccurate boundaries. To overcome this challenge, our proposed approach offered an alternative approach to pooling that mitigated stability issues, thereby improving the overall quality and consistency of the segmentation results. Our approach results demonstrate the effectiveness of our approach to achieve high segmentation performance compared to existing methods. The integration of GANs, wavelet transforms, and the alternative pooling technique showcased the potential for significant advancements in image segmentation accuracy and stability. The proposed approach not only provided a robust solution to stability challenges but also opened up new avenues for exploring the synergistic benefits of combining GANs, wavelet transforms, and alternative pooling strategies in the image segmentation field.

References

[1] Suganyadevi, S., Seethalakshmi, V. Balasamy, K., 2022. A review on deep learning in medical image analy-sis. Int. J Multimed. Info. Retr. 11, 19-38.

[2] Zhou, S. Kevin, Hayit Greenspan, and Dinggang Shen, eds, 2017: Deep learning for medical image analysis. Academic Press.

[3] Kim, M., Yun, J., Cho, Y., Shin, K., Jang, R., Bae, H.J., Kim, N., 2019. Deep Learning in Medical Imaging. Neurospine. 16(4), 657.

[4] Hendee, W. R., Chien, S., Maynard, C. D., Dean, D. J., 2002. The National Institute of Biomedical Imaging and Bioengineering: history, status, and potential impact. Annals of biomedical engineering, 30, 2-10.

[5] Bassi, P. R., Attux, R. 2021. A deep convolutional neural network for COVID-19 detection using chest X-rays. Research on Biomedical Engineering, 1-10.

[6] Bhattacharya, S., Maddikunta, P. K. R., Pham, Q. V., Gadekallu, T. R., Chowdhary, C. L., Alazab, M., Piran, M. J. 2021. Deep learning and medical image processing for coronavirus (COVID-19) pandemic: A survey. Sustainable cities and society, 65, 102589.

[7] Pirimoglu, B., Sade, R., Ogul, H., Kantarci, M., Eren, S., Levent, A. 2016. How can new imaging modalities help in the practice of radiology?. The Eurasian journal of medicine, 48(3), 213.

[8] Smith, J., Johnson, A., Brown, C., 2022. Segmentation Techniques for Chest X-ray Images. Medical Imaging Journal, 15(3), 123-137.

[9] Jacobi, A., Chung, M., Bernheim, A., Eber, C. 2020. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. Clinical imaging, 64, 35-42.

[10] Wong, H. Y. F., Lam, H. Y. S., Fong, A. H. T., Leung, S. T., Chin, T. W. Y., Lo, C. S. Y., Ng, M. Y. 2020. Frequency and distribution of chest radiographic findings in patients positive for COVID-19. Radiology, 296(2), E72-E78.

[11] Magdy, M., Hosny, K.M., Ghali, N.I. et al. 2022. Security of medical images for telemedicine: a systematic review. Multimed Tools Appl 81, 25101–25145.

[12] Muhammad, K., Hussain, T., Tanveer, M., Sannino, G., de Albuquerque, V. H. C. 2019. Cost-effective video summarization using deep CNN with hierarchical weighted fusion for IoT surveillance networks. IEEE Internet of Things Journal, 7(5), 4455-4463.

[13] Long, J., Shelhamer, E., Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, pp. 3431-3440.

[14] Ronneberger, O., Fischer, P., Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, Springer, pp. 234-241.

[15] Yinglei Liang, Yudong Liu, Junyu Dong, and Wei Cheng (2019). Chest X-Ray Segmentation with a U-Net-Based Deep Learning Framework, de Jinfeng Yang.

[16] Xue, Y., Xu, T., Zhang, H., Long, L. R., Huang, X. 2018. Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. Neuroinformatics, 16, 383-392.

[17] Chen, T., Zhai, X., Ritter, M., Lucic, M., Houlsby, N. 2019. Self-supervised gans via auxiliary rotation loss. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12154-12163).

[18] Nahar, N., Soomro, S., Monrat, A. A. 2021. A GAN based Framework for Multi-Modal Medical Image Segmentation.

[19] Long Chen, Jie Chen, Wei Li, Xiaolei Huang, and Pheng-Ann Heng (2019). ”Adversarial Learning for Multi-modal Medical Image Segmentation with SegAN”.

[20] Zeng, Z., Xulei, Y., Qiyun, Y., Meng, Y., Le, Z. 2019. Sese-net: Self-supervised deep learning for segmentation. Pattern Recognition Letters, 128, 23-29.

[21] Zhao, J., Hou, X., Pan, M., Zhang, H. 2022. Attention-based generative adversarial network in medical imaging: A narrative review. Computers in Biology and Medicine, 105948.

[22] El Mansouri, O., El Mourabit, Y., El Habouz, Y., Boujemaa, N., Ouriha, M. 2022. Intelligent System Based on GAN Model for Decision Support in Brain Tumor Segmentation. In: Fakir, M., Baslam, M., El Ayachi, R. (eds) Business Intelligence. CBI 2022. Lecture Notes in Business Information Processing, vol 449. Springer, Cham.

[23] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y. 2020. Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[24] Mirza, M., Osindero, S. 2014. Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

[25] Sawant, Ronit and Shaikh, Asadullah and Sabat, Sunil and Bhole, Varsha, Text to Image Generation using GAN (July 8, 2021). Proceedings of the International Conference on IoT Based Control Networks & Intelligent Systems.

[26] Sandfort, V., Yan, K., Pickhardt, P.J. et al. 2019. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Sci Rep 9, 16884 .

[27] S. A. Israel et al., Generative Adversarial Networks for Classification, 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 2017, pp. 1-4.

[28] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, 2012. Geoffrey: ImageNet Classification with Deep Convolu-tional Neural Networks. In: Neural Information Processing Systems.

[29] Peng SL, Li BB, Hu XY. Wavelet and filter bank design: theory and application. Beijing: Tsinghua univer-sity press, 2017, pp. 95–220.

[30] Li, Y., Zhang, Y., Zhang, H., Yang, Y. 2020. GAN-Based Image Segmentation Model with Wavelet Transforms.

[31] Ran QW. Wavelet transform and fractional Fourier transform theory and applications. Harbin: Harbin Institute of Technology Press, 2001, pp. 1-3.

[32] Luo Y, Li JP, Cheng LZ, et al. Wavelet packet transform and the design cost function. Math Theory Appl 2011; 31(3): 65-70.

[33] Smith, J., Johnson, A., Brown, C. (2022). Introduction to Wavelet Transforms and Image Compression. Digital Signal Processing Journal, 15(3), 123-137.

[34] Mittal, A.; Hooda, R.; Sofat, S. Lung field segmentation in chest radiographs: A historical review, current status, and expectations from deep learning. IET Image Process 2017, 11, 937–952.

[35] Bosdelekidis, V.; Ioakeimidis, N.S. Lung Field Segmentation in Chest X-rays: A Deformation-Tolerant Procedure Based on the Approximation of Rib Cage Seed Points. Appl. Sci. 2020, 10, 6264.

[36] Jaeger, S.; Candemir, S.; Antani, S.; Wang, Y.X.J.; Lu, P.X.; Thoma, G. Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quant. Imaging Med. Surg. 2014, 4, 475.

[37] M. E. H. Chowdhury, T. Rahman, A. Khandakar, R. Mazhar, M. A. Kadir, Z. B. Mahbub, K. R. Islam, M. S. Khan, A. Iqbal, N. A. Emadi, M. B. I. Reaz, M. T. Islam, 2020, Can AI Help in Screening Viral and COVID-19 Pneumonia, IEEE Access, vol. 8, pp. 132665-132676.

[38] F. Luisier, T. Blu and M. Unser,March 2011. Image Denoising in Mixed Poisson–Gaussian Noise.

[39] XUN, Siyi, LI, Dengwang, ZHU, Hui, et al. 2022, Generative adversarial networks in medical image segmentation: A review. Computers in biology and medicine, vol. 140, p. 105063.

[40] YIN, Ming, LIU, Wei, SHUI, Jun, et al.2012, Quaternion wavelet analysis and application in image denoising. Mathematical Problems in Engineering, 2012, vol.

Real-Time Monitoring and Analysis Through Video Surveillance and Alert Generation for Prompt and Immediate Response

Akshat Kumar[1], Renuka Agrawal[2], Akshra Singh[3], Aaftab Noorani[4], Yashika Jaiswal[5], Preeti Hemnani[6], Safa Hamdare[7]

Department of Computer Science and Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune, India[1, 2, 3, 4, 5]

Department of Electronics and Telecommunication Engineering, SIES Graduate School of Technology, Mumbai, India[6]

Department of Computer Science, Nottingham Trent University, Nottingham, NG11 8NS, UK[7]

***Abstract*—The efficacy of Closed-Circuit Television systems (CCTV) in residential areas is often linked to the lack of real-time alerts and rapid response mechanisms. Enabling immediate notifications upon the identification of irregularities or aggressive conduct can greatly enhance the possibility of averting serious incidents, or at the very least, significantly mitigate their impact. The integration of an automated system for anomaly detection and monitoring, augmented by a real-time alert mechanism, is now a critical necessity. The proposed work presents an advanced methodology for real-time detection of accidents and violent activities, incorporating a sophisticated alarm system that not only triggers instant alerts but also captures and stores video frames for detailed post-event analysis. MobileNetV2 is utilized for spatial analysis due to its computational efficiency compared to other Convolutional Neural Networks (CNN) architectures, while Visual Geometry Group 16 (VGG16) enhances model accuracy, especially on large-scale datasets. The integration of Bi-directional Long Short-Term Memory (BiLSTM) strengthens temporal continuity, significantly reducing false alarms. The proposed system aims to improve both safety and security by enabling authorities to intervene timely to incidents. Combining rapid computation with high detection accuracy, the proposed model is ideally suited for real-time deployment across both urban and residential settings.**

Keywords*—Rapid response; anomaly detection; MobileNetV2; VGG16; BiLSTM

I. Introduction

The growth in crime and the need for protection especially in the residential compounds has led to the installation of CCTV systems in the compound; the need to enhance safety and control of incidents has also boosted the usage of CCTV systems in residential areas. One typical use of these devices is for real-time monitoring and recording of events, which enable the authorities to respond to any emergent incidents such as violent criminal incidents or accidents occurring in premises [1]. The major drawback with the conventional method of monitoring is that several instances go unnoticed even with the increasing use of the CCTV in commercial and residential areas [2, 3]. Manual monitoring system is vulnerable to human mistakes, weariness, and slow response times. Besides this, there is feedback delay and lost opportunities for timely intervention and prompt action required in case of occurrence of incidents. The necessity for automatic real-time anomaly detection is highlighted by the growing number of CCTV systems installed in homes and in residential societies [4, 5]. Delays in emergency response often stem from the late detection or lack of reporting of violent incidents and traffic accidents in residential premises. This underscores the need for a lightweight, scalable system designed to rapidly detect abnormal events and automatically raise notifications to designated authorities. Implementing such a system would enhance civilian safety and significantly improve traffic management efficiency.

Traditional methods of monitoring, which rely on human operators, are time-consuming and often inefficient, as operators may not be able to keep up with multiple camera feeds or notice critical details in the video. As human fatigue sets in, the effectiveness of these systems further diminishes. Given the vast volume of data generated by modern CCTV networks, manual monitoring is no longer feasible. As a result, automated tools for tasks such as object recognition, classification, and anomaly detection have been developed using deep learning models like Residual Networks (ResNet) and Densely Connected Networks (DenseNet). While these models offer high accuracy, they face challenges in real-time applications due to their computational demands. It has been noted that both ResNet and DenseNet have high accuracy, though, the computing complexity of these two models are massively different. Due to the tensed structure and millions of parameters ResNet cannot be used effectively in real-time scenario and requires a huge amount of RAM and computation power. While DenseNet employs few parameters, it calls for several convolutions per layer and thus, has high processing intensity. These aspects make such models unsuitable for real-time applications since in real-life one has to decide based on the available information, which in turn is valuable in cases where the amount of resources is limited, but the decision has to be made as soon as possible. Earlier automated techniques frequently failed to give real-time alerts and relied on computationally heavy and expensive frameworks. Furthermore, earlier automated systems such as ResNet and DenseNet are not suited for real-time use because of their high processing cost [6]. One of them is ResNet, which helps in achieving high accuracy, but makes the model computationally

expensive. ResNet layers utilize typical convolutions, which demand many multiplications and adds [7]. The deep architecture with many layers translates to many parameters. For instance, ResNet-50 has approximately 25 million parameters. The extensive depth and numerous operations in ResNet contribute to high latency time and computation complexity when employed for anomaly detection [8, 9]. DenseNet connects each layer to every other layer in a feed-forward manner, ensuring maximum information flow between layers [10]. The dense connectivity increases the complexity of the network. Despite the efficient parameter consumption achieved through feature reuse, DenseNet still requires a substantial number of convolutions. The necessity to retain and process feature maps from all previous layers raises memory and computing requirements [11].

The proposed methodology addresses these challenges by providing a scalable and efficient solution for real-time violence detection in residential areas. This system uses MobileNetV2, a convolutional neural network for spatial feature extraction, considerably minimizes the computational overhead while maintaining the accuracy compared to previous models [12]. It uses depth wise separable convolutions, which decompose a standard convolution into two simpler operations: depth wise convolution and pointwise convolution. Furthermore, the model uses BiLSTM networks for temporal sequence analysis for violence detection. This enables the model to better comprehend the context and course of events, minimizing false predictions thus increasing the accuracy of violence detection. By focusing on temporal coherence, it is ensured that the model not only accurately recognizes violent acts but also distinguishes them from nonviolent activity [13].

Despite advances in anomaly detection, most existing research focuses solely on detecting accidents or unusual events, often overlooking the crucial step of raising alerts for timely intervention in violent incidents. Many studies primarily enhance the performance of models like CNNs for anomaly detection, but they neglect the integration of alert systems for quick response. For example, Trilles et al. [14] discuss the use of AIoT for anomaly detection but focus on detection rather than alerting. Similarly, Chandrakar et al. [15] improve moving object detection and tracking for traffic surveillance, but their work does not address automated response mechanisms. Ullah et al. [16] combine CNNs and BiLSTM networks for real-time anomaly detection, but their focus is on feature extraction and classification, not alerting systems. Kamble et al. [17] propose a smart surveillance system for anomaly detection but exclude alert mechanisms. Wang et al. [18] utilize DenseNet for anomaly detection, but their work also lacks a real-time alerting system. While these studies focus on detection accuracy, they fail to address the critical need for real-time alerts, a limitation that impedes their practical application in real-world security settings.

The prime objective of the proposed work is to design a lightweight model for detecting similar anomalies in residential areas. Besides an alert mechanism is also incorporated, within the model to raise alert to inform concerned authorities for prompt action and also to mitigate the effect of incidents occurring in residential premises. Hence, to meet the needs that require real-time monitoring, automated systems that can promptly generate the alert to enable authorities respond to occurrences are needed.

This paper aims to compare various models and techniques to determine which offers the best performance for anomaly detection in the form of accidents or violence occurring in residential areas. The rest of this paper is structured as follows: the literature review in Section II, the methodology is detailed in Section III, results findings are shown in Section IV, followed by model comparison and analysis discussion in Section V, and the final conclusion with future research directions are provided in Section VI.

II. Literature Review

The domain of video surveillance and security has seen significant advancements with the use of various AI related Machine Learning or Deep learning methodologies. These technologies have enabled the development of sophisticated systems capable of real-time behavior analysis, object detection, anomaly detection, and activity recognition.

Khan et al. [3] proposed a method combining Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for behavior analysis, achieving 92% accuracy on the UCF101 dataset. While effective for real-time human action recognition, this approach is constrained by the dataset's limited variety and struggles with complex backgrounds. Similarly, Lindemann et al. [19] utilized Long Short-Term Memory (LSTM) networks for activity recognition, achieving 90% accuracy with the KTH dataset. Although LSTMs excel at distinguishing activities, they face challenges with overlapping actions and require large training datasets.

For anomaly detection, Reddy [10] applied VGG16-based neural networks on the Avenue Dataset, achieving a high Area Under the Curve (AUC) of 88%. However, this method struggles with anomaly diversity and distinguishing between minor and major anomalies. Raiyn and Toledo [13] leveraged Generative Adversarial Networks (GANs) for pattern recognition, effectively augmenting training data using the CIFAR-10 dataset. Despite its efficacy, GANs demand extensive training and may produce unrealistic samples.

In facial recognition, Wang et al. [18] employed self-supervised learning models, achieving 95% accuracy on the LFW dataset. These models perform well without labeled data but are sensitive to occlusions and lighting variations. For image classification, Huang et al. [20] introduced DenseNet-based models that demonstrated 96% accuracy on the ImageNet dataset, showcasing high performance. However, they require substantial computational resources and are prone to overfitting on smaller datasets.

In object detection, Piekarski et al. [21] utilized YOLO (You Only Look Once), achieving 83% mean Average Precision (mAP) on the COCO dataset. While efficient for real-time detection, YOLO struggles with small or overlapping objects and shares the computational demands of more complex models. Hassan et al. [22] explored Faster R-CNN for object detection, achieving 84% mAP on the Pascal VOC dataset. This method offers high precision but requires significant computational power and exhibits slower inference times compared to single-shot detectors.

While these methodologies achieve remarkable results, their limitations highlight the need for improved algorithms to handle complex scenarios, reduce computational overhead, and enhance generalizability. Table I summarizes the methodologies, domains, datasets, performances, and limitations discussed in this section.

TABLE I. WORK DONE ON ANOMALY DETECTION

Ref. No.	Methodology Used	Domain	Data Set Used	Performance	Outcome/Limitations
[3]	Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)	Behavior Analysis	UCF101	92% Accuracy	Real-time human action recognition; limited by dataset variety; performance may degrade with complex backgrounds
[10]	Autoencoders	Anomaly Detection	Avenue Dataset	88% AUC	High accuracy in anomaly detection; limited by anomaly diversity and type; difficulty in distinguishing between minor and major anomalies.
[13]	Generative Adversarial Networks (GANs)	Pattern Recognition	CIFAR-10	Improved data generation	Effective for training data augmentation; requires extensive training; potential for generating unrealistic samples
[18]	Self-Supervised Learning Models	Facial Recognition	LFW	95% accuracy	High accuracy without labeled data; sensitive to occlusions and variations in lighting conditions
[19]	Faster Region-based Convolutional Neural Network (R-CNN)	Object Detection	Pascal VOC	84% mAP	High accuracy for object detection; slower inference time compared to single-shot detectors; requires significant computational resources
[20]	Long Short-Term Memory (LSTM)	Activity Recognition	KTH Dataset	90% accuracy	Accurate activity recognition; struggles with overlapping activities and requires large training dataset
[21]	YOLO (You Only Look Once)	Object Detection	COCO	83% mAP	Enhanced real-time object detection; struggles with small or overlapping objects; requires significant computational resources.
[22]	Residual Network (ResNet)	Image Classification	ImageNet	96% accuracy	Highly accurate image classification; requires large computational resources and training data; potential overfitting on small datasets

III. METHODOLOGY

Using CCTV video data, the suggested traffic anomaly detection system effectively detects and classifies traffic incidents in real-time by utilizing deep learning models. The proposed model for anomaly detection in residential areas aims to detect dual anomalies occurring in residential areas, one is violence detection, and another is accidents occurring in residential areas. The system uses a step-by-step methodology that includes object detection, geographical and temporal feature extraction, and data preprocessing.

To overcome the challenges of computational complexity and real-time anomaly detection the suggested method employs MobileNetV2, a lightweight CNN architecture designed specifically for deep learning on resource constraint devices. Fig. 1 outlines the methodology steps for both whereas Fig. 2 and Fig. 3 depict violence and accident detection model respectively.

For violence detection, an ensemble model of BiLSTM for temporal feature extraction and MobileNetV2 for Spatial feature extraction is used. For accident detection the ensemble model employed is a combination of VGG16 and MobileNetV2. Further for violence detection 800 samples of violence and nonviolence in the form of video are taken for training and testing ensemble model proposed for detection. On the other hand, accident detection dataset consists of 990 files in the form of frames as images taken from public available dataset for accidents occurring in residential areas. To overcome the challenges of computational complexity and real-time anomaly detection the suggested method employs MobileNetV2, a lightweight CNN architecture designed specifically for deep learning on resource constraint devices. Hence, MobileNetV2 as a model has lesser computational complexity than typical CNNs and therefore ideal for anomaly detection workloads involving real time processing. Due to all of this breakdown, MobileNetV2 emerges as a perfect real-time CCTV footage processing option since it cuts all parameters and computation needs greatly. Another advantage of MobileNetV2 architecture is that it is capable of utilizing pre-trained weights which are trained on large datasets such as ImageNet.

As mentioned earlier, anomaly detection of proposed model includes violence detection and accident detection in residential areas. The first type of anomaly being detected is accidents occurring in residential societies specifically in parking areas. The proposed model can obtain the high-level spatial features from the CCTV feed employing these pre-trained weights and the technique will then be able to effectively identify the outliers such as the violent crime or the traffic accident. Furthermore, the lightweight structure of the model also gives it flexibility in its operational environment and enables its deployment in scenarios that have hardware limitations such as in embedded systems and edge computing systems. The system further improves the MobileNetV2 by adding extra convolutional layers that helps in the feature extraction process. After characteristics of space have been extracted, the model is able to group identified patterns into pre-defined categories. For In traffic monitoring the categories could be “Accident” “No Accident” where the two main categories which will be flagged will be “Violence” and “Non-Violence” for residential areas surveillance. This allows the system to monitor CCTV footage by itself and generate real-time alerts when an anomaly is

detected thus ensuring that the authorities act promptly to prevent further damage from happening. Fig. 2 shows the Methodology involve in violence detection for anomaly detection in residential areas.

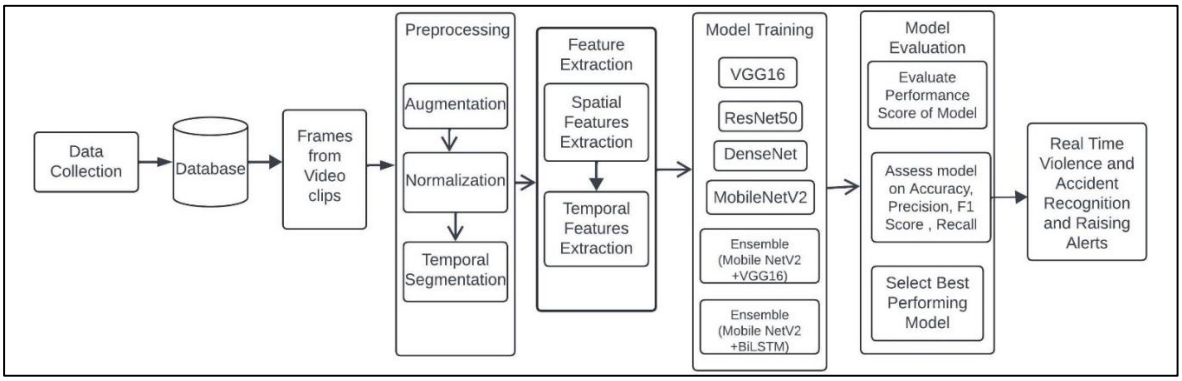


Fig. 1. Methodology of proposed anomaly detection in residential areas.

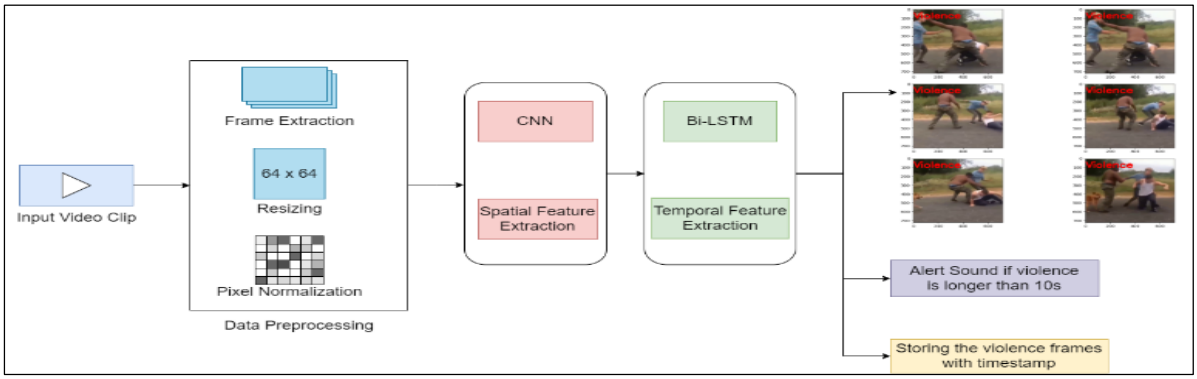


Fig. 2. Architecture of violence detection.

To have a comprehensive analysis the proposed model which focuses on anomaly detection and raising alerts considered dataset in the form of video clips for Violence detection and in the form of Frames from video for Accident detection. This ensures that system will be able to raise alerts on detection of anomaly, irrespective if the input is in the form of image or video. Further for violence detection the system proposes an ensemble model of MobileNetV2 and BiLSTM for spatial and temporal features extraction [23]. However for accident detection the proposed model utilizes ensemble model of MobileNetV2 and VGG16 for features extraction [19]. The architecture for accident detection models is given in the images Fig. 3:

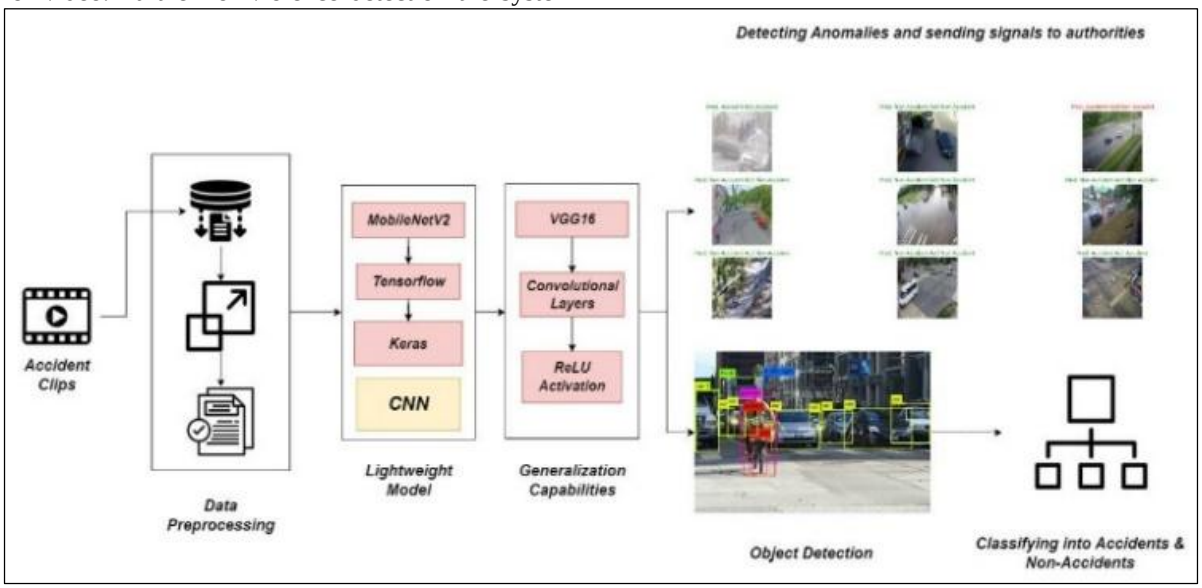


Fig. 3. Architecture of accident detection.

A. Data Preprocessing

The "Real Life Violence Situations Dataset" from Kaggle was used, which contained 800 MP4 files of both violent and non-violent clips with maximum length of 5 seconds. Violent clips were typically real street fight situations with people fighting bare-handed or with items such as sticks and rods. While the Non-Violent clips were various human activities like sports, eating, walking and other ordinary activities. This diversified dataset allowed the model to generalize more reliably and distinguish between violent and nonviolent scenarios. Preprocessing is considered as a critical step in preparing the video data for both model training and testing. By extracting frames at regular intervals, it was ensured that the frames are uniformly distributed across the video. The proposed model extracted 16 frames per video, ensuring it efficiently captures the temporal properties of video content, allowing the model to comprehend the development of activities across time without being overloaded by redundant information.

The first step to process videos into one scene is to extricate frames from the video stream. This is going to be necessary for the later steps of temporal analysis. The first step in the pre-processing of an input image is swapping the color channels, i.e., RGB channel rearranging to make the image compatible with Keras. The next step is the resizing of the input image to a fixed size, i.e., 224 × 224, without taking the aspect ratio of the image into consideration. In the last step in the pre-processing of an image, mean subtraction is applied to the image, where the mean is calculated for all pixel values in the image and the mean is subtracted from the pixel values. The reason behind this step is to make a standardization to which the other parameters such as weights and biases can refer to. The pre-processing of an image makes the image ready to be provided to a CNN model. In the testing phase of this research, the same set of techniques was applied. The testing of the CNN model was performed on videos, so while testing, the frames of the videos were looped, and all the frames were subjected to the same pre-processing as the training images.

Another technique used in the pre-processing phase of training is data augmentation. This technique is very popular while working with images, as it helps create more data, which helps the computation models to be able to generalize better. In this research, multiple data augmentation techniques were used, i.e., shift, rotation, shear, Flip, etc. Data augmentation was not applicable in the testing phase, as there was no need to multiply data and generate more. Raw frames from the video have various shapes each sample provides a set of raw frames with varying size, so before feeding those frames to deep learning models, the frames are resized to have the fixed, standard shape. The pixel values for the frames which have been retrieved are then scaled to fit into an agent ratio which is mostly a range from 0 to 1. This in turn means that consistencies of the intensity values are preserved, enhancing the performance of subsequent CNN based models.

B. Lightweight Model

MobileNetV2 model is used to get spatial features for the frames that were normalized from the frames extracted from the videos. MobileNetV2 is a light CNN model with high accuracy so it used in real-time applications such as traffic anomaly detection and violence recognition in residential areas. MobileNetV2's architecture is the depth wise separable convolution that reduces the amount of computation required while maintaining the extraction of features accurate. It captures immobile objects in a scene such as cars, traffic flow and other elements of the scene.

Then there is Batch Normalization followed by ReLU as the activation functions after the convolutional layers. This guarantees the model's non-linearities and promotes quicker training convergence: This guarantees the model's non-linearities and promotes quicker training convergence:

$$z = ReLU\ (BN\ Wconv * X)$$

Where Wconv are convolution weights, X is input, and BN refers to Batch Normalization.

C. Generalization using (VGG16)

To measure the temporal features; the features that are spatial are first extracted using mobile NetV2; then the VGG16 is used. They work to provide a perspective of the deeper convolutional layers of the VGG16, to identify how objects within frames of a video move. Dependencies between Space and Time: The weights provided by the VGG16 architecture help in model's ability to detect intricate temporal patterns, which are crucial for detecting such anomalies as mishaps or erratic movements.

$$ht = \sigma\ (WhhHt\text{-}1 + WxhXt)$$

Where Whh are weights for temporal connections, and Wxh captures spatial-temporal dependencies as shown in Fig. 4.

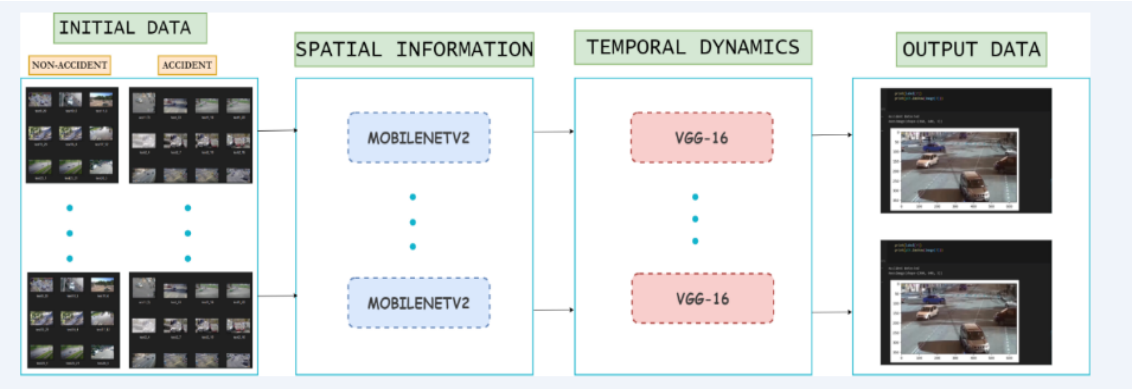


Fig. 4. Extracting temporal and spatial features for anomaly detection.

D. Object Identification

A variety of CNN based object detection algorithms detect objects in each video frame such as vehicles, pedestrians, and others traffic elements. The objects detected have a rectangular frame placed around and the movement of these objects is recorded over time. For this purpose, movement of the observed objects is compared between frames to identify any odd, unexpected behavior which could consist of crash occurrences, traffic congestions, or irregular driving patterns.

1) Accidents Classification and Anomaly Detection: Based on the space and time information gathered all the events depicted in the video as either accidents or non-accidents. In this case, the CNN and the Bi-LSTM models are very important in the identification of traffic pattern such as sudden halts, collisions or erratic movements indicating an accident [20]. The system monitors the occurrence of any irregularities in the flow and activity of the traffic on the road. For this, the technology identifies features that are likely to be abnormal such as times when the car suddenly stops, unpredictably veers or has a crash. Once an irregularity or accident has occurred, a real time alert is made to facilitate the correction of the problem. This may lead to an alert to the right authorities to ensure the situation does not happen again hence ensuring that it act as a stop gap measure.

2) Violence Detection and Time-Based Alerts: For events like violence or aggressive behavior (as depicted in the violence detection figure), if such events last longer than 10 seconds, the system triggers an alert sound to draw attention. The violence frames, along with timestamps, are stored for further analysis or evidence, which could be used in legal or administrative actions.

3) Anomaly Detection in violence cases Temporal features: While extraction of spatial characteristics is helpful in extracting any inconsistencies in a certain frame of a CCTV film, the temporal characteristics of the frames are as important as the former in order to detect anomalies. That is; it might take several frames to gather adequate information in traffic accident detection to help determine that an accident had occurred. As similarly with this, in distinguishing between an aggressiveness and mere touching or coming contact within violent activity detection, the events over time is important.

This difficulty can be addressed by the suggested system Bi-directional Long Short-Term Memory (BiLSTM) networks that are made especially to deal with sequential data. With the help of BiLSTM networks, the model is able to look at temporal sequences in forward as well as in the backward direction thereby adding to the understanding of context and flow of such sequences [19].

IV. Results

A comprehensive evaluation strategy was used to validate the effectiveness of the proposed model for detecting violent activities in residential areas. This involved training and testing the model on the "Real Life Violence Situations Dataset" from Kaggle. An 80-20 ratio was used to divide the dataset into training and testing sets. The testing set was used to assess the model's performance after it had been trained using the training set. The confusion matrix which can be seen in Fig. 5 was used to validate the model's efficiency with a primary focus on accuracy and recall in the Table II.

TABLE II. Model Evaluation Metrics for Violence Detection

Classes	Precision	Recall	F1 Score
Non Violent	0.98	0.89	0.94
Violent	0.94	0.99	0.91

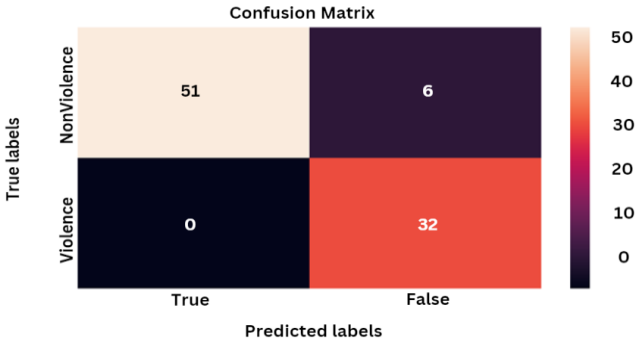


Fig. 5. Confusion matrix for violence classification.

Performance Parameters obtained from ensemble model of MobileNetV2 and VGG16 for accident classification is shown in Table III and the confusion matrix for calculation of performance for the same is shown in Fig. 6.

TABLE III. Model Evaluation Metrics for Accident Detection

Classes	Accuracy	Precision	Recall	F1Score
Non-Accident	0.96	0.96	0.84	0.93
Accident	0.96	0.95	0.91	0.89

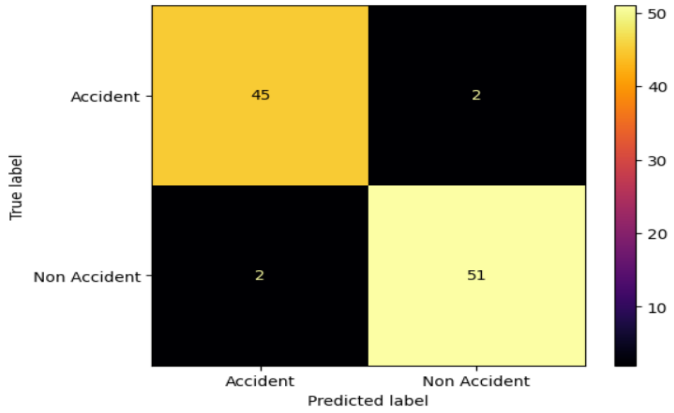


Fig. 6. Confusion matrix for accident classification.

The accuracy depicts the overall accuracy of the model in classifying the video clips correctly. A plot of Loss and Accuracy of the ensemble model for anomaly detection is shown in Table IV. The model's accuracy improved consistently throughout epochs, and the loss consistently decreased with the number of epochs.

TABLE IV. ACCURACY AND LOSS OF THE PROPOSED MODEL OVER EPOCHS

Epoch	Loss	Accuracy	Validation Accuracy
1	0.62	0.55	0.50
3	0.50	0.70	0.62
5	0.40	0.82	0.69
7	0.32	0.89	0.75
9	0.25	0.92	0.80
11	0.15	0.96	0.82
13	0.10	0.96	0.85
15	0.06	0.98	0.86
17	0.04	0.98	0.88
19	0.02	0.99	0.89

For real-time validation, the model was combined with OpenCV and evaluated on live video feeds to imitate real-world settings. The cv2 library was implemented to analyze the video frames. The model demonstrated high accuracy and low loss values for violent/non-violent action recognition, making it suitable for real-time applications.

V. DISCUSSION

The proposed model demonstrates exceptional performance in identifying violent and non-violent activities, achieving an overall test accuracy of 93.25%. Its Recall score of 0.99 for the violent class underscores its effectiveness in identifying nearly all instances of violence, a critical feature for anomaly detection systems in safety-critical environments. This aligns with prior research emphasizing the importance of high recall for detecting relevant instances of violence [22,23].

The model was further validated in real-time using OpenCV, where the cv2 library facilitated frame-by-frame video analysis to simulate real-world scenarios. This integration tested the model's ability to detect violent behavior in live settings, ensuring the classification process remained accurate and prompt. The high accuracy, low loss values, and strong classification metrics achieved in real-time settings validate its practical applicability and reliability for violence/non-violence recognition, as supported by prior findings in similar validation studies [24].

Fig. 7 compares the ROC curves of three models: BiLSTM, MobileNetV2, and the ensemble approach. The BiLSTM model, with an AUC of 0.88, effectively captures temporal features but falls short in spatial feature representation, limiting its ability to distinguish classes [25]. In contrast, MobileNetV2, which specializes in spatial feature extraction, achieves a superior AUC of 0.95. The ensemble model leverages the strengths of both, achieving a near-optimal AUC of 0.98. This performance reflects the ensemble model's ability to comprehensively extract temporal and spatial features, ensuring high classification accuracy across diverse scenarios[26].

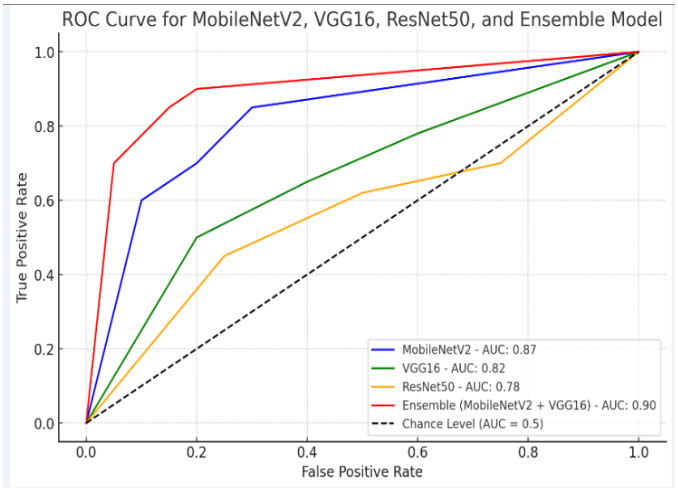


Fig. 7. ROC Curve for comparison of different models.

The comparative analysis in Table V highlights the superiority of the proposed model over existing solutions. Competing models, such as Faster R-CNN with RegNet+, EfficientNetB3 with Cascaded Convolution, and CNN-BiLSTM autoencoder, exhibit limitations, including reliance on low-quality datasets, inability to handle edge cases, and lack of real-time alert mechanisms. The proposed model surpasses these challenges by achieving 96% accuracy and raising alerts within 10 seconds of anomaly detection. Furthermore, its reliance on publicly available datasets enhances its adaptability, addressing limitations seen in prior research.

The dataset was stratified into an 80% training set and a 20% test set to ensure robust model training and unbiased performance evaluation. The confusion matrix reveals an accuracy of 96% and a recall score of 0.90 for accident detection, underscoring the model's efficiency in identifying activities associated with accidents. By raising alerts post-detection, the model further enhances its utility for real-time safety applications.

Thus, the proposed ensemble model effectively combines temporal and spatial feature extraction capabilities, achieving state-of-the-art performance in violence and accident detection. Its high recall and accuracy metrics, superior ROC performance Fig. 7, and unique real-time alert mechanism position it as a robust solution for enhancing community safety. Comparative insights from Table V affirm its versatility, addressing the shortcomings of existing systems and establishing it as a critical tool for anomaly detection in real-world settings. These findings align with prior research, validating its potential as a scalable and reliable anomaly detection system.

TABLE VI. COMPARATIVE ANALYSIS OF PROPOSED MODEL WITH OTHERS

Ref No.	*Dataset Used*	*Model Used*	*Data Balancing*	*Alerts Raised*	*Accuracy (In %)*	*Limitation*
[27]	Animated Customized dataset for Violence detection	Faster RCNN, RegNet+	Yes- Adam Optimzer	No	80.8	Dataset quality can be improved as it consists of animation videos. Also inference time can be further reduced
[28]	MVTec-AD, Railway Track Foreign Object Detection (TFOD)	Cascaded Convolution Self Attention Efficient NetB3	NA	No	99	Anomaly detected on Railway tracks but the model do not work well on edge points.
[29]	unique RGB+D dataset for Bank ATM anomaly detection	CNN-BiLSTM autoencoder framework	No	No	91	Used for Anomaly detection of ATM without raising alerts.
Proposed Model	Violence Recognition from Videos- Kaggle and Accident detection from CCTV footage Dataset .	Ensemble lightweight model for Violence and accident detection	Yes- SMOTE	Raise alerts after 10 sec of anomaly detection	96	Besides detection of anomaly as accidents or violence activity in residential areas, post detection alerts are also raised. Moreover, datasets used are those available in public domain so effective in varied situations.

VI. CONCLUSION

The use of AI-driven machine learning and deep learning techniques has significantly advanced the field of video surveillance and security. These technologies enable near real-time behavior analysis, object recognition, anomaly detection, and activity recognition, enhancing overall system efficiency. The proposed methodology focuses on detecting accidents and violent activities, generating and storing frames for post-event analysis, and raising real-time alerts for timely intervention. This dual functionality—anomaly detection and alert generation—addresses the critical need for proactive safety measures, particularly in residential areas where CCTV systems are primarily installed to detect and respond to emergencies. Unlike many existing anomaly detection models, which are computationally expensive, lack real-time responsiveness, and fail to trigger alerts, the proposed system is lightweight, efficient, and practical for real-world applications. By integrating an alert mechanism, the model ensures that authorities are notified promptly, reducing the impact of undesirable activities and improving community safety.

Future work will focus on improving the system to handle more complex situations, like detecting early signs of violence or predicting accidents before they happen. The model will also be improved to work better in different environments, such as low lighting or when objects are blocking the view. Additionally, using more types of data, like combining video with sound or sensor information, could make the system better at detecting unusual activities. Lastly, the system will be optimized for use on smaller devices, ensuring it can run quickly and efficiently without needing too much computing power.

REFERENCES

[1] Piza, E.L., 2018. The crime prevention effect of CCTV in public places: A propensity score analysis. Journal of crime and justice, 41(1), pp.14-30. doi.org/10.1080/0735648X.2016.1226931

[2] Piza, E.L., Welsh, B.C., Farrington, D.P. and Thomas, A.L., 2019. CCTV surveillance for crime prevention: A 40-year systematic review with meta-analysis. *Criminology & public policy*, *18*(1), pp.135-159. doi.org/10.1111/1745-9133.12419.

[3] S. W. Khan *et al.*, 'Anomaly Detection in Traffic Surveillance Videos Using Deep Learning', *Sensors*, vol. 22, no. 17, Sep. 2022, doi: 10.3390/s22176563.

[4] Kumar, K.K. and Venkateswara Reddy, H., 2022. Crime activities prediction system in video surveillance by an optimized deep learning framework. Concurrency and Computation: Practice and Experience, 34(11), p.e6852. https://doi.org/10.1002/cpe.6852

[5] Karunarathne, L., 2024. Enhancing Security: Deep Learning Models for Anomaly Detection in Surveillance Videos.

[6] Elmetwally, A., Eldeeb, R. and Elmougy, S., 2024. Deep learning based anomaly detection in real-time video. Multimedia Tools and Applications, pp.1-17. https://doi.org/10.1007/s11042-024-19116-9.

[7] Rezaee, K., Rezakhani, S.M., Khosravi, M.R. and Moghimi, M.K., 2024. A survey on deep learning-based real-time crowd anomaly detection for secure distributed video surveillance. *Personal and Ubiquitous Computing*, *28*(1), pp.135-151.

[8] Joshi, M. and Chaudhari, J., 2022. Anomaly Detection in Video Surveillance using SlowFast Resnet-50. *International Journal of Advanced Computer Science and Applications*, *13*(10).

[9] Qian, H., Zhou, X. and Zheng, M., 2020. Abnormal Behavior Detection and Recognition Method Based on Improved ResNet Model. *Computers, Materials & Continua*, *65*(3).

[10] G. Venkata Rami Reddy, 'DETECTING ABNORMALITIES USING VGG 16 NEURAL NETWORKS: AN ANOMALY DETECTION FRAMEWORK', pp. 1484–1491, 2022, doi: 10.48047/nq.2022.20.4.nq22380.

[11] Rahman, M.M., Afrin, M.S., Atikuzzaman, M. and Rahaman, M.A., 2021, December. Real-time anomaly detection and classification from surveillance cameras using Deep Neural Network. In 2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI) (pp. 1-6). IEEE.

[12] Sivakumar, G., Mogesh, G., Pragatheeswaran, N. and Sambathkumar, T., 2024. Video Anomaly Detection in Crime Analysis using Deep learning Architecture-A survey. *Journal of Trends in Computer Science and Smart Technology*, *6*(1), pp.1-17.

[13] J. Raiyn and T. Toledo, 'Real-Time Road Traffic Anomaly Detection', *J Transp Technol*, vol. 04, no. 03, pp. 256–266, 2014, doi: 10.4236/jtts.2014.43023.

[14] Trilles, S., Hammad, S.S. and Iskandaryan, D., 2024. Anomaly detection based on artificial intelligence of things: A systematic literature mapping. Internet of Things, p.101063. https://doi.org/10.1016/j.iot.2024.101063

[15] Chandrakar R, Miri R, Kushwaha A (2022) Enhanced the moving object detection and object tracking for traffic surveillance using RBF-FDLNN and CBF algorithm. Expert Syst Appl 191:116306, ISSN: 0957-4174. https://doi.org/10.1016/j.eswa.2021.116306

[16] Ullah, W., Ullah, A., Haq, I.U., Muhammad, K., Sajjad, M. and Baik, S.W., 2021. CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks. *Multimedia tools and applications*, *80*, pp.16979-16995.

[17] Kamble, K., Jadhav, P., Shanware, A. and Chitte, P., 2022. Smart Surveillance System for Anomaly Recognition. In *ITM Web of Conferences* (Vol. 44, p. 02003). EDP Sciences.

[18] G. Wang, Z. Guo, X. Wan, and X. Zheng, 'Study on Image Classification Algorithm Based on Improved DenseNet', in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Jun. 2021. doi: 10.1088/1742-6596/1952/2/022011.

[19] B. Lindemann, T. Müller, H. Vietz, N. Jazdi, and M. Weyrich, 'A survey on long short-term memory networks for time series prediction', in *Procedia CIRP*, Elsevier B.V., 2021, pp. 650–655. doi: 10.1016/j.procir.2021.03.088.

[20] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, 'Densely connected convolutional networks', in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Institute of Electrical and Electronics Engineers Inc., Nov. 2017, pp. 2261–2269. doi: 10.1109/CVPR.2017.243.

[21] M. Piekarski, J. Jaworek-Korjakowska, A. I. Wawrzyniak, and M. Gorgon, 'Convolutional neural network architecture for beam instabilities identification in Synchrotron Radiation Systems as an anomaly detection problem', Measurement (Lond), vol. 165, Dec. 2020, doi: 10.1016/j.measurement.2020.108116.

[22] N. Hassan, A. S. M. Miah, and J. Shin, 'A Deep Bidirectional LSTM Model Enhanced by Transfer-Learning-Based Feature Extraction for Dynamic Human Activity Recognition', Applied Sciences (Switzerland), vol. 14, no. 2, Jan. 2024, doi: 10.3390/app14020603.

[23] C. Zeng, D. Zhu, Z. Wang, M. Wu, W. Xiong, and N. Zhao, 'Spatial and temporal learning representation for end-to-end recording device identification', *EURASIP J Adv Signal Process*, vol. 2021, no. 1, Dec. 2021, doi: 10.1186/s13634-021-00763-1.

[24] C Yixin Tang, Yu Chen, Sagar A.S.M. Sharifuzzaman, Tie Li, An automatic fine-grained violence detection system for animation based on modified faster R-CNN, Expert Systems with Applications, Volume 237, Part C,2024, 121691, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2023.121691.

[25] Agrawal, R., Singh, J., Ghosh, S.M. (2020). Performance Appraisal of an Educational Institute Using Data Mining Techniques. In: Iyer, B., Deshpande, P., Sharma, S., Shiurkar, U. (eds) Computing in Engineering and Technology. Advances in Intelligent Systems and Computing, vol 1025. Springer, Singapore. https://doi.org/10.1007/978-981-32-9515-5

[26] Raja, R., Sharma, P.C., Mahmood, M.R. and Saini, D.K., 2023. Analysis of anomaly detection in surveillance video: recent trends and future vision. Multimedia Tools and Applications, 82(8), pp.12635-12651.

[27] Sahay, K.B., Balachander, B., Jagadeesh, B., Kumar, G.A., Kumar, R. and Parvathy, L.R., 2022. A real time crime scene intelligent video surveillance systems in violence detection framework using deep learning techniques. *Computers and Electrical Engineering*, *103*, p.108319.

[28] Liu, R., Liu, W., Duan, M., Xie, W., Dai, Y. and Liao, X., 2024. MemFormer: A memory based unified model for anomaly detection on metro railway tracks. Expert Systems with Applications, 237, p.121509.

[29] Khaire, P. and Kumar, P., 2022. A semi-supervised deep learning based video anomaly detection framework using RGB-D for surveillance of real-world critical environments. Forensic Science International: Digital Investigation, 40, p.301346.

Sentiment Analysis of Web Images by Integrating Machine Learning and Associative Reasoning Ideas

Yuan Fang[1]*, Yi Wang[2]
The College of Art and Media-Xianda College of Economics and Humanities, Shanghai International Studies University, Shanghai, 200000, China[1]
EMC Information Technology, R&D Shanghai Co. Ltd, Shanghai, 200000, China[2]

***Abstract*—To achieve automatic recognition and understanding of image sentiment analysis, the study proposes an image sentiment prediction network based on multi-excitation fusion. This network simultaneously handles multiple excitations, such as color, object, and face, and is designed to predict the sentiment associated with an image. A visual emotion inference network based on scene-object association is proposed using the association reasoning method to describe the emotional associations between different objects. The multi-excitation fusion image sentiment prediction network achieved the highest accuracy of 75.6% when the loss weight was 1.0. The network had the highest accuracy of 76.5% when the object frame data was 10. The average accuracy of the visual sentiment inference network based on scene-object association was 91.8%, which was an improvement of about 3.7% compared to the image sentiment association analysis model. The outcomes revealed that the multi-stimulus fusion method performed better in the image emotion prediction task. The visual emotion inference network based on scene-object association can recognize objects and scenes in images more accurately, and both the scene-based attention mechanism and the masking operation can improve the network performance. This research provides a more effective approach to the field of image sentiment analysis and helps to improve the computer's ability to recognize and understand emotional expressions.**

***Keywords*—Sentiment analysis; multi-excitation fusion; image emotion prediction; associative reasoning; attention mechanism**

I. Introduction

Due to the quick advancement of computer technology, the Internet is now a necessary component of daily life. People can watch videos, browse pictures online, and even communicate with friends thousands of miles away in real time [1]. However, with the popularization of the Internet, many problems have emerged. Online images play an important role in human life, but they also have some negative impacts, such as the pornification of online images, violence and false information [2-3]. In order to solve these problems, sentiment analysis of web images is needed, i.e., computer algorithms are used to determine the emotional attributes of web images. Sentiment representation is a key concept in image sentiment analysis, which refers to the use of a certain way to represent and describe the emotions or feelings expressed in an image, so that computers can understand and process the emotional information of the image. Common sentiment representation methods include extracting sentiment words from textual descriptions of images (such as titles, descriptions, labels, etc.) and calculating the sentiment polarity of the words using sentiment dictionaries (such as SentiWordNet). Encode emotional information using low-level visual features of images, such as local areas, colors, textures, etc., and learn a vectorized emotional representation by integrating visual and textual features of the image with conceptual features in the emotional ontology [4]. However, traditional sentiment analysis methods often rely on hand-designed feature extraction, such as color, shape, and texture, etc., which often fail to comprehensively reflect the emotional attributes of web images. The topic of web image emotion (IE) analysis has witnessed significant advancements in machine learning (ML) methods in recent times. ML-based sentiment analysis uses a large amount of text data with labeled sentiment to train models to automatically analyze the sentiment tendencies in the text, thus helping people to better understand and process large-scale sentiment information [5].

Halim et al. proposed a framework for recognizing sentiment in short texts using email text, employing an ML approach and six sentiment classifications. Experimental results indicated that the framework had better performance in sentiment recognition with an average accuracy of 83% [6]. Britzolakis' group presented a tool for sentiment analysis based on lexicon and ML algorithms and explored alternative implementations and open topics for political sentiment analysis on Twitter. The results demonstrated that the methodology could help readers to understand the field and identify the best options to conduct related work and research [7]. Alasmari et al. utilized ML methods for sentiment analysis of Arabic tweets from Saudi Arabian tourism industry using decision trees, random forests, logistic regression, and Naïve Bayes for three categories of classification. The results showed that logistic regression and Naive Bayes performed the best when dealing with Arabic morphology, achieving 86% accuracy [8]. Chirgaiya et al. used natural language processing techniques to train a classifier model for sentiment classification of movie reviews through feature extraction and ranking. The method's 97.68% accuracy in sentiment classification, as demonstrated by the testing data, was a supplement to the web's current movie rating systems [9]. Using a dataset of Facebook user book reviews and taking demographic data into account, Kumar's group investigated the effects of age and gender on sentiment analysis. Utilizing ML techniques for sentiment analysis, the study produced fresh findings about the effects of sentiment analysis on age and gender [10].

Associative inference refers to finding relationships and correlations between various variables in data, and feature

extraction is the selection and transformation of important features from raw data for subsequent analysis. Liewlom's group proposed a new inference framework for determining association rules without defining values for measure, minimum support, and minimum confidence. The study validated the feasibility and effectiveness of the approach through a tree of association rules found from a cancer dataset that reflected the sequential relationships of 15 items associated with the dataset [11]. Li et al. suggested a pedestrian recognition framework based on attribute mining and inference, which improved the performance of pedestrian re-recognition by designing a spatial channel attention module and utilizing the semantic inference and message passing functions of graph convolutional networks. Experimental results indicated that the method achieved 87.03% accuracy on the Market-1501 dataset [12]. The wear monitoring system that Lin' et al. proposed a rapid Fourier transform to extract features from vibration and auditory signals by using a variety of sensors and feature fusion techniques. Through the use of cross-validation, the system developed with a hierarchical neural network structure and sensor feature fusion demonstrated its efficacy and performance under various tightening torque values and spindle speeds [13]. To achieve automatic clustering processing of grid intrusion features, Zhang et al. proposed a fuzzy c-mean clustering based method for extracting intrusion features. This was combined with a fuzzy association rule scheduling method to reconstruct the structure of intrusion statistical feature sequences, and a global optimization method to achieve automatic clustering processing. The approach could enhance the power grid's resistance to attacks and enable autonomous clustering of intrusion feature extraction, according to the results [14]. Guo et al. suggested a new method to extract feature points based on topological information, which achieves feature point detection of point neighborhood topology by introducing an improved local binary pattern to process the original point cloud. Experiments demonstrated that the method performed robustly in extracting point cloud shape features [15].

In summary, many researchers have conducted various designs and studies for sentiment analysis and associative feature extraction. However, these methods and models still have limitations. For example, emotional representation methods are relatively single and rely on manually designed feature extraction, making it difficult to fully reflect the emotional attributes of images. There are many applications in sentiment analysis, and some algorithms need to improve their accuracy when dealing with complex emotions and cross-cultural scenarios. In addition, sentiment analysis research in different fields and languages is relatively scattered, lacking universality and comparability. To improve the effectiveness of sentiment analysis techniques, the study proposes an approach to Web IE analysis that combines ML and associative reasoning concepts.

The study is divided into five sections, with Section II proposing sentiment analysis methods. Section III is the validation and application analysis of the method. Section III is a discussion of research methods, application analysis, etc. Finally, Section V concludes the paper.

II. Methods and Materials

A. IEPN-MIF-Based Method

The Internet is growing quickly, and as a result, a lot of photos and videos with rich emotional content are shared online. Aiming at the problem of image sentiment analysis, the study proposes an IE prediction network based on multi-excitation perception (IEPN-MIF), as shown in Fig. 1. The network consists of three stages, including incentive selection, feature extraction, and emotion prediction. In the incentive selection stage, using object detection and face detection methods in deep learning, specific emotional incentives are accurately selected, which cover aspects such as color I_g, objects I_s, and faces I_e. Then, the feature extraction stage is entered, which synchronously extracts different emotional features from different stimuli. Finally, in the emotion prediction stage, the hierarchical cross-entropy (CE) loss function is used to optimize the emotion prediction results, while fully combining the inherent hierarchical structure of emotions to distinguish between simple negative samples and difficult negative samples. The research also innovatively proposes a new hierarchical CE loss function to further optimize the performance of the whole network and achieve more accurate and efficient IE prediction based on multi-stimulus perception [16].

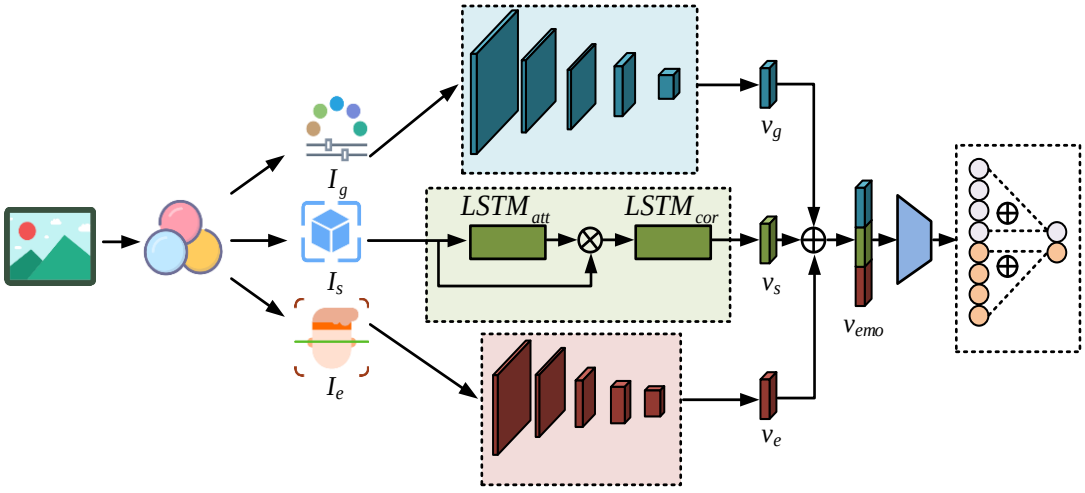


Fig. 1. Image emotion prediction network based on multi-incentive fusion.

In emotion feature extraction, the study extracts color features and other global features in images by constructing a global network. To balance computational efficiency and accuracy, the global network makes use of the ResNet-50 network, which has five convolutional layers and a global average pooling layer [17]. The features output from the last convolutional layer are shown in Eq. (1).

$$F_g = FCN_{res}(I_g) \tag{1}$$

In Eq. (1), the final convolutional layer outputs features as F_g and the full convolutional network as FCN_{res}. The final extracted global features are shown in Eq. (2).

$$G_g = G_{avg}(F_g) \tag{2}$$

In Eq. (2), the global feature is G_g and the fully convolved global average pooling layer is G_{avg}. To better mine the correlation between different objects, the object features are considered as a sequence data and a long short-term memory (LSTM) has been used to carve out this dependency [18]. The study also designed an emotion-specific semantic network for mining semantic associations between different objects and thus inferring IEs. The network consists of two LSTM layers, i.e., attention and association LSTM. The computational procedure of Attention LSTM is shown in Eq. (3).

$$h_t^{att} = LSTM_{att}(x_t^{att}, h_{t-1}^{att}) \tag{3}$$

In Eq. (3), the output of the attention LSTM is h_t^{att} and the input vector is x_t^{att}. The output of the attention LSTM at time t is calculated by the attention LSTM module based on the input vector at time t and the output of the previous time (t-1). The attention output at the current time depends on the current input and the attention state of the previous time. The weight $a_{i,t}$ of the attention module is calculated as shown in Eq. (4).

$$a_{i,t} = w_a \tanh(W_i f_i + W_h h_t^{att}) \tag{4}$$

In Eq. (4), the module learnable parameters are w_a, W_i, and W_h, respectively, and the object features are f_i. The weighted object features are calculated as shown in Eq. (5).

$$f_{att} = \sum_{i=1}^{N} a_{i,t} f_i \tag{5}$$

The weighted object feature f_{att} in Eq. (5) is sent to the associative LSTM. The output vector h_t^{cor} in the associative LSTM is computed as shown in Eq. (6).

$$h_t^{cor} = LSTM_{cor}(x_t^{cor}, h_{t-1}^{cor}) \tag{6}$$

In Eq. (6), the association LSTM input vector is x_t^{cor}, and the output vector of the previous moment is h_{t-1}^{cor}. By feeding object excitations into the attention LSTM and association LSTM separately, the possible redundancy of information between multiple objects can be reduced and the semantic correlations between different objects can be mined [19]. Therefore, the output of the semantic network can be regarded as the higher-level semantic features of the object excitations, as shown in Eq. (7).

$$G_s = h_T^{cor} \tag{7}$$

In Eq. (7), the last moment of LSTM is T, and the semantic features extracted from the object excitation are G_s. In order to extract the face expression features, the study adopts ResNet-18 as the base network to construct the expression network, and the expression features are calculated G_e if shown in Eq. (8).

$$G_e = \begin{cases} G_{avg1}(FCN_{res1}(I_e)), \exists I_e \\ 0, \quad others \end{cases} \tag{8}$$

In Eq. (8), the fully connected layer (FCL) of the expression network is FCN_{res1}, and the average pooling layer is G_{avg1}. Under the given expression category, the expression features are calculated by the ResNet-18 network layer, while in other cases the feature values are 0. The study analyzes three typical emotion incentives (color, object, and face), and designs a dedicated network to extract their emotion features (global features, semantic features, and expression features). These features are independent and complementary to each other, and together determine the final emotion classification result. In order to perform sentiment prediction, these several sentiment features will be spliced, and the final generated sentiment features are shown in Eq. (9).

$$G_{emo} = Concate[G_g, G_s, G_e] \tag{9}$$

In Eq. (9), the splicing operation is $Concate$. The final generated sentiment feature G_{emo} will be input into the subsequent sentiment classifier. In traditional categorization tasks, CE loss is widely used and has made breakthroughs in several tasks. However, sentiment categories are not completely independent from each other, but there is an inherent hierarchical relationship. An eight-category psychological model—which comprises the emotion categories of enthusiasm, amazement, contentment, anger, disgust, sadness, and fear—will serve as the foundation for the dataset that the study will construct. Positive and negative emotions are the emotional polarities of the emotion categories. In the eight-category emotion model, the first four emotions are positive emotions and the last four belong to negative emotions. The eight-category emotion model is shown in Fig. 2.

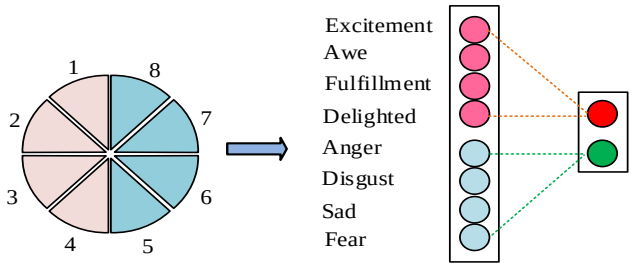


Fig. 2. Eight classification of the emotion model.

After cascading the sentiment features, they are sent to the classifier and activation function respectively to perform operations to obtain the sentiment vector as shown in Eq. (10).

$$p_{emo}(i|G_{emo},W)=\frac{\exp(w_i G_{emo})}{\sum_{i=1}^{C}\exp(w_i G_{emo})} \quad (10)$$

In Eq. (10), the sentiment vector is p_{emo}, and the sentiment type is C. The learnable parameter in the sentiment classifier is W, and its constituent elements are w_i. The traditional CE loss cannot distinguish between positive samples and negative samples, so this study proposes the auxiliary polarity loss and the hierarchical CE loss to solve this problem. Among them, the auxiliary polarity loss is used to distinguish between simple negative samples and difficult negative samples, while the stratified CE loss combines sentiment loss and polarity loss to obtain more accurate sentiment prediction. The final obtained stratified CE loss L_{CE} is shown in Eq. (11).

$$L_{CE}=L_{emo}+\lambda L_{pol} \quad (11)$$

In Eq. (11), the affective loss is L_{emo}, the polarity loss is L_{pol}, and the equilibrium setting hyperparameter is λ.

B. *Image Emotion Classification Based on Associative Reasoning*

When studying IE, one must focus on the scenes and objects it contains and use reasoning to understand IE in the context of how the two interact. The study suggests a scene object-interrupted visual emotion reasoning network (SOLVER) based on this, as seen in Fig. 3.

In the operating mechanism of the SOLVER network, it first uses a faster region convolutional neural network (Faster R-CNN) object detector to extract semantic concepts and visual features of various objects, and then filters and transforms these extracted features. On this basis, sentiment maps are constructed based on object features to represent emotional associations between different objects. Then, a graph neural network (GNN) is used for inference, connecting different object nodes by sentiment edges to generate object features enriched with emotions [20]. In addition, the study also designed a scene-object fusion module, which utilizes the attention mechanism (AM) to fuse object features based on scene features, while constructing the mutual relationship between the scene and objects. In the specific operation, the Faster R-CNN object detector is used to select a set of target candidate regions, pre-trained on the dataset, and finally output attribute categories. Through these steps, the SOLVER network is able to complete its complex operations and feature extraction process. As illustrated in Fig. 4, the sentiment graph is constructed by outputting the ten objects with the highest confidence ranking as its nodes. Each sentiment image is represented by a sequence of object semantic concepts, corresponding confidence scores, and VFs following Faster R-CNN.

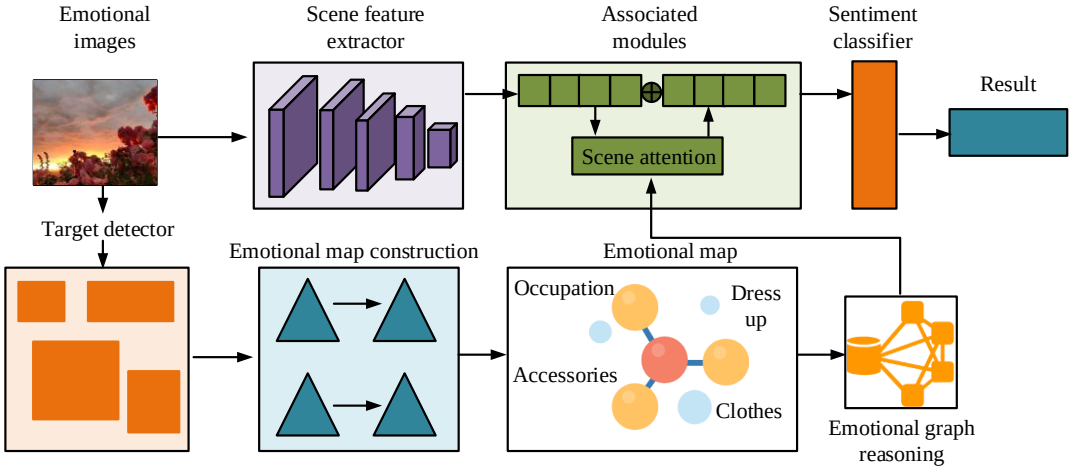


Fig. 3. SOLVER network.

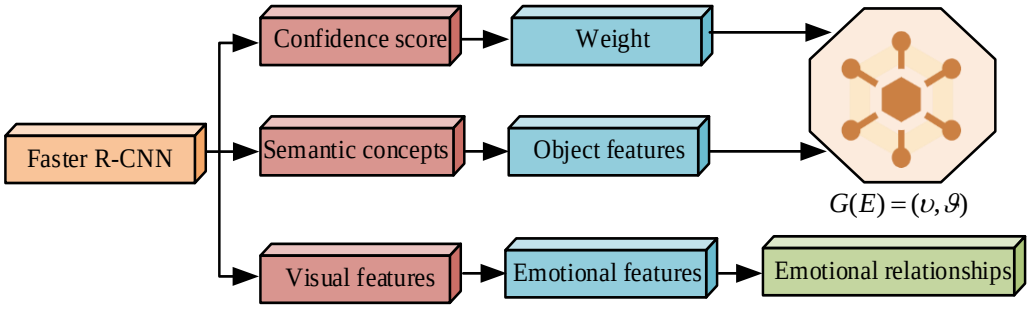


Fig. 4. The construction of emotion map.

The edges of the emotion graph are constructed by VFs and the nodes of the emotion graph are constructed by semantic features, and the emotion VF vector v_i^e is constructed as shown in Eq. (12).

$$v_i^e = \ell_1(W_e v_i + b_e) \tag{12}$$

In Eq. (12), the learnable embedding matrix is W_e, the embedding bias is b_e, the nonlinear regular function is ℓ_1, and the object VFs are v_i. The emotional relationship between different objects is shown in Eq. (13).

$$r_{i,j}^e = \phi(v_i^e)^T \cdot \varphi(v_j^e) \tag{13}$$

In Eq. (13), the sentiment relation of two objects in the same picture is $r_{i,j}^e$, and the two sets of embedding functions are ϕ and φ, respectively. The composition of the sentiment map vector is shown in Eq. (14).

$$G(E) = (\upsilon, \vartheta) \tag{14}$$

In Eq. (14), the emotion graph is $G(E)$, the objects containing different semantic features are υ, and the emotion relationship between different objects is ϑ and is described by an affinity matrix. Then, nodes with a confidence level lower than 0.3 are filtered out by setting a threshold for the confidence level, and the edges connected to the redundant nodes are subjected to a masking operation to reduce the information redundancy of the nodes. Finally, the masked affinity matrix is obtained by weighting the masked matrix to the affinity matrix to describe the sentiment relationship between different objects. Subsequently, the GCN will reason about the sentiment graph to exchange and disseminate information through the structured graph structure. In the design of GCN, residual structures are incorporated to better maintain the original node characteristics. Object features are propagated over the whole sentiment graph based on sentiment relationships in multilayer GCN, where each node is updated based on itself and its neighbors [21]. The multilayer GCN structure iterates the object features based on the sentiment relations to realize the inference of the sentiment graph. The result of emotion graph inference is shown in Eq. (15).

$$O'^{(l)} = f(O'^{(l-1)}, R^{e'}) \tag{15}$$

In Eq. (15), the relationship function between different nodes is f, the output of the last GCN layer is $O'^{(l)}$ and the input edge features are $R^{e'}$. The emotion-enhanced object features are formed by modeling different emotional relationships between objects. Scene is regarded as an additional motivator in the emotional arousal process, which greatly affects the image's emotional tone. When doing an IE analysis, scene features shouldn't be overlooked [22]. Consequently, as illustrated in Fig. 5, the study suggests a scene-based AM. Using ResNet-50 as a base network to build a scene network, this technique mines the emotional connections between objects and scenes, taking an emotional image as input and producing the scene attributes of that image.

Based on scene features and emotion-enhanced object features, AM first projects scene features onto the same embedding space to reduce the difference between scene features and object features. Then the attention weight of each object feature is computed by emotion association to fuse the scene features and object features to establish deep emotional relationships. The attentional weights are calculated as shown in Eq. (16).

$$a_i = \sigma(F_s(f_{sce}) \cdot F_o(o_i')) \tag{16}$$

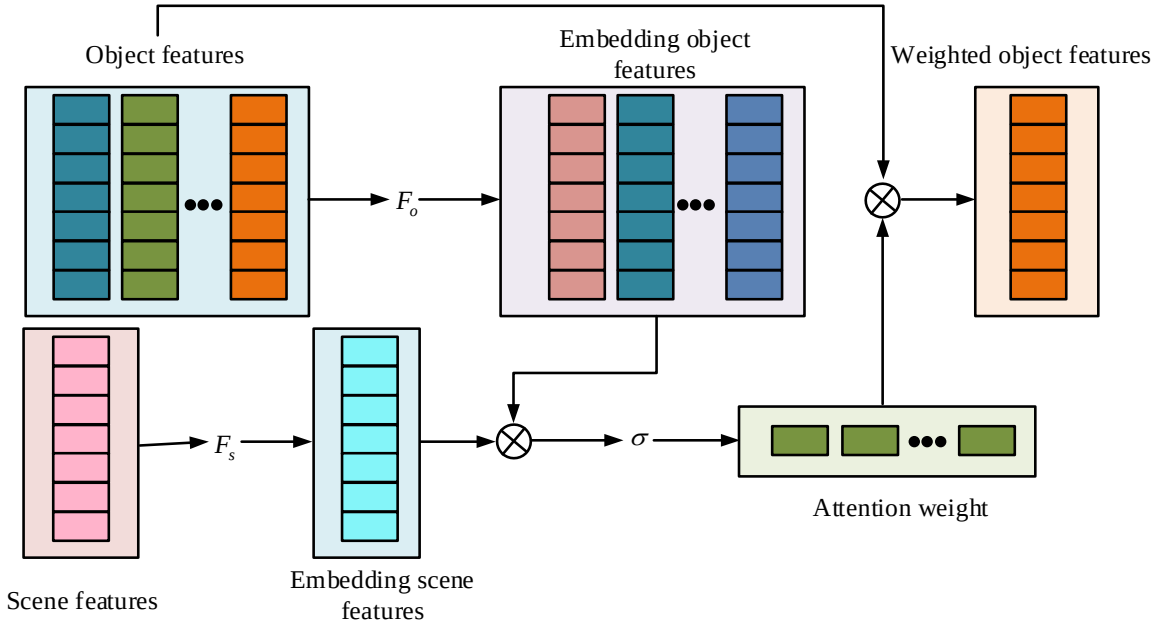


Fig. 5. Scenario-based attention mechanism.

In Eq. (16), the attention weight is a_i, and the activation function is σ. The embedded scene feature input is F_s, and the scene feature is f_{sce}. The embedded object feature is $o_i^{'}$, and the emotionally augmented object feature is F_o. The weighted object feature calculation is shown in Eq. (17).

$$f_{obj} = \sum_{i=1}^{M} a_i o_i^{'} \tag{17}$$

In Eq. (17), the weighted object feature is f_{obj} and the number of objects is M. The emotion feature after cascading the scene and the weighted object feature is shown in Eq. (18).

$$f_{emo} = Concate(f_{sce}, f_{obj}) \tag{18}$$

In Eq. (18), the cascaded sentiment feature is f_{emo}. The study interacts and fuses the scene and object features via AM to obtain weighted object features. Next, the scene and object features are cascaded to obtain the emotion features, which are used for subsequent emotion prediction along with a learnable weight matrix. The results of the sentiment prediction are compared with the sentiment class labels in the dataset and the network is optimized by CE loss.

III. Results

A. IEPN-MIF based Application Analysis

The experiments are performed on a computer based on the Pytorch framework with an Intel(R) Xeon (R) CPU E5-2640 2.40 GHz and an NVIDA GeForce GTX TITAN GPU (12G RAM) with Linux as the operating system. The experimental environment provides powerful computational capabilities to support the analysis and processing of large-scale datasets, which provides a good foundation for the training and testing of deep learning models. Three datasets are used for the experiments, namely, the FI dataset, the IAPS dataset and the ArtPhoto dataset. Table I displays the breakdown of the dataset. 2000 photos are chosen at random as the training set and 1000 images as the test set from the IAPS dataset. 2000 photographs are chosen at random as the test set and 6000 images as the training set from the Artphoto dataset. 15,000 photos are chosen at random as the training set and 7,000 images as the test set from the F1 dataset.

The experiments compare CNN-LSTM, LSTM, convolutional neural network (CNN), and the semantic emotion model proposed in 2023, and the various methods are applied in the small-scale dataset in order to validate the efficacy of the research proposed image emotion prediction network with multi-incentive fusion (IEPN-M). The accuracy and recall in IAPS are shown in Fig. 6.

Fig. 6 (a) shows the accuracy comparison in the dataset IAPS, the average accuracy of the study proposed IEPN-MIF is 80.3%, which improves about 18.6%, 29.5%, 25.6%, and 15.3% compared to the CNN-LSTM, LSTM, CNN models, and semantic sentiment models, respectively. Fig. 6 (b) shows the recall comparison in the dataset IAPS, and the average recall of the proposed IEPN-MIF is 83.2%, which is improved by about 10.3%, 14.3% 15.6%, and 9.8% compared to CNN-LSTM, LSTM, CNN model and semantic sentiment model, respectively. The results show that the multi-stimulus fusion methods have better performance in IE prediction tasks. The accuracy comparison of different methods on Artphoto and F1 datasets is shown in Fig. 7.

TABLE I. Dataset Division

Data set	Training set	Test set	Total number
IAPS	20000	1000	3000
Artphoto	6000	2000	8000
FI	15000	7000	23000

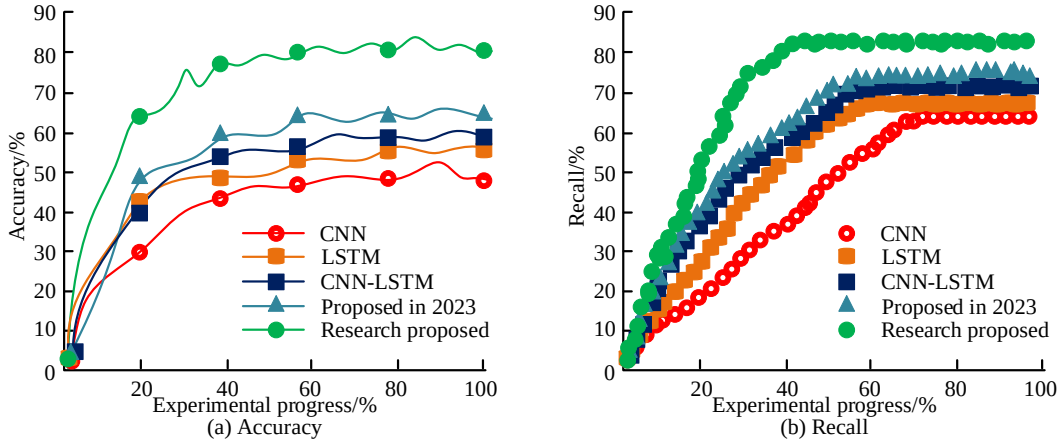


Fig. 6. Accuracy and recall in the dataset IAPS.

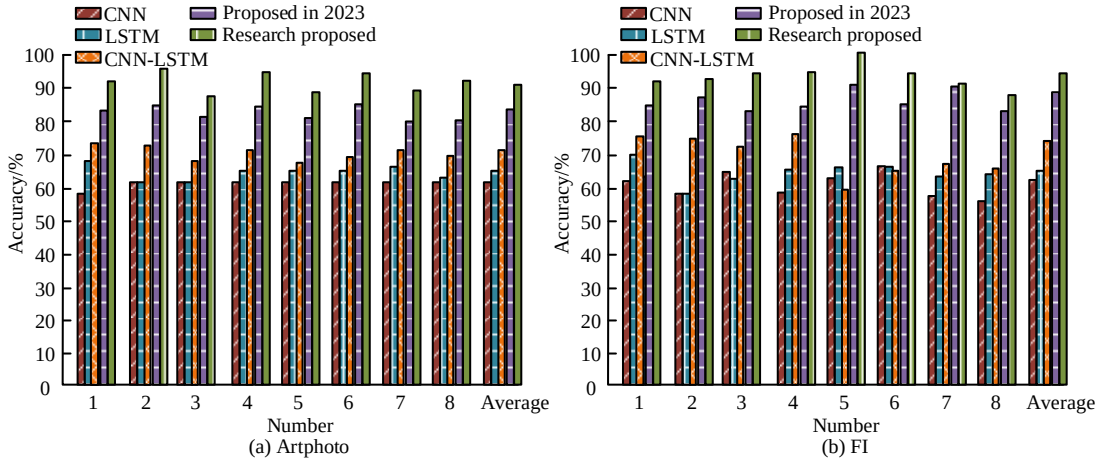


Fig. 7. Accuracy in the Artphoto and FI datasets.

TABLE II. RESULTS OF IMAGE EMOTION PREDICTION NETWORK ABLATION EXPERIMENTS FOR MULTI-EXCITATION FUSION

Global network		Semantic network		Expression network	Accuracy
RGB	Y	LSTM	FCL		
×	√	×	×	×	59.3%
√	×	×	×	×	67.2%
√	×	√	×	×	71.6%
√	×	×	√	×	62.7%
√	×	√	×	√	91.3%

In Fig. 7, 1-8 denote eight emotions, respectively. Fig. 7 (a) shows the accuracy comparison in the Artphoto dataset, and the average accuracy of IEPN-MIF is 91.3%, which improves about 19.6%, 23.8%, 27.8%, and 5.7% compared to CNN-LSTM, LSTM, CNN model and semantic sentiment model, respectively. Fig. 7(b) shows the accuracy comparison in the FI dataset, and the average accuracy of the IE prediction method with multi-incentive fusion is 93.6%, which improves about 20.7%, 30.1%, 27.5%, and 4.7% compared to CNN-LSTM, LSTM, CNN model, and semantic sentiment model, respectively. The outcomes demonstrate the increased accuracy and greater potential of the multi-excitation fusion IE prediction approach in the emotion identification domain. Table II displays the outcomes of the IEPN-MIF ablation trials.

In Table II, the global, semantic, and expression networks each contribute to sentiment prediction to varying degrees. When the three sub-networks act alone, the global network has the highest accuracy of 67.2%, indicating the strong representational power of global features. In the absence of color features, the three-branch network outperforms the two-branch network only to a lesser extent, while the three-branch network considering color features performs superiorly. The results show that by setting up the LSTM layer and the FCL for comparison, the LSTM layer structure can better mine the semantic information between different objects, and thus predict emotions more accurately. The addition of the expression network further improves the accuracy of the network, with a final accuracy of 91.3%. The results of the impact of hyperparameters on the network performance are shown in Fig. 8.

Fig. 8 (a) shows the effect of loss weights on network performance, and the accuracy of IEPN-MIF first increases and then decreases as the loss weights increase. The highest accuracy of 75.6% is achieved when the loss weight is 1.0. Fig. 8 (b) displays the effect of the object frames on the object performance, with the increase of the number of object frames, the accuracy of IEPN-MIF gradually increases and then decreases and stabilizes. When the object frame data is 10, the network has the highest accuracy rate of 76.5%. In conclusion, the model performs best when the IEPN-MIF loss weight and object frame count are set to 1.0 and 10, respectively.

B. Application Analysis of Image Emotion Classification based on Associative Reasoning

To validate the performance of the SOLVER network proposed in the study, the experiments use Faster R-CNN, GCN and the IE correlation analysis model based on hierarchical graph convolutional network proposed in 2023 as comparisons. The comparison of the accuracy of the different methods on the IAPS and Artphoto datasets is shown in Fig. 9.

Fig. 9 (a) shows the accuracy comparison in the IAPS dataset, where the average accuracy of SOLVER network is 92.1%, which is an improvement of about 29.8%, 25.3, and 7.4% compared to the Faster R-CNN, GCN, and IE correlation analysis models, respectively. Fig. 9 (b) shows the accuracy comparison in Artphoto dataset, the average accuracy of SOLVER network is 91.8%, which improves about 32.6%, 31.7, and 3.7% compared to Faster R-CNN, GCN and IE correlation analysis model, respectively. This indicates that the SOLVER network is able to recognize objects and scenes in images more accurately because of the more optimized inference method used in the SOLVER network. The results of the ablation experiments and hyperparametric analysis of the SOLVER network are shown in Fig. 10.

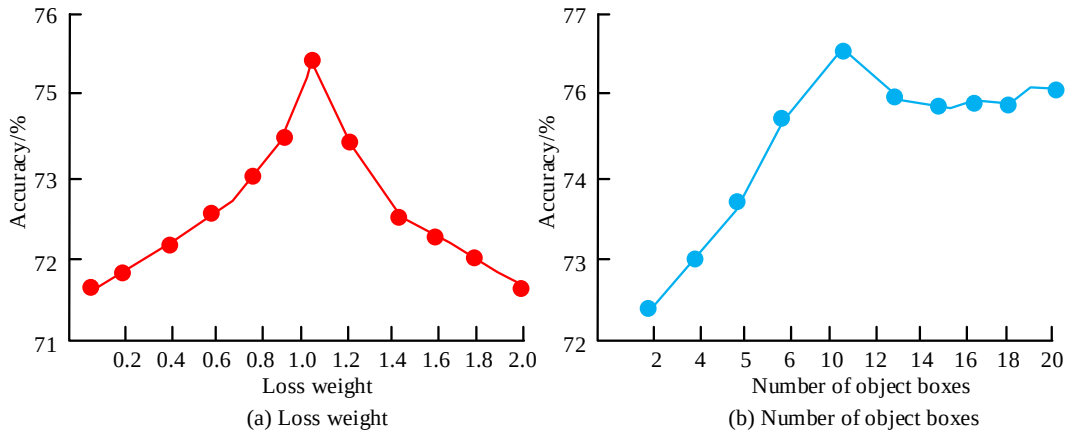


Fig. 8. Results of the influence of hyperparameters on network performance.

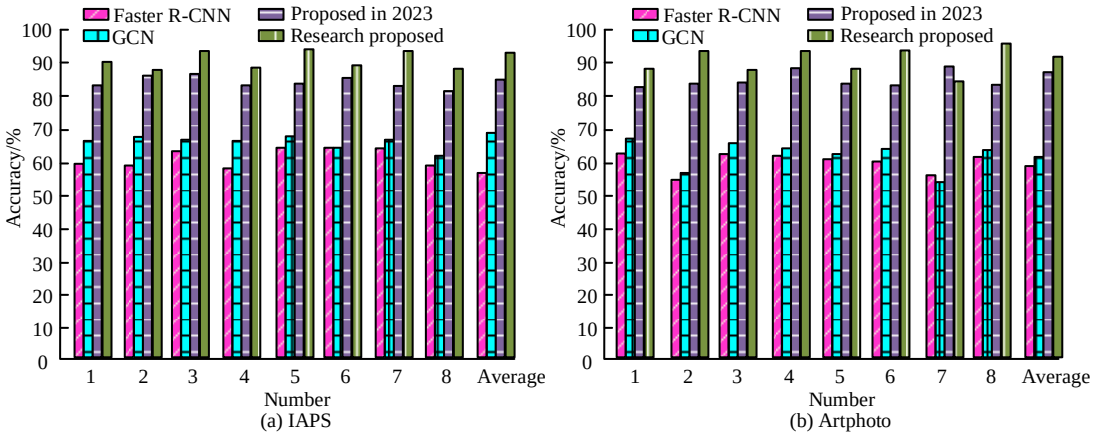


Fig. 9. Comparison of accuracy in the IAPS and Artphoto datasets.

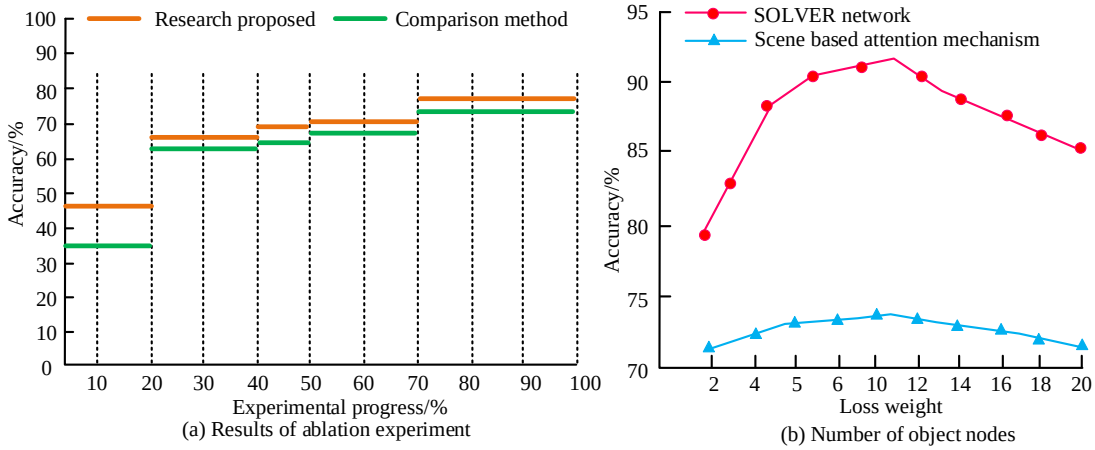


Fig. 10. Results of the ablation experiments and the hyperparameter analysis of the SOLVER network.

The results of the ablation tests are displayed in Fig. 10(a). In the 0%–10% phase, several objects result in a greater performance improvement than a single item. Two separate parameter embedding functions perform better than one embedding function in the 10%–40% stage. The mask operation can enhance performance even more in the 40%–50% range. Scene branching and scene-based AM both significantly contribute to emotion classification. In the 50%-70% stage, scene features direct the fusion process of object features. The results of the hyperparameter analysis are displayed in Fig. 10(b), where the accuracy of the scene-based AM and SOLVER network steadily rises as the number of object nodes increases until stabilizing. The scene-based AM and SOLVER networks get the maximum accuracy rates of 73.6% and 91.9%, respectively, when there are eleven object nodes. Therefore, the number of object nodes should be set to 11 to ensure network performance.

IV. Discussion

The proposed network image sentiment analysis method demonstrates high performance. This method fully utilized the advantages of ML in automatically recognizing image features, and deeply analyzed emotional expression through the idea of associative reasoning, achieving significant results in the field of sentiment analysis. In different emotional scenarios, this method could accurately identify various emotions such as joy, anger, sadness, and happiness, and achieve efficient emotional judgment in complex backgrounds. This was due to the fact that ML algorithms can effectively extract image features and improve the accuracy of emotion recognition. Meanwhile, the introduction of associative reasoning enabled this method to combine contextual information for more detailed analysis and judgment of emotional expression, thereby demonstrating high performance in different emotional scenarios. However, this fusion method also had certain limitations. Firstly, due to the ambiguity and diversity of emotional expression in network images, ML models might find it difficult to fully capture all key features during training, resulting in uncertainty in sentiment analysis results. Second, some environmental factors present in the associative reasoning process might affect the accuracy of sentiment analysis, resulting in results that are somewhat limited by human cognition. In addition, with the continuous changes in the expression of emotions in network images, ML and associative reasoning models needed to be constantly updated and optimized to adapt to new emotional trends.

These results are not limited to the dataset used, and the principles and methods behind them are applicable to other image and video datasets with rich emotional information. For example, in areas such as social media analysis, movie sentiment scoring, and advertising effectiveness evaluation, IEPN-MIF networks can accurately predict the emotional content of images by extracting emotional features from multiple sources, such as colors, objects, and faces. The SOLVER network can be applied to intelligent surveillance systems to support safety warnings and behavior understanding by analyzing the emotional relationship between scenes and objects. In the future, these networks are expected to be widely used in practical applications such as emotional interactive robots and personalized recommendation systems, thereby enhancing user experience and system intelligence.

V. Conclusion

To achieve accurate recognition and classification of IE, the study proposed an IEPN-MIF based network which was capable of utilizing different emotional incentives, including color, object, and face. It can achieve accurate prediction of emotions through deep learning techniques. Secondly, the study proposed SOLVER network to classify the emotion of images by fusing scene features and object features. The outcomes indicated that the average accuracy of the proposed IEPN-MIF in the dataset IAPS was 80.3%, which was improved by about 18.6%, 29.5%, 25.6%, and 15.3% compared to the CNN-LSTM, LSTM, CNN models and the semantic sentiment model, respectively. As the loss weight increased, the accuracy of IEPN-MIF first increased and then decreased. The highest accuracy of 75.6% was achieved when the loss weight was 1.0. The average accuracy of the SOLVER network in the IAPS dataset was 92.1%, which improved about 29.8%, 25.3, and 7.4% compared to the Faster R-CNN, GCN and IE correlation analysis models, respectively. Scene-based AM can effectively fuse scene and object features and classify emotions. Finally, the results of ablation experiments and hyperparameter analysis indicated that the number of object nodes should be set to 11 to ensure network performance. This method had great potential in practical applications, such as in the fields of intelligent interaction, psychology, and advertising marketing, where more accurate emotion recognition and classification could be achieved by analyzing IEs. Meanwhile, the scene based AM could effectively integrate scene and object features, and classify emotions, which provided the possibility for further improving the accuracy of IE recognition. The shortcoming of this research was that only datasets were used for the experiments. Therefore, the results only reflected the performance on these datasets. Future research directions can revolve around the following aspects. First, explore more advanced neural network architectures. Although CNN and LSTM networks have achieved good results in image sentiment analysis, the latest architectures such as Transformer AMs may provide better performance. Second, more effective loss functions should be developed, such as contrast loss or cosine similarity loss. Future research should test the methods on more diverse datasets to ensure their generalizability. Finally, consider multimodal sentiment analysis. Image sentiment analysis typically focuses only on visual content, and combining modal information, such as text associated with images or background audio, can improve analysis accuracy. With this dependency, future research can further expand the field of image sentiment analysis, develop more accurate and robust methods for understanding the emotional content of online images, and make greater contributions to the development of artificial intelligence technology.

References

[1] Hosseinalipour A, Ghanbarzadeh R. A novel metaheuristic optimisation approach for text sentiment analysis. International journal of machine learning and cybernetics, 2023, 14(3):889-909.

[2] Liu X, Zhang Z, Zhang G. Using improved feature extraction combined with RF-KNN classifier to predict coal and gas outburst. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2023, 44(1):237-250.

[3] Sun X, Cai Z. Research on an Eye Control Method Based on the Fusion of Facial Expression and Gaze Intention Recognition. Applied Sciences, 2024, 14(22): 10520-10542.

[4] Bhaumik G, Govil M C. SpAtNet: A spatial feature attention network for hand gesture recognition. Multimedia Tools and Applications, 2024, 83(14): 41805-41822.

[5] Sun Y, Guo Q, Zhao S, Chandran K, Fathima G. Context-aware augmented reality using human-computer interaction models. Journal of Control and Decision, 2024, 11(1): 1-14.

[6] Halim Z, Waqar M, Tahir M. A machine learning-based investigation utilizing the in-text features for the identification of dominant emotion in an email. Knowledge-Based Systems, 2020, 208(15):106443-106459.

[7] Britzolakis A, Kondylakis H, Papadakis N. A Review on Lexicon-Based and Machine Learning Political Sentiment Analysis Using Tweets. International Journal of Semantic Computing, 2021, 14(4):517-563.

[8] Alasmari W A, Abdelhafez H A. Twitter Sentiment Analysis for Reviewing Tourist Destinations in Saudi Arabia Using Apache Spark and Machine Learning Algorithms. Journal of computer sciences, 2022, 18(3):210-221.

[9] Chirgaiya S, Sukheja D, Shrivastava N, Rawat R. Analysis of sentiment based movie reviews using machine learning techniques. Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 2021, 41(5):5449-5456.

[10] Kumar S, Gahalawat M, Roy P P, Dogra D P. Exploring Impact of Age and Gender on Sentiment Analysis Using Machine Learning. Electronics, 2020, 9(2):374-387.

[11] Liewlom P. Alternative Rule Reasoning: Association Rule Tree Reasoning with a Constraining Rule Ascertained using a Reasoning Framework in 2D Interestingness Area. IAENG Internaitonal journal of computer science, 2021, 48(3):619-633.

[12] Li C, Yang X, Yin K, Chang Y, Wang Z, Yin G. Pedestrian re-identification based on attribute mining and reasoning. IET Image Processing, 2021, 15(11):2399-2411.

[13] Lin Y R, Lee C H, Lu M C. Robust tool wear monitoring system development by sensors and feature fusion. Asian Journal of Control: Affiliated with ACPA, the Asian Control Professors, Association, 2022, 24(3):1005-1021.

[14] Zhang L, Li Y, Yi J, Wang J. Research On The Application Of Network Intrusion Feature Extraction In Power Network. International Journal of Autonomous and Adaptive Communications Systems, 2021, 14(4):342-353.

[15] Guo B, Zhang Y, Gao J, Li C, Hu Y. SGLBP: Subgraph-based Local Binary Patterns for Feature Extraction on Point Clouds. Computer Graphics Forum: Journal of the European Association for Computer Graphics, 2022, 41(6):51-66.

[16] Mutanov G, Karyukin V, Mamykova Z. Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms. Computers, Materials and Continua, 2021, 69(1):913-930.

[17] Kelvin Leong, Anna Sung. An Exploratory Study of How Emotion Tone Presented in A Message Influences Artificial Intelligence (AI) Powered Recommendation System. Journal of Technology & Innovation, 2023, 3(2): 80-84.

[18] Zhao Y, Guo M, Sun X, Chen X, Zhao F. Attention-based sensor fusion for emotion recognition from human motion by combining convolutional neural network and weighted kernel support vector machine and using inertial measurement unit signals. IET signal processing, 2023, 17(4): 12201-12212.

[19] Ou Z, Wang H, Zhang B, Liang H, Hu B, Ren L, Liu Y, Zhang Y, Dai C, Wu H, Li W, Li X. Early identification of stroke through deep learning with multi-modal human speech and movement data. Neural Regeneration Research, 2024, 20(1): 234-241.

[20] Purohit J, Dave R. Leveraging Deep Learning Techniques to Obtain Efficacious Segmentation Results. Archives of Advanced Engineering Science, 2023, 1(1):11-26.

[21] Chinthamu N, Karukuri M. Data Science and Applications. Journal of Data Science and Intelligent Systems, 2023, 1(1): 83-91.

[22] Shanqing Z, Yujie C, Yiheng M, Jianfeng L, Li L, Rui B. A multi-level feature weight fusion model for salient object detection. Multimedia Systems, 2023, 29(3): 887-895.

Deep Learning for Coronary Artery Stenosis Localization: Comparative Insights from Electrocardiograms (ECG), Photoplethysmograph (PPG) and Their Fusion

Mohd Syazwan Md Yid[1], Rosmina Jaafar[2], Noor Hasmiza Harun[3], Mohd Zubir Suboh[4], Mohd Shawal Faizal Mohamad[5]

Dept. Electrical, Electronic & Systems Engineering-Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia[1, 2]

Medical Engineering Technology Section, British Malaysian Institute, Universiti Kuala Lumpur, Gombak, Malaysia[1, 3, 4]

Dept. of Medicine, Hospital Canselor Tuanku Muhriz, Cheras, Kuala Lumpur, Malaysia[5]

***Abstract*—Coronary artery stenosis (CAS) is a critical cardiovascular condition that demands accurate localization for effective treatment and improved patient outcomes. This study addresses the challenge of enhancing CAS localization through a comparative analysis of deep learning techniques applied to electrocardiogram (ECG), photoplethysmograph (PPG), and their combined signals. The primary research question centers on whether the fusion of ECG and PPG signals, analyzed through advanced deep learning architectures, can surpass the accuracy of individual modalities in localizing stenosis in the left anterior descending (LAD), left circumflex (LCX), and right coronary arteries (RCA). Using a dataset of 7,165 recordings from CAS patients, three models—CNN, CNN-LSTM, and CNN-LSTM-ATTN—were evaluated. The CNN-LSTM-ATTN model achieved the highest localization accuracy (98.12%) and perfect AUC scores (1.00) across all arteries, demonstrating the efficacy of multimodal signal integration and attention mechanisms. This research highlights the potential of combining ECG and PPG signals for non-invasive CAS diagnostics, offering a significant advancement in real-time clinical applications. However, limitations include the relatively small dataset size and the focus on single-lead ECG and PPG signals, which may affect the generalizability to broader populations. Future studies should explore larger datasets and multi-lead signal integration to further validate the findings.**

***Keywords*—Coronary artery stenosis; deep learning; ECG; PPG; ECG-PPG fusion; CNN; LSTM; attention mechanism**

I. INTRODUCTION

Coronary artery stenosis, and other cardiovascular diseases, remain a major cause of death globally, and thus present a major public health concern [1]. Accurate and timely diagnosis of these conditions, as well as continuous monitoring, are crucial for effective treatment and management, ultimately improving patient outcomes [2]. Coronary artery stenosis, a common and serious manifestation of cardiovascular disease, has been the subject of extensive research, with a focus on developing accurate, non-invasive, and accessible detection methods that can enable early intervention and better disease management strategies [3], [4]. This research emphasis underscores the importance of advancing the field of cardiovascular disease diagnosis and monitoring to address this pressing global health concern.

ECG have long been used in the diagnosis and monitoring of cardiac conditions, and recent advancements in machine learning have shown promising results in ECG-based detection and classification of various cardiovascular diseases. For example, changes in the ECG waveform, such as ST-segment depression or T-wave inversion, can be indicative of myocardial ischemia caused by coronary artery stenosis [5]. A study by [6] explores cardiovascular disease (CVD) prediction using machine learning techniques on ECG and physiological data, finding that an artificial neural network (ANN) model achieves the highest predictive accuracy (90%) by utilizing significant parameters such as the R-R interval, RMSSD, blood pressure, and cholesterol levels, highlighting its potential as a non-invasive diagnostic tool for early CVD detection. Deep learning models trained on ECG data have demonstrated the ability to detect and localize specific patterns associated with different regions of coronary artery disease, such as LAD artery, LCX artery, or RCA obstructions [7].

In contrast, photoplethysmographic is an opto-electrical technique that uses light to quantify hemodynamic changes that is an important aspect of cardiovascular analysis. PPG signals can capture blood volume changes in the peripheral vasculature, which can be indicative of changes in the cardiovascular system, such as those associated with coronary artery disease [8], [9]. For example, [10] investigates photoplethysmography (PPG) as a non-invasive alternative to assess coronary artery disease (CAD) severity, finding that a Discriminant Analysis classifier achieved 88.46% accuracy in detecting severe stenosis, thus highlighting PPG's potential for CAD pre-diagnosis in resource-limited or pandemic-impacted environments. In addition, the analysis of PPG waveforms has shown potential in detecting and monitoring conditions like coronary artery stenosis, as it can provide insights into the vascular dynamics and hemodynamic changes related to this cardiovascular disorder [11].

Deep learning, a powerful subset of machine learning, has emerged as a promising approach for the analysis of various biomedical signals, including electrocardiograms and photoplethysmography [12]. These techniques have the ability to extract complex patterns and features from the data, enabling accurate classification and localization of cardiovascular conditions, such as coronary artery stenosis [13]. Among the deep learning methods, Convolutional Neural Networks have demonstrated good efficiency for identification of both ECG and PPG signals for diagnostics of coronary artery disease location [14]. By leveraging the hierarchical feature extraction capabilities of CNNs, researchers have been able to develop robust models for identifying characteristic patterns in cardiac data that are indicative of coronary artery stenosis [15].

LSTM networks, have also shown promise in the analysis of ECG and PPG signals for the detection and monitoring of coronary artery disease [16]. LSTM models overcome the limitation of capturing long range of temporal dependencies and hence suitable for processing these continuous physiological signals [17]. Some recent works have proved that LSTM networks can capture ECG and PPG patterns related to coronary artery stenosis, indicating that deep learning structures might contribute for the localization and diagnosis of this cardiovascular disease [18].

Moreover, it is also understood that the application of attention-based mechanisms in deep learning systems can potentially improve the assessment and visualization of coronary artery disease based on these multiple modal signals [19]. Attention-based architectures, such as Transformer models, have shown promising results in healthcare applications by allowing the models to focus on the most relevant features and patterns in the data, which can be crucial for the precise localization of coronary artery stenosis [20].

The combination of ECG and PPG data may provide a more comprehensive understanding of the cardiovascular system, potentially leading to improved accuracy in the localization of coronary artery stenosis. This study aims to conduct a comparative analysis of deep learning approaches using ECG, PPG, and the combined ECG-PPG modalities to enable accurate and non-invasive localization of coronary artery stenosis, which could enhance early diagnosis and management of this critical cardiovascular condition.

In the following sections, this paper details the methodology for signal acquisition, dataset preparation, and preprocessing, followed by the development and evaluation of three deep learning architectures: CNN, CNN-LSTM, and CNN-LSTM-ATTN. A comprehensive comparative analysis of these models using ECG, PPG, and combined ECG-PPG signals is presented, highlighting the advantages of multimodal signal fusion and attention mechanisms. Finally, the results are discussed in relation to prior studies, and conclusion was made with insights into the clinical implications of our findings and potential directions for future research. This structure aims to provide readers with a clear roadmap of the study, fostering deeper engagement with the content.

II. Related Works

Tao et al. [21] developed an automatic ischemic heart disease (IHD) detection and localization system using magnetocardiography (MCG) signals and machine learning methods. They used 164 features derived from the MCG recordings and divide into three groups, which are time domain feature, frequency domain feature, information theory feature, and compared the performance of many classifiers. Their ensemble model of SVM and XGBoost achieved high performance in IHD detection, with an accuracy of 94.03%, precision of 86.56%, recall of 94.78%, and an AUC of 0.98. For stenosis localization in the LAD, LCX, and RCA, they employed 18 time-domain features with XGBoost, achieving accuracies of 74%, 68%, and 65%, respectively. The study demonstrated that features related to T-wave repolarization synchronicity and magnetic field patterns were critical in both detection and localization, providing a non-invasive, fast, and accurate tool for clinical use.

In their work, Huang et al. [22] propose the creation of an AI-based ECG algorithm that will help predict and indicate the location of angiography-verified CAD. The study employs a CNN model to examine 12-lead ECG data of patients with CAD who have been verified through ICA. The dataset consists of clinical data from 2303 CAD patients and ECG data of 1053 healthy patients as well as 12,954 ECG records. The CNN model provided an AUC of 0.869 to identify CAD, and specific AUC of 0.885, 0.816, 0.776 to detect stenosis in LAD, LCX, RCA respectively. The AUC of the model reached 0.973 in the case when ECGs demonstrated features of myocardial ischemia. The study proves that the AI-based algorithm on the ECG signal as the primary diagnostic tool can be effective and non-invasive for identifying severe CAD and localized stenosis.

Roopa and Harish [23] proposed an automated system for localizing thrombus in coronary arteries using 12-lead ECG signals and an Information Fuzzy Network (IFN). Their method utilizes ECG feature extraction techniques, including the Stockwell Transform and Nearest-Neighbor Interpolation, to identify key features like ST-segment deviations, time intervals, and peak amplitudes in the ECG waveform. An initial rule-based system is then used to separate ischemic and non-ischemic signals and to determine the culprit artery, which might be LAD, RCA, LCX or another artery. This experimental study showed that the proposed system has an accuracy of 92.30%; sensitivity of 87.50%; and specificity of 100%; thus, it could be a valuable noninvasive solution for diagnosing coronary artery blocks and supporting clinical decision making.

Previous studies on CAD localization have primarily focused on using single physiological signals, including MCG and ECG, alongside traditional machine learning models like XGBoost and CNNs. However, the integration of multimodal signals, particularly ECG and PPG, remains underexplored. Furthermore, while advanced deep learning techniques, such as LSTM networks and attention mechanisms, are increasingly recognized for their ability to capture both spatial and temporal dependencies, their application in CAD localization is limited.

The existing literature often lacks the incorporation of attention-based models that can dynamically focus on the most relevant signal features, potentially enhancing diagnostic accuracy. Additionally, few comparative analyses have been conducted to evaluate the relative effectiveness of ECG, PPG, and their combined modalities for CAD localization. Moreover, most studies have focused on offline models, with limited exploration of real-time clinical applications. This study addresses these gaps by employing deep learning models that integrate CNNs, LSTMs, and attention mechanisms to analyze multimodal signals, offering a comprehensive and more accurate approach to CAD localization with potential for real-time clinical implementation.

III. Materials and Methods

A. Study Population

This research involved a cohort of patients diagnosed with significant coronary artery disease, as confirmed through angiography. Participants, aged 20 to 65 years, were selected based on the presence of severe stenosis, which was quantitatively assessed using coronary angiography. Only individuals with no prior history of CAD were included. The study received ethical clearance from the Research Ethics Committee of Universiti Kebangsaan Malaysia (UKMPPI/111/8/JEP-2020-806), and all participants provided written informed consent.

B. Data Collection

The study followed a protocol for recording both ECG and PPG signals concurrently from participants, as illustrated in Fig. 1. A total of 60 patients with confirmed significant coronary artery disease through angiography were included, resulting in a dataset containing 7,156 simultaneous single-lead ECG and PPG recordings. The patients were categorized into three groups based on their angiography findings: All those patients who had stenosis in LAD, LCX, and RCA arteries were included in this study.

In the LAD group, 27 patients provided 3,884 concurrent single-beat ECG and PPG recordings. The LCX group consisted of 16 patients, contributing 1,565 recordings, while the RCA group included 17 patients, resulting in 1,707 recordings. Fig. 2 displays sample single-beat ECG and PPG signals for each artery group (LAD, LCX, and RCA).

The dataset was derived from the MAX86150EVS ECG/PPG module, which recorded standard single-lead ECG and PPG signals at a 400 Hz sampling rate for 10 minutes per patient. Signal processing techniques, such as baseline wander removal, smoothing, and segmentation, were applied to isolate individual cardiac cycles and enhance the quality of single-beat signals. This resulted in a final dataset of 7,165 simultaneous single-beat ECG and PPG waveforms, each consisting of 187 data points per sample. This dataset is used in the next sections to develop and evaluate various deep learning models.

C. Dataset Preparation and Preprocessing

Prior to proceeding with the deep learning models, the collected data is partitioned into three distinct datasets: the ECG dataset, the PPG dataset, and the combined ECG and PPG dataset. These datasets share the same number of samples and targets. The purpose of this division is to enable the training and evaluation of deep learning models on the individual modalities as well as the combined modality. For the combined ECG and PPG dataset, each sample comprises a concatenation of both the ECG and PPG signals into a single feature vector. This integration aims to leverage the complementary information present in the ECG and PPG signals to potentially enhance the overall performance of the CAD stenosis localization task. Fig. 3 shows examples of the concatenated ECG and PPG signals belonging to the classes, namely LAD, LCX, and RCA.

In the experiment, the datasets are split into subsets with the training, validation, and test data that should not overlap, 70% of the dataset for training, 10% for validation and the remaining 20% for testing. It is noteworthy to mention that the training and test sets are completely disjoint in terms of patient-wise separation, adhering to best practices for evaluating deep learning models. All three groups of datasets are subjected to the same train-validation-test split ratios.

The three datasets seem to be unbalanced, with the LAD group containing a significantly larger number of samples compared to the LCX and RCA groups. Such distribution skew could result into biased deep learning models and performance of the models when trained and tested on the datasets [22]. As a result, to address this problem, we resorted to applying down sampling where we sampled a proportionate sample of the majority class to create a new smaller set of samples that was equivalent in size to the other two classes. Fig. 4 shows bar chart representations of the distribution of samples before and after the down-sampling process for each group.

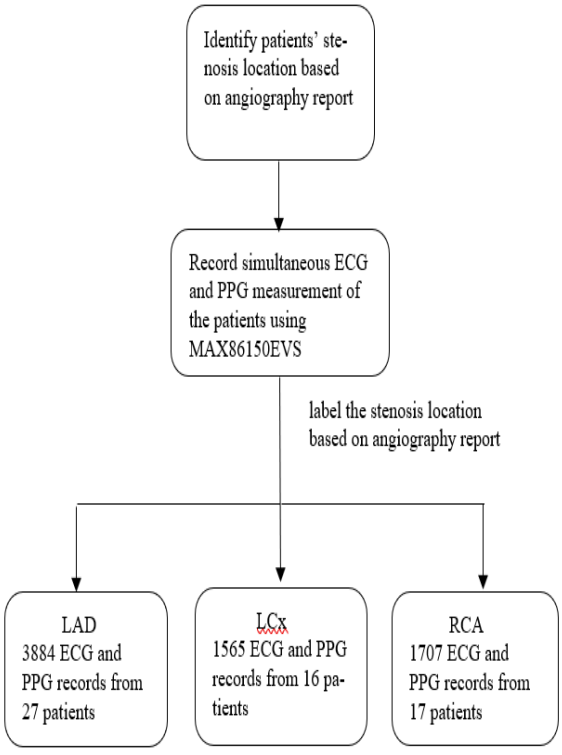


Fig. 1. Data collection procedure.

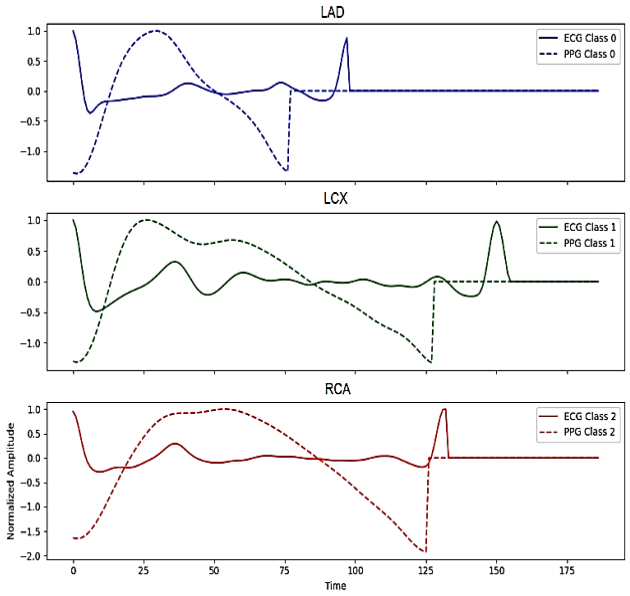


Fig. 2. ECG and PPG data samples for a patient with blockages in the LAD, LCX, and RCA arteries.

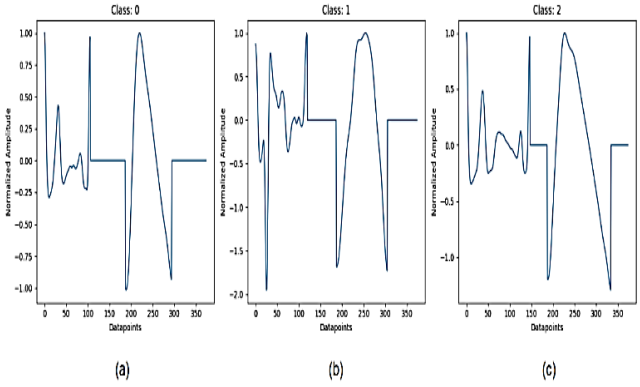


Fig. 3. Combined ECG and PPG samples for a patient with blockages in (a) LAD, (b) LCX, and (c) RCA arteries.

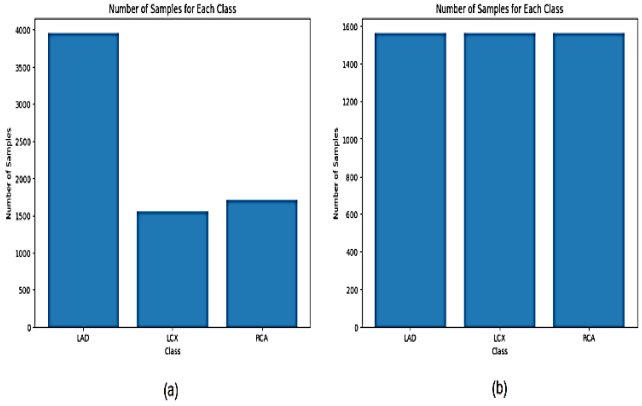


Fig. 4. (a) Dataset showing class imbalance prior to the balancing process, and (b) Dataset with balanced classes after the data balancing procedure.

D. Model Build-up

In this study, a series of deep learning models were developed to classify CAD stenosis location using ECG, PPG the combination of both signals. These models were developed using Google Colaboratory (Colab), a cloud-based platform that provides access to powerful GPU resources and comprehensive deep learning libraries, such as TensorFlow and PyTorch. To effectively process the complex time-series data, three models were explored: a CNN-based model, a hybrid CNN-LSTM model, and an advanced CNN-LSTM-ATTN model. Each architecture was designed to leverage different aspects of deep learning, namely spatial feature extraction, temporal dependency modeling, and attention-based focus mechanisms, to improve classification accuracy.

The initial model utilizes a CNN, designed to automatically extract features from raw ECG and PPG signals. This is a simple sequence model starting with one-dimensional convolutional layer (Conv1D) with 32 filters and 3×1 kernel, that is then passed through a max-pooling (MaxPooling1D) to minimize the dimensionality of the formed feature maps. The convolutional layers identify local features in the input signals; max-pooling reduces dimensionality and masks significant characteristics. Further, subsequent feature maps are subjected to flatten layer and then fed to the fully connected layer containing 128 units with ReLU activation followed by softmax output layer for classification into three diagnostic categories. This CNN model efficiently captures spatial patterns in the signals, establishing a strong baseline for coronary artery disease classification.

The second model enhances the CNN by incorporating LSTM layers to capture temporal dependencies in the sequential ECG and PPG data. Following the convolutional and max-pooling layers, two LSTM layers, each with 50 units, are added. The first LSTM layer is set to return sequences, enabling the second LSTM layer to process both short- and long-term dependencies within the signal data. This combination of CNN and LSTM layers allows the model to extract spatial features while also understanding their temporal progression, thereby boosting its ability to classify coronary artery disease based on dynamic signal changes. A final dense layer with softmax activation generates a probability distribution across the three diagnostic categories.

The third and most advanced architecture incorporates an attention mechanism to further enhance the model's ability to focus on the most relevant portions of the input signals. The model begins with a convolutional layer with 32 filters and a max-pooling layer, followed by an LSTM layer with 50 units and return_sequences=True. After the LSTM layer, an attention mechanism is applied to dynamically weigh the importance of each time step in the signal, allowing the model to focus on the most significant information. This attention mechanism improves the interpretability of the model by highlighting the critical sections of the ECG and PPG signals that contribute to the classification. The attention-weighted features are then flattened and passed to a softmax output layer for final classification. The architectural representation of the model is represented in Fig. 5 after performing data balancing process and the overall flow of the study is represented in Fig. 6.

By combining CNNs for spatial feature extraction, LSTMs for temporal modeling, and attention mechanisms for focused learning, these architectures present a robust approach to CAD classification. The CNN-based model provides a solid baseline by capturing local signal patterns, while the CNN-LSTM hybrid improves the model's ability to learn temporal dependencies. The CNN-LSTM-ATTN model further enhances performance by focusing on the most relevant parts of the signals, making it the most comprehensive and accurate architecture for the task of CAD diagnosis based on ECG and PPG signals.

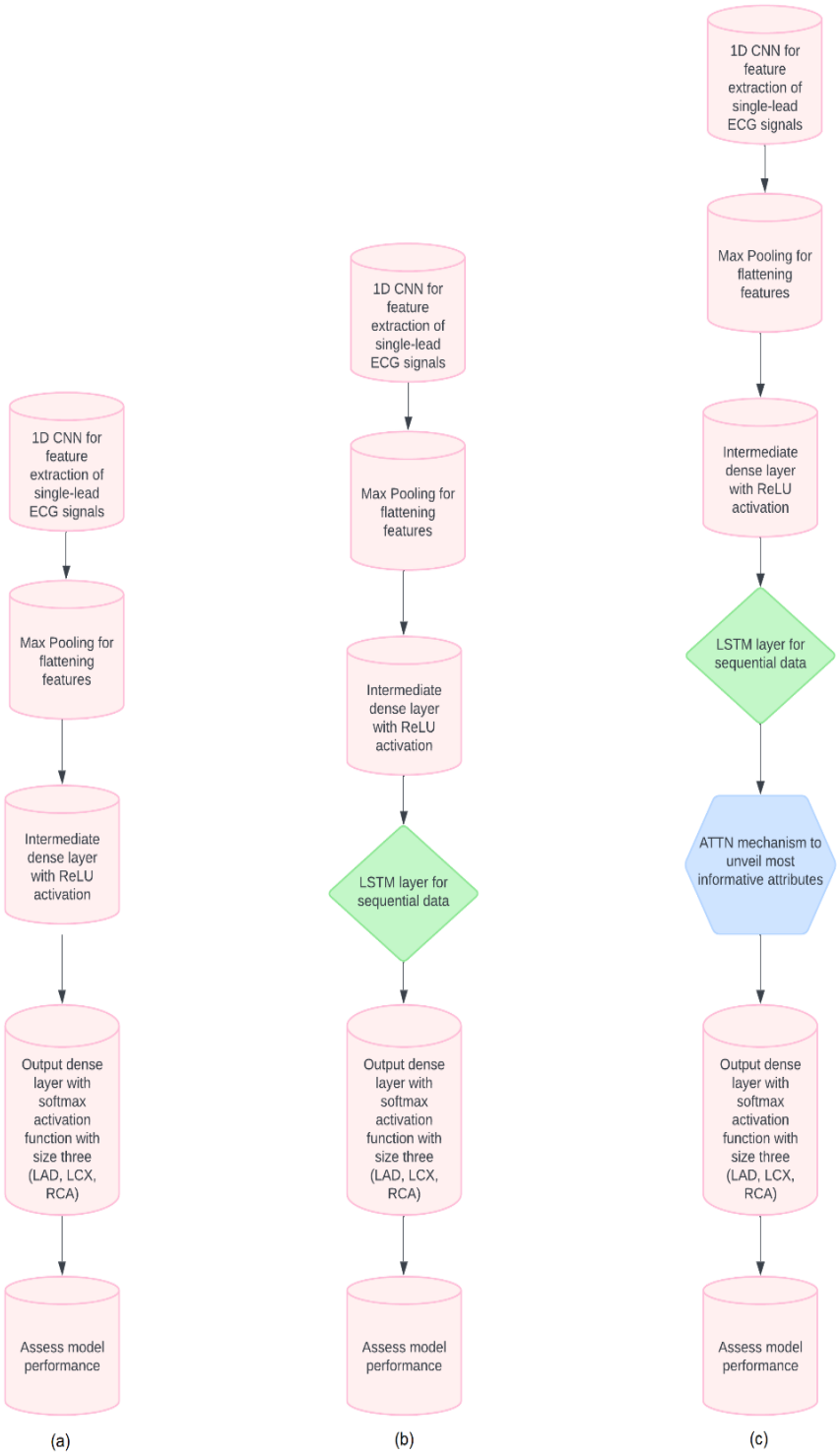


Fig. 5. Architecture for coronary artery blockage localization prediction model for (a) CNN, (b) CNN+LSTM, (c) CNN+LSTM+ATTN.

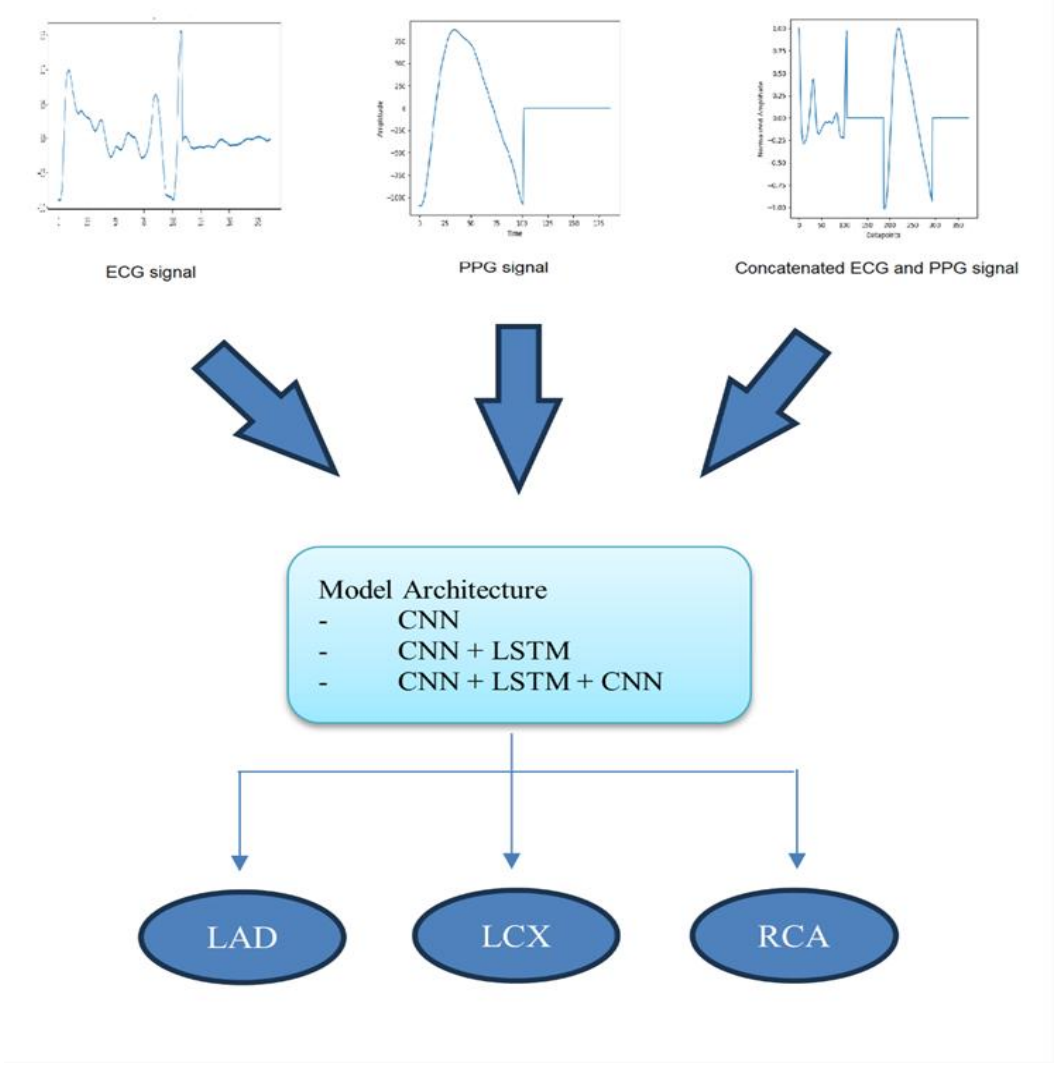


Fig. 6. Model architecture for predicting the localization of coronary artery blockages.

E. Training Process

The training process for each model involves the use of the Adam optimizer with a learning rate of 0.001 and a batch size of 32. The categorical cross-entropy loss function is applied to optimize the models during training. To reduce the risk of overfitting, a dropout layer with a 0.2 rate is introduced after the first fully connected layer in all models. Moreover, the practice of early stopping is used to terminate the learning process when there is a failure to get any improvements in the validation error. The models are trained for up to 100 epochs, with the best-performing model, based on validation accuracy, saved for final evaluation on the test set.

F. Evaluation Metrics

The study aimed to evaluate an AI-enhanced ECG as a means of identifying the location of coronary artery stenosis from single-lead standard ECG recordings, together with comparison based on full lead set. The testing and the performance of the proposed deep learning models was assessed in terms of accuracy metric, area under the receiver operating characteristic curve (AUC-ROC) and the confusion matrix.

The measurement of accuracy is the extent to which samples have been classified correctly; it is the number of correctly predicted over the total samples. This is the ratio of true positive calculated by dividing it by all the positive cases that have been identified as positive by the application. Precision and recall are measurement indicators used in this study. Precision deals with the number of actually positive cases found to be positive by the application, calculated by dividing the true positive value by the actual positive samples. The true positive rate depicts the ratio of the actual positive samples to the total numbers of positive samples that were classified as such the false positive rate depicts the ratio of the negative samples that were classified as positive. Also, the true negative rate expresses the share of real negative patterns that correctly classified and the false negative rate describes the percentage of actual positive patterns that misleading classified as negative.

The area that the curve corresponding to the ROC forms, in short AUC-ROC serves as a rich measure that encapsulates the ability of a classification model in terms of its ability of class separation. Also known as Receiver Operating Characteristic,

this metric measures the true positives along the true negatives across the range of classification thresholds that gives a balanced measure of a model's discriminative abilities.

The confusion matrix is used to display accuracy of the model in terms of the three diagnostics types. It gives an account of the right and wrong classifications made by the model, which aids in determination of its efficiency for each class. This characteristic of the confusion matrix helps to analyze the advantages and disadvantages of the model in terms of identifying the various conditions of coronary artery disease.

IV. Results And Discussion

To enhance the model's performance, we carried out several experiments to assess the effectiveness of various deep learning architectures in classifying CAD stenosis locations using ECG, PPG and combination of ECG and PPG signals. The architectures examined in this study comprised a baseline CNN network, a hybrid CNN-LSTM model, and a CNN-LSTM-ATTN model. Table I provides a summary of the performance metrics, including classification accuracy and AUC, for each of these models on the test set. Additionally, the corresponding ROC curve and confusion matrix based on the best model obtained are illustrated in Fig. 7.

In Table I, the results highlight the comparative performance of three deep learning models for CAD stenosis localization using ECG, PPG, and their combined signals. These models include a CNN, a hybrid CNN-LSTM model, and an advanced CNN-LSTM-ATTN.

TABLE I. Performance Metrics of the Three Models Utilized in this Study

Signal	Model	Accuracy	AUC		
			LAD	***LCX***	***RCA***
ECG	CNN	94.25%	0.99	0.98	0.99
	CNN + LSTM	95.53%	0.99	0.99	0.99
	CNN + LSTM + ATTN	97.61%	0.99	0.98	0.99
PPG	CNN	86.10%	0.95	0.96	0.96
	CNN + LSTM	91.97%	0.97	0.99	0.98
	CNN + LSTM + ATTN	92.25%	0.97	0.98	0.98
Combined ECG & PPG	CNN	94.69%	0.96	0.97	0.97
	CNN + LSTM	92.47%	0.98	0.99	0.98
	CNN + LSTM + ATTN	98.12%	1.00	1.00	1.00

For ECG-based detection, the CNN model achieved an accuracy of 94.25%, AUC scores ranging from 0.98 to 0.99 across the three coronary arteries: LAD, LCX, and RCA. Integrating the LSTM layer improved accuracy to 95.53%, indicating the added value of temporal feature extraction, while AUC scores remained consistently high.

The introduction of the attention mechanism (CNN-LSTM-ATTN) led to a significant performance boost, achieving 97.61% accuracy and maintaining near-perfect AUC values for all three arteries, underscoring the ability of the attention mechanism to focus on the most relevant features in the data.

For PPG signals, the CNN model started with a lower accuracy of 86.10%, yet the inclusion of LSTM and attention mechanisms progressively improved the results. The CNN-LSTM model raised accuracy to 91.97%, and the CNN-LSTM-ATTN model further increased it to 93.32%, while AUC scores improved, especially for the LCX and RCA arteries.

When the ECG and PPG signals were combined, the results demonstrated the most substantial improvement. The CNN model reached 94.69% accuracy, and while the CNN-LSTM model saw a slight dip in accuracy to 92.47%, the addition of the attention mechanism significantly enhanced performance, resulting in 98.12% accuracy and perfect AUC scores of 1.00 for all three coronary arteries. This illustrates the clear advantage of combining both signal modalities, which, coupled with advanced deep learning techniques, maximized diagnostic accuracy and precision.

Overall, the results indicate that while individual signals (ECG or PPG) provide valuable diagnostic insights, combining both signals with sophisticated deep learning architectures, especially with attention mechanisms, offers superior performance in localizing coronary artery stenosis. The CNN-LSTM-ATTN model demonstrates exceptional potential for clinical application, offering a non-invasive, highly accurate method for detecting blockages in major coronary arteries.

Table II presents a comparative analysis between the best model obtained from the study and three previous studies—Tao et al. [21], Huang et al. [22], and Roopa and Harish [23]—in terms of accuracy and Area Under the Curve (AUC) for coronary artery stenosis (CAS) localization. The proposed model, which employs a CNN-LSTM-ATTN architecture integrating both ECG and PPG signals, achieves an impressive overall accuracy of 98.12%, substantially outperforming earlier models. In terms of AUC, the model demonstrates exceptional performance, achieving perfect scores of 1.00 for detecting stenosis in LAD, LCX, and RCA. However, Tao et al.'s XGBoost based model, which employed magnetocardiography, yielded lower AUC of 0.74, 0.68, and 0.65 for these arteries, respectively. Huang et al.'s CNN model, only with ECG signals, achieved the AUCs of 0.89, 0.82, and 0.78 for LAD, LCX, and RCA respectively. At the same time, Roopa and Harish proposed an ECG-based model that yielded an accuracy of 92.3%, but AUC values were not disclosed. The exceptional performance of the proposed model achieved because the proposed model relies on the fusion of the ECG and PPG signals; the two methods improve both the feature extraction and temporal models. Moreover, attention mechanisms help to pay more attention to useful signal characteristics, which in turn contributes to better definition of the location of blockages in coronary arteries.

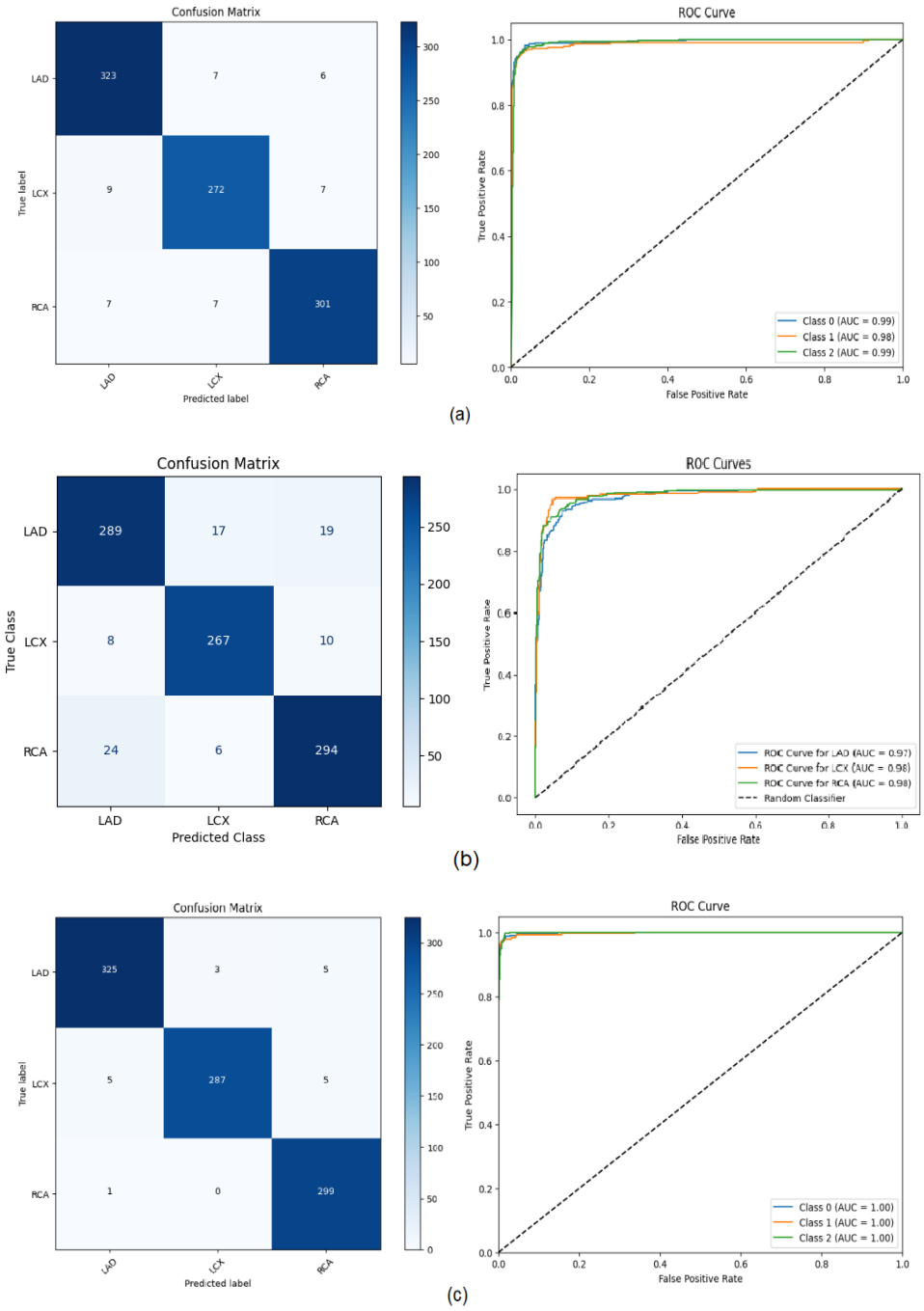


Fig. 7. Confusion matrices and ROC curves of the best models obtained (CNN+LSTM+ATTN) for (a) ECG, (b) PPG, and (c) combination of ECG and PPG signals.

TABLE II. COMPARISON OF PERFORMANCE WITH PRIOR STUDIES

Author, Year	Data	AI Model	Acc.(%)	AUC		
				LAD	*LCX*	*RCA*
Tao et al., 2018 [21]	MCG	XGBoost	NA	0.74	0.68	0.65
Huang et al., 2022 [22]	12 lead ECG	InceptionV3		0.89	0.82	0.78
Roopa and Harish, 2019 [23]	12 lead ECG	IFN	92.3	NA		
Proposed work	Combined simultaneous single lead ECG and PPG	CNN + LSTM + ATTN	98.12	1.00	1.00	1.00

V. CONCLUSION

This paper proposes a method to diagnose and examine coronary artery stenosis (CAS) using ECG and PPG signals, further aided by deep learning algorithms. The comparison between three models—CNN, CNN-LSTM, and CNN-LSTM-ATTN—verifies that the effects of applying progressive AI structures upon CAS localization are prominent. The results demonstrate that both ECG and PPG signals are informative individually; however, when combined and processed through the CNN-LSTM-ATTN model, the highest classification accuracy of 98.12% and AUC scores of 1.00 were achieved for stenosis in LAD, LCX, and RCA.

These findings underscore the necessity of fusing multiple physiological signals to enhance CAS localization predictability and highlight the effectiveness of the CNN-LSTM-ATTN model in capturing spatial and temporal characteristics of multiple physiological signals while selectively attending to essential features. This approach shows great potential for non-invasive diagnosis of coronary artery diseases and the localization of obstructive lesions, which current clinical imaging techniques may inadequately address.

However, the study is not without limitations. The dataset used was relatively small, with data collected from a single hospital, which may limit the generalizability of the findings. Additionally, the study focused solely on single-lead ECG and PPG signals, excluding the potential benefits of multi-lead configurations or other physiological signals. Future research should explore larger, more diverse datasets and investigate the integration of additional modalities to further validate and enhance the proposed method. Moreover, real-time clinical implementation remains a challenge that warrants further development to ensure the practicality and reliability of the approach in routine healthcare settings.

In conclusion, while the proposed method demonstrates promising results for non-invasive CAS diagnostics, addressing these limitations will be crucial for broader adoption and impact in clinical practice.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by the Ministry of Higher Education Malaysia through the Trans-disciplinary Research Grant Scheme (TRGS/1/2019/UKM/01/4/3).

REFERENCES

[1] A. Joshi and M. Shah, “Coronary Artery Disease Prediction Techniques: A Survey,” in Lecture Notes in Networks and Systems, Springer Science and Business Media Deutschland GmbH, 2021, pp. 593–604. doi: 10.1007/978-981-16-0733-2_42.

[2] L. Zhang et al., “Global, Regional, and National Burdens of Ischemic Heart Disease Attributable to Smoking From 1990 to 2019,” J Am Heart Assoc, vol. 12, no. 3, Feb. 2023, doi: 10.1161/JAHA.122.028193.

[3] B. Zhu and G. He, “Myocardial infarction localization and blocked coronary artery identification using a deep learning method,” in Proceedings - 11th International Conference on Prognostics and System Health Management, PHM-Jinan 2020, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 514–519. doi: 10.1109/PHM-Jinan48558.2020.00100.

[4] M. Z. Suboh, R. Jaafar, N. A. Nayan, S. F. Mohamad, N. H. Harun, and H. A. Hamid, “TCTAP A-080 A Comprehensive Analysis on Severity of Stenosis Detection in Coronary Arteries Using Synchronized Electrocardiogram and Photoplethysmogram,” J Am Coll Cardiol, vol. 81, no. 16, p. S53, Apr. 2023, doi: 10.1016/j.jacc.2023.03.109.

[5] K. C. Siontis, P. A. Noseworthy, Z. I. Attia, and P. A. Friedman, “Artificial intelligence-enhanced electrocardiography in cardiovascular disease management,” Jul. 01, 2021, Nature Research. doi: 10.1038/s41569-020-00503-2.

[6] N. A. Nayan et al., “Cardiovascular disease prediction from electrocardiogram by using machine learning,” International journal of online and biomedical engineering, vol. 16, no. 7, pp. 34–48, 2020, doi: 10.3991/ijoe.v16i07.13569.

[7] T. Reasat and C. Shahnaz, “Detection of inferior myocardial infarction using shallow convolutional neural networks,” in 2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC), IEEE, Dec. 2017, pp. 718–721. doi: 10.1109/R10-HTC.2017.8289058.

[8] N. Mangathayaru, B. P. Rani, V. Janaki, L. S. Kotturi, M. Vallabhapurapu, and G. Vikas, “Heart Rate Variability for Predicting Coronary Heart Disease using Photoplethysmography,” in 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), IEEE, Oct. 2020, pp. 664–671. doi: 10.1109/I-SMAC49090.2020.9243316.

[9] K. Azudin, K. B. Gan, R. Jaafar, and M. H. Ja’afar, “The Principles of Hearable Photoplethysmography Analysis and Applications in Physiological Monitoring–A Review,” Jul. 01, 2023, Multidisciplinary Digital Publishing Institute (MDPI). doi: 10.3390/s23146484.

[10] M. Zubir Suboh, R. Jaafar, N. Anuar Nayan, N. Hasmiza Harun, M. Shawal Faizal Mohamad, and H. Abdul Hamid, “ALTERNATIVE METHOD TO PRE-DIAGNOSED CORONARY ARTERY DISEASE USING PHOTOPLETHYSMOGRAPHY: A LESSON FROM COVID-19 PANDEMIC (Kaedah Alternatif untuk Diagnosis Awal Penyakit Arteri Koronari melalui Fotopletismografi: Satu Pengajaran daripada Pandemik COVID-19)”, doi: 10.17576/JH-2024-1601-03.

[11] F. M. Dias et al., “Quality Assessment of Photoplethysmography Signals For Cardiovascular Biomarkers Monitoring Using Wearable Devices,” Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.08766

[12] M. Syazwan et al., “Deep Learning-Driven Localization of Coronary Artery Stenosis Using Combined Electrocardiograms (ECGs) and Photoplethysmograph (PPG) Signal Analysis,” 2024. [Online]. Available: www.ijacsa.thesai.org

[13] H. M. Rai, K. Chatterjee, A. Dubey, and P. Srivastava, “Myocardial Infarction Detection Using Deep Learning and Ensemble Technique from ECG Signals,” in Lecture Notes in Networks and Systems, Springer Science and Business Media Deutschland GmbH, 2021, pp. 717–730. doi: 10.1007/978-981-16-0733-2_51.

[14] B. Au, F. Warner, P. 1, S.-X. Li, H. Krumholz, and M. D. Sm, “Automated Characterization of Stenosis in Invasive Coronary Angiography Images with Convolutional Neural Networks.”

[15] M. Zreik, R. W. van Hamersvelt, J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Isgum, “A Recurrent CNN for Automatic Detection and Classification of Coronary Artery Plaque and Stenosis in Coronary CT Angiography,” IEEE Trans Med Imaging, vol. 38, no. 7, pp. 1588–1598, Jul. 2019, doi: 10.1109/TMI.2018.2883807.

[16] A. Makhir, M. H. El Yousfi Alaoui, and L. Belarbi, “Comprehensive Cardiac Ischemia Classification Using Hybrid CNN-Based Models,” International journal of online and biomedical engineering, vol. 20, no. 3, pp. 154–165, 2024, doi: 10.3991/ijoe.v20i03.45769.

[17] J. Laitala et al., “Robust ECG R-peak detection using LSTM,” in Proceedings of the 35th Annual ACM Symposium on Applied Computing, New York, NY, USA: ACM, Mar. 2020, pp. 1104–1111. doi: 10.1145/3341105.3373945.

[18] R. Xiao, Y. Xu, M. M. Pelter, D. W. Mortara, and X. Hu, “A Deep Learning Approach to Examine Ischemic ST Changes in Ambulatory ECG Recordings.,” AMIA Jt Summits Transl Sci Proc, vol. 2017, pp. 256–262, 2018.

[19] I. Adalioglu, M. Ahishali, A. Degerli, S. Kiranyaz, and M. Gabbouj, “SAF-Net: Self-Attention Fusion Network for Myocardial Infarction Detection using Multi-View Echocardiography,” Sep. 2023, [Online]. Available: http://arxiv.org/abs/2309.15520

[20] Z. Zhao, "Transforming ECG Diagnosis:An In-depth Review of Transformer-based DeepLearning Models in Cardiovascular Disease Detection," Jun. 2023, [Online]. Available: http://arxiv.org/abs/2306.01249

[21] R. Tao et al., "Magnetocardiography-Based Ischemic Heart Disease Detection and Localization Using Machine Learning Methods," IEEE Trans Biomed Eng, vol. 66, no. 6, pp. 1658–1667, Jun. 2019, doi: 10.1109/TBME.2018.2877649.

[22] P. S. Huang et al., "An Artificial Intelligence-Enabled ECG Algorithm for the Prediction and Localization of Angiography-Proven Coronary Artery Disease," Biomedicines, vol. 10, no. 2, Feb. 2022, doi: 10.3390/biomedicines10020394.

[23] C. K. Roopa and B. S. Harish, "Automated ecg analysis for localizing thrombus in culprit artery using rule based information fuzzy network," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 6, no. 1, pp. 16–25, 2020, doi: 10.9781/ijimai.2019.02.001.

Unlocking the Potential of Cloud Computing in Healthcare: A Comprehensive SWOT Analysis of Stakeholder Readiness and Implementation Challenges

Alaa Abas Mohamed
Information System Department, Al-Neelain University, Khartoum, Sudan

***Abstract*—The adoption of cloud computing in healthcare holds the potential to revolutionize healthcare delivery, particularly in developing regions. Despite its promise of scalability, cost-effectiveness, and improved data management, challenges such as digital literacy gaps, infrastructure deficiencies, and security concerns hinder its implementation. This study evaluates the readiness for adopting cloud computing in Sudan's healthcare sector through a comprehensive SWOT analysis. Findings reveal that 93.75% of patients are willing to learn electronic health systems (EHS), yet 53.12% prefer paper records, indicating trust issues. Among medical staff, 34.38% report poor digital literacy, and 46.88% cite limited access to technology as a barrier. Ministry of Health employees highlight poor infrastructure (33.33%) and limited resources (30%) as significant obstacles. By identifying strengths, weaknesses, opportunities, and threats, this research provides actionable recommendations for overcoming these barriers. The findings contribute to the ongoing discourse on digital health transformation, offering insights into fostering trust in cloud technologies for enhanced healthcare outcomes.**

***Keywords*—Cloud computing; SWOT; strength; weakness; opportunities; threat**

I. Introduction

The healthcare sector is undergoing significant digital transformation, with cloud computing emerging as a cornerstone of these advancements. Cloud-based solutions offer scalable, flexible, and cost-effective systems that improve patient care, enhance data management, and facilitate collaboration among healthcare providers (Rimal, Choi, & Lumb, 2017) [1]. Recent studies highlight that cloud computing is pivotal in facilitating the exchange of critical health information, such as in dialysis management, enabling better patient outcomes and streamlined care delivery (Ehteshami et al., 2024) [9]. This potential is particularly critical in developing regions like Sudan, where healthcare systems face persistent challenges, including resource constraints, inadequate infrastructure, and limited access to advanced technologies (Alotaibi, 2020; Liu, Wu, & Sun, 2019) [4], [3]. Furthermore, the COVID-19 pandemic underscored the need for resilient and technology-driven healthcare systems to effectively respond to emerging challenges (Jalali & Kaiser, 2018) [2]. Despite its promise, adopting cloud computing in healthcare is hindered by significant barriers. Key issues include limited digital literacy among healthcare professionals, insufficient infrastructure, and concerns regarding the security and privacy of sensitive patient data (Griebel et al., 2015) [7]. These challenges hinder the implementation of electronic health systems (EHS), which are critical to leveraging cloud computing's full potential. Additionally, the integration of the Internet of Medical Things (IoMT) with cloud systems offers transformative potential by enabling real-time monitoring and enhancing connectivity across healthcare systems, but this requires overcoming technical and infrastructural challenges (Sarıkaya & Dicle, 2024) [10]. Addressing these barriers is vital to creating a robust, secure, and efficient healthcare ecosystem that meets the needs of patients and providers alike. This study evaluates the readiness and feasibility of adopting cloud computing in Sudan's healthcare sector through a comprehensive SWOT analysis. By analyzing the perspectives of patients, medical staff, and Ministry of Health employees, the research identifies the strengths, weaknesses, opportunities, and threats associated with cloud adoption. The findings contribute to the literature by offering actionable recommendations tailored to Sudan's unique challenges and opportunities. The remainder of this paper is structured as follows. Section II outlines the methodology, detailing the data collection process and SWOT framework used. Section III reviews previous studies on cloud computing in healthcare to contextualize the findings. Comparative analysis is given in Section V. Section IV presents the results of the SWOT analysis, while Section VI discusses the implications and offers recommendations. Finally, Section VII concludes with key insights and suggestions for future research. By focusing on readiness and barriers to adopting cloud computing, this study addresses a critical gap in the literature and provides actionable insights for healthcare policymakers aiming to enhance digital transformation efforts.

II. Methodology

Cloud computing in healthcare presents multifaceted challenges that require a structured and comprehensive approach to understanding and addressing them effectively. This study employs a SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis framework, which is a strategic planning tool widely used to assess both internal and external factors influencing a system or organization. The SWOT framework is particularly suitable for evaluating the

readiness and feasibility of cloud computing adoption, as it allows for a holistic examination of enablers and barriers from multiple stakeholder perspectives. Recognizing the critical importance of data security in cloud adoption, Alshar'e et al. (2024) [11] highlight blockchain technology as an effective solution for enhancing data security and mitigating privacy risks, which underscores the necessity of addressing security concerns in this study. The following steps outline the systematic approach used to conduct the research:

1) Defining the scope and objectives: The primary objective of this study is to evaluate the readiness for adopting cloud computing in Sudan's healthcare sector. This involves identifying key strengths, weaknesses, opportunities, and threats from three stakeholder groups: patients, medical staff, and Ministry of Health employees.

2) Survey design and data collection: Surveys were designed to capture both qualitative and quantitative data. Each stakeholder group was asked specific questions regarding:

a) Technological readiness such as comfort with using electronic systems).

b) Perceived challenges such as digital literacy and infrastructure.

c) Willingness to adopt electronic health systems.

A structured questionnaire was used to ensure consistency across responses. The data was collected over a predefined period to ensure accuracy and reliability.

3) SWOT Analysis framework: Responses were categorized into the four SWOT dimensions:

a) Strengths: Positive attributes that support cloud computing adoption.

b) Weaknesses: Internal limitations that could hinder adoption.

c) Opportunities: External conditions that could facilitate adoption.

d) Threats: External risks that could undermine success.

The categorization was performed using a systematic coding process based on predefined criteria.

4) Comparative analysis across stakeholder groups: ata from patients, medical staff, and Ministry of Health employees were compared to identify common themes and divergent perspectives. This comparative analysis helped highlight stakeholder-specific needs and challenges.

5) Synthesis of findings and recommendations: Based on the SWOT analysis, actionable recommendations were developed to address key barriers and leverage existing opportunities.

III. Previous Studies

Cloud computing has been extensively studied for its potential to improve healthcare systems globally. A study by Rimal et al. (2017) [1] found that cloud computing offers improved efficiency, flexibility, and cost-effectiveness in managing patient records and delivering telemedicine services. Similarly, Jalali and Kaiser (2018) [2] identified that cloud-based health systems enable faster data retrieval and collaboration between healthcare providers.

However, barriers such as security concerns and limited infrastructure have been highlighted by Liu et al. (2019) [3], who argued that trust in cloud systems must be built through rigorous data protection measures and staff training. Alotaibi (2020) [4] explored the challenges of adopting cloud solutions in healthcare, finding that staff resistance to change and lack of technical expertise were common obstacles in low-resource settings.

These studies provide valuable insights into the general benefits and challenges of cloud computing in healthcare and form the basis for analysing the specific conditions in Sudan.

IV. SWOT Analysis

A. Patients

1) Strengths:

a) Comfort with technology: 50% of patients are "very comfortable" using technology, while 93.75% are willing to learn how to use electronic health systems if it improves their healthcare experience.

b) Implication: This demonstrates a high level of technological readiness among patients, which can facilitate the adoption of cloud computing systems.

2) Weaknesses:

a) Preference for paper records: 53.12% of patients prefer paper records, reflecting distrust in digital systems.

b) Implication: Trust-building initiatives will be needed to overcome this preference, especially regarding concerns about data security and reliability.

3) Opportunities:

a) Need for additional support: 87.5% of patients believe they need more support to use electronic health systems effectively.

b) Implication: Providing educational resources and support can bridge the gap between readiness and effective adoption.

4) Threats:

a) Resistance to digital records: The preference for paper records (53.12%) poses a threat to the full implementation of EHS, as it suggests potential resistance to change.

5) Gender divide: There may be differences in comfort with technology between genders, indicating a need for tailored training.

B. Medical Staff

1) Strengths:

a) Adaptability to technology: 37.5% of medical staff adapt "Somewhat well," and 25% adapt "Well" to new technologies.

b) Implication: Medical staff are generally open to adopting electronic systems, which is crucial for the success of cloud computing initiatives.

2) Weaknesses:

a) Low digital literacy: 34.38% of medical staff rated their digital literacy as "Poor," while 28.13% rated it as "Fair."

b) Implication: Training programs are necessary to enhance the digital competency of staff to effectively use EHS.

3) Opportunities:

a) Demand for additional training: 93.75% believe that additional training is necessary for EHS to succeed.

b) Implication: This presents a clear opportunity for healthcare institutions to implement targeted training programs.

4) Threats:

a) Limited access to technology: 46.88% of staff reported that limited access to technology is a significant barrier to adopting EHS.

b) Implication: Without addressing this issue, the successful adoption of cloud computing could be limited to certain facilities with better infrastructure.

C. Ministry of Health Employees

1) Strengths:

a) Strong support from MOH: 36.67% of Ministry employees rated the Ministry's support for EHS as "High."

b) Implication: Institutional support is a key enabler for the transition to cloud computing in healthcare.

c) Stakeholder engagement: 96.67% believe that stakeholder engagement is crucial for the success of EHS.

d) Implication: The recognition of the importance of collaboration across different sectors can accelerate the digital transformation process.

2) Weaknesses:

a) Technological infrastructure: 33.33% rated the technological infrastructure as "Poor," indicating a significant gap that must be addressed to ensure the success of EHS.

b) Implication: Investment in modern infrastructure is essential to support cloud-based systems.

3) Opportunities:

a) Providing more training: 36.67% believe that additional training is needed to support the transition to EHS.

b) Implication: Investment in capacity-building initiatives is crucial for the success of digital transformation in healthcare.

4) Threats:

a) Limited financial resources: 30% identified limited financial resources as a key challenge to implementing EHS.

b) Implication: Securing adequate funding is essential to support the infrastructure and training needs required for cloud adoption.

V. Comparative Analysis

A comparison of the perspectives from patients, medical staff, and Ministry of Health employees reveals common themes of willingness to adopt EHS but highlights the need for support and infrastructure improvements. Table I provides a detailed breakdown of responses across stakeholder groups, highlighting key metrics such as comfort with technology, willingness to learn, main weaknesses, opportunities, and threats. Fig. 1 illustrates these metrics visually, showing that while 93.75% of patients and medical staff are willing to learn electronic health systems (EHS), comfort with technology varies significantly, with only 50% of patients feeling very comfortable compared to 37.5% of medical staff and 36.67% of MOH employees. Weaknesses, such as 53.12% of patients preferring paper records and 34.38% of medical staff reporting poor digital literacy, are notable barriers to adoption. Opportunities are evident, as shown in Fig. 2, with all three groups identifying a need for resources and support. 87.5% of patients, 93.75% of medical staff, and 96.67% of MOH employees emphasize the importance of additional resources to enable successful implementation. However, threats such as resistance to digital records (affecting 53.12% of patients) and limited financial resources (affecting 30% of MOH employees) underscore the challenges ahead. Fig. 3 further compares survey responses across the groups, presenting a comprehensive view of the disparities and commonalities in stakeholder perspectives. This highlights the critical areas that need to be addressed, including investment in digital literacy programs, improved access to technology, and sustained stakeholder engagement.

TABLE I. A Comparison of the Perspectives from Patients, Medical Staff, and Ministry of Health Employees

Category	Patients	Medical Staff	MOH Employees
Comfort with Technology	50% are "Very comfortable."	37.5% adapt "Somewhat well," 25% "Well."	36.67% rate support as "High."
Willingness to Learn	93.75% willing to learn EHS.	93.75% agree continuous engagement is key.	96.67% agree on stakeholder engagement.
Main Weakness	53.12% prefer paper records.	34.38% rated digital literacy as "Poor."	33.33% rated infrastructure as "Poor."
Opportunities	87.5% need support to use EHS.	93.75% demand additional support/resources.	96.67% believe more resources are needed.
Main Threat	Resistance to digital records (53.12%).	Limited access to technology (46.88%).	Limited financial resources (30%).

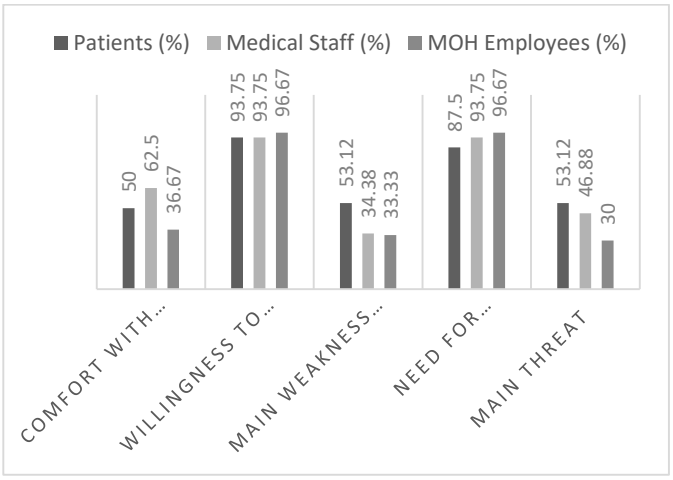


Fig. 1. A comparison of the perspectives from patients, medical staff, and Ministry of Health employees.

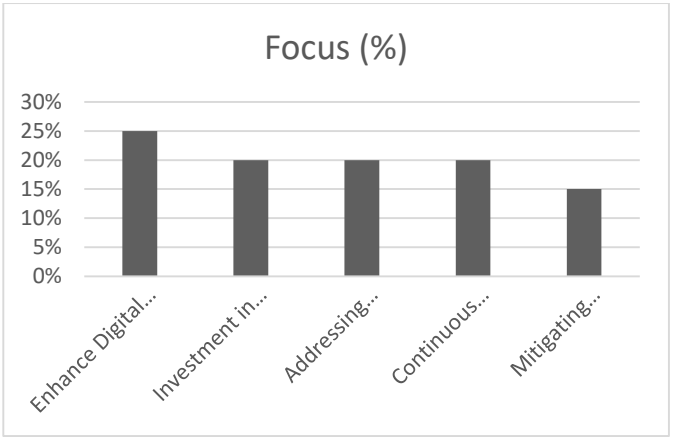


Fig. 2. The focus areas in the recommendations.

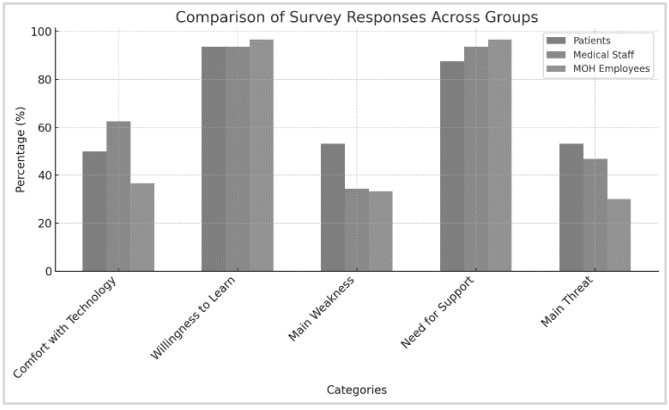


Fig. 3. Comparison of survey response across groups.

VI. Recommendations

1) Digital literacy training: Given the low digital literacy among medical staff (34.38% rated as "Poor"), it is critical to implement targeted training programs. These programs should be designed to improve competency in using electronic health systems (EHS) and address the specific technological needs of both healthcare professionals and patients. Tailored workshops, online modules, and continuous support could enhance digital readiness.

2) Investment in technological infrastructure: 33.33% of Ministry of Health employees reported poor technological infrastructure in healthcare facilities. Investing in modern, reliable infrastructure is essential to support cloud-based systems. This includes upgrading hardware, ensuring stable internet connections, and implementing secure cloud platforms to handle sensitive patient data. Adequate infrastructure is a foundational element for cloud computing in healthcare.

3) Addressing resistance to digital records: 53.12% of patients prefer paper records, indicating a significant trust issue with digital systems. To overcome this resistance, awareness campaigns should be launched to educate patients and staff about the benefits of electronic records, including security, accessibility, and efficiency. Building trust in digital systems is critical to ensuring successful adoption. Highlighting the security features and showing real-world examples of the improved outcomes of using cloud systems could help ease concerns.

4) Continuous support and stakeholder engagement: 96.67% of Ministry of Health employees and 93.75% of medical staff agree that continuous engagement and support are essential for the success of EHS. To ensure a smooth transition to cloud computing, healthcare institutions must provide ongoing technical assistance, regular feedback sessions, and open communication channels between healthcare providers, patients, and stakeholders. Stakeholder engagement ensures that concerns are addressed in real time, and that users feel supported throughout the process. This could include dedicated technical support teams, helplines, and regular system updates based on user feedback to improve system functionality and user satisfaction.

VII. Conclusion

The adoption of cloud computing in healthcare presents a transformative opportunity to modernize healthcare systems, improve patient outcomes, and enhance operational efficiency. This study evaluated the readiness and feasibility of adopting cloud computing in Sudan's healthcare sector through a comprehensive SWOT analysis, capturing the perspectives of patients, medical staff, and Ministry of Health employees. The findings highlight significant strengths, such as the willingness of 93.75% of patients to learn electronic health systems and the adaptability of medical staff. However, challenges, including low digital literacy among healthcare professionals (34.38% rated as poor), limited technological infrastructure, and concerns about data security, remain significant barriers. Actionable recommendations have been proposed to address these issues, such as targeted training programs, investment in infrastructure, and trust-building initiatives. Despite the contributions of this study, several limitations should be acknowledged. First, the reliance on survey-based data collection may introduce self-reporting biases, as respondents' perceptions may not fully represent their actual capabilities or behaviors. Second, the study's focus on Sudan's healthcare system limits the generalizability of the findings to other

regions or contexts with differing levels of technological readiness and resources. Third, the SWOT analysis, while effective in providing a comprehensive overview, does not explore the causal relationships between the identified factors or quantify their relative impact. Finally, the study primarily assesses the readiness for cloud adoption and does not investigate the long-term outcomes or sustainability of such initiatives. Acknowledging these limitations provides a balanced perspective on the study's findings. Nevertheless, the insights presented offer valuable guidance for policymakers and healthcare institutions seeking to adopt cloud computing. Future research should aim to address these limitations by exploring longitudinal studies, employing mixed-method approaches, and extending the analysis to diverse healthcare contexts to further validate and enhance the understanding of cloud adoption.

REFERENCES

[1] Rimal, B. P., Choi, E., & Lumb, I. (2017). A taxonomy and survey of cloud computing systems. Future Generation Computer Systems, 29(1), 137-150.

[2] Jalali, S., & Kaiser, J. (2018). Cloud computing for health data management: A case study and review. Health Informatics Journal, 24(3), 309-320.

[3] Liu, Y., Wu, H., & Sun, Y. (2019). Security and privacy challenges in cloud computing environments for healthcare systems. Journal of Cloud Computing, 8(1), 1-13.

[4] Alotaibi, M. B. (2020). Barriers to implementing cloud computing in healthcare organizations. Journal of Healthcare Management, 65(2), 142-155.

[5] Kuo, A. M. H. (2011). Opportunities and challenges of cloud computing in healthcare. Journal of Medical Internet Research, 13(3), e67.

[6] Rodriguez, J., & Silva, D. (2020). Cloud computing in healthcare: Perspectives, benefits, and security concerns. Health Information Science and Systems, 8(2), 22-30.

[7] Griebel, L., Prokosch, H. U., Köpcke, F., Toddenroth, D., Christoph, J., Leb, I., & Sedlmayr, M. (2015). A scoping review of cloud computing in healthcare. BMC Medical Informatics and Decision Making, 15(1), 17.

[8] Zhang, R., & Liu, L. (2010). Security models and requirements for healthcare application clouds. Health Informatics Journal, 16(1), 56-62.

[9] A. Ehteshami, M. Esmailzadeh, et al., "A Scoping Review of Cloud Computing Solutions in Enhanced Dialysis Information Exchange," Journal of Evidence-Based Healthcare Policy and Management Education, 2024.

[10] A. Sarıkaya And F. Dicle, "Internet of Medical Things and Smart Health Systems and Its Reflections on Application: A Systematic Review," 4th Internatıonal Home Care Congress , no.7817, İstanbul, Turkey, pp.66-68, 2023.

[11] M. Alshar'e, K. Abuhmaidan, F. Y. H. Ahmed, and A. Abualkishik, "Assessing Blockchain's Role in Healthcare Security: A Comprehensive Review," Informatica, vol. 48, no. 1, pp. 1-14, 2024.

An Novel Approach Based on Information Relevance Perspective and ANN for Predicting the Helpfulness of Online Reviews

Nur Syadhila Bt Che Lah[1], Khursiah Zainal-Mokhtar[2]
Department of Computer and Information Science, Universiti Teknologi PETRONAS, Perak, Malaysia[1]
Research Innovation Centre (RIC), Universiti Teknologi PETRONAS, Perak, Malaysia[2]

***Abstract*—This study presents a novel approach to predicting the helpfulness of online reviews using Artificial Neural Networks (ANNs) focused on information relevance. As online reviews significantly influence consumer decision-making, it is critical to understand and identify reviews that provide the most value. This research identifies four key textual features namely content novelty, content specificity, content readability, and content reliability, that contribute to perceived helpfulness and incorporates them as primary inputs for the ANN model. Datasets of Amazon reviews are analyzed, and various preprocessing steps are employed to ensure data quality. Reviews are classified as helpful or unhelpful based on helpful vote thresholds, with experiments conducted across multiple helpful vote thresholds to determine the optimal threshold value. Performance was evaluated using accuracy, precision, recall, and F1 scores, with the best-performing classifier achieving 74.34% accuracy at a helpful vote threshold of 12 votes. These results highlight the potential of information relevance-based criteria to enhance the accuracy of online review helpfulness prediction models.**

***Keywords*—Review helpfulness; online reviews; information relevance; review novelty; review readability; review specificity; Artificial Neural Networks**

I. INTRODUCTION

The emergence of internet has revolutionized life, bringing significant changes essential to people's daily life. In the past, tasks such as making purchases had to be done face-to-face. In this traditional market settings, consumers acquire new information via advertisements, brochures and word-of-mouth (WOM) about various products and services before making purchases. In the context of online reviews, consumer often reads multiple reviews before making purchasing decision. Online reviews have impacted not only consumers but also business, platforms and reviewers. Reviews can help consumers reduce uncertainty [1] regarding the quality of a product or service by offering firsthand insights and experiences from other users. For businesses, positive reviews can lead to increased sales as they build trust and credibility [2]. Platforms benefit from the continuous creation of online reviews and building consumer trust, which attracts more consumers. In turn, reviewers gain recognition from their peers and receive gifts and special promotion on platforms. Such incentives encourage them to continue providing useful reviews [3].

As the number of reviews rapidly increasing, platforms face the challenge of managing this large amount of information to ensure that consumers can easily access to the most relevant and helpful review. To address the issue, platforms have introduced feedback mechanisms that allow consumers to vote for posted reviews that they considered helpful. Readers are more likely to trust the statements if it has been marked as helpful by other consumers [4]. However, helpful vote is a manual process and could result in helpful reviews being ignored. Moreover, it is unclear to potential customers whether previous customers marked a review as "helpful" before using a product or service or after having used it. The criteria for what constitute a "helpful" review is not strictly defined and thus can be difficult to assess.

Many studies have utilized various theories such as Elaboration Likelihood Model (ELM) and Information Adoption Model (IAM) to capture numerous factors that affect review helpfulness [5-9]. In the context of IAM, it suggests that information helpfulness is influenced by argument quality and source credibility. Recent research has revealed inconsistencies on how to properly judge the quality of arguments in reviews. Some experts believe that certain parts of a review are difficult to measure objectively and can be changed depending on the situation [10]. Also, many traditional ways of measuring argument quality might not work well for evaluating personal, subjective aspects of the review content [11]. As a result, relevance has been suggested as an important factor, with its importance depending on the specific decision a reader is trying to make [12]. Information relevance is recognized as a key determinant of information diagnostic for gaining a better understanding of consumers' opinions and their relevance to electronic WOM communications [13]. Online reviews are perceived as relevant when businesses provide information that aligns with consumers' expectations [14].

Previous studies have typically examined factors that contribute to review helpfulness including, text sentiment [15, 16], review depth [17, 18], readability [19-21], novelty [22], credibility [2, 23], specificity [24], reliability [19] and reading enjoyment [25]. However, the research on this topic is quite scattered and inconsistent. Different studies focus on various parts of quality text or use different methods, which makes it hard to get a clear and complete picture. As a result, there is no definitive list of key factors related to information relevance, leading to an incomplete and fragmented understanding of the subject.

Drawing from this observation, this study introduces four key textual features based on information relevance perspective of content novelty, content specificity, content readability and content reliability that potentially contribute to the helpfulness of reviews. These factors represent essential dimensions of information relevance, to enhance quality of online review. Novelty introduces previously unmentioned perspectives, while specificity provides detailed information tailored to the reader's needs. Readability ensures that the information is easily understood, and reliability supports the trustworthiness and accuracy of the content. Therefore, the proposed features can help to better understand how the subjective qualities of reviews impact their helpfulness and influence customer purchasing decisions.

Next, this study adopts a 'threshold' approach to identify helpful reviews by categorizing them based on the number of helpful votes received. The concept of helpful votes threshold provides an innovative way to filter reviews, ensuring that reviews deemed helpful by a larger number of users are given prominence. By setting these thresholds, platforms can prioritize reviews that have resonated with consumers, thereby helping potential buyers make informed decisions more quickly. For instance, reviews exceeding a certain number of helpful votes can be classified as "helpful," allowing the system to highlight feedback that users have found insightful and trustworthy.

Although results did not show a dramatic improvement in classification accuracy, the threshold approach offers practical benefits. It enables a structured and automated system for identifying helpful content, reducing the reliance on manual helpfulness voting and minimizing the risk of helpful reviews being overlooked. The threshold system also aids in understanding how different levels of helpful votes correlate with review helpfulness, providing insights that could guide future improvements in review filtering algorithms. Furthermore, this threshold-based method could serve as a foundation for iterative refinements, where feedback from user interactions helps to adjust the threshold levels dynamically, enhancing the platform's ability to deliver relevant and valuable reviews to consumers.

The rest of this paper is organized as follows. The detail literature reviews of various factors or indicators that contribute to review helpfulness and helpful votes threshold impact on model performance is presented in Section II. Section III introduces methodology that integrates four review text characteristics. Section IV presented results and discussion. Lastly, Section V provides the conclusion of this work.

II. Literature Review

Previous works on review helpfulness have demonstrated association between various review characteristics and review helpfulness and serve as primary source of information over other aspects such as reviewer's identity, metadata and product [26-28]. Research indicates that the content quality, clarity, and emotional tone of the text significantly affect its perceived helpfulness as they enhance the review's credibility and relatability to readers [29]. In addition, text-related features such as length and structure can impact engagement levels, making them more critical than metadata or reviewer-related characteristics, which may not consistently correlate with helpfulness [30, 31]. Therefore, emphasizing text-related characteristics allows for a more direct assessment of review helpfulness.

Previous studies on novelty detection have explored various techniques to identify new and unique information within data. These studies aim to identify new information within a document by employing various approaches tailored to different objectives. Various widely used measurement metrics for novelty detection are utilized across document such as Simple New Word Count, Set Difference, TF-IDF scoring, and Cosine Distance [32, 33]. These methods apply a bag-of-words approach, utilizing word counts within a document. Some novelty measures are derived from probabilistic document models [33-35]. The Simple New Word Count measure, which examines the occurrence of novel words in sentences, has been shown to be as effective as probabilistic document models and other bag-of-words-based methods [33]. Recently, Deep Learning methods such as BERT (Bidirectional Encoder Representations from Transformers) have gained popularity for tasks involving semantic textual similarity [36]. One such method, Sentence-BERT [37], utilizes a pre-trained BERT model to generate context-aware text embeddings, which can be employed to assess the similarity between documents. Sentence-BERT was adapted to calculate the novelty measure in the main analysis, and the analyses were replicated using new word pairs and a revised version of the Simple New Word Count measure [22]. The work also demonstrated review novelty impacts on consumers and businesses.

The discussion on specificity primarily focuses on the sentence level. A common definition of specificity, as used by [38-40], refers to the amount of detail within a sentence. Research on specificity utilizes a broad array of features to indicate sentence specificity. While some studies employ a large collection of features, others may rely on just one. Given the significant variation in feature sets, it is logical to analyze the importance of each feature. Based on current knowledge, the simplest prediction method relies on just one feature which is normalized inverse word frequency (IDF). The sum, average, minimum, and maximum IDF values for all words within a sentence were evaluated, with the maximum IDF value found to be the best indicator of specificity [41]. However, relying on a single feature has its limitations, as the predictor may lose effectiveness in tasks involving multiple topics, due to the significant variation in word distribution across different topics. Speciteller [42] is a popular tool for predicting sentence specificity. It generates 17 features, including sentence characteristics and word representations. Typically, researchers combine Speciteller features with their own custom features to build specificity estimators. For instance, the study by [43] combined Speciteller features with online dialogue features, resulting in a model that outperforms Speciteller in predicting specificity in classroom discussions. In contrast, the work by [24] used a model developed by [38] to assess the specificity of sentences in product reviews. They introduced three new metrics: the percentage of specific sentences in a review, the overall specificity of the review, and the balance between specific and general sentences. This study represents the first attempt to approach the helpfulness prediction problem from sentence specificity perspective.

Review readability, which refers to the ease with which a text can be understood, significantly affects the perceived helpfulness of reviews. Consumers are more likely to find a review helpful if they can easily interpret it. Therefore, higher readability generally facilitates better understanding. Like novelty and specificity, numerous features are used to predict readability. The features commonly employed in readability prediction studies are generally consistent. These features were categorized into semantic and syntactic groups, with an analysis of both the words and sentence structures [44]. Syntactic features include sentence length, average number of characters per word, average number of syllables per word and the percentage of various part-of-speech tags. Semantic features involve the frequency of various 1-, 2-, and 3-word sentences in a review. Many studies have utilized various readability indices, including the Simple Measure of Gobbledygook (SMOG), Automated Readability Index (ARI), Gunning Fog Index (GFI), Flesch–Kincaid Grade Level (FKGL), Coleman–Liau Index (CLI), and Flesch–Kincaid Reading Ease (FKRE) to predict review helpfulness [21, 28, 45] found that a hybrid set of features based on linguistic categories, review metadata, readability, and subjectivity offered the best review predictive performance. Review content features, such as readability, were identified as the most effective predictors of helpfulness [28]. Readability, along with linguistic and psychological features, was utilized to predict the helpfulness of movie reviews [45].

The reliability of online reviews has attracted significant attention in recent years. Various studies have explored the factors that influence the reliability of online reviews, and the methods used to assess and enhance this reliability. The Linguistic Inquiry and Word Count (LIWC) program was used to analyze the proportion of positive and negative words in reviews [46, 47]. It has been discovered that the sentiment of a review, whether positive or negative, along with advice for decision-making and claims of expertise, significantly influences the perceived helpfulness of the review [46]. Additionally, previous studies have highlighted that both sentiment orientation (positive or negative) and the writing style of the review (subjective or objective) are key factors in determining its believability [46, 48]. Consequently, research conducted by [19] leveraged these elements as reliability indicators to assess review helpfulness. The detection of spam reviews is also critical in evaluating online review reliability. Early research focused on identifying spam reviews by detecting copied content [49-51]. Various reliability features, such as Kullback-Leibler divergence, syntactic text characteristics, and review semantics, have been employed to distinguish fake reviews from genuine ones. Additionally, several algorithms have been developed to filter out unreliable reviews [52-53].

The concept of helpful votes threshold in the context of online reviews, is a crucial parameter in binary classification systems. It directly impacts the performance of algorithms that distinguish between helpful and unhelpful reviews by setting a boundary that defines what qualifies as "helpful." An inappropriate threshold can lead to the misclassification of reviews, where helpful reviews are categorized as unhelpful or vice versa, ultimately weakening the model's performance and skewing results.

The choice of classification threshold is essential to improve classification precision. Studies, such as those by Ghose and Ipeirotis [64], have found that a threshold where the ratio of helpful votes to total votes equals 0.6 can significantly enhance classification accuracy for review helpfulness on platforms like Amazon. This threshold value, also adopted by researchers such as Krishnamoorthy [21] and Malik and Hussain [69], minimizes the chances of misclassifying helpful reviews as unhelpful and vice versa, improving the reliability of the helpfulness measure.

III. METHODOLOGY

The section introduces methodology for predicting the helpfulness of online reviews, including the collection of product reviews, review characteristics and the helpfulness of a review as shown in Fig. 1.

A. Data Collection

Many e-commerce platforms provide product or service reviews and relevant data. This study is focused on products available in Amazon.com. The data collected included review rating, review title and text, identification number of a product, user identification number, the time a review is posted, number of helpful votes received by a review and verified purchase of a product. Reviews from the year 2022 until the year 2023 were downloaded for this study [54].

B. Data Preparation

This study utilized a dataset consisting of 9,369 helpful reviews and 9,369 unhelpful reviews sourced from the Beauty, Health, and Personal Care categories on Amazon.com. The dataset is consistent across all helpful votes thresholds employed in this experiment. As many e-commerce platforms provide valuable insights through product and service reviews, this research specifically focuses on products available on Amazon.com.

The collected data included various attributes such as review ratings, review titles, review text, product identification numbers, user identification numbers, timestamps of when the reviews were posted, the number of helpful votes each review received, and whether the purchase was verified. The reviews analyzed were collected from the years 2022 to 2023 [54]. To refine the data before feature extraction, non-English text is filtered out from the dataset. Since text containing URLs and HTML tags might point to a promotional site, or competitors, rows of data with these elements are also removed. In addition, text with emojis and emoticons are also eliminated. Then, the rows of data where the text column is empty or contains fewer than 5 words are excluded, as these offer limited information for potential customers [55]. Besides, data with duplicated text are also omitted. Review duplication can potentially occur in three difference situations [56]:

- Duplicate reviews of the same product with a different user identification number.
- Duplicate reviews from the same user ID but on different products.
- Duplicate reviews from different user IDs on different products.

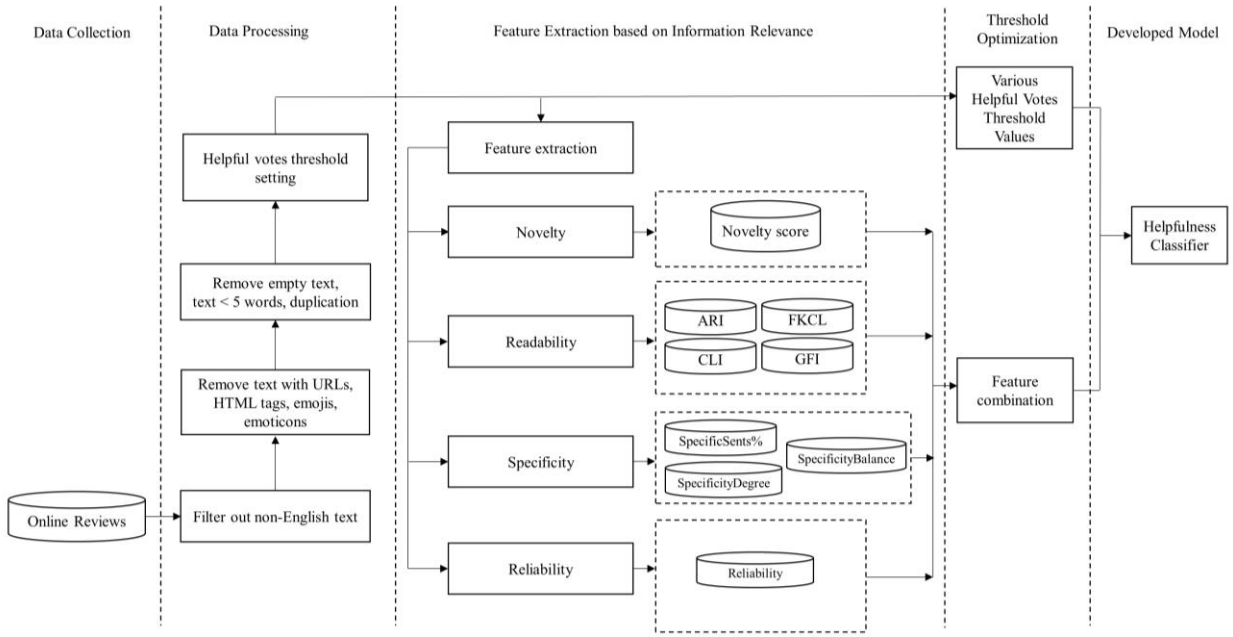


Fig. 1. Research process of this work

In terms of helpful indicator, the study by [57] suggest that highly adopted review helpfulness ratio (number of helpful votes/total number of votes) could lead to highly bias results. Hence, this study utilized the various helpful votes threshold values that is transformed into binary representation.

C. Feature Extraction

1) Content novelty: Some previous studies have described novel information as information that is different from what readers already know or expect [58]. However, most consumers cannot rely solely on their prior knowledge to guide them when making decision in an online environment, especially when purchasing experienced goods [59]. New information and perspective in online reviews may influence consumers to purchase products as it increases consumer's awareness about a product or a service. Consumers highly value knowledge gained from firsthand experience, often spending substantial time and effort searching for and reading reviews to find new information and insights. Therefore, novel information in reviews can be perceived as helpful by consumers.

Empirical evidence indicates that most consumers do not look beyond the first page of search results [60, 61]. Additionally, a consumer survey reveals that most consumers read no more than ten reviews before making a purchase [62]. Hence, content novelty can be defined as the amount of novel information in current review compared with the 10 most recent reviews on a single page. Novelty score for each review is used to determine the amount of novel information in reviews. The method of measuring the amount of novel information in each review is based on method proposed by [22]. First, let r_i represent the focal review and let $C_i = \{r_{i1}, r_{i2}, r_{i3}, \ldots, r_{ij}\}$ represent the comparison set for r_i, where j is the number of reviews in the comparison set (up to 10).

The novelty score for r_i given the comparison set C_i can be expressed as:

$$N(\rho_\iota, X_\iota) = \min_{\rho_{\iota\kappa} \square X_\iota} \left(1 - \chi o \sigma \left(\varepsilon \mu \beta(\rho_\iota), \varepsilon \mu \beta(\rho_{\iota\kappa})\right)\right) \quad (1)$$

Where

- emb(r) represent the embedding of review r using the context-aware representation method.
- $\cos(\text{emb}(r_i), \text{emb}(r_{ik}))$ is the cosine similarity between the embeddings of the focal review r_i and a review r_{ik} in the comparison set C_i.

The final novelty score for the review r_i, considering all possible comparison sets $C_1, C_2, C_3, \ldots, C_n$ is the average of the minimum novelty score across all comparison sets:

$$\Phi \iota \nu \alpha \lambda \ No \varpi \varepsilon \lambda \tau \psi \ \Sigma \chi o \rho \varepsilon \ (\rho_\iota) = \frac{1}{\nu} \sum_{\mu=1}^{\nu} N(\rho_\iota, X_\mu) \quad (2)$$

Where n is the number of different comparison sets considered for r_i and m is the index of the comparison set.

2) Content readability: The readability of a review is another crucial factor that can influence its perceived helpfulness. A review that is highly readable is more likely to be read and voted on by a larger number of users. Readability refers to the ease with which a reader can understand and process a piece of textual information [63]. The readability feature also determines the complexity of any review for the user [45]. Previous research by [64] shows that readability and subjectivity features outperform the lexical features employed by [65]. The work by study [21] shows that the combination of features derived from linguistic categories, readability, review metadata, and subjectivity provide the most accurate predictive performance.

To assess review readability, four grade level readability metrics can be utilized namely (1) Automated Readability Index (ARI), (2) Flesch–Kincaid Grade Level (FKGL), (3) Gunning Fog Index (GFI), and (4) Coleman–Liau Index (CLI) [21].

ARI was selected in this study because it is one of the primary measures used to assess text readability and is less prone to error than other readability measures [66]. The ARI can be calculated using its standard formula:

$$\text{API } \Sigma\chi\text{o}\rho\varepsilon\text{=4.71 } \left(\frac{\nu\upsilon\mu\beta\varepsilon\rho \text{ o}\phi \chi\eta\alpha\rho\alpha\chi\tau\varepsilon\rho\sigma}{\nu\upsilon\mu\beta\varepsilon\rho \text{ o}\phi \omega\text{o}\rho\delta\sigma}\right) +0.5 \left(\frac{\nu\upsilon\mu\beta\varepsilon\rho \text{ o}\phi \omega\text{o}\rho\delta\sigma}{\nu\upsilon\mu\beta\varepsilon\rho \text{ o}\phi \sigma\varepsilon\nu\tau\varepsilon\nu\chi\varepsilon\sigma}\right) - 21.43 \quad (3)$$

By analyzing sentence length, word difficulty, and text cohesion, the FKGL formula measures how challenging readers may find a text. The formula is as follows:

$$\text{K}\Gamma\Lambda\ \Sigma\chi\text{o}\rho\varepsilon = 0.39 \left(\frac{\omega\text{o}\rho\delta\sigma}{\sigma\varepsilon\nu\tau\varepsilon\nu\chi\varepsilon\sigma}\right) - 11.8 \left(\frac{\sigma\psi\lambda\lambda\alpha\beta\lambda\varepsilon\sigma}{\omega\text{o}\rho\delta\sigma}\right) - 15.59 \quad (4)$$

The Gunning Fog Index formula is based on the idea that shorter sentences written in clear, straightforward language receive a better score than longer, more complex sentences. Online reviews that score well on the Gunning Fog Index are likely to be more accessible and comprehensible to a broader audience, enhancing user engagement and improving the overall quality of the review content. The formula is given by

$$\Gamma\Phi\text{I=0.4} \left[\left(\frac{\omega\text{o}\rho\delta\sigma}{\sigma\varepsilon\nu\tau\varepsilon\nu\chi\varepsilon\sigma}\right) + 100 \left(\frac{\chi\text{o}\mu\pi\lambda\varepsilon\xi \ \omega\text{o}\rho\delta\sigma}{\omega\text{o}\rho\delta\sigma}\right)\right] \quad (5)$$

Meanwhile, CLI scores indicate the complexity of a text and are determined using the formula:

$$\text{X}\Lambda\text{I } \Sigma\chi\text{o}\rho\varepsilon\text{=0.0588} \left(\frac{\lambda\varepsilon\tau\tau\varepsilon\rho\sigma}{100 \ \omega\text{o}\rho\delta\sigma}\right) - 0.296 \left(\frac{\sigma\varepsilon\nu\tau\varepsilon\nu\chi\varepsilon\sigma}{100 \ \omega\text{o}\rho\delta\sigma}\right) - 15 \quad (6)$$

3) Content specificity: To assess content specificity in reviews, we adopted a two-step approach that involves calculating a specificity score based on the Normalized Inverse Document Frequency (NIDF) method and then deriving three specific features as outlined in prior research.

Inverse Document Frequency (IDF) is a widely recognized metric that measures the discriminative ability of a term within a document collection. It is defined as the logarithmic ratio of the total number of documents in the collection (n_d) to the number of documents containing the term (known as the term's document frequency, $df(t_i)$), as shown as follow:

$$IDF(t_i) = \log\left(\frac{n_d}{df(t_i)}\right) \quad (7)$$

In this study, we employed the Normalized Inverse Document Frequency (NIDF)The NIDF, defined in equation 8, normalizes with respect to the number of documents not containing the term ($n_d - df(t_i)$) and adds a constant 0.5 to both the numerator and the denominator to moderate extreme values:

$$NIDF(t_i) = \log\left(\frac{n_d - df(t_i) + 0.5}{df(t_i) + 0.5}\right) \quad (8)$$

Commonly used words, such as "the", "and", and "it" are likely to appear in nearly every document and are therefore not particularly discriminative. This lack of discriminative capability is reflected in their low NIDF values. Conversely, terms that occur in only a small number of documents are more useful for distinguishing between documents, resulting in higher NIDF values.

Our assumption is that documents dominated by terms with low NIDF values are less specific than those containing more discriminative terms. Consequently, we define a document specificity score, S_1, as follows:

$$S_1(d) = \frac{1}{l_d} \sum_{t_i \in d} tf(t_i) \cdot \log\left(\frac{n_d - df(t_i) + 0.5}{df(t_i) + 0.5}\right) \quad (9)$$

In this equation, $tf(t_i)$ represents the term frequency of t_i in document d, and l_d denotes the length of document d. The inclusion of l_d in the denominator reduces the impact of varying document lengths on the specificity score.

Following this calculation of the specificity score for each review, we further derived three specific features that have been proposed in previous study [24]. These features aim to assess different aspects of specificity within the context of helpfulness in online reviews.

- SpecificSents% represents the percentage of specific sentences within a review. A sentence is classified as "specific" if its specificity score is 0.5 or higher. This feature is designed to assess whether the number of specific or general sentences in a review affects its perceived helpfulness.
- SpecificityDegree represents the overall specificity of a review. Given the set *P* all sentences in a review and $\sigma(p)$ as the specificity score of a sentence $p \in P$, the specificity degree of the review is defined as:

$$\Sigma\pi\varepsilon\chi\iota\phi\iota\varpi\iota\tau\psi\Delta\varepsilon\gamma\rho\varepsilon\varepsilon = \frac{\sum_{\pi}^{\Pi} \int(\pi)}{|\Pi|} \quad (10)$$

Where |P| is the total number of sentences in the review.

- SpecificityBalance measures the balance between specific and general sentences in a review. Let S be the set of specific sentences (with a specificity degree ≥ 0.5) and G the set of general sentences (with a specificity degree < 0.5) in a review. The specificity balance of a review is calculated as:

$$\Sigma\pi\varepsilon\chi\iota\phi\iota\chi\iota\tau\psi\text{B}\alpha\lambda\alpha\nu\chi\varepsilon = \frac{||\Sigma| - |\Gamma||}{|\Sigma| + |\Gamma|} \quad (11)$$

Where |S| is the number of specific sentences, |G| is the number of general sentences and $||S|-|G||$ represents the absolute difference between these quantities. A value of 0 indicates a perfect balance between specific and general sentences, whereas a value of 1 means the review consists entirely of either specific or general sentences. According to a study by [67], general sentences are vital for high-quality journalism summaries, suggesting that this balance might influence the perception of helpfulness in product reviews as well.

4) Content reliability: Content reliability plays a critical role in determining the trustworthiness of online reviews. For this feature, a binary indicator is used to represent the reliability

of the review, specifically through the presence of a verified purchase. Reviews from verified customers are considered more genuine because they come from individuals who have actually bought and used the product. This authenticity boosts the perceived reliability of the review, as users are more likely to trust feedback from verified buyers, viewing it as a truthful and accurate reflection of the product [68]. By using a binary indicator - where 1 represents a review from a verified purchase and 0 represents a non-verified review—this feature effectively captures the connection between the genuineness of the review and its perceived reliability. Incorporating this binary indicator allows for a more structured evaluation of content reliability, enhancing the accuracy of any model that seeks to assess the trustworthiness and overall value of online reviews.

D. Neural Network Architecture for Helpfulness Prediction

To predict review helpfulness, we developed an Artificial Neural Network (ANN) using a Multi-Layer Perceptron (MLP) classifier. The MLP architecture was carefully tuned to achieve optimal predictive performance by experimenting with the number of hidden neurons and analyzing the network's response to input features.

1) Model architecture: The MLP model receives each review as an input vector of features that capture key aspects related to content quality, readability, novelty, and specificity, which are hypothesized to influence helpfulness. The input layer of the MLP is designed to process these 9 input features (illustrated in Fig. 2), each representing a distinct attribute of the review. This input layer serves as the foundation for the subsequent layers, encoding the feature values as the model begins to learn from the data.

The final MLP model configuration is as follows:

- Input Layer: Accepts a vector of 9 features per review, providing the model with a rich, multi-dimensional view of each review.
- Hidden Layer Size: A single hidden layer with 140 neurons, which was selected as the optimal configuration after experimenting with values of 20, 40, 60, 80, 100, and 140 neurons. This configuration effectively balances complexity and generalization ability.
- Activation Function: The Rectified Linear Unit (ReLU) activation function was used for the hidden layer, providing computational efficiency and mitigating the vanishing gradient problem.
- Solver: The Adam optimizer was employed to train the network. Adam combines the benefits of Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp), ensuring efficient convergence.
- Learning Rate: An initial learning rate of 0.001, which allows the model to learn gradually and converge steadily.
- Epochs: Training was set to a maximum of 2000 iterations, with early stopping to prevent overfitting.
- Random State: A random state of 42 ensures reproducibility across different runs.
- Output Layer: The output layer consists of a single neuron with a sigmoid activation function, which produces a binary output of either 0 or 1 for each review. Here, an output of 1 indicates that the model predicts the review as "helpful" while an output of 0 indicates a prediction of "not helpful".

The optimal configuration of 140 hidden neurons, based on its predictive accuracy, provided the best balance between model complexity and performance. By systematically testing different neuron counts, we identified this structure as the most suitable for the task of review helpfulness prediction.

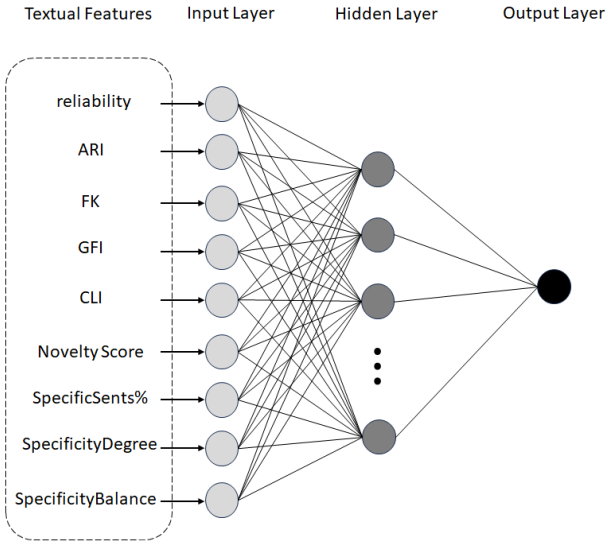


Fig. 2. Layout of MLP model.

2) Performance metrics: To evaluate the effectiveness of our model in classifying online reviews as helpful or unhelpful, we utilized several performance metrics: accuracy, precision, recall, and F1 score. Each of these metrics provides insight into different aspects of model performance, particularly in the context of user-generated content where the classification of reviews can significantly impact consumer decision-making.

Accuracy measures the overall correctness of the model's predictions. It is defined as the ratio of the number of correct predictions to the total number of predictions made. In the context of online reviews, it can be expressed mathematically as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (12)$$

Where

- TP is True Positives (number of helpful reviews correctly classified as helpful).
- TN is True Negatives (number of unhelpful reviews correctly classified as unhelpful).
- FP is False Positives (number of unhelpful reviews incorrectly classified as helpful).

- FN is False Negatives (number of helpful reviews incorrectly classified as unhelpful).

Precision quantifies the accuracy of the positive predictions made by the model, focusing specifically on how many of the predicted helpful reviews are actually helpful. This metric is particularly important in online reviews, where consumers benefit from identifying truly helpful feedback. Precision is calculated as follows:

$$Precision = \frac{TP}{TP+FP} \quad (13)$$

Recall, also known as sensitivity, measures the model's ability to identify all relevant instances of helpful reviews. It assesses how many of the actual helpful reviews were correctly identified by the model. Recall is computed using the formula:

$$Precision = \frac{TP}{TP+FN} \quad (14)$$

The F1 score provides a balance between precision and recall, offering a single metric that captures both aspects of the model's performance. It is especially useful in scenarios where there is an uneven class distribution, such as when the number of helpful reviews significantly differs from unhelpful reviews. The F1 score is calculated as:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (15)$$

In the context of online reviews, these metrics allow us to assess how well our model performs in distinguishing between helpful and unhelpful reviews. A high accuracy indicates that the model correctly identifies most reviews, while high precision ensures that consumers can trust the reviews labeled as helpful. Additionally, high recall signifies that the model successfully captures most helpful reviews, which is crucial for users seeking reliable information. The F1 score serves as a comprehensive measure, ensuring a balanced consideration of both precision and recall, which is vital in enhancing the overall consumer experience in online platforms.

E. Optimization of Helpful Votes Threshold Values

The use of Amazon.com's publicly available dataset, which contains only helpful votes, provides a reliable measure of review quality based on user interaction. Given the time constraints of this study, this dataset allowed for an efficient and effective investigation into the impact of helpfulness thresholds. Future research could expand this investigation by incorporating datasets that include total votes for comparison.

In this study, we examine various thresholds for helpful votes, ranging from more than 2 to more than 12, using intervals of 2 votes. Reviews with helpful votes exceeding the specified threshold are classified as helpful, while those with votes below the threshold are deemed unhelpful. Table I summarizes these thresholds.

The experiment stops at the helpful votes threshold of more than 14 because the data available for that threshold is not consistent with the data used for the previous thresholds (more than 2, 4, 6, 8, 10, and 12). For accurate model performance comparisons, it is crucial that the number of data points remains the same across all thresholds. This consistency ensures that any observed differences in model performance can be attributed to the thresholds themselves rather than variations in data quantity.

TABLE I. HELPFUL AND UNHELPFUL REVIEWS BY HELPFUL VOTES THRESHOLD

Helpful Votes Threshold	Helpful Reviews	Unhelpful Reviews
More than 2	> 2	≤ 2
More than 4	> 4	≤ 4
More than 6	> 6	≤ 6
More than 8	> 8	≤ 8
More than 10	> 10	≤ 10
More than 12	> 12	≤ 12

IV. RESULTS AND DISCUSSION

Fig. 2 illustrates the relationship between test accuracy and the number of hidden neurons in a neural network across different helpful vote thresholds. The x-axis represents the number of hidden neurons, ranging from 20 to 140, while the y-axis displays test accuracy as a percentage. Each line corresponds to a different threshold for the number of helpful votes, ranging from "> 2" to "> 12," offering insights into how the model performs with varying thresholds.

A general trend shows that test accuracy improves slightly with an increasing number of hidden neurons, though the gains are more pronounced for certain thresholds. The performance tends to stabilize around 100–140 hidden neurons for most thresholds, but the overall accuracy is highly dependent on the threshold used.

The helpful vote thresholds have a significant impact on performance. Lower thresholds, particularly "> 2" and "> 4", exhibit the lowest performance, with test accuracies ranging from 66% to 68%. This indicates that when the model includes reviews with very few helpful votes, it struggles to make accurate predictions. In contrast, moderate thresholds such as "> 6" and "> 8" show improved accuracies, ranging between 70% and 72%, but still do not reach the highest performance levels.

The highest test accuracies, around 73–74%, are achieved with higher thresholds such as "> 10" and "> 12". These thresholds indicate that when the model focuses on reviews with a greater number of helpful votes, it is able to generalize more effectively and perform better. Notably, the highest overall accuracy, approximately 74%, occurs at a threshold of "> 12" with 80 hidden neurons. This suggests that reviews with many helpful votes provide the model with more reliable data for classification, possibly due to clearer distinctions between helpful and unhelpful reviews in these subsets.

In summary, higher thresholds for helpful votes (such as "> 10" and "> 12") combined with around 80–100 hidden neurons offer the best classification performance, while lower thresholds lead to poorer model accuracy. The moderate thresholds (Fig. 3) provide a middle ground, but ultimately, the model benefits most from training on reviews with a larger number of helpful votes, which may contain clearer patterns for the network to learn from.

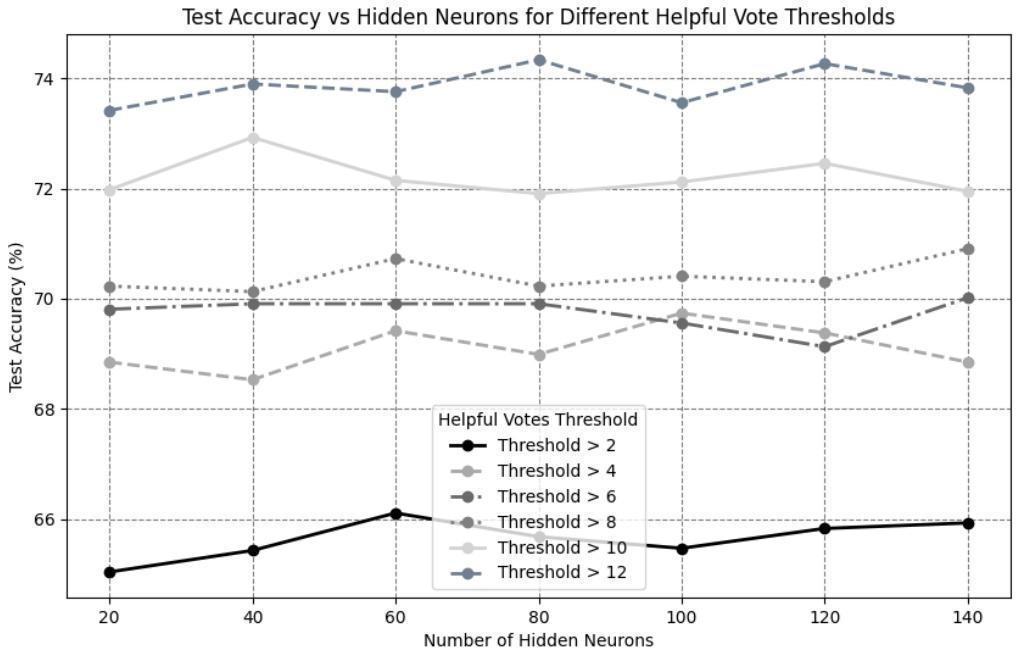


Fig. 3. Results of test accuracy with various helpful votes threshold and number of hidden neurons.

Table II presents model performance across various helpful vote thresholds and hidden neuron configurations, with metrics such as accuracy, precision, recall, and F1 score. These metrics provide a comprehensive evaluation of how well the model classifies reviews as helpful or unhelpful.

TABLE II. THE TEST ACCURACY, PRECISION, RECALL, AND FI SCORE FOR VARIOUS HELPFUL VOTES THRESHOLD AND NUMBER OF HIDDEN NEURONS

Helpful Votes Threshold	Hidden Neuron	Accuracy (%)	Precision	Recall	F1 Score
> 2	60	66.11	0.65	0.70	0.67
> 4	60	69.42	0.68	0.72	0.70
> 6	140	70.02	0.69	0.72	0.71
> 8	140	70.91	0.72	0.69	0.70
> 10	40	72.93	0.72	0.74	0.73
> 12	80	74.34	0.75	0.73	0.74

The model's accuracy improves steadily as the threshold for helpful votes increases. For the lowest thresholds, "> 2" and "> 4", the model achieves accuracy scores of 66.11% and 69.42%, respectively. As the threshold increases to "> 6" and "> 8", accuracy rises to 70.02% and 70.91%. The highest accuracy, 74.34%, is achieved with the "> 12" threshold and 80 hidden neurons. This suggests that reviews with a higher number of helpful votes are easier for the model to classify, possibly due to clearer patterns of helpfulness in more highly voted reviews.

Precision generally improves as the threshold increases, indicating the model's growing ability to correctly identify helpful reviews as the threshold rises. For the "> 2" and "> 4" thresholds, precision starts at 0.65 and 0.68, respectively. It peaks at 0.75 for the "> 12" threshold, demonstrating that the model is better at reducing false positives (i.e., labeling unhelpful reviews as helpful) when working with reviews that have garnered more helpful votes.

The recall metric, which measures the model's ability to correctly identify all helpful reviews, is relatively stable across different thresholds, ranging from 0.69 to 0.74. Interestingly, the highest recall value of 0.74 is achieved at the "> 10" threshold, slightly higher than the recall at the "> 12" threshold (0.73). This suggests that the model is not significantly more likely to miss helpful reviews as the threshold increases.

The F1 score, which balances precision and recall, follows a similar trend to accuracy. It starts at 0.67 for the "> 2" threshold and gradually increases to 0.74 for the "> 12" threshold. This indicates that as the threshold rises, the model becomes better at balancing false positives and false negatives. The highest F1 score (0.74) at the "> 12" threshold demonstrates the model's overall strongest performance with this higher threshold.

The model's performance improves consistently as the threshold for helpful votes increases. Higher thresholds, such as "> 10" and "> 12," yield the best results in terms of accuracy, precision, and F1 score, showing that reviews with more helpful votes provide better training data for classifying helpfulness. In particular, the threshold "> 12" paired with 80 hidden neurons achieves the highest overall accuracy (74.34%) and F1 score (0.74), making it the optimal configuration in this context. However, recall remains relatively stable across thresholds, indicating that the model consistently identifies a high proportion of helpful reviews regardless of the threshold. This analysis suggests that setting a higher helpful votes threshold allows the model to focus on more reliable data, leading to better classification performance. Overall, the findings indicate a robust performance in the model's classification capabilities, highlighting its potential for assisting users in identifying helpful content amidst a large volume of reviews.

The model's performance may be impacted by the presence of both search and experience products within the beauty, health, and personal care category. Search and experience products have distinct characteristics that influence how consumers evaluate reviews. For search products, features like skin type compatibility, ingredients, and fragrance are objective and easy to compare [70], [71], [18], leading consumers to generally agree on these qualities. Consequently, reviews for search products often need a higher number of helpful votes to meet consumer expectations for helpfulness, as they offer limited new insights [72].

On the other hand, experience products involve more subjective aspects, with consumers forming varied opinions based on personal experience [71]. Reviews for these products tend to provide unique, valuable perspectives [73] that consumers find helpful even if they receive fewer helpful votes. This difference means that a dataset containing both types of products may create diversity in review patterns that affect the model's ability to generalize and accurately predict helpfulness. As a result, the model might struggle to perform optimally compared to models trained exclusively on either search or experience product data.

V. Conclusion

This study presents a research process aimed at enhancing the classification of helpful reviews on online platforms, ultimately improving the experience for consumers navigating through vast amounts of user-generated content. The findings demonstrate that the performance of our model significantly improves as the threshold for helpful votes increases, particularly when combined with an optimal number of hidden neurons. The highest test accuracy and F1 score were achieved with thresholds of "> 10" and "> 12," suggesting that reviews receiving greater helpful votes provide clearer patterns for the model to learn from.

While the model achieved balanced F1 scores and demonstrated robustness in distinguishing between helpful and unhelpful reviews, certain limitations emerged, particularly in handling both search and experience products within the beauty, health, and personal care domains. The varied characteristics between these product types - where search products offer more objective qualities, and experience products depend heavily on subjective assessments - introduced challenges in model generalization. This differentiation emphasizes the need for adaptive models that can account for the unique features of each product type, potentially by utilizing additional domain-specific content indicators.

Overall, the findings underscore the value of content-based features in predicting review helpfulness and provide a foundation for future advancements in online review classification systems. By identifying and prioritizing content that is likely to aid consumer decision-making, e-commerce platforms can enhance the user experience, foster trust, and promote customer engagement. This work contributes to the ongoing effort to make consumer review systems more effective and efficient in delivering relevant, trustworthy information to users. Furthermore, this research flow can be integrated into existing e-commerce platforms by enhancing the review sorting mechanisms to prioritize helpful reviews based on consumer preferences. For example, platforms could implement an algorithm that displays reviews with high helpfulness scores at the top of product pages, facilitating quicker decision-making for consumers. Additionally, the framework could offer visual indicators for reviews deemed most helpful, improving user engagement with content that meets their specific needs. By integrating our framework, e-commerce platforms could not only improve the quality of information presented to consumers but also foster greater trust and satisfaction, ultimately leading to enhanced purchasing decisions.

While this study has demonstrated significant advancements in predicting the helpfulness of online reviews using information relevance theory and Artificial Neural Networks (ANN), certain limitations must be acknowledged. Firstly, the dataset utilized focuses exclusively on reviews from a single platform (Amazon) within specific product categories, which may limit the generalizability of the findings to other e-commerce platforms or product types. Additionally, the binary classification approach based on helpful votes thresholds may not fully capture the nuanced perceptions of review helpfulness among diverse consumer groups. Another limitation is the potential bias introduced by the predominance of English-language reviews, which excludes multilingual perspectives and insights. Lastly, the variability in consumer behavior across search and experience products introduces challenges in model generalization, underscoring the need for adaptive or hybrid approaches that consider product-specific factors for improved predictive accuracy. Future research should address these limitations to enhance the robustness and applicability of the proposed model.

VI. Future Work

Future research could explore several directions. First, it would be valuable to explore whether the conclusions drawn from this study apply to data from other online marketplaces beyond Amazon.com, which could broaden the relevance of our findings. Second, expanding the dataset reviews spanning a broader range of years to assess the framework's robustness over time and capture evolving trends in review helpfulness. Third, integrating other advanced natural language processing techniques, such as sentiment analysis and topic modeling, could enhance the model's ability to capture intricate aspects of helpfulness that vary across review types. Additionally, employing adaptive thresholding, which dynamically adjusts based on product type or user interaction data, might yield a more tailored approach to identifying helpful reviews. Finally, investigating the potential of hybrid models that combine content features with reviewer and product metadata could offer a more comprehensive framework for evaluating review helpfulness across various platforms and consumer needs.

References

[1] N. Sahoo, C. Dellarocas, and S. Srinivasan, "The impact of online product reviews on product returns", Inf. Sys. Res., vol. 29, no. 3, pp. 723–738, 2018.

[2] K. Pooja, and P. Upadhyaya, "What makes an online review credible? A systematic review of the literature and future research directions," Manag. Rev. Q., vol. 74, pp. 627–659, 2024.

[3] Y. Wu, L. Chen, E. W. T. Ngai, and P. Wu, "Stimulating positive reviews by combining financial and compassionate incentives," Internet Res., https://doi.org/10.1108/INTR-01-2023-0062, 2024.

[4] J. Li, X. Xu, and E. W. T. Ngai, "How review content, sentiment and helpfulness votes jointly affect trust of reviews and attitude", Internet Res., https://doi.org/10.1108/INTR-01-2023-0025, 2024.

[5] J. Luo, Y. Zhang, Y. Guo, and J. Zhang, "A novel method based on knowledge adoption model and non-kernel SVM for predicting the helpfulness of online reviews," J. Oper. Res. Soc., vol. 75, no. 6, pp. 1205-1222, 2024.

[6] X. Liu, G. A. Wang, W. Fan, and Z. Zhang, "Finding useful solutions in online knowledge communities: A theory-driven design and multilevel analysis," Inf. Syst. Res., vol. 31, no. 3, 2020.

[7] A. H. Huang, K. Chen, D. C. Yen, and T. P. Tran, "A study of factors that contribute to online review helpfulness," Comput. Human Behav., vol. 48, pp. 17–27, 2015.

[8] C. Cheung, M. Lee, and N. Rabjohn, "The impact of electronic word-of-mouth – The adoption of online opinions in online customer communities," Internet Res., vol. 18, no. 3, pp. 229–247, 2008.

[9] S. W. Sussman, and W. S. Siegal, "Informational influence in organizations: An integrated approach to knowledge adoption," Inf. Syst. Res., vol. 14, no. 1, pp. 47-65, 2003.

[10] S. Watts, G. Shankaranarayanan, and A. Even, "Data quality assessment in context: a cognitive perspective," Decis. Support Syst., vol. 48, no. 1, pp. 202–211, 2009.

[11] Y. C. Chen, R. A. Shang, and M. J. Li, "The effects of perceived relevance of travel blogs' content on the behavioral intention to visit a tourist destination," Comput. Hum. Behav., vol. 30, pp. 787–799, 2014.

[12] A. L. Jepsen, "Factors affecting consumer use of the Internet for information search," J. Interact. Mark., vol. 21, no. 3, pp. 21–34, 2007.

[13] R. Filieri, "What makes online reviews helpful? A diagnosticity-adoption Framework to explain informational and normative influences in e-WOM," Journal of Business Research, 68(6), 1261–1270, 2015.

[14] R. Filieri, and F. McLeay, "E-WOM and accommodation: an analysis of the factors that influence travelers' adoption of information from online reviews," J. Travel Res., vol. 53, no. 1, pp. 44-57, 2014.

[15] J. L. Nicolau, Z. Xiang, and D. Wang, "Daily online review sentiment and hotel performance," Int. J. Contemp. Hosp. Manag., vol. 36, no. 3, pp. 790 – 811, 2024.

[16] P. K. Jain, R. Pamula, and G. Srivastava, "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews", Comput. Sci. Rev., vol. 41, pp. 1- 17, 2021.

[17] Y. -H. Cheng, and H. -Y. Ho, "Social influence's impact on reader perceptions of online reviews," J. Bus. Res., vol. 68, no. 4, pp. 883-887, 2015.

[18] S. M. Mudambi, and D. Schuff, "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com," MIS Quarterly, vol. 34, no. 1, pp. 185-200, 2010.

[19] Y. Meng. N. Yang, Z. Qian, and G. Zhang, "What Makes an Online Review More Helpful: An Interpretation Framework Using XGBoost and SHAP Values," J. Theor. Appl. Electron. Commer. Res., vol. 16, no. 3, pp. 466-490, 2020.

[20] N. S. B. C. Lah, A. R. B. C. Hussin, and H. M. Dahlan, "A concept-level approach in analyzing review readership for E-Commerce persuasive recommendation," *Proceedings of the International Conference on Research and Innovation in Information Systems*, Langkawi, Malaysia, pp. 1-5, 2017.

[21] S. Krishnamoorthy, "Linguistic features for review helpfulness prediction," Expert Syst. Appl., vol. 42, pp. 3751-3759, 2015.

[22] D. Y. Ozdemir, "Essays on Novelty in Online Reviews," [Doctoral Thesis, The University of Texas at Dallas], 2023.

[23] A. G. Mumuni, K. O'Reilly, A. MacMillan, S. Cowley, and B. Kelley, "Online Product Review Impact: The Relative Effects of Review Credibility and Review Relevance," J. Internet Commer., vol. 19, no. 2, 2020.

[24] B. Lima, and T. Nogueira, "Novel features based on sentence specificity for helpfulness prediction of online reviews," 8th Brazilian Conference on Intelligent Systems (BRACIS), pp. 84–89, 2019.

[25] Z. Liu, and S. Park, "What makes a useful online review? Implication for travel product websites," Tour. Manag., vol. 47, pp. 140-151, 2015.

[26] S. J. S. Quaderi, and K. D. Varathan, "Identification of significant features and machine learning technique in predicting helpful reviews," PeerJ Comput. Sci., vol. 10, e1745, 2024.

[27] J. Kong and C. Luo, "Do cultural orientations moderate the effect of online review features on review helpfulness? A case study of online movie reviews," J. Retail. Consum. Serv., vol. 73, 2023.

[28] M. S. I. Malik, "Predicting users' review helpfulness: the role of significant review and reviewer characteristics," Soft Comput., vol 24, no. 18, pp. 13913-13928, 2020.

[29] M. Akgül, and A. R. Montazemi, "Online Review Helpfulness: A Literature Review," In M. Khosrow-Pour, D.B.A. (Ed.), Encyclopedia of Information Science and Technology, Sixth Edition, Advance online publication, 2025. https://doi.org/10.4018/978-1-6684-7366-5.ch055

[30] B. Ganguly, P. Sengupta and B. Biswas, "What are the significant determinants of helpfulness of online review? An exploration across product-types," J. Retail. Consum. Serv., vol. 78, 103748, 2024.

[31] X. Li, Q. Li, and J. Kim, "A Review Helpfulness Modeling Mechanism for Online E-commerce: Multi-Channel CNN End-to-End Approach," Appl. Artif. Intell., vol. 37, no. 1, 2023.

[32] T. Ghosal, T. Saikh, T. Biswas, A. Ekbal, and P. Bhattacharyya, "Novelty Detection: A Perspective from Natural Language Processing," Comput. Linguist., vol.48, no. 1, pp. 77–117, 2022.

[33] J. Allan, C. Wade and A. Bolivar, "Retrieval and novelty detection at the sentence level," In Proc. 26th Annual Internat. ACM SIGIR Conf. Res. and Development Inform. Retreival, pp 314–321, 2003.

[34] Zhang, J., Z. Ghahramani, and Y. Yang (2004). A probabilistic model for online document clustering with application to novelty detection. Adv. Neural Inform. Processing Systems 17.

[35] Zhang, Y., J. Callan, and T. Minka (2002). Novelty and redundancy detection in adaptive filtering. In Proc. 25th Annual Internat. ACMSIGIR Conf. Res. and Development Inform. Retrieval, pp. 81–88.

[36] Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference.

[37] Reimers, N. and I. Gurevych (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

[38] W-. J. Ko, G. Durrett, and J. J. Li, "Domain agnostic real-valued specificity prediction," in 33rd AAAI Conf. Artif. Intell., 2019.

[39] A. Louis, and A. Nenkova, "Automatic identification of general and specific sentences by leveraging discourse annotations," In Proceedings of 5th International Joint Conference on Natural Language Processing. Chiang Mai, Thailand: Asian Federation of Natural Language Processing, pp. 605–613, 2011.

[40] A. Louis, and A. Nenkova, "A corpus of general and specific sentences from news," In Proceedings of the Eighth International Conference on Language Resources and Evaluation, pp. 1818–1821, 2012.

[41] R. Zhang, J. Guo, Y. Fan, Y. Lan, J. Xu, and X. Cheng "Learning to Control the Specificity in Neural Response Generation," In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia, pp. 1108–1117, 2018.

[42] J. J. Li, and A. Nenkova, "Fast and accurate prediction of sentence specificity," In Proc of 29th AAAI Conf. Artif. Intell., pp 2281–2287, 2015.

[43] L. Lugini, and D. Litman, "Predicting Specificity in Classroom Discussion," In Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications, pp. 52–61, 2017.

[44] X. Liu, W. B. Croft, P. Oh, and D. Hart, "Automatic recognition of reading levels from user queries," In Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp.548-549, 2004.

[45] M. S. Khan, A. Rizwan, M. S. Faisal, T. Ahmad, M. S. Khan, and G. Atteia, "Identification of Review Helpfulness Using Novel Textual and Language-Context Features," Mathematics, vol. 10, 3260, 2022.

[46] Chua, A.Y.; Banerjee, S. Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. Comput. Hum. Behav. 2016, 54, 547–554.

[47] J. W. Pennebaker, R. J. Booth, and M. E. Francis, “Linguistic inquiry and word count: LIWC”, 2007. www.liwc.net

[48] Rausser, G.C.; Simon, L.; Zhao, J. Rational exaggeration and counter-exaggeration in information aggregation games. Econ. Theory 2015, 59, 109–146.

[49] C. L. Lai, K. Q. Xu, R. Y. K. Lau, Y. Li and D. Song, "High-Order Concept Associations Mining and Inferential Language Modeling for Online Review Spam Detection," 2010 IEEE International Conference on Data Mining Workshops, Sydney, NSW, Australia, pp. 1120-1127, 2010a.

[50] C. L. Lai, K. Q. Xu, R. Y. K. Lau, Y. Li and L. Jing, "Toward a Language Modeling Approach for Consumer Review Spam Detection," 2010 IEEE 7th International Conference on E-Business Engineering, Shanghai, China, pp. 1-8, 2010b.

[51] R. Y. K. Lau, S. Y. Liao, R. C. -W. Kwok, K. Xu, Y Xia, and Y. Li, “Text mining and probabilistic language modeling for online review spam detection,” ACM Trans. Manage. Inf. Syst., vol. 2, no. 4, 2012.

[52] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, “What Yelp Fake Review Filter Might Be Doing?,” Proceedings of the International AAAI Conference on Web and Social Media, vol. 7, no. 1, pp. 409-418, 2013.

[53] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, ‘‘Spotting fake reviews via collective positive-unlabeled learning,’’ in Proc. IEEE Int. Conf. Data Mining, pp. 899–904, 2014.

[54] McAuley Lab, Amazon Reviews’23 [Data set], 2024. https://amazon-reviews-2023.github.io/

[55] S. Xiao, G. Chen, C. Zhang, and X. Li, “Complementary or Substitutive? A Novel Deep Learning Method to Leverage Text-image Interactions for Multimodal Review Helpfulness Prediction,” Expert Syst. Appl., vol. 208, 118138, 2022.

[56] N. Jindal, and B. Liu, “Analyzing and Detecting Review Spam”, in Proc IEEE Int. Conf. Data Min., pp. 547-552, 2007.

[57] X. Guo, G. Chen, C. Wang, Q. Wei, and Z. Zhang, “Calibration of Voting-Based Helpfulness Measurement for Online Reviews: An Iterative Bayesian Probability Approach,” INFORMS J. Comput., vol. 33, no. 1, pp. 246-261, 2021.

[58] M. T. van Kesteren, D. J. Ruiter, G. Fernández, and R. N. Henson, “How schema and novelty augment memory formation,” Trends Neurosci., vol. 35, no. 4, pp. 211-219, 2012.

[59] W. Zhu, J. Mou, and M. Benyoucef, ‘Exploring purchase intention in cross-border E-commerce: A three stage model,” J. Retail. Consum. Serv., vol. 51, pp. 320–330, 2019.

[60] D. R. Fesenmaier, Z. Xiang, B. Pan, and R. Law, “A framework of search engine use for travel planning”, J. Travel Res., vol. 50, no. 6, pp. 587-601, 2011.

[61] B. J. Jansen, and A. Spink, “How are we searching the world wide web? A comparison of nine search engine transaction logs,” Inf. Process. Manag., vol. 42, no. 1, pp. 248-263, 2006.

[62] S. Paget, “Local Consumer Review Survey 2024: Trends, Behaviors, and Platforms Explored,” Accessed 9 August 2024, https://www.brightlocal.com/research/local-consumer-review-survey/.

[63] I. Raoofpanah, C. Zamudio, and C. Groening, “Review reader segmentation based on the heterogeneous impacts of review and reviewer attributes on review helpfulness: a study involving ZIP code data,” J. Retailing Consum. Serv., vol. 72, 103300, 2023.

[64] A. Ghose, and P. Ipeirotis, “Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics,” IEEE Trans. Knowl. Data Eng., vol. 23, pp. 1498–1512, 2011.

[65] Z. Zhang, and B. Varadarajan, “Utility scoring of product reviews,” In Proc of the 15th ACM Int. Conf. Inf. Knowl. Manag., pp. 51–57, 2006.

[66] N. Hu, I. Bose, Y. Gao, and L. Liu, "Manipulation in digital word-of-mouth: A reality check for book reviews", Decis. Support Syst., vol. 50, no. 3, pp. 627-635, 2011.

[67] Annie Louis and Ani Nenkova. Text Specificity and Impact on Quality of News Summaries. In Proceedings of the Workshop on Monolingual Text-To-Text Generation, pages 34–42, 2011.

[68] M. J. Kim, K. K.-C., Park, M. Mariani, and S. F. Wamba, “Investigating reviewers' intentions to post fake vs. authentic reviews based on behavioral linguistic features,” Technol. Forecast. Soc. Change, vol. 198, pp. 122971, 2024.

[69] M.S.I. Malik, and A. Hussain, “An analysis of review content and reviewer variables that contribute to review helpfulness,” Inf. Process. Manage., vol. 54, no. 1, pp. 88–104, 2018.

[70] M. J. Kim, K. K.-C., Park, M. Mariani, and S. F. Wamba, “Investigating reviewers' intentions to post fake vs. authentic reviews based on behavioral linguistic features,” *Technol. Forecast. Soc. Change*, vol. 198, pp. 122971, 2024.

[71] L. Huang, C.-H. Tan, W. Ke, and K.-K. Wei, “Comprehension and assessment of product reviews: a review-product congruity proposition”, *J. Manag. Inf. Syst.*, vol. 30, no. 3, pp. 311–343, 2013.

[72] S. Xiao, G. Chen, C. Zhang, and X. Li, “Complementary or Substitutive? A Novel Deep Learning Method to Leverage Text-image Interactions for Multimodal Review Helpfulness Prediction,” *Expert Syst. Appl.*, vol. 208, 118138, 2022.

[73] Y. Ma, Z. Xiang, Q. Du, and W. Fan, “Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep leaning,” *Int. J. Hosp. Manag.*, vol. 71, pp. 120-131, 2018.

An Advanced Semantic Feature-Based Cross-Domain PII Detection, De-Identification, and Re-Identification Model Using Ensemble Learning

Poornima Kulkarni[1], Cauvery N K[2], Hemavathy R[3]
Department of ISE, RV College of Engineering, Bengaluru, India[1, 2]
Department of CSE, RV College of Engineering, Bengaluru, India[3]

Abstract*—The digital data being core to any system requires communication across peers and human machine interfaces; however, ensuring (data) security and privacy remains a challenge for the industries, especially under the threat of man-in-the-middle attacks, intruders and even ill-intended unauthorized access at warehouses. Almost all digital communication practices embody personally identifiable information (PII) like an individual's address, contact details, identification credentials etc. The unauthorized or ill-intended access to these PII attributes can cause major losses to the individual and therefore it is inevitable to identify and de-identify aforesaid PII elements across digital platforms to preserve privacy. Unfortunately, the diversity of PII attributes across disciplines makes it challenging for state-of-arts to perform PII detection by using a predefined dictionary. The model developed for a specific PII type can't be universally viable for other disciplines. Moreover, applying multiple dictionaries for the different disciplines can make a solution more exhaustive. To alleviate these challenges, in this paper a robust ensemble of ensemble learning assisted semantic feature driven cross-discipline PII detection and de-identification model (EESD-PII) is proposed. To achieve it, a large set of text queries encompassing diverse PII attributes including personal credentials, healthcare data, finance attributes etc. were considered for training based PII detection and classification. The input texts were processed for the different preprocessing tasks including stopping-word removal, punctuation removal, website-link removal, lower case conversion, lemmatization and tokenization. The tokenized text was processed for Word2Vec driven continuous bag-of-word (CBOW) embedding that not only provided latent feature space for analytics but also enabled de-identification to preserve security aspects. To address class-imbalance problems, synthetic minority over-sampling techniques like SMOTE, SMOTE-BL, SMOTE-ENN were applied. Subsequently, the resampled features were processed for the feature selection by using Wilcoxon Rank Sum Test (WRST) method that in sync with 95% confidence interval retained the most significant features. The selected features were processed for Min-Max Normalization to alleviate over-fitting and convergence problems, while the normalized feature vector was classified by using ensemble of ensemble learning model encompassing Bagging, Boosting, AdaBoost, Random Forest and Extra Tree Classifier as base classifier. The proposed model performed a consensus-based majority voting ensemble to annotate each text-query as PII or Non-PII data. The positively annotated query can later be processed for dictionary-based PII attribute masking to achieve de-identification. Though, the use of semantic embedding serves the purpose towards NLP-based PII detection, de identification and re-identification tasks. The simulation results reveal that the proposed EESD-PII model achieves PII annotation accuracy of 99.77%, precision 99.81%, recall 99.63% and F-Measure of 99.71%.

Keywords*—PII Detection; machine learning; natural language processing; artificial intelligence; de-identification

I. INTRODUCTION

The last few years have witnessed significantly high-pace rise in digital data and allied communication. Increasing internet uses has broadened the horizon for digital data communication to serve socialization, business communication, networking etc. [1]. In fact, the modern world with exponentially rising population can't be hypothesized to exist without digital data. The digital data obtained by means of Email, Blogs, reviews, diagnosis details, business communication etc. can have personally identifiable elements including phone number, email ID, bank account details, social security numbers, diagnosis or ailments, insurance details, vehicle number, passport number etc. [1], [2]. The unauthorized and ill-intended intentional or unintentional use of these personally identifiable elements can cause both privacy breaches as well as losses [3-7]. For instance, an intruder can misuse insurance number and diagnosis details to cause cyber frauds and financial losses. Similarly, the unauthorized access to the bank account number, user identity number can give passage to the cyber criminals to cause digital crimes. Even the disclosure of house number to an unexpected intruder might broaden the horizon for stalking etc. [3], [4], [7]. To alleviate these issues, ensuring abstraction or masking of aforesaid personally identifiable elements can be of great significance. Though, the organizations claim to use personally identifiable information (PII) for constructive decisions including market-segmentation, profiling, marketing, business communication [8], [9]; however, the likelihood of PII misuse within organization or outside the mainstream mechanism can't be ruled out [10-13]. The unauthorized users or intruders might exploit one's online digital behaviour including texts, reviews, Email, personal chats etc. to misuse foresaid PII elements. To ensure privacy preserved communication identifying aforesaid PII is really a challenging task [5-7]. It becomes even more difficult to detect PII elements from unstructured and heterogenous data, which is unavoidable in modern digital world [6], [7].

Though, the classical approaches apply dictionary-based methods to detect and identify PII elements in text; however, preparing a large set of dictionaries for each discipline is near

infeasible or highly exhaustive. Moreover, the PII elements vary across the disciplines. For example, the PII elements pertaining to the user's personal identify might differ from the one from healthcare domain. In this case, generalizing the dictionary from one discipline is infeasible for the other and hence a model trained over one type of dictionary can't be used for other [14]. It limits efficacy of these solution towards real-time PII identification tasks. On the other hand, it is really difficult to prepare dictionary for each discipline distinctly. To alleviate these challenges, though a few machine-learning (ML) based solutions are proposed in the past [14]; however, almost all state-of-arts have been prepared for single domain and hence their scalability towards other discipline remains suspicious. To alleviate it, recently, the organization named European Unions suggested General Data Protection Regulation (GDPR) define PII attributes over broader spectrum to enable PII named identify representation (NER) for the different domains or disciplines [7]. However, ensuring both PII element heterogeneity, contextual details and optimal computing remained challenge for industry to ensure PII detection, de-identification and re-identification optimal. It can limit the performance of solution over large heterogenous data environment [15-17]. The lack of contextual details over large heterogenous and unstructured data makes PII detection challenging [15][16][18][19] and complex [20]. Though, to reduce computational exhaustion fast search space method was proposed [20] where expression matching was done over text content to perform PII detection. Ontology methods too can lack contextual dependencies to perform PII detection [17]. Though, the natural language programming (NLP) methods possess greater significance with the ML classifiers; however, ensuring feature optimality is inevitable to achieve better performance [21]. Despite efforts the diversity of PII elements amongst the different discipline such as healthcare [18][19], legal text documents [22], user browsing data and email [23] and academic or non-academic publication [24], make major state-of-arts confined and generalizing a solution over other discipline texts can yield low accuracy. This is because almost all ML-driven solutions perform PII detection for a specific data type or discipline [24], [25]. It clearly indicates that a model designed to perform PII detection in financial data can't be applicable in EHR-related dataset [75]. It motivates researchers to design a cross-domain learning environment, where one can take the inputs text data from the different domains (i.e., healthcare, business communication, financial data etc.) and train a model to have broader knowledge ability for PII detection and classification. Ironically, none of the state-of-art could address this problem so far. Though, the use of NLP with better feature engineering, contextual learning and robust classification can enable a cross-domain PII detection and classification. However, it requires training a robust learning model with maximum possible features encompassing samples from the different domains. Despite numerous studies in the past, none of the solutions could address class-imbalance problem while using ML or NLP methods for PII detection and classification. On the contrary, it is inevitable that the real-time systems might have text queries relatively higher for non-PII than the PII containing queries or sentences. Such skewed data and allied learning might force model(s) yielding false positive or false negative. Therefore, an NLP driven PII detection and classification model requires addressing class-imbalance problem to improve learning and classification results. Additionally, processing a training model over most representative PII text samples with embedded latent or semantic details can make learning more efficient and accurate. These key inferences can be considered as the predominant motivation behind this study.

Considering aforesaid motivations and allied scopes, in this paper a novel and robust ensemble of ensemble learning assisted semantic feature driven cross-discipline PII detection and de-identification model (EESD-PII) is proposed. To achieve it, a large set of text queries encompassing diverse PII attributes including personal credentials, healthcare data, finance attributes etc. were considered for training based PII detection and classification. The input texts were processed for the different pre-processing tasks including stopping-word removal, punctuation removal, website-link removal, lower case conversion, lemmatization and tokenization. The tokenized text was processed for Word2Vec-CBOW embedding that not only provided latent feature space for analytics but also enabled de-identification to preserve security aspects. To address class-imbalance problem, synthetic minority over-sampling techniques like SMOTE, SMOTE-BL, SMOTE-ENN were applied. Subsequently, the resampled features were processed for the feature selection by using WRST algorithm that in sync with 95% confidence interval retained the most significant features. The selected features were processed for Min-Max Normalization to alleviate over-fitting and convergence problems, while the normalized feature vector was classified by using ensemble of ensemble learning model encompassing Bagging, Boosting, AdaBoost, Random Forest and Extra Tree Classifier as base classifier. The proposed ensemble learning model performed consensus-based majority voting ensemble to annotate each text-query as PII or Non-PII data. The positively annotated query was later processed for a dictionary-based PII attribute detection and masking to achieve de-identification. The simulation results reveal that the proposed EESD-PII model achieves PII annotation accuracy of 99.77%, precision 99.81%, recall 99.63% and F-Measure of 99.71%.

The remaining sections in this paper are divided as follows. Section II discusses the related work, while the problem formulation and questions are given in Section III and Section IV, respectively. The overall proposed model is discussed in Section V, while the proposed model is given in Section VI. Conclusion and allied inferences are given in Section VII, while Section VIII discusses the future scope. The references used are given at the end of the manuscript.

II. Related Work

The authors in study [26] applied conditional random field (CRF) and SVM to learn the pre-annotated PII lexical and embedding features to classify PII query. On the other hand, the authors [27] infused the phrase embedding concept to yield semantic details, which was later trained over machine learning to perform named entity detection and classification. On the other hand, the allied challenges in named entity recognition were discussed in study [28], where the authors concluded that the use of text-specific semantic details trained over better learning environment can perform more accurate PII identification. However, the authors failed to indicate their

suitability over the different sets of text inputs. The authors [29] used long and short-term memory (LSTM) deep network to extract and learn contextual features from the input queries to perform PII detection. The recurrent neural networks (RNNs) were applied in [30], [31] as well where the authors used pre-trained transformer-based language models for PII prediction. These approaches couldn't address large PII heterogeneity amongst queries from the different disciplines. Moreover, merely applying cross-validation learning can't generalize the solution for cross-discipline PII detection and classification [32]. In study [33], PrivacyBot was designed, especially to perform privacy sensitive PII detection over unstructured text input. The authors applied machine learning models to train over the text content for classification. Yet, their generalizability over large unstructured muti-source inputs remains suspicious [34]. In [35] deep learning was applied to classify annotated text corpus for PII classification. The authors in study [36] made use of the convolutional neural network to perform named entity recognition in unstructured text data. However, it was completely depending on the local hierarchical features, while a non-lineal environment demands a model to have contextual learning capabilities to yield accurate prediction results. In study [37], the conceptual relationship amongst the unstructured text was applied to perform PII detection from input texts. More specifically, an association extraction approach was applied on Twitter data to assess the relatedness of each word with PII attributes by using pointwise mutual information statistical association rule. The authors in study [38] exploited privacy information across privacy-related topics, including plan for the vacations, alcohol and healthcare conditions to perform PII detection and classification. Despite ML classifier, guaranteeing scalability of solution over large heterogenous inputs makes it suspicious. In [39], a semi-supervised ML [17] was applied for PII detection in healthcare text. The authors in study [17] fused topic modelling, privacy ontology and sentiment score, which was later processed by Naïve Bayes (NB) classifier to detect PII attribute in Twitter data. In study [40], PII dictionary was applied to annotate text elements as PII or non-PII. In study [41], the authors used ontologies with rule-based method to perform PII detection in text. The authors in study [42] designed a text feature learning driven PII detection interface for software-defined networks (SDN) that classifies each query as PII or non-PII. Though, it failed addressing numerous issues including feature optimality, class-imbalance and low accuracy problem. In [43], the authors used dictionary and intra-query structural details to perform PII detection and classification. RenCon [44] was proposed to detect PII in mobile network traffic, where supervised ML was used to annotate each query as PII or Non-PII. A similar work was done in study [5] where the authors applied ML algorithm to learn PII attributes personal identification (ID), phone number, social graph, email, location and biometric ID detail for PII detection. In [45] as well ML was applied where the Person of interest (PoI) and PII attributes were detected from the email dataset. The authors suggested that unlike standalone classifier, the use of ensemble learning method by using decision tree, SVM, neural network and random forest can yield better accuracy. PII detection over healthcare data was done in study [46], yet, despite BigData problem the authors failed in addressing the problem of class-imbalance [47]. The authors in study [48] used random forest algorithm with Simpson index that measured the diversity amongst the PII elements to annotate them as PII or non-PII element. Ontology method was applied in [49], [50] to detect PII and its masking. In [50], ML was applied to solve PII detection as an NLP problem.

In study [70], the authors focused on developing a new Korean dialogic dataset, especially designed for the PII de-identification. For PII detection, the authors used text anonymization benchmark and network intrusion detection dataset, based on which a new de-identification dataset was prepared. Unfortunately, there are very less efforts made towards text data de-identification; though there exists certain works related to the face de-identification in video or image dataset. For instance, in study [71] the authors developed a disentangled representation learning for multiple attributes preserving face de-identification. More specifically, they proposed replacing and restoring variational autoencoders (R 2 VAEs) that disentangle the identity-related factors and the identity-independent factors so that the identity-related information can be obfuscated, while they do not change the identity-independent attribute information. In [72], UU-Net was developed for the reversible face de-identification in visual surveillance video data. Here, the proposed UU-Net model learns jointly by optimizing a public module that receives the raw data and generates the de-identified stream, along with a private module, especially designed for security authorities. The second module receives the public stream and regenerate the original data by exploiting semantic and contextual details, disclosing the actual IDs of the subjects in a scene. This method made use of the conditional generative adversarial network to achieve synthetic faces by preserving pose, lighting, background information and even facial expressions. In study [73], the authors developed de-identification method by applying both human perceptions derived as semantic information and the face recognition models. This method explored the tradeoff between a user misidentifying the original identity with a well-known celebrity and a facial recognition model that tries to identify the original identity. It generated caricature faces of the de-identified faces to ensure that the manipulated faces can be distinguished effortlessly. In study [74], an attribute-preserving face de-identification framework called Enhanced Embedded Auto Encoders was developed. The proposed model embodied three components, privacy removal network (PRN), feature selection network and privacy evaluation network. Here, PRN made the model capable to discard information involving identity privacy and retaining desired face attributes for certain prediction applications. The other modules focused on exploiting global or contextual features to improve learning-based classification accuracy. Despite efforts, ensuring de-identification in text data, especially carrying critical details like electronic healthcare records (EHR) remains a challenge [75]. It broadens the horizon for the researchers to develop a scalable and robust PII detection, classification and deidentification and re-identification [75] model.

III. PROBLEM FORMULATION

This research hypothesizes that the inclusion of PII texts from the different disciplines for improved NLP analytics can enable cross-discipline PII detection and classification. This hypothesis considers optimization on both data as well as

computational front, where initially it embodies PII queries from the different domains which are processed for the different pre-processing tasks including stopping-word removal, punctuation removal, website-link removal, lower case conversion, lemmatization and tokenization. This approach can strengthen the proposed PII detection model to address data heterogeneity and unstructured-ness and hence can make it suitable for the real-world applications. In addition, it can make the proposed model suitable for cross-discipline PII detection and classification. Moreover, unlike syntactical term matching based solutions, to retain sufficiently large contextual details Word2Vec CBOW method is proposed that can provide large latent information to make cross-domain PII detection. It also helps in achieving de-identification by transforming input texts into equivalent embedding matrix, which can later be used to perform de-identification and re-identification. Realizing the at hand computational challenges like local minima and convergence, redundant computation, in addition to the aforesaid data optimization the proposed work focuses on computational aspects as well, where at first WRST significant predictor test is applied over the embedded metrics which retains a sufficiently large set of samples with high representativeness. It can reduce computational cost decisively while ensuring that the performance remains high. The selected samples have been processed for resampling by using SMOTE resampling methods including SMOTE, SMOTE-BL and SMOTEENN methods. Noticeably, these resampling methods intend to alleviate any probable class-imbalance problem and skewed data problem without undergoing iterative hotspot creation as witnessed with random sampling and up-sampling methods. The resampled data has been later processed for Wilcoxon Rank Sum Test (WRST) method that in sync with 95% confidence interval retained the most significant features. Subsequently, Min-Max normalization is performed that maps each input data in the range of 0 to 1, and thus alleviates the problem of over-fitting problem. Finally, the normalized feature vector was classified by using ensemble of ensemble learning model encompassing Bagging, Boosting, AdaBoost, Random Forest and Extra Tree Classifier as base classifier. The proposed ensemble learning model performed consensus-based majority voting ensemble to annotate each text-query as PII or Non-PII data. In this manner, it performs two-class classification which classifies each query as PII or Non-PII. Once annotating query as PII, the predefined dictionaries can be applied to perform search-based method to detect and mask PII element. Though, the proposed model applied word-embedding method that transformed original input text into equivalent embedding metrics and hence preserved privacy and/or de-identification to the data while retaining data-sanity for analytics.

IV. RESEARCH QUESTIONS

In reference to the aforesaid formulations, this research defines certain questions, whose justifiable answers can put foundation for a robust PII detection and classification system. These research questions (RQ) are:

RQ1: Can the use of multi-disciplinary PII text with multi-aspects pre-processing tasks like stopping-word removal, punctuation removal, website-link removal, lower case conversion, lemmatization and tokenization enable NLP to achieve cross-discipline PII detection and classification?

RQ2: Can the use of Word2Vec CBOW embedding, WRST significant predictor test, SMOTE resampling, and Min-Max normalization provide semantically rich feature vector for PII detection and classification?

RQ3: Can the use of ensemble-of-ensemble learning environment encompassing Bagging, Boosting, AdaBoost, Random Forest and Extra Tree Classifier as base classifier to perform PII detection and classification?

RQ4: Can the strategic amalgamation of the solutions RQ1-RQ4 provide a robust cross-discipline PII detection and classification system?

As indicated in RQ1, the use of the text data and allied PII variables (say, artefacts) can enable an AI tool to learn more effectively. Here, training any machine learning or deep model over heterogenous PII elements belonging to the different categories of text inputs can improve learning ability and knowledge. Undeniably, it can improve learning; however, the data heterogeneity over the different input sources, data nature, diversity of presentation etc. can impact overall learning. To address this problem, there is the need of transforming input raw data into corresponding processed structured data for further feature extraction and learning. In this reference, this research defines research question (RQ1) which assesses whether the pre-processing tasks like stopping-word removal, punctuation removal, website-link removal, lower case conversion, lemmatization and tokenization enable NLP to achieve cross-discipline PII detection and classification. In the same manner, to improve scalability even with the reduced computational costs, this research intends to use contextual or latent semantic information, which can enhance learning and classification accuracy. To achieve it, this research makes use of the semantic embedding technologies like Word2Vec embedding. This method intends to transform input (pre-processed text data) into corresponding latent embedding matrix (low-dimensional matrix). To further improve computational efficacy and reliability over non-linear, imbalanced data environment, it performs SMOTE resampling (to alleviate class-imbalance problem), and WRST significant predictor test-based feature selection. These methods altogether can improve feature environment and hence can alleviate any likelihood of class-imbalance, premature convergence, local minima etc. In this reference, this research defines research question RQ2, whether aforesaid methods can improve overall performance towards a run-time PII detection and classification solution. Unlike traditional machine learning methods like decision tree (DT), support vector machine (SVM), naïve bayes (NB), etc. the ensemble learning methods like random forest (RF), AdaBoost, extra tree classifier, XGBoost etc. have performed better for the different data and image classification problems. Being consensus driven prediction solutions, it yields higher accuracy and reliability towards PII detection and classification. Considering it as motivation, this research intends to use only ensemble learning methods, such as the Bagging, Boosting, RF, ETC, XGBoost to perform maximum voting ensemble-based learning and classification. In this method, a total of five base classifiers (here, ensemble classifiers) are considered. This is because, out of five base classifiers, if three of the base classifiers classifies an input as PII and annotates it as 1, it would

be classified as PII. On the other hand, if three base classifiers annotate an input as 0, the final prediction is made as non-PII. Being consensus driven approach or maximum voting-based approach, it can be more reliable towards run-time significances. These aspects are defined in terms of the research question RQ3. Since, the at hand problem is a challenge of text classification, this work considers standard classification related performance parameters such as the accuracy, prediction, recall and F-Measure as performance variable. In this reference, this work defines research question RQ4. Thus, the overall research intends to achieve justifiable answer for these questions (RQ1-RQ4).

V. System Model

This section discusses the overall proposed method and allied implementation. The overall research method encompasses the following key steps:

a) Data Acquisition,
b) Pre-processing,
c) Word2Vec CBOW Embedding,
d) Significant Predictor Test,
e) Resampling,
f) Normalization, and
g) Ensemble(E2E) MVE Classification.

A snippet of the overall proposed model is given in Fig. 1.

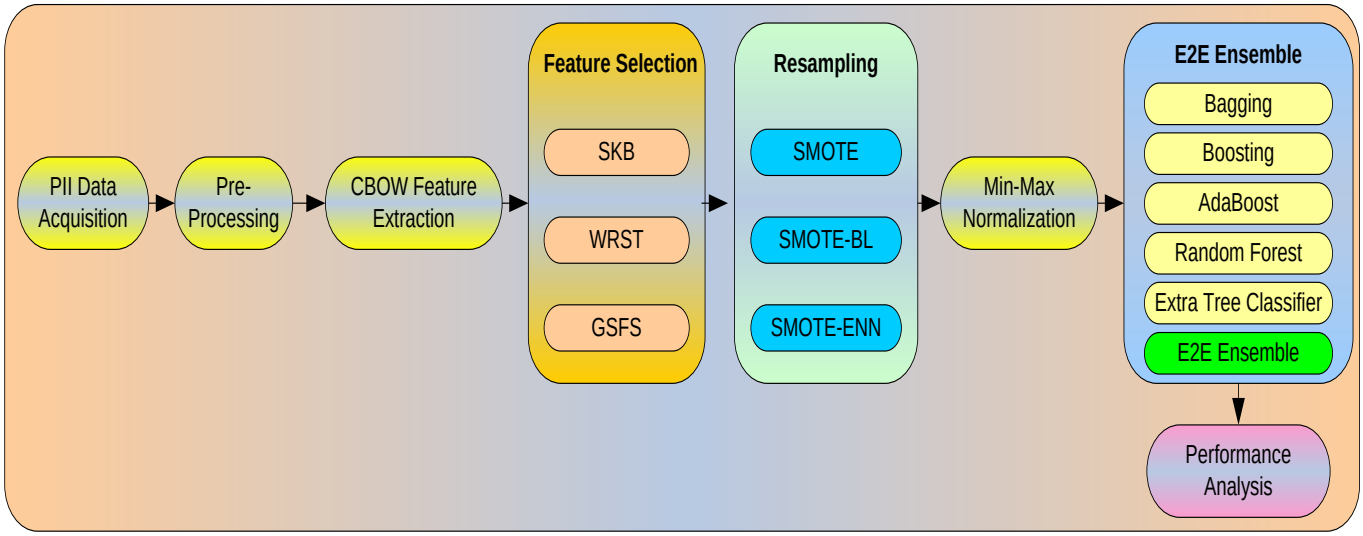


Fig. 1. Proposed EESD-PII model.

The detailed discussion of the overall proposed work is given as follows:

A. Data Acquisition

Considering real-world heterogenous and unstructured text data available across digital communication platforms such as Email, Blogs, social media, data warehouses etc., in this research primary data was synthesized. Moreover, since there exists very limited data available towards cross-discipline PII detection problem, we synthesized own primary data encompassing sufficiently large text queries possessing PII elements from the different domains, including personal credentials, healthcare data, financial data etc. In sync with targeted PII detection and classification problem, Python Faker tool was taken into consideration where a total of 12000 text (reviews) queries were synthesized. Realizing the diversity of PII attributes across different disciplines, a set of PII elements were prepared for each discipline. A snippet of the different PII attributes used for the different discipline is given in Table I.

As depicted in Table I, text queries from three different domains including personal credentials, healthcare credentials and financial credentials were prepared. Here, personal credentials encompassed the key PII attributes as “Name”, “Address”, “Email”, “Aadhar Number”, “PAN”, “SSN”, “Vehicle Number”, and “Phone Number”. Similarly, three different PII credentials including “Health Insurance Number”, “Diagnosis Details”, and “Life Insurance Number” were considered. The other attributes like “bank Account Number”, “Loan Account”, “Balance Details”, “OTP”, “Access PIN” and “Credit Card Number”, were prepared towards financial credentials. We synthesized 4000 queries from each discipline and thus a total of 12000 queries were obtained. To synthesize these reviews Python’s Faker package was applied that embodies different callable functions to generate a large number of unstructured texts. The aforesaid text reviews were generated arbitrarily in such way that it embodies aforesaid PII credentials (Table I) in the different texts and allied disciplines. In other words, each of the queries generated possessed the different PII elements pertaining to the personal credentials, healthcare and finance. Thus, the queries generated provided data heterogeneity while ensuring data privacy and integrity. This approach also serves the purpose of PII privacy preserving and anonymized data preparation. Recalling the fact that the faker generates queries with the aforesaid credentials randomly and hence guaranteeing uniform proportion of each PII attribute remains difficult. In this manner, the data can be hypothesized to be unbalanced and hence can give rise to the skewed learning. It eventually can give rise to the false positive or false negative performance. To alleviate this problem, applying resampling can

be of great significance. The resampling can improve data distribution uniform throughout and hence training a model over such balanced sample can improve learning and classification accuracy.

TABLE I. PII ATTRIBUTES

Discipline	PII Attributes	No. of Samples
Personal Credentials	Name Address Email Aadhar Number Permanent Account Number Social Security Number Vehicle Number Phone Number	4000
Healthcare Credentials	Health Insurance Number Diagnosis Details Life Insurance Number	4000
Financial Credentials	Bank Account Number Loan Account Balance Details One Time Password Access PIN Credit Card Number	4000

B. Pre-processing

In reference to the real-world data scenarios where the likelihood of data-heterogeneity and unstructured-ness can't be ruled out, strengthening data with the different pre-processing tool can be of great significance. Additionally, since the proposed method intends to exploit semantic word-embedding metrics (feature) improving data with certain set of pre-processing tasks can be of great significance. With this motive, we performed the different pre-processing tasks including the removal of stopping words, punctuation marks, Emoji, lower case conversion, lemmatization etc. We apply the following pre-processing tasks to ensure optimal data presentation and input towards targeted PII detection, de-identification and identification tasks.

a) Missing value removal
b) Unicode normalization
c) Emoji removal
d) Website link removal
e) Punctuations and Stopping word removal
f) Converting to the lower case
g) Removing stop-words
h) Lemmatization
i) Tokenization.

A snippet of these pre-processing tasks is given as follows:

1) Missing Value Removal: In real-world data environment, the text queries or reviews, especially collected from email, online Blogs, reviews, social media etc. can have certain broken sentences or incomplete words. There can be the set of words or strings with incomplete sentence representing the missing data element. Any NLP and allied AI-based solution, especially the text-analysis methods might yield inaccurate prediction results with aforesaid missing elements. In fact, the lack of contextual relatedness amongst aforesaid incomplete words or missing elements can impact overall performance. Learning over their allied features can impact learning and classification results. To address this problem, removing broken words and sentences from the input text corpuses can be vital. To achieve it, we applied standard NLTK Python library to remove the incomplete sentences. The overall process of missing value removal ensured that the proposed method examines and retains associations amongst the PII attributes and incomplete sentence(s).

2) Unicode Normalization: Realizing diversity amongst the reviews where the different users can have written the reviews with the different word-constitution, formats, (personalized) writing method, skills and texting designs, the proposed method performed Unicode normalization on each input query or sentence. The aforesaid data heterogeneity (amongst the word constituents and writing style) might impact NLP solution and hence transforming raw input text into a uniform structure is must. To achieve it, we performed Unicode normalization that transferred the input text into the single norm data output for further semantic feature extraction and corresponding de-identification tasks.

3) Emoji Removal: Though, the considered Faker tool didn't introduce any emoji element; however, in real-world data systems the likelihood of emoji (Ex. 😊, 😵) can't be ruled out, especially the modern mobile based text communications. Though, there exists no concrete literature which could prove their efficacy or significance towards NLP problems; however, their presence in raw data can impact learning adversely. In fact, readability of such elements can be difficult and can introduce data ambiguity that eventually can impact overall performance. Considering these facts, we performed emoji removal by using NLTK libraries and rule-based functions. This ability can make the proposed model robust in real-time applications where the targeted PII detection tool can be applied as an interface to detect, de-identify and re-identify PII attributes.

4) Hashtag and URL Removal: In present day digital world, where social media, web-media etc. have been playing decisive role in digital data promotions and allied marketing (say, communication). The users often try to mention the different URLs or web-links pertaining to certain subject matter. In addition, there can be user defined or organization driven hashtag information (i.e., #). The aforesaid URLs and #Hashtag helps the individual or organization gaining broader market space or audiences. Despite such significances, such web-URLs, links and/or #Hashtag don't have any significance towards PII detection and classification. The presence of such elements in texts can impact NLP learning and prediction (accuracy or) efficiency. It makes it inevitable to remove such elements before executing feature extraction and learning. In this paper, we applied rule-based methods to remove different hashtag and web-URLs. The rule-based method helped in identifying the URL components including www, https:// etc.,

that eventually helped in removing aforesaid contents or attributes. We applied Python URL removal function to remove hashtag and URL elements from each input query.

5) Punctuation and Stopping Words Removal: In text data the presence of the non-word symbols or the punctuation (i.e., ",", "?", "!",";", ":", etc.) can't be ruled out. Though, such elements have decisive significance towards contextual role or presentation; however, are not much related to the PII probability. Though, their presence in vicinity of a PII element, for instance, "AA0123456" can't be ruled out. Noticeably, though AA0123456 can be certain PII element; however, "" seems to be not related to the PII and just represents a highlighting element and/or remark. Therefore, removing such elements can be vital, especially when exploiting NLP for PII detection and classification. We applied NLTK library to remove punctuation marks from each text query or sentence. The deployed method applied standard expression and the rule-based approach to detect and remove punctuations from the input texts.

In general, the stop-words represent the terms that add no value to the sentence. In at hand NLP problem, the presence of these words can impact feature purity and its significance towards contextual (or semantic) learnability. Considering this fact, we dropped aforesaid stop-words from the input text corpus without making any change to the original intend behind the text phrase.

6) Lower Case Conversion: The ML algorithms can be case-sensitive and therefore the diversity amongst the case might impact learning efficiency. To cope up with the NLP solution, the relatedness and significance amongst the terms like SMART and smart can be differ and hence can cause contextual ambiguities in the extracted features. To alleviate this problem, the input texts were processed for lower case conversion, where each word or term was converted into lower case. We used Python's inbuilt lower-case conversion function to perform lower-case conversion.

7) Lemmatization: Once converting text data into equivalent lower-case sequence, each query was processed for lemmatization. In this method, the extended word was transformed into the respective root form. It assessed the expected component of the text without losing the original sense behind the word. Noticeably, stemming and lemmatization performs similar task; however, differs the way it (i.e., lemmatization) assesses the context first, which is then followed by the transformation of the extended word to the root word. On the contrary, in stemming the extended characters like "s", "es" etc. at the end of the word are removed that loses the actual intend of the word. The following illustrations depicts the process of stemming and lemmatization methods.

Studies → Lemmatization→ Study,

Studies → Stemming→ Studi.

As depicted above, to retain original intend or originality, lemmatization was performed where each word was transformed into respective original form.

8) Tokenization: Once performing lemmatization of the input text or queries, tokenization was performed that transformed each query into certain set of tokens. In this work, the tokenization method at first transformed input query or sentence into corresponding set of tokens, which were later processed for semantic feature extraction and learning to perform feature learning and classification.

C. Semantic Feature Extraction

Though, in the past a few ML and NLP-based efforts have been made towards PII detection and classification; however, most of the state-of-arts fail in address or exploit contextual feature to perform learning and classification. Additionally, unlike term-matching based approaches the use of latent information or semantic features can improve efficacy, especially in NLP-based solutions including at hand PII detection and classification. Unlike traditional PII detection models, in this work the focus was made on exploiting depth or semantically enriched features or latent features from the input queries is hypothesized. In addition, to presence data sanity, originality with privacy preserving (say, de-identification), word-embedding can be of great significance. More specifically, the use of Word2Vec based Continuous Bag of Words (CBOW) model can yield low-dimensional semantic feature to perform learning and classification for PII detection and classification. Unlike other approaches such as TF-IDF or N-skip gram with $n = 1$, CBOW yields low-dimensional features to perform learning and classification towards PII-detection. The use of Word2Vec CBOW model can be even effective over the data encompassing PII attributes from the different disciplines, including personal credentials, healthcare and finance. With such data diversity, the proposed Word2Vec CBOW model can be vital. Moreover, in the considered dataset there can be limited text elements amongst the non-PII attributes and therefore retaining allied contextual details can provide more insight towards NLP-based PII detection and classification. Training a ML solution with aforesaid contextual and semantic (or latent) information can make PII detection more efficient and accurate over unannotated or minimally annotated inputs. Though, the literatures suggest different semantic extraction methods including TF-IDF, Skip-Gram, Glove, etc.; however, CBOW provides more intrinsic feature with easier implementation over large inputs. Additionally, it yields relatively low-dimensional features with intact feature vector or contextual significance, which can provide suitable feature environment for ML to predict PII in input sentence or query. In reference to these inferences, we applied Word2Vec CBOW method to perform word-embedding over the tokenized inputs for each query. In this method, for each query it generates set of embedding metrics, which is subsequently used for optimization and learning to perform PII detection. A brief of the CBOW method used in this work is given as follows:

In Word2Vec CBOW, we performed word-embedding on the input tokenized inputs (say, text sequences per query). More specifically, Gensim Word2Vec method was applied that transformed the set of tokens into corresponding low-

dimensional semantic embeddings (say, embedding vector). To achieve it, we designed Word2Vec with dual-layer neural network having two hidden layers. It helped achieving more contextually rich but sparser features so as to improve computational aspects. In CBOW method, let, W_{i-1}, W_{i-2}, W_{i+1}, W_{i+2} be the context words retrieved from the sentence or review text. Then, CBOW predicts W_i which is related with the other tokens available in that specific query. The predicted embedding outputs are always related to the target token W_i. The detailed discussion of CBOW embedding is given as follows:

In general, the CBOW method contains two sets of word-embedding vectors, say, "Source-side" and "Target-side" vectors, signifying $v_w, v'_w \in \mathbb{R}^d$ for each sentence or query's tokens. Noticeably, in our proposed method, we obtained $w \in V$ as the tokenized vocabulary once performing tokenization for each text input or sentence. Being Gensim-based embedding, a text window within the input review or sentence contains central token w_0 which generates respective context embedded vector $w_1, \dots, w_C$. For the above stated conditions, (say, text-window), the CBOW loss is measured as in Eq. (1).

$$v_c = \frac{1}{C}\sum_{j=1}^{C} v_{w_j} \tag{1}$$

$$\mathcal{L} = -\,log\,\sigma\left(v'_{w0}\,T_{v_c}\right) - \sum_{i=1}^{k} log\,\sigma\left(-v'_{n_i}\,T_{v_c}\right) \tag{2}$$

In (2) $n_1, \dots, n_k \in V$ be the negative examples obtained from the noise distribution P_n over the input vector V. In Eq. (2), the parameter $\mathcal{L}$ gradient is estimated with respect to the target value v'_{w0}, negative target value v'_{n_i} and average context source (v_c).

$$\frac{\partial \mathcal{L}}{\partial v'_{w0}} = \left(\sigma\left(v'_{w0}\;T_{v_c}\right) - 1\right)v_c \tag{3}$$

$$\frac{\partial \mathcal{L}}{\partial v'_{n_i}} = \left(\sigma\left(v'_{n_i}\;T_{v_c}\right) - 1\right)v_c \tag{4}$$

$$\frac{\partial \mathcal{L}}{\partial v_c} = \left(\sigma\left(v'_{w0}\;T_{v_c}\right) - 1\right)v'_{w0} + \sum_{i=1}^{k}\left(\sigma\left(v'_{n_i}\;T_{v_c}\right) - 1\right)v'_{n_i} \tag{5}$$

Now, deploying Chain-rule method over the source context embedding, the gradient of the predicted word vector, also called the context vector is obtained as per Eq. (6).

$$\frac{\partial \mathcal{L}}{\partial v_{w_j}} = \frac{1}{C}\left[\left(\sigma\left(v'_{w0}\;T_{v_c}\right) - 1\right)v'_{w0} + \sum_{i=1}^{k}\left(\sigma\left(v'_{n_i}\;T_{v_c}\right) - 1\right)v'_{n_i}\right] \tag{6}$$

To address the problem of inaccurate context vector update, in this work context word normalization was performed. To achieve it, we sampled the context window's width arbitrarily in the predefined range of 1 to C_{max} for each target word. Thus, applying this method the embedded metrics for each query was obtained for further processing or eventual analytics task(s). Noticeably, this method not only provided the embedded metrics as semantic feature for learning and classification; but also transformed the original text including PII elements or attributes into equivalent pseudonymized data and hence fulfils the goal of deidentification, without losing data intend and originality towards further analytics. Unlike data masking, our proposed model can provide a suitable reversible data form with certain rule-based approaches. Though, in the proposed work, such tasks are not required as the PII element itself has been transformed into anonymized embedding vector or numeric value.

D. Ensemble Feature Selection

This is the matter of fact that in real-time applications, the volume of data inputs (here, text reviews) can be gigantic and hence training a ML method over such humongous data might give rise to the local minima and convergence and hence can yield inaccurate prediction result. Additionally, it can make overall processing exhaustive. The severity of such problems can be high in NLP domain, especially over the large text inputs. To alleviate these problems, retaining the set of most-significant samples having decisive impact on the (PII) prediction results can be vital. Though, there exists a number of state-of-art feature selection methods; however, in this work we examined applied three different methods, named Select K-Best, Wilcoxon Rank Sum Test and Gini-Score based significant predictor test to retain the optimal set of features, which were horizontally concatenated to yield an ensemble feature vector for further learning and classification. A brief of these methods is given as follows:

1) Select k-Best Feature Selection: The Select k-Best (SKB) method identifies the top-k most significant samples or allied features. We applied Chi-Squared method to perform Select K-Best based most significant feature selection. The applied Chi-Squared feature selection approach measures the level of significance of each feature to retain the top-K features. It is mainly achieved by estimating the value of χ^2statistics in reference to the target class. In this mechanism, each feature is examined separately to assess the relationship between the feature and target class. It acts as a non-parametric test method which compares the different variables for an arbitrarily chosen data. As an independence-test approach, it identified the difference and independence amongst the different arbitrarily selected variables. Thus, it measured a value on the basis of the association between the feature instance and the class it must belong to. For output 0, it states that there exists no association between the feature element and the class. The greater association refers stronger relationship between the feature (say, sample) and the probable class. The use of Chi-Squared method was done by using scikit-learn library which helped retaining the set of most significant samples. In the developed approach, the Chi-square method performed in reference to the information-theoretic feature selection paradigm, where it measured the intuition that the best terms $t_k$for a class c_i are the one distributed amongst the set of positive and negative examples of the class c_i. Chi-square's test can be derived as per Eq. (7).

$$Chi - Square(t_k, c_i) = \frac{N(AD-CB)^2}{(A+C)(B+D)(C+D)} \tag{7}$$

In Eq. (7), N represents the total samples or features available in the input feature vector (for the considered text corpus), and A be the elements in class c_i with t_k. The feature elements carrying t_k in other classes is defined as B. The number of features in class c_i that don't possess any term of t_k

is given by $C.D$ on the other hand, represents the number of feature elements with no term t_k in other classes. Thus, by using above Eq. (7), the proposed method assigned a score value for each feature towards a class (i.e., PII or Non-PII). Eventually, the scores are amalgamated to yield a final score, given as Eq. (8).

$$max\big(Chi - Square(t_k, c_i)\big) \tag{8}$$

Thus, with the obtained score as shown in Eq. (8), it retained the top-K features (or samples) to perform further processing.

2) Wilcoxon Rank Sum Test: It is also called as the significant predictor test which belongs to a type of non-parametric test with independent samples. Functionally, it measures the level of significance of each embedding metrics value and allied probability towards PII or Non-PII query. Here, WRST compares the location of the two samples by using two matched samples. More specifically, Signed Rank Sum test approach was applied to measure the level of significance. Consider, the paired data contains the samples $(X_1, Y_1), \dots, (X_n, Y_n)$, where each sample signifies a pair of measurements. These measurements are converted to the real-numbers, while the paired sample test is converted to the one-sample test by replacing each pair of numbers (X_i, Y_i) by respective difference $X_1 - Y_1$. This method enables ranking of the difference in between the pairs. However, it needs that the data remains on an ordered metric scale. Consider the data for one-sample test be $X_1, \dots, X_n$, then assuming that the samples carry different absolute values, and no sample is equal to zero, then it applies the following steps to examine the level of significance of each data in the sample.

- ✓ Estimate the value $|X_1|, \dots, |X_n|$
- ✓ Sort $|X_1|, \dots, |X_n|$ and use the sorted-list to assign ranks $R_1, \dots, R_n$, where the rank of the smallest measurement remains unit value (i.e., 1). And the rank of the subsequent smallest becomes two and so on.
- ✓ Consider sgn be the sign-function stating $sgn(x) = 1$, if $x > 0$ and $sgn(x) = -1$ if $x < 0$. It provides signed rank sum value T by using following formulation [see Eq. (9)].

$$T = \sum_{i=1}^{N} sgn\,(X_i) R_i \tag{9}$$

- ✓ Retrieve a $p - value$ by estimating T to its distribution under the null-hypothesis.

Here, the ranks can be assigned in such manner that the value R_i remains the number of j for which $|X_j| \leq |X_i|$. Additionally, for $\sigma: \{1, \dots, n\} \rightarrow |X_{\sigma(1)}| \leq |X_{\sigma(n)}|$, then $R_{\sigma(1)} = i$ for all element i. With the obtained $p - value$ for each sample, the level of significance is measured and each sample was annotated as significant or insignificant. Thus, applying this approach, the significant feature elements were retained for further computation.

3) Gini Score based significant feature selection (GSFS): Gini Score measures the level of impurity in a data by applying a function as shown in Eq. (10).

$$m(s) = \sum_{i \neq j} \widehat{P_{si}}\, \widehat{P_{sj}} = 1 - \sum_j \widehat{P_{sj}} \tag{10}$$

This method generalizes the variance impurity, signifying the variance of a distribution belonging to the two class labels i and j. GSFS is also referred as the expected error rate when the class label is selected randomly from the input feature distribution (here, CBOW embedding metrics). The impurity criterion applied to be highly peaked at the same likelihood in comparison to the entropy-based approaches that makes Ginni Index based feature selection suitable for our targeted PII detection and classification problem. In this work, the Eq. (11) was applied to measure the likelihood of a feature (or sample) to be retained for further computation.

$$GSFR(S) = 1 - \sum_{i=1,\dots,m} P_i^2 \tag{11}$$

In Eq. (11), P_i states the likelihood that a tuple in feature set S belongs to the class C_i. Here, we obtained the value of P_i as per $|C_i, S|/|S|$.

Once performing aforesaid feature selection methods over the input embedding metrics, the retained features were horizontally concatenated to yield a composite or fused (say, ensemble) feature vector for further learning and (PII) prediction.

E. Resampling

As discussed in the previous section, in real-world scenario or dataset, the frequency of PII attributes over a large text input can be smaller. In other words, the number of non-PII elements can be higher than the PII elements, signifying class-imbalance problem. The similar problem can be unavoidable in at hand PII classification problem as well. Since, in this work, the total samples comprised a massive 12000 queries, containing 4000 samples from each discipline (i.e., personal credential, health credential and financial credential). However, the number of queries or sentences (say, review) containing PII element were almost 1100, which is smaller than even 10% of the original data size, which is 12000. Though, the data combination is prepared intentionally, emulating or representing the real-time data condition where the PII elements can be of minority class than the non-PII elements (signifying majority class). This data imbalance problem can skew the learning process and hence can yield inaccurate performance. To alleviate this problem, resampling method can be applied which can make effort to retain sufficient balance between the samples belonging to the PII data as well as non-PII queries. Though, in the past, different classical approaches like up-sampling, random sampling methods have been applied; however, such approaches might often result iterative hot-spot problem (i.e., iterative class-imbalance). Noticeably, the classical up-sampling method intends to increase the minority samples to leverage the imbalance, while in down-sampling the majority class samples are reduced. On the other hand, in random sampling technique, the additional samples are appended by selecting samples from the original feature space (i.e., original minority samples). An inappropriate selection of samples might give rise to the aforesaid hot-spot or imbalance data conditions, iteratively. To address this problem, in this work three different improved resampling methods are applied. More specifically, we have applied three different variants of the synthetic minority over-sampling method (SMOTE), named SMOTE, SMOTE-

Boundary line (SMOTE-BL), and SMOTE with Edited Nearest Neighbor (SMOTE-ENN). Unlike traditional resampling techniques, SMOTE method generates synthetic samples depicting most correlated features, without impacting original sample distribution or allied (sample) ratio. Though, in exceedingly high non-linear data or feature space, SMOTE can yield better performance than the classical random sampling or up-sampling methods; however, it undergoes a scenario where there can be the multiple data instances belonging to both minority as well as majority class (it can happen due to feature relatedness or ambiguity). The classical SMOTE doesn't address this problem. Though, it has been solved by using SMOTE-BL. Considering the efficacy of SMOTE-BL method over classical SMOTE, we applied Python Imbalance Data library's SMOTE-BL method to resample the selected features (by using, Select K-Best, WRST and GSFS, distinctly).

In addition to the above discussed SMOTE and SMOTE-BL resampling method, we applied SMOTE-ENN as well that used minority samples as input to generate the synthetic samples. The generated samples were subsequently processed with k-Nearest Neighbor (k-NN) classifier, where Euclidean distance-based k-NN algorithm was taken into consideration. This approach formed a vector between the current samples and the one from obtained k-neighbors. The estimated vector was multiplied with a random number existing in between 0 to 1, which was appended to the original sample to obtain the final synthetic sample as output. Unlike classical SMOTE resampling method where defining class-boundaries can be challenging as there can be certain synthetic minority samples undergoing cross-over or overlap with the majority class. The severity of such events can be more frequent over the large non-linear features. This problem can mis-label the synthetic samples inappropriately, while training over such data might impact overall classification results or can yield false positive or false negative results. To address this problem, we applied an improved SMOTE algorithm called SMOTE-ENN. It possesses an additional computing ability in conjunction with the Edited Nearest Neighbor (ENN) classifier in which the label of each synthetic sample is compared in reference to the vote of its k-NNs neighbors. If it finds any inconsistency between the input sample and corresponding k-NNs, it drops that sample from the synthetic sample set, else it retains the same. Here, the higher $k - value$ enables stringent cleaning and therefore appends original data with the optimal set of synthetic samples to alleviate class-imbalance problem. It also provided consistent set of input features for further learning and classification.

F. Min-Max Normalization

In sync with huge non-linear features, to alleviate any probable over-fitting and convergence problem, we performed feature normalization by using Min-Max Normalization method. The proposed Min-Max normalization method mapped input features in the range of [0 1]. Noticeably, Min-Max normalization was done over each feature set retained after feature selection. Mathematically, we applied equation (12) to perform Min-Max normalization over the input features. As depicted in (12), the variable x_i be the feature element, where $x_i \in N$, which is mapped to the corresponding normalized value signifying x_i'. Here, x_i' is obtained in the range of 0 to 1. In (12), the parameters $min(X)$ and $max(X)$ states the minimum and the maximum values of X, respectively.

$$Norm(x_i) = x_i' = \frac{x_i - min(X)}{max(X) - min(X)} \quad (12)$$

G. Ensemble of Ensemble Classification

Unlike traditional ML-based solutions where the authors have applied merely single or standalone ML algorithm to perform learning and classification, we have designed a robust ensemble of ensemble (E2E) learning method that applies five different ensemble learning methods including Bagging, Boosting, AdaBoost, Random Forest and Extra Tree Classifier algorithms. In the developed E2E model, each classifier performs learning over the normalized features, and provides label for each sentence as 1 and 0, for PII and Non-PII element respectively. Finally, applying the concept of the maximum voting ensemble (MVE), it labels each input query as PII or Non-PII. Noticeably, unlike traditional machine learning methods like decision tree (DT), support vector machine (SVM), naïve bayes (NB), etc. the ensemble learning methods like random forest (RF), AdaBoost, extra tree classifier, XGBoost etc. have performed better for the different data and image classification problems. Being consensus driven prediction solutions, it yields higher accuracy and reliability towards PII detection and classification. Considering it as motivation, this research intends to use only ensemble learning methods, such as the Bagging, Boosting, RF, ETC, XGBoost to perform maximum voting ensemble-based learning and classification. In this method, a total of five base classifiers (here, ensemble classifiers) are considered. This is because, out of five base classifiers, if three of the base classifiers classifies an input as PII and annotates it as 1, it would be classified as PII. On the other hand, if three base classifiers annotate an input as 0, the final prediction is made as non-PII. Being consensus driven approach or maximum voting-based approach, it can be more reliable towards run-time significances. A brief of the different ensemble base classifiers used to design E2E ensemble is given as follows:

1) AdaBoost: It represents a kind of adaptive boosting ensemble learning method that possesses superior instance-wise analysis ability. In this work, to implement AdaBoost method, the allied prerequisite tests were doled-out to the equivalent weight that enables generation or forming certain weak learners. For each cycle of computation, it measures the error rate for the aforesaid weak classifier and eventually the weight for the accurately classified sample is improved, while the weights for the inaccurately classified samples are reduced. Finally, the weak learner becomes the strong learner and classifies each sentence as PII or PII-Free and labels them as "1" and "0", respectively. The gradient boosting method is an improved Boosting ensemble in which the weights for the accurately classified samples are varied gradually, rather with a fixed update. It solves convergence problem.

2) Bagging: In this work, we applied bagging ensemble learning method with two different kernels. More specifically, we applied k-NN algorithm and MNB classifier as the two base classifiers to design Bagging ensemble learning algorithms.

3) Random Forest (RF): It is one of the most efficient ensemble-learning methods that embodies multiple tree-based classifiers. Being a tree-based learning model, each tree generates its own predicted output or vote for the most probable class (for each sentence or allied embedding feature vector). Consider that the input training samples be N, then a sample possessing N cases are selected arbitrarily from the input data (say, input feature vector). The selected samples are subsequently used as training set to constitute a new tree. Let, there be the M input variables, then the best split is applied over m to split the node. Noticeably, in this work the value of m was fixed as constant while performing forest development. Thus, the proposed method develops each tree to the possible largest level or size. Noticeably, unlike other machine learning algorithms, RF requires relatively smaller number of parameters to be calculated during learning and classification and hence serves a lightweight and computationally efficient classifier (solution). In sync with above discussed forest growth process, the RF algorithm can be defined as the amalgamation of the different tree-structures, given as Eq. (13).

$$\{h(x, \theta_k), k = 1,2, \dots i \dots\} \quad (13)$$

In Eq. (13), h represents the classifier function, while the arbitrary vector generated unformal across tress is given by $\{\theta_k\}$. Here, each tree embodies a unit vote for its most probable class towards a unit query or sentence (say, input x). To be noted, the ability to employ multiple DTs where each DT performs respective classification makes proposed RF method more suitable for learning and classification. We applied a bootstrapped subset of training samples to train each tree across the formed forest that uses 70% of the training data, while the remaining data elements are labeled as the out-of-bag samples, which are subsequently applied for inner cross-validation to predict output. Thus, applying this method, we classified each input query or sentence (say, corresponding embedded feature vector) as PII or PII-Free, which were subsequently labelled as "1" or "0", correspondingly.

4) Extra Tree Classifier (ETC): This algorithm forms a cluster of unpruned DTs as per the conventional top-down paradigm. Unlike RF method, ETC algorithm comprises randomization of attribute and cut-point selection when performing node-split. It is also able to form overall randomized trees containing structures which are independent of the outputs of the training sample. In fact, it distinguishes itself from the other tree-based ensemble algorithms because of the two key factors. The first is that it splits nodes by choosing cut-points arbitrarily and the second it applies the complete training sample to for forest (say, tree growth). The predicted output from the encompassing trees is amalgamated to yield final PII prediction result by using MVE method. The key approach behind the ETC method is that the complete randomization of the cut-point and attribute altogether with ensemble averaging that minimizes the variance in more efficient way than the weaker randomization methods applied in other machine learning approaches. Additionally, the use of the original training samples minimizes the probability of bias and therefore accomplishes higher accuracy for classification. Applying this ensemble approach each sentence was classified as PII and PII-Free, which was labelled as "1" and "0", respectively.

Once obtaining labelled output from each base (ensemble) classifier, for each query or sentence, the proposed MVE method estimated the consensus, signifying the maximum vote for each sentence (i.e., PII or Non-PII). As the proposed model applied five different ensemble ML classifier as base classifier, a query with minimum three 1 was classified as PII and the one with three 0 was annotated as the non-PII query. In this manner, the proposed model applied consensus (say, Maximum Voting) score to predict each query as PII or non-PII.

The use of this approach can help identifying a set of those queries having PII elements, which can not only enable further PII element detection and identification, but can also make computation fast due to reduced search space (or queries). Once identifying a query as PII, dictionary-based technique(s) can be applied to perform specific PII attribute detection from each sentence or query. Though, this research considers it a scope for future efforts, and hence is beyond the research scope. Simulation results and allied inferences are given in the subsequent section.

VI. Results and Discussion

In this work, the key emphasis is made on developing a robust cross-discipline PII detection and classification system for privacy preserved digital data transmission and storage. In this work, the PII detection and classification task is solved as an NLP problem, where it exploits semantic features from the input queries to learn and predict PII elements in each query. The overall proposed model emphasizes on addressing almost major challenges in at hand PII detection problem. In other words, we made effort to address both data challenges (feature engineering) as well as classification so as to ensure a reliable and optimized solution. In this reference, this work contributed a robust ensemble of ensemble learning assisted semantic feature driven cross-discipline PII detection and de-identification model (EESD-PII). Being a machine learning driven NLP solution, we have tried to exploit PII attributes from the different disciplines, which provide sufficiently large feature space to learn and hence predict PII queries in real-world application. Moreover, realizing class-imbalance, convergence, local minim and over-fitting problems, different computational optimization measures were taken into consideration that ensured optimal data (or feature) as well as computational component to achieve a robust cross-discipline PII detection and classification solution. In sync with the real-world data conditions where the likelihood of data heterogeneity, unstructured-ness ss etc. can' be ruled out, we at first performed pre-processing over the input cross-discipline datasets. Noticeably, we synthesized a total of 12000 text queries containing PII attributes from three different categories including personal credentials ("Name", "Address", "Email", "Aadhar Number", "PAN", "SSN", "Vehicle Number", and "Phone Number"), healthcare credentials ("Health Insurance Number", "Diagnosis Details", and "Life Insurance Number") and financial credentials ("Bank Account Number", "Loan Account", "Balance Details", "OTP", "Access PIN" and "Credit

Card Number"). Here, the key motive was to strengthen the proposed PII detection model to detect and classify each input text query irrespective of its subject matter or allied discipline. To address aforesaid problem of data heterogeneity and unstructured-ness different pre-processing tasks were performed including missing value removal, Unicode normalization, removal or emoji, website link, punctuations, stop-words, and lower-case conversion. Additionally, lemmatization and tokenization were performed, where the earlier one helped in retaining the root-intends or contextual information, while the use of tokenization helped tokenizing each query to exploit corresponding semantic features for learning and prediction. The tokenized outputs were passed to the CBOW feature extraction that obtained semantically enriched feature vector to perform learning and prediction. Noticeably, we applied CBOW as Word2Vec embedding method to retain maximum possible contextual feature even in low-dimensional feature which helped retaining sufficiently large contextual feature vector even with the low computational cost. Subsequently, CBOW feature was processed for feature selection, where the different algorithms including Chi-Square based Select K-Best (SKB) method, WRST algorithm and Ginni Score based feature selection (GSFS) methods. Unlike traditional feature selection methods, we designed an ensemble feature model by concatenating the selected features from the SKB, WRST and GSFS methods. Thus, the combined feature vector was later applied for feature resampling. In this work, we applied three different resampling methods including SMOTE, SMOTE-BL and SMOTE-ENN. Noticeably, in the proposed model, we executed these three resampling methods individually over the selected features. It helped addressing class-imbalance problem. Subsequently, the resampled features were then processed for Min-Max normalization that mapped each input in the range of 0 to 1 and thus alleviated any likelihood of convergence and over-fitting problem. The normalized outputs were passed to the E2E classifier which was designed by using five different ensemble learning methods named Bagging, Boosting, AdaBoost, Random Forest, and Extra Tree Classifier. We applied consensus oriented MVE method to perform learning and classification.

The proposed model was developed using Python Notebook, and the simulation was done on Google Co-laboratory platform, which helped reducing implementation complexity and cost. The simulation was done on a central processing unit configured with Microsoft Window operating system operating with 8 GB RAM and Intel i5 processor. To quantify the performance of the proposed PII detection and classification system, we retrieved confusion metrics in terms of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). These parameters were used to measure performance in terms of accuracy, precision, recall and F-Measure, which were obtained by using following equations (Table II).

TABLE II. PERFORMANCE PARAMETERS

Parameter	Mathematical Expression
Accuracy	$\frac{(TN+TP)}{(TN+FN+FP+TP)}$
Precision	$\frac{TP}{(TP+FP)}$
Recall	$\frac{TP}{(TP+FN)}$
F-measure	$2.\frac{Recall.Precision}{Recall+Precision}$

Observing the implementation schematic (Fig. 1), it can be found that the overall proposed model encompasses the different feature selection, resampling and classification models; though, the proposed model defines the term "proposed" with the CBOW feature extraction, ensemble feature selection, resampling, Min-Max normalization and E2E MVE-based classification. It signifies that the proposed model encompasses sub-algorithms for the different phases, including feature selection, resampling and classification. Considering this fact, we at first examined the performance with the different feature selection model, feature resampling and classifiers. Subsequently, with the best performing model the comparison was done with the existing state-of-arts. In this reference, the overall performance characterization is done in terms of intra-model characterization and inter-model characterization. The detailed discussion of the results and allied inferences is given in the subsequent sections.

A. Intra-Model Characterization

In this section, the performance comparison with the different feature selection methods, feature resampling and classifiers is given. Noticeably, in this work we applied single Word2Vec CBOW feature extraction method, which is subsequently processed for the different feature selection, feature resampling, and ensemble classifiers. In this reference, this research performs relative performance with the different algorithms. The key purpose of this assessment is to identify the best performing set of algorithms towards targeted PII attribute detection and classification solution. The detailed discussion of the results obtained and allied inferences is given as follows:

In the proposed model, the CBOW features are fed as input to the three different feature selection methods, including SKB, WRST and GSFS method. Table III presents the simulated performance outcomes with the different feature selection methods. The outputs obtained reveal that with the CBOW input feature, SKB feature selection model exhibits the accuracy of 94.91%, while WRST and GSFS method yielded PII prediction accuracy of 95.03% and 94.89%, respectively. Similarly, the F-Measure outputs were measured as 94.93%, 95.08% and 95.10%, for SKB, WRST and GSFS feature selection methods, correspondingly. The simulation results signify that the intrinsic features driven approaches whether WRST and GSFS methods yield better performance towards PII prediction. The precision performance by SKB, WRST and GSFS methods were obtained as 94.86%, 94.97% and 95.21%. It clearly indicates that the WRST method or GSFS can be suitable towards at hand text-review driven PII prediction solution. Noticeably, these results (Table III to Table V) have been obtained by using proposed E2E MVE ensemble learning classifier.

TABLE III. PERFORMANCE WITH THE DIFFERENT FEATURE SELECTION METHODS

Feature	Feature Selection	Performance (%)			
		Accuracy	Precision	Recall	F-Measure
CBOW	SKB	94.91	94.86	95.01	94.93
	WRST	95.03	94.97	94.82	95.08
	GSFS	94.89	95.21	95.00	95.10

Table IV presents the efficiency with the different resampling methods. More specifically, three different resampling methods were applied towards targeted PII detection and prediction model. The simulation results reveals that the CBOW features with SMOTE, SMOTE-BL and SMOTE-ENN methods exhibited the accuracy of 97.77%, 97.93% and 99.35%, respectively. The simulation results also reveal that the SMOTE, SMOTE-BL and SMOTE-ENN methods achieves precision of 99.63%, 99.81% and 99.88%, respectively. These algorithms exhibit the recall of 99.37%, 99.63% and 99.94%, correspondingly. The depth assessment also exhibits that the F-Measure performance by SMOTE, SMOTE-BL and SMOTE-ENN resampling method exhibited the F-Measure of 99.49%, 99.71% and 99.90%, respectively. The simulation clearly depicts that the proposed SMOTE-ENN method exhibits the superior performance than the other approaches. Recalling the theoretical aspects where SMOTE-ENN methods possess superior efficacy than the classical random sampling, up-sampling and even classical SMOTE variants. The results obtained confirm efficacy of the proposed SMOTE-ENN model towards at hand PII detection and prediction model.

TABLE IV. PERFORMANCE WITH THE DIFFERENT FEATURE RESAMPLING METHODS

Feature	Feature Resampling	Performance (%)			
		Accuracy	Precision	Recall	F-Measure
CBOW	SMOTE	97.77	99.63	99.37	99.49
	SMOTE-BL	97.93	99.81	99.63	99.71
	SMOTE-ENN	99.35	99.88	99.94	99.90

The simulation results with the different ensemble learning classifiers including the proposed E2E MVE ensemble model are given in Table V. Here, the key motive is to assess relative efficacy of the different base classifiers and assess whether the proposed E2E ensemble with MVE concept can yield superior performance or not. In this reference, the accuracy, precision, recall and F-Measure performance obtained are given in Table V. As depicted in Table V, Bagging ensemble method exhibits the accuracy of 97.89%, while Boosting, AdaBoost, Random Forest and Extra Tree Classifier algorithms exhibit the accuracy of 96.68%, 96.72%, 98.67% and 98.81%, respectively. On the contrary, the proposed E2E MVE ensemble method which applied aforesaid ensemble learning methods as base classifier achieved the PII prediction of 99.77%. The precision performance by Bagging, Boosting, AdaBoost, Random Forest and Extra Tree Classifier algorithms was measured as 94.37%, 96.61%, 94.99%, 94.99% and 96.01%, respectively. The precision performance by the proposed E2E MVE method exhibited the recall of 99.63%, which is higher than other standalone ensemble learning classifiers. The F-Measure performance by Bagging, Boosting, AdaBoost, Random Forest and Extra Tree Classifier algorithms were obtained as 95.45%, 96.69%, 95.88%, and 96.065, respectively. The proposed E2E MVE method exhibited the F-Measure of 99.71%, which is superior than any other approaches.

TABLE V. PERFORMANCE WITH THE DIFFERENT CLASSIFIERS

Feature	Feature Resampling	Performance (%)			
		Accuracy	Precision	Recall	F-Measure
CBOW	Bagging	97.89	94.37	96.57	95.45
	Boosting	96.68	96.61	96.69	96.69
	AdaBoost	96.72	94.99	96.80	95.88
	Random Forest	98.67	94.99	96.80	95.88
	Extra Tree Classifier	98.81	96.01	96.13	96.06
	Proposed E2E	**99.77**	**99.81**	**99.63**	**99.71**

Taking into consideration of the overall results and allied inferences, it can be quantified that the proposed model encompassing CBOW semantic features, WRST feature selection, SMOTE-ENN resampling, and E2E MVE ensemble method exhibits the accuracy of 99.77%, precision 99.81%, recall 99.63% and F-Measure of 99.71%. Here onwards we call this method as the proposed model and thus with this specific performance, we have performed the relative performance comparison with other state of arts, which is given in the subsequent sections.

B. Inter-Model Characterization

In sync with the results obtained and allied inferences, we define proposed model as the PII detection solution encompassing CBOW feature extraction, WRST feature selection, SMOTE-ENN resampling, Min-Max normalization and E2E MVE ensemble classifier. Here onwards, we define accuracy, precision, recall and F-Measure performance of the proposed model as 99.77%, 99.81%, 99.63% and 99.71%, respectively. This section discusses the relative performance with the other state-of-arts, as given in Table VI.

This is the matter of fact that in the very few efforts have been made for PII detection, where the different algorithms have been applied for feature extraction, and classification (say, learning). A few approaches have used deep learning as well to learn over the local deep features pertaining to each PII related query to perform (PII) detection and classification. However, their inability to address class-imbalance problems, convergence and local minima etc. make then inferior towards a robust and optimal PII detection and classification solution. Moreover, the different existing approaches have applied the different datasets

and hence generalizing solutions for the same data becomes difficult, especially when the data availability is a challenge. Despite this we have compared the efficacy of our proposed model with other state of arts. Noticeably, in sync with the above discussions (i.e., Inter-Model Assessment), we have identified CBOW feature extraction followed by CCRA feature selection, SMOTE-ENN resampling, Min-Max normalization and ETC classifier as the optimal solution towards the targeted PII detection and classification system. And therefore, respective performance (say, accuracy 99.77%, precision 99.81%, recall 99.63% and F-Measure of 99.71%) is considered as the proposed method towards targeted PII detection and classification system. In this reference, a comparison between the proposed method and other state-of-arts is given in Table II. Noticeably, in the past the different state-of-arts have applied the different performance variables for efficacy analysis, we considered a common performance variable F-Measure to assess relative efficiency. Furthermore, the considered state-of-arts represent the different approaches towards NER (Named Entity Recognition) analysis that possess similar intends as that of PII detection.

TABLE VI. INTER-MODEL PERFORMANCE COMPARISON

Reference	Method	F-Measure (%)
[51]	Bi-LSTM-CRF +CNN	93.5
[52]	Bi-LSTM-CRF+ELM (Extreme Learning Machine) + BERT+ Flair	93.38
[53]	Bi-LSTM + Flair embeddings	93.09
[54]	BERT+ BI-LSTM	92.80
	BERT Base	92.4
[55]	Bi-LSTM + multi-task	92.61
[56]	Bi-LSTM + BERT + ANN-LM	92.22
[57]	Word-Embedding + ANN-LM	91.93
[58]	Bi-LSTM + CRF +Auto-encoders+ Lexical Features	91.89
[59]	Bi-LSTM + CRF + Lexical Features	91.73
[60]	LM- LSTM +CRF	91.71
[61]	Bi-LSTM+ CRF	91.62
[62]	Hybrid Semi-Markov +CRF	91.38
[63]	IXA-Pipes	91.36
[64]	CCNN+ WLSTM+ CRF	91.35
[65]	CRF +GRU	91.26
[66]	BI-LSTM +CRF	90.94
[67]	Glove	88.30
[68]	CBOW	88.20
[69]	Bi-LSTM + Glove + Flair	85.51
Proposed	**CBOW + WRST + SMOTE-ENN + Min-Max Normalization + E2E MVE**	**99.71**

In sync with the results obtained (Table VI), it can be found that the major at hand efforts towards NER or applied PII detection methods have exploited deep learning methods with the different feature embedding techniques such as CBOW [68], Glove [67], Flair [69]. Though, the major state-of-arts have directly applied varied deep learning methods such as LSTM, Bi-LSTM, CNN etc., to perform two-class classification by using Softmax classifier. However, most of these methods fail in addressing contextual details within the input text corpus. Moreover, none of these approaches could address inherent challenges of NLP including unstructured data, class-imbalance, local minima and convergence. Such limitations might make any solution suspicious, especially when the solution is supposed to be applied over real-world BigData. Unlike aforesaid approaches, in this work, we emphasized on addressing above stated issues so as to contribute a generalizable optimal PII detection and classification system. Moreover, a few approaches like [56][57] tried to extract local features from the sequential text embeddings for further learning and classification by using machine learning methods (i.e., ANN-LM); however, the highest F-Measure observed was 92.2%, which is almost 7% lower than the proposed method. Here, the impact of feature engineering such as class-imbalance (SMOTE-ENN), feature selection (WRST) and normalization can't be ruled out. Noticeably, these feature improvement methods helped our proposed model to achieve features with minimum redundant data or non-redundant feature elements. It helped in alleviating pre-mature convergence. SMOTE-ENN helped in addressing class-imbalance problem, while the use of Min-Max normalization resolved over-fitting problem. Thus, the improvement in data or allied feature enabled our proposed method to exhibit superior over other state-of-arts. The overall results confirm superiority of the proposed method over other state-of-arts available in PII detection domain. The depth assessment affirms positive response or answers for the research questions (RQ1-RQ4), as discussed in Section IV.

VII. CONCLUSION

Unlike existing PII detection approaches where the authors have either applied predefined dictionaries to detect and classify (PII) text for a specific dataset or discipline, in this work a robust multi-disciplinary PII detection method was developed. To achieve it, the PII detection, de-identification and re-identification tasks were solved as an NLP problem. In this reference, unlike predefined dictionary or syntactical learning-based solutions, the texts encompassing PII and normal queries related to the different disciplines including personal credential, healthcare, business communication details etc. were processed for NLP processing that eventually enabled a cross-discipline PII detection system. Additionally, this research made effort to alleviate at hand computational limitations including class-imbalance, local minima and convergence and over-fitting. More specifically, novel and robust ensemble of ensemble learning assisted semantic feature driven cross-discipline PII detection and de-identification model (EESD-PII) was developed in this work. In sync with cross-discipline PII detection and classification task, a large set of text queries

possessing PII elements belonging to the personal credentials, healthcare data, financial texts and queries etc. were used for training. The collected cross-discipline texts were processed for pre-processing tasks like the removal of stopping-words, punctuations, URL-link, lower case conversion, lemmatization and tokenization. The tokenized text-sequences were processed for continuous bag-of-word (CBOW) word-embedding that provided semantic feature space for further learning. The use of CBOW embedding provided contextually-rich feature vector to ensure better learning and PII detection (and/or classification). Realizing real-world scenarios where the frequency of PII terms is very small in comparison to the non-PII elements (signifying severe class-imbalance problem), SMOTE resampling methods including SMOTE, SMOTE-BL and SMOTE-ENN were applied. This approach helped in alleviating the class-imbalance problem and hence curse of dimensionality, thus helping the model to achieve higher learning accuracy. Subsequently, Wilcoxon Rank Sum Test (WRST) method was applied to retain the most representative samples in reference to the 95% confidence interval. The retained features were processed for Min-Max Normalization which helped alleviating any over-fitting and convergence problems. Finally, a robust ensemble of ensemble (E2E) learning classifiers was designed by using five different ML algorithms encompassing Bagging, Boosting, AdaBoost, Random Forest and Extra Tree Classifier as base classifier. This approach not only helped in achieving more reliable outcome but also alleviated performance diversity problem. Noticeably, the use of tokenization and word-embedding helped achieving de-identification goals as well, without losing data essence for further learning and classification. The proposed E2E model performed majority voting ensemble to annotate each text-query as PII or Non-PII data. The simulation results reveal that the proposed EESD-PII model achieves PII annotation accuracy of 99.77%, precision 99.81%, recall 99.63% and F-Measure of 99.71%. In future, the authors can design AI driven multi-disciplinary dictionary with PII-sensitive masking approach to improve de-identification; though, the proposed model retains aspects of de-identification and re-identification to serve real-time decisions.

VIII. Future Scope

This is the matter of fact that the proposed method has yielded superior and generalizable performance towards PII detection and classification; however, it was mainly based on exploiting semantic features trained over ensemble machine learning classifiers. The use of deep features can also be explored. The recent development in advanced deep structures, especially designed to address long-term dependency problems like multi-attention-based Bi-LSTM, Bi-GRU, residual networks etc. can be applied as ensemble feature structure. It can help to exploit more contextual details with no probability of any gradient vanishing or gradient explode over large deep structure running over non-linear data space. In this reference, the researchers can focus on applying deep (hybrid) feature models for feature extraction and learning to achieve better outputs. The use of GAN can also be considered in future to improve de-identification and word-restructuring.

References

[1] A. Acquisti, L. Brandimarte, G. Loewenstein, “Privacy and human behavior in the age of information”, Science 347(6221), 2015, pp. 509-514.

[2] M. Callahan, “Us dhs handbook for safeguarding sensitive personally identifiable information. Washington, DC, 2012.

[3] Personally Identifiable Information (PII) Guidebook, Personally Identifiable Information Working Group of the Indiana Executive Council on Cybersecurity, January, 2021, pp. 1-33. [Accessed on 27 April 2022]

[4] Y. Pan, B. Stackpole, L. Troell, “Computer forensics technologies for personally identifiable information detection and audits”, ISACA, vol. 02, 2010, pp. 1-7. [Available http://scholarworks.rit.edu/article/999]

[5] P. Kulkarni and N. K. Kauvery, “Personally Identifiable Information (PII) Detection in the Unstructured Large Text Corpus using Natural Language Processing and Unsupervised Learning Technique”, (IJACSA) International Journal of Advanced Comp. Science and Applications, Vol. 12, No. 9, 2021.

[6] X. Feng, Y. Feng and A. Asante, “A Systematic Approach of Impact of GDPR in PII and Privacy”, International Journal of Engineering Science Invention, 10 (1), pp. 05-14.

[7] ICO (2016) “Overview of the General Data Protection Regulation (GDPR)”. https://ico.org.uk/for-organisations/data-protectionreform/overview-of-the-gdpr/ [Accessed 14/3/2021].

[8] A. K. Makhija, “Deep Learning Application – Identifying PII (Personally Identifiable Information) to Protect”, Journal of Accounting, Finance, Economics, and Social Sciences Vol.5, No.3, 2020, pp.49-55.

[9] C. Cadwalladr and E. Graham-Harrison, “Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach", The guardian, 2018, pp. 17-22.

[10] Narayanan, A., & Shmatikov, V. (2010). Myths and fallacies of “personally identifiable information.” Communications of the ACM, 53(6), 24.

[11] Al-Zaben, N., Hassan Onik, M. M., Yang, J., Lee, N.- Y., & Kim, C.-S. (2018). General Data Protection Regulation Complied Blockchain Architecture for Personally Identifiable Information Management. 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE). doi:10.1109/iccecome.2018.8658586.

[12] B. T. Welderufael, S. Jetzabel, and P. Sebastian, “Challenges in Detecting Privacy Revealing Information in Unstructured Text”, pp. 1-8. [Accessed on 23 April 2023].

[13] H. C. Kum, S. Ahalt, “Privacy by design: understanding data access models for secondary data. American Medical Informatics Association (AMIA) Joint Summits on Translation Science and Clinical Research Informatics (2013).

[14] Shah, R., Valera, M.: Survey of sensitive information detection techniques: The need and usefulness of machine learning techniques [Accessed on 27 April 2023].

[15] C. Posey, U. Raja, R. E. Crossler, and A. J. Burns, “Taking stock of organisations’ protection of privacy: categorising and assessing threats to personally identifiable information in the USA”. European Journal of Information Systems, 2017, 26(6), 585–604.

[16] C. Tikkinen-Piri, A. Rohunen, and J. Markkula, “EU General Data Protection Regulation: Changes and implications for personal data collecting companies”, Computer Law & Security Review, 2017, 34(1), pp. 134–153.

[17] A. C. Islam, J. Walsh, R. Greenstadt, “Privacy detective: Detecting private information and collective privacy behaviour in a large social network. In: Proceedings of the 13th Workshop on Privacy in the Electronic Society, 2014. [Accessed on 27 April 2022]

[18] L. Xiao-Bai and Q., “Anonymizing and sharing medical text records”, Information Systems Research, 2017, Vol. 28(2), pp. 332–352.

[19] C. A. Kushida, D. A Nichols, R. Jadrnicek, R. Miller, J. K. Walsh, and K. Griffin, Strategies for de-identification and anonymization of electronic health record data for use in multi-center research studies”, Medical care, 2012, Vol. 50(Suppl): S82.

[20] T. Aura, T. A. Kuhn, and M. Roe, “Scanning electronic documents for personally identifiable information. In Proceedings of the 5th ACM workshop on Privacy in electronic society, 2006, pp. 41–50.

[21] P. Ongsulee, “Artificial intelligence, machine learning and deep learning”, 2017 15th International Conference on ICT and Knowledge Engineering, 2007. doi:10.1109/ictke.2017.8259629.

[22] Arttu Oksanen, J Tuominen, E Mäkelä, M Tamper, Aki Hietanen, and Eero Hyvönen. 2019. Semantic finlex: Transforming, publishing, and using finnish legislation and case law as linked open data on the web. Knowledge of the Law in the Big Data Age, 317:212–228.

[23] Ahmadreza Mosallanezhad, Ghazaleh Beigi, and Huan Liu. 2019. Deep reinforcement learning-based text anonymization against private-attribute inference. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 2360–2369. Association for Computational Linguistics.

[24] Fadi Hassan, Josep Domingo-Ferrer, and Jordi Soria-Comas. 2018. Anonymization of unstructured data via named-entity recognition. In International conference on modelling decisions for artificial intelligence, pages 296–305. Springer.

[25] Filip Gralinski, Krzysztof Jassem, Michał Marci´nczuk, and Paweł Wawrzyniak. 2009. Named entity recognition in machine anonymization. Recent Advances in Intelligent Information Systems, pages 247–260.

[26] Gang Luo, Xiaojing Huang, Chin-Yew Lin, and Zaiqing Nie. 2015. Joint named entity recognition and disambiguation. In Proceedings of the 2015 Conf. on Empirical Methods in Natural Language Proc., page 879–888, USA. Association for Computational Linguistics.

[27] Alexandre Passos, Vineet Kumar, and Andrew McCallum. 2014. Lexicon infused phrase embeddings for named entity resolution. In Proceedings of the Eighteenth Conference on Comp. Natural Language Learning, pages 78–86.

[28] Lev Retinov and Dan Roth. 2009. Design challenges and misconceptions in named entity recognition. In Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL), page 147–155, USA. Association for Computational Linguistics.

[29] Jason PC Chiu and Eric Nichols. 2016. Named entity recognition with bidirectional lstm-cnns. In Transactions of the Association for Computational Linguistics, volume 4, pages 357–370. MIT Press.

[30] Zhenjin Dai, Xutao Wang, Pin Ni, Yuming Li, Gangmin Li, and Xuming Bai. 2019. Named entity recognition using BERT-Bi LSTM-CRF for Chinese electronic health records. In 2019 12th international congress on image and signal processing, biomedical engineering and informatics, pages 1–5. IEEE.

[31] Kathleen C Fraser, Isar Nejadgholi, Berry De Bruijn, Muqun Li, Astha LaPlante, and Khaldoun Zine El Abidine. 2019. Extracting umls concepts from medical text using general and domain-specific deep learning models. arXiv preprint arXiv:1910.01274.

[32] Xin, Y. et al. (2018). Machine learning and deep learning methods for cybersecurity. IEEE Access, 6, 35365-35381.

[33] Tesfay, W. B., Serna, J., & Rannenberg, K. (2019). PrivacyBot: Detecting Privacy Sensitive Information in Unstructured Texts. 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). doi:10.1109/snams.2019.8931855

[34] Jinlan Fu, Pengfei Liu, and Qi Zhang. 2020. Rethinking generalization of neural models: A named entity recognition case study. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 7732–7739.

[35] Apruzzese, G., Colajanni, M., Ferretti, L., Guido, A., & Marchetti, M. (2018). On the effectiveness of machine and deep learning for cyber security. 2018 10th International Conference on Cyber Conflict (CyCon).

[36] Tesfay, W. B., Serna, J., & Rannenberg, K. (2019). PrivacyBot: Detecting Privacy Sensitive Information in Unstructured Texts. 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS). doi:10.1109/snams.2019.8931855.

[37] Wang, Z., Quercia, D., S_eaghdha, D.O.: Reading tweeting minds: Real-time analysis of short text for computational social science. In: Proceedings of the 24th ACM Conference on Hypertext and social media (2013)

[38] Mao, H., Shuai, X., Kapadia, A.: Loose tweets: An analysis of privacy leaks on twitter. In: Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society (2011)

[39] Jindal, P., Gunter, C.A., Roth, D.: Detecting privacy-sensitive events in medical text. In: Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, BCB '14 (2014)

[40] Gill, A.J., Vasalou, A., Papoutsi, C., Joinson, A.N.: Privacy dictionary: a linguistic taxonomy of privacy for content analysis. In: Proceedings of the SIGCHI conf. on human factors in computing sys. 2011, pp. 3227-3236.

[41] Zhang, N.J., Todd, C.: A privacy agent in context-aware ubiquitous computing environments. In: IFIP International Conference on Communications and Multimedia Security. Springer (2006)

[42] Pape, S., Serna-Olvera, J., Tesfay, W.: Why open data may threaten your privacy. In: Workshop on Privacy and Inference, co-located with KI (September 2015)

[43] Y. Liu, H. H. Song, I. Bermudez, A. Mislove, M. Baldi, and A. Tongaonkar, "Identifying personal information in internet traffic," in Proceedings of the 2015 ACM on Conference on Online Social Networks, COSN '15, (New York, NY, USA), pp. 59–70, ACM, 2015.

[44] J. Ren, A. Rao, M. Lindorfer, A. Legout, and D. Choffnes, "Recon: Revealing and controlling pii leaks in mobile network traffic," in Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, MobiSys '16, (New York, NY, USA), pp. 361–374, ACM, 2016.

[45] D. Noever “The Enron Corpus: Where the Email Bodies are Buried?”, arXiv preprint arXiv:2001.10374, 2020.

[46] M.D. Bader, S.J. Mooney, A.G. Rundle, “Protecting personally identifiable information when using online geographic tools for public health research”, Am J Public Health, pp. 206-208, 2016.

[47] A. Alnemari, R.K. Raj, C.J. Romanowski, S. Mishra, “Protecting personally identifiable information (PII) in critical infrastructure data using differential privacy”, In IEEE International Symposium on Technologies for Homeland Security (HST), pp. 1-6, 2019.

[48] A. Majeed, F. Ullah, S. Lee, “Vulnerability-and diversity-aware anonymization of personally identifiable information for improving user privacy and utility of publishing data”, Sensors, vol.17(5), pp.1059, 2017.

[49] J. Venkatanathan, V. Kostakos, E. Karapanos, J. Gonçalves, “Online disclosure of personally identifiable information with strangers: Effects of public and private sharing, Interacting with Comp., vol. 26(6):614-26, 2014.

[50] W.B. Tesfay, J.M. Serna, and S. Pape, “Challenges in Detecting Privacy Revealing Information in Unstructured Text”, In PrivOn@ ISWC, 2016.

[51] Alexei Baevski, Sergey Edunov, Yinhan Liu, Luke Zettlemoyer, and Michael Auli. Clozedriven pretraining of self-attention networks. In 2019 Conference on Empirical Methods in Natural Language Processing, 2019.

[52] Jana Straková, Milan Straka, and Jan Hajic. Neural architectures for nested NER through linearization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5326–5331, Florence, Italy, July 2019. Association for Computational Linguistics.

[53] Alan Akbik, Duncan Blythe, and Roland Vollgraf. Contextual string embeddings for sequence labeling. In COLING 2018: 27th International Conference on Computational Linguistics, pages 1638–1649, 2018.

[54] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In NAACL-HLT 2019: Annual Conference of the North American Chapter of the Association for Computational Linguistics, pages 4171–4186, 2019.

[55] Kevin Clark, Minh-Thang Luong, Christopher D. Manning, and Quoc Le. Semi-supervised sequence modeling with cross-view training. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 1914–1925, Brussels, Belgium, October-November 2018. Association for Computational Linguistics.

[56] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[57] Matthew E. Peters, Waleed Ammar, Chandra Bhagavatula, and Russell Power. Semi supervised sequence tagging with bidirectional language models. CoRR, abs/1705.00108, 2017.

[58] Minghao Wu, Fei Liu, and Trevor Cohn. Evaluating the utility of hand-crafted features in sequence labelling. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 2850–2856, Brussels, Belgium, Oct-Nov 2018. Association for Comp. Linguistics.

[59] Abbas Ghaddar and Philippe Langlais. Robust lexical features for improved neural network named-entity recognition. CoRR, abs/1806.03489, 2018.

[60] Liyuan Liu, Jingbo Shang, Frank F. Xu, Xiang Ren, Huan Gui, Jian Peng, and Jiawei Han. Empower sequence labeling with task-aware neural language model. CoRR, abs/1709.04109, 2017.

[61] Jason P. C. Chiu and Eric Nichols. Named entity recognition with bidirectional lstm-cnns. CoRR, abs/1511.08308, 2015.

[62] Zhixiu Ye and Zhen-Hua Ling. Hybrid semi-Markov CRF for neural sequence labeling. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Vol. 2), pages 235–240, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[63] Rodrigo Agerri and German Rigau. Robust multilingual named entity recognition with shallow semi-supervised features. Artificial Intelligence, 238:63 – 82, 2016.

[64] Jie Yang and Yue Zhang. NCRF++: An open-source neural sequence labeling toolkit. In Proceedings of ACL 2018, System Demonstrations, pages 74–79, Melbourne, Australia, July 2018. Association for Computational Linguistics.

[65] Zhilin Yang, Ruslan Salakhutdinov, and William W. Cohen. Transfer learning for sequence tagging with hierarchical recurrent networks. CoRR, abs/1703.06345, 2017.

[66] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. CoRR, abs/1603.01360, 2016.

[67] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. Glove: Global vectors for word representation. In Empirical Methods in Natural Language Processing (EMNLP), pages 1532–1543, 2014.

[68] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. In Yoshua Bengio and Yann LeCun, editors, 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, Arizona, USA, May 2-4, 2013, Workshop Track Proceedings, 2013.

[69] Carlos Jorge Augusto Pereira da Silva, “Detecting and Protecting Personally Identifiable Information through Machine Learning Techniques”, Faculdade De Engenharia Da Universidade Do Porto, July 27, 2020.

[70] L. Fei, Y. Kang, S. Park, Y. Jang, J. Lee and H. Kim, "KDPII: A New Korean Dialogic Dataset for the Deidentification of Personally Identifiable Information," in IEEE Access, vol. 12, pp. 135626-135641, 2024.

[71] M. Gong, J. Liu, H. Li, Y. Xie and Z. Tang, "Disentangled Representation Learning for Multiple Attributes Preserving Face Deidentification," in IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 1, pp. 244-256, Jan. 2022.

[72] H. Proença, "The UU-Net: Reversible Face De-Identification for Visual Surveillance Video Footage," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 2, pp. 496-509, Feb. 2022.

[73] L. Laishram, J. T. Lee and S. K. Jung, "Face De-Identification Using Face Caricature," in IEEE Access, vol. 12, pp. 19344-19354, 2024.

[74] J. Liu, Z. Zhao, P. Li, G. Min and H. Li, "Enhanced Embedded AutoEncoders: An Attribute-Preserving Face De-Identification Framework," in IEEE Internet of Things Journal, vol. 10, no. 11, pp. 9438-9452, 1 June1, 2023.

[75] B. Shickel, P. J. Tighe, A. Bihorac and P. Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589-1604, Sept. 2018.

Risk Assessment for Geological Exploration Projects Based on the Fuzzy-DEMATEL Method

Zhenhua Yang[1], Hua Shi[2]*, Ning Tian[3], Juan Bai[4], Xiaoyu Han[5]
801 Institute of Hydrogeology and Engineering Geology, Shandong Provincial Bureau of Geology & Mineral Resources, Jinan, Shandong 250014, China[1, 3, 4, 5]
Shandong Communication & Media College, Jinan, Shandong 250200, China[2]

***Abstract*—This paper briefly introduces the analytic hierarchy process (AHP) method and uses the fuzzy decision-making and trial evaluation laboratory (DEMATEL) method to adjust the index weight in it. The geological exploration project of Qingdao undersea tunnel project in Shandong Province was selected as the subject of case study. Firstly, the fuzzy-DEMATEL method was used to analyze the degree of influence between different risk factors in the project and the types of risk factors. Then, the AHP method divided the risk factors and calculated their weight. Finally, the influence parameters calculated by the fuzzy-DEMATEL method was employed to adjust the weight of the indicators in the AHP method. The fuzzy-DEMATEL analysis obtained the driving, conclusion, and transitional risk factors. It was found from the analytic results of the AHP method that the construction supervision unit's qualification risk, management mechanism, and awareness risk had the greatest impact on the risk of the project, and the overall risk level of the project was 2.1 points.**

Keywords*—Geological exploration project; analytic hierarchy process; DEMATEL; fuzzy theory; risk assessment

I. Introduction

With the extensive development of geological exploration projects around the world, its risk management has become the key to ensure the smooth progress and successful implementation of projects [1]. Geological exploration projects are faced with a variety of risk factors due to their technical complexity, long cycle, large investment, and changing environment [2]. Geological exploration projects have a wide range of risk sources, including but not limited to technical risk, market risk, environmental risk, management risk, financial risk, etc. Therefore, in the risk management assessment of geological exploration projects, it is necessary to first identify the source of risk and then set corresponding indicators for investigation and analysis [3]. In order to solve the complexity and uncertainty problems in the risk assessment of geological exploration projects, this paper introduced the fuzzy decision-making and trial evaluation laboratory (DEMATEL) method [4]. The DEMATEL method is a structural modeling method used to visualize the structure of complex causal relationships and can calculate the influence degree of each element in the system on other elements. The fuzzy DEMATEL method introduces the fuzzy mathematics theory and is used to deal with the problems of complex systems with fuzziness and uncertainty [5]. The advantage of the fuzzy DEMATEL method in the risk assessment of geological exploration projects lies in that this method can reveal the mutual influence and causal relationship among various risk factors, which is helpful to identify the key risk factors and potential risk chains. Secondly, by calculating indicators such as the influence degree, influenced degree, centrality, and cause degree of each risk factor, the fuzzy DEMATEL method can provide a quantitative basis for risk assessment, enabling decision-makers to understand the distribution and severity of risks more intuitively [6]. Finally, based on the assessment results, it can also provide targeted risk management suggestions for the project team and help formulate effective risk response strategies. Liu et al. [7] proposed a risk assessment method that combined fuzzy weighted average and fuzzy decision-making trial with evaluation laboratory to sort the failure risks in system failure mode and effects analysis. Mentes et al. [8] proposed an integrated approach to identify and evaluate the driving factors, including the geographical location at the time of the accident and the failure mode leading to the death on the cargo ship. Sangaiah et al. [9] used a hybrid fuzzy multi-criteria decision-making method to effectively identify and rank significant software project risks. The evaluation results showed that compared with the existing software project risk evaluation methods, the fuzzy comprehensive evaluation method was effective and accurate. The above-mentioned related studies have all conducted relevant analyses on how to evaluate risks. Some adopted the method of fuzzy weights to evaluate risks, while others focused on the identification of related factors affecting risks. This paper used the analytic hierarchy process (AHP) to assess the risks of geological exploration projects and utilizes the Fuzzy DEMATEL method to adjust the hierarchical weights in the AHP method, thereby improving the accuracy of risk assessment. This paper briefly introduces the AHP method and uses the fuzzy DEMATEL method to adjust the indicator weights in the AHP method. A case study was performed on the geological exploration project of an undersea tunnel project in Qingdao, Shandong Province. The structure of this article is: abstract - introduction - introduction of AHP and fuzzy DEMATEL method - case analysis - discussion - conclusion.

II. AHP Method and Fuzzy-DEMATEL Method

With the rapid development of economy, in order to adapt to the rapid growth of population, a variety of basic livelihood projects continue to be established. The specifications of these livelihood projects are large or small, but they require a certain area of land, so before the implementation of livelihood projects, it is necessary to carry out geological exploration of the construction area, in order to improve safety [10]. The main

purpose of geological exploration projects is to understand the geological conditions of the construction area. The geological conditions are unknown during the implementation of geological exploration projects [11], which means that there are risks in projects. In addition, geological exploration projects will be affected by various risk factors such as market, management mode, and finance during operation. Therefore, geological exploration projects also need to carry out risk management assessment, so as to reduce the risks of projects.

The AHP method is one of the many methods that can analyze the risk of geological exploration projects. As a multi-criteria decision-making technology [12], the AHP method can divide the problem into small problems at multiple levels and then carry out qualitative and quantitative analyses on the small problems. The steps of the AHP method are as follows. (1) The risk sources of geological exploration projects are divided into different hierarchies, and the risk indicators of each hierarchy are obtained. ② Starting from the lowest risk indicator, the pairwise judgment matrix is constructed for the different indicators of each hierarchy. ③ The indicator weight is calculated according to the pairwise judgment matrix [13], and the consistency check is used to adjust the weight. After the weight of the risk indicators is obtained, the score of each risk indicator can be collected by questionnaire, and the risk level is calculated based on it.

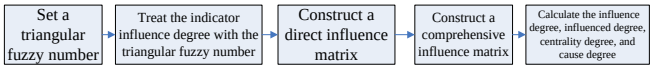


Fig. 1. The flow of the fuzzy-DEMATEL method.

When the AHP method is used, the key is to build a suitable pairwise judgment matrix to calculate the weight of indicators. In this process, the AHP method usually regards each indicator as an independent indicator, but in actual geological exploration projects, there are more or less mutual influences among risk indicators, which will affect the result of risk assessment [14]. Therefore, DEMATEL is adopted in this paper to revise the weights of the AHP method. The function of DEMATEL is to evaluate the influence degree of one indicator on other indicators and then to correct the weights in the AHP method. However, DEMATEL also needs to build an indicator influence matrix when calculating the influence degree of indicators, and the element values in the matrix are also obtained by manual evaluation. Therefore, a triangular fuzzy number [15] is introduced to process the element values in the matrix, thereby minimizing the subjective influence. The steps are shown in Fig. 1.

1) A triangular fuzzy number is set [16], and the expression of triangular fuzzy number A is: $A=(l,m,r)$, where l is the minimum value, m is the most likely value, and r is the maximum value. This paper adopts the form of manual scoring to evaluate the influence degree of a risk indicator on other indicators and sets an A for each influence degree level. The greater the influence degree, the closer A is to 1; otherwise, the closer A is to 0. For example, the A of the evaluation level of "no influence" is set as $(0,0.1,0.3)$.

2) The evaluation level of the influence degree of the indicator given by the manual score is converted into a triangular fuzzy number according to the setting.

3) Direct influence matrix B is constructed, and the matrix is expressed as:

$$B=\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1j} \\ b_{21} & b_{22} & \cdots & b_{2j} \\ \vdots & \vdots & \vdots & \vdots \\ b_{i1} & b_{i2} & \cdots & b_{ij} \end{bmatrix}, \quad (1)$$

where b_{ij} represents the value of the influence of risk indicator i on risk indicator j after fuzzy processing. Its calculation formula is:

$$\begin{cases} b_{ij}=\dfrac{\sum_{k=1}^{k} b_{ij}^{k}}{k} \\ b_{ij}^{k}=\min l_{ij}^{k}+x_{ij}^{k}\Delta_{\min}^{\max} \\ \Delta_{\min}^{\max}=\max r_{ij}^{k}-\min l_{ij}^{k} \\ x_{ij}^{k}=\dfrac{x_{lsij}^{k}(1-x_{lsij}^{k})+x_{rsij}^{k}x_{rsij}^{k}}{1-x_{lsij}^{k}+x_{rsij}^{k}} \\ x_{lsij}^{k}=\dfrac{x_{mij}^{k}}{1+x_{mij}^{k}-x_{lij}^{k}} \\ x_{rsij}^{k}=\dfrac{x_{rij}^{k}}{1+x_{rij}^{k}-x_{mij}^{k}} \end{cases}, \quad (2)$$

where x_{lij}^{k}, x_{mij}^{k}, and x_{rij}^{k} are triangular fuzzy numbers of the influence degree between risk indicators i and j given by rater k after standardization, x_{lsij}^{k} and x_{rsij}^{k} are the left and right standard values of the triangular fuzzy number of the influence between indicators i and j given by rater k, x_{ij}^{k} is the total standard value of the triangular fuzzy number of the influence between indicators i and j given by rater k [17], $\Delta_{\min}^{\max}$ is the difference between the largest maximum value and the smallest minimum value among the triangular fuzzy numbers given by all raters, and b_{ij}^{k} is the clear value between indicators i and j given by rater k [18].

4) The formula of the comprehensive influence matrix is:

$$\begin{cases} C=\dfrac{B}{\max_{1\le i\le n}\sum_{j=1}^{n} b_{ij}} \\ D=C\cdot(E-C)^{-1} \end{cases}, \quad (3)$$

where C is the direct influence matrix after normalization, D is the comprehensive influence matrix, and E is the identity matrix [19].

5) The comprehensive influence parameter of the risk indicators is calculated according to D:

$$\begin{cases} d_i = \sum_{j=1}^{n} d_{ij} \\ e_i = \sum_{j=1}^{n} d_{ji} \\ f_i = d_i + e_i \\ g_i = d_i - e_i \end{cases}, \quad (4)$$

where d_i is the influence degree of risk indicator i, e_i is the influenced degree, f_i is the degree of centrality, and g_i is the degree of cause [20].

After obtaining the comprehensive influence parameter of risk indicators through the above steps, the degree of centrality can be used to adjust the weight of indicators calculated in the AHP method. The adjustment formula is:

$$W_i = \frac{\omega_i f_i}{\sum_{i=1}^{n} \omega_i f_i}, \quad (5)$$

where ω_i is the weight of risk indicator i in the AHP method and W_i is the weight of risk indicator adjusted by the fuzzy-DEMATEL method.

III. Case Analysis

A. Case Overview

This paper took the geological exploration project of an undersea tunnel project in Qingdao, Shandong province as a case analysis. The project is located at Jiaozhou Bay in Qingdao. The main purpose of the undersea tunnel project is to directly connect the economic zones on both sides of Jiaozhou Bay, so as to improve the efficiency of freight and passenger transport and further promote economic development.

In the geological exploration project, the method of drilling was used to survey the seabed geology of Jiaozhou Bay, and a total of 16 drilling points were set up. The ship-borne drilling platform was used at the drilling point. In the process of drilling, the ship-borne drilling platform was fixed to the drilling point by using multiple ship anchors. Then, the drilling casing was driven vertically into the sea bed by traction until the casing reached the stable layer.

B. Methods for Risk Analysis of the Geological Exploration Project

Firstly, 20 experts from the related field were invited to identify and summarize the risk sources of the geological exploration project. The risk hierarchical structure of the geological exploration project was constructed, as shown in Fig. 2.

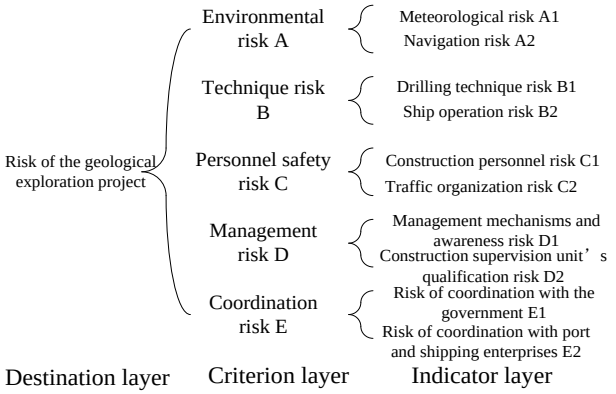


Fig. 2. Hierarchical structure.

After that, 20 experts were distributed the criterion-level rating table and the indicator level rating table. They marked the importance degree of the risk factors in the criterion level and the indicator level through the rating tables. Taking the criterion level score table as an example, as shown in Table I, the five risk factors in the criterion level were compared pairwise, and "1, 2, 3... 8, 9" was used to evaluate the relative importance between two factors. "1" indicates that the two factors are equally important; the larger the value, the more important the former factor is than the latter factor. When the former and latter factors are exchanged during comparison, the reciprocal of the value is used. The same is true for the indicator level score table, but the difference is that the risk indicators of the indicator level for pairwise comparison needs to belong to the same criterion level indicator. According to the rating tables, the judgment matrix of the criterion layer and indicator layer was constructed successively, and the weight of risk factors of the criterion layer and indicator layer was calculated based on it. At the same time, consistency check was used to determine whether the weight is reasonable. If not, the score in the judgment matrix was re-adjusted, and the weight was calculated until passing the consistency check.

TABLE I. Indicator of Importance Rating Table

Risk factor	Factor A	Factor B	Factor C	Factor D	Factor E
Factor A	1				
Factor B		1			
Factor C			1		
Factor D				1	
Factor E					1

After calculating the weight of the risk factors through the above steps, the fuzzy-DEMATEL method was used to calculate the degree of influence between the risks. First of all, the 20 experts were given the risk factor influence degree score table, which compared all risk factors in the indicator layer pairwise. The evaluation of the influence degree from weak to strong was divided into five levels, and a triangular fuzzy number was set for each level. After that, equation (2) was used to deblur the rating table, and the average score of the influence degree was calculated. A direct influence matrix was constructed according to the average score of the influence degree in the score table, and then the centrality degree of each risk factor was calculated following the steps mentioned above.

The weight of the risk factor obtained by the AHP method was adjusted accordingly.

Finally, a questionnaire was designed according to the indicator layer in the hierarchical structure given by the AHP method, and the score of each indicator was set from 0 to 10 according to the risk degree from low to high. Then, the 20 experts scored the risk factor indicators in the questionnaire.

IV. Analysis of Results

This paper used the fuzzy-DEMATEL method to analyze the degree of mutual influence of risk factors in the geological exploration project. After a series of calculations, comprehensive influence matrix D of risk factors in the geological exploration project is shown in Table II. The causal relationship diagram of risk factors drawn according to Table III is shown in Fig. 3. Matrix D reflects the comprehensive influence of a risk factor on other risk factors. A scatter plot of the distribution of risk factors in a plane was obtained by using the centrality degree as the x-axis and cause degree as the y-axis. After introducing the average centrality degree of risk factors, the plane was divided into four quadrants. It can be seen that meteorological risk A1, management mechanism and awareness risk D1, and construction supervision unit's qualification risk D2 were in the second quadrant, belonging to driving risk factors. Drilling technique risk B1, risk of coordination with the government E1, and risk of coordination with port and shipping enterprises were in the third quadrant, belonging to conclusion risk factors. Navigation risk A2, ship operation risk B2, construction personnel risk C1, and traffic organization risk C2 were in the fourth quadrant, belonging to transitional risk factors.

TABLE II. Comprehensive Influence Matrix D of Risk Factors in the Geological Exploration Project

	A1	A2	B1	B2	C1	C2	D1	D2	E1	E2
A1	0.148	0.493	0.308	0.510	0.570	0.504	0.272	0.210	0.257	0.339
A2	0.111	0.274	0.261	0.336	0.353	0.331	0.198	0.164	0.189	0.350
B1	0.086	0.228	0.229	0.360	0.415	0.316	0.203	0.156	0.183	0.227
B2	0.107	0.408	0.340	0.375	0.501	0.477	0.246	0.193	0.230	0.373
C1	0.099	0.264	0.332	0.378	0.392	0.413	0.353	0.204	0.308	0.287
C2	0.115	0.463	0.294	0.514	0.522	0.420	0.262	0.247	0.267	0.424
D1	0.086	0.221	0.206	0.384	0.443	0.317	0.214	0.150	0.175	0.218
D2	0.119	0.330	0.444	0.550	0.617	0.501	0.351	0.231	0.310	0.359
E1	0.064	0.181	0.130	0.190	0.199	0.316	0.117	0.101	0.144	0.155
E2	0.079	0.347	0.169	0.259	0.253	0.363	0.146	0.124	0.141	0.236

TABLE III. Measurement of the Comprehensive Influence Degree of Various Risk Factors in the Geological Exploration Project

	Influence degree	Influenced degree	Centrality degree	Degree of cause	Ranking of centrality degree
A1	3.610	1.014	4.624	2.596	9
A2	2.567	3.209	5.776	0.642	4
B1	2.401	2.713	5.114	0.312	6
B2	3.252	3.855	7.107	0.603	3
C1	3.028	4.265	7.293	1.237	2
C2	3.528	3.958	7.486	0.430	1
D1	2.413	2.361	4.774	0.052	8
D2	3.811	1.779	5.590	2.032	5
E1	1.597	2.203	3.800	0.606	10
E2	2.117	2.967	5.084	0.850	7

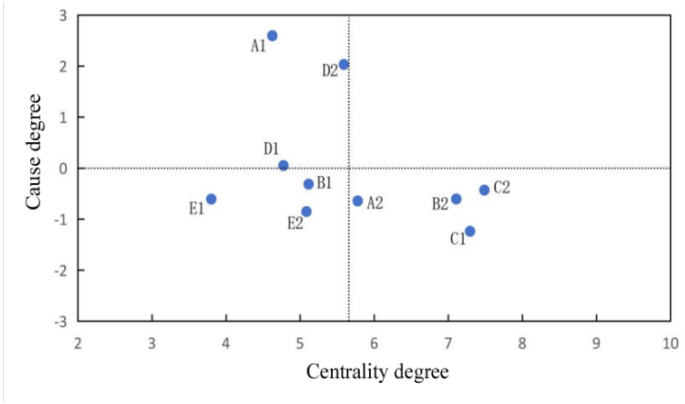


Fig. 3. Causal relationship diagram of risk factors in the geological exploration project.

Then, the AHP method was used to carry out hierarchical analysis on the risk of the geological exploration project, and the weight of risk indicators was calculated and adjusted using the centrality degree. The results and the scores of each risk indicator collected by questionnaire are shown in Table IV. After the adjustment, the weight distribution of risk indicators changed, among which construction supervision unit's qualification risk D2 had the highest weight, management mechanism and awareness risk D1 had the second-highest weight, and risk of coordination with the government E1 had the lowest weight. After combining the adjusted weight with the average score of the corresponding risk indicators, the overall risk score of the geological exploration project risk was 2.1, indicating that the overall risk of the geological exploration project was low.

TABLE IV. THE HIERARCHICAL ANALYSIS STRUCTURE OF THE RISK OF THE GEOLOGICAL EXPLORATION PROJECT AND THE WEIGHT DISTRIBUTION AND AVERAGE SCORES OF INDICATORS BEFORE AND AFTER ADJUSTMENT

Destination layer	Criterion layer	Indicator layer	Initial weight	Centrality degree	Weight after adjustment	Average score
The risk of the geological exploration project	A	A1	0.180	4.624	0.154	1.2
		A2	0.120	5.776	0.128	1.3
	B	B1	0.040	5.114	0.038	1.2
		B2	0.060	7.107	0.079	1.3
	C	C1	0.050	7.293	0.067	1.1
		C2	0.050	7.486	0.069	1.1
	D	D1	0.200	4.774	0.176	3.6
		D2	0.200	5.590	0.206	3.5
	E	E1	0.045	3.800	0.032	1.1
		E2	0.055	5.084	0.052	1.0

V. DISCUSSION

During the construction of infrastructure, it is common to conduct a geological survey of areas where the facility will be constructed in order to ensure the safety of the building facility as well as the construction personnel [21]. Geological survey projects often face many risk factors due to complicated technology, long period, large investment, and changing environment [22]. In order to ensure the safe implementation of geological exploration projects, it is necessary to evaluate and analyze the risk factors involved in these projects. Some studies related to risk assessment are reviewed. Lin et al. [23] analyzed the risk of heavy metal pollution in the Beibu Gulf. The results of the analysis using principal component analysis, positive matrix factor model, and mercury isotope method showed that heavy metal pollution mainly came from industrial pollution sources, including petrochemical, coal combustion, metal and metalloid processing, leather tanning, and human activities, among which anthropogenic pollution sources accounted for approximately 70% of all pollution. Pan [24] established a wind power output power model based on the output characteristics of wind power and established probability models of generating units, lines, and loads considering the uncertainty of other system components to evaluate system operation risks. Based on the "source-sink" landscape theory, Zhao et al. [25] established the location-weighted landscape contrast index and non-point source pollution risk index, in order to study the pollution risk of Baihua Lake in Guiyang City. The evaluation results were compared with the measured water quality data and field investigation results to verify the reliability of this method. This paper used the AHP method to perform qualitative and quantitative analysis of the risk factors in geological exploration projects. Firstly, through the analysis of the risk sources of projects, the hierarchical structure of the risk of projects was constructed, experts were invited to score the importance between two indicators in the hierarchy, and the weight of these indicators was calculated. In addition, in order to reduce the subjectivity brought by the expert score, this paper used the fuzzy-DEMATEL method to calculate the degree of mutual influence between the risk factors and adjusted the weight of the risk factor indicators in the AHP method. Then, a case study was conducted using the geological exploration project of an undersea tunnel project in Qingdao, Shandong province as the subject. The fuzzy-DEMATEL method analyzed the influence degree between the risk factors and the types of risk factors in the project, and then the AHP method divided the risk factors and calculated their weight. Finally, the influence parameters calculated by the fuzzy-DEMATEL method were used to adjust the weight of the indicators in the AHP method, and moreover, the risk level of the project was scored.

Case analysis results showed that meteorological risk A1, management mechanism and awareness risk D1, and construction supervision unit's qualification risk D2 belonged to the driving risk factors, of which A1 and D2 had high cause degree but low centrality degree. Meteorological risk and construction supervision unit's qualification risk were external risks, and D1 could affect the factors in the project, but the influence was small. Drilling technique risk B1, risk of coordination with the government E1, and risk of coordination with port and shipping enterprises E2 belonged to the conclusion risk factors, which had low cause and centrality degrees. These risk factors were the internal risks of the project, which were easy to be affected by other factors, but not easy to influence other factors. Navigation risk A2, ship operation risk B2, construction personnel risk C1, and traffic organization risk C2 belonged to transitional risk factors. These risk factors had low cause degree but high centrality degree, indicating that these internal factors of the project were susceptible to the influence of other factors and not easy to impact other factors.

In the analysis results of the AHP method, the construction supervision unit' qualification risk D2 had the highest weight, followed by management mechanism and awareness risk D1, and risk of coordination with the government had the lowest weight, indicating that the risk caused by the construction supervision unit' qualification, management mechanism and awareness had the greatest impact on the risk of the entire project. Moreover, the risk score of the project was calculated by combining the score given by the experts for the risk indicators with the corresponding weight, and the result was 2.1, suggesting that this project was at a relatively low risk level.

VI. CONCLUSION

This paper briefly introduces the AHP method and uses the fuzzy-DEMATEL method to adjust the indicator weight in the AHP method. A case study was performed using the geological exploration project of an undersea tunnel project in Qingdao, Shandong Province. Firstly, the fuzzy-DEMATEL method was employed to analyze the influence degree between different risk factors and the types of risk factors in the project. Then, the AHP method divided the risk factors and calculated their weight. Eventually, the indicator weight in the AHP method was regulated using the influence parameters calculated by the fuzzy-DEMATEL method. Meteorological risk A1, management mechanism and awareness risk D1, and construction supervision unit's qualification risk D2 belonged to the driving risk factors, drilling technique risk B1, risk of

coordination with the government E1, and risk of coordination with port and shipping enterprises E2 belonged to the conclusion risk factors. Navigation risk A2, ship operation risk B2, construction personnel risk C1, and traffic organization risk C2 belonged to the transitional risk factors. In the analysis results of the AHP method, the construction supervision unit's qualification risk (D2) had the highest weight, the management mechanism and awareness risk (D1) was the second, and the risk of coordination with the government (E1) had the lowest weight. The score of the overall risk level of the project was 2.1.

REFERENCES

[1] Yin, L. Pan, and X. Li, "A Novel Multi-Criteria Decision-Making Approach for Intellectual Property Risk Assessment in Crowdsourcing Design," J. Circuits Syst. Comp., vol. 31, pp. 1-25, 2022.

[2] M. Khalilzadeh, H. Shakeri, and S. Zohrehvandi, "Risk identification and assessment with the fuzzy DEMATEL-ANP method in oil and gas projects under uncertainty," *Proc. Comput. Sci*., vol. 181, pp. 277-284, 2021.

[3] O. Ahmadi, S. B. Mortazavi, H. A. Mahabadi, and M. Hosseinpouri, "Development of a dynamic quantitative risk assessment methodology using fuzzy DEMATEL-BN and leading indicators," *Process Saf. Environ*., vol. 142, pp. 15-44, 2020.

[4] M. Sága, I. Kuric, I. Klačková, and D. Więcek, "Comparison of risk assessment approaches and analyzes used in technical transport systems," *Transp. Res. Proc.*, vol. 74, pp. 516-521, 2023.

[5] M. Li, S. Tian, C. Huang, W. Wu, and S. Xin, "Risk Assessment of Highway in the Upper Reaches of Minjiang River under the Stress of Debris Flow," *J. Geosci. Environ. Protect.*, vol. 09, pp. 21-34, 2021.

[6] S. Dhulipala, "Gaussian Kernel Methods for Seismic Fragility and Risk Assessment of Mid-Rise Buildings," *Sustainability*, vol. 13, pp. 1-25, 2021.

[7] H. C. Liu, J. X. You, Q. L. Lin, and H. Li, "Risk assessment in system FMEA combining fuzzy weighted average with fuzzy decision-making trial and evaluation laboratory," *Int. J. Comput. Integ. M*., vol. 28, pp. 701-714, 2015.

[8] A. Mentes, H. Akyildiz, M. Yetkin, and N. Turkoglu, "A FSA based fuzzy DEMATEL approach for risk assessment of cargo ships at coasts and open seas of Turkey," *Safety Sci*., vol. 79, pp. 1-10, 2015.

[9] A. K. Sangaiah, O. W. Samuel, X. Li, M. Abdel-Basset, and H. Wang, "Towards an efficient risk assessment in software projects–Fuzzy reinforcement paradigm," *Comput. Electr. Eng*., vol. 71, pp. 833-846, 2018.

[10] X. Li, R. Wan, B. Wu, G. Meng, and H. Chen, "Risk Assessment of the construction of metro stations with open cut method based on fuzzy comprehensive evaluation method," *IOP Conf. Ser.: Earth Environ. Sci.*, vol. 636, pp. 1-7, 2021.

[11] T. Hao, X. Zheng, H. Wang, K. Xu, and Y. Yu, "Development of a method for weight determination of disaster-causing factors and quantitative risk assessment for tailings dams based on causal coupling relationships," *Stoch. Env. Res. Risk A*., vol. 37, pp. 749-775, 2023.

[12] G. F. Can, and P. Toktaş, "A novel fuzzy risk matrix based risk assessment approach," *Kybernetes*, vol. 47, pp. 1721-1751, 2018.

[13] S. M. Hatefi, and J. Tamoaitien, "An integrated fuzzy dematel-fuzzy anp model for evaluating construction projects by considering interrelationships among risk factors," *J. Civ. Eng. Manag*., vol. 25, pp. 114-131, 2019.

[14] S. Liu, X. Guo, and L. Zhang, "An Improved Assessment Method for FMEA for a Shipboard Integrated Electric Propulsion System Using Fuzzy Logic and DEMATEL Theory," *Energies*, vol. 12, pp. 1-18, 2019.

[15] S. Sharifi, R. Shahoei, B. Nouri, R. Almvik, and S. Valiee, "Effect of an education program, risk assessment checklist and prevention protocol on violence against emergency department nurses: A single center before and after study," *Int. Emerg. Nurs.*, vol. 50, pp. 1-6, 2020.

[16] Y. Sun, T. Zhang, S. Ding, Z. Yuan, and S. Yang, "Risk assessment of high-speed railway CTC system based on improved game theory and cloud model," *Railw. Sci*., vol. 3, pp. 388-410, 2024.

[17] Y. Q. Sang, H. Li, and H. H. Jin, "Safety risk assessment of aircraft EWIS based on the improved combined weight and cloud model," *Aircr. Eng. Aerosp. Tec*., vol. 95, pp. 1470-1482, 2023.

[18] Z. Huang, F. Gao, X. Li, and M. Jiang, "Static and dynamic fuzzy assessment methods for the collapse risk of mountain tunnel construction," *J. Intell. Fuzzy Syst*., vol. 45, pp. 7975-7999, 2023.

[19] J. Li, and K. Xu, "A combined fuzzy DEMATEL and cloud model approach for risk assessment in process industries to improve system reliability," *Qual. Reliab. Eng. In*t., vol. 37, pp. 2110-2133, 2021.

[20] R. Deb, and S. Roy, "A Software Defined Network information security risk assessment based on Pythagorean fuzzy sets," *Expert Syst. Appl*., vol. 183, pp. 1-17, 2021.

[21] A. Tokai, and T. Nakakubo, "Risk Assessment and Management Methodologies for Applying Unusual Chemical Releases Derived from Disaster and Accident," *Jpn. J. Risk Anal*., vol. 29, pp. 195-197.

[22] K. Pang, C. Cheng, H. Zhao, Y. Ma, B. Dong, and J. Hu, "Simultaneous Analysis and Risk assessment of Quizalofop, Acifluorfen, Bentazone and Its Metabolites residues in Peanut and Straw under field conditions of China," *Microchem. J*., vol. 164, pp. 1-9, 2021.

[23] H. Lin, W. Lan, Q. Feng, X. Zhu, T. Li, R. Zhang, H. Song, Y. Zhu, and B. Zhao, "Pollution and ecological risk assessment, and source identification of heavy metals in sediment from the Beibu Gulf, South China Sea," *Mar. Pollut. Bull.*, vol. 168, pp. 112403, 2021.

[24] W. Pan, "Research on power system risk assessment considering large-scale wind and solar access," *J. Phys.: Conf. Ser.*, vol. 2215, pp. 012024, 2022.

[25] W. Zhao, Z. Zhou, and L. Q. Li, "Risk assessment of non-point source pollution in karst reservoirs based on 'source – sink' landscape theory," *Water Sci. Technol.*, vol. 22, pp. 6094-6110, 2022.

Blockchain-Based Financial Control System

Tedan Lu
School of Business, Jiangxi Modern Polytechnic College, Nanchang 330095, China

***Abstract*—In order to solve the problems of data security and low efficiency of information transmission in traditional financial control systems, this paper discusses in depth the application of blockchain technology in financial control systems. In order to optimize the performance of the traditional financial control system, this paper introduces blockchain technology into it and analyzes the structure and function of the financial control system. By constructing a blockchain-based financial data collection, information exchange and security consensus mechanism, a more efficient financial control system is designed, which can significantly improve the cost efficiency, shorten the audit cycle and enhance the data security. In the model, resource allocation within the financial control system is optimized, information exchange is more efficient, and a consensus mechanism is established. The experimental results prove that the model simplifies data entry and storage, reduces the workload of financial staff and improves transparency. The study bridges the gap between blockchain and traditional financial frameworks and advances the development of modern financial control systems.**

***Keywords*—Blockchain technology; financial control system; resource allocation; information exchange; consensus mechanism**

I. Introduction

With the development of information technology, internet technology has been widely applied in various industries. As one of the important contents of internal management control in enterprises, financial data has a large number of human factors that affect the recording forms of accounting vouchers, books, and other records in traditional financial data entry methods [1-3]. Encrypting the original accounting vouchers is a necessary prerequisite to ensure the authenticity and completeness of accounting information [4]. As the size and complexity of enterprise financial data grows, traditional financial control systems are exposed to many limitations. It faces serious challenges in data recording accuracy, processing efficiency and security, and is difficult to adapt to the rapidly changing business environment. With the rise of blockchain technology, its decentralization, transparency, and tamperability have brought a new dawn to the field of financial control. However, blockchain still has many problems in integrating with existing financial infrastructure and meeting regulatory compliance. Therefore, this study aims to deeply analyze the shortcomings of the traditional financial control system, explore the optimization path of blockchain technology integration with it, and construct a new financial control model. This article provides a deep understanding of the decentralized structure and analyzes the development and application of blockchain technology based on existing literature, in order to establish a long-term development strategy for optimizing and improving the performance of traditional financial control systems.

*Corresponding Author

Blockchain technology has obvious advantages in optimizing transmission architecture and improving system characteristics. Analyzing the application and development of blockchain technology in various fields is an important basis for studying its feasibility. The research on traditional blockchain has not been investigated from a technical and application perspective. Javaid conducted a comprehensive investigation into blockchain technology to fill this gap [5]. Xie explored the application theory of blockchain in the financial field, and designed a financial management control platform based on blockchain technology to better summarize the development direction of blockchain technology, providing a case study for analyzing the role of blockchain technology in the supply management process [6]. Zhang conducted research on financial control systems in the financial industry from three perspectives: data, rules, and applications, and provided constructive suggestions for promoting the development of blockchain technology in the field of financial control [7]. With the outdated performance of existing financial control systems and increasingly fierce market competition, Markos Zachariadis explored the process of building a distributed billing structure using blockchain technology and provided insights from a critical perspective [8]. Francesca Antonucci took advantage of the transparency and relatively low transaction costs of blockchain technology to construct control systems in an encrypted and distributed form to achieve new optimization solutions [9]. The above plan analyzes the feasibility of blockchain technology application in various fields, opening up new directions for the optimization and development of control systems.

The quality of financial control systems is influenced by many factors. In order to determine the reasons for the insufficient performance of traditional financial control systems and the effective improvement of advanced information technology in financial control systems, it is necessary to explore them. Ratmi Dewi studied the impact of the use of internal financial control systems on the quality of financial statements and reflected on the application of advanced information technology in financial control systems [10]. With the advancement of information technology stimulating the relevant advantages of financial systems in the financial industry, Ulrich Bindseil reviewed the impact of effective control of financial systems on the financial system, and compared and discussed the development direction and optimization trend of financial control systems [11]. Javad Oradi summarized the internal weaknesses of financial control systems, providing a research case for reducing audit costs and improving accounting efficiency [12]. In the process of examining the internal control strength of financial control systems, SAPUTRA Komang Adi Kurniawan found that organizational culture and human resource capabilities are one

of the key factors affecting the performance of financial control systems [13]. In the process of traditional access to financial control systems, a large amount of private data is generated. Due to the inability of traditional access control methods to ensure the security of financial data transmission, the current system optimization trend has made prospects for the development and application of blockchain technology in financial systems [14-16]. The above research only provides a brief analysis of the factors that affect financial control systems and the direction in which information technology can improve financial systems, and further research is needed.

In order to improve the performance of financial control systems and enhance the confidentiality and convenience of financial information data transmission, this article studied the architecture and workflow of traditional financial control systems, and analyzed the factors that affect the operational mode of financial control systems. In response to the shortcomings and shortcomings of traditional financial control systems, improvement suggestions were proposed. Based on blockchain technology, an anonymous financial control credit mechanism was constructed, and a decentralized financial control system was established by combining consensus algorithms. In this optimized financial control system, point-to-point information exchange can greatly improve system efficiency, and distributed data storage can reduce information acquisition costs while also improving the security and objectivity of financial data.

Compared with other existing solutions, the financial control system constructed based on blockchain technology in this paper shows significant advantages in many aspects. It solves the problems of the traditional financial control system, such as difficult to distinguish the subject, easy to tamper with information, and limited resource allocation, etc. By optimizing the structure and functions, such as establishing an anonymous crediting mechanism, adopting distributed storage and encryption technology, it achieves a significant increase in cost efficiency, which can significantly reduce the cost of cross-border transactions; significantly shortens the auditing cycle and reduces the risk of accounting fraud; and enhances the security of data and effectively prevents the tampering of data.

Main Contributions:

The main contributions of this paper can be summarized as follows:

1) Significant improvement in cost efficiency: The blockchain-based financial control model constructed in this paper has made a breakthrough in cost efficiency. Through experimental verification, the model can significantly reduce cross-border transaction costs and save a lot of money for enterprises.

2) Significant reduction in audit cycle: By optimizing the audit process, the model in this paper significantly reduces the audit cycle of transactions, improves the efficiency of financial data processing, and reduces the risk of accounting fraud.

3) Enhanced data security: Using the distributed storage and encryption technology of blockchain, the model in this paper enhances the security of financial data. Through comparative experiments, its excellent performance in preventing data tampering and improving data storage security is verified.

4) Combination of theory and practice: This paper not only theoretically analyzes the application of blockchain technology in the field of financial control, but also verifies the effectiveness of the model through actual cases and experiments, providing practical guidance for the application of blockchain technology in financial control systems.

II. Defects in Traditional Financial Control Systems

The financial control system is an important control component in the business process of an enterprise, which is a program that adjusts and constrains the financial activities of the enterprise to achieve commercial value. Traditional financial control systems have key functions such as financial prediction, financial decision-making, financial analysis, and financial evaluation to process and manage financial data, and have smooth and systematic operation. If the financial control system of an enterprise falls into chaos and cannot effectively control and manage financial data, it would reduce the management efficiency of the enterprise and leave the entire business in a disorderly state [17-19]. In short, financial control systems play a crucial role in curbing fraudulent economic activities and ensuring the safety and integrity of economic assets. Analyzing and summarizing the shortcomings of traditional financial control systems is the foundation for building a scientific and efficient financial control system.

A. Main Body of Financial Control System

In the application process of traditional financial control systems, reasonable control is based on the financial flow information, goods transaction information, and raw material procurement information generated in enterprise economic activities. Financial decisions in economic activities are comprehensively determined by various stakeholders, regulating the relationships between various operating entities in the economic plan, and forming a complete functional network. However, this makes it difficult to distinguish the controlling entities of the financial system [20]. In order to strengthen the information management of the financial control system, this article analyzes the control subjects of the financial system, hoping to make a unified plan for the control subjects of the financial system, and arrange and process financial information and economic processes.

The control subject of traditional financial systems should have the functions of financial departments and work arrangements, and should be in a central position in the governance structure of enterprises, with absolute discourse power in decision-making on financial conditions and economic activities. Faced with complex sources of financial information, the information management capabilities of traditional financial control systems are limited. Therefore, the subject of financial control should exist in the system as a multi-level state, in order to manage and process financial data with higher efficiency in modern enterprise economic activities.

B. Financial Control System Structure

With the rapid development of technology, the form of capital flow in the enterprise economy is also undergoing a huge transformation, and the transformation process from the real economy to the digital economy is breathtaking. Electronic vouchers are more convenient and easier to manage compared to paper vouchers. However, the digital form of currency has led to an explosive growth of information in financial control systems. Compared to traditional forms of currency, data in financial information management is prone to tampering, and data storage and transmission processes are no longer reliable [21-23]. In the new situation of modern capital flow, the establishment of a comprehensive financial control system cannot only be limited to plans and considerations.

With the changes and development of capital flow forms, the management philosophy of the financial control system should be updated, and ultimately implemented in the structural transformation of the financial control system [24-25]. Due to the inability of traditional financial control systems to ensure the confidentiality and objectivity of financial information storage and transmission processes, this article tentatively utilizes blockchain technology to optimize its structure, such as constructing a reasonable anonymous credit mechanism and decentralized system structure, and efficiently handling problems throughout the entire production and operation process with a sounder financial control system structure. The structure of the financial control system constructed in this paper is shown in Fig. 1:

C. Resource Allocation of Financial Control System

From the perspective of system resource allocation, the financial control system should solve problems encountered in financial information processing difficulties, risk response, and business models with more economical and reasonable resource allocation plans [26]. The resource allocation plan for the financial control system must include all levels of participation in enterprise economic activities to achieve the initial goal of integrating resources and covering the entire company. Under the influence of a comprehensive financial control system resource allocation plan, the means of financial control are no longer limited to a single audit and inspection in traditional financial control systems.

With the advancement of modern economic construction, financial control systems have put forward higher performance requirements. A financial control system with principles, competitiveness, and objectivity is conducive to the healthy development of the economic market and the integration of the financial field and control system. Analyzing the workflow and structural characteristics of traditional financial control systems can effectively identify the application deficiencies of traditional financial control systems in modern economic activities. This article summarizes the shortcomings of traditional financial control systems and proposes effective solutions to address these shortcomings. This innovation and development have been made to achieve efficient management of financial information in financial control systems and promote the progress of modern financial information management work.

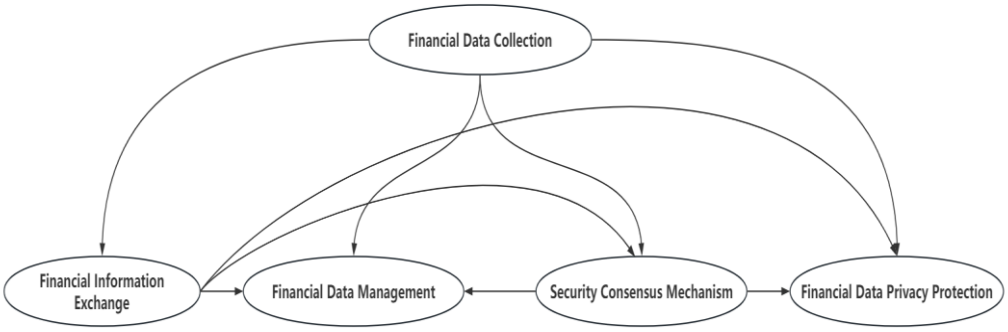


Fig. 1. System structure diagram.

III. Basic Scheme of Financial Control System

In order to achieve the goal of optimizing traditional financial control systems, this article tentatively combines blockchain technology to analyze the structure and functions of financial control systems. The financial control system not only has accounting and bookkeeping functions, but also assists leaders in decision-making and management. It can also analyze financial data through financial information processing software to generate financial statements that reflect the operation of enterprise economic activities. A more reasonable and scientific financial control system structure can better collect, analyze, and predict financial information.

A. Financial Data Collection

The collection of financial data by the financial control system is a prerequisite for its operation. Good data collection capabilities can provide raw data support to the financial control system and ensure its normal operation. This article is based on blockchain technology to encrypt and save financial data, enabling timely generation of corresponding timestamps and authentication signatures during the process of uploading and downloading financial data in the system. Before extracting financial data, proof of work must be provided for the financial data in the blockchain. The financial data collection function based on blockchain makes the operation of the financial control system more transparent and secure. Each system operator can be recorded in real-time while performing any operation on financial data, including various commercial transaction data and daily operational flow information, as well as transaction records of enterprises, distributors, and even individuals, which can be stored in encrypted form in the blockchain, providing necessary technical support for the normal operation of the financial control system.

B. Financial Information Exchange

In the financial control system, the main purpose of constructing an interaction layer is to process various financial information generated in enterprise economic activities. While funds flow, the financial control system based on blockchain technology abandons the traditional form of paper currency transactions and instead uses encrypted digital currencies similar to Bitcoin for transactions on the blockchain. Every process of financial information interaction reflects a real enterprise economic transaction, and the initiation and end of each transaction are recorded in encrypted form in the blockchain. This ensures the security of financial information while also ensuring the objectivity of financial information, and records various details and processes of the transaction in detail.

C. Financial Data Management

In the enterprise's economic activities, the management is responsible for the interactive operation of recording the capital flow and financial data in the form of cryptocurrency, and assisting the financial information interaction layer to process and manage the relevant financial data while recording the main economic transaction records. In an optimized financial control system based on blockchain technology, management can apply corresponding types of cryptocurrency systems to different types of transactions, enabling enterprises to record and manage all processes related to economic activities from the beginning to the end. Due to the fact that all fund transfer records in the financial control system are stored on the blockchain, it ensures that the financial data during the transaction process and fund flow conversion cannot be tampered with by third parties, ensuring the reliability of encrypted financial data storage and greatly enhancing the company's ability to control financial data. The working interface of the financial control system is shown in Fig. 2.

D. Security Consensus Mechanism

Based on the application of blockchain technology, a security consensus mechanism is established in the financial control system. Due to the susceptibility to network latency during peer-to-peer network access, it is common for nodes in the blockchain to receive financial data at different times. Therefore, this article applies a randomly parallel blockchain security consensus algorithm to establish a security consensus mechanism, which unifies the receiving time of each node in the blockchain during transactions. The member who first completes the transaction using the consensus algorithm records the transaction time, thereby obtaining the identity of the bookkeeper and receiving accounting rewards.

E. Financial Data Privacy Protection

In order to improve the privacy of financial data, the optimized financial control system in this paper tentatively refers to blockchain homomorphic encryption technology to optimize it. In the practical application of optimizing financial control systems, each node in the blockchain has the same copy of financial data, thus achieving a balanced treatment of the objectivity and privacy of the transaction process in the entire enterprise's economic activities. In the process of point to point transactions, it is not necessary to fully read financial data to ensure the objectivity and privacy of the transaction, while also preventing illegal hackers from invading and stealing.

In the face of increasingly complex economic transaction activities, this article analyzes the workflow and infrastructure of the financial control system in order to build a more reliable financial control system and achieve more efficient control of financial data for companies. The working structure diagram of the optimized financial control system is shown in Fig. 3. The above section introduces the basic scheme and functions of a financial control system designed based on blockchain technology in this article. However, further experiments are needed to verify the working performance and application effectiveness of the optimized financial control system.

Fig. 2. Financial control system work interface.

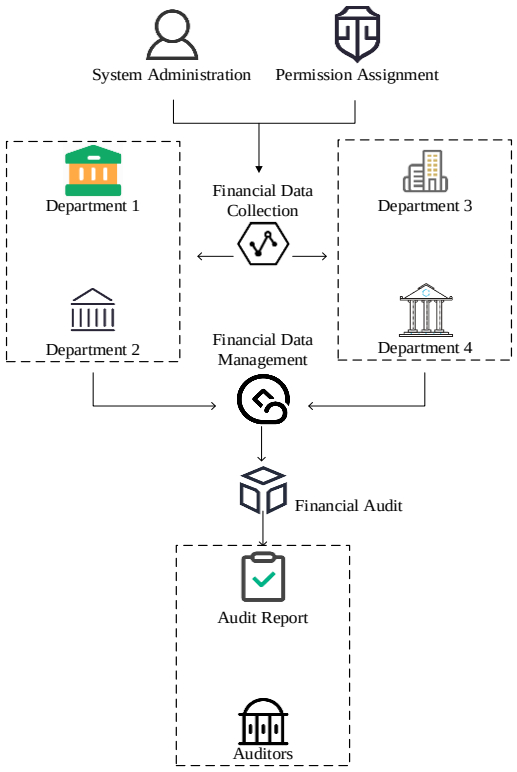


Fig. 3. Optimizing the work structure of financial control system.

IV. Performance Comparison Experiment of Financial Control System

A. Cross Border Transaction Cost Comparison

The coordination of financial information is an important function of financial control systems, and improving the information coordination performance of financial control systems is the key to enhancing the competitiveness of enterprises. Cost control in cross-border transactions is an important indicator of a company's strength. Low cost capital operation often provides enormous convenience for enterprises. How to open up a broader economic market with more efficient cost control is a serious challenge to the information coordination ability of financial control systems. The indicators for cost calculation by the control center are shown in Table I.

To explore the impact of blockchain technology on financial control systems and verify that optimizing financial control systems can reduce the cost of cross-border transactions by coordinating financial information, this article conducted an application experiment on a financial control system in a certain e-commerce enterprise. 60 e-commerce transaction orders in the same batch were randomly sampled as experimental data and divided into two groups, with 20 transaction orders for children's toys, mother and baby products, and women's and men's clothing. The experiments that applied traditional financial control systems and optimized financial control systems were divided into a control group and an experimental group, and the average transaction costs of 60 transaction orders in the two groups of experiments were compared. Fig. 4(a) and Fig. 4(b) are comparative analyses of transaction costs between traditional and optimized system applications.

Enterprises often face high capital operating costs when facing overseas markets, and the operational processes in economic transactions are complex and have high transaction fees. Efficient and cleverly structured financial control systems can help simplify the complex process of economic transactions, making transactions that originally required multiple stages simpler. This article is based on the blockchain network to optimize the financial control system. In practical applications, it crosses the clearing process of import and export banks and intermediate third parties, and directly conducts point-to-point transactions in the blockchain network. Fund transfer is carried out in the form of encrypted digital currency, reducing the complexity of transactions and improving the operational efficiency of funds. As shown in Fig. 4(a), the average cost of transaction orders for children's toys, mother and baby products, and women's and men's clothing in the control group with traditional financial control systems was 60000, 70000, and 70000 yuan, respectively. As shown in Fig. 4(b), the average cost of transaction orders for children's toys, mother and baby products, and women's and men's clothing in the experimental group with the application of the optimized financial control system was 50000 yuan, 50000 yuan, and 40000 yuan, respectively. Compared to the transaction orders with the application of traditional financial control systems, the transaction orders with the optimized financial control system had significantly reduced costs.

TABLE I. Indicators Involved in Cost Coordination Responsibility Report

Projects	Indicators			Projects	Indicators		
Material Costs	Budget	Actual	Differences	Uncontrollable costs	Budget	Actual	Differences
Labor Costs	Budget	Actual	Differences	Depreciation of equipment	Budget	Actual	Differences
Variable Costs	Budget	Actual	Differences	Other	Budget	Actual	Differences

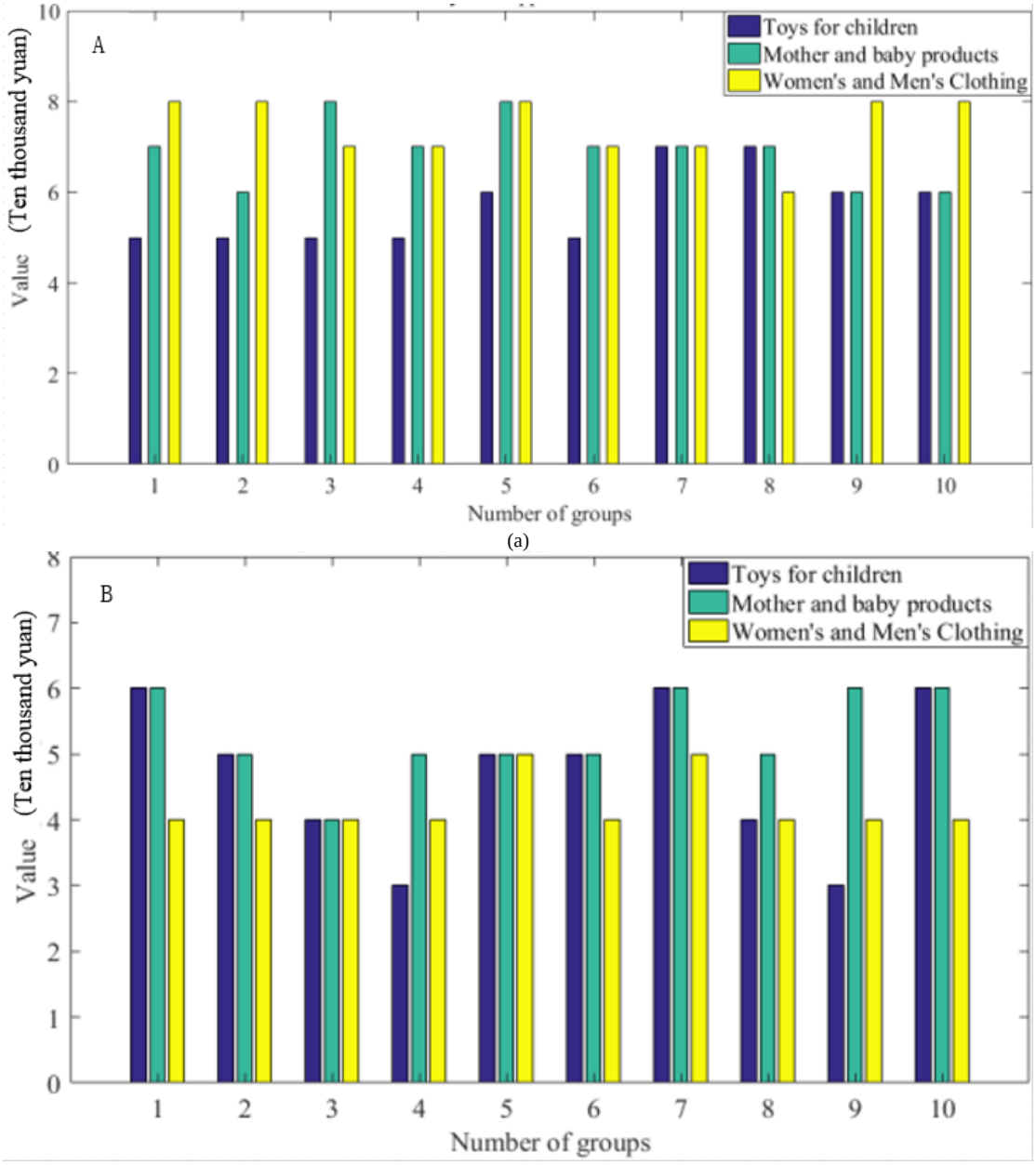


Fig. 4. (a) Traditional transaction cost diagram, (b) Optimized system transaction cost diagram.

B. *Comparison of Audit Process Cycles*

Traditional financial control systems have complex transaction processes for enterprise economic activities, such as counterparty confirmation, financial data analysis, information registration, system auditing, etc. The efficiency of decision-making plans for economic activities is not high, and the audit process of accounting bookkeeping, confirmation, and final generation of accounting reports after transactions are completed increases the risk of accounting fraud. To verify the positive role of optimizing financial control systems in the audit process cycle, this article conducted a comparative experiment. 80 transactions were divided into control and experimental groups, with 40 orders for children's toys and 40 orders for mother and baby products. The experiment using traditional financial control systems was set as the control group, while the experiment using optimized financial control systems was set as the experimental group. The audit cycles of transactions in two sets of experiments were recorded and compared. As shown in Fig. 5, a comparative analysis of audit cycles between traditional and optimized systems is presented.

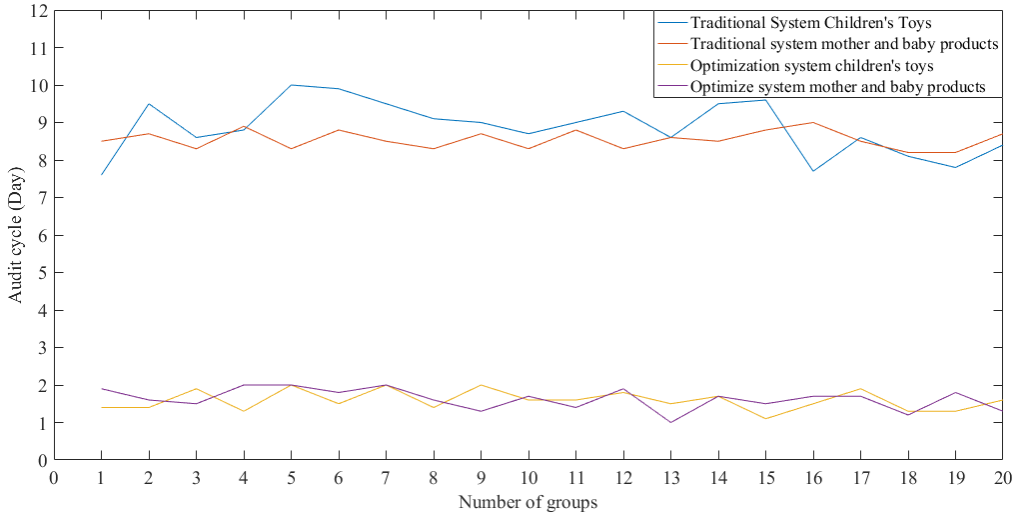


Fig. 5. Comparison of audit cycles between traditional and optimized systems.

As can be seen from the image in Fig. 5, in the traditional financial control systems, transactions of economic activities of enterprises must be reviewed and approved by central institutions. Without the allocation and supervision of control centers, transaction chaos can easily occur in complex audit processes. Due to the overly complex audit process of traditional financial control systems, errors in financial information exchange often occur, such as duplicate payments, which can cause varying degrees of confusion in the transaction records of both parties. The optimized financial control system based on blockchain technology architecture in this article can record transaction information in real-time, and the arrangement of blocks in the blockchain network can be sequentially recorded. Only then can the initial audit information be recognized by other nodes in the network, and there is no risk of confusion in the process of auditing financial information. As shown in Fig. 4, in the control group with the application of traditional financial control systems, the average audit period for children's toy type transaction orders was 8.9 days, and the average audit period for mother and baby product type transaction orders was 8.6 days. In the experimental group that applied the optimized financial control system, the average audit period for children's toy type transaction orders and mother and baby product type transaction orders was 1.6 days. From the comparison of data in Fig. 4, it can be seen that the optimized financial control system based on blockchain technology has greatly improved the audit process compared to traditional financial control systems, greatly refining important links in the audit process, skipping or even removing the process of relative shortcomings, greatly increasing the efficiency of the financial data audit process.

C. Comparison of Financial Data Security

As a financial control system, ensuring the security of financial data storage is crucial. The process of managing and controlling financial data in traditional financial control systems is uniformly planned by the control center, and the top management formulates a rough process to process and store financial data in a top-down structure. In the complex financial data processing process, the security of financial data cannot be fully guaranteed. The security protection of financial data is not comprehensive, leading to difficulties in perfect handover between various departments in the enterprise during accounting and auditing, and even causing huge commercial loopholes and economic losses. The optimized financial control system based on blockchain technology in this article adopts a distributed encrypted storage method for financial data, greatly improving the security of financial data. In order to verify the positive effect of blockchain technology on financial data storage security performance, security comparison experiments were conducted. This article simulated hackers invading and tampering with the financial data storage modules of traditional and optimized financial control systems, in order to verify the security level of financial control system storage of financial data. Establishing a logically rigorous security evaluation standard is the prerequisite and foundation for verifying the security of financial control systems. In each intrusion experiment, the security of the financial control system was evaluated, with an upper limit of 12, 1-3 being poor, 4-6 being average, 7-9 being good, and 10-12 being excellent. The results of each intrusion were recorded as the evaluation basis. The recorded results of 15 intrusion experiments are shown in Table II.

The blockchain network adopts decentralized auditing and accounting, with each node in the network participating together and copying financial data separately. Each node only determines the ownership of accounting rights based on whether the transaction was completed first and recorded, which is recognized by all nodes. The optimized financial control system based on blockchain network is different from the traditional classified accounting structure of financial control systems, but adopts a distributed accounting structure. Each node in the fully running blockchain network can access transaction data that has been confirmed to exist, and can independently check and verify the integrity of the branch chain. Due to the immutability and persistence of the optimized financial control system based on blockchain technology, the financial data stored in the optimized financial control system is also more secure. As shown in Table 2, after 15 simulated intrusion experiments, in the control group experiment that applied the traditional financial control system, the intrusion was successful twice with an average evaluation index of 7. In the experimental group experiment that applied the optimized financial control system, all simulated intrusions

failed with an average evaluation index of 11. According to data comparison, the optimized financial control system based on blockchain technology architecture has an improved security performance evaluation index of four compared to traditional financial control systems in terms of financial data storage.

TABLE II. RECORD ANALYSIS OF INTRUSION EXPERIMENTS

	Traditional Financial Control System		Optimize financial control systems	
Number of experiments	Evaluation Index	Intrusion results	Evaluation Index	Intrusion results
1 time	7	Failed	12	Failed
2 time	5	Succeed	9	Failed
3 time	7	Failed	12	Failed
4 time	8	Failed	11	Failed
5 time	8	Failed	12	Failed
6 time	9	Failed	10	Failed
7 time	7	Failed	11	Failed
8 time	8	Failed	10	Failed
9 time	5	Succeed	12	Failed
10 time	7	Failed	9	Failed
11 time	8	Failed	12	Failed
12 time	7	Failed	12	Failed
13 time	6	Failed	11	Failed
14 time	9	Failed	10	Failed
15 time	8	Failed	11	Failed

V. RESULT AND DISCUSSION

Compared with traditional systems, the blockchain based financial control system proposed in this article has achieved significant improvements in all aspects. Through experimental verification, the model can effectively reduce cross-border transaction costs. In the e-commerce enterprise experiment, after applying the optimized system, the average cost of different product transaction orders significantly decreased. In terms of audit cycle, the optimization system based on blockchain technology greatly shortens the audit process. The traditional system has a relatively long review cycle for different types of transactions, while the optimized system greatly shortens the review cycle and improves the efficiency of financial data processing. In terms of data security, optimizing the distributed encrypted storage method of the system makes financial data more secure. After multiple intrusion experiments, traditional systems have a certain success rate of intrusion with a low average evaluation index, while optimized systems can effectively resist all simulated intrusions with a high average evaluation index. These results demonstrate the effectiveness and superiority of the proposed model.

With the vigorous development of the digital economy, blockchain technology has gradually entered the social perspective and played an important role in multiple fields, especially in the economic field, where there is an urgent need for strong support from advanced technology. The management system and classified accounting structure of traditional financial control systems are difficult to enable rapid interaction and transmission of financial information between relevant departments, and traditional system solutions are prone to hacker intrusion and tampering, leading to frequent examples of major economic accidents. Due to the advantages of distributed data storage structure and immutability, blockchain technology can achieve secure and efficient processing of financial data, ensure the authenticity of financial data and the transparency of economic transaction processes, and promote the work efficiency and management level of financial control systems.

This article analyzed the workflow and structural architecture of traditional financial control systems, tentatively applied blockchain technology to optimize the financial control system, and designed a more excellent financial control system. The performance and application effectiveness of the optimized financial control system were verified through comparative experiments. The results showed that the optimized financial control system in this article had more precise cost control and a safer financial data storage structure, ensuring the normal progress of enterprise economic activities while saving a lot of time and costs. In the future, this paper will conduct in-depth research on the compatibility issues that the current system may face with existing financial infrastructure, explore effective integration strategies, and ensure that the system can smoothly integrate into the current financial ecosystem. Thus promoting the application and popularization of blockchain financial control systems in a wider range of fields, providing more efficient, secure, and intelligent solutions for enterprise financial control.

REFERENCES

[1] Xu Xinxin. Financial Sharing Mode of Leather Enterprises Based on Blockchain Technology. China Leather,2023,52(1):48-51.

[2] Hou Cheng. Impact and Challenge Analysis of Blockchain Technology on Enterprise Financial Sharing Center[J]. Jiangsu Commercial Forum,2022(12):70-73.

[3] Meng Miao. The Application of Blockchain Technology in the Financial Management of State-owned Enterprises under the Background of Mixed Reform. Economig Rfsearch Cuide,2022(1):77-80.

[4] Lin, Jiangxiong. Design of enterprise financial early warning model based on complex embedded system. Microprocessors and microsystems,2021,80(Feb.):103532.1-103532.7.

[5] Javaid M, Haleem A, Singh R P, et al. A review of Blockchain Technology applications for financial services. BenchCouncil Transactions on Benchmarks, Standards and Evaluations, 2022, 2(3): 100073-100091.

[6] Peng Xie, Qiang Chen, Ping Qu, Jianping Fan, Zhijun Tang,Peng Xie, etal. "Research on financial platform of railway freight supply chain based on blockchain." Smart and Resilient Transportation 2.2 (2020): 69-84.

[7] Li Zhang, Yongping Xie, Yang Zheng, Wei Xue, Xianrong Zheng, Xiaobo Xu. "The challenges and countermeasures of blockchain in finance and economics." Systems Research and Behavioral Science 37.4 (2020): 691-698.

[8] Markos Zachariadis,Garrick Hileman,Susan V. Scott. "Governance and control in distributed ledgers: Understanding the challenges facing blockchain technology in financial services." Information and Organization 29.2 (2019): 105-117.

[9] Francesca Antonucci, Simone Figorilli, Corrado Costa, Federico Pallottino, Luciano Raso, Paolo Menesatti. "A review on blockchain applications in the agri-food sector." Journal of the Science of Food and Agriculture 99.14 (2019): 6129-6138.

[10] Ratmi Dewi,Jan Hoesada. "The effect of government accounting standards, internal control systems, competence of human resources, and use of information technology on quality of financial

statements." International Journal of Innovative Research and Advanced Studies (IJIRAS) 7.1 (2020): 4-10.

[11] Ulrich Bindseil. "Central bank digital currency: Financial system implications and control." International Journal of Political Economy 48.4 (2019): 303-335.

[12] Javad Oradi, Kaveh Asiaei, Zabihollah Rezaee. "CEO financial background and internal control weaknesses." Corporate Governance: An International Review 28.2 (2020): 119-140.

[13] SAPUTRA Komang Adi Kurniawan, SUBROTO,Bambang, RAHMAN Aulia Fuad, SARASWATI Erwin. "Financial management information system, human resource competency and financial statement accountability: a case study in Indonesia." The Journal of Asian Finance, Economics and Business 8.5 (2021): 277-285.

[14] Dezhi Han, Yujie Zhu, Dun Li, Wei Liang, Alireza Souri, Kuan-Ching Li. "A blockchain-based auditable access control system for private data in service-centric IoT environments." IEEE Transactions on Industrial Informatics 18.5 (2021): 3530-3540.

[15] Sebahattin Demirkan, Irem Demirkan, Andrew McKee. "Blockchain technology in the future of business cyber security and accounting." Journal of Management Analytics 7.2 (2020): 189-208.

[16] Li X, Wang J, Yang C. Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy. Neural Computing and Applications, 2023, 35(3): 2045-2058.

[17] Ren S. Optimization of Enterprise Financial Management and Decision-Making Systems Based on Big Data. Journal of Mathematics, 2022, 2022(1): 1708506-1708517.

[18] Kazan G, Kocamış T U. Assessing the impact of blockchain technology on internal controls within the COSO framework. J. Corp. Gov. Insur. Risk Manag, 2023, 10(1): 86-95.

[19] Izwan Amsyar, Ethan Christopher, Arusyi Dithi,Amar Najiv Khan, Sabda Maulana. "The Challenge of Cryptocurrency in the Era of the Digital Revolution: A Review of Systematic Literature." Aptisi Transactions on Technopreneurship (ATT) 2.2 (2020): 153-159.

[20] Adam Faturahman,Vertika Agarwal, Chandra Lukita. "Blockchain technology-the use of cryptocurrencies in digital revolution." IAIC Transactions on Sustainable Digital Innovation (ITSDI) 3.1 (2021): 53-59.

[21] Wang Lanfei, Yang Yiwen. Design of financial software automatic control system based on BP neural network. Electronic Design Engineering,2024,32(20):163-167.

[22] Wang Wei. Design of Financial Risk Automatic Control System Based on Cloud Computing Technology. Techniques of Automation and Applications,2024,43(7):137-140.

[23] Li Jianfa, Zhang Xiaofang. Exploration of Practical Application of Thinking in Unit Financial Management System. Economig Rfsearch Cuide,2024(11):91-94.

[24] Ding Xin. Risk Control Financial Security Management Strategies for Plastic Enterprises. Plastic Additives,2024(3):82-86.

[25] Kong Ping, Wu Fei. Design of Financial Information Management System Based on Association Rules. Techniques of Automation and Applications,2022,14(7):166-169.

[26] Li Huwei. Analysis and design of financial data mining system based on fuzzy clustering. Expert systems: The international journal of knowledge engineering,2024,41(5):e13031.1-e13031.13.

User Interface Design of SEVIMA EdLink Platform for Facilitating Tri Kaya Parisudha-Based Asynchronous Learning

Agus Adiarta[1]*, I Made Sugiarta[2], Komang Krisna Heryanda[3],
I Komang Gede Sukawijana[4], Dewa Gede Hendra Divayana[5]
Department of Electrical Education, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia[1, 4]
Department of Mathematics Education, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia[2]
Department of Management, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia[3]
Department of Informatics Education, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia[5]

***Abstract*—This research aims to show the user interface design of the SEVIMA EdLink platform to facilitate *Tri Kaya Parisudha*-based asynchronous learning in the nuances of independent learning. This research used the Research and Development method with the Borg & Gall development model, which focused on several stages, including research and field data collection, planning, design development, initial trial, and revision of the initial trial results. The number of respondents involved in the initial trial of the user interface design was two education experts, two informatics experts, 40 teachers of Tourism Vocational Schools in Bali, and 60 students of Tourism Vocational Schools in Bali. The data collection tool for the initial trial of the user interface design was a questionnaire consisting of ten questions. The analysis was conducted by comparing the effectiveness percentage of the user interface design with the effectiveness categorization standard referring to the five scales. The results showed that the user interface design of the SEVIMA EdLink platform was effective in facilitating *Tri Kaya Parisudha*-based asynchronous learning. The impact of this research on stakeholders in the field of education is the existence of new information related to the existence of an online learning platform called SEVIMA EdLink, which is integrated with an asynchronous learning strategy, independent learning policy, and Balinese local wisdom.**

***Keywords*—Design user interface; SEVIMA EdLink; asynchronous; Tri Kaya Parisudha; independent learning**

I. Introduction

One of the efforts made by Tourism Vocational School to improve the quality of learning and the character of its students in the frame of implementing the independent learning policy is to determine the right learning strategy. One of the learning strategies used to make this happen is asynchronous learning. However, an asynchronous learning strategy has still not been effectively implemented. This is because the teacher does not directly meet the students like in-class learning. In addition [1], in asynchronous learning, teachers are also constrained to assess each student's character and quality of learning objectively. Teachers can easily assess the character and quality of student learning if they can interact with students directly in the classroom. An innovation is needed to solve the problems associated with such asynchronous learning. One of the innovations is adapting the SEVIMA EdLink platform into asynchronous learning based on the concept of Tri Kaya Parisudha. SEVIMA EdLink [2] is a learning platform that can be freely obtained from the internet. This platform provides facilities to make the learning process asynchronous, which refers to the cognitive, affective, and psychomotor domains [3]. The concept of Tri Kaya Parisudha consists of three parts [4], namely Manacika (thinking well), Wacika (saying well), and Kayika (doing well). Manacika [5] can be used as a foundation in asynchronous learning to determine the quality of student learning, especially in the cognitive domain. Wacika [5] determines student character (affective domain). Kayika [5] is used as a basis for determining the quality of student learning, especially in the psychomotor domain. The innovation will run well if followed by implementation. The initial effort made to implement the innovation was to design the user interface design of the SEVIMA EdLink platform based on Tri Kaya Parisudha. Based on this, the research question is: How is the user interface design of the SEVIMA EdLink platform to facilitate asynchronous learning based on Tri Kaya Parisudha in the nuances of independent learning?

II. Literature Review

Some of the research behind this study includes research by Utomo & Ahsanah [6], which shows that online learning using the Edmodo application is very effective. This is based on the results obtained from the field trial: giving a pretest and posttest and seeing the quantity of student communication in the forum and chat through the Edmodo application. The obstacle of Utomo & Ahsanah's research is that it has not shown the visualization of online learning in terms of material content and test questions used in measuring the quality of students' abilities and characters. Research by Soesanto et al. [7] showed the aspects of measuring the effectiveness of the blended learning model based on user activities in accessing content. The constraint of Soesanto et al.'s research has not shown the completeness of blended learning regarding test questions used in measuring the user's ability and character. Research by Sela et al. [8] shows that the blended learning model assisted by Google Classroom is very effective based on the percentage of learning implementation results. The limitation of Sela et al. research is that it has not shown the completeness of blended learning regarding material content

*Corresponding Author

and test questions used in measuring the quality of students' abilities and characters. The research of Anggraeni et al. [9] showed the measurement of the level of effectiveness of the blended learning model based on independent and collaborative asynchronous activities. The research constraint of Anggraeni et al. is that it has not shown the completeness of blended learning in terms of material content and test questions used in measuring the quality of students' abilities and characters.

Papadakis' research [10] states that people can access educational content for free through platforms that provide online learning content. The limitation of Papadakis' research is that it does not specifically explain the platforms that can be used for the online learning he means. Papadakis et al.'s research [11] shows the combined power of cloud technology and augmented reality in supporting the educational process. The limitation of Papadakis et al.'s research is that it has not shown in detail the process of combining cloud technology and augmented reality that can support the educational process, especially in measuring the quality of student's character.

III. Method

A. Research Approach

The approach of this research was development. This research used the Research and Development method, with a research development model, namely Borg & Gall, which consists of 10 stages of development [12], [13], [14], [15], including (1) research & field data collection; (2) planning; (3) design development; (4) initial trial; (5) revision of initial trial results; (6) field trial; (7) revision of field trial results; (8) usage trial; (9) final product revision; (10) desimination and implementation of the final product. Specifically for this 2024 year research, several stages were carried out, including (1) research and field data collection, (2) planning, (3) design development, (4) initial trial, and (5) revision of the initial trial results.

B. Research Subjects

The subjects in this study were determined using the Purposive Sampling technique. This technique was conducted by selecting research subjects initially determined based on the subject's direct relationship with the SEVIMA Edlink platform to facilitate Tri Kaya Parisudha-based asynchronous learning. The subjects involved in this 2024 year research were two education experts, two informatics experts, 40 Tourism Vocational School teachers in Bali, and 60 Tourism Vocational School students in Bali who will be involved in conducting the initial trial. All subjects involved have obtained official and valid consent from each individual without any coercion or conflict of interest.

C. Object and Location of Research

The object of research was the main topic that must be studied and researched in depth. The object of this study was the design of the SEVIMA Edlink platform to facilitate asynchronous learning based on Tri Kaya Parisudha in the nuances of independent learning. This research is implemented in several Tourism Vocational Schools spread across six regencies in Bali Province. The six regencies are Gianyar, Tabanan, Buleleng, Klungkung, Badung, and Denpasar.

D. Research Data Collection Instruments

The instruments used in collecting data in this research are questionnaires. The questionnaires were used to obtain primary data in the form of quantitative data from respondents as a basis for making decisions about the percentage level of effectiveness of the design of the SEVIMA Edlink platform used in asynchronous learning based on Tri Kaya Parisudha in the nuance of independent learning at some Tourism Vocational School in Bali.

E. Data Analysis Techniques

The technique used to analyze the collected data was quantitative descriptive technique through descriptive percentage calculation. The results of the descriptive percentage calculation were used as the basis for interpreting the results of the research on the design of the SEVIMA Edlink platform used in the asynchronous learning based on Tri Kaya Parisudha in the nuanced of independent learning to improve the learning outcomes of Tourism Vocational School students in Bali in the cognitive domain, affective, and psychomotor domains. The formula for calculating the descriptive percentage is as follows [16], [17], [18], [19], [20], [21].

$$P = \frac{f}{N} \times 100\% \quad (1)$$

Notes:

P = Effectiveness percentage

f = Total acquisition value

N = Maximum total value

The percentage results obtained from that formula are then converted into the following Table I [22], [23], [24], [25], [26].

TABLE I. Categorization Standards Referring to Five's Scale

Category of Effectiveness	Percentage of Effectiveness (%)	Follow-up
Poor	0-54	Revision
Less	55-64	Revision
Moderate	65-79	Revision
Good	80-89	No Revision
Excellence	90-100	No Revision

IV. Results and Discussion

A. Results

Some of the research results in 2024 were focused on five stages that refer to the Borg and Gall model. The five stages include 1) research and field data collection stage, 2) planning, 3) design development, 4) initial trial, and 5) revision of the initial trial results. The research results based on these five stages can be shown as follows.

1) The results of the research and field data collection stage: At this stage, data related to several things supporting asynchronous learning for entrepreneurship subjects in the Tourism Vocational School was obtained. The data intended includes 1) the content of entrepreneurship material given to students, 2) examples of test questions based on the concept of Tri Kaya Parisudha, and 3) the features needed in the SEVIMA EdLink platform to support learning asynchronous

in the nuances of independent learning. The data related to entrepreneurship material content can be seen in Table II. Examples of test questions based on the Tri Kaya Parisudha concept can be seen in Table III. The features provided in the SEVIMA EdLink platform to support the occurrence of asynchronous learning based on Tri Kaya Parisudha in the nuances of independent learning can be seen in Table IV.

TABLE II. CONTENT OF ENTREPRENEURSHIP MATERIALS PROVIDED TO STUDENTS OF TOURISM VOCATIONAL SCHOOL IN BALI

No	Content Material
1	Identify entrepreneurial attitudes and behaviors
2	Prestigious work attitudes and behaviors
3	Problem solution formulation
4	Development of spirit entrepreneurship
5	Build commitment to self and others
6	Business risks
7	Decision making

TABLE III. EXAMPLES OF ENTREPRENEURSHIP TEST QUESTIONS BASED ON THE CONCEPT OF TRI KAYA PARISUDHA

No	Content Material	Test Questions	
1	Identify entrepreneurial attitudes and behaviors	*Manacika*	Explain the difference between entrepreneurship, self-employment, and entrepreneurship!
		Wacika	Make a sound recording that shows an entrepreneur who can appreciate the work and experience of others as input for his/her development!
		Kayika	Make a video recording that shows the attitude of an honest and realistic entrepreneur!
2	Prestigious work attitudes and behaviors	*Manacika*	Explain the meaning, purpose, and benefits of presentative work behavior!
		Wacika	Make a voice recording that shows an entrepreneur's ability to maintain their emotional self while realizing presentative work behavior!
		Kayika	Make a video recording that shows the attitude of an entrepreneur who always works to get ahead!
3	Problem solution formulation	*Manacika*	Explain the meaning and difference between problems and non-problems!
		Wacika	Make a voice recording that shows an entrepreneur negotiating well when solving a problem!
		Kayika	Make a video recording that shows the attitude of an entrepreneur who always has how to determine alternative solutions to problems!
4	Development of spirit entrepreneurship	*Manacika*	Explain factors that influence entrepreneurial morale!
		Wacika	Make a voice recording that shows the attitude of an entrepreneur who can influence work morale!
		Kayika	Make a video recording that shows an entrepreneur being able to inspire the spirit of entrepreneurship!
5	Build commitment to self and others	*Manacika*	Explain the factors that show a person is highly committed to their entrepreneurial activities!
		Wacika	Make a voice recording that shows the attitude of an entrepreneur who can maintain a high commitment to self-control!
		Kayika	Create a video recording that shows an entrepreneur who is punctual in his work environment and daily life!
6	Business Risks	*Manacika*	Describe the types of risk in a business!
		Wacika	Make a voice recording that shows the attitude of an entrepreneur in avoiding/minimizing risk in business!
		Kayika	Make a video recording that shows the attitude of an entrepreneur when handling risks in business !
7	Decision making	*Manacika*	Explain the steps of decision-making in an enterprise!
		Wacika	Make a voice recording that shows the attitude of an entrepreneur when giving instructions to subordinates!
		Kayika	Make a video recording that shows the attitude of an entrepreneur in making a decision fairly!

TABLE IV. FEATURES PROVIDED IN THE SEVIMA EDLINK PLATFORM TO SUPPORT ASYNCHRONOUS LEARNING BASED ON TRI KAYA PARISUDHA IN INDEPENDENT LEARNING

No	Features
1	Class
2	Material
3	Task
4	Info
5	Quiz
6	Event
7	Survey
8	Comments

2) The result of planning stage: Data about the number of people involved, personal job descriptions, and the time needed to complete this research was obtained at this stage. The total time prepared for data collection and the revision of the trial results of the user interface design of the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning in nuances of independent learning was 30 days. The complete data related to this research planning can be seen in Table V.

TABLE V. DETAILS OF THE NUMBER OF PERSONAL JOB DESCRIPTIONS AND COMPLETION TIME USER INTERFACE DESIGNS OF SEVIMA EDLINK PLATFORM TO FACILITATE TRI KAYA PARISUDHA-BASED ASYNCHRONOUS IN THE NUANCES OF INDEPENDENT LEARNING

No	Total Personal	Personal Job Description	Time (Day)
1	6	Field data collection	6
2	3	Designing user interface of SEVIMA EdLink Platform to facilitate *Tri Kaya Parisudha*-based asynchronous learning	8
3	104	Initial testing of user interface design	12
4	3	Revision of initial trial results	4
Total	**116**		**30**

3) The result of the design development stage: Referring to some of the features provided in the SEVIMA EdLink platform to support the occurrence of asynchronous learning based on Tri Kaya Parisudha in the nuances of independent learning in some Tourism Vocational Schools in Bali, and the research planning that has been shown in Table V, then the initial design of the user interface of the SEVIMA Edlink platform used in asynchronous learning based on *Tri Kaya Parisudha* in the nuances of independent learning at several

Tourism Vocational School in Bali can be done. The user interface design of this platform was taken directly from the SEVIMA EdLink platform. The results of the initial design of the SEVIMA Edlink platform used in asynchronous learning based on *Tri Kaya Parisudha* in the nuances of independent learning can be seen in Fig. 1 to Fig. 8.

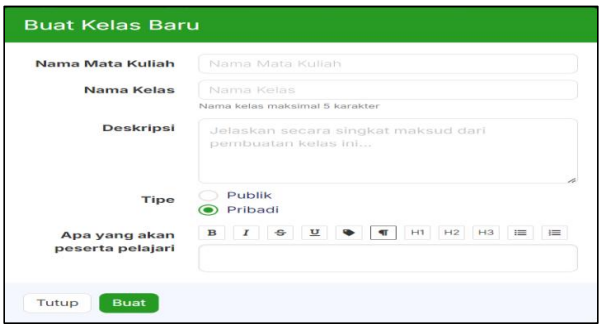


Fig. 1. Design showing facilities for creating classes (in bahasa format).

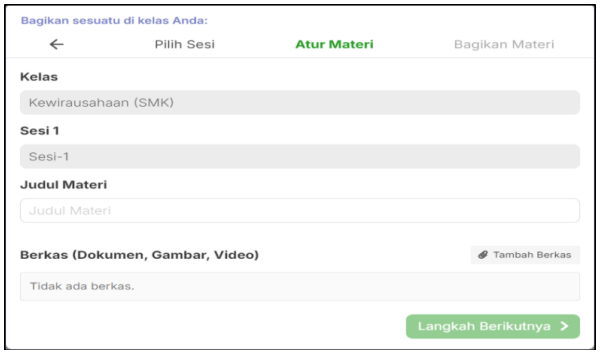


Fig. 2. Design showing facilities for entering course materials (in bahasa format).

Fig. 3. Design showing facilities for entering tasks (in bahasa format).

Fig. 4. Design showing facilities for entering news/information related to learning (in bahasa format).

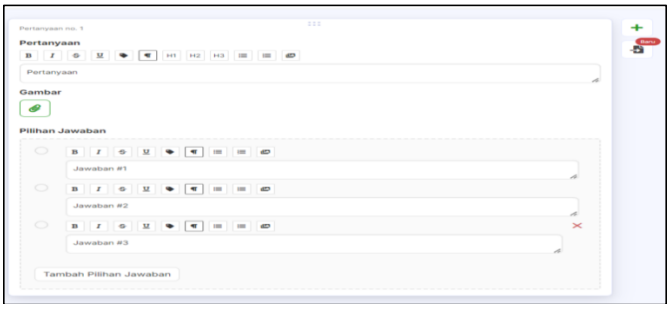


Fig. 5. Design showing the facility for entering test questions (in bahasa format).

Fig. 6. Design showing facilities for entering events/programs related to learning (in bahasa format).

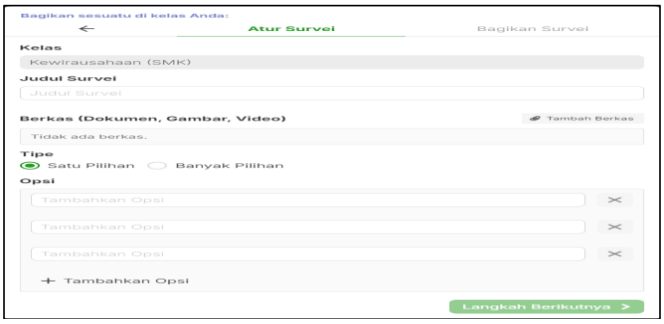


Fig. 7. Design showing the facility for entering questionnaire items for a survey (in bahasa format).

Fig. 8. Design showing the facility to enter comments (in bahasa format).

4) The result of the initial trial stage: Based on the initial design shown in Fig. 1 to 8, an initial trial of the design was conducted. 104 respondents participated in the initial trials. The results of the initial trial can be seen in Table VI.

TABLE VI. TEST RESULTS OF THE USER INTERFACE DESIGN OF THE SEVIMA EDLINK PLATFORM TO FACILITATE TRI KAYA PARISUDHA-BASED ASYNCHRONOUS LEARNING IN THE NUANCES OF INDEPENDENT LEARNING

Respondent	Items-										Σ	Percentage of Effectiveness (%)
	1	2	3	4	5	6	7	8	9	10		
Respondent-1	5	5	5	5	4	5	4	5	4	5	47	94.00
Respondent-2	4	5	4	4	4	5	5	4	5	4	44	88.00
Respondent-3	5	4	4	5	4	4	4	4	4	5	43	86.00
Respondent-4	4	4	4	5	4	5	5	5	5	5	46	92.00
Respondent-5	4	5	5	4	4	4	5	4	5	4	44	88.00
Respondent-6	4	5	5	4	5	5	4	5	4	4	45	90.00
Respondent-7	4	4	4	5	5	4	5	4	4	4	43	86.00
Respondent-8	4	4	4	4	4	5	5	4	5	5	44	88.00
Respondent-9	4	4	5	5	4	4	4	4	5	5	44	88.00
Respondent-10	4	4	4	5	4	4	4	4	4	4	41	82.00
Respondent-11	5	4	4	5	5	4	4	4	4	4	43	86.00
Respondent-12	5	5	4	4	4	4	5	4	4	5	44	88.00
Respondent-13	4	5	4	5	5	4	5	4	4	4	44	88.00
Respondent-14	4	5	4	4	5	4	5	5	4	4	44	88.00
Respondent-15	4	4	5	5	4	5	4	5	5	4	45	90.00
Respondent-16	4	4	4	4	5	5	4	4	4	4	42	84.00
Respondent-17	4	5	5	4	4	4	4	5	4	5	44	88.00
Respondent-18	4	4	5	4	4	4	4	5	4	5	43	86.00
Respondent-19	4	4	5	5	4	4	4	4	5	4	43	86.00
Respondent-20	5	4	4	4	4	5	4	4	4	5	43	86.00
Respondent-21	5	4	5	5	4	5	4	5	5	5	47	94.00
Respondent-22	5	4	4	5	4	5	5	4	5	4	45	90.00
Respondent-23	4	5	5	4	4	5	5	4	5	5	46	92.00
Respondent-24	4	5	4	4	5	4	4	4	4	4	42	84.00
Respondent-25	5	4	5	5	4	5	5	4	4	5	46	92.00
Respondent-26	4	4	4	4	4	5	5	4	4	5	43	86.00
Respondent-27	5	4	4	4	4	4	4	5	4	5	43	86.00
Respondent-28	5	5	4	4	4	4	4	4	5	4	43	86.00
Respondent-29	4	4	4	4	4	4	5	4	5	4	42	84.00
Respondent-30	4	5	5	4	5	5	4	5	4	5	46	92.00
Respondent-31	4	5	5	4	5	5	4	5	5	4	46	92.00
Respondent-32	4	4	4	4	4	4	5	4	4	5	42	84.00
Respondent-33	4	4	4	4	4	4	4	5	5	5	43	86.00
Respondent-34	4	4	5	4	4	5	5	5	4	4	44	88.00
Respondent-35	4	4	4	4	4	4	5	4	5	4	42	84.00
Respondent-36	5	4	4	5	4	4	5	5	4	4	44	88.00
Respondent-37	5	5	4	5	5	4	4	4	4	5	45	90.00
Respondent-38	4	4	4	4	4	4	4	5	4	5	42	84.00
Respondent-39	4	5	5	4	5	4	5	4	5	4	45	90.00
Respondent-40	4	4	5	4	5	5	4	4	5	4	44	88.00
Respondent-41	5	5	4	4	5	5	4	4	4	5	45	90.00
Respondent-42	5	4	4	5	4	4	4	4	4	4	42	84.00
Respondent-43	4	5	5	4	5	5	4	4	4	4	44	88.00
Respondent-44	4	4	4	4	5	5	4	5	5	4	44	88.00
Respondent-45	4	4	4	4	4	4	5	5	5	4	43	86.00
Respondent-46	5	4	4	4	4	4	4	4	4	4	41	82.00
Respondent-47	4	4	4	4	4	5	4	4	4	5	42	84.00
Respondent-48	5	5	4	5	5	4	5	4	5	5	47	94.00
Respondent-49	5	5	4	5	5	4	5	4	4	4	45	90.00
Respondent-50	4	4	4	4	4	5	4	4	4	5	42	84.00
Respondent-51	4	4	4	4	4	4	5	5	4	5	43	86.00
Respondent-52	4	5	4	4	5	5	4	5	4	4	44	88.00
Respondent-53	4	4	4	4	4	5	4	4	4	4	41	82.00
Respondent-54	4	4	5	4	4	5	4	4	4	4	42	84.00
Respondent-55	5	4	5	5	4	4	4	4	5	5	45	90.00
Respondent-56	4	4	5	4	5	5	4	4	5	5	45	90.00
Respondent-57	5	4	4	5	4	5	5	4	4	4	44	88.00
Respondent-58	5	4	5	4	5	4	5	5	4	4	45	90.00
Respondent-59	4	4	4	4	4	4	4	5	4	5	42	84.00
Respondent-60	4	4	4	4	5	5	4	4	4	4	42	84.00
Respondent-61	5	4	4	5	4	4	4	4	4	4	42	84.00
Respondent-62	4	4	4	4	4	4	4	4	5	4	41	82.00
Respondent-63	4	5	4	4	4	4	4	4	4	4	41	82.00
Respondent-64	4	5	5	4	4	4	5	4	4	4	43	86.00
Respondent-65	5	4	5	4	5	4	4	5	4	5	45	90.00
Respondent-66	5	5	4	4	5	4	5	4	4	5	45	90.00

Respondent	Items-										Σ	Percentage of Effectiveness (%)
	1	2	3	4	5	6	7	8	9	10		
Respondent-67	5	5	4	4	4	4	4	4	5	4	43	86.00
Respondent-68	4	4	4	4	4	4	4	4	4	5	41	82.00
Respondent-69	4	5	5	4	5	4	4	5	5	5	46	92.00
Respondent-70	4	5	5	4	4	4	4	4	5	4	43	86.00
Respondent-71	4	4	4	5	4	5	4	4	5	5	44	88.00
Respondent-72	4	4	4	4	4	5	5	4	4	4	42	84.00
Respondent-73	4	4	5	4	5	4	5	4	5	4	44	88.00
Respondent-74	5	5	4	5	5	5	4	4	5	4	46	92.00
Respondent-75	5	5	4	5	5	5	4	4	4	4	45	90.00
Respondent-76	4	4	5	4	4	4	4	4	4	5	42	84.00
Respondent-77	4	4	4	5	5	5	4	4	4	4	43	86.00
Respondent-78	4	5	5	5	5	5	4	5	5	4	47	94.00
Respondent-79	4	4	5	4	4	4	5	5	5	4	44	88.00
Respondent-80	4	5	5	4	4	4	4	4	4	4	42	84.00
Respondent-81	4	4	5	4	4	5	4	4	4	4	42	84.00
Respondent-82	5	5	4	4	5	4	5	4	5	4	45	90.00
Respondent-83	5	5	4	5	5	4	5	4	4	4	45	90.00
Respondent-84	4	4	4	4	4	5	4	4	4	4	41	82.00
Respondent-85	4	4	4	4	4	4	5	5	4	5	43	86.00
Respondent-86	4	5	4	4	5	5	5	4	4	5	45	90.00
Respondent-87	4	4	4	4	4	5	4	5	5	4	43	86.00
Respondent-88	4	4	5	4	5	5	4	5	4	4	44	88.00
Respondent-89	5	4	5	4	4	5	4	5	5	4	45	90.00
Respondent-90	4	5	5	5	4	4	4	5	5	4	45	90.00
Respondent-91	5	5	5	5	4	4	5	4	4	4	45	90.00
Respondent-92	5	4	4	4	5	5	4	5	5	4	45	90.00
Respondent-93	5	4	4	4	4	4	4	5	5	4	43	86.00
Respondent-94	5	4	4	5	4	4	4	4	4	5	43	86.00
Respondent-95	4	4	5	4	5	4	4	4	4	4	42	84.00
Respondent-96	4	5	5	4	5	4	4	4	5	4	44	88.00
Respondent-97	4	4	4	5	4	4	5	5	4	5	44	88.00
Respondent-98	4	4	4	4	5	4	5	5	4	5	44	88.00
Respondent-99	4	4	5	5	5	4	4	4	5	4	44	88.00
Respondent-100	4	4	4	5	4	4	4	4	4	5	42	84.00
Respondent-101	5	4	5	5	4	4	4	5	5	5	46	92.00
Respondent-102	5	5	4	4	4	5	4	4	5	4	44	88.00
Respondent-103	5	4	4	4	4	4	4	4	5	5	43	86.00
Respondent-104	5	4	4	4	5	5	5	4	4	4	44	88.00
Average												**87.38**

Respondents made several suggestions during the initial trial of the user interface design of the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning. These suggestions, which can be seen in Table VII, were used to improve the user interface design.

5) Revision stage of the initial trial results: Referring to the respondents' suggestions in Table VII, it was necessary to revise the user interface design of the SEVIMA EdLink Platform to facilitate Tri Kaya Parisudha-based asynchronous learning. Three research teams made revisions. The revised user interface design can be seen in Fig. 9 to Fig. 14.

TABLE VII. SUGGESTIONS FROM RESPONDENTS WHO WERE GIVEN ON THE INITIAL PILOT TEST

No	Respondents	Suggestions
1	Respondent-10	Show the facility to create *Manacika*-based test questions in detail
2	Respondent-16	Demonstrate the feature to set question weight and the minimum passing score for *Manacika*-based test
3	Respondent-24	Show the facility to create *Wacika*-based tasks in detail
4	Respondent-42	Indicate the existence of a feature to randomize test questions based on *Manacika.*
5	Respondent-50	Show the facility to create *Kayika*-based tasks in detail
6	Respondent-63	Show the feature to set a time limit for *Manacika*-based test questions
7	Respondent-84	Demonstrate the feature of setting question weights on tests to measure students' cognitive abilities
8	Respondent-95	Show the feature to randomize cognitive test questions

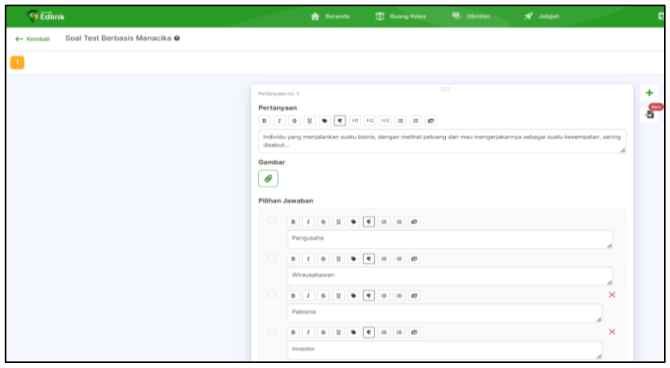


Fig. 9. Design showing the facilities for creating manacika-based test questions in detail (in bahasa format).

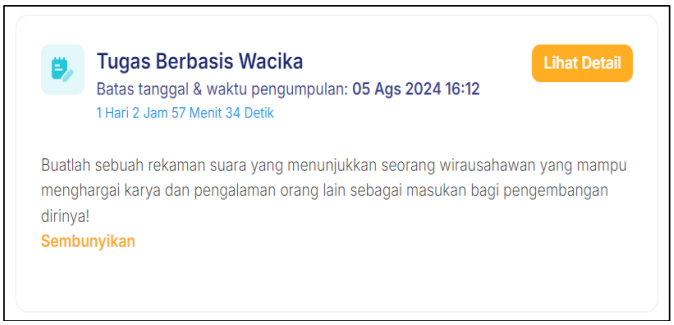


Fig. 10. Design showing the facilities for creating wacika-based tasks in detail (in bahasa format).

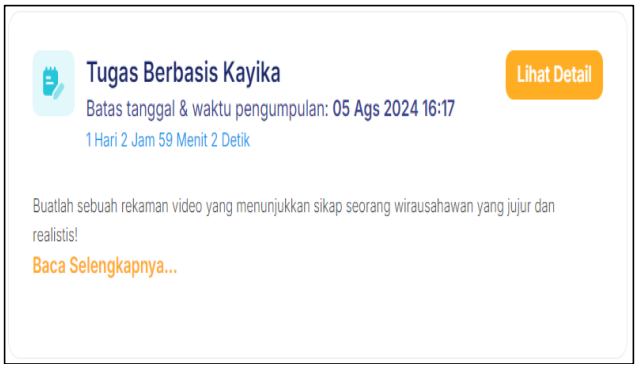


Fig. 11. Design showing the facilities for creating kayika-based tasks in detail (in bahasa format).

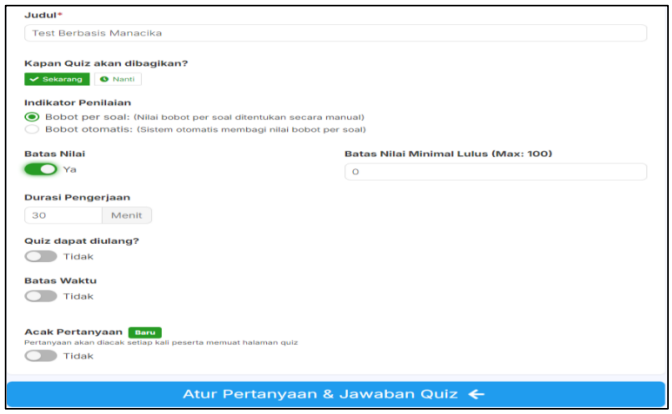


Fig. 12. Design showing features for setting question weights and minimum passing limits for manacika-based tests (in bahasa format).

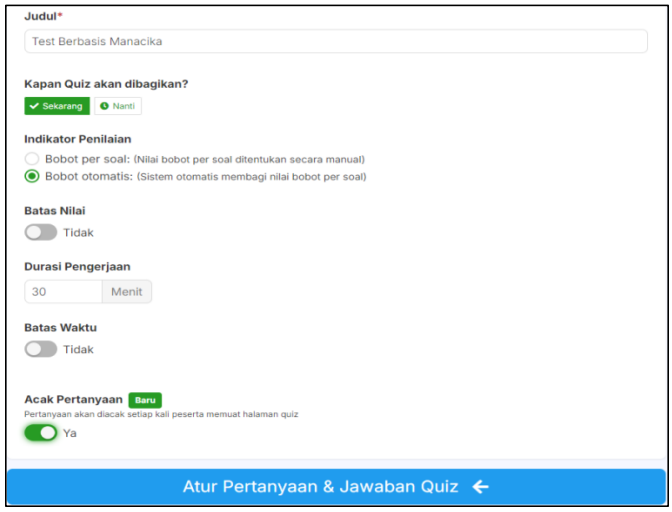


Fig. 13. Design showing the existence of a feature to randomize manacika-based test questions (in bahasa format).

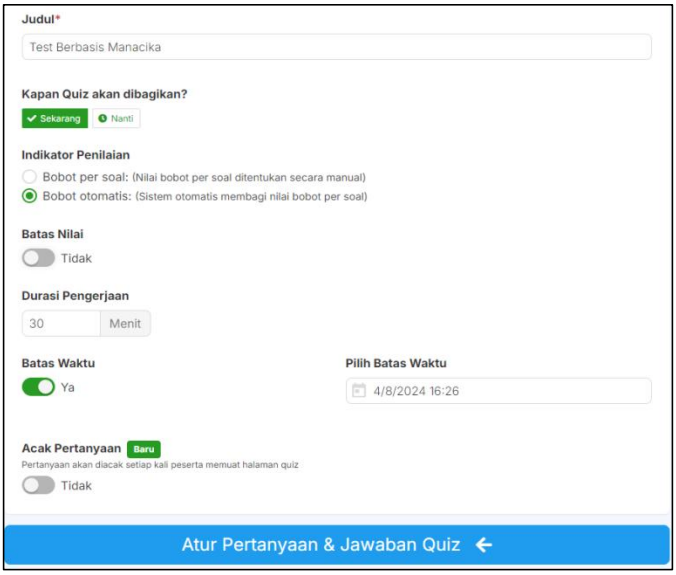


Fig. 14. Design showing features to set time limits for completing test questions based on manacika (in bahasa format).

B. Discussion

The effectiveness of the SEVIMA EdLink platform user interface design to facilitate Tri Kaya Parisudha-based asynchronous learning was included in the category of good effectiveness level. This was because the average percentage of effectiveness was 87.38%. When referring to the effectiveness categorization standard, which refers to a five-scale in Table I, it was evident that the average percentage of effectiveness was included in the good category. The underlying reason for testing the effectiveness of the SEVIMA EdLink platform user interface design is to show developers and users on a wider scale that the design has been tested and is ready to be developed into the application product creation stage.

There were ten questions used in the initial trial of the user interface design of the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning. Item-1 is related to the suitability of the design form that shows the

facility to create a class. Item-2 is related to the suitability of the design form which indicates the presence of facilities for entering the subject matter. Item-3 is related to the suitability of the design form that shows the facilities for entering Wacika and Kayika concept-based tasks. Item-4 is related to the suitability of the design form that shows the presence of facilities for entering learning information/news. Item-5 is related to the suitability of the design form that shows the facility to create Manacika concept-based test questions. Item-6 is related to the suitability of the design form that shows the facility to enter events/programs related to learning. Item-7 is related to the suitability of the design form that shows the facility to enter survey question items. Item-8 is related to the suitability of the design that showed the facility to enter comments. Item-9 is related to the ease of use of the SEVIMA EdLink platform in realizing Tri Kaya Parisudha-based asynchronous learning in the nuances of independent learning. Item-10 is related to data storage security in the SEVIMA EdLink platform to realize Tri Kaya Parisudha-based asynchronous learning in the nuances of independent learning.

The entire user interface of the SEVIMA EdLink platform, which facilitates Tri Kaya Parisudha-based asynchronous learning, uses the Indonesian language format. This is because the SEVIMA EdLink platform is an online learning platform developed by one of the limited liability company in Indonesia, namely PT. Sentra Vidya Utama (SEVIMA). This platform is intended to support the online learning process in Indonesia.

Fig. 9 is the revised user interface design of the SEVIMA EdLink Platform to facilitate Tri Kaya Parisudha-based asynchronous learning to follow up on the suggestion from respondent 10. Fig. 9 shows the facility to create Manacika-based test questions in detail. Figure 10 is the revised user interface design of the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning to follow up on the suggestions from respondent 24. Fig. 10 shows the facility to create Wacika-based tasks in detail. Fig. 11 is the revised user interface design of the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning to follow up on the suggestions from respondent 50. Fig. 11 shows the facility to create Kayika-based tasks in detail. Fig. 12 is the revised user interface design of the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning to follow up on the suggestions from respondent-16 and respondent-84. Fig. 12 shows the ability to set question weights and minimum passing scores for Manacika-based tests to measure students' cognitive abilities. Fig. 13 is the revised user interface design of the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning to follow up on the suggestions from respondent-42 and respondent-95. Fig. 13 shows the existence of a feature to randomize test questions based on Manacika in order to measure students' cognitive abilities. Fig. 14 is the revised user interface design of the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning to follow up on the suggestion from respondent-63. Fig. 14 shows the presence of a feature to set a time limit for tests based on Manacika.

This research has been able to answer the limitations of Utomo & Ahsanah' research [6], research by Soesanto et al. [7], research by Sela et al. [8], and Anggraeni et al.'s research. [9], by showing the user interface design of an online learning platform called SEVIMA EdLink, which is integrated with asynchronous learning strategies, independent learning policies, and Balinese local wisdom. This platform is appropriate for use as an asynchronous learning facility to improve student learning outcomes in the cognitive, affective, and psychomotor domains. This research in principle has the same characteristics, objectives, and concepts as research by Arvianti & Wahyuni [27], research by Fajri & Saputri [28], research by Hikmawati et al. [29], research by Ramdhani et al. [30], research by Supiatman et al. [31], and research by Yulianto et al. [32], which internalized the concept of local wisdom into the online learning process.

The novelty of this research is in the form of an innovative learning platform user interface design that combines the SEVIMA EdLink platform, asynchronous learning strategy, independent learning policy, and Balinese local wisdom, namely Tri Kaya Parisudha. This user interface design shows the SEVIMA EdLink platform used in facilitating asynchronous learning based on Tri Kaya Parisudha in the nuances of independent learning. Therefore, the presence of this platform can provide practical impacts that support the improvement of learning outcomes of Tourism Vocational School students in Bali in the cognitive (knowledge), affective (character), and psychomotor (skills) domains in the nuances of independent learning.

Although this research has novelty, but in reality this research also has limitations. The limitation of this research is that it has not been shown in detail how to operate the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning. This is because the research is focused only on the design of the user interface.

V. Conclusion

In general, the results of this research have produced a user interface design for the SEVIMA EdLink platform used in facilitating asynchronous learning based on Tri Kaya Parisudha in nuances independent learning with good quality. This user interface design shows the visualization form of Sevima EdLink learning platform, asynchronous learning strategy, independent learning policy, and Balinese local wisdom, namely Tri Kaya Parisudha, to give birth to a unique and distinctive form of learning platform in order to improve the learning outcomes of Tourism Vocational School students in Bali in the cognitive, affective, and psychomotor domains. The future work that needs to be done to overcome the obstacles of this research is to show how to operate the SEVIMA EdLink platform to facilitate Tri Kaya Parisudha-based asynchronous learning at the Tourism Vocational School in Bali. The impact of this research on stakeholders in the field of education is the existence of new information related to the existence of an online learning platform called SEVIMA EdLink, which is integrated with an asynchronous learning strategy, independent learning policy, and Balinese local wisdom. Educational stakeholders can use this online learning platform to present an asynchronous learning process based on Balinese local culture to improve the learning outcomes of Tourism Vocational

School students in Bali in the cognitive, affective, and psychomotor domains.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the Directorate General of Research and Development, Ministry of Education, Culture, Research and Technology of the Republic of Indonesia, who had to provide this research funding. This research was funded and completed on time based on research grant with main contract number: 081/E5/PG.02.00.PL/2024 and derivative contract number: 343/UN48.16/LT/2024.

REFERENCES

[1] S. Fabriz, J. Mendzheritskaya, and S. Stehle, "Impact of synchronous and asynchronous settings of online teaching and learning in higher education on students' learning experience during covid-19," *Frontiers in Psychology*, Vol. 12, pp. 1–16, 2021, doi: 10.3389/fpsyg.2021.733554.

[2] A. N. Trisna, and Jismulatif, "Implementation of edlink as e-learning media in teaching-learning online process during covid-19 at english department, universitas lancang kuning," *Talenta Conference Series: Local Wisdom, Social, and Arts (LWSA)*, vol. 7, no. 2, pp. 1–11, 2024, doi: 10.32734/lwsa.v7i2.2042.

[3] M. Suswandari, "The Influence of the discipline of online assignment collection assisted by edlink sevima on the learning outcomes," *Journal of Education Technology*, vol. 5, no. 4, pp. 596–602, 2021, doi: 0.23887/jet.v5i4.41015.

[4] I. G. Suwindia, and N. N. K. Wati, "The effect of moderate leadership based on Tri Kaya Parisudha and social attitude on employees' performance," *Vidyottama Sanatana: International Journal of Hindu Science and Religious Studies*, vol. 5, no. 2, pp. 255–263, 2021, doi: 10.25078/ijhsrs.v5i2.3040.

[5] N. N. K. Wati, and I. G. Suwindia, "The value of Tri Kaya Parisudha local wisdom as a form of religious education transformation," *International Journal of Interreligious and Intercultural Studies*, vol. 5, no. 2, pp. 85–97, 2022, doi: 10.32795/ijiis.vol5.iss2.2022.3021.

[6] D. T. P. Utomo, and F. Ahsanah, "The implementation of bichronous online learning: EFL students' perceptions and challenges," *Journal of English Language Teaching*, vol. 11, no. 2, pp. 134–147, 2022, doi: 10.15294/elt.v11i2.54273.

[7] R. H. Soesanto, J. E. Bermuli, and B. Mumu, "Implementation of blended learning models during the pandemic: A perception of prospective teachers," *Journal of Education and Technology*, vol. 5, no. 4, pp. 875–886, 2022, doi: 10.29062/edu.v5i4.362.

[8] O. Sela, F. Azhar, and U. Samanhudi, "Asynchronous learning model (its implementation via google classroom)," *ELT-Lectura*, vol. 9, no. 2, pp. 227–242, 2022, doi: 10.31849/elt-lectura.v9i2.11041.

[9] D. Anggraeni, L. A. Zahra, and R. A. Shoheh, "Pembelajaran blended learning berbasis schoology pada mata kuliah pendidikan agama Islam," *TARBAWY: Indonesian Journal of Islamic Education*, vol. 7, no. 1, pp. 56–69, 2020, doi: 10.17509/t.v7i1.21735.

[10] S. Papadakis, "MOOCs 2012-2022: An overview," *Advances in Mobile Learning Educational Research*," vol. 3, no. 1, pp. 682–693, 2023, doi: 10.25082/AMLER.2023.01.017.

[11] S. Papadakis, A. E. Kiv, H. M. Kravtsov, V. V. Osadchyi, M. V. Marienko, O. P. Pinchuk, M. P. Shyshkina, O. M. Sokolyuk, I. S. Mintii, T. A. Vakaliuk, L. E. Azarova, L. S. Kolgatina, S. M. Amelina, N. P. Volkova, V. Y. Velychko, A. M. Striuk, and S. O. Semerikov, "Unlocking the power of synergy: the joint force of cloud technologies and augmented reality in education," *In Proceedings of the 10th Workshop on Cloud Technologies in Education (CTE 2021) and 5th International Workshop on Augmented Reality in Education (AREdu 2022)*, Ukraine, on May 23, pp. 1–23, 2022.

[12] E. Faridah, I. Kasih, S. Nugroho, and T. Aji, "The effectiveness of blended learning model on rhythmic activity courses based on complementary work patterns," *International Journal of Education in Mathematics, Science and Technology*, vol. 10, no. 4, pp. 918–934, 2022, doi: 10.46328/ijemst.2618.

[13] Muhlis, A. Raksun, I. P. Artayasa, G. Hadiprayitno, and A. Sukri, "Developing context-based teaching materials and their effects on students' scientific literacy skills," *Pegem Journal of Education and Instruction*, vol. 14, no. 1, pp. 226–233, 2024, doi: 10.47750/pegegog.14.01.25.

[14] K. Rusmulyani, I. M. Yudana, I. N. Natajaya, and D. G. H. Divayana, "E-Evaluation based on CSE-UCLA model refers to glickman pattern for evaluating the leadership training program," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 279–294, 2022, doi: 10.14569/IJACSA.2022.0130534.

[15] T. Wulandari, A. Widiastuti, Nasiwan, J. Setiawan, M. R. Fadli, and Hadisaputra, "Development of learning models for inculcating Pancasila values," *International Journal of Evaluation and Research in Education*, vol. 12, no. 3, pp. 1364–1374, 2023, doi: 10.11591/ijere.v12i3.25687.

[16] D. G. H. Divayana, P. W. A. Suyasa, and A. Adiarta, "Improvement of experts' weights based on Tat Twam Asi in the TOPSIS method as a supporting parameter for optimization of blended learning evaluation results," *AIP Conference Proceedings*, vol. 2590, no. 1, pp. 1–7, 2023, doi: 10.1063/5.0106205.

[17] C. Timbi-Sisalima, M. Sánchez-Gordón, J. R. Hilera-Gonzalez, and S. Otón-Tortosa, "Quality assurance in e-learning: A proposal from accessibility to sustainability," *Sustainability*, vol. 14, no. 5, pp. 1–27, 2022, doi: 10.3390/su14053052.

[18] D. G. H. Divayana, I. G. Sudirtha, and I. K. Suartama, "Digital test instruments based on wondershare-superitem for supporting distance learning implementation of assessment course," *International Journal of Instruction*, vol. 14, no. 4, pp. 945–964, 2021, doi: 10.29333/iji.2021.14454a.

[19] N. Wijana, N. N. Parmithi, I. G. A. Wesnawa, I. M. Ardana, I. W. E. Mahendra and D. G. H. Divayana, "The Measurement of Rare Plants Learning Media using Backward Chaining Integrated with Context-Input-Process-Product Evaluation Model based on Mobile Technology," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 8, pp. 265–277, 2018, doi: 10.14569/IJACSA.2018.090834.

[20] D. G. H. Divayana, A. Adiarta, and P. W. A. Suyasa, "Implementation of Discrepancy Evaluation Application Based on TOPSIS-TTA," *TEM Journal*, vol. 12, no. 4, pp. 2613–2624, 2023, doi: 10.18421/TEM124-73.

[21] D. G. H. Divayana, P. W. A. Suyasa, and I. B. G. S. Abadi "Physical design development for evaluate digital library application based on modified CSE-UCLA with weighted product," *The 3rd Annual Applied Science and Engineering Conference (AASEC 2018)*, Bandung, Indonesia, MATEC Web of Conferences, Vol. 197, pp. 1–6, doi: 10.1051/matecconf/201819715003.

[22] F. S. Cakir, and Z. Adiguzel, "Analysis of leader effectiveness in organization and knowledge sharing behavior on employees and organization," *SAGE Open*, vol. 10, no. 1, pp. 1–14, 2020, doi: 10.1177/2158244020914634.

[23] R. Firmansyah, D. M. Putri, M. G. S. Wicaksono, S. F., A. A. Putri, Widianto, and M. R. Palil, "Educational transformation: An evaluation of online learning due to covid-19," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 7, pp. 61–76, 2021, doi: 10.3991/ijet.v16i07.21201.

[24] J. McGowan, B. Attal, I. Kuhn, L. Hinton, T. Draycott, G. P. Martin, and M. Dixon-Woods, "Quality and reporting of large-scale improvement programmes: A review of maternity initiatives in the english NHS, 2010–2023," *BMJ Quality & Safety*, vol. 2023, no. 1, pp. 1–12, 2023, doi: 10.1136/bmjqs-2023-016606.

[25] N. M. Ratminingsih, L. P. P. Mahadewi, and D. G. H. Divayana, "ICT-Based Interactive Game in TEYL: Teachers' Perception, Students' Motivation, and Achievement," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, no. 9, pp. 190–203, 2018, doi: 10.3991/ijet.v13i09.8170.

[26] D. G. H. Divayana, "User Interface Design of CIPP-WP Evaluation Application Based on Delphi," *2022 International Conference on Assessment and Learning (ICAL)*, Bali, Indonesia, pp. 1–5, 2022, doi: 10.1109/ICAL50372.2022.10075604.

[27] I. Arvianti, and A. Wahyuni, "The effectiveness of local wisdom-based integrative thematic english education games in 2013 curriculum," *Parole: Journal of Linguistics and Education*, vol. 10, no. 1, pp. 62–71, 2020, doi: 10.14710/parole.v10i1.62-71.

[28] D. R. Fajri, and S. W. Saputri, "Development of english learning model based on local wisdom with blended learning approach using macromedia flash at SMPN Satu Atap Tunda Island," *JOLLT Journal of Languages and Language Teaching*, vol. 9, no. 4, pp. 422–431, 2021, doi: 10.33394/jollt.v9i4.4098.

[29] Hikmawati, I. W. Suastra, K. Suma, and A. A. I. A. R. Sudiatmika, "Online lectures with local wisdom context: efforts to develop students' higher-order thinking skills," *International Journal of Evaluation and Research in Education (IJERE)*, vol. 13, no. 2, pp. 943–951, 2024, doi: 10.11591/ijere.v13i2.25744.

[30] I. S. Ramdhani, A. Muhyidin, and S. Hidayat, "The effectiveness of android teaching materials based on local wisdom in improving students' writing skill," *IJORER: International Journal of Recent Educational Research*, vol. 5, no. 3, pp. 549–560, 2024, doi: 10.46245/ijorer.v5i3.588.

[31] L. Supiatman, Y. Aryni, P. L. P. Sari, M. O. Siahaan, R. A. Nasution, and D. M. Sari, "Mapping of integrated local wisdom to develop instructional material," *Journal of English Language and Education*, vol. 8, no. 2, pp. 212–223, 2023, doi: 10.31004/jele.v8i2.455.

[32] F. Yulianto, Winarno, and M. Indriayu, "Audiovisual learning media based on local wisdom values of the baduy tribe community to grow student character," *International Journal of Elementary Education*, vol. 7, no. 1, pp. 43–53, 2023, doi: 10.23887/ijee.v7i1.54930.

Deep Learning-Optimized CLAHE for Contrast and Color Enhancement in Suzhou Garden Images

Chuanyuan Li[1], Ziyun Jiao[2]
School of Architecture, Sanjiang University, Nanjing, Jiangsu 210000, China[1]
School of Architecture, Sanjiang University, Nanjing, Jiangsu 210000, China[2]

***Abstract*—Suzhou gardens are renowned for their unique color palettes and rich cultural significance. This study introduces a deep learning-optimized Contrast Limited Adaptive Histogram Equalization (CLAHE) method to enhance image contrast and improve color extraction accuracy in Suzhou garden images. An initial collection of 18,502 images was refined to 11,526 high-quality images from a single dataset. A pre-trained VGG16 convolutional neural network was used to extract image features, which were then employed to dynamically optimize the CLAHE parameters, thereby preserving the original color tones while enhancing contrast. The optimized CLAHE achieved significant improvements in the Structural Similarity Index (SSIM) by 24.69 percent and in the Peak Signal-to-Noise Ratio (PSNR) by 24.36 percent, and a reduction in Loss of Edge (LOE) by 36.62 percent,compared to the standard CLAHE. Additionally, enhanced structural detail and color complexity were observed. High-Resolution Network (HRNet) was utilized for semantic segmentation, enabling precise color feature extraction. K-means clustering was used to identify key color characteristics and complementary relationships among the primary and secondary colors in Suzhou gardens. A mathematical model capturing these relationships was developed to form the basis of a color palette generator, which can be applied to digital archiving, cultural preservation, aesthetic education, and virtual reality.**

Keywords*—Deep Learning-Optimized CLAHE; image contrast enhancement; color extraction; Suzhou gardens; VGG16; semantic segmentation

I. Introduction

Suzhou gardens are widely regarded as quintessential examples of Chinese classical garden design, embodying rich cultural and historical values. Zhang and Lian [1] describe these gardens as masterpieces that harmonize architecture, water bodies, and vegetation, reflecting cultural ethos across dynasties [2]. Jiang et al. [3] emphasize their role as cultural heritage sites illustrating aesthetic principles and socio-economic changes from the Tang to Qing Dynasties. He and Chu [4] further highlight the significance of both tangible and intangible heritage values in contemporary urban development.

However, digitally preserving the visual authenticity of these heritage sites presents significant challenges. Variations in lighting conditions across different images can lead to inconsistent color representation and contrast, complicating accurate analysis and preservation efforts. Moreover, existing color extraction methods often fail to capture the full spectrum of colors inherent in the intricate designs of Suzhou gardens, resulting in incomplete analyses that do not fully reflect the gardens' aesthetic and cultural richness.

Nallaperumal et al. [5] note difficulties in maintaining visual fidelity, while Kulkarni et al. [6] observe that color extraction methods struggle to capture the diverse color ranges in complex environments like Suzhou gardens due to shading, lighting variations, and color shifts. Additionally, Chen and Gu [7] point out that many current studies emphasize individual color properties rather than conducting comprehensive color analyses, limiting our understanding of the full color spectrum that contributes to the gardens' aesthetic and cultural essence.

To address these issues, we propose an optimized Contrast Limited Adaptive Histogram Equalization (CLAHE) method enhanced by deep learning techniques. Traditional CLAHE relies on fixed parameter settings, which may not be optimal for the diverse and complex images of Suzhou gardens [8]. By leveraging VGG16, a deep convolutional neural network, we dynamically optimize CLAHE parameters—specifically the Clip Limit and Tile Grid Size—based on high-level image feature extraction [9]. This optimization allows CLAHE to adaptively enhance contrast while preserving original color tones, addressing the limitations of fixed parameter selection.

Furthermore, we incorporate High-Resolution Network (HRNet) for semantic segmentation to isolate garden-related elements from background noise, thereby refining color feature extraction [10].

This study advances digital heritage preservation and sustainable design by introducing a deep learning-optimized CLAHE method for color enhancement in Suzhou garden images. Enhanced color accuracy facilitates digital archiving and cultural preservation while promoting the integration of traditional aesthetics into sustainable design practices. These contributions provide valuable insights for eco-friendly applications in cultural conservation and modern design.

The main contributions of this study are summarized as follows:

1) Deep learning-enhanced image enhancement: We develop a novel method that dynamically optimizes CLAHE parameters using VGG16, significantly improving contrast and color preservation in Suzhou garden images.

2) Advanced semantic segmentation: We incorporate HRNet for precise segmentation of garden-related features, reducing noise and enhancing the accuracy of color extraction.

3) Color clustering analysis: Using K-means clustering in Lab color space, we identify primary and secondary color characteristics, uncovering complementary color relationships unique to Suzhou garden aesthetics.

4) Digital preservation and design applications: We create a color palette generator based on the identified color

relationships, providing practical tools for digital archiving, cultural preservation, aesthetic education, and virtual reality applications.

This paper is organized as follows: Section II presents an overview of related work, focusing on the traditional CLAHE method and advancements in semantic segmentation and color extraction techniques. Section III elaborates on the research methodology, including the optimization of CLAHE parameters using deep learning, semantic segmentation with HRNet, and color clustering. Section IV discusses the experimental setup, results, and analysis. Finally, Section V concludes with the implications of this study, its limitations, and potential future directions.

II. Overview of Related Work

A. CLAHE Color Correction Technique

Contrast Limited Adaptive Histogram Equalization (CLAHE) is an image processing technique designed to enhance local contrast, thereby improving overall image quality. Unlike conventional methods such as Global Histogram Equalization (GHE), Gamma Correction, and Retinex algorithms, which apply uniform adjustments across the entire image, CLAHE operates on small sections independently, providing optimal representation for each region with greater efficiency.

CLAHE has been widely adopted across various fields of image processing, demonstrating its effectiveness in numerous applications. For instance, Kim et al. [11] combined CLAHE with Retinex to improve color uniformity while reducing image interference, significantly enhancing image quality. In medical semantic segmentation, Nizamani et al. [12] showed that CLAHE can significantly improve segmentation accuracy. Additionally, CLAHE has been utilized as a preprocessing step in Convolutional Neural Networks (CNNs) to enhance the accuracy of identifying agricultural diseases, as demonstrated by Sayyid [13].

Further advancements in CLAHE have focused on noise suppression and detail enhancement in various imaging tasks, leading to superior performance [14]. Similarly, Soniminde and Biradar [15] applied CLAHE in multi-scale image fusion, resulting in improved image clarity and contrast. Its application has extended to enhancing complex images, as shown by Karthikha and Jamal [16]. In low-light conditions, Yuan et al. [17] utilized CLAHE to increase target detection accuracy. Ren and Xu [18] developed an improved CLAHE algorithm for enhancing dot-matrix invisible code images, effectively improving contrast in low-quality images. He et al. [19] applied CLAHE to low-illumination image processing in shield tunnels, achieving remarkable results and showcasing CLAHE's versatility in image enhancement tasks.

Despite its widespread adoption, traditional CLAHE methods are often hindered by fixed parameter settings that do not account for diverse lighting conditions and complex color distributions, particularly in culturally significant images like those of Suzhou gardens. Recent studies by Rahman et al. [20], [21], [22], [23], [24] have proposed advanced image enhancement models that adaptively address uneven illumination, noise suppression, and color preservation. These works highlight the importance of leveraging techniques like multiscale decomposition, Retinex models, and dynamic parameter optimization to handle diverse imaging scenarios effectively.

Inspired by these approaches, our study introduces a novel method that leverages VGG16, a deep convolutional neural network, to dynamically optimize CLAHE parameters. By utilizing VGG16 for high-level feature extraction, our method adjusts CLAHE's Clip Limit and Tile Grid Size based on each image's unique characteristics. This deep learning-based optimization addresses the challenge of parameter selection in CLAHE, allowing it to adaptively enhance contrast while preserving original color tones. This approach is particularly advantageous for processing complex scenes, such as those found in Suzhou gardens.

B. Semantic Segmentation and Color Extraction

High-Resolution Network (HRNet) is a robust deep neural network architecture tailored for semantic segmentation, excelling in maintaining high-resolution representations while integrating multi-scale features. Enhancements such as attention mechanisms have significantly improved HRNet's segmentation accuracy, as demonstrated by Lai [25] in breast ultrasound imaging and Jin et al. [26] in landslide segmentation from remote sensing data. Liu et al. [27] utilized HRNet's superior localization to capture intricate details in retinal vessel images by introducing deformable convolutions, while Yan et al. [9] improved semantic segmentation with a boundary detail enhancement module. Kim et al. [28] further enhanced HRNet's effectiveness across various datasets through attention modules, and Sadeghi et al. [29] confirmed HRNet's superiority in high-resolution semantic segmentation through comparative analyses.

Following segmentation, K-means clustering is employed for color extraction due to its effectiveness in identifying dominant colors by minimizing the distance between data points and cluster centroids. Zhu et al. [30] demonstrated K-means' effectiveness in cultural heritage analysis by clustering colors in traditional Yi costumes. Kristanto et al. [31] enhanced segmentation accuracy in microbial images by combining K-means with Gabor filters for texture and color feature extraction. Jardim et al. [32] utilized K-means with the watershed algorithm for complex segmentation tasks, while Abernathy and Celebi [33] developed an enhanced online K-means algorithm that improved color quantization performance through incremental processing and better initialization. Bhuvanya and Kavitha [34] integrated K-means with other clustering techniques to advance image feature extraction and classification accuracy. Additionally, Kalaivani and Vimaladevi [35] improved endmember extraction in hyperspectral images by leveraging the relationship between K-means and the Pixel Purity Index, showcasing its versatility across various image analysis domains.

Integrating optimized CLAHE for color correction, HRNet for precise semantic segmentation, and K-means for effective color clustering enhances the accuracy of color extraction. This composite approach provides a solid foundation for digitally preserving the heritage of Suzhou gardens through improved color extraction techniques, enabling detailed analysis and accurate representation of their unique color palettes.

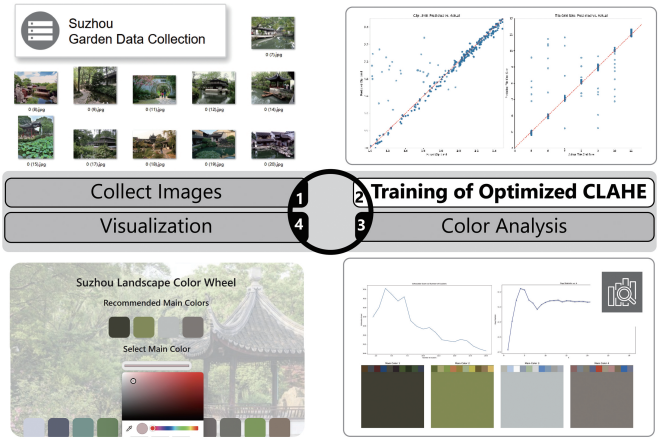


Fig. 1. Methodological roadmap.

III. Research Methodology

The methodological roadmap of this study is shown in Fig. 1.

This study integrates an optimized Contrast Limited Adaptive Histogram Equalization (CLAHE) method with a High-Resolution Network (HRNet) and K-means clustering to enhance contrast and color extraction in Suzhou garden images. To address lighting variations and incomplete color analyses, the optimized CLAHE is dynamically refined using VGG16 for effective contrast adjustment, while HRNet is employed for multi-scale semantic segmentation. K-means clustering is then applied to the segmented images to identify key color patterns and relationships.

A. Training of Optimized CLAHE

CLAHE's performance critically depends on its parameters: Clip Limit and Tile Grid Size. The Clip Limit controls maximum local contrast, reducing noise amplification, while the Tile Grid Size determines the level of local detail by segmenting the image into blocks. Traditional regression-based methods [36] have optimized these parameters; however, this study employs a deep learning-based approach for more adaptive enhancement. For the discrete Tile Grid Size, we treated optimization as a classification problem with 13 classes, corresponding to Tile Grid Sizes ranging from 8 to 32 in increments of 2. This allows the model to predict the most suitable Tile Grid Size category for each image effectively. The continuous Clip Limit is predicted using a linear activation function for fine-grained adjustments.

1) Role of VGG-16 in parameter optimization: VGG16, renowned for its deep convolutional architecture and effective feature extraction [37], is employed to capture high-level image features such as brightness, contrast, and fine-grained textures characteristic of Suzhou garden images. Leveraging pre-trained weights from extensive training on large datasets like ImageNet, VGG16 facilitates efficient transfer learning, enabling the model to generalize effectively across diverse image types while minimizing the need for extensive retraining.

VGG16 was chosen for its simplicity and proven effectiveness in feature extraction tasks. Compared to more recent architectures like ResNet or EfficientNet, VGG16 offers a balanced depth and computational efficiency, making it ideal for our parameter optimization without introducing excessive complexity. Its robust feature representation from ImageNet pre-training generalizes well to various garden images.

2) Feature extraction and parameter selection: The VGG16 model processes each image to extract high-dimensional features, which serve as inputs for optimizing CLAHE parameters. These features are passed through a global average pooling layer to reduce dimensionality, ensuring computational efficiency while retaining essential information:

$$f_{\text{VGG}} = \text{GlobalAveragePooling2D}(\text{VGG16}(x)) \tag{1}$$

Image contrast is quantified by calculating a brightness histogram and contrast ratio:

$$l_h = \frac{\text{calcHist}([Y], [0], \text{None}, [256], [0, 256])}{\sum_{i=0}^{255} \text{calcHist}([Y], [0], \text{None}, [256], [0, 256])_i} \tag{2}$$

$$\text{contrast} = \frac{\text{std}(Y)}{\text{mean}(Y)} \tag{3}$$

To enhance dataset diversity and improve model robustness, a series of data augmentation techniques were employed:

- Rotations: Random rotations within $\pm 20^\circ$ to simulate diverse perspectives and improve spatial feature recognition.
- Flipping: Horizontal and vertical flips, each applied with a probability of 50%, enhancing the model's robustness to symmetric structures.
- Brightness and Contrast Adjustments: Random changes within $\pm 20\%$ to account for lighting variations.
- Scaling: Random resizing between 80% and 120% of the original image dimensions to accommodate scale diversity.
- Noise Addition: Gaussian noise was applied to simulate real-world interference and improve noise tolerance.

These augmentation techniques were specifically designed to address the unique visual complexity and dynamic conditions of Suzhou garden images. Experimental results revealed an improvement of over 10% in validation accuracy when augmentation was applied, demonstrating its critical role in enhancing model performance.

3) Model construction and training: Fig. 2 illustrates the workflow of the model training and evaluation process. It begins with input data (images) that undergo data preparation, including train-validation splitting and data augmentation. The model initialization uses the VGG16 architecture with random weight initialization. In the training loop, processes such as optimization, forward and backpropagation, learning rate reduction, and early stopping are applied. Once trained, the

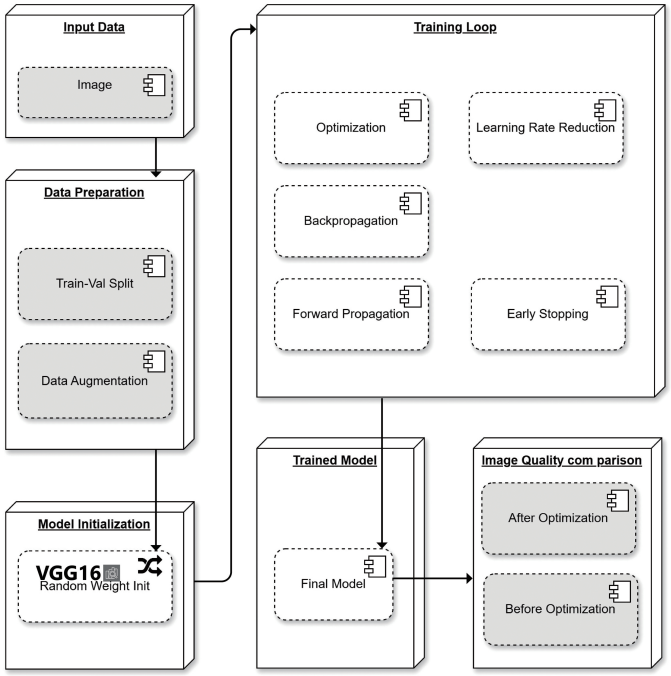


Fig. 2. Model training and evaluation workflow.

final model is evaluated by comparing image quality before and after optimization.

The model architecture comprises an input layer followed by fully connected layers to reduce the dimensionality of the feature space:

$$x_{i+1} = \text{ReLU}(\text{Dense}(x_i, \text{units} = n)) \tag{4}$$

Residual connections are incorporated to prevent gradient vanishing:

$$x_{\text{new}} = \text{add}([x_{\text{prev}}, \text{Dense}(x_{\text{prev}}, 512, \text{'relu'})]) \tag{5}$$

Using Early Stopping and ReduceLROnPlateau, the model outputs continuous predictions for optimized Clip Limit and Tile Grid Size:

$$\text{clip\_limit\_output} = \text{Dense}(1, \text{activation} =' \text{linear}')(x) \tag{6}$$

$$\text{tile\_grid\_size\_output} = \text{Dense}(13, \text{'softmax'})(x) \tag{7}$$

The neural network was trained over 200 epochs with a batch size of 64, utilizing 80% of the data for training and 20% for validation. After VGG16-based feature extraction, the network architecture included fully connected layers with batch normalization and dropout regularization to ensure stability during training. Residual connections were employed to retain

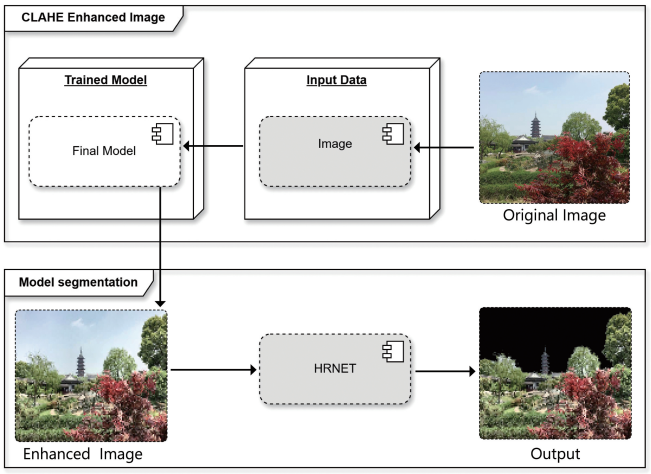


Fig. 3. HRNet flowchart.

essential information across layers, particularly at the 512-dimensional layer. This enhances gradient flow and improves learning efficiency.

The Adam optimizer was used with an initial learning rate of 0.001, which was dynamically adjusted using ReduceLROnPlateau based on the root mean squared error (RMSE) of the validation loss for the outputs. Early Stopping was applied with a patience of 20 epochs to prevent overfitting, restoring the best model weights.

B. Application of HRNet in Semantic Segmentation

HRNet is applied to CLAHE-optimized images for multi-scale feature extraction, preserving high-resolution details. Four parallel convolutional branches process features at varying scales, which are then fused for refined segmentation:

$$F_{\text{fusion}} = \sum_{i=1}^{N} \alpha_i \cdot F_i \tag{8}$$

The fused output enables precise identification of relevant garden regions, enhancing the accuracy of color feature extraction. Final segmentation is defined as:

$$Y_{\perp}^{*} = \text{argmax}(\text{softmax}(F_{\text{fusion}})) \tag{9}$$

To improve segmentation accuracy, we applied morphological operations to remove small non-garden-related elements such as sky regions and transient objects like people and cars. Additionally, connected component analysis was performed to retain only the largest contiguous garden regions, ensuring the exclusion of irrelevant areas. The HRNet flowchart is shown in Fig. 3.

C. Color Clustering and Analysis

The segmented RGB images are converted to Lab color space for better perceptual uniformity in color representation:

$$L = 116 \cdot f\left(\frac{Y}{Y_n}\right) - 16,$$
$$a = 500 \cdot \left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right),$$
$$b = 200 \cdot \left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \quad (10)$$

K-means clustering identifies primary and secondary colors, measuring distances using Euclidean metrics:

$$\text{Distance}(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \quad (11)$$

Cluster centers are updated iteratively, with Euclidean distance determining dominant color clusters:

$$\mu_j = \frac{1}{n_j}\sum_{i=1}^{n_j} x_i \quad (12)$$

The distance between primary and secondary colors is calculated as:

$$\text{distance} = \sqrt{(L_1 - L_2)^2 + (a_1 - a_2)^2 + (b_1 - b_2)^2} \quad (13)$$

Linear regression is then applied to model color relationships:

$$Y = \beta_0 + \beta_1 \cdot \text{MC } L + \beta_2 \cdot \text{MC } a + \beta_3 \cdot \text{MC } b + \beta_4 \cdot \text{SC } L + \beta_5 \cdot \text{SC } a + \beta_6 \cdot \text{SC } b \quad (14)$$

D. Validation and Visualization of Cluster Numbers

The optimal number of clusters is validated using Gap Statistic and Silhouette Score for robust clustering results. Principal Component Analysis (PCA) reduces color data dimensionality to visualize clusters effectively:

$$X_{\text{pca}} = X \cdot V_k \quad (15)$$

PCA-reduced color data undergoes K-means clustering to form primary and secondary color clusters, providing a structured analysis of color distributions in Suzhou garden images.

IV. Experiments and Results

A. Experimental Setup

1) Optimized CLAHE processing:

a) Suzhou garden data collection: An automated function was developed to collect images from the internet, specifically targeting names associated with famous classical gardens or landmarks in Suzhou. The selection criteria required that these sites be recognized as World Cultural Heritage sites. These include "Canglang Pavilion," "Huanxiu Mountain Villa," "Lingering Garden," "Couple's Retreat Garden," "Lion Grove Garden," "Retreat and Reflection Garden," "Garden of the Master of the Nets," "Art Garden," and "Humble Administrator's Garden." Additionally, the keywords "Classical gardens of Suzhou," "Suzhou Garden," and "Suzhou Classical Gardens" were used as supplements. Based on these search terms, images were collected from Baidu Images, resulting in a total of 18,502 images. The original image data are shown in Table I.

TABLE I. Original Image Data

Keywords	Number of Pictures
Classical Gardens of Suzhou	352
Suzhou Garden	145
Canglang Pavilion	1758
Huanxiu Mountain Villa	1263
Lingering Garden	1853
Couple's Retreat Garden	1581
Lion Grove Garden	1508
Classical Suzhou Gardens	1623
Suzhou Yuanlin	1862
Retreat and Reflection Garden	1380
Garden of the Master of the Nets	1755
Art Garden	1595
Humble Administrator's Garden	1827

An Average Hash (AHash) algorithm was employed to identify and remove duplicate images from the collected dataset by computing the hash value for each image and deleting any identical photos. The formula for calculating the average hash is:

$$\text{average\_hash} = \sum_{i=0}^{n} 2^i \cdot I(v_i > \text{mean}) \quad (16)$$

Following this, all images were manually reviewed to remove any non-Suzhou garden-related photos from the dataset. After this refinement process, the final number of preprocessed images was 11,526.

b) Training and evaluation of optimized CLAHE: The processed images of Suzhou gardens underwent necessary preprocessing steps before model training. The model was trained using the formulas provided in Eq. (1) to (7) and evaluated on a validation set. The evaluation process involved tracking changes in loss and error metrics throughout the training process and assessing regression performance using metrics such as Mean Squared Error (MSE) and R-squared (R^2).

c) Comparison of predicted and actual values: For the Clip Limit, most data points align closely with the red dashed line, indicating that the model achieves high predictive

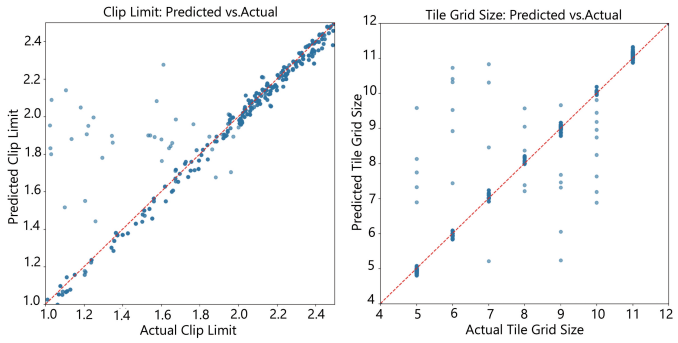


Fig. 4. Comparison of predicted value and actual value.

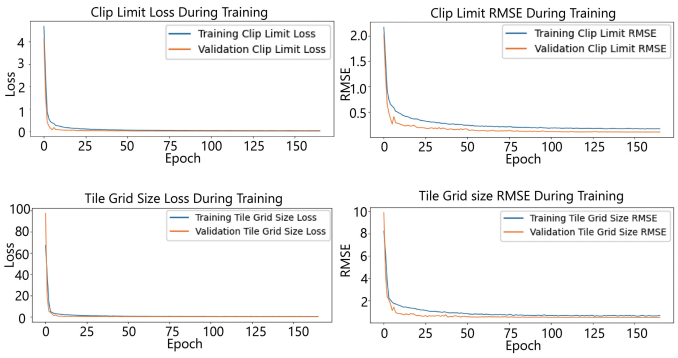


Fig. 5. Training history chart.

accuracy in the majority of cases. A few data points exhibit significant deviations, suggesting that the model may struggle with predictions under certain extreme conditions or specific scenarios. For the Tile Grid Size, the predicted values generally follow the trend of the actual values, though some deviations are noted at higher grid sizes. Despite these outliers, they are relatively rare, indicating that the model's predictions are accurate in most situations. The analysis data is shown in Fig. 4.

d) Training history chart analysis: The training and validation loss, along with Root Mean Squared Error (RMSE), decreased rapidly during the initial stages of training and gradually stabilized, indicating that the model effectively learned the data features early on and converged progressively. The application of data augmentation techniques further contributed to minimizing overfitting and improving generalization. Notably, rotations and flips enhanced the model's ability to generalize across diverse spatial features, while brightness and contrast adjustments improved its robustness to varying lighting conditions. The Training History information is shown in Fig. 5.

The model demonstrates strong performance in predicting the clip_limit parameter, achieving minimal prediction error and exhibiting a concentrated distribution of predictions, reflecting its robust generalization capability. While the prediction of the tile_grid_size parameter is generally good, it exhibits some limitations at larger grid sizes, suggesting areas for potential optimization in future work.

e) Comparison of CLAHE and optimized CLAHE: To objectively evaluate the performance of the optimized CLAHE method against the standard CLAHE, we applied three key image quality metrics: Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Loss of Edge (LOE). These metrics were calculated by comparing the CLAHE-enhanced images with the original images in the dataset.

SSIM measures similarity between two images by considering luminance, contrast, and structural details, where higher values indicate better structural preservation. PSNR, on the other hand, quantifies the ratio between the maximum possible signal power and the noise power; higher PSNR values imply better image quality and less noise. LOE evaluates the preservation of edge details in the images, where lower values indicate better edge retention.

The average SSIM, PSNR, and LOE values for all images were computed for both the standard CLAHE and Optimized CLAHE methods, as shown in Table II.

TABLE II. AVERAGE SSIM, PSNR, AND LOE VALUES FOR STANDARD AND OPTIMIZED CLAHE

Metric	Standard CLAHE	Optimized CLAHE	Improvement (%)
SSIM	0.4437	0.5532	+24.69%
PSNR (dB)	13.22	16.45	+24.36%
LOE	0.2314	0.1466	-36.62%

The results indicate that Optimized CLAHE achieved significant improvements over standard CLAHE, with a 24.69% increase in SSIM (from 0.4437 to 0.5532), a 24.36% improvement in PSNR (from 13.22 dB to 16.45 dB), and a 36.62% reduction in LOE (from 0.2314 to 0.1466). This demonstrates that Optimized CLAHE not only enhances contrast but also better preserves structural integrity and significantly reduces edge loss, thereby maintaining critical image details.

Fig. 6 illustrates overall performance comparisons for SSIM, PSNR, and LOE metrics. These enhanced metrics confirm that Optimized CLAHE contributes to more vivid and accurate color representation in the images, capturing the intricate details and richness of Suzhou gardens.

The evaluation results strongly suggest that the Optimized CLAHE model provides superior performance, making it a promising approach for applications requiring high-quality image enhancement.

2) HRNet Processing and Color Extraction: After applying the Optimized CLAHE processing to all images, semantic segmentation was performed using Eq. (8) and (9) to remove non-garden-related elements such as people and cars. To ensure the accuracy of color extraction, the sky was excluded before proceeding to subsequent steps.

When executing the calculations according to Eq. (10) to (12), the maximum number of clusters generated by the algorithm for each image was limited to 30. All extracted color data were converted into a DataFrame format and saved. The generated data includes the LAB values of the primary and secondary colors for each image, and Table III provides information about an image named "Image1".

Eq. (13) and (14) generate the color clustering formulas as follows:

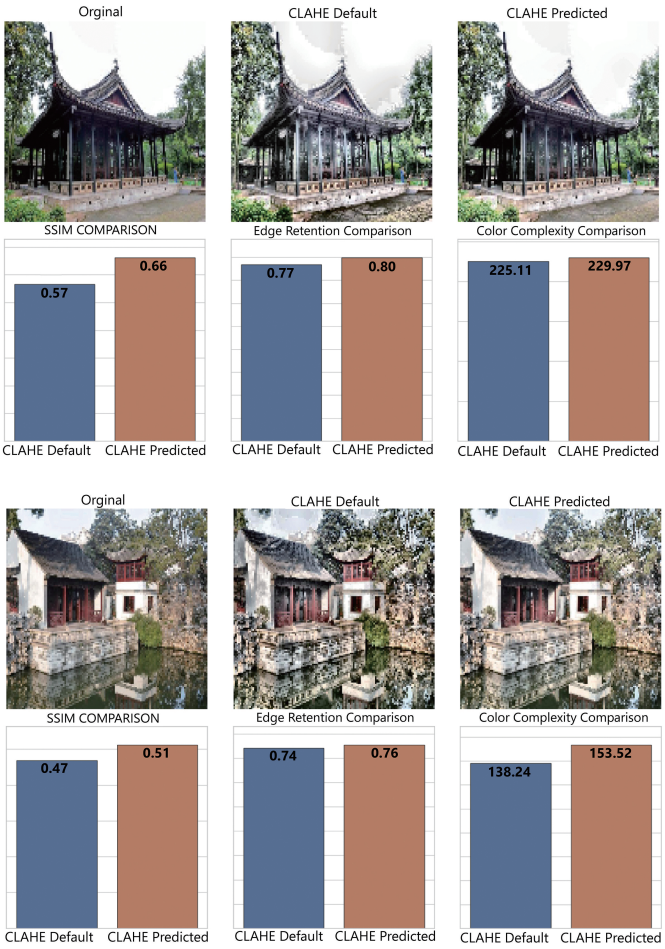


Fig. 6. Performance comparison of CLAHE and enhanced CLAHE.

TABLE III. COLOR CARD LAB VALUES - IMAGE 1

L	a	b	Ratio
34.1446	0.4313	9.8101	0.0463
87.6003	-0.2872	-10.7976	0.0334
17.2502	3.0355	2.4672	0.0717
64.0184	3.5201	14.0537	0.0256
56.8409	-23.9924	34.4455	0.0158
97.1075	-1.0001	-1.1957	0.0837
53.0612	10.1856	21.9371	0.0180
76.0309	3.2525	-0.4037	0.0474
37.9928	3.8353	-6.3526	0.0169
57.2163	2.3148	-4.1065	0.0217
46.6206	5.3030	3.2603	0.0373
86.9956	3.5583	3.1048	0.0817
44.1516	0.0378	12.3873	0.0390
40.4602	9.3195	15.2071	0.0118
55.0158	2.8220	5.9573	0.0556
68.0085	7.8041	27.5084	0.0391

$$Y = 35.78 + 0.19 \cdot \mathrm{MC}L + 0.002 \cdot \mathrm{MC}a - 0.28 \cdot \mathrm{MC}b - 0.22 \cdot \mathrm{SC}L + 0.10 \cdot \mathrm{SC}a + 0.46 \cdot \mathrm{SC}b \quad (17)$$

The range of Y is [0.0,163.32] and its data distribution is shown in the figure below. To control the range, the values between the 25th and 75th percentiles are selected as the

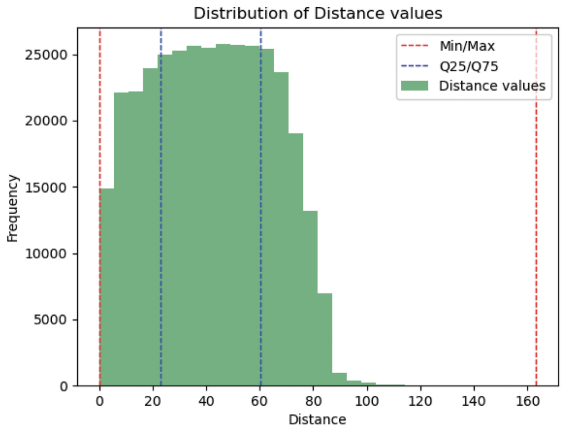


Fig. 7. Y Value range definition.

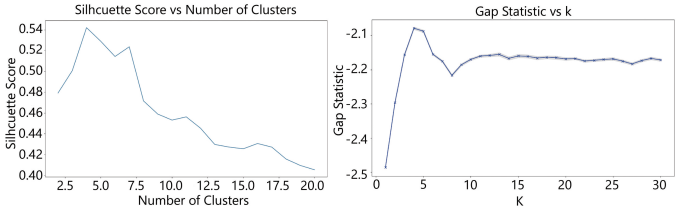


Fig. 8. Cluster quantity analysis.

controlled range for Y. The final range of Y is [22.84, 60.38]. The relevant information is shown in Fig. 7.

The Gap Statistic and Silhouette Score were compared for different numbers of clusters, revealing that the optimal number of clusters is 4. This suggests that the clustering effect is best when the number of primary colors is 4. The Cluster Quantity Analysis is illustrated in Fig. 8.

Using Eq. (15), the reduced-dimensional color data was clustered using the K-means algorithm to identify the main color cluster centers (primary colors). For each primary color, the corresponding secondary colors were further clustered using the K-means algorithm, with the cluster centers selected to represent these secondary colors. To aid in visualization, the maximum number of clusters for secondary colors corresponding to each primary color was limited to 10, ensuring clarity and manageability in the resulting visual representation. Information called "Color Card LAB Values" is shown in Tables IV, V, VI, and VII. Each support color represents a secondary color that complements the main color within its group, contributing to the overall color harmony observed in Suzhou gardens.

The final visualization image generated by this process is shown in Fig. 9.

B. Results and Analysis

Using Eq. (17), our analysis reveals that the color design of Suzhou classical gardens exhibits a complementary relationship in terms of lightness (L) and the blue-yellow channel (b). Specifically, there is a positive or negative correlation between "Main Color L" and "Support Color L" as well as between

TABLE IV. COLOR CARD LAB VALUES - GROUP 1

Color Result Group 1			
Color Type	**L**	**a**	**b**
Support Color 1	28.84	8.37	22.32
Support Color 2	30.20	-1.83	-3.60
Support Color 3	22.73	21.81	11.61
Support Color 4	26.59	4.82	-35.18
Support Color 5	29.95	-2.03	8.03
Support Color 6	15.79	4.41	-0.66
Support Color 7	33.93	-15.91	21.87
Support Color 8	19.57	-6.58	10.27
Support Color 9	32.30	-14.70	15.67
Support Color 10	26.51	0.44	-10.59
Main Color 1	26.77	-1.55	7.34

TABLE V. COLOR CARD LAB VALUES - GROUP 2

Color Result Group 2			
Color Type	**L**	**a**	**b**
Support Color 1	43.19	-14.52	31.60
Support Color 2	71.68	-8.94	37.19
Support Color 3	62.60	-24.40	47.16
Support Color 4	59.29	23.75	44.69
Support Color 5	56.44	-1.80	49.71
Support Color 6	73.83	-13.11	24.92
Support Color 7	80.02	-12.28	67.12
Support Color 8	39.95	-0.67	31.99
Support Color 9	53.12	3.85	29.67
Support Color 10	81.91	-1.46	57.86
Main Color 2	57.56	-16.58	35.54

TABLE VI. COLOR CARD LAB VALUES - GROUP 3

Color Result Group 3			
Color Type	**L**	**a**	**b**
Support Color 1	77.76	-2.90	6.01
Support Color 2	84.69	-0.38	-14.41
Support Color 3	98.18	-0.23	0.43
Support Color 4	67.34	-0.59	-8.70
Support Color 5	77.32	-12.04	17.55
Support Color 6	88.17	-0.26	-0.14
Support Color 7	62.07	10.53	-50.65
Support Color 8	66.76	-2.14	-18.09
Support Color 9	68.37	-1.52	3.36
Support Color 10	74.33	0.15	-4.63
Main Color 3	80.49	-1.61	0.24

TABLE VII. COLOR CARD LAB VALUES - GROUP 4

Color Result Group 4			
Color Type	**L**	**a**	**b**
Support Color 1	47.20	16.61	12.50
Support Color 2	54.80	-1.07	-5.90
Support Color 3	53.24	-1.60	13.72
Support Color 4	42.84	6.98	-24.07
Support Color 5	46.77	54.44	47.14
Support Color 6	65.06	-1.26	17.93
Support Color 7	63.90	17.13	11.22
Support Color 8	47.68	1.44	18.59
Support Color 9	44.59	14.93	-53.98
Support Color 10	50.82	4.07	-14.29
Main Color 4	51.47	0.99	5.99

"Main Color b" and "Support Color L." The Euclidean distance of color differences predominantly falls within the range of [22.84, 60.38], which is considered the "harmony" standard in garden color design.

This harmony is achieved through the subtle control of contrast in secondary colors, reflecting the traditional philosophy of "harmony between humans and nature" and the concept of "calm and introspection." Suzhou gardens create a rich visual experience by uniting the stability of primary colors with the diversity of secondary colors. The study also indicates that secondary colors, particularly those in the blue-yellow spectrum (such as tones of sunlight, water surfaces, and autumn leaves), significantly influence the overall visual experience and play a crucial role in determining the color difference distance (Y). In contrast, the influence of the main color's a value is found to be minimal.

Building on these findings, we developed a color palette generator program designed to automatically create color schemes that align with the design principles of Suzhou gardens. This tool provides practical applications for heritage preservation, garden design, aesthetic education, and virtual reality. The development program is shown in Fig. 10.

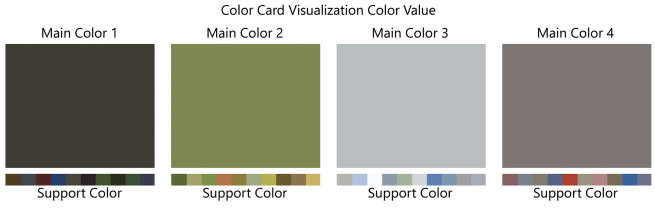


Fig. 9. Color card visualization of color value.

C. Limitations

1) Clustering method: While K-means clustering effectively identified dominant color clusters, its assumption of spherical clusters may not capture the nuanced color variations inherent in Suzhou gardens. Alternative methods like Gaussian Mixture Models (GMM) or DBSCAN could potentially model more complex color distributions.

2) Parameter optimization robustness: The deep learning-based optimization of CLAHE parameters, though effective, may require further refinement to handle extreme lighting conditions or highly complex scenes beyond the current dataset.

3) Dataset diversity: The dataset, while extensive, may still lack certain variations in garden scenes, such as seasonal changes or rare architectural elements, potentially limiting the universality of the model.

D. Future Work

Future research could address these limitations by exploring more sophisticated clustering techniques and improving the robustness of parameter optimization. Expanding the dataset to include a wider variety of garden scenes and conditions would further enhance the model's generalizability. Additionally, integrating reinforcement learning for real-time parameter adjustments and exploring end-to-end deep learning models for simultaneous enhancement and color extraction present promising directions. Applying this methodology to other culturally significant heritage sites could also validate its versatility and effectiveness across diverse contexts.

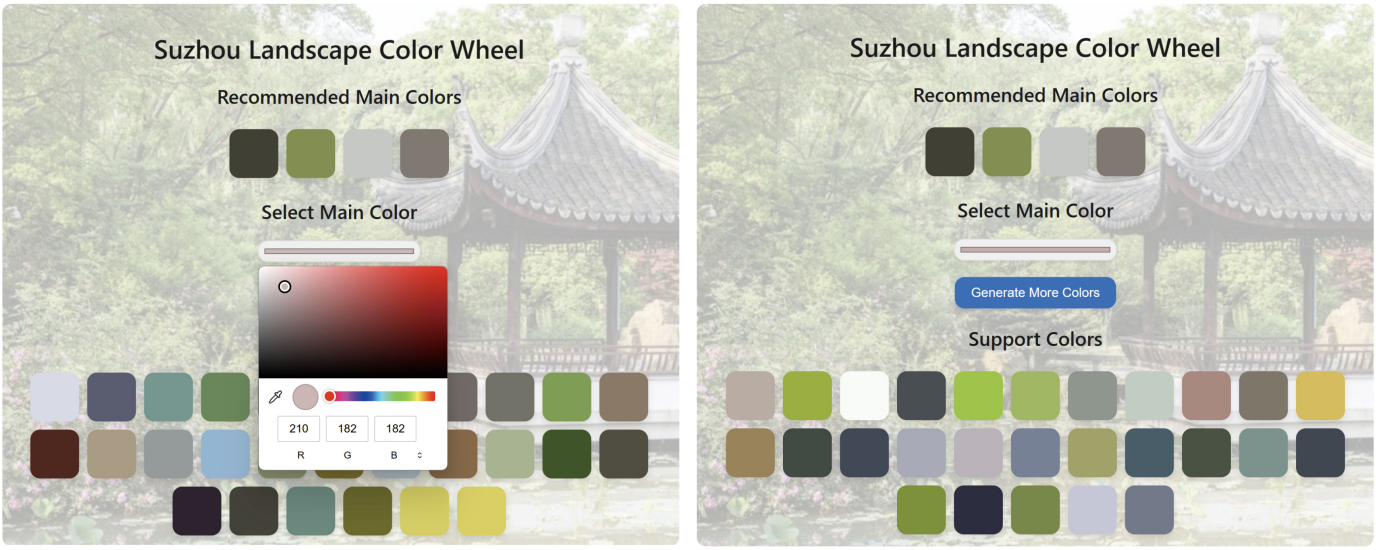


Fig. 10. Color palette generator development program

V. Conclusion

This study explored the integration of an optimized Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm for enhancing image contrast and accurately extracting colors from images of Suzhou gardens. The findings demonstrated the potential of this approach to improve the visualization of fine details, preserve structural integrity, and capture a richer spectrum of colors in heritage images. Specifically, the optimized CLAHE showed a 24.69% improvement in Structural Similarity Index (SSIM), a 24.36% increase in Peak Signal-to-Noise Ratio (PSNR), and a 36.62% reduction in Loss of Edge (LOE), alongside enhanced color complexity and edge preservation.

The use of VGG16 for dynamic parameter optimization allowed CLAHE to adapt to each image's unique characteristics, ensuring consistent enhancement across varied lighting and shading conditions. High-Resolution Network (HRNet) further refined the segmentation process, isolating garden-related elements and enhancing the accuracy of color feature extraction. K-means clustering effectively identified primary and secondary color clusters, revealing complementary color relationships that align with the traditional aesthetics of Suzhou gardens.

Despite these advancements, some limitations remain. The reliance on the K-means algorithm for color clustering, while effective for basic categorizations, may not fully capture the nuanced color variations characteristic of Suzhou gardens. Additionally, while deep learning (DL) was utilized to optimize CLAHE parameters, further refinement is needed to make this approach robust across diverse imaging scenarios. Future work could explore more sophisticated clustering techniques, such as Gaussian Mixture Models (GMMs) or advanced deep learning methods, to provide a more comprehensive understanding of color relationships in heritage contexts.

The findings of this study have several implications for future research. First, there is potential to expand the application of this method beyond Suzhou gardens to other complex imaging scenarios, such as urban landscapes, natural reserves, or historic façades. Such explorations could validate the generalizability and utility of the proposed techniques across diverse cultural heritage sites. Second, further advancements in adaptive parameter optimization for CLAHE, possibly using reinforcement learning or other machine learning approaches, could enhance the adaptability and effectiveness of image enhancement techniques in real-time applications.

Overall, this study contributes to the field of cultural heritage preservation by demonstrating the value of integrating advanced image processing techniques with deep learning algorithms to enhance the analysis and documentation of classical Chinese gardens. By addressing current limitations and exploring new research directions, future studies can further refine these methods and expand their applicability, ultimately advancing the digital preservation and understanding of cultural heritage.

This work lays a foundation for more sophisticated tools that not only preserve the aesthetic qualities of heritage sites but also enhance their accessibility and comprehensibility for future generations.

Financial Disclosure

National Natural Science Foundation Project (51408337). Universities' philosophy and Social Sciences Research Projects in Jiangsu Province (KZ2022017)

Conflict of Interest

The authors declare no potential conflict of interests.

Code and Data Availability

The code related to this study is publicly available on GitHub at: https://github.com/andrew849039/optimized_clahe

REFERENCES

[1] Y. Zhang and Z. Lian, "Research on the distribution and scale evolution of Suzhou gardens under the urbanization process from the Tang to the Qing dynasty," *Land*, vol. 10, no. 3, p. 281, 2021.

[2] X. Yi, "The tangible and intangible value of the Suzhou classical gardens," in *Proceedings of the 16th ICOMOS General Assembly and International Symposium 'Finding the Spirit of Place—Between the Tangible and the Intangible'*, p. 29, Quebec, QC, Canada, 2008.

[3] J. Jiang, et al., "Urban heritage conservation and modern urban development from the perspective of the historic urban landscape approach: A case study of Suzhou," *Land*, vol. 11, no. 8, p. 1251, 2022.

[4] Q. He and C.-H. H. Chu, "A new shadow removal method for color images," in *International Conference Paper*, 2013.

[5] K. Nallaperumal, et al., "An analysis of suitable color space for visually plausible shadow-free scene reconstruction from single image," in *2013 IEEE International Conference on Computational Intelligence and Computing Research*, 2013.

[6] K. Kulkarni, P. Patil, and S. G. Kanakaraddi, "Multi-modal colour extraction using deep learning techniques," in *2022 Fourth International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, 2022.

[7] Y. Chen and F. Gu, "Quantitative analysis of traditional Chinese color," *BioResources*, vol. 18, no. 3, 2023.

[8] S. Sahu, et al., "An approach for de-noising and contrast enhancement of retinal fundus image using CLAHE," *Optics Laser Technology*, vol. 110, pp. 87–98, 2019.

[9] G. Yan, et al., "Enhancing building segmentation in remote sensing images: Advanced multi-scale boundary refinement with mbr-HRNet," *Remote Sensing*, vol. 15, no. 15, p. 3766, 2023.

[10] J. Wang, et al., "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3349–3364, 2020.

[11] Y.-J. Kim, D.-M. Son, and S.-H. Lee, "Retinex jointed multiscale CLAHE model for HDR image tone compression," *Mathematics*, vol. 12, no. 10, p. 1541, 2024.

[12] A. H. Nizamani, et al., "Advance brain tumor segmentation using feature fusion methods with deep U-Net model with CNN for MRI data," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 9, p. 101793, 2023.

[13] M. F. N. Sayyid, "Klasifikasi penyakit daun jagung menggunakan metode CNN dengan image processing HE dan CLAHE," *Jurnal Teknik Informatika dan Teknologi Informasi*, vol. 4, no. 1, pp. 86–95, 2024.

[14] X. Zhao, et al., "Application and analysis of medical image processing based on improved CLAHE," in *International Conference on Artificial Intelligence in China*, Springer Nature Singapore, Singapore, 2023.

[15] N. V. Soniminde and M. Biradar, "Improving wavelet-based image fusion with weighted average of high boost and CLAHE," in *2024 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, IEEE, 2024.

[16] R. Karthikha and D. N. Jamal, "Enhancing colonoscopy image quality with CLAHE in the Gastrolab dataset," in *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, IEEE, 2023.

[17] Z. Yuan, et al., "CLAHE-based low-light image enhancement for robust object detection in overhead power transmission system," *IEEE Transactions on Power Delivery*, vol. 38, no. 3, pp. 2240–2243, 2023.

[18] M. Ren and J. Xu, "An improved CLAHE image enhancement algorithm for dot matrix invisible code," in *Eighth International Conference on Electronic Technology and Information Science (ICETIS 2023)*, vol. 12715, SPIE, 2023.

[19] Z. He, et al., "Multi-scale fusion for image enhancement in shield tunneling: A combined MSRCR and CLAHE approach," *Measurement Science and Technology*, vol. 35, no. 5, p. 056112, 2024.

[20] Z. Rahman, Y.-F. Pu, M. Aamir, S. Wali, and Y. Guan, "Efficient image enhancement model for correcting uneven illumination images," *IEEE Access*, vol. 8, pp. 109038–109053, 2020, doi: 10.1109/ACCESS.2020.3001206.

[21] Z. Rahman, Y.-F. Pu, M. Aamir, and S. Wali, "Structure revealing of low-light images using wavelet transform based on fractional-order denoising and multiscale decomposition," *The Visual Computer*, vol. 37, no. 5, pp. 865–880, 2021.

[22] Z. Rahman, Z. Ali, I. Khan, M. I. Uddin, Y. Guan, and Z. Hu, "Diverse image enhancer for complex underexposed image," *Journal of Electronic Imaging*, vol. 31, no. 4, p. 041213, 2022.

[23] Z. Rahman, M. Aamir, Z. Ali, A. K. J. Saudagar, A. AlTameem, and M. Khan, "Efficient contrast adjustment and fusion method for underexposed images in industrial cyber-physical systems," *IEEE Systems Journal*, vol. 17, no. 4, pp. 5085–5096, 2023.

[24] Z. Rahman, J. A. Bhutto, M. Aamir, Z. A. Dayo, and Y. Guan, "Exploring a radically new exponential Retinex model for multi-task environments," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 7, p. 101635, 2023.

[25] S. Lai, "GMS-uHRNet: A global multi-scale spatial U-HRNet for breast ultrasound semantic segmentation," in *International Conference on Image, Signal Processing, and Pattern Recognition (ISPP 2024)*, vol. 13180, SPIE, 2024.

[26] Y. Jin, X. Liu, and X. Huang, "EMR-HRNet: A multi-scale feature fusion network for landslide segmentation from remote sensing images," *Sensors*, vol. 24, no. 11, p. 3677, 2024.

[27] J. Liu, et al., "HRD-net: High resolution segmentation network with adaptive learning ability of retinal vessel features," *Computers in Biology and Medicine*, vol. 173, p. 108295, 2024.

[28] J.-S. Kim, et al., "E-HRNet: Enhanced semantic segmentation using squeeze and excitation," *Electronics*, vol. 12, no. 17, p. 3619, 2023.

[29] N. Sadeghi, et al., "Comparing the semantic segmentation of high-resolution images using deep convolutional networks: SegNet, HRNet, CSE-HRNet and RCA-FCN," *Journal of Information Systems and Telecommunication (JIST)*, vol. 4, no. 44, p. 359, 2023.

[30] H. Zhu, et al., "Application of K-means algorithm in Yi clothing color," in *International Conference on Internet of Things and Machine Learning (IoTML 2021)*, vol. 12174, SPIE, 2022.

[31] S. P. Kristanto, L. Hakim, and D. Yusuf, "K-means clustering segmentation on water microbial image with color and texture feature extraction," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, pp. 1317–1324, 2022.

[32] S. Jardim, J. António, and C. Mora, "Graphical image region extraction with K-means clustering and watershed," *Journal of Imaging*, vol. 8, no. 6, p. 163, 2022.

[33] A. Abernathy and M. E. Celebi, "The incremental online K-means clustering algorithm and its application to color quantization," *Expert Systems with Applications*, vol. 207, p. 117927, 2022.

[34] R. Bhuvanya and M. Kavitha, "Image clustering and feature extraction by utilizing an improvised unsupervised learning approach," *Cybernetics and Information Technologies*, vol. 23, no. 2, pp. 3–19, 2023.

[35] S. Kalaivani and M. R. Vimaladevi, "Enhancing endmember extraction using K-means clustering and pixel purity index," in *2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN)*, IEEE, 2023.

[36] G. F. C. Campos, et al., "Machine learning hyperparameter selection for Contrast Limited Adaptive Histogram Equalization," *EURASIP Journal on Image and Video Processing*, pp. 1–18, 2019.

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

Surface Roughness Prediction Based on CNN-BiTCN-Attention in End Milling

Guanhua Xiao[1], Hanqian Tu[2], Yunzhe Xu[3], Jiahao Shao[4], Dongming Xiang[5]*
Department of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China[1,2,4,5]
Department of QIXIN HONOR SCHOOL, Zhejiang Sci-Tech University, Hangzhou 310018, China[3]

***Abstract*—Surface roughness is a pivotal indicator of surface quality for machined components. It directly influences the performance and lifespan of manufactured products. Precise prediction of surface roughness is instrumental in refining production processes and curtailing costs. However, despite the use of identical processing parameters, the final surface roughness would be different. Thus, it challenges the effectiveness of traditional prediction models based solely on processing parameters. Current prevalent approaches for surface roughness prediction rely on handcrafted features, which require expert knowledge and considerable time investment. To address these challenges, we comprehensively consider the advantages of various deep learning methods and propose a novel end-to-end architecture. It synergistically integrates convolutional neural networks (CNN), bidirectional temporal convolutional networks (BiTCN), and attention mechanism, termed the CNN-BiTCN-Attention (CBTA) architecture. This architecture leverages CNN for automatic spatial feature extraction from signals, BiTCN to capture temporal dependencies, and the attention mechanism to focus on important features related to surface roughness. Experiments are conducted with popular deep learning methods on the public ACF dataset, which includes vibration, current, and force signals from the end milling process. The results demonstrate that the CBTA model outperforms other compared models. It achieves exceptional prediction performance with a mean absolute percentage error as low as 0.79% and an R^2 as high as 99.81%. This validates the effectiveness and superiority of CBTA in end milling surface roughness prediction.**

***Keywords*—Surface roughness prediction; end milling; CNN-BiTCN-Attention; deep learning**

I. Introduction

End milling is a machining process, which utilizes the cutting edges of a rotating cylindrical cutter to remove material from a workpiece. This method is extensively used for the production of parts with intricate shapes and stringent precision requirements. The control of surface roughness during end milling is essential for ensuring product quality, as it directly affects the appearance and wear resistance of parts as well as the compatibility with other components [1]. Hence, it is necessary to predict the surface roughness in end milling. This holds significant importance for optimizing machining parameters and enhancing both the efficiency of the process and the quality of the workpiece.

Methods for predicting surface roughness in end milling can be categorized into three main types: physical modeling, statistical analysis, and artificial intelligence based approaches. Firstly, the physical modeling methods are highly reliant on expert experience and prior knowledge, and they may struggle to accurately depict actual conditions due to the complex nonlinear characteristics of the end milling process [2]. Secondly, statistical analysis methods, which mostly consider only the influence of machining parameters, fail to address the situations where surface roughness differs despite identical machining parameters during the actual end milling [3]. Finally, the AI-based methods mainly focus on the impact on surface roughness from machining parameters, handcrafted features, or single signal. Moreover, their architectures are relatively simple, and handcrafted features also depend on the expertise of researchers [4].

In light of the above analysis, there is a need to propose an end-to-end deep learning method that fuses multiple signals for surface roughness prediction. After a thorough consideration of the advantages and disadvantages of various deep learning methods, we introduce the CNN-BiTCN-Attention (CBTA) architecture for the first time to predict surface roughness. This architecture employs convolutional neural network (CNN) for feature extraction, Bidirectional Temporal Convolutional Network (BiTCN) to capture long-term dependencies in signals, and incorporates an attention mechanism to allocate reasonable weights to different signals and features. Experiments are conducted on the public ACF dataset, which is collected during end milling of 45# steel and includes vibration, current, and force signals, each with three components. In this study, CBTA is compared with three popular deep learning models, and their performance is evaluated when using both multi-signal fusion and single-signal inputs, in order to assess the robustness and effectiveness of our proposed model.

The remainder of this paper is organized as follows: Section II presents related works. Section III introduces the surface roughness prediction model and data processing methods. Section IV details the experimental design, results, and discussion. The main conclusions of this study are presented in Section V.

II. Related Work

The traditional prediction of end milling surface roughness is primarily achieved through physical modeling methods. These methods analyze and establish mathematical models based on the physical phenomena occurring during the machining process. Feng et al. [5] proposed an analytical model to predict surface roughness during laser-assisted end milling of Inconel 718, based on tool movement and elastic response of the workpiece. Zhang et al. [6] developed an analytical model for surface roughness in particle-reinforced metal matrix composites. Their model utilized an undeformed chip thickness approach based on the Rayleigh probability distribution. Wang

*Corresponding authors.

et al. [7] developed a prediction model for surface roughness in milling. This model combined the effects of elastoplastic deformation, cutter parameters, microhardness, cutting force, and material properties with geometric and mechanical models. Jiang et al. [8] explored the influence of process parameters on cutting forces and surface roughness. They proposed a mathematical model to predict the milling forces and surface roughness of carbon fiber reinforced polymers, through a combination of theoretical and experimental approaches.

Statistical regression analysis is also a classic technique for predicting surface roughness, typically focusing on the relationship between machining parameters and surface roughness. Huang et al. [9] proposed a grey online modeling surface roughness monitoring system for end milling, based on grey theory and bilateral optimal fitting methods. Misaka et al. [10] employed the Co-Kriging method in conjunction with measurement data and analytical model to predict surface roughness in Computer Numerical Control (CNC) turning. The measurement data included cutting speed, feed rate, depth of cut, and acceleration. Gao et al. [11] developed an empirical model based on multiple regression analysis for dry face turning of AZ31B magnesium alloy. The model enabled the prediction of surface roughness through cutting speed, feed rate, and cutting depth. Sekulic et al.[12] utilized Response Surface Methodology (RSM) to predict surface roughness in ball-end milling of hardened steel, with spindle speed, feed per tooth, axial depth, and radial depth as input parameters.

In addition, machine learning and deep learning methods are also applied to surface roughness prediction especially with the rapid advancement of artificial intelligence. Chen et al. [13] proposed a backpropagation neural network (BPNN) to predict surface roughness in end milling, which utilizes spindle speed, feed rate, cutting depth, and milling distance as inputs. Li et al. [14] used cutting parameters and tool wear as input variables to propose a milling surface roughness prediction method which based on particle swarm optimization least squares support vector machine (PSO-LSSVM). Wu et al. [15] extracted time-domain and frequency-domain features from vibration signals through statistical calculations, and then input these features and cutting parameters into an Artificial Neural Network (ANN) to predict surface roughness of S45C steel. Shehzad et al. [16] introduced a CNN-LSTM model designed for the online monitoring of surface roughness in copper workpieces during ultraprecision fly cutting. This model employs vibration signals for its predictions and underwent a robustness assessment through validation cohort analysis. Guo et al. [17] collected an ISSA-optimized Deep Belief Network (DBN) model for surface roughness prediction. They gathered vibration and force signals during the milling process of P20 die steel, and subsequently extracted and filtered the time-frequency domain features of these signals to serve as inputs for the model.

In conclusion, the surface roughness prediction methods presented in the aforementioned literatures have achieved remarkable results in certain specific scenarios. Their research primarily relies on machining parameters or handcrafted features for prediction, while the use of fused signals and automatic feature extraction remains lacking. These methods suffer from issues such as neglect of dynamic factors, time-consuming processes, and an inability to comprehensively describe the entire machining process. To address these challenges, we propose an end-to-end deep learning architecture that fuses multiple signals.

III. Methodology

A. Surface Roughness Prediction Model

Deep learning models do not rely on handcrafted features, and directly utilize raw signals as input. Given the substantial volume and complex characteristics of vibration, current, and force signals, the model must possess capabilities for dimensionality reduction and feature extraction. Since these signals are time series data, the model requires strong temporal analysis capabilities. Additionally, these signals have different importance for surface roughness, and it is necessary to assign appropriate weights to the model. Based on these considerations, we propose the CBTA architecture, as illustrated in Fig. 1. Below is an introduction to each component of CBTA.

1) Convolutional Neural Network block: CNN is employed for dimensionality reduction and automatic feature extraction from signals in this paper. CNN is widely utilized in computer vision and speech recognition due to its robust spatial feature extraction capability. In the realm of intelligent manufacturing, applications of CNN include bearing remaining useful life prediction [18], tool wear estimation [19], and surface roughness prediction [20]. The CNN block comprises an input layer, convolutional layers, and pooling layers. The input layer receives the signal data. The convolutional layers contain a set of convolutional kernels, and one-dimensional convolution kernels are adopted in order to process time-series data. After the convolutional layer generates feature maps, the pooling layer uses the max pooling operation to reduce the number of network parameters and retain key features. By stacking multiple convolutional and pooling layers, CNN can extract deep features from signals and filter out redundant features. Furthermore, rectified linear unit (ReLU) [21] is utilized as the activation function to avoid overfitting and enhance convergence speed. The expressions for the convolution and pooling layers are as follows:

$$\mathrm{Y}_{m,k} = f\left(\sum_{i=1}^{n} X_i * w + b\right) \tag{1}$$

$$\mathrm{Z}_{m,l} = \max\left(\mathrm{Y}_{m,k}\right) \tag{2}$$

where $Y_{m,k}$ and $Z_{m,l}$ represent the outputs of the convolutional layer and the pooling layer, respectively; f denotes the activation function ReLU, X_i signifies the number of samples, w refers to the size of the convolutional kernel, and b represents the bias vector.

2) Bidirectional Temporal Convolutional Network block: TCN was proposed by Bai et al. [22], and has been proven superior to Long Short-Term Memory (LSTM) in various fields such as multivariate time series analysis and natural language processing [23]. Inspired by Bidirectional Long Short-Term Memory (BiLSTM) [24], we introduce BiTCN for surface roughness prediction. The deep features extracted by CNN are fed into BiTCN for analysis. As depicted in Fig. 1, BiTCN consists of a pair of parallel TCNs, termed Forward

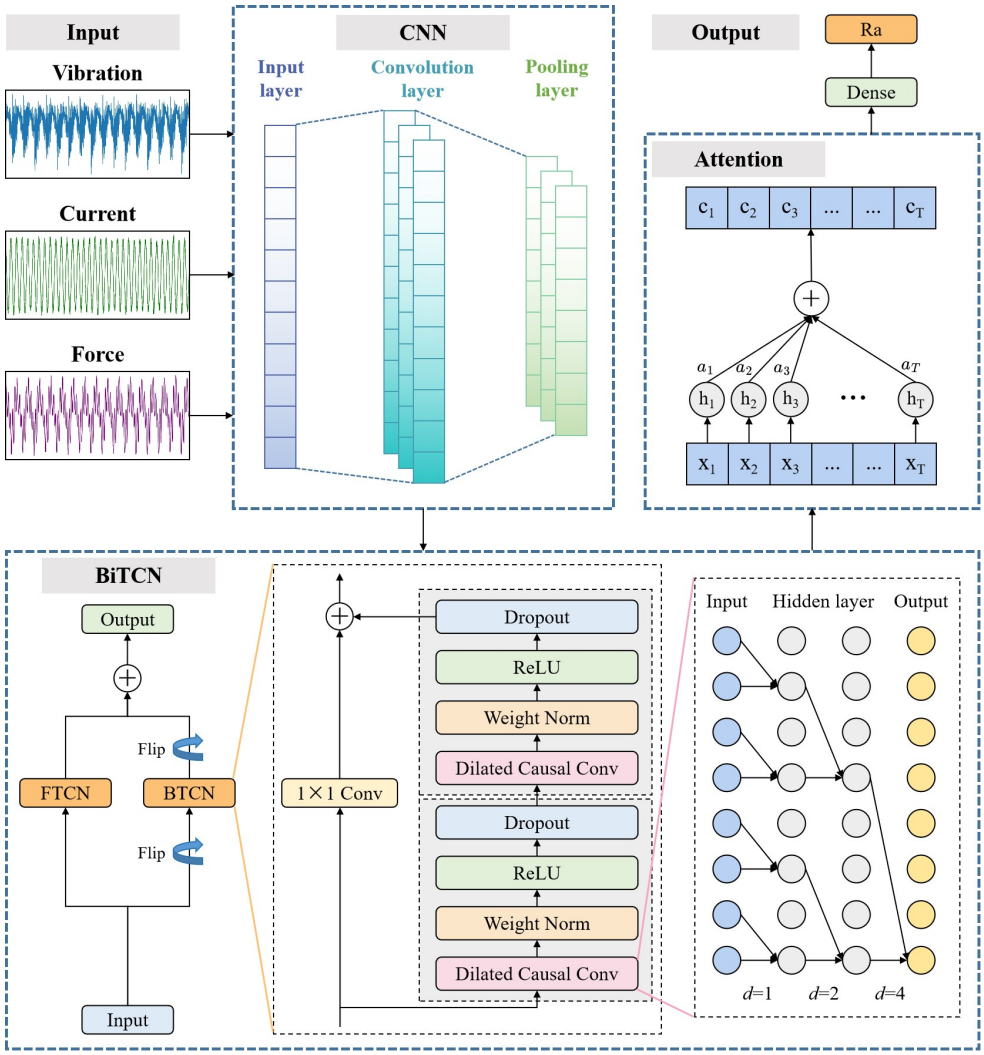


Fig. 1. CBTA architecture.

TCN (FTCN) and Backward TCN (BTCN), respectively. Their primary distinction lies in the order of input sequence, i.e. FTCN processes the sequence in the forward direction, while BTCN processes it in the reverse direction. Apart from this difference, the structures of FTCN and BTCN are identical.

TCN adheres to two principles: the output length of the network equals the input length, and future information cannot be leaked to the past. To satisfy these principles, 1D fully-convolutional network (1D-FCN) and causal convolution are introduced. The 1D-FCN ensures that the input and output lengths of each hidden layer remain identical. The causal convolution restricts the leakage of future information by simply padding the beginning and end of the time series, and then discards the excess padding values after convolution. It also limits the sliding direction solely from the past to the future. However, the modeling length of simple causal convolution is constrained by the size of the convolutional kernel. To establish long-range dependencies, numerous layers would need to be stacked, which can complicate the model and increase the risk of overfitting. To solve this issue, dilated convolution is introduced. The formula for dilated convolution is shown in Eq. (3).

$$F(s) = (X * df)(s) = \sum_{i=0}^{k-1} f(i) \cdot X_{s-d \cdot i} \tag{3}$$

In Eq. (3), $X = (x_1, x_2, \ldots, x_l)$ represents the feature sequence output by the CNN, where l is the sequence length. The symbol $*$ denotes the convolution operation, d is the dilation factor, k is the filter size, and $s - d \cdot i$ indicates the direction of the past. The dilation factor can be viewed as the sampling stride between adjacent filters, typically recommended to be 2^n, where n is the number of hidden layers. When d is equal to 1, the dilated convolution degrades to a regular convolution. The increase of d allows the TCN to achieve a broader receptive field, as illustrated in Fig. 1. This mechanism enables the network to capture long-range dependencies efficiently, and enhance its ability to analyze complex temporal patterns in the data.

Additionally, as a deep neural network model, even with the use of dilated causal convolution, TCN may still encounter

some issues such as vanishing gradients or network degradation. To address them, TCN incorporates residual blocks [25] to enhance its stability and generalization ability. This residual block comprises two layers of dilated causal convolution, where each layer utilizes weight normalization [26] for normalization and employs the ReLU activation function to expedite convergence. Moreover, dropout is applied for regularization. The output of residual function F is then added to the input x of residual block, as shown in Eq. (4).

$$O = Activation(x + F(x)) \tag{4}$$

Since the number of channels in x and $F(x)$ may not be consistent, the TCN employs a 1×1 Conv. This ensures that x matches the shape of $F(x)$ after a simple transformation.

3) Attention mechanism block: Traditional neural networks treat different features equally, which makes it difficult to distinguish important features. However, different features of different signals have varying degrees of correlation with surface roughness. To solve this problem, we introduced an attention mechanism. It assigns weights to enable the model to focus on features that are more important for surface roughness and reduce attention to less important information. For the computation of the attention mechanism, we adopt the commonly used *Bahdanau* algorithm with the following formulas:

$$e_{ij} = \tanh\left(W_1 * h_i + W_2 * h_j + b\right) \tag{5}$$

$$a_{ij} = softmax(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k=1}^{T} \exp(e_{ik})} \tag{6}$$

$$c_i = \sum_{j=1}^{T} a_{ij} * h_j \tag{7}$$

where, e_{ij} represents the degree of match between hidden layer states h_i and h_j, W is a learnable weight parameter, b is a bias vector, a_{ij} denotes the attention weight of the network, and c_i is the final output of the attention layer. By means of the attention mechanism, the computational efficiency of the model is enhanced, and the influence of noise and outliers is suppressed.

B. Data Processing Method

The public ACF dataset [27], derived from Shang et al.'s work, is employed for surface roughness prediction. This dataset captures end-milling operations on a 45# steel workpiece, which is performed with a four-flute carbide tool on a CNC machining center. It encompasses vibration, current, and force signals from the end milling process, along with corresponding surface roughness measurements. Each signal contains three directional components. All signals were sampled at a frequency of 20kHz to ensure synchronous and consistent data collection.

In this paper, our data processing is conducted on the ACF dataset, which involves two main steps: data segmentation and sample amplification, as shown in Fig. 2.

1) Data segmentation: Since the tool does not contact the workpiece before the actual milling process starts and after it ends, the raw signals consist of redundant data with values close to zero, as depicted in Fig. 2(a). Hence, it is necessary to identify and remove these redundant data segments and retain only valid segments from the actual machining process.

The method to extract valid data segments involves several steps. First, the original X-direction vibration data is divided into multiple segments through a fixed-length window. Secondly, the standard deviation for each segment is calculated according to Eq. (8). Thirdly, segments with standard deviation above a certain threshold are selected as valid data segments. Finally, valid segments from the other raw data are extracted based on the same positions. As shown in Fig. 2(b), the standard deviation of redundant data segments are significantly lower than those of valid data segments. Therefore, the threshold is set to a value slightly lower than the standard deviation of valid data segments. This method ensures that only the segments containing relevant machining activity are retained for further analysis, and enhance the precision and reliability of the data.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}} \tag{8}$$

2) Sample augmentation: Even after data segmentation, a single valid data segment still contains a large number of data points, which may hinder the learning speed and efficiency of deep learning models. Moreover, the quantity of samples can also impact the performance of these models. When the original sample size is limited, to improve the prediction accuracy of the model, sample augmentation can be employed to generate additional samples [28]. Specifically, the valid data segments are divided into N sub-signals using a fixed time step window without overlapping, as shown in Fig. 2(d). Each sub-signal is subsequently treated as a sample. By setting N to 50 and 100, we obtained two different datasets, namely ACF-50 and ACF-100, which divided the valid data segments into 50 and 100 equal parts respectively.

IV. Experiments

After data processing, we conduct experiments on the datasets ACF-50 and ACF-100, in order to evaluate the effectiveness of the CBTA model in predicting surface roughness. Each dataset is randomly split in a 4:1 ratio, with 80% used as the training set and 20% as the test set. In these experiments, CNN, BiTCN, and LSTM-Attention (LSTM-A) are employed as baseline models for comparison. The performances of models are evaluated by three common metrics: Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Coefficient of Determination (R^2). Their formulas are shown in Eq. (9), (10), and (11). Smaller values of RMSE and MAPE, along with larger values of R2, indicate better model performance.

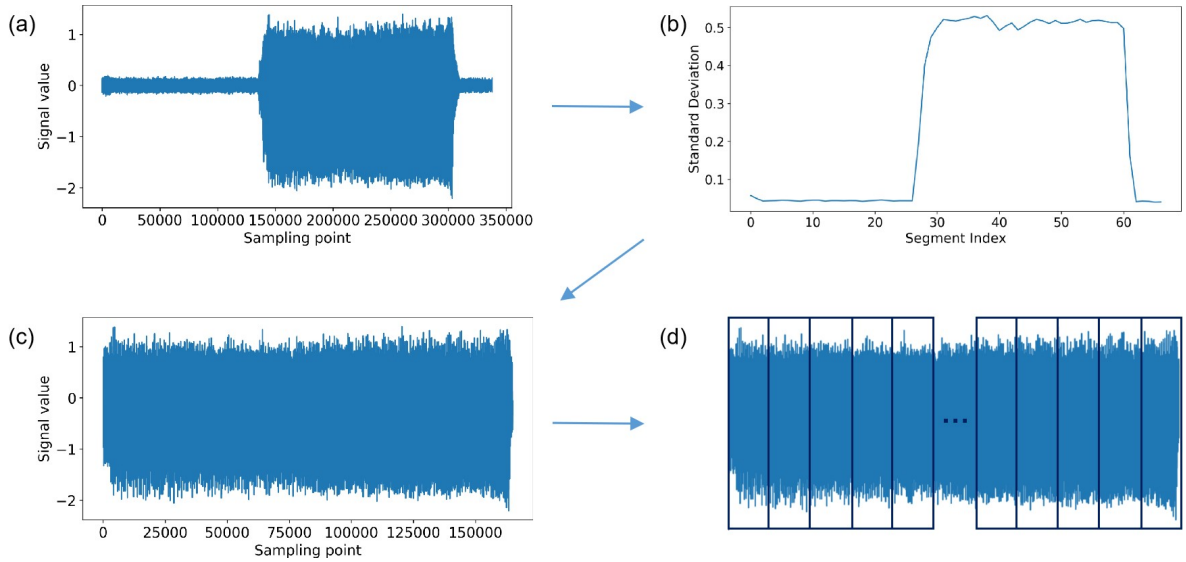


Fig. 2. Data segmentation and sample augmentation.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y_i})^2} \tag{9}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y_i}}{y_i}\right| \tag{10}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y_i})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{11}$$

A. Model Parameters Settings

The experiments are conducted with Python 3.8 programming language and TensorFlow 2.9.0 framework. All models share the same common hyperparameters, i.e. a batch size of 32, a learning rate of 0.0005, the Adam optimizer, and a training epoch count of 200. All data are standardized before being input into the models, in order to eliminate the impact of different scales on the training process.

Table I presents the network structure of LSTM-A, while the structure of CBTA is detailed in Table II. The core network structures of the CNN and BiTCN models are identical to the CNN block and BiTCN block in CBTA, respectively. In our study, the number of stacks of residual blocks in TCN is set to 1, and the kernel initializer uses glorot uniform. According to the receptive field calculation formula (Eq. (12)) of TCN, the dilations in CBTA are set to [1, 2, 4, 8, 16] to ensure that the receptive field can fully cover the input sequence. As for BiTCN model, since there is no CNN block for data dimensionality reduction, the dilations are set to [1, 2, 4, 8, 16, 32, 64].

$$R_{field} = 1 + 2 \cdot (K_{size} - 1) \cdot N_{stack} \cdot \sum_{i} d_i \tag{12}$$

TABLE I. NETWORK STRUCTURE OF LSTM-A

Block	Layer	Units	Activation
LSTM block	LSTM	32	Tanh
Attention block	Attention	16	-
Output block	Dense	-	-

B. Training Loss

Training loss plot is crucial for evaluating model performance. Fig. 3 illustrates the training losses of various models on ACF-100. To prevent overfitting, an early stopping strategy was employed during the training process in our experiments. This strategy halts iterations when the training loss of models ceases to decrease. Both CNN and BiTCN demonstrate rapid convergence rates, and achieve convergence within 50 epochs. CNN stops training early due to the activation of the early stopping mechanism, while BiTCN's loss continues to decline slightly. LSTM-A has a slower convergence rate, and stabilizes around 90 epochs. It also triggers early stopping. CBTA exhibits the fastest convergence rate, and it achieves convergence within 30 epochs with the lowest loss.

C. Prediction Results

To investigate the effectiveness of various models in utilizing fused multi-process signals for prediction, we employed vibration, current, and force signals as inputs. Fig. 4 showcases the performance of different models on the ACF-50 and ACF-100 test sets. The prediction results and absolute errors of each model on ACF-50 are illustrated in Fig. 5, with only 50 samples displayed due to space limitations. Evidently, CBTA achieves the best performance on both datasets, followed by LSTM-A, while CNN and BiTCN exhibit similar effectiveness. Moreover, CBTA achieves RMSE, MAPE, and R^2 values of 0.0844, 2.76%, and 98.48% respectively on the ACF-50 dataset, while 0.0284, 0.79%, and 99.81% on the ACF-100 dataset. These results indicate that our proposed method

TABLE II. NETWORK STRUCTURE OF CBTA

Block	Layer	Filters	Kernel size	Units	Strides	Padding	Activation
CNN block	Conv1D	64	3	-	1	Same	ReLU
	MaxPool1D	2	-	-	2	-	-
	Conv1D	32	3	-	1	Same	ReLU
	MaxPool1D	2	-	-	2	-	-
BiTCN block	FTCN	16	16	-	-	Causal	ReLU
	BTCN	16	16	-	-	Causal	ReLU
Attention block	Attention	-	-	16	-	-	-
Output block	Dense	-	-	-	-	-	-

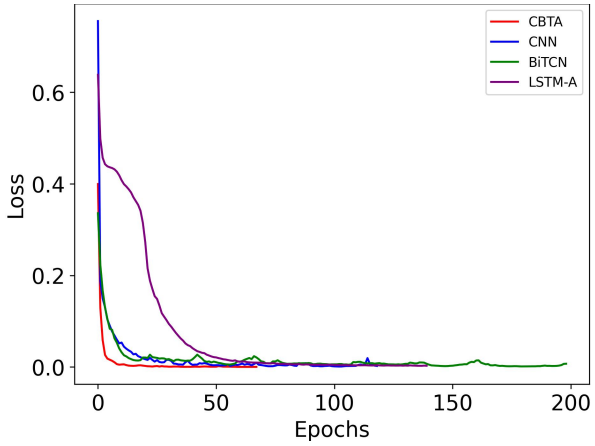


Fig. 3. Training loss plot for all models.

achieves satisfactory prediction accuracy, and demonstrate outstanding performance and robust prediction ability.

D. Comparison Under Single Signal Input

In order to evaluate the performance of each model with a single signal input, we conducted experiments on the ACF-100 dataset for predicting surface roughness with only vibration data. The results are shown in Fig. 6. It is evident that the performance of models declined to varying degrees when solely vibration signal is used, compared to the performance when employing a fusion of vibration, current, and force signals as inputs. However, CBTA is the least affected, and still achieves excellent results, with MAPE of 1.80% and R^2 value of 99.09%. This demonstrates that the fusion of multiple process signals contributes to improve accuracy, and also validates the effectiveness and robustness of CBTA in surface roughness prediction.

E. Result Discussion

Given the above experimental results, along with different model frameworks, the superiority of CBTA is analyzed from the following aspects:

- The CNN block possesses remarkable capabilities in feature extraction. In the milling process, the generated vibration, current, and force signals are intricate and multidimensional. They encompass not only spatial features, such as waveforms and frequency distributions, but also temporal features, like trends and periodicity. Compared with BiTCN and LSTM-A models, the CNN block, through its unique convolutional and pooling layer structure, can automatically extract local features related to surface roughness from the raw data. These features serves as input variables in subsequent blocks.
- The BiTCN block enables the model to comprehend temporal dependencies within the data. In comparison to CNN and LSTM-A models, TCN expands the network's receptive field through dilated causal convolutions, which allows it to more effectively capture long-term dependencies across time steps. Moreover, the bidirectional architecture of the BiTCN block enables the model to harness information from the entire signal sequence, encompassing both antecedent and subsequent data. This capability not only bolsters the precision of predictions but also amplifies the model's robustness.
- The attention mechanism enhances CBTA to concentrate on the critical segments of the input sequence. During the milling process, certain fluctuations in the vibration signal may arise from external disturbances rather than actual changes in surface quality. Compared to CNN and BiTCN models, the attention mechanism helps CBTA discern those irrelevant fluctuations by assigning weights to different time steps or features. This allows the model to concentrate on the information most relevant to surface roughness. This targeted method not only enhances the reliability of the model but also equips it with greater resilience against noise and irrelevant data.
- Each component offers complementary advantages. The integration of CNN, BiTCN and attention mechanism equips our model with the ability to handle complex nonlinear relationships between input signals and surface roughness, as well as to model temporal dependencies. This hierarchical approach enables CBTA to conduct a profound and meticulous analysis of the milling process and ultimately achieve precise roughness predictions.

V. CONCLUSION

This study investigates a deep learning-based approach for surface roughness prediction in end milling with the fusion of multiple process signals. Initially, through data segmentation and sample augmentation, two distinct datasets are developed

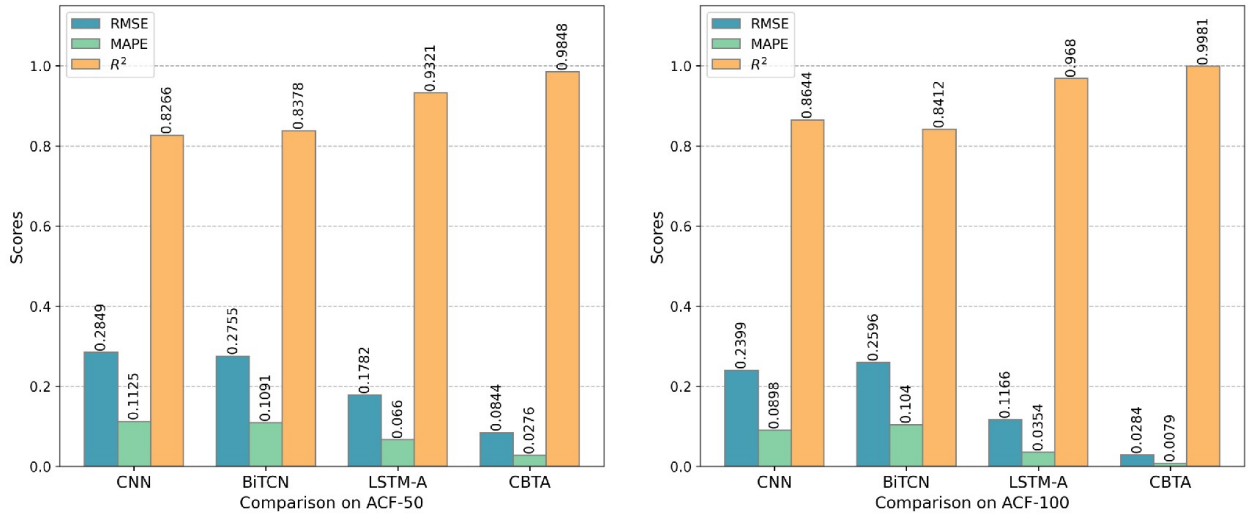


Fig. 4. Comparison of different models on the ACF-50 and ACF-100 test sets.

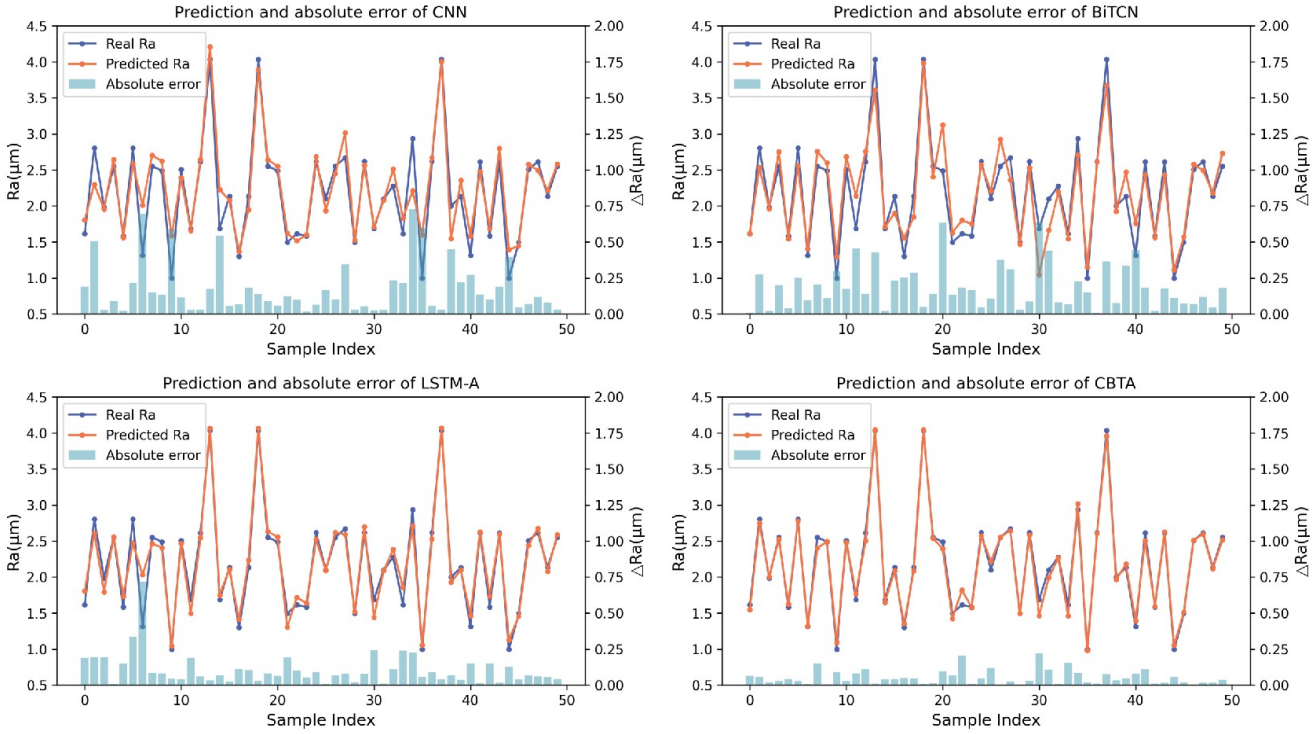


Fig. 5. Surface roughness prediction results and absolute errors of each model on the ACF-50 test set.

from the ACF dataset for model training and evaluation. Subsequently, the complementary strengths of CNN, BiTCN and attention mechanism lead to the proposal of an end-to-end CNN-BiTCN-Attention architecture (CBTA). This architecture possesses the capabilities to extract spatial features, capture temporal dependencies and allocate appropriate weights.

Three different deep learning architectures are utilized in experiments for comparison with CBTA. The results show that CBTA achieves high prediction accuracies of 98.48% and 99.81% on the two datasets, respectively, along with low MAPE values of 2.76% and 0.79%. These metrics indicate that CBTA outperforms other models in terms of precision. Additionally, the effectiveness of surface roughness prediction is examined by using only vibration signal. The results show that CBTA maintains excellent prediction results although the performance of other models decline, which highlight its robustness and effectiveness.

Despite its excellent performance on public datasets, CBTA still has some limitations. For instance, as a deep learning architecture, it lacks interpretability, which hinders fault diagnosis and process optimization in industrial applications. Additionally, as the performance of machining equipment may

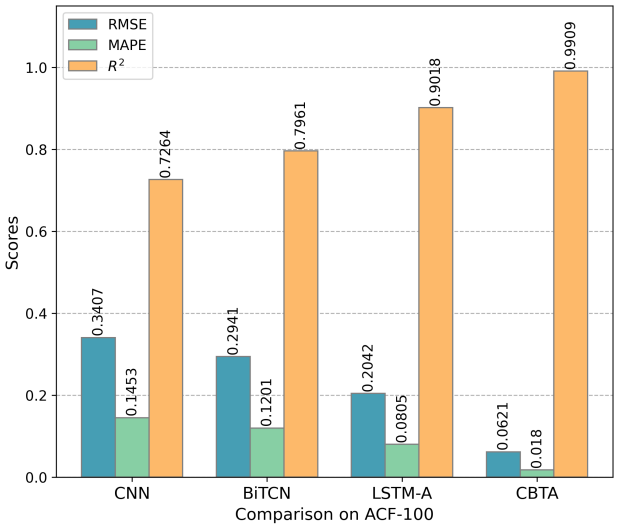


Fig. 6. Comparison of model prediction performance on the ACF-100 dataset using only vibration signal as input.

change over long-term operation, the prediction accuracy of the pre-trained model could decline. In such cases, it becomes necessary to adjust model parameters or retrain the model to maintain its effectiveness.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No. 62002328), and the Fundamental Research Funds of Zhejiang Sci-Tech University (Grant No. 24232123-Y).

REFERENCES

[1] K. Vipindas and J. Mathew, "Wear behavior of tialn coated wc tool during micro end milling of ti-6al-4v and analysis of surface roughness," *Wear*, vol. 424, pp. 165–182, 2019.

[2] A. M. Zain, H. Haron, and S. Sharif, "Application of ga to optimize cutting conditions for minimizing surface roughness in end milling machining process," *Expert Systems with Applications*, vol. 37, no. 6, pp. 4650–4659, 2010.

[3] N. Karkalos, N. Galanis, and A. Markopoulos, "Surface roughness prediction for the milling of ti–6al–4v eli alloy with the use of statistical and soft computing techniques," *Measurement*, vol. 90, pp. 25–35, 2016.

[4] H. Yang, H. Zheng, and T. Zhang, "A review of artificial intelligent methods for machined surface roughness prediction," *Tribology International*, p. 109935, 2024.

[5] Y. Feng, T.-P. Hung, Y.-T. Lu, Y.-F. Lin, F.-C. Hsu, C.-F. Lin, Y.-C. Lu, X. Lu, and S. Y. Liang, "Surface roughness modeling in laser-assisted end milling of inconel 718," *Machining Science and Technology*, 2019.

[6] Z. Zhang, P. Yao, J. Wang, C. Huang, R. Cai, and H. Zhu, "Analytical modeling of surface roughness in precision grinding of particle reinforced metal matrix composites considering nanomechanical response of material," *International Journal of Mechanical Sciences*, vol. 157, pp. 243–253, 2019.

[7] B. Wang, Q. Zhang, M. Wang, Y. Zheng, and X. Kong, "A predictive model of milling surface roughness," *The International Journal of Advanced Manufacturing Technology*, vol. 108, pp. 2755–2762, 2020.

[8] J. Xiaohui, G. Shan, Z. Yong, H. Shirong, and L. Lei, "Prediction modeling of surface roughness in milling of carbon fiber reinforced polymers (cfrp)," *The International Journal of Advanced Manufacturing Technology*, vol. 113, pp. 389–405, 2021.

[9] P. B. Huang, H.-J. Zhang, and Y.-C. Lin, "Development of a grey online modeling surface roughness monitoring system in end milling operations," *Journal of Intelligent Manufacturing*, vol. 30, pp. 1923–1936, 2019.

[10] T. Misaka, J. Herwan, O. Ryabov, S. Kano, H. Sawada, N. Kasashima, and Y. Furukawa, "Prediction of surface roughness in cnc turning by model-assisted response surface method," *Precision Engineering*, vol. 62, pp. 196–203, 2020.

[11] H. Gao, B. Ma, R. P. Singh, and H. Yang, "Areal surface roughness of az31b magnesium alloy processed by dry face turning: An experimental framework combined with regression analysis," *Materials*, vol. 13, no. 10, p. 2303, 2020.

[12] M. Sekulic, V. Pejic, M. Brezocnik, M. Gostimirović, and M. Hadzistevic, "Prediction of surface roughness in the ball-end milling process using response surface methodology, genetic algorithms, and grey wolf optimizer algorithm," *Advances in Production Engineering & Management*, vol. 13, no. 1, pp. 18–30, 2018.

[13] C.-H. Chen, S.-Y. Jeng, and C.-J. Lin, "Prediction and analysis of the surface roughness in cnc end milling using neural networks," *Applied Sciences*, vol. 12, no. 1, p. 393, 2021.

[14] B. Li and X. Tian, "An effective pso-lssvm-based approach for surface roughness prediction in high-speed precision milling," *Ieee Access*, vol. 9, pp. 80 006–80 014, 2021.

[15] T. Wu and K. Lei, "Prediction of surface roughness in milling process using vibration signal analysis and artificial neural network," *The International Journal of Advanced Manufacturing Technology*, vol. 102, no. 1, pp. 305–314, 2019.

[16] A. Shehzad, X. Rui, Y. Ding, J. Zhang, Y. Chang, H. Lu, and Y. Chen, "Deep-learning-assisted online surface roughness monitoring in ultraprecision fly cutting," *Science China Technological Sciences*, pp. 1–16, 2024.

[17] M. Guo, J. Zhou, X. Li, Z. Lin, and W. Guo, "Prediction of surface roughness based on fused features and issa-dbn in milling of die steel p20," *Scientific Reports*, vol. 13, no. 1, p. 15951, 2023.

[18] D. Yao, B. Li, H. Liu, J. Yang, and L. Jia, "Remaining useful life prediction of roller bearings based on improved 1d-cnn and simple recurrent unit," *Measurement*, vol. 175, p. 109166, 2021.

[19] F. Aghazadeh, A. Tahan, and M. Thomas, "Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process," *The International Journal of Advanced Manufacturing Technology*, vol. 98, pp. 3217–3227, 2018.

[20] A. P. Rifai, H. Aoyama, N. H. Tho, S. Z. M. Dawal, and N. A. Masruroh, "Evaluation of turned and milled surfaces roughness using convolutional neural network," *Measurement*, vol. 161, p. 107860, 2020.

[21] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 315–323.

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

[23] S. Gopali, F. Abri, S. Siami-Namini, and A. S. Namin, "A comparison of tcn and lstm models in detecting anomalies in time series data," in *2021 IEEE International Conference on Big Data (Big Data)*. IEEE, 2021, pp. 2415–2420.

[24] S.-H. Chien, B. Sencer, and R. Ward, "Accurate prediction of machining cycle times and feedrates with deep neural networks using bilstm," *Journal of Manufacturing Systems*, vol. 68, pp. 680–686, 2023.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[26] T. Salimans and D. P. Kingma, "Weight normalization: A simple reparameterization to accelerate training of deep neural networks," *Advances in neural information processing systems*, vol. 29, 2016.

[27] S. Suiyan, "Dataset for digital twin paper," Feb. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7643683

[28] X. Yan, Y. Liu, and M. Jia, "Multiscale cascading deep belief network for fault identification of rotating machinery under various working conditions," *Knowledge-Based Systems*, vol. 193, p. 105484, 2020.

Enriching Sequential Recommendations with Contextual Auxiliary Information

Adel Alkhalil
Department of Information and Computer Science, College of Computer Science and Engineering,
University of Ha'il, Ha'il, 81481, Saudi Arabia

***Abstract*—Recommender Systems (RS) play a key role in offering suggestions and predicting items for users on e-commerce and social media platforms. Sequential recommendation systems (SRS) leverage the user's previous interaction history to forecast the next user-item interaction. Although deep learning methods like CNNs and RNNs have enhanced recommendation quality, current models still face challenges in accurately predicting future items based on a user's past behavior. Transformer-based SRS have shown a significant performance boost in generating accurate recommendations by using only item identifiers which are not sufficient to generate meaningful and relevant results. These models can be improved by incorporating descriptive features of the items, such as textual descriptions. This paper proposes a transformer-based SRS, ConSRec, Contextual Sequential Recommendations, that incorporates auxiliary information of the items, such as textual features, along with item identifiers to model user behavior sequences for producing more accurate recommendations. ConSRec builds upon the BERT4Rec model by integrating auxiliary information through sentence representations derived from the textual features of items. Extensive experiments conducted on several benchmark datasets demonstrate substantial improvements compared to other advanced models.**

***Keywords*—Recommender system; sequential recommendation; auxiliary information; sentence transformer; sentence embedding**

I. Introduction

A Recommender system (RS) is designed to predict user preferences based on user intent, which can vary over time [2], [1], [3], [4]. To better capture users' dynamic preferences, several sequential recommendation (SR) techniques have been recently introduced. These methods utilize a user's past interaction history to predict the next item they are likely to engage with [7], [6], [5], [8], [9].

Earlier SR models relied on Markov Chain models to capture user preferences and predict the next item in the sequence [10], [6]. With the rise of deep learning, many SR models have transitioned to using neural network architectures like RNNs [12], [11] and CNNs [13]. Later, attention-based Transformer models [14], [32] were introduced to address SR problems, such as SASRec [8] employs a uni-directional attention mechanism, which limits its ability to capture comprehensive user preferences. BERT4Rec [9] addresses this by adopting a bi-directional transformer architecture to learn contextual relationships from both directions. However, both models rely primarily on item identifiers and fail to incorporate auxiliary contextual data, such as textual descriptions or reviews, which are critical for improving recommendation accuracy, especially under sparse data conditions.

KeBERT4Rec [15] incorporated keywords along with item identifiers in BERT4Rec model by concatenating the keyword representation with items by using one-hot encoding to generate the keyword vector, which did not capture the contextual meaning of keywords. Although, these SR models show significant performance gain, however, they do not exploit contextual features to generate meaningful representations.

Incorporating such contextual auxiliary information into SR models not only improves recommendations, particularly under sparse situations but also has significant practical implications in real-world applications. For instance, in e-commerce, enriching recommendations with textual features like product descriptions can enhance personalized shopping experiences, leading to increased user satisfaction and higher sales conversion rates. Similarly, in media streaming services, leveraging textual metadata such as genre descriptions or user reviews can better align recommendations with user preferences, enhancing user engagement. To achieve this, we propose a model called ConSRec, Contextual Sequential Recommendations, a modification of the BERT4Rec model, which includes contextual descriptions. The proposed model, ConSRec, leverages a transformer-based architecture to incorporate contextual auxiliary information, such as textual descriptions and user reviews, into sequential recommendation tasks. This design makes it highly effective in addressing key challenges like sparse user-item interactions and the inability of existing models to utilize rich contextual data. By integrating multi-head self-attention mechanisms, ConSRec ensures the effective fusion of sequential and contextual data, providing a robust solution to improve recommendation accuracy. The major objectives of the paper can be outlined as below:

- Introduce a model that incorporates the user sequences with contextual item descriptions using Masked Language Model (Close Objective Task) and bidirectional transformers.
- Generate meaningful representations using contextual description of the items using Sentence-BERT.
- Evaluation and performance comparison of the model against existing leading models.

The article is structured as follows: Section II provides a review of the relevant literature, while Section III offers a detailed explanation of the proposed model. Section IV covers the experimental evaluation of the work, and Section V concludes the paper.

II. Literature Review

Sequential recommendation system is a type of RS that exploits the user interaction sequences to infer the successive

item [17]. The aim of SR is to recommend future product by considering historical behavior of users. This historical behavior is also known as next item prediction. Earlier, the SRS were introduced using Markov Chains (MC) models for capturing sequential patterns from the user historical preferences [18], [10], [19]. The next item preferred by the users are predicted depending upon the last item, thus interpreting only the adjoining sequential behavior.

Recurrent Neural Network (RNN) based models exploiting Gated Recurrent Unit (GRU) [20], [12] and Long Short-Term Memory (LSTM) [39] have showed substantial performance gain for SR [25], [21], [22], [11], [12], [24], [23]. RNNs enforce rigid sequential patterns for encoding user preferences for making predictions. Besides RNN, a number of Convolutional Neural Network (CNN) [26], [27] based RS have also been introduced that also target problems related to the sequential recommendation. For example, Tang and Wang *et al.* [13] exploit CNN for capturing local sequential features using more recent behaviors.

Recently, Transformer models based on the attention mechanism [14] have achieved outstanding results in various deep learning tasks, including text classification [28], image captioning [29], and machine translation [30]. Initially developed for natural language processing, Transformers have also transformed the field of sequential recommendation (SR) [17] by leveraging the encoder component to process sequential data. This converts the sequence of items, representing the user's interaction history, into a sequence of vector representations [14]. The input sequence of items is first embedded and then concatenated with positional embeddings capturing the item's position in the sequence.

TABLE I. COMPARISON OF SR MODELS

Model	Pros	Cons
GRU4Rec	GRU model of RNN with ranking-based loss function.	Aimed for session-based recommendation systems using RNN.
SASRec	First unidirectional self-attention sequential model based on Transformers for next-item recommendation.	Uses only left-to-right attention, limiting its ability to learn hidden representations bidirectionally.
BERT4Rec	Bidirectional model employing Transformer architecture with multi-head self-attention to analyze user behavior sequences.	Lacks incorporation of additional information to produce meaningful predictions.
KeBERT4Rec	Extension of BERT4Rec leveraging keywords alongside item identifiers for next-item prediction.	Keyword representations are not derived using contextual embedding techniques.
FDSA	Segregated attention blocks exploit items and their features to predict the next item.	Heterogeneous item characteristics make it challenging to determine user preferences accurately.
S3Rec	Self-supervised model utilizing attribute data to learn correlations among items.	Does not use descriptive information for generating meaningful representations.

Most of recent sequential models shown in Table I follows the transformer architecture comprising of encoder block [31], [8], [9], [15] and using the item identifiers for next item recommendation. A feature level deeper self attentive model [16] introduced by T. Zhang *et al* exploits segregated attention blocks for items and their associated features to predict next item. In [33] proposed S3Rec, a self supervised SR model that utilized the attribute data of item to learn the correlation among them. KeBERT4Rec [15] leverages the keyword by integrating them with item identifier for the prediction of next item in sequence.

Existing sequential recommendation models, such as SASRec and BERT4Rec, rely primarily on item identifiers and focus on implicit feedback for next-item prediction. SASRec's uni-directional attention mechanism limits its ability to fully capture sequential dependencies, while BERT4Rec's bi-directional architecture, although more robust, still neglects auxiliary information like textual descriptions and user reviews. These omissions reduce the models' effectiveness in scenarios with sparse data or ambiguous user-item interactions, which necessitate richer contextual representation. To address this we proposed ConSRec, a model that combines auxiliary information and item identifiers to create embeddings using the Sentence BERT [34] embedding technique. This enhances item recommendation and prediction accuracy by capturing contextualized representations.

III. METHODOLOGY

The suggested framework "ConSRec - Contextual Sequential Recommendation System" is depicted in Fig. 1. The proposed paradigm is developed based upon Transformer architecture that adapted the deep bidirectional BERT model for SR prediction task (Fig. 2).

Fig. 1. Proposed methodology outline diagram.

A. System Overview

Before passing the sequence of items to the model proposed, the auxiliary features of these items are taken as input to the Sentence-BERT. This auxiliary information is the textual description of the items in the form of sentences that are processed to extract the contextual dense feature representation. These dense embedding are extracted prior to training phase to reduce the model training time.

Subsequently, during the training process, the auxiliary information's embeddings of items within a sequence are extracted and then passed to the embedding layer. These embeddings are subsequently combined with the item's embedding and positional embedding to capture the sequential behavior of the items. Only the encoder part of Transformer is used to compute the hidden representation using self attention mechanism for each item.

This resultant concatenated item's representations of sequence are then processed through stack of Transformer layer from [14] where hidden features for each item are calculated simultaneously at each layer. These layers share information bidirectionally across each position in hierarchical manner. After processing through all layers, a final learned hidden representation is projected at output layer that contemplated the future item recommendation for a user.

Several experiments were carried out on three benchmark datasets—MovieLens-1M, MovieLens-20M, and Amazon Beauty—to validate the effectiveness of the proposed model. The model's architecture includes an embedding layer, a transformer layer, and an output layer.

B. Mathematical Formulation of Proposed Model

Let set of users be shown mathematically as $\mathcal{U}$ = $\{u_1, u_2, u_3, ..., u_{|\mathcal{U}|}\}$ $\mathcal{M}$ = $\{m_1, m_2, m_3, ..., m_{|\mathcal{M}|}\}$ be the set of items. For each item, there is some item description (auxiliary information) that is in textual form denoted as $\mathcal{TD}$ = $\{des_1, des_2, des_3, ..., des_{|\mathcal{M}|}\}$. The items interacted in the sequence $\mathcal{S}$ in historical order for a user u is denoted as $\mathcal{S}$ = $\{m_1, m_2, m_3, ..., m_n\}$ where m_n is a particular item from $\mathcal{M}$, the user has acted upon previously. Given the sequence history $\mathcal{S}$, the objective of the SRS is to anticipate the future item m_n+1, the model will predict as:

$$\mathcal{P}(m_n\text{+1} = m|\mathcal{S})$$

C. Embedding Layer

The recommendation model in [9] makes use of the positional embedding along with the item identifier embedding to maintain the sequence of the items, thus memorizing the sequential order of the input. However, the pair alone does not describe the contextual representation of the input. It also does not recommend contextually especially under sparse conditions. To overcome this limitation, ConSRec incorporates additional auxiliary information based on contextualized description of items. The model utilizes the Sentence-BERT [34] for capturing contextual representation of the item descriptions. The architecture of Sentence-BERT for extracting sentence embedding is depicted in Fig. 3.

Sentence-BERT first utilizes BERT to generate word/ token embedding. Input in the form of sentences or text of various length is injected to the selected SBERT model, that generates contextualized word embedding for all input tokens in the sentence. Secondly, these word embedding are passed through a pooling layer to generate a fixed-sized vector representation. Among various pooling options available, the model utilizes the mean pooling in which mean of all contextualized token embedding is calculated to produce a fixed dimensional output embedding vector. Given the item descriptions of various length of all items, $\{\mathcal{TD}\}$ as input, the model produces 384 dimensional dense vector representation $\{Emb_{\mathcal{TD}}\}$ as in Eq. 1. These 384 dimensional embeddings are then used along with the item identifier and position embedding to produce information rich vector representations as shown in Eq. 2.

$$SBERT(\{\mathcal{TD}\}) = \{Emb_{\mathcal{TD}}\} \tag{1}$$

In the proposed model, *d* dimensional embedding layer is constructed by summing up the item identifier embedding, the position embedding and the additional auxiliary information (item description) extracted from $\{Emb_{\mathcal{TD}}\}$. Thus, for a given item m_i, the input embedding matrix $\mathcal{EM}$ is formulated by adding the corresponding item embedding E_m, position embedding E_{pos} and textual description embedding E_{des} as:

$$\mathcal{EM}_m = E_m + E_{pos} + E_{des} \tag{2}$$

D. Transformer Layer

The summed embedding $\mathcal{EM}$ becomes the input to the transformer layer that iteratively calculates the hidden representations of each item at each layer.

The structure of transformer layer or simply the encoder layer is build using the "multi-head attention". The layer piles up multiple encoder blocks [14] each consisting of "Multi-Head Self Attention" sub-layer and a "Position-wise Feed Forward Network". Given that $\mathcal{E}^{\curlywedge} = [\mathcal{EM}_{\curlywedge}^{\curlywedge}, \mathcal{EM}_{\curlywedge}^{\curlywedge}, ...\mathcal{EM}_{\approx}^{\curlywedge}]$ depict the dense vector embedding of the user sequence to the transformer, multi head self attention layer, MHSA is defined as:

$$MHSA(\mathcal{E}^l) = [h_1; h_2; ....h_h]W^0 \tag{3}$$

$$h_i = Attn(\mathcal{E}^l W_i^Q, \mathcal{E}^l W_i^K, \mathcal{E}^l W_i^V) \tag{4}$$

where W_i^Q, W_i^K and $W_i^V \in \mathbb{R}^{dxd/h}$ are the three learnable projection weight matrices and $W_i^0 \in \mathbb{R}^{dxd}$. $\mathcal{E}^l W_i^Q$, $\mathcal{E}^l W_i^K$, $\mathcal{E}^l W_i^V$ are the three linear transformation of input vector representation $\mathcal{E}'$ for Query, Key and Value (Q,K,V) vectors. Here, the attention function is scaled dot product [14] computed as:

$$Attn(Q, K, V) = \sigma\left(\frac{QK^T}{\sqrt{d/h}}\right)V \tag{5}$$

where the Query, Key and Value matrices are denoted by Q, K, V respectively and σ is the softmax function. Let MHSA at the l^{th} layer be S_i. Since, the MHSA block is based on linear projections, thus, the non-linearity to the MHSA is empowered by applying position-wise feed-forward network layer, PFN on all MHSA(S_i) separately.

$$PFN = [FNL(S_1^l)^T, FNL(S_2^l)^T .....FNL(S_n^l)^T] \tag{6}$$

$$FN(S_i) = GELU\big(S_i W^{(1)} + b^{(1)}\big)W^{(2)} + b^{(2)} \tag{7}$$

A smoother GELU activation function is used inline with BERT [5] and OpenAI GPT [40]. $W^{(1)}, b^{(1)}, W^{(2)}$, and $b^{(2)}$ are hyper-parameters communicated at all layers. Complexity of the model is reduced using residual connection at each sub layer. Dropout is applied followed by layer Normalization, LNorm. Thus, the sub-layer output at each level is $LNorm(x + Dropout(sublayer(x)))$. Input at each layer is denoted by x in the LNorm and represented as:

$$\mathcal{E}^l = Trm\big(\mathcal{E}^{l-1}\big), \quad \forall i \in [1, 2, ...L] \tag{8}$$

$$whiteA = Dropout\big(PFN(S_i^{l-1})\big) \tag{9}$$

$$Trm(\mathcal{E}^{l-1}) = LNorm\big(S_i^{l-1} + A\big) \tag{10}$$

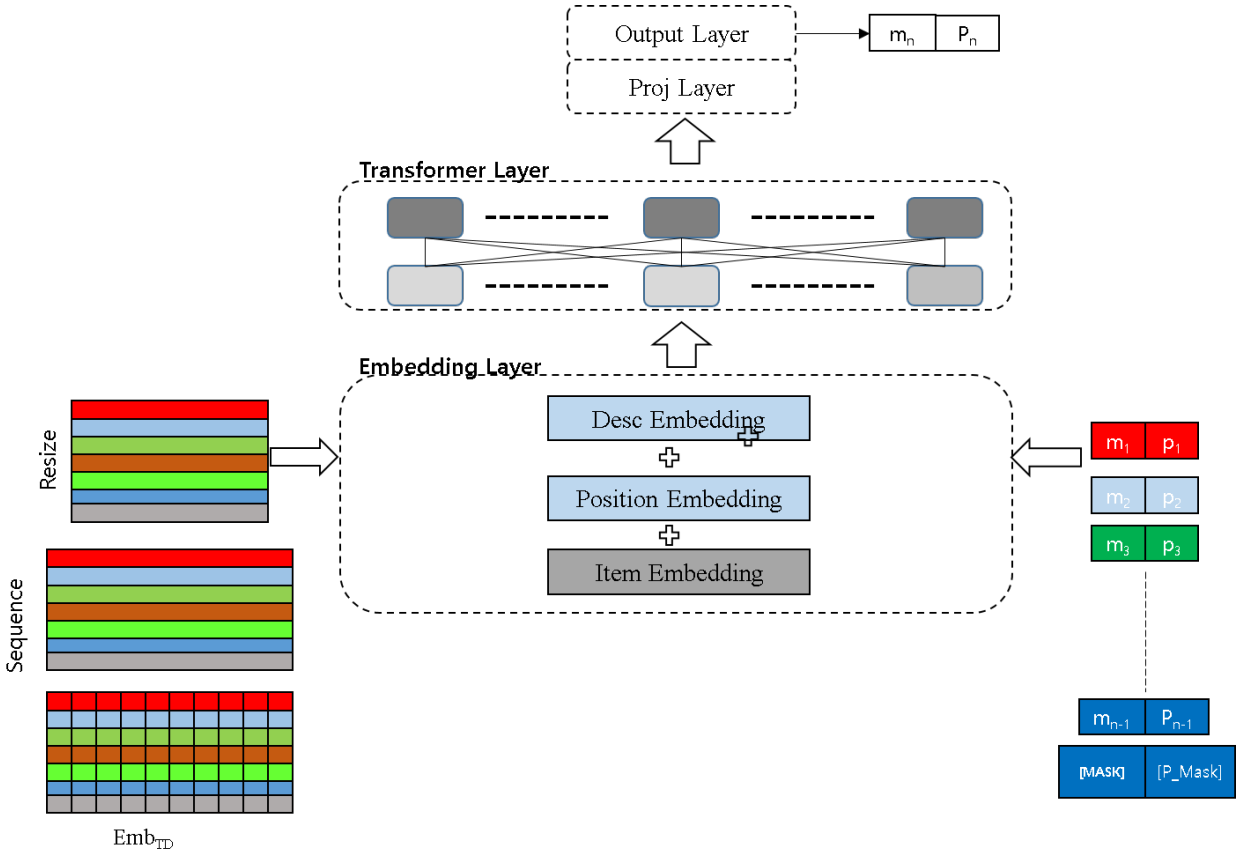


Fig. 2. Model architecture of contextual sequential RS.

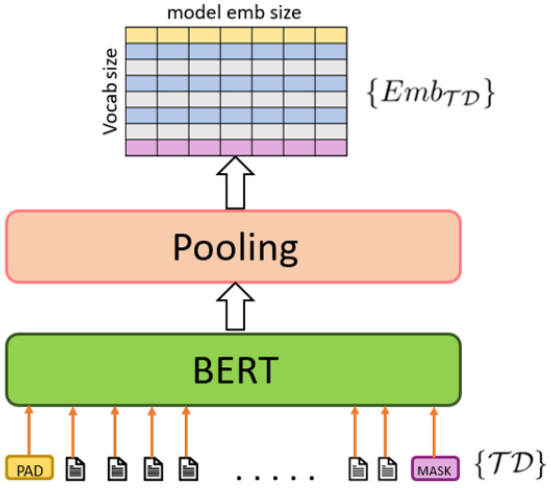


Fig. 3. Design architecture of sentence-BERT.

$$whiteB = Dropout\big(MHSA(\mathcal{E}^{l-1})\big) \tag{11}$$

$$S_i^{l-1} = LNorm\big(\mathcal{E}^{l-1} + B\big) \tag{12}$$

E. Output Layer

After passing through L layers and shared representations bidirectionally over each position in hierarchical manner, a final learned hidden representation $\mathcal{E}^{\mathcal{L}}$ is projected at output layer for all input item sequences. Considering the last item m_n in the sequence is masked, m_n is anticipated using embedding sequence $\mathcal{E}^{\mathcal{L}}$ that is depicted in Fig. 2. The last layer applies linear transformation twice followed by softmax function to predict the masked item.

$$whiteG = GELU\big(\mathcal{E}_t^L W^P + b^P\big) \tag{13}$$

$$P(m) = softamx\big(G(\mathcal{E}\mathcal{M}^T) + b^O\big) \tag{14}$$

where b^P and b^O are the bias at projection layer and W^P is the projection matrix. $\mathcal{E}\mathcal{M}^T$ is the item embedding matrix comprising of item identifier, positional and auxiliary information embedding. Here, shared item embedding is applied to minimize model size and relieve over-fitting.

IV. Experimental Configurations and Results Analysis

The datasets used to evaluate the proposed model and their preparation followed by experiment setup, evaluation metrics and performance comparison are presented in this section.

Most of the State-of-the-art SRS are trained on benchmark datasets that includes MovieLens, Amazon – Beauty. To effectively compare the improvements in the proposed model, these models are chosen. Moreover, auxiliary information of these datasets are also available and can easily be incorporated in these datasets. The auxiliary information, movie plot summary, of MovieLens dataset has been obtained through IMDbPy and for the Beauty Dataset, the description is chosen as a auxiliary information which is extracted from the meta file associated with the beauty dataset. Any item lacking auxiliary information has been ignored.

A. Dataset Pre-processing

Performance of proposed model is demonstrated through experiments carried out on three benchmark datasets including movielens-1m , movieLens-20m (ml-1m[1] and ml-20m[2]) and Amazon-Beauty[3] as described below:

1) MovieLens: A well-known dataset most commonly used for evaluating the performance of SRS. MovieLens ratings dataset contains the user id, item id (IDs of the movies from "movies" table), ratings and timestamp for movie ratings from each user. The auxiliary information (movie plot summary) for MovieLens is extracted through IMDbPY[4] using the **ImdbId** unique identifier, thus, making it information rich dataset.

2) Amazon - Beauty: It is a set of dataset comprising of reviews of a number of products extracted from "Amazon.com". The data is split into multiple datasets based upon product categories on Amazon. In our experiments, "Beauty" category is chosen that has a "rating" and a "meta" file. To incorporate the auxiliary information in the "rating" dataset, "description" of each product is extracted from the "meta" dataset.

The following Table II summarizes the dataset statistics.

TABLE II. DATASETS STATISTICS

Dataset	#users	#items	#interactions	Sparsity
ML-1m	6,040	3,706	1.0m	95.16%
ML-20m	138,493	26,744	20.0m	95.53%
Beauty	22,363	12,101	0.2m	99.93%

B. Evaluation Metrics

To measure the overall SR behavior, widely used leave-one-out strategy [8], [9], [15] is employed. The last item in each user's sequence is used for testing for every user, the second-to-last item is used for validation, and the remaining interaction items are utilized for training. For fair assessment, commonly used sampling practice [8], [9], [15] i.e. the ground truth object is coupled item with 100 negative items in test set based on how popular they are.

For evaluating all methods, "Normalized Discounted Cumulative Gain" (NDCG), "Hit Ratio" (HR) and "mean reciprocal rank" (MRR) are calculated. Higher values of these metrics depicts how better the recommendation performance is. Hit Ratio (HR) is used for measuring the ranking accuracy by comparing the test item set (T) with the ranked list. Mathematically it is expressed as:

$$HR@K = \frac{Numberof Hits@K}{|T|} \tag{15}$$

HR@K calculates the number of hits in a K-sized list. A hit occurs if the item tested is available in ranked list. Whereas the relative position of that item is assessed using NDCG in the ranked list. It assigns higher scores if the item is present at top position in the list. Mathematically it is evaluated by following formula:

$$NDCG@K = N_K \sum_{j=1}^{K} \frac{2^{z_j} - 1}{log_2(j+1)} \tag{16}$$

where N_K is the normalizer and z_j being the item's graded relevance at position j. We compute both the metrics of every test user items and then take their mean.

C. Baselines

For performance comparisons, we consider the following methods.

1) BPR-MF: [35] This model is the first one that uses the Bayesian personalized ranking loss for the optimization of matrix factorization (MF).

2) NCF: [36] This model utilizes MLP for capturing the item sequence interacted by user instead of using inner product in MF.

3) FPMC: [6] Combines MF with first-order MC to capture the long-term preferences of the user.

4) GRU4Rec: [12] It models the user click sequences using RNN-GRU for session based recommendation.

5) SASRec: [8] It is a unidirectional (left-to-right) self attentive model for next item prediction.

6) BERT4Rec: [9] This top of the line model uses bidirectional self attentive blocks and Cloze [37] masking for the recommendation task.

7) KeBERT4Rec: [15] This model extends BERT4Rec [9] by integrating keywords as additional input layer.

D. Implementation Details

The proposed model is trained on machine having 16 GB RAM and NVIDIA GTX 3080Ti (11GB). The training of proposed model is done using Adam Optimizer [38] with initial learning rate lr) of 0.001 and weight decay of 0.01. The hidden dimension is set to 128, dropout to 0.1 and 200 value used for maximum sequence length for MovieLens datasets and 50 value for Amazon Beauty. The masking probability of 0.15 is set for ML-20m and ML-1m. A 256 of batch size is used to train the proposed model.

The code provided by the corresponding authors of the respective baselines models were executed on the same machine. The optimized settings for hyperparameter values are used for all baseline models. The *hidden dimensionality* is tested from {64,128,256}, *dropout* from {0.1-0.9}, l_2 *regularizer* from {0-0.0001}.

[1] https://grouplens.org/datasets/movielens-1m/
[2] 1https://grouplens.org/datasets/movielens-20m/
[3] 3http:/jmcauley.ucsd.edu/data/amazon/
[4] https://imdbpy.github.io/

E. Comprehensive Performance Analysis

Tables III, IV and V presents the optimized outcomes of each baseline models on benchmark datasets. The highest scores in each table are shown in bold, while the 2nd place scores are underlined. The last row in each table displays how the proposed model performs in comparison to the best baseline model. The advantage of FPMC over BPR-MC is that it sequentially models users' previous records. From this observation, the importance of taking sequential pattern in consideration for recommendation systems can be ascertained.

Comparing the sequential baseline models, SASRec model outperforms GRU4Rec and FPMC on all benchmark datasets. This observation demonstrate that use of transformer based self attention models are more accurate than using traditional mechanisms. However, SASRec performance fall behind as compare to BERT4Rec which depicts that bidirectional model like BERT4Rec is more powerful as compared to unidirectional model like SASRec. BERT4Rec is a SR model that relies only on the item identifiers for the purpose of generating representation/ embedding. This model ignores the auxiliary information that is already provided with the datasets. However, KeBERT4Rec, a variant of BERT4Rec, has modified the representation by adding keywords describing the items e.g. Genre of movie.

TABLE III. Comprehensive Performance Analysis of Proposed Model with Referenced Models for Next Item Recommendations on ML-1m Dataset

ML-1m				
Metric	**HR@5**	**HR@10**	**NDCG@5**	**NDCG@10**
BPR-MF	0.2866	0.4301	0.1903	0.2365
NCF	0.1932	0.3477	0.1146	0.1640
FPMC	0.4297	0.5946	0.2885	0.3439
GRU4Rec	0.4673	0.6207	0.3196	0.3627
SASRec	0.5434	0.6629	0.3980	0.4368
BERT4-Rec	0.5876	0.6970	0.4454	0.4818
KeBERT4-Rec	0.5873	0.7651	0.5134	0.5488
ConSRec	**0.6690**	**0.7761**	**0.5287**	**0.5633**
Improvement	13.91%	1.44%	2.98%	2.64%

TABLE IV. Comprehensive Performance Analysis of Proposed Model with Referenced Models for Next Item Recommendations on ML-20m Dataset

ML-20M				
Metric	**HR@5**	**HR@10**	**NDCG@5**	**NDCG@10**
BPR-MF	0.2128	0.3538	0.1332	0.1786
NCF	0.1358	0.2922	0.0771	0.1271
FPMC	0.3601	0.5201	0.2239	0.2895
GRU4Rec	0.4657	0.5844	0.3091	0.3637
SASRec	0.5727	0.7136	0.4208	0.4665
BERT4-Rec	0.6325	0.7473	0.4967	0.5340
KeBERT4-Rec	0.8770	0.9450	0.7250	0.7470
ConSRec	**0.9863**	**0.9981**	**0.7687**	**0.8237**
Improvement	12.46%	5.62%	6.03%	10.27%

TABLE V. Comprehensive Performance Analysis of Proposed Model with Referenced Models for Next Item Recommendations on Beauty Dataset

Beauty				
Metric	**HR@5**	**HR@10**	**NDCG@5**	**NDCG@10**
BPR-MF	0.1209	0.1992	0.0814	0.1064
NCF	0.1305	0.2142	0.855	0.1124
FPMC	0.1387	0.2401	0.0902	0.1211
GRU4Rec	0.1315	0.2343	0.0812	0.1074
SASRec	0.1934	0.2653	0.1436	0.1633
BERT4-Rec	0.2207	0.3025	0.1599	0.1862
KeBERT4-Rec	0.3751	0.4753	0.2841	0.3164
ConSRec	**0.3884**	**0.5321**	**0.3261**	**0.3394**
Improvement	3.55%	11.95%	14.78%	7.27%

TABLE VI. Analysis on the Incorporation of Auxiliary Information

Model	**Metrics**	**BERT4Rec**	**ConSRec***	**ConSRec**
Beauty	**HR@10**	0.3025	0.3321	0.4631
	NDCG@10	0.1862	0.1922	0.3120
	MRR	0.1701	0.1653	0.2581
ML-1m	**HR@10**	0.6970	0.7023	0.7761
	NDCG@10	0.4818	0.4953	0.5633
	MRR	0.4254	0.4308	0.4484

TABLE VII. Impact of Using Contextual Embedding Technique

Dataset	**Metric**	**One-Hot Encoding**	**Word2Vec**	**Doc2Vec**	**SBERT**
Ml-1m	HR@1	0.3502	0.3601	0.3643	0.3672
	HR@5	0.6563	0.6589	0.6607	0.6690
	HR@10	0.7651	0.7690	0.7740	0.7761
	NDCG@5	0.5134	0.5198	0.5203	0.5287
	NDCG@10	0.5488	0.5590	0.5619	0.5633
	MRR	0.4322	0.4381	0.4443	0.4484
Beauty	HR@1	0.1897	0.1906	0.2012	0.2038
	HR@5	0.3432	0.3671	0.3874	0.3884
	HR@10	0.4983	0.5012	0.5296	0.5321
	NDCG@5	0.3079	0.3160	0.3256	0.3261
	NDCG@10	0.3187	0.3251	0.3361	0.3394
	MRR	0.2263	0.2476	0.2509	0.2517

The addition of this keywords embedding in the model makes KeBERT4Rec perform better than BERT4Rec. Thus, suggesting that incorporating some kind of side information along with item can improve the recommender's performance. It is evident from Table III that outcomes of all the sequential models like GRU4Rec, BERT4Rec, SASRec etc outperformed the non-sequential models like BPR-MC and NCF on dataset ml-1m. Our model outperformed in all baseline metrics showing the accuracy of model by incorporating the additional auxiliary information.

Result depicted in Table IV also indicates the importance of using Sentence-BERT to incorporates the contextual meaning of addition auxiliary information alongwith the item identifiers. Our model outperforms all baseline models. ConSRec gains an improvement of 5.62% on HR@10 and 10.27% on NDCG@10.

In accordance with the results, on the beauty dataset, Table V shows that our proposed model, ConSRec clearly

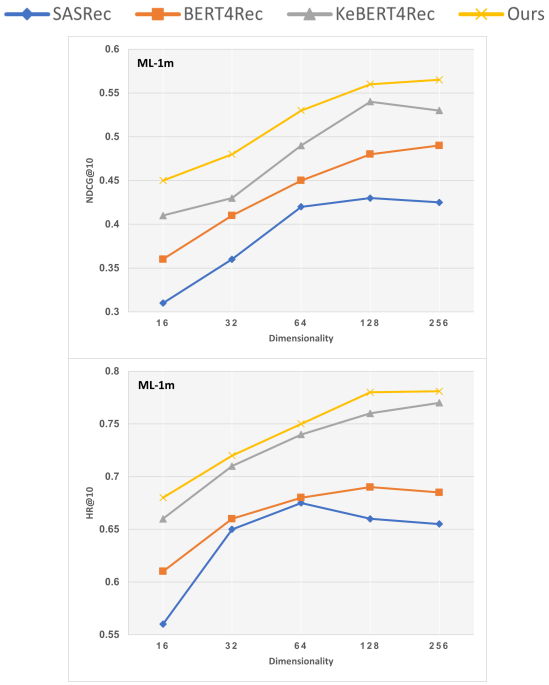


Fig. 4. Hidden dimensionality, d impact on NDCG@10 and HR@10 for ml-1m.

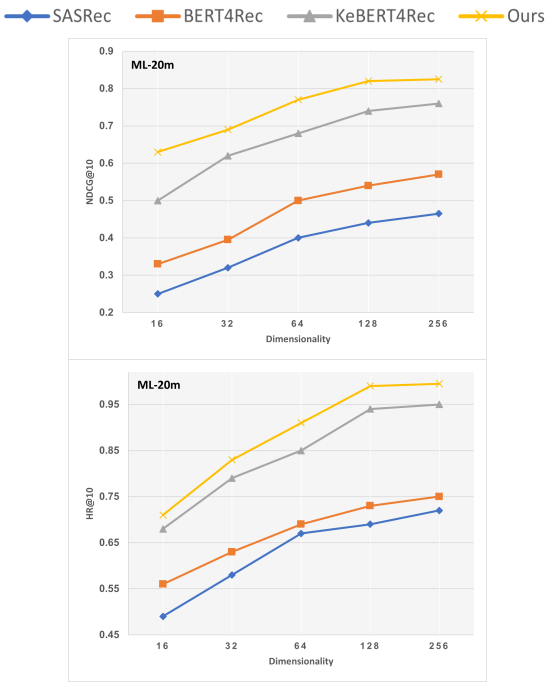


Fig. 5. Hidden dimensionality, d impact on NDCG@10 and HR@10 for ml-20m.

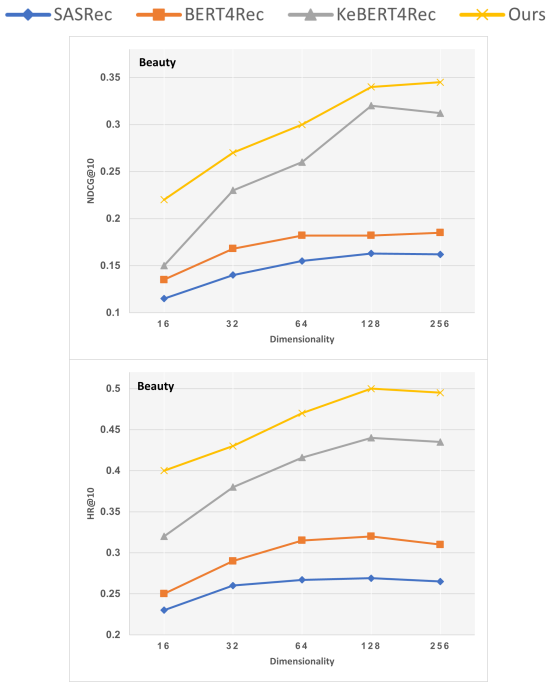


Fig. 6. Hidden dimensionality, d impact on NDCG@10 and HR@10 for beauty.

outperforms all baseline methods. The proposed model gains an average improvement of 11.95% on "HR@10" and 7.27% on "NDCG@10" as compared to the best baselines.

F. Evaluating Effect of "Hidden Dimensionality"

The hidden dimensionality *d* has a great impact on the performance of recommendation system that is studied in this section. Fig. 4, 5 and 6 exhibits the values of HR@10 and NDCG@10 on different baseline sequential model by varying hidden dimensionality *d* ranges between 16,32,64,128,256. The remaining of the hyperparameters are constant and kept to their optimal values.

The performance of ml-20m on varying dimensionality has very less impace on NDCG@10 and HR@10 as depicted in Fig. 5. However, bigger value of hidden dimensionality doesn't always yield accurate results.

It is obvious from Fig. 6 that with the increase of dimensionality, the graph of each model converges. However, improved model performance is not always achieved with bigger value of hidden dimensionality, particularly on sparse datasets, such as Beauty.

G. Impact of Integrating Auxiliary Information

As stated earlier, BERT4Rec only incorporated the embeddings of item identifiers along with its positional embedding in the input layer. However, in the proposed model, these embeddings are further enhanced and improved by incorporating additional contextual embedding layer of auxiliary information which is in the form of sentences.

The embeddings of this auxiliary information is extracted through Sentence-Transformer model that generates contextualized word embedding for all input tokens in the sentence.

To visualize this impact of using contextual information along with item identifier, the proposed model is initially trained by excluding the contextual information (ConSRec*). The results are then compared with the ConSRec i.e. by integrating side information. It is evident from the results that auxiliary information can enhance the productivity of SR system. Only the results on ml-1m and beauty dataset with batch size 128 are reported above in Table VI due to space limitations which clearly depicts that excluding the side information from proposed model degrades the performance.

H. Ablation Study

Finally, to visualize the impact of incorporating auxiliary information, some ablation experiments were conducted. Sentence-Transformer is used to train the proposed model which is a pre-trained model for generating embedding of item's side information. To analyze the impact of using contextual embedding instead of traditional techniques, the proposed model is tested using one hot encoding techniques to generate textual embedding.

As depicted in Table VII, By the use of Sentence Transformer to generate embeddings, the results of proposed model on ml-1m and beauty datasets outperforms all other non-contextual methods like word2vec, doc2vec, etc. This also emphasize the use of meaningful and context embedding generating technique for model training to produce relevant results.

V. Conclusion

Self Attention and Transformer based recommendation system have proven to be more precise and accurate as compared to traditional RS. In this paper, a transformer based sequential RS have been proposed that enhances recommendation accuracy by incorporating contextual auxiliary information of items in a sequence. The paper also uses contextual auxiliary information of items, such as descriptions or reviews, to enhance the recommendations. A contextual based pre-trained model sentence-transformer is used to generate meaningful embedding of auxiliary information. The experiments on various datasets show significant improvements over the state-of-the-art models.

However, ConSRec reliance on textual features as well as generalizability across highly diverse datasets beyond the domain of e-commerce and media streaming services poised limits to its capability. In future integration of additional multimodal data sources to further improve its performance and robustness.

References

[1] Zhiwei Liu, Mengting Wan, Stephen Guo, Kannan Achan, and Philip S Yu. 2020. Basconv: aggregating heterogeneous interactions for basket recommendation with graph convolutional neural network. In *Proceedings of the 2020 SIAM International Conference on Data Mining.* SIAM, 64–72.

[2] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *SIGIR.* 165–174.

[3] Liu, Zhiwei, et al. 2021. "Contrastive self-supervised sequential recommendation with robust augmentation." In *arXiv preprint arXiv:2108.06479.*

[4] Latifi, Sara, Dietmar Jannach, and Andrés Ferraro. 2022. "Sequential recommendation: A study on transformers, nearest neighbors and sampled metrics." In *Information Sciences* 609 (2022) : 660-678.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *CoRR abs/1810.04805*

[6] S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, 2010. "Factorizing personalized markov chains for next-basket recommendation," *in Proceedings of the 19th International Conference on World Wide Web, ser. WWW '10.* ACM, 2010, p. 811–820.

[7] J. Tang and K. Wang, 2018. "Personalized top-n sequential recommendation via convolutional sequence embedding," *in Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining,* ser. WSDM '18. ACM, 2018, p. 565–573.

[8] W.-C. Kang and J. McAuley, 2018. "Self-attentive sequential recommendation," in *2018 IEEE International Conference on Data Mining (ICDM).* IEEE, 2018, pp. 197–206.

[9] F. Sun, J. Liu, J. Wu, C. Pei, X. Lin, W. Ou, and P. Jiang, "Bert4rec: Sequential recommendation with bidirectional encoder representations from transformer," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, ser. CIKM '19. ACM, 2019, p. 1441–1450

[10] Guy Shani, David Heckerman, and Ronen I Brafman. 2005. An MDP-based recommender system. *Journal of Machine Learning Research* 6, Sep (2005), 1265-1295.

[11] Balázs Hidasi and Alexandros Karatzoglou. 2018. Recurrent Neural Networks with Top-k Gains for Session-based Recommendations. In *Proceedings of CIKM.* ACM, New York, NY, USA, 843–852

[12] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk, "Session-based recommendations with recurrent neural networks," *in 4th International Conference on Learning Representations*

[13] Jiaxi Tang and Ke Wang. 2018. Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding. In *Proceedings of WSDM.* 565–573.

[14] Vaswani, et. al. 2017 *Attention Is All You Need.* 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA.

[15] Elisabeth Fischer, Daniel Zoller, Alexander Dallmann, and Andreas Hotho. 2020. Integrating Keywords into BERT4Rec for Sequential Recommendation. In *KI 2020: Advances in Artificial Intelligence*.

[16] T. Zhang, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, D. Wang, G. Liu, and X. Zhou. 2019. Feature-level Deeper Self-Attention Network for Sequential Recommendation. In *IJCAI 2019. 4320–4326*.

[17] Aleksandr Petrov and Craig Macdonald. 2022. A Systematic Review and Replicability Study of BERT4Rec for Sequential Recommendation. In *Proc. RecSys*

[18] Ruining He and Julian McAuley. 2016. Fusing similarity models with markov chains for sparse sequential recommendation. In *2016 IEEE 16th International Conference on Data Mining (ICDM).* IEEE, 191–200.

[19] Steffen Rendle, Christoph Freudenthaler, and Lars Schmidt-Thieme. 2010. Factorizing personalized markov chains for next-basket recommendation. In *Proceedings of the 19th international conference on World wide web.* 811–820.

[20] Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Proceedings of EMNLP. Association for Computational Linguistics,* 1724–1734.

[21] Qiang Liu, Shu Wu, Diyi Wang, Zhaokang Li, and Liang Wang. 2016. Contextaware sequential recommendation. In *2016 IEEE 16th International Conference on Data Mining (ICDM).* IEEE, 1053–1058.

[22] Qiang Liu, Shu Wu, Liang Wang, and Tieniu Tan. 2016. Predicting the next location: A recurrent model with spatial and temporal contexts. In *Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30*

[23] Tim Donkers, Benedikt Loepp, and J"urgen Ziegler. 2017 Sequential User-based Recurrent Neural Network ecommendations. In *Proceedings of RecSys,* ACM, New York, NY, USA, 152–160.

[24] Massimo Quadrana, Alexandros Karatzoglou, Balázs Hidasi, and Paolo Cremonesi. 2017. Personalizing Session-based Recommendations with Hierarchical Recurrent Neural Networks. In *Proceedings of RecSys. ACM, New York, NY, USA, 130–137*

[25] Qiang Cui, Shu Wu, Qiang Liu, Wen Zhong, and Liang Wang. 2018. MV-RNN: A multi-view recurrent neural network for sequential recommendation. *IEEE Transactions on Knowledge and Data Engineering 32, 2 (2018),* 317–331.

[26] Fajie Yuan, Alexandros Karatzoglou, Ioannis Arapakis, Joemon M Jose, and Xiangnan He. 2019. A simple convolutional generative network for next item recommendation. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 582–590.

[27] Xu Chen, Hongteng Xu, Yongfeng Zhang, Jiaxi Tang, Yixin Cao, Zheng Qin, and Hongyuan Zha. 2018. Sequential Recommendation with User Memory Networks. In *Proceedings of WSDM. ACM, New York, NY, USA, 108–116.*

[28] Juyong Jiang, Jie Zhang, and Kai Zhang. 2020. Cascaded Semantic and Positional Self-Attention Network for Document Classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics, Online, 669–677.

[29] K. Xu, J. Ba, R. Kiros, K. Cho, A. C. Courville, R. Salakhutdinov, R. S. Zemel, and Y. Bengio, 2015. "Show, attend and tell: Neural image caption generation with visual attention," in *ICML.*

[30] Jianling Wang, Kaize Ding, Liangjie Hong, Huan Liu, and James Caverlee. 2020. Next-item Recommendation with Sequential Hypergraphs. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 1101–1110.

[31] Kyeongpil Kang, Junwoo Park, Wooyoung Kim, Hojung Choe, and Jaegul Choo. 2019. Recommender system using sequential and global preference via attention mechanism and topic modeling. In *Proceedings of the 28th ACM international conference on information and knowledge management*. 1543–1552.

[32] Potter, Michael, Hamlin Liu, Yash Lala, Christian Loanzon, and Yizhou Sun. 2022. "GRU4RecBE: A Hybrid Session-Based Movie Recommendation System *(Student Abstract).*"

[33] Zhou, Kun, Hui Wang, Wayne Xin Zhao, Yutao Zhu, Sirui Wang, Fuzheng Zhang, Zhongyuan Wang, and Ji-Rong Wen. 2020. "S3-rec: Self-supervised learning for sequential recommendation with mutual information maximization." In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, pp. 1893-1902.

[34] Reimers, Nils, and Iryna Gurevych. 2019. "Sentence-bert: Sentence embeddings using siamese bert-networks." *arXiv preprint arXiv:1908.10084 (2019).*

[35] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, 2009. "BPR: bayesian personalized ranking from implicit feedback," in *UAI, 2009*.

[36] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural Collaborative Filtering. In *Proceedings of WWW. 173–182.*

[37] Wilson L. Taylor. 1953. Cloze procedure: a new tool for measuring readability. Journalism & Mass Communication Quarterly 30 (1953), 415–433.

[38] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." In *Proceedings of ICLR*.

[39] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Computation 9, 8 (Nov. 1997), 1735–1780.

[40] Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. 2018. *Improving language understanding by generative pre-training.* In OpenAI Technical report.

On the Context-Aware Anomaly Detection in Vehicular Networks

Mohammed Abdullatif H. Aljaafari
Department of Management Information Systems-School of Business,
King Faisal University, Alhufof 31982, Saudi Arabia

***Abstract*—Transportation systems are moving towards autonomous and intelligent vehicles due to advancements in embedded systems, control algorithms, and wireless communications. By enabling connectivity among vehicles, a vehicular network can be developed which offers safe and efficient driving applications. Security is a major challenge for vehicular networks as application reliability depends on it. In this paper, we highlight the security challenges faced by a vehicular network especially related to jamming and data integrity attacks. Such attacks cause major disruptions in the wireless connectivity of users with the centralized servers. We propose a context-aware anomaly detection technique for vehicular networks that considers factors such as signal strength, mobility, and data pattern to find abnormal behaviors and malicious users. We further discuss how an intelligent learning system can be developed using efficient anomaly detection. We implement a vehicular network scenario with malicious users and provide simulation results to highlight the performance gain of the proposed technique. We also highlight several appropriate future opportunities related to the security of vehicular network applications.**

***Keywords*—Fog computing; load balancing; task offloading**

I. Introduction

Intelligent Transportation Systems (ITS) are an important component of smart cities. The goal of ITS is to introduce intelligence, connectivity, and control capabilities in the vehicles such that driving can be more safer and comfortable [1]. A vehicular network is an important part of ITS as it provides wireless communication between different components of ITS [2], [3], [4], [5], [6], [7], [8]. The major advantage of vehicular networks is that it can continuously monitor the location of all vehicles within the city and guide the vehicle to make intelligent driving decisions [9], [10], [11]. This can be particularly useful for managing city level traffic flow and accident free driving on the roads [12], [13], [14], [15], [16], [17].

Security and privacy are important aspects of future wireless communications and vehicular networks [18], [19], [20], [21], [22], [23], [24]. Without secure communications, many vehicular network applications can be compromised and sensitive data can be leaked. Thus, security is an important part of reliable communications [25], [26], [27], [28].

Vehicular networks suffer from many attacks by intruders and malicious nodes [29], [30], [31]. These attacks impact the quality of wireless communications, the confidentiality of data transmission, the integrity of messages shared between nodes, and the availability of information among all the devices. By ensuring all these requirements of secure communications, applications' reliability can be significantly improved [9].

Few cyber attacks that target vehicular applications include data integrity attacks, eavesdropping, and jamming. In data integrity attacks, malicious users send wrong information to the server so that data used for making decisions can be polluted [32], [33], [34]. Since most decisions made by the applications rely on the quality of data and its accuracy, it is critical to detect these types of attacks [35], [36], [37].

In eavesdropping attacks, the malicious user captures the data sent between the vehicles and the server [38], [39], [40]. This can be dangerous for applications where sensitive information can be captured and used to impact the vehicle safety [41], [42], [43]. In jamming attacks, malicious users transmit signals to cause interference for other users in the network. This reduces reliability of communications as signal to noise ratio is reduced [44], [45], [46], [47], [48], [49].

To overcome these challenges in the security of vehicular network applications, cryptographic schemes such as Elliptic Curve Digital Signature (ECDSA) are commonly used. By adding extra encryption information to the original message, it is made secure and can only be decoded by users with the correct key information. While ECDSA can make the information and data transmission secure, it is still impacted by cyber-attacks. Particularly, all messages are verified for correct signature information at the receiver. As a result, a lot of time is wasted verifying fake messages by malicious users. This impacts the Quality of Service (QoS) of applications as a large end-to-end delay is incurred.

The data generated by the malicious users can be considered an anomaly or abnormal behavior [50], [51], [52], [53], [54]. The anomalies need to be efficiently detected in a secure system and malicious user data should be recorded for future attack prevention. The goal of anomaly-based security is to complete the process of detection in a quick time and with good accuracy [55], [56], [57], [58].

In this paper, we present a new framework for anomaly detection and verification in vehicular network applications. The key idea of the proposed framework is to develop a trust table in the server for all associated vehicular nodes in the network. The framework uses contextual information related to message physical parameters, mobility of the user, and data patterns to compute trust and detect anomalies. Moreover, the framework also includes an anomaly verification procedure so that the trust values in the trust table can be updated periodically on the server. Simulation results in MATLAB provide a detailed performance evaluation of the proposed technique compared to the other recent techniques in the literature. Finally, we also present several research challenges and future opportunities in the area of vehicular network security.

Table I below shows the anomaly detection techniques in the literature.

II. Literature Review

Security and privacy in vehicular and IoT systems have been explored in several papers in the literature. Similarly, anomaly detection techniques have been part of many proposals. We discuss some of these techniques in this section.

The work in [59] proposes a novel anomaly detection technique for IoT networks. The major problem addressed in this paper is related to class imbalance i.e. when normal data is much larger than the abnormal one or vice versa. The data set considered in this paper is Network Security Laboratory-Knowledge Discovery and Data Mining Tools Competition (NSL-KDD). The technique used in this paper is reinforcement learning in which actions are classification of input data into normal and malicious categories. The states in the paper are the data type. As the data in an IoT network can be of different types. Hence, the state takes into account the data category. The reward function in the paper is anomaly prediction accuracy. The proposed work shows better accuracy, recall, and F1 score.

In [60], the authors focus on a network related to Industrial IoT (IIoT). The goal is to detect cyber attacks and a federated learning-based approach is used in this regard. The major advantage of the federated learning approach is its privacy preservation as data is only shared locally. In the proposed work, universal anomaly detection is achieved with the help of different local Anomaly Detection Centers (ADC). Moreover, anomalous ADCs are also detected with the proposed technique. There is also an appeal procedure reserved for users that are declared anomalous. The accuracy and throughput of the proposed technique are shown to be better than other related techniques.

The work in [61] deals with traffic flow monitoring applications for Software Defined Networks (SDN). The goal of the technique is to choose traffic flow monitoring granularity. There exists a tradeoff between accuracy and network load. More accurate techniques may incur a large network load, hence a balancing technique is needed. The proposed technique uses Deep-Q learning to detect anomalies. The proposed technique achieves quick and optimal policy evaluation. Particularly against the Denial of Service (DDoS) attacks, the proposed technique performs efficiently. The accuracy and detection time of the proposed technique has been shown to perform better than other techniques in the literature.

In [56], the authors considered an Internet of Vehicles (IoV) scenario where vehicles share information about the surrounding traffic. Parameters such as traffic density, vehicles in an emergency, vehicle speeds, etc. are shared with infrastructure Road Side Units (RSUs). Malicious users launch data integrity attacks i.e., they change the information about traffic density and send wrong information. As a result, the traffic density estimation is corrupted and wrong decisions are made. To overcome this problem, the authors propose an isolation forest-based anomaly detection algorithm. The anomalies are verified using probe messages that are sent to vehicles in the neighborhood of malicious users. A communication mechanism is also designed to share the verification information. Results in terms of accuracy, recall, and F1 score of the proposed technique is enhanced.

The social networks are considered as part of the work in [62]. The main problem addressed is related to feature learning and combining information from the neighborhood. Based on feature gathering, Graph Neural Network (GNN) technique is proposed that uses GNN based encoder for feature learning. For efficient training, pattern mining algorithms are used by the proposed technique. A novel loss function is also proposed in the work. Metrics such as precision, recall, and F1 score are shown to be improved as compared to other techniques.

The work in [63] deals with improving the security of the Domain Name System (DNS). The key idea used is to make the system topology aware and consider the structural properties of the network. The technique used is the exponential random graph model and the topology is converted into the graph format. A time series analysis is also applied for anomaly detection. The autoregressive moving average is used for the time series analysis. The precision and recall of the system are improved as compared with other techniques.

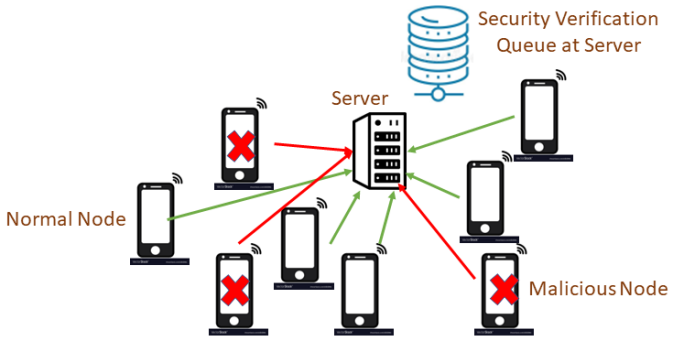


Fig. 1. Considered system model.

III. System Model

In this paper, we consider a wireless transmission scenario where several vehicular nodes are sharing their data with a centralized server or Road Side Unit (RSU) as shown in Fig. 1. This data can be traffic mobility information, tasks for offloading, service query requests, etc. The server authenticates each vehicular user using traditional cryptographic mechanisms. Moreover, vehicular users also share their messages using digital signature schemes.

Several malicious vehicle nodes are also present in the vicinity of the normal vehicle nodes and servers. These users can carry out several types of attacks such as jamming signal transmissions, eavesdropping, data integrity attacks, etc. These attacks can significantly disrupt the data security and reliability of the wireless system.

The server receives messages from both normal vehicle users and malicious vehicle users. These messages are placed in a queue and a digital signature verification algorithm is applied to messages one by one. The messages from malicious users result in additional queuing delays at the server. Some messages may be discovered to be malicious after the signature

TABLE I. ANOMALY DETECTION TECHNIQUES IN THE LITERATURE

Network	Main Idea	Technique Used	Results
IoT [59]	Anomaly Detection Class imbalance problem NSL-KDD data set	Reinforcement Learning Action - Classify input data State - Data type Reward - Prediction accuracy	Accuracy Recall F1 score
IIoT [60]	Detect cyber attacks Federated learning Privacy preservation	Universal anomaly detection Anomalous ADC detection Appeal procedure for users	Accuracy Throughput
SDN [61]	Traffic flow monitoring Monitoring granularity Accuracy vs network load	Deep Q leaning Quick optimal policy evaluation DDoS attack detection	Accuracy Detection time
IoV [56]	Data Integrity attacks Traffic density information Information checking	Isolation Forest Verify anomalies Neighborhood verification	Accuracy Recall F1 score
Social networks [62]	Feature learning Combine neighborhood information	Graph Neural Network (GNN) GNN encoder for feature learning Pattern mining algorithms for training Novel loss function	Precision Recall F1 score
DNS [63]	Topology aware Structural properties of network	Exponential Random Graph Model Time series analysis Auto regressive moving average	Precision Recall

verification, but the time spent in their verification causes an extra delay for the messages.

In addition, some malicious messages have the correct digital signature but are launched to disrupt the wireless transmissions. These messages pass through the digital signature verification stage but impact the system latency. Without an efficient anomaly detection technique, these messages can not be detected on the run. Even their malicious behavior is not learned due to the absence of an anomaly detection mechanism and can not be detected in the future.

IV. PROPOSED ANOMALY DETECTION FRAMEWORK

The proposed anomaly detection framework is shown in Fig. 2. The server maintains an anomaly detection module in its storage. The anomaly detection module consists of several blocks which are explained in the following subsections. The goal of the proposed framework is to shortlist and sideline anomalies so that server does not waste time in signature verification of malicious user messages.

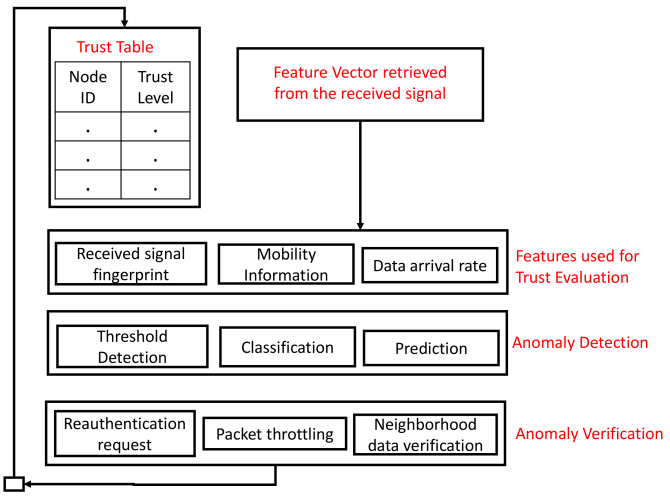


Fig. 2. Proposed context-aware anomaly detection framework.

A. Trust Table

Each server maintains a trust table that contains trust levels computed for each vehicle in the network. The trust table is a lookup table that contains two columns and several rows. The columns include information about node ID and trust level. The rows are equal to the number of vehicles in the network of the server.

A new vehicle that is entered in the coverage range of the server is registered in the trust table once it starts sharing its data with the server. Based on the data received from the vehicle nodes, the trust table is updated on regular basis i.e., the trust values are updated.

A timer value is used by the server to keep a check on the current status of the vehicle nodes in the network. Vehicles that do not send data for a long time are removed from the trust table. The reason for this is that either the vehicles have moved out of the coverage range of the server or the device has been turned off. As a result, the trust look-up table is shortened by the removal of such devices. This reduces the search time for a device's trust from the look-up table.

Trust table is evaluated based on fuzzy logic as shown in Table II. We explain the features used for developing the trust table in the next subsections.

TABLE II. TRUST EVALUATION BASED ON FUZZY LOGIC

Trust Feature	Trust Value Threshold		
	Low	Medium	High
Received Signal Fingerprint (F_{inc})	≥ 0.6	0.3-0.6	0-0.3
Mobility Information (T_r^{inc})	0-100 ms	100 ms-500 ms	500 ms-1 s
Data Arrival Rate (D_A)	$\geq 20p/s$	10-20 p/s	0-10 p/s

B. Trust Evaluation

Trust evaluation is the most critical block of the anomaly detection module. Trust in our proposal is computed based on several features considering the general behavior of malicious vehicles. The proposal is context-aware in the sense that the environment, data transmission, and mobility factors related to vehicular network are considered while evaluating trust.

The trust evaluation in our proposal is based on both the historic long-term data of the used features as well as the current feature vector. An average value of the features used in our proposal is maintained by the server. The advantage of storing long-term data is to evaluate the trend of the features and efficiently detect anomalies.

The following vehicle features are used for the trust evaluation in the proposed framework. The evaluation mechanism of each feature and its rationale are presented below. It is to be noted that these are physical features that are obtained using the obtained signal. These features are selected before the signature verification mechanism so that malicious messages are not verified.

1) Received signal-based fingerprint: The received signal can provide useful identification of devices in the network. In this regard, Signal Interference and Noise Ratio (SINR) and Angle of Arrival (AoA) retrieved from the signal can act as a fingerprint for a device. The received power from a device can be calculated as follows:

$$P_r = G_c \frac{P_t}{d^\gamma} \tag{1}$$

Here P_r is the received power, G_c is the channel gain, P_t is the transmitted power, d is the distance between the transmitter and the receiver, and γ is the path loss index. The value of the path loss index can vary between 2 and 4 depending on the scenario. The channel gain also varies depending on the channel conditions. In many scenarios, multi-path fading introduces varying channel gains depending on the amount of reflection and refraction from different objects.

The SINR at the server can be computed from the following equation:

$$SINR = \frac{P_r}{I + N_O} \tag{2}$$

where

$$I = \sum_{v=1}^{M} P_{r,v} \tag{3}$$

From Eq. (2), SINR is evaluated from the ratio of received power P_r and total noise in the system. There are two components of total noise. One is the background noise N_O and the other is the total interference. If the server receives v signals at the same time, the total interference can be computed from Eq.(3). This is the summation of all signals transmitted from those v vehicles.

Using the received signal, a map between physical location and user is developed. Any malicious user that is sending wrong location information to launch jamming attacks may be identified. The fingerprint can also be used to find areas that are critical concerning security and have more density of malicious users.

As shown in Table II, we map the received signal finger print value to the trust value using three categories; low, medium and high. The server tracks the fingerprint value of each vehicle based on its current location as follows:

$$F = \frac{SINR}{d} \tag{4}$$

The incremental change in fingerprint F_{inc} is given as

$$F_{inc} = \frac{|F_{new} - F_{old}|}{F_{old}} \tag{5}$$

If F_{inc} is greater than 0.6, the corresponding trust value can be mapped to a low value. Similarly, a medium trust value means a F_{inc} value between 0.3 and 0.6. Lastly, a high trust value means a low F_{inc} value of less than 0.3. The rationale behind this approach is that malicious vehicles will be sending

either wrong location information in its message or using high transmit power from its correct location for the attack. This factor is captures in the metric F_{inc}.

2) Mobility information: Mobility is an important parameter that can provide useful information about the normal behavior of nodes. Malicious users may periodically change their position to transmit their information so that they are off the radar. As a result, the percentage change in location of nodes is determined and those devices whose position is changing randomly in an abnormal manner is shortlisted for anomaly evaluation.

Since mobility information can only be obtained once the message is verified and decoded, it is not possible to get this information at the physical level. To overcome this problem, signal power is used as an estimate of the position of the sender. In literature, some techniques can estimate the location of the sender using received signal power [64]. The sender location is estimated in the form of distance bins rather than the exact longitude and latitude.

The distance bin of the initial location of the sender is noted based on the last message sent to the server. The node ID of the sender is noted down once the signature of the message is verified. A local map for node ID and its distance bin is stored in the server. Once, the same node changes its location and distance bin by a certain threshold for the next message, the node is marked as a possible malicious node and short-listed for the verification phase.

In Table II we use time to reach from vehicle to server (T_r) as the metric to evaluate change in sudden location of the vehicle. T_r is evaluated as follows:

$$T_r = \frac{d}{speed} \tag{6}$$

The incremental change in T_r^{inc} is given as

$$T_r^{inc} = |\frac{T_r^{new} - T_r^{old}|}{T_r^{old}} \tag{7}$$

A low trust value refers to a high T_r^{inc} which means either the speed of the vehicle or the location of the vehicle obtained from the received message is suddenly changed.

3) Data arrival rate: The data arrival rate can also provide information about the possible malicious behavior of a node. Malicious users may try to jam the network by sending repeated requests for task computation or sending incorrect data messages as part of data integrity attacks. The data inter-arrival rate for a particular node is computed as the difference between the time when the current message is received and the time when the last message was sent. This time is computed once the messages are verified for digital signature correction.

To overcome this problem, the inter-arrival time between two messages is checked regularly. Once a certain arrival rate is crossed, those nodes is marked as malicious for checking. Besides data arrival rate, message size can also be checked as malicious users may transmit large packets to negatively impact the bandwidth usage and increase the signature verification time.

4) Data variation: The data variation is a critical metric that can be used to detect anomalies specifically in the cases of malicious users that are sending wrong unrelated information. An example of this data integrity attack could be a traffic management application where wrong traffic density values are shared with the server. As a result, data variation sent from sensors located in a geographical location should raise alarms and should be checked for anomalies. In Table II, we utilize a data arrival rate of greater than 20 packets per second (p/s) to be marked as a value with low trust. This is because generally, the packet frequency of vehicular network varies from 1 to 10 packets/second. Hence, we give a higher trust value to nodes which conform with this requirement and send packets within the allowed range of 1-10 packets per second.

C. Anomaly Detection

In this phase, nodes that are marked as anomalous or malicious are noted down and a list is maintained. This list includes all the nodes that are on the radar for anomaly detection and a further investigation of their status is needed. For maintaining this list, techniques such as threshold detection, cumulative distribution function evaluation, standard deviation, etc. is applied to the features data from the previous block.

There exists a trade-off between the accuracy of anomaly detection and the time for evaluating the anomaly. If the threshold is kept too strict, only a few users may be marked as anomalous and anomaly verification is done quickly. However, many malicious nodes may be missed. On the other hand, if the threshold is kept too loose, many users will be short-listed for anomaly verification. This will need more message overhead and time needed to verify each anomaly. As a result, extra bandwidth and time will be needed.

In this work, we utilize fuzzy logic to evaluate anomalies. We first calculate trust levels based on individual features as described in Table II. Finally, we utilize majority voting to find the overall trust level. This means that the trust level is categorized as high if the trust based on at least two of the features are evaluated as high. Data from all nodes which have a low and medium trust levels are marked as anomalous.

D. Anomaly Verification

In this block, the shortlisted malicious users is tracked for further verification so that actual anomalies are picked up. For this, short-listed users send a message to inform them about their entry into the malicious node list. Those nodes may need to re-authenticate themselves before sending any further messages. All those nodes that can re-authenticate are allowed to start transmitting their messages again.

Another way to verify anomalies used in the proposed technique is packet throttling requests. The users are asked to reduce their data rates. All users that follow this request are allowed to continue sending messages. However, those nodes that do not follow the instructions and keep on sending a large number of messages which are marked as malicious users who are launching jamming attacks.

One technique to verify anomalies is neighborhood verification. This is particularly useful for data integrity attacks. The area of those nodes that are sending wrong information

and marked as possible malicious nodes are noted down. The list of neighboring nodes are sent probe messages to find if the information sent by malicious nodes is correct or not. This technique however is suited for those applications in which the sensed information in the neighborhood is nearly the same, for example, temperature monitoring in an area, average vehicle traffic density on a road segment, etc.

TABLE III. SIMULATION PARAMETERS

Simulation Parameter	Value
Number of vehicular nodes	200
Packet Generation Rate	10 per second
Number of Malicious users	50-200
Security Algorithm	ECDSA
Hash Function Used	SHA-256
Message size	7400 bits
Security overhead	1600 bits
Total Packet Size	9000 bits
Security Verification Time	0.5ms
Packet expiry time	3ms

V. PERFORMANCE EVALUATION

In this section, the performance evaluation of the Proposed Technique (PT) is presented. The results are compared with two other techniques. The first one is the Outlier Detection, Prioritization, and Verification (ODPV) algorithm which uses an isolation forest algorithm to detect anomalies. The second algorithm is the K-Nearest Neighbor (KNN) algorithm which uses majority voting to make decisions about the anomaly.

A. Simulation Model

The simulation model is developed in MATLAB and its parameters are given in Table III. The number of vehicular nodes is taken as 200. The packet generation rate of each vehicular node is 10 packets per second. The number of malicious users is 50-200. The malicious users also generate packets at the same rate as the malicious users with wrong information about the sensing parameters and causing data integrity attacks. Also, the malicious users continuously change their positions. As a result, received signal strength, mobility, and data variation factors come into play.

The security algorithm used is the Elliptic Curve Digital Signature (ECDSA) algorithm and the hash function used is the Secure Hashing Algorithm (SHA) with a key size of 256 bits. The message size generated by vehicular devices is 7400 bits and the security overhead for the ECDSA algorithm is 1600 bits [9]. The total packet size is 9000 bits. The security verification time for each packet is 0.5ms. The packet expiry time is 3ms.

B. Results

For results, we take the following four metrics defined as follows:

1) Security verification time: This is the time needed to verify the digital signature as per the ECDSA algorithm. The security verification time depends on the key bit size used by the ECDSA algorithm. Moreover, the security verification time increases with the number of messages.

2) Total delay: The delay is the sum of the transmission time and security verification time. The transmission time depends on the message size and data rate.

3) Percentage of packets expired: The percentage of packets expired includes the percentage of packets that could not be verified within the time limit of 3ms.

4) Shortlisted packets and anomaly packets: The shortlisted packets include the number of packets that were originally shortlisted by the anomaly detection technique. The anomaly packets are the number of packets that were declared as anomaly after verification.

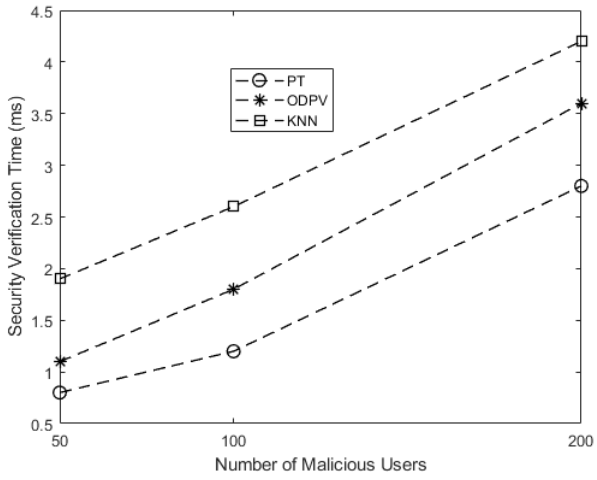


Fig. 3. Security verification time vs Number of malicious users.

Fig. 3 shows the security verification time of packets as the number of malicious users is increased from 50 to 200. The results show that the security verification time of the proposed technique is up to 1ms better than the other two techniques namely, ODPV and KNN. This is due to using context-aware parameters for trust evaluation and anomaly detection. As a result, the anomaly packets are identified after the first few iterations, and the overall network load on the server is reduced. This results in reduced security verification time.

In Fig. 4, the plot of malicious users vs total delay is shown for the three algorithms. The proposed technique has the least delay in packet transmission and hence, is very useful for vehicular applications. As compared to ODPV and KNN, the delay values of the proposed technique show up to 50% reduction. As the number of malicious users increased up to 200, the results of the proposed technique are much better than the other two techniques.

Fig. 5 shows the results of the percentage of packets expired for the three algorithms. The proposed technique has excellent results with only less than 10% packet expiry rate

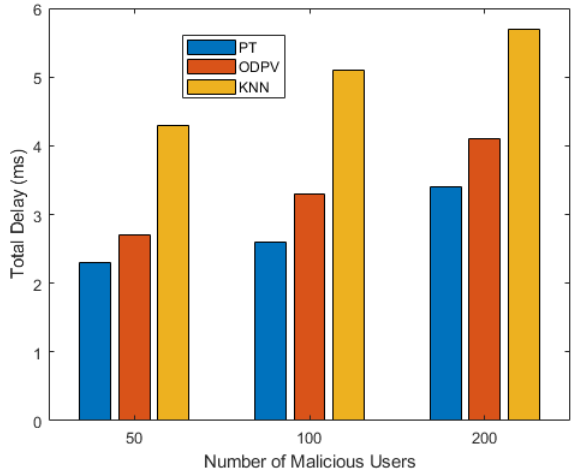


Fig. 4. Total delay vs Number of malicious users.

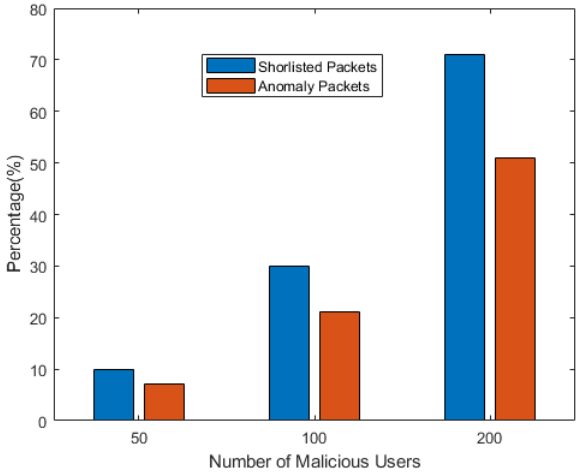


Fig. 6. Shortlisted packets vs Anomaly packets by the proposed technique.

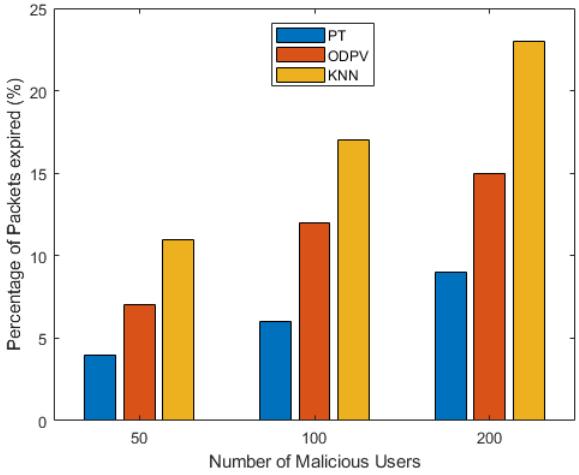


Fig. 5. Percentage of packets expired vs Number of malicious users.

as compared to the ODPV and KNN that has up to 15% and 22% packets expiry rate.

Lastly, Fig. 6 shows the anomaly detection accuracy performance of the proposed technique. The proposed technique almost accurately shortlists the packets which are anomalous when the number of malicious users is 50. As the malicious users increased, the difference between shortlisted packets and anomaly packets increased. This is because, at a lower number of nodes, the anomalous packets are easier to detect after a few iterations. As the number of packets is increased, less number of packets meet all the context-related conditions and hence fall into the likely malicious category. However, after verification, the anomaly packets out of the shortlisted ones are picked up by the algorithm.

VI. Future Opportunities

In this section, future opportunities are provided related to the challenges in the security of vehicular applications.

A. Learning based Algorithms

Machine learning-based algorithms can be designed to adaptively vary the trust levels using context-aware data from different vehicular devices. Appropriate neural networks can be developed to predict the users which are malicious or geographical areas where malicious users reside. Moreover, reinforcement learning algorithms can also be useful for optimal security-related policy design such as priority-based security verification of incoming messages. Similarly, anomaly detection can also be improved using classification algorithms based on data of malicious and normal users. Related with the current work, a learning algorithm can be used to find the appropriate threshold for anomaly detection based on incremental changes in the fingerprint, location and speed values of the vehicles.

B. Graph based Algorithms

Graph theory algorithms are very useful for evaluating the trust of vehicular devices. For example, metrics such as centrality and degree of nodes can be taken into account when evaluating the trust. Stable matching algorithms can also be used for sending authentication requests by vehicular devices to appropriate servers depending on factors such as data rate, and load on the servers. Similarly, breadth-first search and depth-first search algorithms can be used for searching anomalies from the data sets, and also for finding an efficient route for reaching the nodes for data verification. Graph theory algorithms can also be used to verify anomalies from vehicles with high centrality values or vehicles with highest trust values. This will improve the accuracy of the verification process.

C. Blockchain Techniques

Besides, trust-based anomaly detection techniques, blockchain security mechanisms can also be used to enhance the privacy of data sharing. With blockchain, smaller key sizes of the ECDSA algorithm can be used as added security will be provided with blockchain. However, a major challenge in designing such techniques will be the delay in mining the blockchain nodes and appropriate consensus

algorithms. The proposed work of anomaly detection can also be used to find malicious mining nodes which can disrupt the security of blockchain. Moreover, new context-aware algorithms considering different features can be designed for the particular blockchain mining application.

VII. Conclusion

In this paper, we present an overview of anomaly detection for vehicular applications. We discuss recent work in the literature related to anomaly detection. We present a novel framework for anomaly detection which uses context-related information such as physical signal parameters to classify anomalies without security verification. The proposed technique also includes an anomaly verification phase and develops a trust table for each node in the network. Finally, we highlight several important research directions in the area of security of vehicular and anomaly detection.

References

[1] M. A. Javed, T. N. Nguyen, J. Mirza, J. Ahmed, and B. Ali, "Reliable communications for cybertwin driven 6g iovs using intelligent reflecting surfaces," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2022.

[2] G. Xie, K. Yang, C. Xu, R. Li, and S. Hu, "Digital twinning based adaptive development environment for automotive cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1387–1396, 2022.

[3] W. Duo, M. Zhou, and A. Abusorrah, "A survey of cyber attacks on cyber physical systems: Recent advances and challenges," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 5, pp. 784–800, 2022.

[4] X. Zhang, H. Ma, and C. K. Tse, "Assessing the robustness of cyber-physical power systems by considering wide-area protection functions," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 12, no. 1, pp. 107–114, 2022.

[5] X. Ning and J. Jiang, "Design, analysis and implementation of a security assessment/enhancement platform for cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1154–1164, 2022.

[6] M. Hussain, N. Ali, and J.-E. Hong, "Deepguard: a framework for safeguarding autonomous driving systems from inconsistent behaviour," *Automated Software Engineering*, vol. 29, no. 1, p. 1, 2022.

[7] J. Valinejad, L. Mili, C. N. van der Wal, and Y. Xu, "Environomic-based social demand response in cyber-physical-social power systems," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 69, no. 3, pp. 1302–1306, 2022.

[8] S. V. Thiruloga, V. K. Kukkala, and S. Pasricha, "Tenet: Temporal cnn with attention for anomaly detection in automotive cyber-physical systems," in *2022 27th Asia and South Pacific Design Automation Conference (ASP-DAC)*, 2022, pp. 326–331.

[9] M. A. Javed, E. B. Hamida, and W. Znaidi, "Security in intelligent transport systems for smart cities: From theory to practice." *Sensors*, vol. 16, no. 6, p. 879, July 2016.

[10] D. Zhao, C. Liu, G. Xu, Z. Ding, H. Peng, J. Yu, and J. Han, "A security enhancement model based on switching edge strategy in interdependent heterogeneous cyber-physical systems," *China Communications*, vol. 19, no. 2, pp. 158–173, 2022.

[11] D. Cheng, J. Shang, and T. Chen, "Finite-horizon strictly stealthy deterministic attacks on cyber-physical systems," *IEEE Control Systems Letters*, vol. 6, pp. 1640–1645, 2022.

[12] J. Zhang, L. Pan, Q.-L. Han, C. Chen, S. Wen, and Y. Xiang, "Deep learning based attack detection for cyber-physical system cybersecurity: A survey," *IEEE/CAA Journal of Automatica Sinica*, vol. 9, no. 3, pp. 377–391, 2022.

[13] F. Li, R. Xie, B. Yang, L. Guo, P. Ma, J. Shi, J. Ye, and W. Song, "Detection and identification of cyber and physical attacks on distribution power grids with pvs: An online high-dimensional data-driven approach," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 10, no. 1, pp. 1282–1291, 2022.

[14] G. Sun, T. Alpcan, B. I. P. Rubinstein, and S. Camtepe, "Securing cyber-physical systems: Physics-enhanced adversarial learning for autonomous platoons," in *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part III*. Berlin, Heidelberg: Springer-Verlag, 2023, p. 269–285.

[15] C. Qian, X. Liu, C. Ripley, M. Qian, F. Liang, and W. Yu, "Digital twin-cyber replica of physical things: Architecture, applications and future research directions," *Future Internet*, vol. 14, no. 2, 2022. [Online]. Available: https://www.mdpi.com/1999-5903/14/2/64

[16] R. Müller, F. Kessler, D. W. Humphrey, and J. Rahm, "Data in context: How digital transformation can support human reasoning in cyber-physical production systems," *Future Internet*, vol. 13, no. 6, 2021. [Online]. Available: https://www.mdpi.com/1999-5903/13/6/156

[17] M. Hussain and J.-E. Hong, "Enforcing safety in cooperative perception of autonomous driving systems through logistic chaos map-based end-to-end encryption," in *2022 16th International Conference on Open Source Systems and Technologies (ICOSST)*, 2022, pp. 1–6.

[18] G. Li, C. Lai, R. Lu, and D. Zheng, "Seccdv: A security reference architecture for cybertwin-driven 6g v2x," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 5, pp. 4535–4550, 2022.

[19] M. A. Javed and E. B. Hamida, "On the interrelation of security, qos, and safety in cooperative its," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 7, pp. 1943–1957, July 2017.

[20] S. Soderi and R. De Nicola, "6g networks physical layer security using rgb visible light communications," *IEEE Access*, vol. 10, pp. 5482–5496, 2022.

[21] H. Guo, J. Li, J. Liu, N. Tian, and N. Kato, "A survey on space-air-ground-sea integrated network security in 6g," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 53–87, 2022.

[22] M. A. Javed, S. Zeadally, and Z. Hamid, "Trust-based security adaptation mechanism for vehicular sensor networks," *Computer Networks*, vol. 137, pp. 27 – 36, 2018.

[23] D. P. Moya Osorio, I. Ahmad, J. D. V. Sánchez, A. Gurtov, J. Scholliers, M. Kutila, and P. Porambage, "Towards 6g-enabled internet of vehicles: Security and privacy," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 82–105, 2022.

[24] M. Hussain, N. Ali, and J.-E. Hong, "Vision beyond the field-of-view: A collaborative perception system to improve safety of intelligent cyber-physical systems," *Sensors*, vol. 22, no. 17, 2022. [Online]. Available: https://www.mdpi.com/1424-8220/22/17/6610

[25] S. A. Soleymani, S. Goudarzi, M. H. Anisi, Z. Movahedi, A. Jindal, and N. Kama, "Pacman: Privacy-preserving authentication scheme for managing cybertwin-based 6g networking," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4902–4911, 2022.

[26] H. Cao, J. Du, H. Zhao, D. X. Luo, N. Kumar, L. Yang, and F. R. Yu, "Toward tailored resource allocation of slices in 6g networks with softwarization and virtualization," *IEEE Internet of Things Journal*, vol. 9, no. 9, pp. 6623–6637, 2022.

[27] M. A. Javed, E. B. Hamida, A. Al-Fuqaha, and B. Bhargava, "Adaptive security for intelligent transport system applications," *IEEE Intelligent Transportation Systems Magazine*, vol. 10, no. 2, pp. 110–120, 2018.

[28] P. Vijayakumar, M. Azees, S. A. Kozlov, and J. J. P. C. Rodrigues, "An anonymous batch authentication and key exchange protocols for 6g enabled vanets," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 2, pp. 1630–1638, 2022.

[29] T.-V. Le, C.-F. Lu, C.-L. Hsu, T. K. Do, Y.-F. Chou, and W.-C. Wei, "A novel three-factor authentication protocol for multiple service providers in 6g-aided intelligent healthcare systems," *IEEE Access*, vol. 10, pp. 28 975–28 990, 2022.

[30] N. Fotiou, V. A. Siris, G. Xylomenos, and G. C. Polyzos, "Iot group membership management using decentralized identifiers and verifiable credentials," *Future Internet*, vol. 14, no. 6, 2022. [Online]. Available: https://www.mdpi.com/1999-5903/14/6/173

[31] D. Marabissi, L. Mucchi, and A. Stomaci, "Iot nodes authentication and id spoofing detection based on joint use of physical layer security and machine learning," *Future Internet*, vol. 14, no. 2, 2022. [Online]. Available: https://www.mdpi.com/1999-5903/14/2/61

[32] D. An, F. Zhang, Q. Yang, and C. Zhang, "Data integrity attack in dynamic state estimation of smart grid: Attack model and countermeasures," *IEEE Transactions on Automation Science and Engineering*, pp. 1–14, 2022.

[33] H. Yu, Q. Hu, Z. Yang, and H. Liu, "Efficient continuous big data integrity checking for decentralized storage," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 1658–1673, 2021.

[34] H. Wu, B. Zhou, and C. Zhang, "Secure distributed estimation against data integrity attacks in internet-of-things systems," *IEEE Transactions on Automation Science and Engineering*, pp. 1–14, 2021.

[35] Y. Zhao, X. Gong, F. Lin, and X. Chen, "Data poisoning attacks and defenses in dynamic crowdsourcing with online data quality learning," *IEEE Transactions on Mobile Computing*, pp. 1–1, 2021.

[36] T. Hong and A. Hofmann, "Data integrity attacks against outage management systems," *IEEE Transactions on Engineering Management*, vol. 69, no. 3, pp. 765–772, 2022.

[37] Y. Luo, L. Cheng, Y. Liang, J. Fu, and G. Peng, "Deepnoise: Learning sensor and process noise to detect data integrity attacks in cps," *China Communications*, vol. 18, no. 9, pp. 192–209, 2021.

[38] X. Zhong, C. Fan, and S. Zhou, "Eavesdropping area for evaluating the security of wireless communications," *China Communications*, vol. 19, no. 3, pp. 145–157, 2022.

[39] ——, "Eavesdropping area for evaluating the security of wireless communications," *China Communications*, vol. 19, no. 3, pp. 145–157, 2022.

[40] B. Li, Y. Yao, H. Zhang, Y. Lv, and W. Zhao, "Energy efficiency of proactive eavesdropping for multiple links wireless system," *IEEE Access*, vol. 6, pp. 26 081–26 090, 2018.

[41] B. Li, Y. Yao, H. Zhang, and Y. Lv, "Energy efficiency of proactive cooperative eavesdropping over multiple suspicious communication links," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 1, pp. 420–430, 2019.

[42] G. Savva, K. Manousakis, and G. Ellinas, "Eavesdropping-aware routing and spectrum/code allocation in ofdm-based eons using spread spectrum techniques," *Journal of Optical Communications and Networking*, vol. 11, no. 7, pp. 409–421, 2019.

[43] J. Moon, S. H. Lee, H. Lee, and I. Lee, "Proactive eavesdropping with jamming and eavesdropping mode selection," *IEEE Transactions on Wireless Communications*, vol. 18, no. 7, pp. 3726–3738, 2019.

[44] H. Pirayesh and H. Zeng, "Jamming attacks and anti-jamming strategies in wireless networks: A comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 767–809, 2022.

[45] S. Hu, D. Yue, C. Dou, X. Xie, Y. Ma, and L. Ding, "Attack-resilient event-triggered fuzzy interval type-2 filter design for networked nonlinear systems under sporadic denial-of-service jamming attacks," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 1, pp. 190–204, 2022.

[46] J. Liu, X. Wang, S. Shen, Z. Fang, S. Yu, G. Yue, and M. Li, "Intelligent jamming defense using dnn stackelberg game in sensor edge cloud," *IEEE Internet of Things Journal*, vol. 9, no. 6, pp. 4356–4370, 2022.

[47] L. Zhao, H. Xu, J. Zhang, and H. Yang, "Resilient control for wireless cyber-physical systems subject to jamming attacks: A cross-layer dynamic game approach," *IEEE Transactions on Cybernetics*, vol. 52, no. 4, pp. 2599–2608, 2022.

[48] A. Cetinkaya, H. Ishii, and T. Hayakawa, "Effects of jamming attacks on wireless networked control systems under disturbance," *IEEE Transactions on Automatic Control*, pp. 1–1, 2022.

[49] J. Villain, V. Deniau, C. Gransart, A. Fleury, and E. P. Simon, "Characterization of ieee 802.11 communications and detection of low-power jamming attacks in noncontrolled environment based on a clustering study," *IEEE Systems Journal*, vol. 16, no. 1, pp. 683–692, 2022.

[50] C. Zhang, W. Zuo, P. Yang, Y. Li, and X. Wang, "Outsourced privacy-preserving anomaly detection in time series of multi-party," *China Communications*, vol. 19, no. 2, pp. 201–213, 2022.

[51] C. Huang, Z. Yang, J. Wen, Y. Xu, Q. Jiang, J. Yang, and Y. Wang, "Self-supervision-augmented deep autoencoder for unsupervised visual anomaly detection," *IEEE Transactions on Cybernetics*, pp. 1–14, 2021.

[52] C. Huang, J. Wen, Y. Xu, Q. Jiang, J. Yang, Y. Wang, and D. Zhang, "Self-supervised attentive generative adversarial networks for video anomaly detection," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2022.

[53] W. A. Yousef, I. Traoré, and W. Briguglio, "Un-avoids: Unsupervised and nonparametric approach for visualizing outliers and invariant detection scoring," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 5195–5210, 2021.

[54] X. Ma, J. Wu, S. Xue, J. Yang, C. Zhou, Q. Z. Sheng, H. Xiong, and L. Akoglu, "A comprehensive survey on graph anomaly detection with deep learning," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2021.

[55] W. Wang, Z. Wang, Z. Zhou, H. Deng, W. Zhao, C. Wang, and Y. Guo, "Anomaly detection of industrial control systems based on transfer learning," *Tsinghua Science and Technology*, vol. 26, no. 6, pp. 821–832, 2021.

[56] M. A. Javed, M. Z. Khan, U. Zafar, M. F. Siddiqui, R. Badar, B. M. Lee, and F. Ahmad, "Odpv: An efficient protocol to mitigate data integrity attacks in intelligent transport systems," *IEEE Access*, vol. 8, pp. 114 733–114 740, 2020.

[57] W. Wang, W. Song, Z. Li, B. Zhao, and B. Zhao, "A novel filter-based anomaly detection framework for hyperspectral imagery," *IEEE Access*, vol. 9, pp. 124 033–124 043, 2021.

[58] Y.-H. Nho, S. Ryu, and D.-S. Kwon, "Ui-gan: Generative adversarial network-based anomaly detection using user initial information for wearable devices," *IEEE Sensors Journal*, vol. 21, no. 8, pp. 9949–9958, 2021.

[59] X. Ma and W. Shi, "Aesmote: Adversarial reinforcement learning with smote for anomaly detection," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 943–956, 2021.

[60] X. Wang, S. Garg, H. Lin, J. Hu, G. Kaddoum, M. Jalil Piran, and M. S. Hossain, "Toward accurate anomaly detection in industrial internet of things using hierarchical federated learning," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7110–7119, 2022.

[61] T. V. Phan, T. G. Nguyen, N.-N. Dao, T. T. Huong, N. H. Thanh, and T. Bauschert, "Deepguard: Efficient anomaly detection in sdn with fine-grained traffic flow monitoring," *IEEE Transactions on Network and Service Management*, vol. 17, no. 3, pp. 1349–1362, 2020.

[62] T. Zhao, T. Jiang, N. Shah, and M. Jiang, "A synergistic approach for graph anomaly detection with pattern mining and feature learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2393–2405, 2022.

[63] M. Tsikerdekis, S. Waldron, and A. Emanuelson, "Network anomaly detection using exponential random graph models and autoregressive moving average," *IEEE Access*, vol. 9, pp. 134 530–134 542, 2021.

[64] E. B. Hamida and M. A. Javed, "Channel-aware ECDSA signature verification of basic safety messages with k-means clustering in VANETs," in *Proc. IEEE Intl. Conf. on Advanced Information Networking and Applications*, 2016, pp. 1–8.

TLDViT: A Vision Transformer Model for Tomato Leaf Disease Classification

Sami Aziz Alshammari
Department of Information Technology-Faculty of Computing and Information Technology
Northern Border University, Saudia Arabia

***Abstract*—Accurate and efficient diagnostic methods are essential for crop health monitoring due to the substantial impact of tomato leaf diseases on crop yield and quality. Traditional machine learning models, such as convolutional neural networks (CNNs), have shown promise in plant disease classification; however, they often require extensive data preprocessing and struggle with complex variations in leaf appearance. This study introduces TLDViT (Tomato Leaf Disease Vision Transformer), a Vision Transformer model specifically designed for the classification of tomato leaf diseases. TLDViT reduces the need for preprocessing by learning disease-specific features directly from raw images, leveraging Vision Transformers' ability to capture long-range dependencies within images. We evaluated TLDViT on the Plant Village Dataset, which includes healthy and diseased samples across multiple classes. For comparative analysis, two Vision Transformer models, ViT-r50-l32 and ViT-l16-fe, were tested. Among these, ViT-r50-l32 achieved the highest performance, surpassing both ViT-l16-fe with an accuracy of 98%. These findings highlight TLDViT's potential as an effective tool for crop health monitoring and automated plant disease diagnosis.**

***Keywords*—Tomato Leaf Disease; Vision Transformer (ViT); crop health monitoring; plant disease classification**

I. Introduction

Agriculture is fundamental to global food security, and enhancing crop health management is crucial for maintaining production and reducing economic losses. Tomato (Solanum lycopersicum) is among the most extensively farmed crops globally, however it is very vulnerable to several foliar diseases, such as early blight, late blight, and leaf mold. These illnesses, mostly induced by pathogens including fungus, bacteria, and viruses, result in substantial decreases in production and quality [1], [2]. Accurate early detection and classification of these illnesses is essential for facilitating prompt and focused therapies, which may help reduce future transmission and harm. Conventional techniques for diagnosing plant diseases depend significantly on manual visual assessment and expert expertise, which are labor-intensive, expensive, and susceptible to subjective inaccuracies [3]. Advances in artificial intelligence (AI) and machine learning (ML) have shown potential to overcome these constraints through the automation of disease diagnosis. Convolutional neural networks (CNNs), a prevalent deep learning methodology, have shown efficacy in recognizing intricate patterns in plant disease imagery, attaining high accuracy in classification tests across several crop illnesses [4]. The authors of [3] used CNNs on an extensive dataset of plant diseases, achieving classification accuracies of 90% across 26 distinct crops. In [4], authors introduced advanced CNN architectures to improve the categorization of plant diseases, particularly those impacting tomatoes, but with considerable preprocessing and computing demands. These studies highlight the promise of CNNs while also exposing significant obstacles, including their reliance on large labeled datasets, susceptibility to overfitting, and constraints in capturing non-local connections in images [5]. Researchers have investigated different models and strategies to enhance the resilience and efficiency of plant disease classification systems, addressing these constraints. The author in [6] integrated handmade features with deep learning models, enhancing the robustness of CNNs against variability in image data, while [7] supplemented restricted datasets with synthetic images to elevate CNN performance. Notwithstanding these efforts, CNN-based models exhibit constraints in their ability to apprehend global spatial linkages within images, a factor that is especially critical in plant disease categorization, where symptoms may appear in non-contiguous areas on the leaf.

In recent years, Vision Transformers (ViT) have surfaced as a formidable alternative to CNNs for identification of images tasks [8], [9], [10], [11]. In contrast to CNNs, which depend on local convolutional filters for hierarchical feature extraction, ViT use self-attention processes to capture long-range relationships over the whole image. The global attention mechanism enables ViT to comprehend spatial connections from a comprehensive viewpoint, making them especially adept at image processing tasks that need acute sensitivity to spatial intricacies. The author in [12] shows that ViT may get superior performance on extensive image classification datasets, surpassing CNNs in both precision and efficiency. In [13], the author emphasized the promise of ViT in applications necessitating intricate spatial analysis, including medical imaging and remote sensing. The author in [14] used Vision Transformers for agricultural disease detection, proving their efficacy in identifying disease patterns in crops such as rice and wheat; nevertheless, research on their application to tomato leaf diseases is still scarce.

This paper presents TLDViT (Tomato Leaf Disease Vision Transformer), a Vision Transformer model particularly developed for the classification of tomato leaf diseases, motivated by recent breakthroughs. Our methodology utilizes the ViT architecture's capacity to capture long-range relationships, allowing it to identify nuanced and intricate disease patterns that CNNs may overlook. TLDViT, in contrast to CNN-based methods that need considerable preprocessing and data augmentation, is designed to immediately learn disease-specific features from minimally processed images, enhancing its adaptability and efficiency for practical agricultural applications.

Our study provides multiple contributions to the field of automated plant disease identification. The proposal introduces

TLDViT, a new Vision Transformer model tailored for the classification of tomato leaf diseases. Furthermore, two Vision Transformer models, ViT-r50-l32 and ViT-l16-fe, were employed to establish a comparative framework, ensuring that all models were trained on the Plant Village Dataset for consistency and robustness. A comprehensive comparison of model performances demonstrated that TLDViT exhibits superior accuracy compared to CNN-based methods and the two Vision Transformer models, underscoring its efficacy in this context. The study illustrates the benefits of Vision Transformers in agricultural diagnostics, emphasizing their sensitivity to spatial details, which is crucial for precise disease identification. These contributions enhance the application of Vision Transformers in plant disease detection and establish a basis for wider use in agricultural diagnostics.

The rest of the paper is structured as follows: Section II introduces the Literature Review, where we discuss related work on plant disease classification, highlighting the advantages and limitations of existing deep learning approaches, including Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs). The proposed methodology for classifying tomato leaf diseases is presented in Section III, which also covers the TLDViT model architecture, training procedures, and data preparation. The findings and discussion, together with performance comparisons and analysis, are presented in Section IV. Finally, Section V concludes the paper and outlines future research directions.

II. Literature Review

The latest developments in deep learning have markedly improved systems for the identification and categorization of plant diseases. Numerous research have investigated alternate methods to enhance the precision and efficacy of these systems. Initial approaches mostly depended on manually produced features and traditional machine learning methodologies [15], [16]. Convolutional neural networks (CNNs) exhibit exceptional performance in image-based illness classification; yet, their dependence on considerable preprocessing and difficulty in capturing global picture dependencies provide significant problems [17]. Hybrid models that combine CNNs with alternative architectures have shown enhanced robustness to fluctuations in picture quality [18], [19]. The advent of Vision Transformers (ViTs) has offered a persuasive alternative to CNNs for agricultural applications. Vision Transformers use self-attention processes to record long-range dependencies, allowing for the analysis of complex spatial patterns in pictures [20], [21]. Applications of Vision Transformers (ViTs) in the identification of plant diseases, including those affecting rice, wheat, and grapes, have shown enhanced efficacy relative to conventional Convolutional Neural Networks (CNNs) [22]. Recent research has used transfer learning with transformer architectures to address data scarcity challenges in agricultural datasets [23]. The integration of transformers with real-time systems and edge devices is becoming prevalent, with the objective of implementing disease detection models directly in agricultural fields for practical use [24]. Nevertheless, few research has concentrated especially on tomato leaf diseases, highlighting the need for a specialized Vision Transformer model to fill this void.

III. Proposed Approach-Based Tomato Leaf Disease Classification

This section describes our method to diagnosing tomato leaf illnesses using TLDViT (Tomato Leaf Disease Vision Transformer), a hybrid Vision Transformer model designed to capture both localized and global patterns in leaf photos. TLDViT combines ResNet-50's feature extraction powers with Vision Transformers' self-attention capabilities, resulting in a robust tool for detecting and recognizing disease signs in tomato leaves. We describe the major processes in our technique below, which include data preparation, model construction, training, and evaluation.

A. Data Preprocessing

Data preprocessing is a crucial phase to guarantee the quality and uniformity of the pictures used for training the TLDViT model. The dataset consists of images of tomato leaves, classified into six categories: Healthy, Bacterial Spot, Early Blight, Late Blight, Septoria Leaf Spot, and Yellow Leaf Curl Virus. The dataset used for this study is the publicly available Plant Village Dataset [25], which provides a comprehensive set of labeled images representing various plant diseases. This dataset is widely used for plant disease classification tasks and offers high-quality images that ensure accurate training and evaluation of the TLDViT model. All images are scaled to 224x224 pixels to standardize input dimensions, so minimizing computing effort while preserving enough information for precise categorization. Each pixel intensity is standardized to the interval [0, 1], enhancing the stability of the training process and facilitating more effective model learning. To improve the model's resilience and mitigate overfitting, many data augmentation methods are used, such as rotation, horizontal flipping, and brightness modifications. These changes create variances in the dataset, allowing TLDViT to generalize well across diverse lighting and ambient circumstances, which is essential for practical use.

Fig. 1 depicts the class distributions of tomato leaf disease images before to and after to data augmentation. Before augmentation (blue bars), the dataset comprised a total of 10,958 images unevenly allocated among six categories: Bacterial Spot (1,925 images), Early Blight (1,702 images), Healthy (1,920 images), Late Blight (1,705 images), Septoria Leaf Spot (1,745 images), and Yellow Leaf Curl Virus (1,961 images). Following augmentation (orange bars), the dataset dramatically increased to 13,603 images, enhancing class equilibrium. The post-augmentation dataset comprises 2,084 photos of Bacterial Spot, 2,352 images of Early Blight, 2,358 photographs of Healthy specimens, 2,267 images of Late Blight, 2,140 images of Septoria Leaf Spot, and 2,402 images of Yellow Leaf Curl Virus. This augmentation approach guarantees a more equitable dataset, which is essential for training machine learning models to generalize proficiently across all categories.

The used dataset is partitioned into three subsets: training (70%), validation (20%), and test (10%), allowing an equitable assessment of the model's efficacy on novel data. The training set is used to train the model, the validation set is applied for hyperparameter optimization and monitoring throughout training, and the test set offers a conclusive evaluation of the model's classification accuracy across six categories.

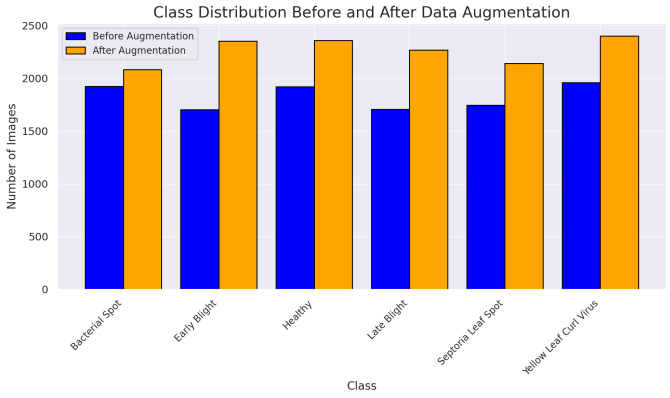


Fig. 1. Tomato class distribution.

B. TLDViT Model (Tomato Leaf Disease Vision Transformer)

The Vision Transformer (ViT) architecture, as presented in Fig. 2, has been customized to classify tomato leaf maladies into six categories: Healthy, Bacterial Spot, Early Blight, Late Blight, Septoria Leaf Spot, and Yellow Leaf Curl Virus. The model is highly effective in capturing complex disease patterns by combining local feature extraction with global context modeling.

1) Image patch division and flattening: The first phase is partitioning each input picture of a tomato leaf into a grid of smaller segments. A 224 x 224-pixel image may be divided into 32 x 32 pixel patches, resulting in 49 patches arranged in a 7 x 7 grid. Subsequently, each patch is transformed into a one-dimensional vector. This patch-based method collects intricate local details inside each leaf segment, enabling the model to identify localized disease indicators such as spots, discolorations, and texture alterations unique to certain illnesses.

2) Linear projection of flattened patches: The flattened patches undergo a linear projection layer, converting each patch into a high-dimensional vector appropriate for further processing in the Vision Transformer. This transformation produces a series of patch embeddings that preserve the localized features of each patch while mapping them into a higher-dimensional space, allowing the model to interpret the picture as a sequence instead of a grid.

3) Positional embedding and class token: Positional embeddings are included into each patch embedding to maintain the spatial configuration of patches inside the original image. The positional embeddings provide the model with data on the location of each patch inside the image, which is essential for comprehending spatial links among illness symptoms. Furthermore, an additional learnable class token is attached to the series of patch embeddings. The class token engages with the patch embeddings throughout the transformer layers and ultimately retains the information required to generate the final classification label.

4) Transformer encoder: The fundamental component of the design is the Transformer Encoder, including numerous layers that integrate self-attention mechanisms with feed-forward neural networks. Each encoder layer has many essential components: The Multi-Head Self-Attention mechanism, which allows the model to concentrate on several sections of the leaf concurrently, therefore capturing both local intricacies and overarching patterns within the picture. This multi-head attention enables the model to discern intricate relationships across patches, facilitating the identification of disease-related patterns that may be distributed over different areas of the leaf. Normalization (Norm) layers are used to stabilize the learning process and mitigate overfitting by guaranteeing that inputs to each layer possess a standardized distribution, facilitating model convergence and enhancing generalization. Subsequent to the self-attention mechanism, a Multi-Layer Perceptron (MLP) introduces non-linear changes to the representations, therefore augmenting the model's capacity to discern intricate patterns pertinent to each illness type and boosting classification precision. This encoder architecture enables the model to analyze the whole image as a series of patches, use self-attention to discern correlations both internally and externally among the patches. This is especially beneficial in the categorization of plant diseases, where symptoms may manifest in scattered patterns or as nuanced textural alterations on the leaf.

5) MLP head and classifier: The class token's embedding is supplied into a Multi-Layer Perceptron (MLP) head and subsequently into a classifier, which generates the final prediction, after passing through the transformer layers. The classifier attributes the image to one of the six classes, thereby designating the leaf as healthy or indicating the specific type of disease present. The MLP head is the ultimate stage in the processing process, incorporating all the information acquired by the transformer layers to provide a precise diagnosis.

This ViT architecture is particularly effective for the classification of tomato leaf diseases because it can manage both local and global image features. The self-attention mechanism enables the model to interpret relationships across the entire image, while the patch-based approach captures detailed visual features within small sections, which is essential for identifying disease-specific symptoms. The distinction between diseases that may appear visually similar but have unique patterns or spread across various areas of the leaf is dependent on the combination of local and global context.

C. Training Methodology

In order to optimize the performance of TLDViT, we implement a systematic training approach that incorporates exhaustive evaluation techniques, optimization, and regularization. In order to reduce prediction errors across all disease classes, categorical cross-entropy loss is implemented, as this is a multi-class classification problem. The Adam optimizer is employed with an initial learning rate of 0.0001, which is progressively reduced by a learning rate scheduler as training progresses. This approach assists in the stabilization of the model and the enhancement of convergence, while also preventing overshooting. The high model complexity necessitates the use of regularization methods, such as dropout layers inside the transformer component, to prevent overfitting. In order to avoid superfluous epochs and overfitting, early stopping is sometimes used. This involves monitoring validation accuracy and ending training as performance stabilizes. To achieve a happy medium between computing efficiency and enough iteration for learning, the model is trained with a batch size of 64 for 25 epochs.

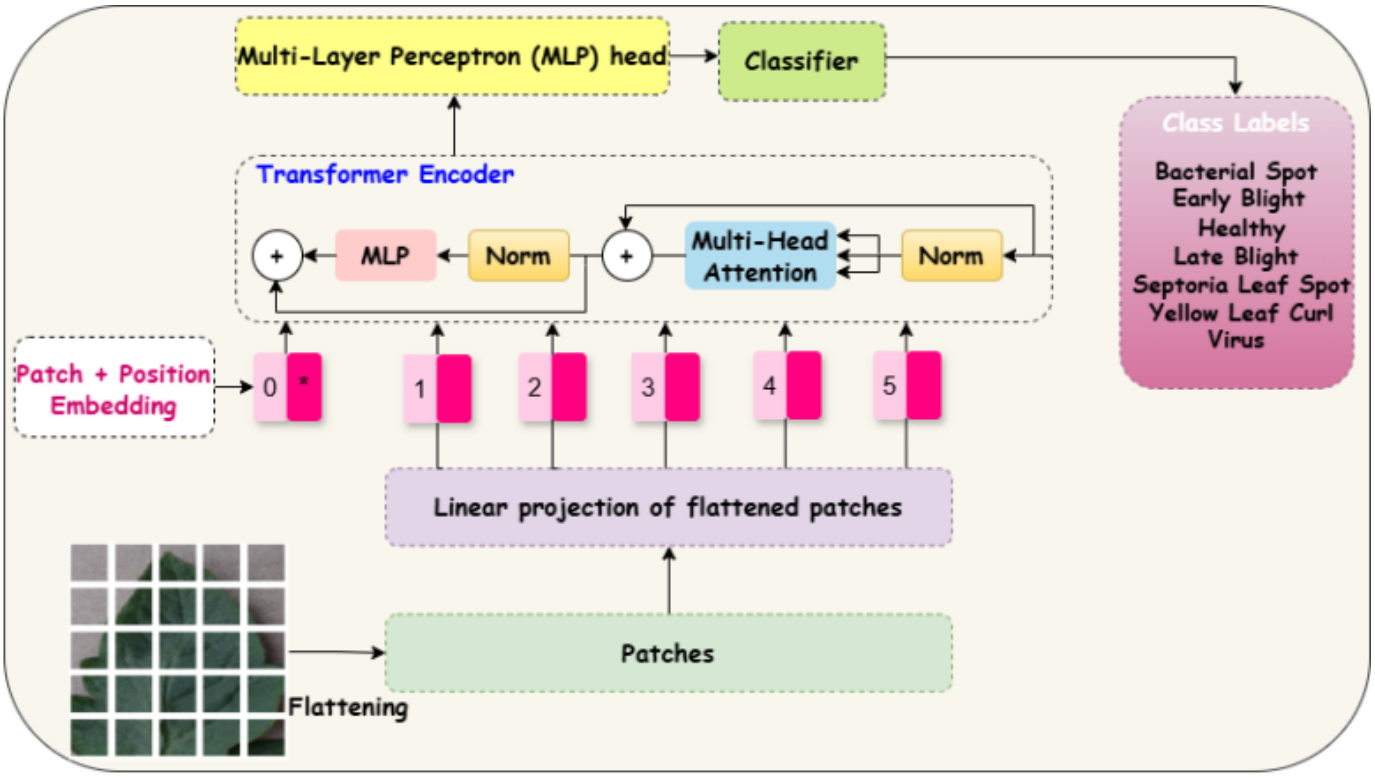


Fig. 2. TLDViT architecture.

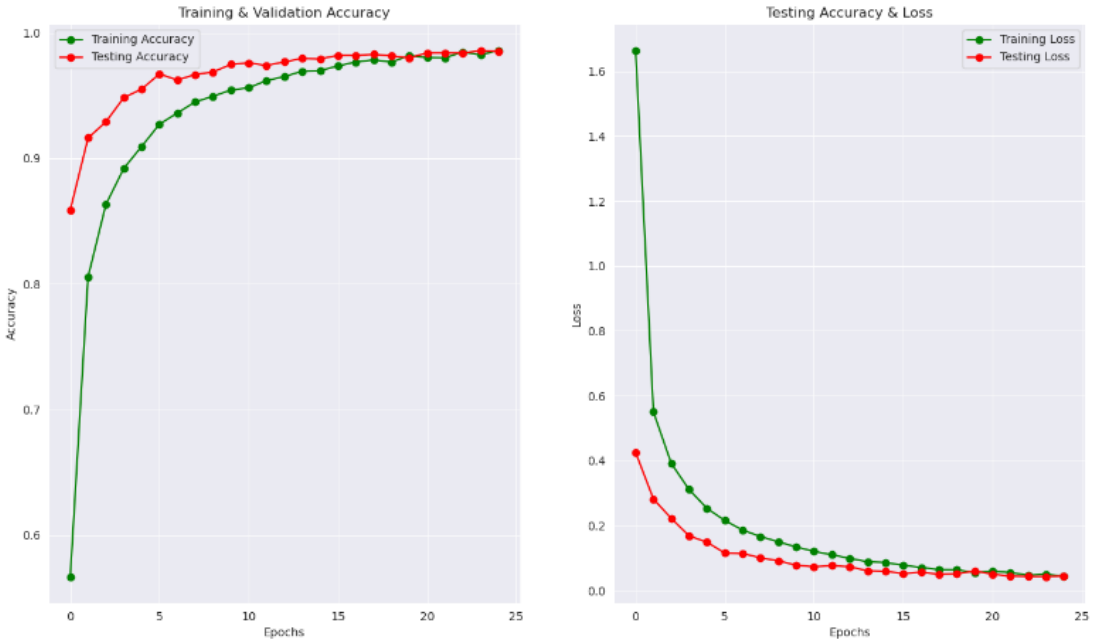


Fig. 3. Accuracy and loss curve of ViT-r50-l32 model.

We evaluate TLDViT's efficacy in accurately categorizing tomato leaf diseases using the following metrics: accuracy, F1-score, precision, recall, and confusion matrix. Making ensuring the model satisfies all criteria, this assessment checks it thoroughly. Additionally, we use ROC curves as a measure to evaluate the model's performance across multiple thresholds, especially when distinguishing across interrelated illness types. To find out how well the model can distinguish between classes at various decision thresholds, we build ROC curves for each class and then measure the area under the curve (AUC). The following equations introduced the performance evaluations:

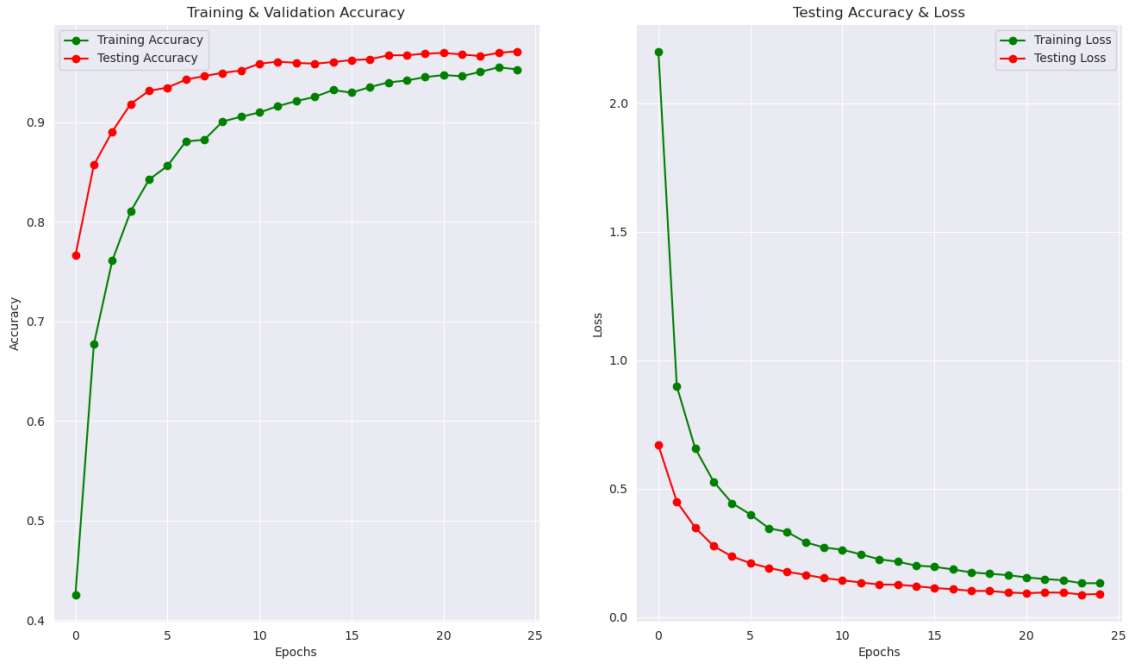


Fig. 4. Accuracy and loss curve of ViT-l16-fe model.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

$$F_1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

- True Positives (TP): Correctly predicted positive samples.
- True Negatives (TN): Correctly predicted negative samples.
- False Positives (FP): Incorrectly predicted positive samples.
- False Negatives (FN): Incorrectly predicted negative samples.

IV. Experimental Results

The experimental findings of the proposed TLDViT model for categorizing tomato leaf diseases into six labels are shown in this section. These categories are Healthy, Bacterial Spot, Early Blight, Late Blight, Septoria Leaf Spot, and Yellow Leaf Curl Virus. For the purpose of conducting a complete evaluation, we evaluate the performance of the model using a number of different measures, such as accuracy, precision, recall, F1-score, and area under the ROC curve (AUC).

A. Classification Accuracy and Loss

Fig. 3 shows that training and testing accuracy increase with epochs in the ViT-r50-l32 model, stabilizing at high values. Within a few epochs, the model converges, with training accuracy around 99% and testing accuracy close behind. Regularly decreasing loss curves for training and testing indicate good learning without overfitting. The tight alignment of training and testing performance implies that ViT-r50-l32 can generalize to new data and discriminate tomato leaf disease classes. Fig. 4 shows that ViT-l16-fe has slower convergence and larger loss values during training, suggesting a weaker generalization capacity than ViT-r50-l32. Although ViT-l16-fe is accurate, it lacks the stability and minimum loss of ViT-r50-l32. These findings show that ViT-l16-fe is effective but may not capture disease-specific aspects as well as ViT-r50-l32.

The results of this study reveal that ViT-r50-l32 outperforms ViT-l16-fe in accuracy and loss measures, evidenced by its fast convergence, elevated final accuracy, and reduced loss levels throughout training and testing. The exceptional efficacy of ViT-r50-l32 indicates that its integration of a ResNet-50 backbone with Vision Transformer layers is especially adept at collecting complex illness characteristics, enabling more precise differentiation across disease categories. The constant and consistent performance shown in both training and testing reinforces the resilience of ViT-r50-l32.

B. Performance Comparison of Tomato Leaf Disease

Table I shows the classification performance of two models, ViT-r50-l32 and ViT-l16-fe, on tomato leaf disease categories, including Precision, Recall, F1-Score, and Overall Accuracy. F1-Scores of 0.95 or better are achieved by the ViT-r50-l32 model in all categories. It has excellent precision and recall for "Late Blight" and "Septoria Leaf Spot," an F1-Score of 1.00 for both, and an accuracy of 0.98, showing good generalization. Though scoring lower in several areas, the ViT-l16-fe model performs well. For "Late Blight" it has an F1-Score of 0.94 owing to a minor loss in accuracy, but it has good precision and recall across most classes, especially for "Yellow Leaf Curl Virus" with 1.00 precision. Though somewhat lower than ViT-r50-l32, ViT-l16-fe has solid classification performance with an accuracy of 0.97. In conclusion, both models have good accuracy and F1-Scores across all categories, although ViT-r50-l32 may be preferable for tomato leaf disease classification in this dataset.

TABLE I. PERFORMANCE CLASSIFICATION OF TLDVIT MODELS

Model	Class	Precision	Recall	F1-Score
ViT-r50-l32	Bacterial Spot	0.93	0.98	0.95
	Early Blight	1.00	0.96	0.98
	Late Blight	1.00	1.00	1.00
	Septoria Leaf Spot	1.00	1.00	1.00
	Yellow Leaf Curl Virus	0.96	1.00	0.98
	Healthy	1.00	0.96	0.98
	Overall Accuracy		0.98	
ViT-l16-fe	Healthy	0.96	1.00	0.98
	Bacterial Spot	0.98	0.92	0.95
	Early Blight	0.98	1.00	0.99
	Late Blight	0.90	0.98	0.94
	Septoria Leaf Spot	1.00	0.95	0.98
	Yellow Leaf Curl Virus	1.00	0.96	0.98
	Overall Accuracy		0.97	

Fig. 5 displays the Precision-Recall (PR) curve for the ViT-r50-l32 model, which classified tomato leaf diseases well. Each curve symbolizes a disease: Healthy, Bacterial Spot, Early Blight, Late Blight, Septoria Leaf Spot, and Yellow Leaf Curl Virus. The legend shows each category's average accuracy score. Due to its near-perfect accuracy and recall across all classes, the model can reliably categorize each illness type without substantial false positives or negatives. All classes' aggregate performance is tinted blue, with an average accuracy of 0.997. The model excels on "Early Blight" and "Yellow Leaf Curl Virus," scoring 1.000 in precision-recall, while the remaining classes score 0.995–0.999. This curve shows the ViT-r50-l32 model's resilience and accuracy, making it ideal for identifying and differentiating tomato leaf diseases.

Fig. 6 illustrates the Multi-Class Receiver Operating Characteristic (ROC) curve for the ViT-r50-l32 model, which is the most effective model for classifying tomato leaf diseases. Each curve denotes one of the six disease categories: Healthy, Bacterial Spot, Early Blight, Late Blight, Septoria Leaf Spot, and Yellow Leaf Curl Virus. The values presented in the legend represent the area under the curve (AUC) for each category. The model attains an AUC score of 1.00 for each class and for the micro-average ROC curve, demonstrating optimal performance in differentiating among the various disease categories. The ROC curve indicates that the model can achieve a true positive rate (sensitivity) of 1.0 while keeping The false positive rate near 0 across all categories. The observed accuracy indicates that ViT-r50-l32 is a reliable

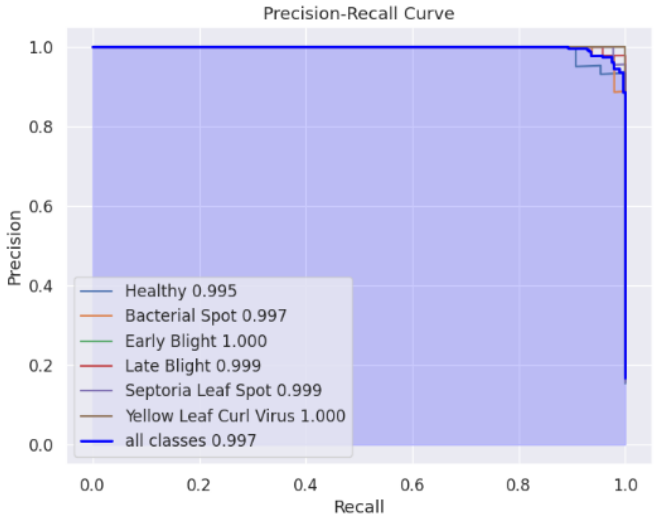


Fig. 5. Precision-Recall (PR) curve of ViT-r50-l32 model.

model for classifying tomato leaf diseases, as evidenced by the absence of misclassifications in the ROC metrics. The diagonal dashed line indicates a random classifier (AUC = 0.5), while the model's ROC curves positioned significantly above this line demonstrate its robust predictive performance. The optimal AUC scores demonstrate the model's high accuracy and robustness, positioning it as an effective tool for the detection.

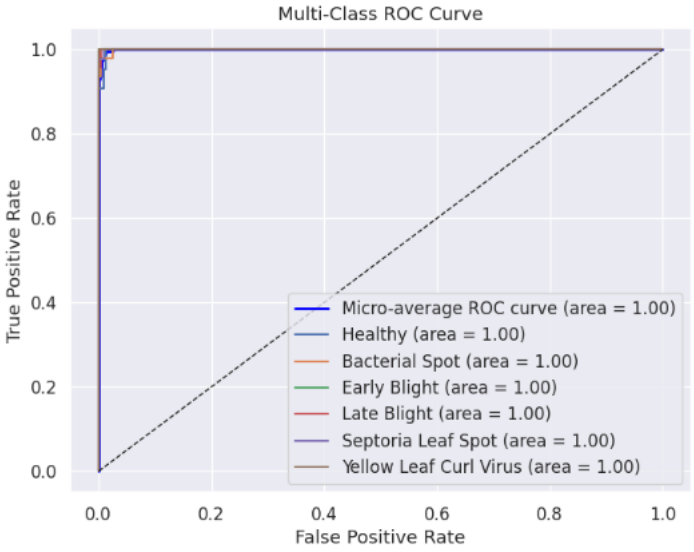


Fig. 6. ROC curve of ViT-r50-l32 model.

The confusion matrix for the ViT-r50-l32 tomato leaf disease classification model shows high true positive counts for each disease category, as depicted in Fig. 7. Bacterial Spot has 42 accurate categories and a few Healthy misclassifications. Early Blight is properly categorized 46 times, mostly in Late Blight. One Early Blight occurrence was misclassified, whereas Late Blight had 49 proper classifications. With 47 and 44 accurate classifications and no substantial misclassifications, Septoria Leaf Spot and Leaf Curl Virus perform well.

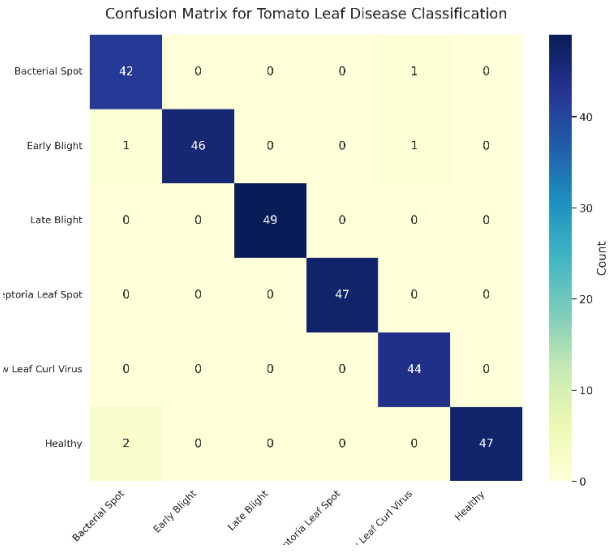


Fig. 7. Confusion matrix of ViT-r50-l32 model for differentiation of tomato leaf diseases.

Healthy is accurately labeled 47 times, with Bacterial Spot misclassified. Some classifications, such Healthy and Bacterial Spot, are somewhat confusing, suggesting model refining or hyperparameter tweaks might improve classification accuracy.

In Fig. 8, the ViT-r50-l32 model classifies six tomato leaf types from the Plant Village Dataset, including healthy and sick samples, using test images. The model accurately distinguishes Bacterial Spot, which has dark, irregular spots; Early Blight, which has concentric rings on yellowed areas; Healthy leaves, which are uniformly green and symptom-free; Late Blight, which has large, darkened lesions; Septoria Leaf Spot, which has small, circular lesions with light centers and dark edges; and Yellow Leaf Curl Virus, which shows curled, yellowed edges The model can distinguish these groups, suggesting its potential for early illness identification and treatment.

Fig. 8. Classification results of ViT-r50-l32 model.

C. Comparative Study

Table II presents a comparative analysis of tomato leaf disease classification models, emphasizing the performance metrics of various methodologies. The authors [26] used a CNN-based model (Inception-V3 and DenseNet-121), attaining an accuracy of 95.08%, with precision, recall, and F1-score metrics closely matched at 95.10%, 95.05%, and 95.07%, respectively. In addition, the authors in [27] introduced the TomFormer model, which amalgamates transformer-based architectures, achieving an accuracy of 87%, with somewhat reduced precision (87.50%), recall (86.50%), and F1-score (87.00%) relative to CNN-based models. Our proposed model, ViT-r50-l32, utilizes Vision Transformers to attain exceptional performance, achieving an accuracy of 98%, precision of 98.30%, recall of 98.33%, and an F1-score of 98.20%, thereby illustrating its efficacy and resilience in tomato leaf disease classification tasks.

TABLE II. Comparative Study with Related Approaches

Study	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
[26]	95.08	95.10	95.05	95.07
[27]	87	87.50	86.50	87.00
TLDViT Model	98	98.30	98.33	98.20

V. Conclusion

This paper presents TLDViT, a Vision Transformer model explicitly developed for classifying tomato leaf diseases using images from the Plant Village Dataset. TLDViT exhibited effective classification capabilities across six categories: Bacterial Spot, Early Blight, Healthy, Late Blight, Septoria Leaf Spot, and Yellow Leaf Curl Virus. In our assessments, we used two Vision Transformer models ViT-r50-l32 and ViT-l16-fe for comparative analysis. Among them, ViT-r50-l32 surpassed the other model, demonstrating enhanced accuracy and resilience across the illness categories. These findings underscore TLDViT's potential, in conjunction with ViT-r50-l32, for facilitating the early detection and control of crop diseases, which is essential for sustainable agriculture and food security. We propose the development of a mobile or field-deployable application for real-time disease diagnostics, facilitating the rapid identification of tomato leaf diseases by farmers and agronomists on-site, hence enabling prompt intervention and management.

Future work will optimize TLDViT for mobile and edge devices for real-time crop health monitoring in the field. We also want to combine this model into a mobile or field-deployable application for real-time disease diagnostics to help farmers and agronomists quickly identify tomato leaf diseases and control them. Other efforts include domain adaptation to improve model performance in varied environmental settings and adding new plant species and disease categories to the dataset.

Acknowledgment

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through project no. NBU-FFR-2024-2119-04

References

[1] J. L. Bargul and N. Ghanbari, "Detection of leaf diseases in tomato using machine learning approaches: A review," *International Journal of Plant Pathology*, vol. 12, no. 3, pp. 150–160, Sep. 2020.

[2] N. Ghanbari and A. R. Smith, "An analysis of disease patterns in tomato leaves using advanced imaging techniques," *Plant Disease Analysis*, vol. 45, no. 2, pp. 75–85, Feb. 2021.

[3] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, p. 1419, Sep. 2016.

[4] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, Jan. 2018.

[5] Y. Li, X. Ma, Y. Qiao, and J. Shang, "Plant disease detection based on convolutional neural network," *Cluster Computing*, vol. 22, no. 2, pp. 2593–2602, Jun. 2019.

[6] A. D. S. Ferreira, D. M. Freitas, G. G. da Silva, H. Pistori, and M. T. Folhes, "Weed detection in soybean crops using convnets," *Computers and Electronics in Agriculture*, vol. 143, pp. 314–324, Oct. 2017.

[7] J. G. A. Barbedo, "Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease recognition," *Computers and Electronics in Agriculture*, vol. 153, pp. 46–53, Aug. 2018.

[8] H. Kim and J. Lee, "Vit-smartagri: Vision transformer and smartphone-based plant disease detection for smart agriculture," *Agronomy*, vol. 14, no. 2, p. 327, Feb. 2024.

[9] M. Ali, R. Khan, and D. Patel, "A multitask learning-based vision transformer for plant disease localization and classification," *International Journal of Machine Learning and Cybernetics*, vol. 15, no. 3, pp. 987–1001, Mar. 2024.

[10] R. Gupta, L. Singh, and P. Choudhury, "Plant disease detection using vision transformers on multispectral natural environment images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, p. 102205, Jan. 2024.

[11] K. Mehta and F. Alzahrani, "Early betel leaf disease detection using vision transformer and deep learning algorithms," *Journal of Ambient Intelligence and Humanized Computing*, vol. 15, no. 1, pp. 115–126, Feb. 2024.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, and e. a. T. Unterthiner, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, Oct. 2021.

[13] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-end object detection with transformers," in *European Conference on Computer Vision (ECCV)*, Aug. 2020, pp. 213–229.

[14] J. Chen, D. Liu, and Y. Zhang, "Application of vision transformers in agricultural disease detection: Case studies on rice and wheat," *Agricultural Informatics Journal*, vol. 7, no. 4, pp. 234–244, Apr. 2022.

[15] J. Ma, Z. Zhou, Y. Wu, and X. Zheng, "Deep convolutional neural networks for automatic detection of agricultural pests and diseases," *Computers and Electronics in Agriculture*, vol. 151, pp. 83–90, 2018.

[16] A. Singh, B. Ganapathysubramanian, A. Singh, and S. Sarkar, "Machine learning for high-throughput stress phenotyping in plants," *Trends in plant science*, vol. 23, no. 10, pp. 883–898, 2018.

[17] A. Fuentes, S. Yoon, and S. Kim, "Automated crop disease detection using deep learning: A review," *Computers and Electronics in Agriculture*, vol. 142, pp. 361–370, 2017.

[18] A. Mishra, S. Hossain, and A. Sadeghian, "Image processing techniques for detection of leaf disease," *Journal of Agricultural Research*, vol. 11, pp. 134–145, 2017.

[19] A. Rangarajan, R. Purushothaman, and A. Ramesh, "Diagnosis of plant leaf diseases using cnn-based features," *Journal of Image Processing*, vol. 32, pp. 123–135, 2018.

[20] M. Khan, S. Amin, and M. Bilal, "Transformers in computer vision: A survey for plant disease recognition," *Computer Vision Research*, vol. 15, pp. 231–249, 2022.

[21] W. Liu, J. Zhang, and Q. Wang, "Transformer-based architectures for image classification in agricultural disease detection," *Information Processing in Agriculture*, vol. 9, no. 3, pp. 412–423, 2022.

[22] X. Zhang and Y. Huang, "Plant disease recognition based on vision transformers: A case study of grapevine leaf diseases," *IEEE Access*, vol. 10, pp. 24 256–24 267, 2022.

[23] M. Jiang and W. Li, "Plant disease detection using vision transformers with transfer learning," *Agricultural Informatics*, vol. 8, pp. 87–101, 2023.

[24] C. Feng and M. Wu, "Edge computing for real-time plant disease detection using lightweight transformer models," *Computers and Electronics in Agriculture*, vol. 210, p. 108330, 2023.

[25] D. P. Hughes and M. Salathé, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv preprint arXiv:1511.08060*, Nov. 2015.

[26] M. Yasin and N. Fatima, "Comparative performance evaluation of cnn models for tomato leaf disease classification," *arXiv preprint arXiv:2312.08659*, 2023.

[27] A. Khan and S. Ahmad, "Tomformer: A fusion model for early and accurate detection of tomato leaf diseases using transformers and cnns," *arXiv preprint arXiv:2312.16331*, 2023.

Hybrid Approach of Classification of Monkeypox Disease: Integrating Transfer Learning with ViT and Explainable AI

MD Abu Bakar Siddick[1], Zhang Yan*[2], Mohammad Tarek Aziz[3], Md Mokshedur Rahman[4], Tanjim Mahmud[5], Sha Md Farid[6], Valisher Sapayev Odilbek Uglu[7], Matchanova Barno Irkinovna[8], Atayev Shokir Kuranbaevich[9] Ulugbek Hajiev[10]

Department of Computer Science and Technology, Beijing Institute of Technology (BIT), Beijing, China[1,2,4]
Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Bangladesh[3]
Department of Computer Science and Engineering, Rangamati Science and Technology University, Bangladesh[5]
Department of Technology, Wilmington University, Delaware, United States[6]
Department of General-Professional Science, Mamun University, Khiva, Uzbekistan[7]
Urgench State Pedagogical Institute, Khorezm, Uzbekistan[8]
Urgench State University, Khorezm, Uzbekistan[9,10]

***Abstract*—Human monkeypox is a persistent global health challenge, ranking among the most common illnesses worldwide. Early and accurate diagnosis is critical to developing effective treatments. This study proposes a comprehensive approach to monkeypox diagnosis using deep learning algorithms, including Vision Transformer, MobileNetV2, EfficientNetV2, ResNet-50, and a hybrid model. The hybrid model combines ResNet-50, MobileNetV2, and EfficientNetV2 to reduce error rates and improve classification accuracy. The models were trained, validated, and tested on a specially curated monkeypox dataset. EfficientNetV2 demonstrated the highest training accuracy (99.94%), validation accuracy (97.80%), and testing accuracy (97.67%). ResNet-50 achieved 99.87% training accuracy, 99.85% validation accuracy, and 97.18% testing accuracy. MobileNetV2 reached 95.47% training accuracy, with validation and testing accuracies of 79.51% and 78.18%, respectively. Designed to mitigate overfitting, the Vision Transformer achieved 100% training accuracy, 87.51% validation accuracy, and 99.41% testing accuracy. Our hybrid model yielded 99.33% training accuracy and 99.09% testing accuracy. The Vision Transformer emerged as the most promising model due to its robust performance and high accuracy, followed closely by the hybrid model. Explainable AI (XAI) techniques, such as Grad-CAM, were applied to enhance the interpretability of predictions, providing visual insights into the classification process. The results underscore the potential of Vision Transformer and hybrid deep learning models for accurate and interpretable monkeypox diagnosis.**

***Keywords*—Monkeypox; vision transformer; hybrid model; transfer learning; explainable artificial intelligence**

I. Introduction

In order to manage outbreaks, it is essential to detect monkeypox accurately and promptly. Vision transformer models, transfer learning[1], [2], and deep learning [3], [4] provide effective ways to improve diagnostic accuracy from image data. Clinical examination and laboratory testing are frequently used in traditional diagnostic techniques, however, they can be time-consuming and less available in remote or underdeveloped places. Because models can be taught to accurately identify the distinctive characteristics of monkeypox lesions, deep learning [5], [6] makes monkeypox detection quicker and more scalable [7]. Transfer learning, which enables models to use pre-trained weights from sizable datasets, is particularly advantageous when it comes to monkeypox because there is a dearth of labeled data. Even with a limited monkeypox dataset [8], accurate models may be deployed thanks to this technique, which builds on previously obtained knowledge to enable faster training and higher accuracy. The detection method is further improved by the employment of vision transformers, which are renowned for their capacity to grasp intricate spatial relationships by processing images as patches. Vision transformers, as opposed to conventional convolutional networks [3], [9], [10], are able to identify global patterns in the image, which enables the model to concentrate on certain lesion features that could otherwise go unnoticed. This capacity is especially crucial for preventing misdiagnosis, improving patient outcomes, and distinguishing monkeypox from other skin disorders that look similar. When combined, these deep learning techniques [11] not only provide a more affordable and easily available diagnostic option, but they also aid in early detection and containment initiatives, which has a big influence on public health by enabling quick action in the event of an outbreak.

The main objective of image-based monkeypox disease detection is to facilitate early, precise, and easily accessible diagnosis, which is crucial for efficient outbreak management and patient treatment. Healthcare professionals can swiftly detect monkeypox lesions by using image-based AI models, which enables the early isolation and treatment of infected people to stop the spread of the disease [12], [13]. Another important goal is to provide diagnostic tools in settings with low resources and remote locations, where access to standard lab-based testing may be difficult [14], [15]. To avoid misdiagnosis and guarantee that patients receive the right care, models that can reliably differentiate monkeypox from other comparable skin disorders, including chickenpox or measles,

must be developed. Since there is a dearth of picture data related to monkeypox, it is crucial to use transfer learning [16] to build trustworthy models from small datasets. This will enable pre-trained models to identify distinctive characteristics of monkeypox. Additionally, by offering a scalable method for monitoring and forecasting outbreaks, AI-based monkeypox detection can help public health initiatives by facilitating prompt reaction and containment [17]. These goals support a strong and workable approach to controlling monkeypox, boosting readiness for infectious disease risks, and improving health outcomes.

The main contributions of this study are:

1) Improved model resilience, a bespoke dataset was created using random images that simulated a variety of real-world situations.
2) We enhanced feature extraction and detection by using ResNet-50's deep residual connections to categorize monkeypox lesions with accuracy.
3) MobileNetV2 was utilized for lightweight, effective detection, making diagnostics accessible on mobile or low-resource devices in remote locations.
4) EfficientNetV2 was used to maximize efficiency by striking a balance between decreased computation and detection accuracy for effective model operation.
5) We enhanced the model's capacity to distinguish monkeypox from related skin disorders by using Vision Transformer to gather global and patch-based picture information.
6) We created a hybrid model with enhanced diagnostic accuracy and dependability by combining ResNet-50, MobileNetV2, and EfficientNetV2.
7) We interpreted model predictions using explainable AI methodologies, offering visual justifications for clear, reliable diagnostics.

In this study, the previous research on existing work is described in Section II, and the proposed working framework is then detailed in Section III. The later Section IV denotes the result analysis, and finally, the conclusion and future plan are explained in the last Section V of this paper.

II. Previous Studies

In recent years, researchers have been trying to prevent the monkeypox disease and they are finding different types of solutions. Hence, some related publications are found online about the solutions to human monkeypox detection. We mentioned and explained some of them.

Bala et al. [18] proposed a deep CNN-based monkeypox disease detection system. Their study summary is that MonkeyNet, a novel deep learning-based model, was created to identify monkeypox from skin images. Its accuracy was 93.19% on the original dataset and 98.91% on an augmented dataset. In order to facilitate model training and testing, this study made the "Monkeypox Skin Images Dataset (MSID)" publicly available. Grad-CAM graphics help doctors diagnose monkeypox early and accurately by highlighting affected areas.

Dahiya et al. [19] explained monkeypox disease detection using a deep learning model. They used CNN and YOLO V5. Using the monkeypox dataset online, they obtain 98.18% accuracy in image classification.

Haque et al. [20] described to find out the monkeypox disease from the images with deep-transfer learning and attention mechanisms. They used online images and obtained a validation accuracy of 83.89%.

Sitaula et al. [21] proposed a method to find out the monkeypox virus detection with seven deep learning model as well as their ensemble method. They used publicly available data and applied seven pre-trained models initially. Later, to improve the performance they used an ensemble method of deep learning. The highest accuracy of their proposed work is 87.13%.

Ali et al. [8] proposed a method to detect human monkeypox from online collected image data. Their study was deep learning-based. Using online data, their classification rate is 82.96%.

Rahman et al. [22] explained federated and deep learning-based monkeypox disease detection from private limited data. Their study summary is-Accurate identification of monkeypox is difficult because it was deemed a global public health emergency after the COVID-19 epidemic. For efficient monkeypox classification, this paper suggests a safe, federated learning system that makes use of deep learning models such as MobileNetV2, Vision Transformer, and ResNet-50. With 97.90% accuracy using the ViT-B32 model, the method improves data security while guaranteeing accurate disease classification.

Azar et al. [23] proposed a deep neural network-based system to detect monkeypox from the images. Their study summary-this study created a deep learning model based on DenseNet201 to identify skin scans as either normal, chickenpox, monkeypox, or measles in response to the 2022 outbreak. The model performed exceptionally well, attaining 95.18% accuracy in a four-class scenario and 97.63% accuracy in a two-class scenario. To enhance model interpretability and help clinicians trust and comprehend the decision-making process, LIME and Grad-CAM were used. This model performs better than previous research, particularly in F1-Score, and provides information on the afflicted skin areas that are essential for diagnosis.

Altun et al. [7] proposed a method to detect monkeypox from sensor-based data with deep-transfer learning. Their work summary is that-in order to target possible pandemic scenarios, this study sought to create a deep learning-based monkeypox detection algorithm that is both quick and accurate. VGG19, DenseNet121, ResNet-50, EfficientNetV2, MobileNetV3, and Xception models were used to create a new CNN model with hyperparameter tuning and transfer learning. With an F1-score of 0.98, AUC of 0.99, accuracy of 0.96, and recall of 0.97, the optimized MobileNetV3 model exhibited the greatest performance, proving the usefulness of deep learning in quick disease classification.

Ahsan et al. [24] demonstrate a deep learning-based monkeypox disease detection from input data. This study evaluated six deep learning models, including Inception ResNetV2 and Mobile NetV 2, for early illness detection utilizing transfer learning in light of widespread worries about monkeypox as a possible pandemic danger. The altered models demonstrated

their diagnostic capabilities with accuracies ranging from 93% to 99%. By identifying important characteristics linked to the development of monkeypox, LIME was used to improve model transparency.

Uysal et al. [25] explained a hybrid deep learning method to identify monkeypox disease from image data. In order to identify monkeypox from skin photos in a multi-class dataset (monkeypox, chickenpox, measles, and normal), this study created a hybrid AI model by merging CSPDarkNet, InceptionV4, MnasNet, MobileNetV3, RepVGG, SE-ResNet, Xception, and LSTM. Following data augmentation, the hybrid model demonstrated strong performance in differentiating monkeypox from related diseases, with 87% test accuracy and a Cohen's kappa value of 0.8222.

Saleh et al. [26] proposed a new approach to detect monkeypox from image data. They used AI, Chimp algorithm etc. The two-phase AI-based Human Monkeypox Detection (HMD) strategy is presented in this article as a means of early monkeypox detection. Weighted Naïve Bayes, Weighted K-Nearest Neighbors, and a deep learning model through weighted voting are all combined in the Detection Phase (DP) to create an Ensemble Diagnosis (ED) model. The first phase, the Selection Phase (SP), uses an Improved Binary Chimp Optimization (IBCO) algorithm for optimal feature selection. With an accuracy of 98.48%, precision of 91.1%, and recall of 88.91%, HMD outperforms contemporary diagnostic techniques.

Almufareh et al. [27] explained how to detect monkeypox from two different datasets with transfer learning. As a safer substitute for conventional PCR testing, this work suggests a non-invasive, computer-vision-based approach for detecting monkeypox by employing deep learning to analyze skin lesion photos. The method's high sensitivity, specificity, and balanced accuracy, as established by tests on the MSLD and MSID datasets, make it an attractive option for general usage, particularly in places with inadequate lab infrastructure. IoT and AI are used in this method to provide safe, contactless diagnostics.

Table I is the summary of the mentioned work that was published in 2022, 2023, and 2024. After analyzing, we see that the previous work has limitations in some cases, such as that they almost used deep learning, transfer learning, and a hybrid model to classify the monkeypox data without explainability. But, we proposed a new method named Vision Transformer and a hybrid model with explainable AI with the best accuracy [see Table I]. So, our work is superior to theirs because our proposed work has the best accuracy from them. Particularly, we reached 100% accuracy using the vision transformer model without any overfitting and 99.33% using deep hybrid learning. So, we can say that our proposed work is the best work till now.

III. Methods

The overall workflow diagram of monkeypox disease detection and classification is illustrated in Fig. 1. In this part, we will discuss data collection, preprocessing, image augmentation, image separation, using of different types of deep and transfer learning models with details and proposed framework, and finally, we will explain explainable AI results applied to input image and predicted image. From, data collection to the result performance of each model, the sequential explanation is included.

A. Data Collection

We collected the image dataset from the online website such as Kaggle and we customized the data for later use https://www.kaggle.com/datasets/mdmokshedurrahman/monkeypox-image-dataset. This dataset has a total of six classes with one directory. We separated the data for training and testing. The total amount of image data is 7,532 and the classes are "Chickenpox", "HFMD", "Measles", "Healthy", "Cowpox", and "Monkeypox". The images are in different color modes. The sample dataset is shown in Fig. 2.

B. Image Preprocessing

Preprocessing images is an essential step in getting data ready for visual transformer models and transfer learning, particularly in applications like the diagnosis of monkeypox disease [28]. To start, pictures are gathered and their sizes are standardized to guarantee consistency, usually shrinking them to 224x224 pixels [29]. Convergence during model training is accelerated by normalization, which involves scaling pixel values to a range between 0 and 1 or standardizing them to have a mean of 0 and a standard deviation of 1.

C. Image Augmentation

In machine learning applications [30], [31] such as transfer learning, ensemble learning [32], and vision transformers, image augmentation is crucial for enhancing model performance by producing a variety of data variants. Rotation, flipping, scaling, cropping, and other techniques expand the amount of the dataset, which improves model generalization and lowers overfitting, particularly in small or unbalanced datasets [33]. Augmentation enables pre-trained models to successfully adjust to new datasets in transfer learning. By encouraging each model to concentrate on unique features, applying different augmentations across models for ensemble learning lowers prediction variance. Augments like patch shuffling and random cropping within vision transformers (ViTs) improve the model's resistance to visual fluctuations by enhancing its capacity to capture global patterns. Augmentation improves generalization overall, allowing models to better handle changes in real-world data. In our dataset, we applied the some rules for image augmentation [34]. We set the parameter as follows:

```
train-datagen= image.ImageDataGenerator(
rescale=1./255,
shear-range=0.2,
zoom-range=0.2,
horizontal-flip=True)
test-dataset=image.ImageDataGenerator(rescale=1./255)
```

D. Image Partitioning

After finishing the preprocessing and augmentation method to the image data, we separate the data for training, validation, and testing[35], [36]. We partitioned the data as follows:
total image for training: 5,273
total image for validation: 2,259 and
total image for validation: 2,259
That is, we separated the total images into three categories.

TABLE I. SUMMARY OF THE RELATED WORK

Reference	Dataset	Used Methods	Accuracy	XAI
[18]	MSID Dataset	Deep-CNN	98.91%	Yes
[19]	Monkeypox detection dataset	Deep Learning, YOLOV5	98.18%	No
[20]	Online image data	Deep learning and Attention Mechanism	83.89%	No
[21]	Online dataset	Deep learning ensembles	87.13%	Yes
[8]	Collected from online portal	Deep learning	82.96%	No
[22]	Their own data	Deep learning and federated learning	97.90%	No
[23]	Kaggle data	Deep neural network	97.63%	Yes
[7]	Real-time data	Deep-transfer learning	99%	No
[24]	Puclic data	Deep learning	99%	Yes
[25]	Puclicly availabe data	Hybrid Deep Learning	87%	No
[26]	Public dataset	AI, Chimp Algorithm and DL	98.48%	No
[27]	MSLD, MSID Online data	Transfer learning	94%	No
Proposed Approach	Online recent data	Vision Transformer, Hybrid Model	100% for ViTs, 99.33% for Hybrid	Yes

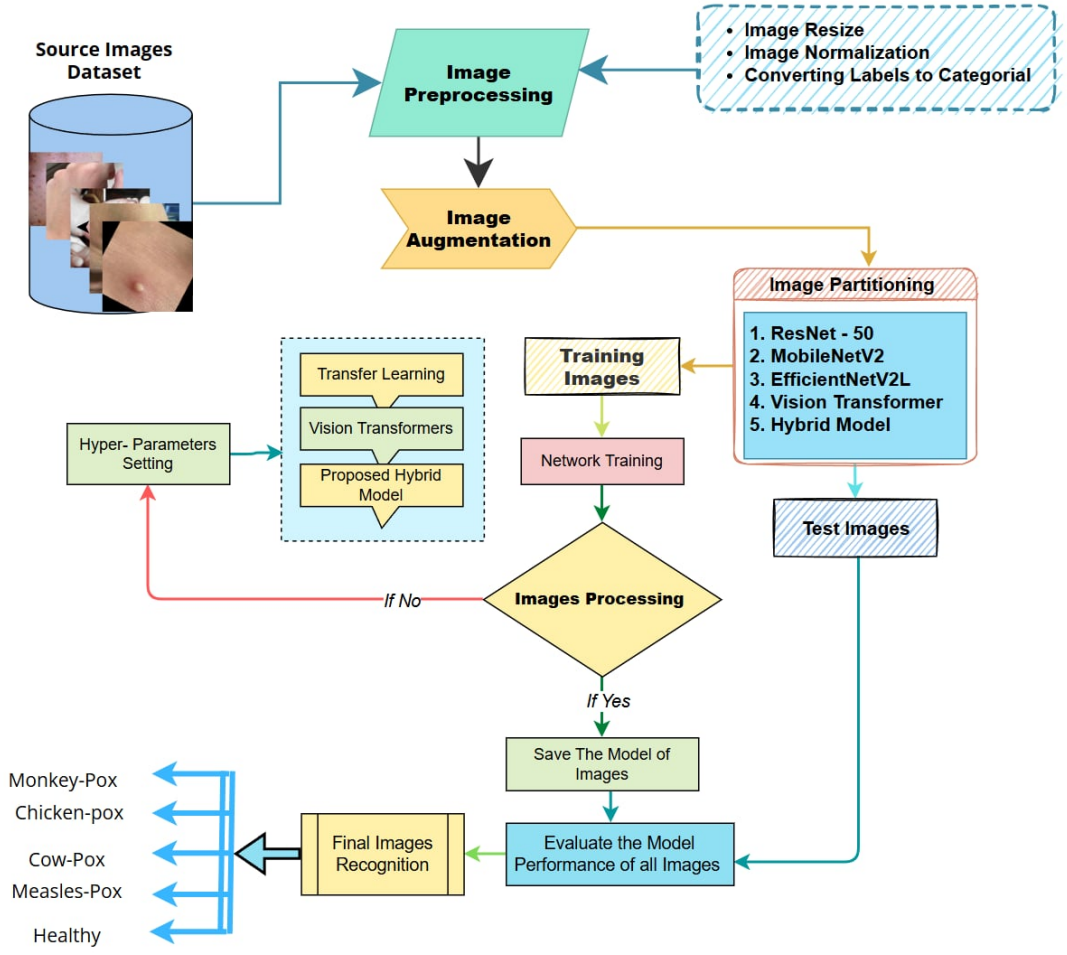


Fig. 1. System architecture.

E. Proposed Neural Network Framework

For image classification, deep learning models are suitable. These models can detect the disease more accurately. In this study, we applied some deep learning models such as EfficientNetV2 and MobileNetV2 [37]. A separate explanation is below.

1) EfficientNetV2: Monkeypox may be successfully detected by picture analysis using EfficientNetV2, a cutting-edge deep-learning model created for image classification applications [38]. By employing a compound scaling technique that consistently increases network depth, width, and resolution, EfficientNetV2's fundamental concept is its capacity to strike a compromise between accuracy and processing efficiency. The model is perfect for processing

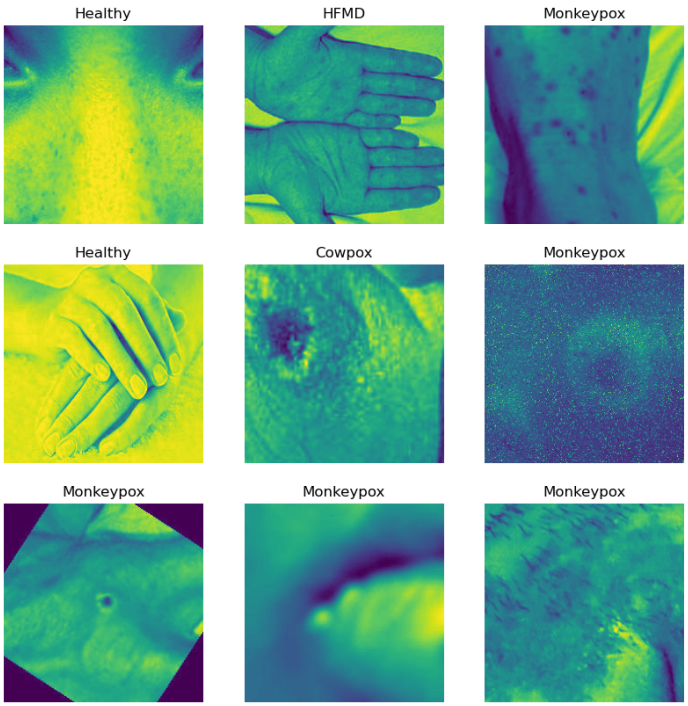


Fig. 2. Dataset sample.

medical images where accuracy is crucial because of its architecture, which allows it to extract complex information from images at a reduced computational cost. EfficientNetV2 may be optimized on a dataset of photos of monkeypox by utilizing transfer learning, which enables it to pick up unique patterns and characteristics that are suggestive of the illness. Its capacity to generalize from sparse data is further improved by its sophisticated training methods, such as progressive training, which begins with images of lower quality and progressively raises the resolution during training. In order to facilitate early diagnosis and prompt action in clinical settings, EfficientNetV2 provides a potent solution for precisely identifying monkeypox lesions in pictures. In this study, we used monkeypox image data to detect and classify the disease from those images. Total images were separated into six classes. In this model, we set the parameters as follows:
batch-size = 32
img-height = 224
img-width = 224
lr-rate = 1e-3
lr-mode = 'cos'
epochs = 30
For the training, we set the parameter as follows"
validation-split=0.2,
subset="training",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size
For the validation, we have,
data0dir,
validation0split=0.2,
subset="validation",
seed=123,
image0size=(img0height, img0width),
batch-size=batch-size

2) MobileNetV2: The lightweight deep learning model MobileNetV2 was created especially for mobile and edge devices, which makes it ideal for real-time applications like identifying monkeypox in medical photos. MobileNetV2's fundamental concept is based on depthwise separable convolutions, which drastically cut down on the number of parameters and calculations needed to analyze data quickly without sacrificing accuracy. Through a sequence of linear bottleneck layers that promote effective information flow and the retention of significant visual details, this architecture improves the model's capacity to learn key elements from images. MobileNetV2 may be optimized on a specific dataset of monkeypox photos by using transfer learning techniques, which will allow it to recognize the distinct patterns and traits linked to the illness. Because of its small size, the model may be used on gadgets with minimal processing power, like smartphones or portable medical imaging equipment, guaranteeing that medical practitioners can make good use of it in a variety of contexts. Consequently, MobileNetV2 offers a quick and easy way to identify monkeypox, which helps with prompt diagnosis and efficient treatment of the illness. In this study, we set the parameter as follows:
batch-size = 32
img-height = 224
img-width = 224
lr-rate = 1e-3
lr-mode = 'cos'
epochs = 15

For the training data, we used,
train-split=0.2,
subset="training",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size

For the validation of data, we difine,
validation-split=0.2,
subset="validation",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size

3) ResNet-50: An excellent option for identifying monkeypox in a six-class image dataset is ResNet-50, a potent transfer learning architecture that performs exceptionally well in image classification tasks. ResNet-50's fundamental concept is its creative use of residual connections, which mitigate the vanishing gradient issue that sometimes arises in very deep networks while enabling the model to learn intricate features [39]. These residual connections make it easier to train deeper networks by allowing gradients to have direct paths during backpropagation, which enhances the model's capacity to recognize complex patterns in images. ResNet-50 can be refined in the context of monkeypox detection using a broad dataset that comprises several classes associated with the disease, such as distinct lesion phases or other skin disorders. This flexibility improves classification accuracy

by enabling the algorithm to pick up subtle visual cues that distinguish monkeypox from related illnesses. ResNet-50 is also well-suited for managing sparse or unbalanced datasets, which are typical in medical imaging, due to its resilience to overfitting. In the end, ResNet-50 offers a dependable method for precisely identifying monkeypox lesions by utilizing its depth and architectural improvements, which helps with prompt diagnosis and efficient patient treatment in clinical settings [40]. In this study, we used the below parameter for ResNet-50 model training and also validation.
batch-size = 32
img-height = 224
img-width = 224
lr-rate = 1e-3
lr-mode = 'cos'
epochs = 30
For the training, we set the parameter as follows:
validation-split=0.2,
subset="training",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size
For the validation, we used the parameter list as follows:
validation-split=0.2,
subset="validation",
seed=123,
image-size=(img-height, img-width),
batch-size=batch-size

The basic architecture of the ResNet 50 model for this study is illustrated in Fig. 3.

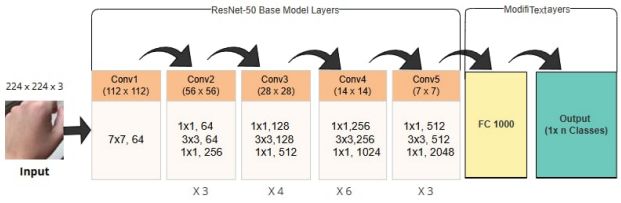


Fig. 3. Proposed ResNet 50 model architecture.

4) Vision Transformer: By adapting transformer-based designs, which were initially created for natural language processing, to visual data, the Vision Transformer (ViT) model(google/vit-base-patch16-224) offers a fresh method for image analysis [41]. ViT interprets images as a series of tiny patches, collecting global dependencies throughout the image, in contrast to convolutional neural networks (CNNs), which rely on local feature extraction [42]. It is especially useful for differentiating intricate visual patterns linked to illnesses like monkeypox because of its capacity to comprehend spatial relationships. ViT can learn the distinct visual indicators of monkeypox lesions, such as shape, texture, and distribution, across different phases and classes by training on a collection of monkeypox images. This method is useful for detecting monkeypox because it enables the model to understand both little details and more general contextual patterns, which helps it distinguish monkeypox from other skin disorders that are similar. Furthermore, ViT can concentrate on pertinent image regions thanks to its attention mechanism, which could improve interpretability in medical diagnostics. All things considered, Vision Transformer offers a potential tool for detecting monkeypox by fusing high accuracy with knowledge of the spatial patterns that characterize the illness [43]. In our study, we follow the working mechanism of the Vision Transformer model shown in Fig. 4.

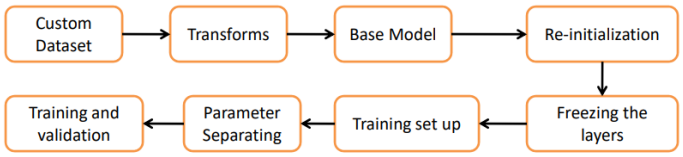


Fig. 4. Working procedure of the vision transformer model.

Initially, we customized the dataset for the transformer. The parameter set as:
transform = transforms.Compose([
transforms.Resize((224, 224)),
transforms.ToTensor(),
transforms.Normalize(mean=[0.5, 0.5, 0.5],
std=[0.5, 0.5, 0.5]),])

After transforming, we defined the base model and re-initialized the features. Freezing the layers, we did training setup and parameter dividing for the training and validation.

5) Hybrid Model: To obtain better performance and more accuracy, we used a hybrid version of three models. The hybrid model is generated from averaging the output from ResNetV2, MobilNetV2, and EfficientNetV2 [44]. This hybrid model technique detects and classifies monkeypox across six different image classes by averaging the outputs of three sophisticated deep-learning architectures: ResNetV2, MobileNetV2, and EfficientNetV2. Every one of these types has special advantages: While MobileNetV2 offers lightweight efficiency, making it highly responsive and appropriate for real-time processing with limited CPU resources, ResNetV2's residual connections enable the capture of intricate image features and deep hierarchical patterns. In contrast, EfficientNetV2 offers a balanced scaling technique that simultaneously modifies network depth, width, and resolution to improve accuracy and efficiency. The hybrid model leverages the combined advantages of each architecture by averaging the predictions from these three networks. This integration lessens the biases present in any one model, producing a more robust and balanced outcome that is particularly helpful for managing the many visual traits of monkeypox lesions in various classes [45]. Even with the variances found in a medical picture dataset, the hybrid model can function effectively thanks to the averaging technique, which may enhance generalization and lower mistakes. In the end, this team approach improves the accuracy and dependability of monkeypox detection, offering a complete tool to assist medical professionals in diagnosing and categorizing the illness. In this combined model, we set the features for the training data and also the same for validation as follows:
validation-split=0.2,
subset="training",

seed=123,
image-size=(224, 224),
batch-size=32)
For the base model, the parameters are:
weights='imagenet',
include-top=False,
input-shape=(224, 224, 3)

The basic organization of the proposed hybrid model is illustrated in Fig. 5.

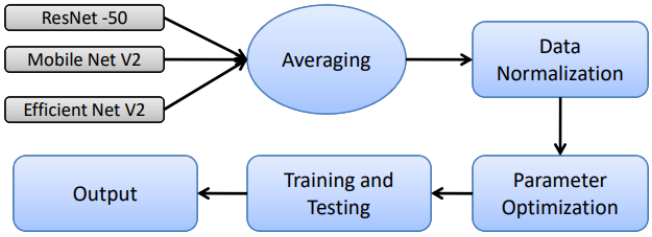


Fig. 5. Working procedure of proposed hybrid model.

In Fig. 5, for the section of Normalization, we set the parameter as:
Rescaling=1./255,
map(lambda x, y: (normalization-layer(x), y)).cache().prefetch(buffer-size=tf.data.AUTOTUNE)
We optimized the some parameters such as:
Input(shape=(224, 224, 3))
optimizers.Adam(learning-rate=1e-3),
losses.SparseCategoricalCrossentropy(from-logits=True),
metrics=['accuracy']
epochs=30

F. Explainable AI Approach

We use Grad-CAM (Gradient-weighted Class Activation Mapping) to show the relevant areas in monkeypox images as part of our Explainable AI feature extraction method for monkeypox detection [46]. The interpretability technique Grad-CAM improves the transparency of deep learning models by assisting in determining which aspects of an image have the greatest influence on a model's prediction. This method highlights the main characteristics in the images used for monkeypox categorization across several classes by implementing Grad-CAM for a ResNet-based model [21]. First, a Grad-CAM class is created, which initializes a gradient model by choosing the network's last convolutional layer and attaching it to the output layer of the model. Grad-CAM computes the gradient of the output class (predicted as monkeypox or another) in relation to the feature mappings in the final convolutional layer after an input picture has been run through the model. The features that are important for the model's categorization are shown by these gradients [47]. A heatmap highlighting the significant regions of the image that influenced the model's prediction is produced after pooling the gradients and appropriately weighting the feature maps. In order to create a superimposed visualization, the calculated heatmap is enlarged to the original image proportions, colored using a "jet" colormap, and then superimposed on the original image. Medical practitioners may more easily validate the model's focus areas and comprehend predictions thanks to this overlay, which shows the portions of the image that the algorithm looks for in order to detect monkeypox. By making the model's decision-making process more clear, these explainable strategies increase confidence in the model's application for medical imaging diagnosis and categorization of monkeypox.

IV. Results

In this section, we discussed the result of the deep learning model in Monkeyfox disease detection [16]. Particularly, we explored the results of EfficientNetV2 and MobileNetV2.

From the above part, we know that EfficientNetV2 is used in the monkeypox image dataset and has a total of six classes. The basic parameter details of EfficientNetV2 is shown in Table II.

TABLE II. EfficientNetV2 Parameter Details

Layer (type)	Output Shape	Param
input-layer-18 (Input-Layer)	(None, 224, 224, 3)	0
efficientnetv2-l (Functional)	(None, 7, 7, 1280)	117,746,848
conv2d-7 (Conv2D)	(None, 7, 7, 512)	5,898,752
global-average-pooling2d-7	(None, 512)	0
dense-14 (Dense)	(None, 256)	131,328
dense-15 (Dense)	(None, 6)	1,542
Total params:	123,778,470	0
Trainable params:	6,031,622	0
Non-trainable params:	117,746,848	

In this model,
Training accuracy is 99.94%
Training loss is 0.38%.
The validation accuracy is 97.80% and
The validation loss is 6.72%.
The testing accuracy is: 97.67%
The testing loss is: 6.94%.
The training accuracy and validation accuracy are shown in Fig. 6 and the training loss and validation loss are shown in Fig. 7.

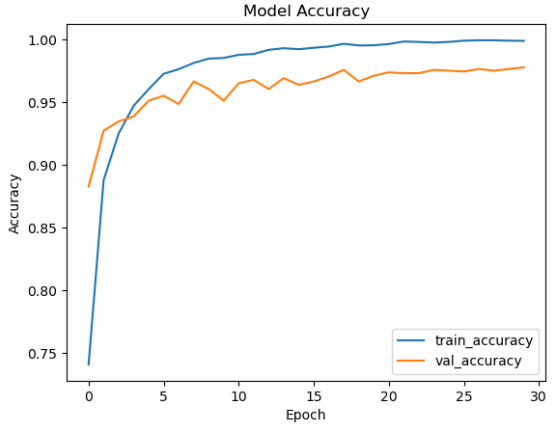


Fig. 6. Train accuracy vs. Validation accuracy of EfficientNetV2.

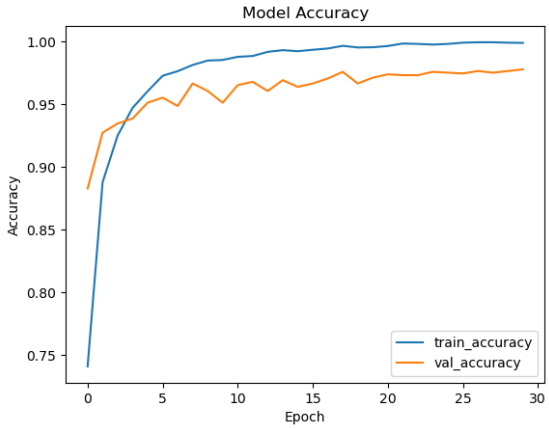


Fig. 7. Train loss vs. Validation loss of EfficientNetV2.

The classification report of the EfficientNetV2 is illustrated in Table III.

TABLE III. CLASSIFICATION REPORT OF EFFICIENTNETV2 MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.95	0.96	0.95	140
Chickenpox	0.95	0.96	0.95	140
Cowpox	1.00	0.97	0.98	120
HFMD	0.99	0.98	0.98	345
Healthy	0.99	0.99	0.99	251
Measles	0.99	0.94	0.96	98
Monkeypox	0.97	0.99	0.98	549
accuracy	0	0	0.98	1503
macro avg	0.98	0.97	0.97	1503
weighted avg	0.98	0.98	0.98	1503

We used MobileNetV2 in image data for detecting monkeypox. After applying the model, we have the following parameter list shown in Table IV.

TABLE IV. MOBILENETV2 PARAMETER DETAILS

Layer (type)	Output Shape	Param
input-layer-14 (Input Layer)	(None, 224, 224, 3)	0
mobilenetv2-1.00-224	(None, 7, 7, 1280)	2,257,984
conv2d-5 (Conv2D)	(None, 7, 7, 512)	5,898,752
global-average-pooling2d-5	(None, 512)	0
dense-10 (Dense)	(None, 256)	131,328
dense-11 (Dense)	(None, 6)	1,542

In this model, we got,
Training accuracy is 95.47%
Training loss is 16.2%.
The validation accuracy is 79.51% and
The validation loss is 64.62%.
The testing accuracy is: 78.18%
and the testing loss is: 20%.
The classification report of MobileNetV2 is shown in Table V.

We used one Transfer Learning model named ResNet-50 to detect the monkeypox from the image data. After applying this model, we have the model parameter summary shown in Table VI.

TABLE V. CLASSIFICATION REPORT OF MOBILENETV2 MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.95	0.96	0.95	140
Chickenpox	1.00	0.97	0.98	140
Cowpox	1.00	0.97	0.98	120
HFMD	0.99	0.98	0.98	345
Healthy	0.99	0.99	0.99	251
Measles	0.99	0.94	0.96	98
Monkeypox	0.97	0.99	0.98	549
accuracy	0	0	0.98	1503
macro avg	0.98	0.97	0.97	1503
weighted avg	0.98	0.98	0.98	1503

TABLE VI. RESNET-50 PARAMETER DETAILS

Layer (type)	Output Shape	Param
input-layer-12 (Input Layer)	(None, 224, 224, 3)	0
ResNet-50 (Functional)	(None, 7, 7, 2048)	23,587,712
conv2d-4 (Conv2D)	(None, 7, 7, 512)	9,437,696
global-average-pooling2d-4	(None, 512)	0
dense-8 (Dense)	(None, 256)	131,328
dense-9 (Dense)	(None, 6)	1,542

Training accuracy is 99.87%%
Training loss is 0.43%.
The validation accuracy is 99.85% and
The validation loss is 0.39%.
The testing accuracy is: 97.18%
and the testing loss is: 4%.
The classification report of this model is shown in Table VII and the ROC Curve of this model is shown in Fig. 8.

TABLE VII. CLASSIFICATION REPORT OF RESNET-50 MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.97	0.86	0.91	140
Cowpox	0.97	0.98	0.98	120
HFMD	0.98	0.97	0.98	345
Healthy	0.97	0.98	0.97	251
Measles	0.96	0.95	0.95	98
Monkeypox	0.96	0.98	0.97	549
accuracy	0	0	0.97	1503
macro avg	0.97	0.95	0.96	1503
weighted avg	0.97	0.97	0.97	1503

The accuracy and loss curve of this model is illustrated in Fig. 9.

Using Vision Transformer (google/vit-base-patch16-224) for detecting monkeypox disease, we have the following results.
The number of epochs: 20
Training accuracy is 100%
Training loss is 0.00%.
The validation accuracy is 87.51% and
The validation loss is 0.37%.

The classification report is shown in Table VIII and the confusion matrices of this model is shown in Fig. 10 where true label vs predicted label and actual label vs. predicted lable is illustrated. The multiclass ROC Curve and Precision-recall curve is explained in Fig. 11 and 12.

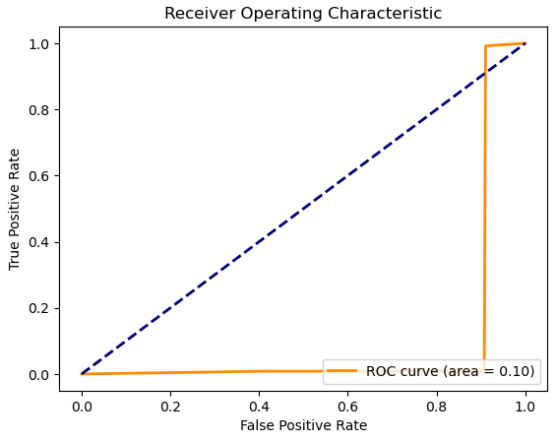


Fig. 8. The ROC curve for ResNet-50 model.

TABLE VIII. CLASSIFICATION REPORT OF VISION TRANSFORMER MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.64	0.86	0.73	110
Cowpox	0.96	0.93	0.94	91
HFMD	0.94	0.97	0.96	229
Healthy	0.95	0.86	0.90	175
Measles	0.96	0.89	0.92	80
Monkeypox	0.88	0.84	0.87	445
accuracy	0	0	0.88	1130
macro avg	0.89	0.89	0.89	1130
weighted avg	0.89	0.89	0.89	1130

We combined three models to improve the accuracy and reduce the error rate as well as loss amount. Hence, ResNet-50, MobileNetV2, and EfficientNetV2 models are averaging into a single unit named as Hybrid model. After applying this technique, we have the following results:
The number of epochs: 30
Training accuracy is 99.33%
Training loss is 2.11%.
The validation accuracy is 90.09% and
The validation loss is 40.62%.

The summary of the parameter is listed in Table IX and the classification report of this model is shown in Table X.

TABLE IX. HYBRID MODEL PARAMETER DETAILS

Layer (type)	Output Shape	Param	Connected to
input-layer-14 (Input Layer)	(None, 224, 224, 3)	0	-
functional-15	(None, 6)	24,113,798	input-layer-32[0...
functional-16	(None, 6)	2,587,462	input-layer-32[0...
functional-17	(None, 6)	118,076,3...	input-layer-32[0...
average-1 (Average)	(None, 6)	0	functional-15[0]... functional-16[0] functional-17[0]...

The accuracy and loss curve is shown in Fig. 13 and the confusion matrix is shown in Fig. 14.

TABLE X. CLASSIFICATION REPORT OF HYBRID MODEL

Class-Name	Precision	Recall	F1-Score	Support
Chickenpox	0.94	0.77	0.85	140
Cowpox	0.99	0.87	0.92	120
HFMD	0.98	0.85	0.91	345
Healthy	0.94	0.91	0.93	251
Measles	0.83	0.83	0.83	98
Monkeypox	0.84	0.98	0.91	549
accuracy	0	0	0.90	1503
macro avg	0.92	0.77	0.89	1503
weighted avg	0.91	0.90	0.90	1503

The results summary of the proposed model are shown in Table XI.

TABLE XI. RESULT SUMMARY OF USED METHODS

Model	Training Accuracy	Validation Accuracy	Testing Accuracy
EfficientNetV2	99.94%	97.80%	97.67%
MobileNetV2	95.47%	79.51%	78.18%
ResNet-50	99.87%	99.85%	97.18%
Vision Transformer	100%	87.51%	99.41%
Hybrid Model	99.33%	90.09%	99.09%

A. Exploring Grad-CAM

We applied explainable AT to the predicted image to explain it based on trained images. If we use any monkeypox-positive image for the explanation, then based on the training and predicted value, the machine can explain the image using the heat-map method [48]. Using the Grad-CAM, the system can explain the input image for clearance. One suitable example is shown in Fig. 15. Especially, this image is the monkeypox positive input image, the system will use a heat map to analyze and explain it. Finally, the system is successful, saying that it is the monkeypox positive image [49]. Table XII shows evaluation metrics for Grad-CAM for hybrid model.

B. Comparison with Previous Studies

The comparison presented in Table XIII highlights the effectiveness of the proposed models in achieving state-of-the-art performance for monkeypox diagnosis. Our study demonstrated superior accuracy with Vision Transformers (ViTs) achieving 99.41% and the hybrid model achieving 99.09%. These results outperform many of the existing studies, such as [18] (98.91%) and [7] (99%), showcasing the robustness of our approach.

The incorporation of Vision Transformers proved particularly impactful due to their ability to capture global dependencies within the input data, which is critical for nuanced image classification tasks. The hybrid model further enhanced performance by combining the strengths of ResNet-50, MobileNetV2, and EfficientNetV2, enabling better feature extraction and classification accuracy.

Compared to prior studies, such as [20] and [8], which reported lower accuracies of 83.89% and 82.96%, respectively, our models exhibited a significant improvement. Additionally, while models like [21] and [23] incorporated Explainable AI (XAI) techniques, their accuracies (87.13% and 97.63%) were

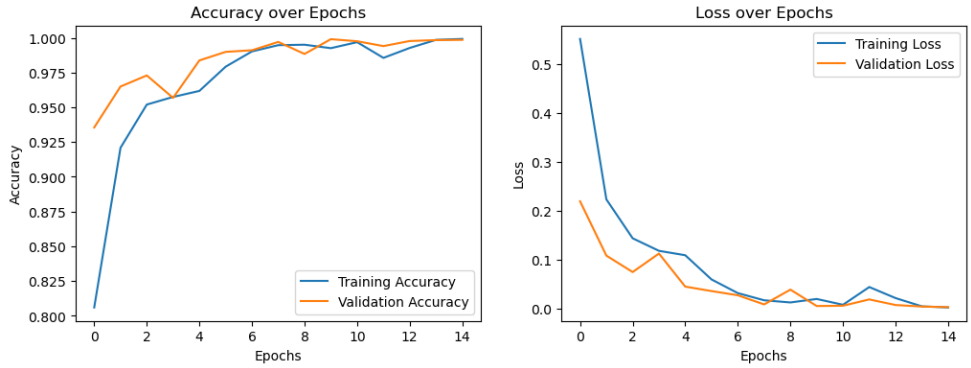


Fig. 9. Accuracy and loss curve of ResNet-50 model.

TABLE XII. EVALUATION METRICS FOR GRAD-CAM FOR HYBRID MODEL

Evaluation Metric	Grad-CAM	Explanation
Ground Truth Mask Overlap	97%	Percentage of overlap between the ground truth mask and the highlighted region.
Feature Coverage	0.97	Proportion of the image covered by the highlighted features in the explanation.
Relevant Activation	96%	Percentage of activation in relevant areas.
Feature Relevance	0.96	Relevance of features in the explanation, corresponding to the model's decision.
Similarity (Mean Absolute Error)	0.89	Mean absolute error between the predicted and actual class.
Consistency Error	0.89	Error in consistency when input is perturbed or modified.

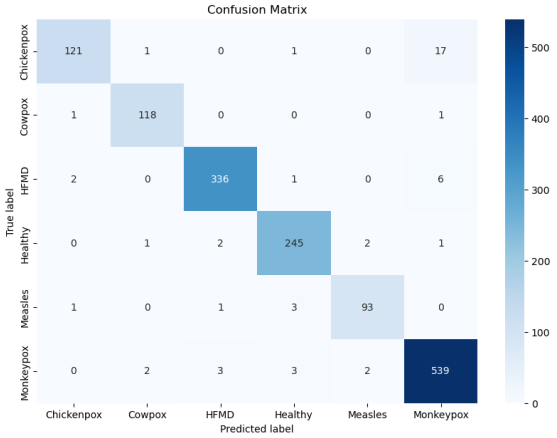


Fig. 10. Confusion matrix of vision transformer in true label vs. Predicted label.

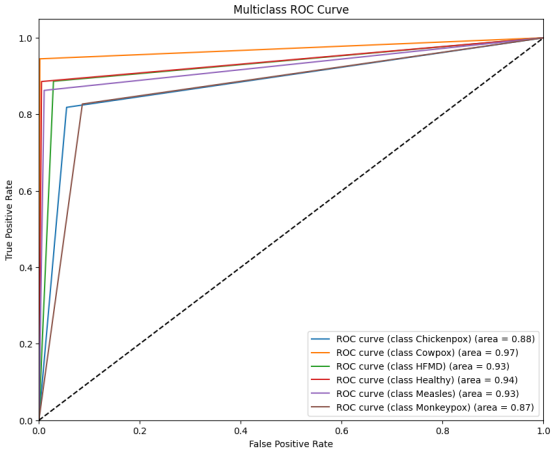


Fig. 11. ROC Curve for vision transformer model.

lower than ours, demonstrating that our integration of Grad-CAM not only enhanced interpretability but also maintained high performance.

Explainability remains a critical aspect of monkeypox diagnosis, as accurate predictions alone are insufficient in sensitive medical applications. Our use of Grad-CAM enabled a detailed understanding of model decisions, providing visual insights into key features contributing to the classification. This is a step forward in building trust and transparency in AI-based healthcare solutions, addressing concerns in studies like [24] and [26], which either did not integrate XAI or lacked detailed visualization.

V. CONCLUSION AND FUTURE RESEARCH

This study demonstrates the potential of deep learning models, particularly Vision Transformer and hybrid approaches, in achieving accurate and interpretable monkeypox diagnosis. Among the models tested, the Vision Transformer emerged as the most effective, achieving high accuracy across training, validation, and testing phases while maintaining robustness against overfitting. The hybrid model, combining ResNet-50, MobileNetV2, and EfficientNetV2, also delivered competitive performance, highlighting the benefits of leveraging diverse architectural strengths. The integration of Grad-CAM enhanced the interpretability of the models, providing valuable insights into their decision-making processes, a critical requirement for clinical applications. These findings highlight the role of AI-driven solutions in enabling early and

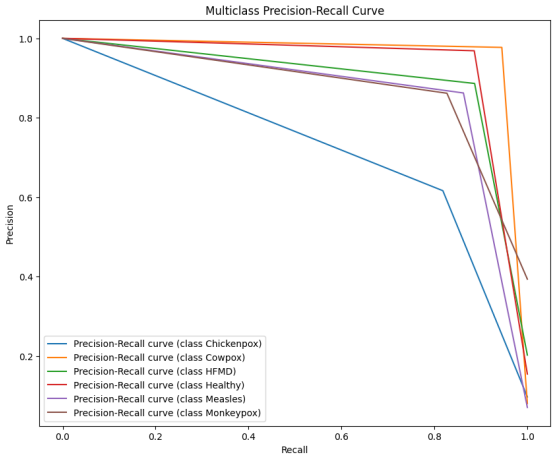


Fig. 12. Precision-recall curve for vision transformer model.

TABLE XIII. COMPARISON WITH PREVIOUS STUDIES

Reference	Accuracy	XAI
[18]	98.91%	Yes
[19]	98.18%	No
[20]	83.89%	No
[21]	87.13%	Yes
[8]	82.96%	No
[22]	97.90%	No
[23]	97.63%	Yes
[7]	99%	No
[24]	99%	Yes
[25]	87%	No
[26]	98.48%	No
[27]	94%	No
Our Study	99.41% for ViTs, 99.09% for Hybrid	Yes

precise monkeypox diagnosis, thereby aiding timely containment and treatment. One of the limitations of this study is that the dataset was taken from an online publication from the clinical sector, and to detect the proper place of monkeypox, the segmentation method can be applied in real-time image data. Future research will cover this technique. Future research should focus on enhancing the generalizability of the proposed models by expanding the dataset to include diverse populations, imaging conditions, and clinical real-time data. Additionally, improving model efficiency for deployment in resource-constrained environments will be crucial for enabling widespread adoption. Incorporating other explainability techniques, such as LIME or SHAP, could provide deeper insights into model predictions, fostering greater trust among clinicians. Exploring federated learning frameworks may further enhance privacy and scalability, allowing collaborative training across institutions without compromising data security. Longitudinal studies spanning various demographic and clinical contexts will help validate model reliability over time. Moreover, integrating multi-modal data, such as clinical biomarkers and patient metadata, could improve diagnostic accuracy and provide a more holistic understanding of monkeypox.

DATA AVAILABILITY

The used datasets are open-access and referenced in this manuscript.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] S. Das, T. Mahmud, D. Islam, M. Begum, A. Barua, M. Tarek Aziz, E. Nur Showan, L. Dey, and E. Chakma, "Deep transfer learning-based foot no-ball detection in live cricket match," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, p. 2398121, 2023.

[2] S. Umme Habiba, F. Tasnim, M. S. Hasan Chowdhury, M. K. Islam, L. Nahar, T. Mahmud, M. S. Kaiser, M. S. Hossain, and K. Andersson, "Early prediction of chronic kidney disease using machine learning algorithms with feature selection techniques," in *International Conference on Applied Intelligence and Informatics*. Springer, 2023, pp. 224–242.

[3] S. U. Habiba, T. Mahmud, S. R. Naher, M. T. Aziz, T. Rahman, N. Datta, M. S. Hossain, K. Andersson, and M. Shamim Kaiser, "Deep learning solutions for detecting bangla fake news: A cnn-based approach," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 107–118.

[4] M. T. Aziz, T. Mahmud, N. Datta, M. Maskat Sharif, N. U. A. Khan, S. Yasmin, M. D. N. Uddin, M. S. Hossain, and K. Andersson, "A state-of-the-art review of machine learning in cybersecurity data science," in *Innovations in Cybersecurity and Data Science*. Singapore: Springer Nature Singapore, 2024, pp. 791–806.

[5] S. R. Naher, S. Sultana, T. Mahmud, M. T. Aziz, M. S. Hossain, and K. Andersson, "Exploring deep learning for chittagonian slang detection in social media texts," in *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET*. IEEE, 2024, pp. 1–6.

[6] T. Mahmud, K. Barua, K. Chakma, R. Chakma, N. Sharmen, M. S. Kaiser, M. S. Hossain, M. S. Hossain, and K. Andersson, "Exploring the effectiveness of region-based cnns in skin cancer diagnosis," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 371–389.

[7] M. Altun, H. Gürüler, O. Özkaraca, F. Khan, J. Khan, and Y. Lee, "Monkeypox detection using cnn with transfer learning," *Sensors*, vol. 23, no. 4, p. 1783, 2023.

[8] S. N. Ali, M. T. Ahmed, J. Paul, T. Jahan, S. Sani, N. Noor, and T. Hasan, "Monkeypox skin lesion detection using deep learning models: A feasibility study," *arXiv preprint arXiv:2207.03342*, 2022.

[9] T. Mahmud, T. Akter, M. T. Aziz, M. K. Uddin, M. S. Hossain, and K. Andersson, "Integration of nlp and deep learning for automated fake news detection," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2024, pp. 398–404.

[10] N. A. Chowdhury, T. Mahmud, A. Barua, N. Basnin, K. Barua, A. Iqbal, M. S. Hossain, K. Andersson, M. S. Kaiser, M. S. Hossain *et al.*, "A novel approach to detect stroke from 2d images using deep learning," in *International Conference on Big Data, IoT and Machine Learning*. Springer, 2023, pp. 239–253.

[11] M. T. Aziz, J. Sikder, T. Rahman, A. D. Del Mundo, S. F. Faisal, and N. U. A. Khan, "Covid-19 detection from chest x-ray images using deep learning," *The Seybold Report*, vol. 17, pp. 706–718, 2022.

[12] M. H. Ali, T. Mahmud, M. T. Aziz, M. F. B. A. Aziz, M. S. Hossain, and K. Andersson, "Leveraging transfer learning for efficient classification of coffee leaf diseases," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–6.

[13] T. Mahmud, N. Datta, R. Chakma, U. K. Das, M. T. Aziz, M. Islam, A. H. M. Salimullah, M. S. Hossain, and K. Andersson, "An approach for crop prediction in agriculture: Integrating genetic algorithms and machine learning," *IEEE Access*, 2024.

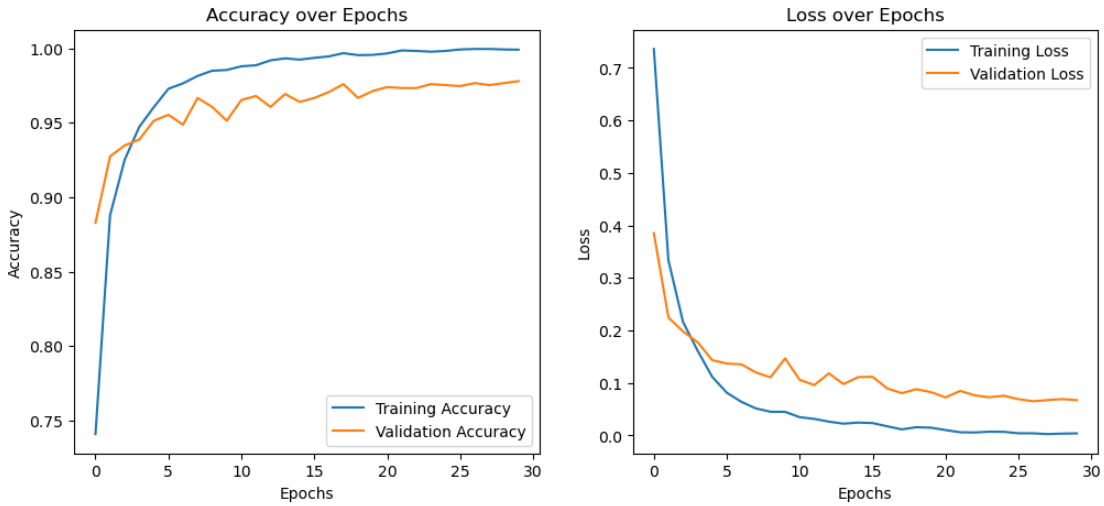


Fig. 13. Accuracy and loss curve for hybrid model.

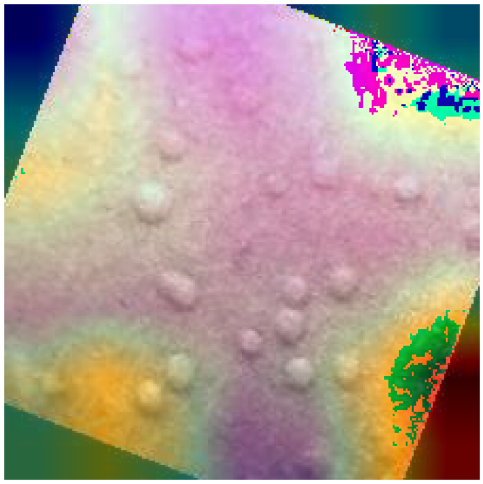


Fig. 15. Explainable AI predicted output image by Grad-CAM for hybrid model.

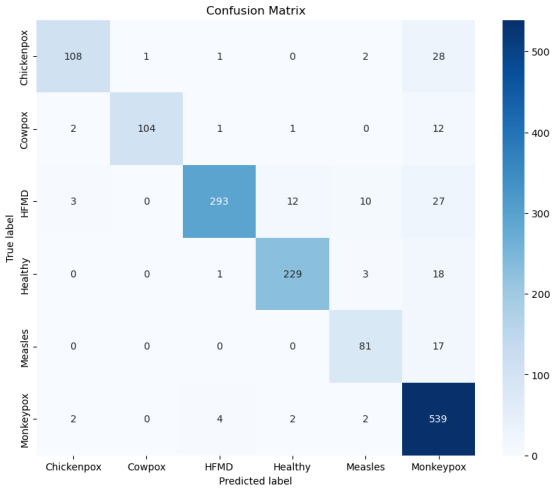


Fig. 14. The confusion matrix of hybrid model.

[14] T. Mahmud, T. Akter, S. Anwar, M. T. Aziz, M. S. Hossain, and K. Andersson, "Predictive modeling in forex trading: A time series analysis approach," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2024, pp. 390–397.

[15] N. Datta, T. Mahmud, M. T. Aziz, R. K. Das, M. S. Hossain, and K. Andersson, "Emerging trends and challenges in cybersecurity data science: A state-of-the-art review," in *2024 Parul International Conference on Engineering and Technology (PICET)*. IEEE, 2024, pp. 1–7.

[16] T. Mahmud, I. Hasan, M. T. Aziz, T. Rahman, M. S. Hossain, and K. Andersson, "Enhanced fake news detection through the fusion of deep learning and repeat vector representations," in *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2024, pp. 654–660.

[17] M. T. Aziz, T. Mahmud, N. Datta, M. M. Sharif, N. U. A. Khan, S. Yasmin, M. N. Uddin, M. S. Hossain, and K. Andersson, "A state-of-the-art review of machine learning in cybersecurity data science."

[18] D. Bala, M. S. Hossain, M. A. Hossain, M. I. Abdullah, M. M. Rahman, B. Manavalan, N. Gu, M. S. Islam, and Z. Huang, "Monkeynet: A robust deep convolutional neural network for monkeypox disease detection and classification," *Neural Networks*, vol. 161, pp. 757–775, 2023.

[19] N. Dahiya, Y. K. Sharma, U. Rani, S. Hussain, K. V. Nabilal, A. Mohan, and N. Nuristani, "Hyper-parameter tuned deep learning approach for effective human monkeypox disease detection," *Scientific Reports*, vol. 13, no. 1, p. 15930, 2023.

[20] M. E. Haque, M. R. Ahmed, R. S. Nila, and S. Islam, "Human monkeypox disease detection using deep learning and attention mechanisms," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 1069–1073.

[21] C. Sitaula and T. B. Shahi, "Monkeypox virus detection using pre-trained deep learning-based approaches," *Journal of Medical Systems*, vol. 46, no. 11, p. 78, 2022.

[22] D. Kundu, M. M. Rahman, A. Rahman, D. Das, U. R. Siddiqi, M. G. R. Alam, S. K. Dey, G. Muhammad, and Z. Ali, "Federated deep learning for monkeypox disease detection on gan-augmented dataset," *IEEE Access*, 2024.

[23] A. Sorayaie Azar, A. Naemi, S. Babaei Rikan, J. Bagherzadeh Mohasefi, H. Pirnejad, and U. K. Wiil, "Monkeypox detection using deep neural networks," *BMC Infectious Diseases*, vol. 23, no. 1, p. 438, 2023.

[24] M. M. Ahsan, T. A. Abdullah, M. S. Ali, F. Jahora, M. K. Islam, A. G. Alhashim, and K. D. Gupta, "Transfer learning and local interpretable model agnostic based visual approach in monkeypox disease detection and classification: A deep learning insights," *arXiv preprint arXiv:2211.05633*, 2022.

[25] F. Uysal, "Detection of monkeypox disease from human skin images

with a hybrid deep learning model," *Diagnostics*, vol. 13, no. 10, p. 1772, 2023.

[26] A. I. Saleh and A. H. Rabie, "Human monkeypox diagnose (hmd) strategy based on data mining and artificial intelligence techniques," *Computers in Biology and Medicine*, vol. 152, p. 106383, 2023.

[27] M. F. Almufareh, S. Tehsin, M. Humayun, and S. Kausar, "A transfer learning approach for clinical detection support of monkeypox skin lesions," *Diagnostics*, vol. 13, no. 8, p. 1503, 2023.

[28] Y. Zhang, J. Wang, J. M. Gorriz, and S. Wang, "Deep learning and vision transformer for medical image analysis," p. 147, 2023.

[29] R. Karthik, V. Thalanki, and P. Yadav, "Deep learning-based histopathological analysis for colon cancer diagnosis: A comparative study of cnn and transformer models with image preprocessing techniques," in *International Conference on Intelligent Systems Design and Applications*. Springer, 2023, pp. 90–101.

[30] M. T. Aziz, T. Mahmud, M. K. Uddin, S. N. Hossain, N. Datta, S. Akther, M. S. Hossain, and K. Andersson, "Machine learning-driven job recommendations: Harnessing genetic algorithms," in *International Congress on Information and Communication Technology*. Springer, 2024, pp. 471–480.

[31] T. Mahmud, M. Ptaszynski, and F. Masui, "Leveraging explainable ai and sarcasm features for improved cyberbullying detection in multilingual settings," in *2024 IEEE Digital Platforms and Societal Harms (DPSH)*. IEEE, 2024, pp. 1–8.

[32] S. Barman, M. R. Biswas, S. Marjan, N. Nahar, M. H. Imam, T. Mahmud, M. S. Kaiser, M. S. Hossain, and K. Andersson, "A two-stage stacking ensemble learning for employee attrition prediction," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 119–132.

[33] D. Sehgal and I. Saini, "Gan-based image augmentation and comparative analysis of various cnn models for monkeypox detection," in *2024 First International Conference on Electronics, Communication and Signal Processing (ICECSP)*. IEEE, 2024, pp. 1–7.

[34] N. E. Khalifa, M. Loey, and S. Mirjalili, "A comprehensive survey of recent trends in deep learning for digital images augmentation," *Artificial Intelligence Review*, vol. 55, no. 3, pp. 2351–2377, 2022.

[35] P. Dey, T. Mahmud, K. M. Foysol, N. Sharmen, M. S. Hossain, and K. Andersson, "Hybrid deep transfer learning framework for humerus fracture detection and classification from x-ray images," in *2023 4th International Conference on Intelligent Technologies (CONIT)*. IEEE, 2024, pp. 1–6.

[36] P. Dey, T. Mahmud, M. S. Chowdhury, M. S. Hossain, and K. Andersson, "Human age and gender prediction from facial images using deep learning methods," *Procedia Computer Science*, vol. 238, pp. 314–321, 2024.

[37] M. H. Imam, N. Nahar, R. Bhowmik, S. B. S. Omit, T. Mahmud, M. S. Hossain, and K. Andersson, "A transfer learning-based framework: Mobilenet-svm for efficient tomato leaf disease classification," in *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, 2024, pp. 693–698.

[38] S. Vats, J. P. Bhati, A. Singla, V. Kukreja, and R. Sharma, "Advanced image classification on intel datasets using optimized efficientnet and mobilenetv2," in *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*. IEEE, 2024, pp. 1–4.

[39] M. B. Sahaai, G. Jothilakshmi, D. Ravikumar, R. Prasath, and S. Singh, "Resnet-50 based deep neural network using transfer learning for brain tumor classification," in *AIP Conference Proceedings*, vol. 2463, no. 1. AIP Publishing, 2022.

[40] T. Tian, L. Wang, M. Luo, Y. Sun, and X. Liu, "Resnet-50 based technique for eeg image characterization due to varying environmental stimuli," *Computer Methods and Programs in Biomedicine*, vol. 225, p. 107092, 2022.

[41] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu *et al.*, "A survey on vision transformer," *IEEE transactions on pattern analysis and machine intelligence*, vol. 45, no. 1, pp. 87–110, 2022.

[42] M. Aloraini, "An effective human monkeypox classification using vision transformer," *International Journal of Imaging Systems and Technology*, vol. 34, no. 1, p. e22944, 2024.

[43] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, C. Xu, and W. Gao, "Pre-trained image processing transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 12 299–12 310.

[44] T. Mahmud, M. T. Aziz, M. K. Uddin, K. Barua, T. Rahman, N. Sharmen, M. Shamim Kaiser, M. Sazzad Hossain, M. S. Hossain, and K. Andersson, "Ensemble learning approaches for alzheimer's disease classification in brain imaging data," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 133–147.

[45] S. Maqsood, R. Damaševičius, S. Shahid, and N. D. Forkert, "Moxnet: Multi-stage deep hybrid feature fusion and selection framework for monkeypox classification," *Expert Systems with Applications*, vol. 255, p. 124584, 2024.

[46] G. M. Idroes, T. R. Noviandy, T. B. Emran, and R. Idroes, "Explainable deep learning approach for mpox skin lesion detection with grad-cam," *Heca Journal of Applied Sciences*, vol. 2, no. 2, pp. 54–63, 2024.

[47] A. Akram, A. A. Jamjoom, N. Innab, N. A. Almujally, M. Umer, S. Alsubai, and G. Fimiani, "Skinmarknet: an automated approach for prediction of monkeypox using image data augmentation with deep ensemble learning models," *Multimedia Tools and Applications*, pp. 1–17, 2024.

[48] A. Chaddad, J. Peng, J. Xu, and A. Bouridane, "Survey of explainable ai techniques in healthcare," *Sensors*, vol. 23, no. 2, p. 634, 2023.

[49] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma, P. N. Bokoro, and R. Sharma, "Explainable ai for healthcare 5.0: opportunities and challenges," *IEEE Access*, vol. 10, pp. 84 486–84 517, 2022.

Explainable Deep Transfer Learning Framework for Rice Leaf Disease Diagnosis and Classification

Md Mokshedur Rahman[1], Zhang Yan[2], Mohammad Tarek Aziz[3],
MD Abu Bakar Siddick[4], Tien Truong[5], Md. Maskat Sharif[6],
Nippon Datta[7], Tanjim Mahmud[8], Renzon Daniel Cosme Pecho[9], Sha Md Farid[10]
Department of Computer Science and Technology,
Beijing Institute of Technology (BIT), Beijing, China[1,2,4]
Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Bangladesh[3,7]
Department of Economics and Cognitive Science, University of California at Berkeley, Berkeley, United States[5]
Department of Electronics and Telecommunication Engineering,
Chittagong University of Engineering and Technology, Bangladesh[6]
Department of Computer Science and Engineering,
Rangamati Science and Technology University, Rangamati-4500, Bangladesh[8]
Diagnostic Imaging Center, Lima, Peru[9]
Department of Technology, Wilmington University, Delaware, United States[10]

***Abstract*—Rice plays a vital role in the food stock. But sometimes this crop leaf falls into disease. And, the amount of food consumed will decrease due to leaf disease. So, discovering the rice leaf disease is necessary to improve rice productivity. Currently, many researchers use deep learning methods to solve this problem. Unfortunately, their research results were less accurate. In this paper, we construct transfer learning models to diagnose and categorize illnesses affecting rice leaves. To further improve the model performance, we construct three ensemble learning models to combine various architectures. In order to bring transparency to the disease diagnostic process, we explore the explainable AI (XAI) problem of the visual object detector and integrate Gradient-weighted Class Activation Mapping (Grad-CAM) into three ensemble models to generate explanations for individual object detections for assessing performance. The results of Ensemble Learning indicate that merging different architectures can be effective in disease diagnosis, as evidenced by their best accuracy of 99.78% which is better than other state-of-the-art works. This research demonstrates that the integration of deep learning and transfer learning models yields improved prediction interpretability and classification accuracy of rice leaf disease. So, we established a dependable method of deep, transfer, and ensemble learning for the diagnosis of diseases affecting rice leaves.**

***Keywords*—Rice leaf; ensemble-learning; explainable AI; disease diagnosis; transfer learning**

I. Introduction

Over half of the world's population depends on rice production as their primary staple, making it essential for global food security [1], [2], [3]. Because it sustains livelihoods, particularly in Asia where it is a significant source of employment and revenue, it has a significant impact on people [4], [5]. Efficient rice cultivation has a crucial role in maintaining and enhancing lives globally by ensuring food stability, economic well-being, and socio-economic development [6].

Finding rice leaf disease is a vital job in agriculture since rice is one of the most important crops in the world and feeds millions of people as a staple diet. Brown spot, leaf blast, and bacterial blight are just a few of the diseases that can severely limit rice crop productivity, resulting in financial losses and food poverty. By using targeted treatments or preventive measures, farmers can take appropriate corrective action to preserve crop health and maximize production. Early and precise detection of these diseases using rice leaf images can assist farmers in doing just that. The scalability and effectiveness of image analysis in rice leaf disease detection make it a valuable tool. Conventional approaches to disease identification are laborious, arbitrary, and prone to mistakes as they frequently depend on the manual inspection of specialists. However, a quick, reliable, and scalable solution is provided by automated detection that makes use of computer vision and machine learning algorithms. Advanced models have demonstrated significant potential in accurately classifying rice leaf illnesses from images. These models include ensemble learning models (e.g. combining VGG16, ResNet50, InceptionV3, and EfficientNet) and deep learning architectures (e.g. CNNs) [7], [8]. By including Explainable AI methods such as Grad-CAM, these models become more visible and users are able to see which leaf portions are responsible for the predictions made by the model [9]. This guarantees that academics and farmers alike can rely on the technology and obtain practical knowledge about the well-being of rice crops, which in turn promotes improved disease control and more environmentally friendly farming methods [10], [11].

By including Explainable AI methods such as Grad-CAM, these models become more visible and users are able to see which leaf portions are responsible for the predictions made by the model. This guarantees that academics and farmers alike can rely on the technology and obtain practical knowledge about the well-being of rice crops, which in turn promotes improved disease control and more environmentally friendly farming methods [12].

The main goal of rice leaf disease identification using rice leaf images is to create an automated, precise, and effective system that can recognize and categorize illnesses that impact

rice harvests. Reduced reliance on costly, time-consuming, and error-prone manual expert inspection is the aim of utilizing sophisticated image processing and machine learning approaches [13], [14], [15]. By early detection of illnesses including bacterial blight, brown spot, and leaf blasts, this method hopes to give farmers the information they need to take preventative measures that can stop the disease's progress and lessen crop loss. Another goal is to provide a scalable solution that can be implemented at various agricultural scales and geographical locations to increase the overall productivity and sustainability of rice growing. The system employs deep learning models [16], [17], [18], such as CNNs, in conjunction with ensemble learning techniques that integrate the capabilities of models like VGG16, ResNet50, and InceptionV3, to achieve a high degree of disease detection accuracy while maintaining resilience in a range of environmental circumstances and image characteristics. The goal also includes improving these models' interpretability by using Explainable AI techniques like Grad-CAM, which let users see the specific regions of the rice leaf that the algorithm targeted for prediction. This openness strengthens farmer confidence in the system and gives them a greater knowledge of the health of their crop [19], which in turn improves disease control procedures and promotes more sustainable agricultural results.

The key contributions of this paper are as follows:

1 We explored the disease of rice leaf images from two datasets using separate deep and transfer learning models.
2 We customized and applied three ensemble learning techniques from deep and transfer learning.
3 We proposed an ensemble learning model with the highest accuracy and lowest data loss rate.
4 We used Explainable AI to evaluate the input image and with the proposed ensemble learning.
5 Finally, we proposed new algorithms for generated ensemble learning algorithms.

The remaining parts of this paper are formatted as follows: Section II explains the previous studies of existing work that were published recently. While Section III describes the research technique used in this work, Section IV provides the experimental results of this work. Section V demonstrates the conclusion and future work of this study.

II. Prior Studies

Many researchers have published solutions for rice leaf disease diagnosis using different types of algorithms such as deep learning, machine learning, ensemble learning, etc. Some recent publications are mentioned here.

S. Ghosal and K. Sarkar et al. [20], proposed a VGG-16-based CNN architecture to detect different rice leaf diseases accurately and they used their own collected dataset containing about 500 images and achieved 92.46% of accuracy. In the paper et al.[21], authors applied several image processing techniques such as RGB to HSV conversion, background subtraction, segmentation, etc., and then implemented an automated system using a deep neural network for rice leaf disease detection and achieved an average accuracy of 92% using their self-collected dataset of 209 images. J. Chen, D. Zhang, Y. A. Nanehkaran, and D. Li et al. [22] proposed a system for the detection of various rice leaf diseases combining DenseNet pre-trained with ImageNet and Inception module on an image dataset collected by Fujian Institute of Subtropical Botany, Xiamen, China and achieved an outstanding accuracy of no less than 94.07% for each type of disease category.

M. A. Islam et al. [23] worked with four types of paddy disease to detect it early and accurately. They applied several deep learning CNN models such as VGG-19, Inception-Resnet-V2, ResNet-101, and Xception and their experimental result shows that Inception-Resnet-V2 performed better with 92.68% accuracy. Several image processing techniques [24] were applied by the authors to extract important features from images that describe the most significant characteristics and then classified the images as rice leaf disease category using XGBoost and SVM algorithms and got about 86.58% accuracy. They created their own dataset for the experiment and used a public dataset for testing. They extract features from images using several image processing techniques [25] and applied CNN models VGG16, ResNet50, and DenseNet121 to detect rice leaf disease accurately and 91.63% accuracy is achieved by the model DenseNet121.

To detect rice leaf disease automatically considering various leaf sizes author applied deep learning-based CNN model ResNet and YOLOv4 [26] on a public dataset of 4960 images and YOLOv4 models show better performance with mAP value 91.14%. In the paper et al. [27], the author applied various CNN techniques for rice leaf disease detection such as VGG16, VGG19, Xception, ResNet, and the 5-layer convolution model, and finally, it is shown that the 5-layer convolution model achieved the highest accuracy 78.2% in compare to other models. CNN models DenseNet121, DenseNet169, MobileNetV2, and VGG16 are employed [28] on the public dataset from Kaggle containing the 5932 images for rice leaf disease detection, and DenseNet169 and moboleNetV2 show the highest performance with 94.30% accuracy. In the paper et al. [29] collected 1500 images from Feni, Bangladesh to detect rice leaf disease and applied CNN model YOLOv5 and achieved 76% mAP value.

A new machine learning approach Nu-SVM model is employed [30] on a Kaggle dataset to detect rice leaf disease and the experimental result shows 52.12% to 53.81% of validation accuracy. In the paper et al. [31], the author employed various filter-based feature transformation techniques for rice leaf disease detection accurately. They used a public rice leaf dataset from Kaggle and it showed that in the experiment the KNN model achieved the highest balance accuracy of 90%. DNet-SVM: XAI is proposed [32] by the authors to detect sugarcane disease detection and they used a public sugarcane dataset from kaggle. DNet-SVM: XAI detects and predicts sugarcane disease early and explains its prediction reason. They also applied another deep learning model such as VGG16, VGG19, and Inception, etc., and compared the result with the proposed model. The proposed model outperformed in comparison to other models. Rice crop disease detection is very important and to detect rice crop disease early author applied the CNN model for detection and LIME to explain its interpretability [33]. The experimental result shows that the proposed models achieved 91.60% accuracy.

Deep learning models VGG16, SqueezeNet, and InceptionV3 were employed in [34] for rice leaf disease detection

and the proposed model SqueezeNet achieved the highest accuracy of 93%. P. Kulkarni and S. Shastri at el. [35] proposed a novel deep learning-based CNN model that is applied for the automatic detection of rice leaf disease. They used a public dataset from Kaggle and the author achieved about 95% accuracy in the experiment. In the classification of corn leaf diseases, the proposed VGG16 model augmented with LRP [36] performed better than earlier cutting-edge models. The outcomes of the simulation showed that the model not only produced findings with a high degree of accuracy but also highlighted important areas in the images that were classified. The authors introduce an improved YOLOv8 [37] that combines EIoU loss and α-IoU loss to replace the original Box Loss function and enhance the rice leaf disease detection system's performance. Finally, they compare YOLOv8 performance with YOLOv5 and YOLOv7 and it is shown that their proposed model performed better with 89.90% accuracy.

The limitation of previous studies was less accurate. In most of the publications, they used pretrained models of deep learning and transfer learning to detect and classify rice leaf disease. The use of Explainable AI was rare.

We employed the "Rice Leaf Disease Detection" dataset from Kaggle and some effective deep learning methods such as CNN, VGG-16, and InceptionV3 in our suggested work shown in Table I, which exhibits higher accuracy than earlier studies. Machine learning, deep learning, and advanced image processing techniques have greatly expanded the field of rice leaf disease detection and recognition. These advancements improve prompt intervention and detection accuracy. However, issues like environmental unpredictability and dataset restrictions still exist. In order to enhance sustainable rice cultivation and further improve detection systems, future research should focus on these concerns.

III. Proposed Framework and System Architecture

The proposed working flow diagram of rice leaf disease diagnosis and classification is illustrated in Fig. 1 where data collection to result in findings is described sequentially.

A. Dataset

For experimental implementation, we collected image datasets from different types of online sources. We used two datasets from online. One is "Rice Leafs Disease Dataset"[1]. This data has a total of two different directories as training and validation with 6 classes. The total images contain 2,627 in six classes. The classes are Bacterial Leaf Blight, Brown Spot, Healthy, Leaf Blast, Leaf Scald, and Narrow Brown Spot. Each class has 350 images for training and 88 images for validation.

Another dataset name is "Rice Leaf Diseases Detection"[2]. This dataset is released at the beginning of 2024. This dataset is also divided into two divisions training and validation. Each directory has a total of 10 classes such as bacterial-leaf-blight, brown-spot, healthy, leaf-blast, leaf-scald, narrow-brown-spot, neck-blast, rice-hispa, sheath-blight, and tungro. Each class has a total of 1,385 images for training and 350 for validation. The total dataset has 17, 350 images for both training and validation. Fig. 2 and 3 describe the sample dataset for two different types of images.

[1]https://www.kaggle.com/datasets/dedeikhsandwisaputra/rice-leafs-disease-dataset

[2]https://www.kaggle.com/datasets/loki4514/rice-leaf-diseases-detection

B. Feature Extraction and Image Processing

To remove noise, reduce dimensions, and make it suitable for model training, we change the shape of the images into 224*224 dimensions after converting the grayscale. To extract the features of the images, the Lanczos interpolation method. We later normalized the images by dividing 255.0 between the pixel values of 0 and 1.

Basically, two feature extraction methods were used here [38]. These are: (1) Shaped based technique, and (2) Transform based Technique.

The dataset is labeled using One-hot encoding techniques for each class in two datasets. Data labeling makes it easy to train and validate the model. Based on the unique class name, the images are labeled with a one-hot encoding method in this study [39].

C. Image Preprocessing

As the preprocessing part of the data, we made changes to the images as:

1. Resize of the images into 224*224
2. Set batch size=32 and classmode='categorical' for the multi-class classification.
3. Array conversion with the help of Numpy.
4. Input shape is 224*224*3.

Such preprocessed images are shown in Fig. 4; where six images from six classes are combined into a single form [40].

D. Data Augmentation

We augmented the dataset 1 to increase the number of images [41], [42]. The increased amount of data will increase the detection and training accuracy of the model [43]. If the dataset is vast then augmentation is not needed but dataset 1, has only 2,627 images. That's why, we used the augmentation method for dataset increasing. We set the parameter details in the augmentation part as follows:

rescale=1.0/255.0,

horizontal-flip=True,

zoom-range=0.2,

shear-range=0.2

But dataset-2 has around 17,350 image data. This amount is enough for the model training. However, dataset-2 has satisfactory data, so, it does not need to augment this data [44]. We ignored to augment. The augmented images for dataset 1 are shown in Fig. 5, where six individual images from six classes are merged into a single image.

TABLE I. COMPARATIVE ANALYSIS BETWEEN EXISTING WORK WITH PROPOSED WORK

Reference	Dataset	Used Methods	Accuracy	XAI
[20]	Private (500)	CNN, VGG-16	92.46%	No
[21]	Private (209)	DNN, KNN	Avg. 92.6%	No
[22]	Private (500) Fujian Institute of Subtropical Botany, Xiamen, China	DENS-INCEP, VGGNet-19, ResNet-50, DenseNet-201, InceptionV3, VGG19-SVM	94.07%	No
[24]	UCI rice leaf disease dataset	XGBoost, SVM	86.58%	No
[25]	Private (386)	VGG16, ResNet50, DenseNet121	91.63%	No
[26]	Rice Leaf Dataset (4960)	ResNet, YOLOv4	—	No
[27]	Kaggle Dataset (1600)	VGG16, VGG19, Xception, ResNet, 5-layer Convolution	78.20%	No
[28]	Kaggle Dataset (5932)	DenseNet121, DenseNet169, MobileNetV2, VGG16	74.30%	No
[29]	Private Dataset (1500)	YOLOv5	—	No
[30]	Rice Leafs Dataset from Kaggle	Nu-SVM	53.81%	No
[31]	Kaggle Dataset	RFC, KNN, LDA, HGBC etc.	90%	No
[32]	Sugarcane (14000) from Kaggle	DNet-SVM, VGG16,VGG19, ResNet, Inception, DenseNet, DNet-SVM	94%	Yes
[33]	Kaggle Dataset	CNN and LIME	90.60%	Yes
[34]	Rice Leaf Dataset	VGG16, SqueezeNet, and InceptionV3	93%	No
[35]	Kaggle Dataset	CNN	95%	No
[36]	Corn Leaf Dataset (4188)	VGG16, LRP	94%	Yes
[37]	Private Dataset (1634)	YOLOv8	89.90%	No
Proposed Work	Rice Leaf Diseases Diagnosis	Deep-transfer learning ensembles	99.78%%	Yes

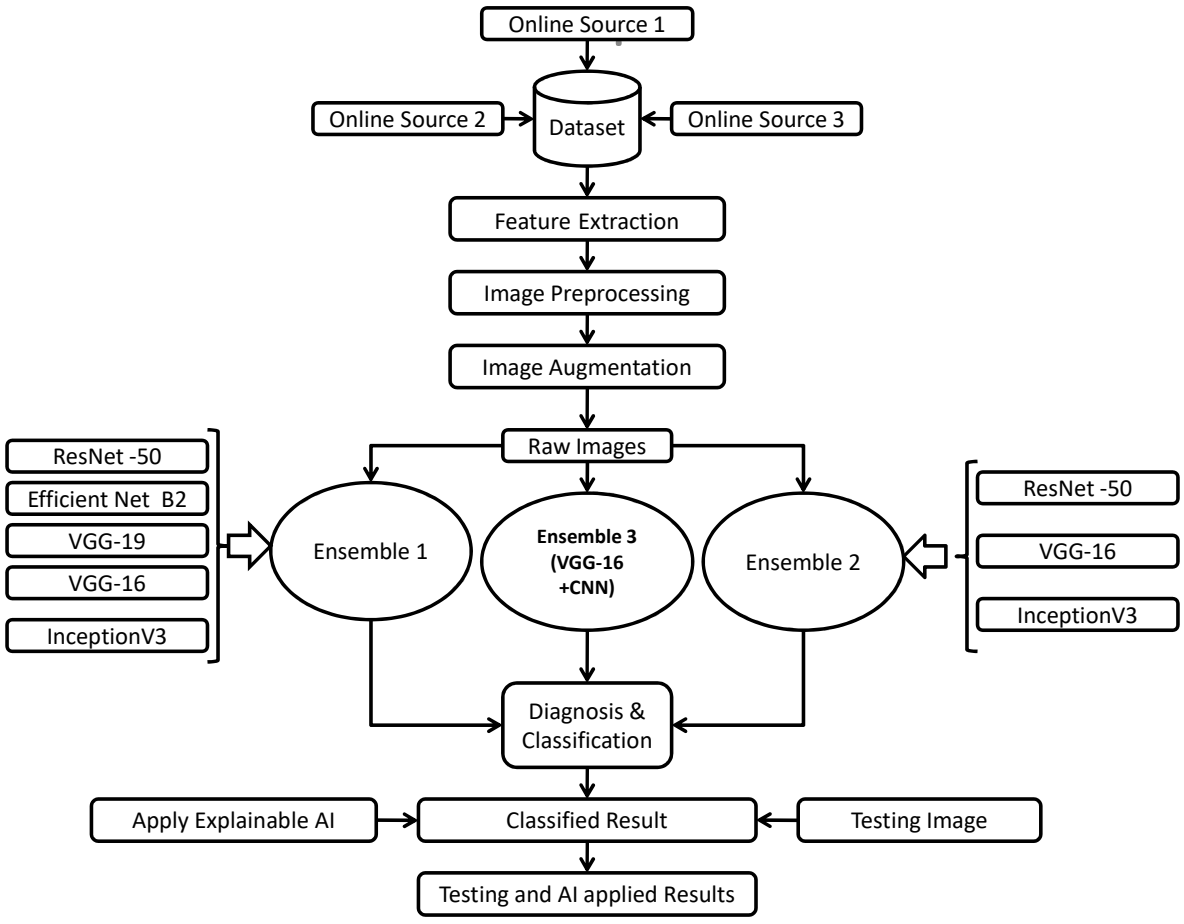


Fig. 1. Proposed system architecture.

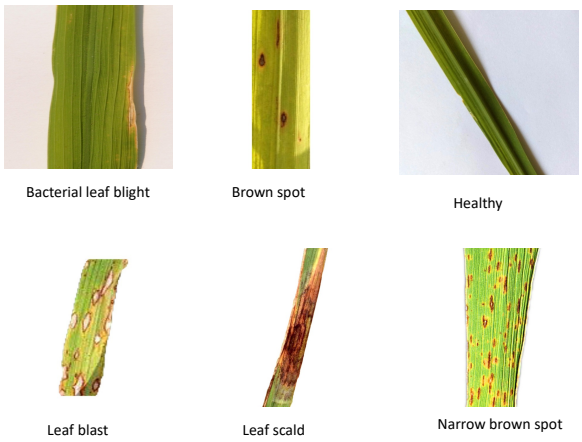


Fig. 2. Dataset-1 sample with six classes.

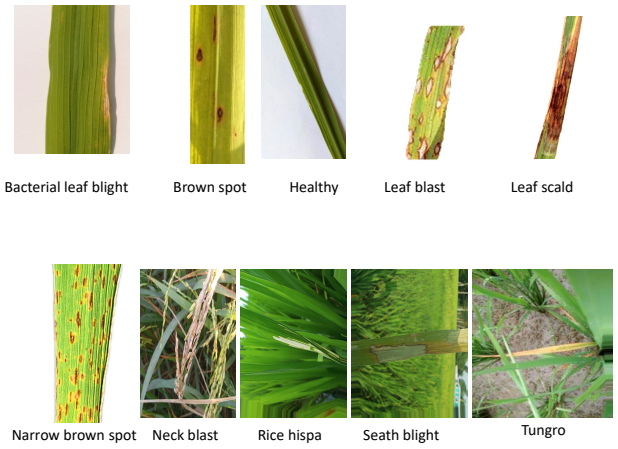


Fig. 3. Dataset-2 sample with 10 classes.

E. Feature Extraction Based on Deep Learning

For model training and validation, we used a deep learning model CNN in this study. This model is suitable for image classification. Preprocessed and augmented images were trained and validated by this model. CNN model is applied to two datasets separately. It is constructed with an input layer, a fully connected layer, some convolutional layers, and max pooling layers [45]. The input layer takes images as input, convolutional layers use a 3*3 filter or kernel for image filtering. The Max pooling layer receives the output from the convolutional layer and processes it. After processing, the output will go through the fully connected layer. The fully connected layer combines the output from the max-pooling layer and makes a single form in the Dense layer. Dense layers flatten the previous layers for making a single form or it is used for combining [46], [16]. The basic organization of the CNN architecture of rice leaf image classification is shown in Fig. 6. CNN model parameter details are shown in Table II.

TABLE II. CNN Model Parameter Details in Every Layer

Layer (type)	Output Shape	Parameter
input-1 (InputLayer)	[(None, 224, 224, 3)]	0
block1-conv1(Conv2D)	(None, 224, 224, 64)	1792
block1-conv2(Conv2D)	(None, 224, 224, 64)	36928
block1-pool(MaxPooling2D)	(None, 112, 112, 64)	0
block2-conv1(Conv2D)	(None, 112, 112, 128)	73856
block2-conv2(Conv2D)	(None, 112, 112, 128)	147584
block2-pool(MaxPooling2D)	(None, 56, 56, 128)	0
block3-conv1(Conv2D)	(None, 56, 56, 256)	295168
block3-conv2(Conv2D)	(None, 56, 56, 256)	590080
block3-conv3(Conv2D)	(None, 56, 56, 256)	590080
block3-pool(MaxPooling2D)	(None, 28, 28, 256)	0
block4-conv1(Conv2D)	(None, 28, 28, 512)	1180160
block4-conv2(Conv2D)	(None, 28, 28, 512)	2359808
block4-conv3(Conv2D)	(None, 28, 28, 512)	2359808
block4-pool(MaxPooling2D)	(None, 14, 14, 512)	0
block5-conv1(Conv2D)	(None, 14, 14, 512)	2359808
block5-conv2(Conv2D)	(None, 14, 14, 512)	2359808
block5-conv3(Conv2D)	(None, 14, 14, 512)	2359808
block5-pool(MaxPooling2D)	(None, 7, 7, 512)	0
flatten(Flatten)	(None, 25088)	0
dense (Dense)	(None, 2)	50178
Total params:	0	14,764,866
Trainable params:	0	50,178
Non-trainable params:	0	14,714,688

F. Feature Extraction Based on Transfer Learning

To implement the proposed work, we used total five transfer learning algorithms such as VGG16, VGG19, ResNet-50, InceptionV3, and EfficientNetV2-M. These models apply to two different datasets separately. For vast datasets, transfer learning models are suitable for the best accurate training and validation. Most of the algorithms work are similar way. However, we described them separately below.

1) VGG16: It is the variants of the deep learning model and updated version of the CNN model. This model has basic 16 layers. That is why, it is called VGG16. It consists of some layers such as the Input layer, Convolution layers, Max-pooling layer, Dense layer, and Output layer [47]. A 3*3 laplacian mask was applied here. The image input shape is 224*224 with RGB Channel. Then input image is processed by convolution and max pooling layer [48], [49]. Finally, layers are combined into a single layer named as dense layer. It produces the final output of the model. During trainable. all layers are frozen. The activation function is "softmax" used here [50]. The parameters for the VGG16 is set as:

optimizer='rmsprop',

loss='categorical-crossentropy',

metrics=['accuracy']

The summary parameter details of VGG16 is illustrated in Table III.

2) VGG19: VGG19 is also a transfer learning model, used for classification. It has basic 19 layers. So, it is called, VGG19 [51]. It can classify a total of 1,000 classes of objects. So, this model is highly applicable to vast datasets. It is the incremental version of VGG16 [52], [53]. We used two separate datasets. The first dataset has 2,627 images. The second dataset has 10 image classes and around 17,350 image data. So, we used

Fig. 4. Preprocessed image sample.

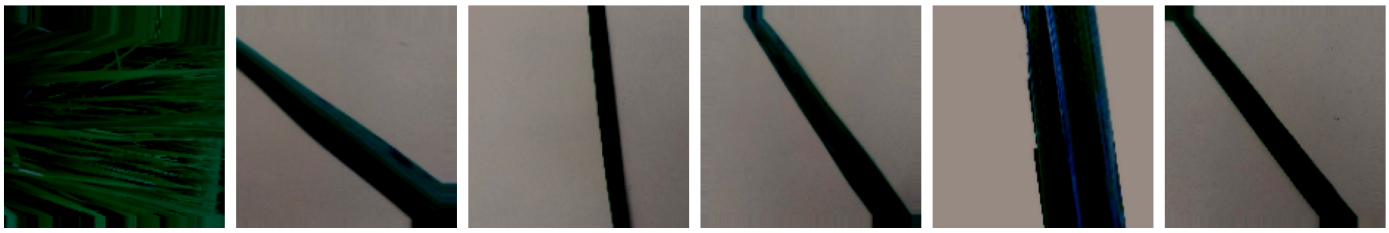


Fig. 5. Augmented images for dataset 1.

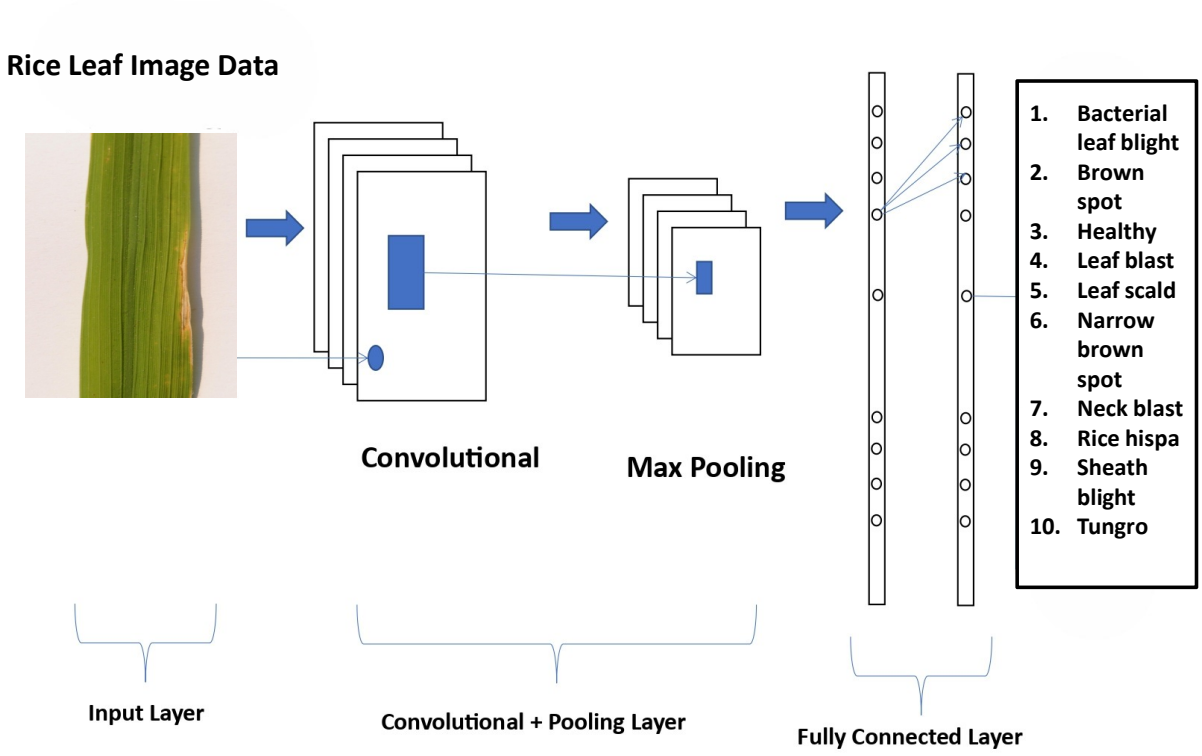


Fig. 6. CNN Architecture for rice leaf disease diagnosis.

TABLE III. VGG16 MODEL PARAMETER DETAILS

Layer (type)	Output Shape	Parameter
vgg16	(Functional) (None, 7, 7, 512)	14714688
flatten-1 (Flatten)	(None, 25088)	0
dense-2 (Dense)	(None, 512)	12845568
dropout-1 (Dropout)	(None, 512)	0
dense-3 (Dense)	(None, 10)	5130
Total params:	0	27565386
Trainable params:	0	12850698)
Non-trainable params:	0	14714688

it to get the highest accuracy. This model can classify the object with high accuracy. It takes the input image as 224*224 in standard format. This model consists of some layers. The layers are Convolution, Max pooling, fully connected, input, and dense layer. The dense layer is used to flatten the previous layers and it provides the output image [54]. The parameter details of VGG19 are shown in Table IV.

TABLE IV. VGG19 MODEL PARAMETER DETAILS IN EVERY LAYER

Layer (type)	Output Shape	Parameter
input-2 (InputLayer)	[(None, 224, 224, 3)]	0
block1-conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1-conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1-pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2-conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2-conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2-pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3-conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3-conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3-conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3-conv4 (Conv2D)	(None, 56, 56, 256)	590080
block3-pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4-conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4-conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4-conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4-conv4 (Conv2D)	(None, 28, 28, 512)	2359808
block4-pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5-conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5-conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5-conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block5-conv4 (Conv2D)	(None, 14, 14, 512)	2359808
block5-pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten-1 (Flatten)	(None, 25088)	0
dense-1 (Dense)	(None, 6)	150534
Total params:	0	20174918
Trainable params:	0	150534
Non-trainable params:	0	20024384

3) InceptionV3: This model is another powerful transfer learning model for image data classification. There are eleven inception modules, two max-pooling layers, five convolutional layers, one average pooling layer, one fully connected layer, and one max-pooling layer in InceptionV3[55]. We used this model for rice leaf disease diagnosis and classification. For two separate datasets, we applied this model. Though it takes more time than other models, it can be classified accurately than other models. The basic parameter details of this algorithm are lengthy. So, we are ignoring the parameter details. It has a total of 48 layers [56].

4) ResNet-50: A deep convolutional neural network (CNN) architecture with 50 layers, ResNet-50 is intended for computer vision and image recognition applications. Recursive learning, also known as "skip connections", was presented, in which the network learns residuals, or the variations between input and output layers [57]. This facilitates more effective training of deeper models by addressing the vanishing gradient issue that arises in very deep networks. Convolutional, pooling, and fully connected layers are the building blocks of ResNet-50, which are arranged in residual blocks [58]. It has achieved state-of-the-art results in several vision tasks and has been widely utilized for transfer learning, having been trained on big datasets such as ImageNet. Reputably, the model strikes a balance between computational efficiency and depth. We used this model for two separate datasets in rice leaf image classification.

5) EfficientNetV2-M: EfficientNetV2-M is a powerful transfer learning algorithm, used for large-scale image data. A member of the EfficientNet family, EfficientNetV2-M is renowned for striking an ideal balance in image recognition tasks between computational efficiency and model performance. By adopting a more sophisticated scaling technique and methodically increasing breadth, depth, and resolution, it improves accuracy and speed over the original EfficientNet. Depthwise separable convolutions and more sophisticated methods such as Fused-MBConv are combined by EfficientNetV2-M to minimize computing expenses without sacrificing precision. Large-scale picture classification tasks are especially well-suited for this model, which offers shorter training periods and smaller model sizes than its predecessors. It is adaptable for a range of computer vision applications because it has been pretrained on big datasets like. However, to classify the rice leaf images, we used this model for two separate datasets. Due to the lengthy of layers, we are ignoring the parameter details of this model [59].

To implement the working process and evaluate the performance, we proposed a new algorithm followed in Algorithm 1.

G. Ensemble Learning

Using the pretrained deep and transfer learning models, we got the disease detection accuracy to be more than 95% but not reach 99.99%. We tried multiple ensembling techniques because it was unknown to us which ensemble would be the proper model for this type of work. So, we used three types of ensemble methods [60], [51]. To improve the training and validation accuracy by more than 95%, sometimes we used some models in a single structure known as ensemble learning. Bagging and Boosting are the commonly used ensemble techniques for image classification [61]. The main purpose of ensemble learning is:

1 improving the training and validation accuracy,
2 decrease the data loss amount,
3 find out the optimum solution of time and space complexity,
4 find out the proper detection, prediction, and diagnosis,
5 increase the system speed, etc.

However, in this study, we ensembled deep learning and transfer learning to find out the above requirements. We made three ensembles from six separate models such as CNN, VGG16, VGG19, InceptionV3, ResNet-50, EfficientNetV2-M [62]. Three ensembles are described below.

Algorithm 1 Proposed Algorithm for Rice Leaf Disease Detection

#Rice Leaf Images used as Dataset
$DataX : Training, Validation, Testing$
$InputImage \leftarrow (256, 64, 3)$
$Algorithms :\leftarrow CNN, InceptionV3, ResNet50,$
$VGG16, VGG19, EfficientNetV2M$
$ITR \leftarrow NumberofIterations$
$ACC \leftarrow Performancematrix$
$ALG \leftarrow Numberofalgorithms$
#Feature Extraction using Deep and Transfer Learning Algorithms
$ML \leftarrow ModelLayer$
$M \leftarrow Epochs$
#Deep and Transfer Learning Model Training:
for 1 to M **do**
  for each ML **do**
    for each Sample in X **do**
      calculate A from X by Conv. Process
    end for
  end for
end for
Ensemble Learning Call
$E1 \leftarrow Ensemblelearning1$
$E2 \leftarrow Ensemblelearning2$
$E3 \leftarrow Ensemblelearnin3$
#Number of Iteration
for 1 to ITR **do**
  1. Train Model with N Number of batch size
  2. Feature Extraction through hidden layers
  3. Forward propagations
  4. Backword propagation for updating weights
  5. Model validation with validation data to check overfitting
end for
Model Evaluation:
1. Evaluate the model with test data
2. Store the model performance in Acc variable

Algorithm 2 Proposed Algorithm for Ensemble Learning-1

$ITR \leftarrow NumberofIterations$
$ACC \leftarrow Performancematrix$
$ALG \leftarrow Numberofalgorithms$
#Ensemble Models
$E1 \leftarrow FusionofALG$
$ALG \leftarrow ResNet50, EfficienNetV2M, InceptionV3,$
$VGG16, VGG19$
for 1 to ITR **do**
  1. Preprocess the input features of fusion model
  2. Customize the layers in the model
  3. Freezing the base model
  4. Optimize the parameters in the model
  5. Train the model by dataset
  6. Model validation with validation data to check overfitting.
end for

#Number of Iteration
for 1 to ITR **do**
  1. Train Model with N Number of batch size
  2. Feature Extraction through hidden layers
  3. Forward propagations
  4. Backword propagation for updating weights
  5. Model validation with validation data to check overfitting
end for

1) Ensemble Learning-1: Ensemble Learning-1 is developed from transfer learning models. To make this, we used 5 transfer learning models [63]. They are VGG16, VGG19, InceptionV3, ResNet-50, and EfficientNetV2-M. After bagging them, we got a new model named "Ensemble learning-1". Though it requires more time than other single models, it can classify and diagnose the disease more correctly than other normal models [64], [65]. The basic structure of the newly developed ensemble model-1 is illustrated in Fig. 7 and the parameter list is shown in Table V. Algorithm 2 is the proposed new algorithm for ensemble learning-1 used in this study. It is developed based on the fusion of VGG16, VGG19, ResNet-50, InceptionV3, and EfficientNetV2-M [66], [67].

2) Ensemble Learning-2: Another transfer learning combination of the three models is developed into a single structure known as "ensemble learning-2". We made it, after voting the VGG16, Inception V3, and ResNet-50 [68]. This structure is simpler than Ensemble learning-1. Also, it is suitable for high accuracy of the model training and validation. Though its parameter list is vast, hence it can classify the image data and can detect the diagnosis[69]. The basic structure of ensemble-2 is described in Fig. 8 and the parameter list is shown in Table VI. Algorithm-3 describes the combined algorithm proposed for this model. It is the appropriate model in this structure.

3) Ensemble Learning-3: Ensemble learning-3 is developed from one deep learning and one transfer learning model. By CNN and VGG16 combinations, ensemble-3 is made [70]. It saves memory and space complexity also. Its structure is simple and easy to implement [71], [72]. The basic structure is shown in Fig. 9 and the parameter list is also shown in Table VII. Algorithm 4 explained the basic working process of this fusion model from deep learning and transfer learning. We proposed the algorithm in this stage for rice leaf disease diagnosis at high accuracy and it is more effective now [73].

IV. Result Analysis and Discussion

In this section, we will discuss different types of algorithm performance applied in our study. Particularly, we will explain training, validation, and testing accuracy for each model as well as loss. Graphical representation also will be described here such as Plot details, Curves [74], Confusion Matrix, Classification report, Correctly classified and misclassified images, AI-based testing image report, etc. After all, a comparison of

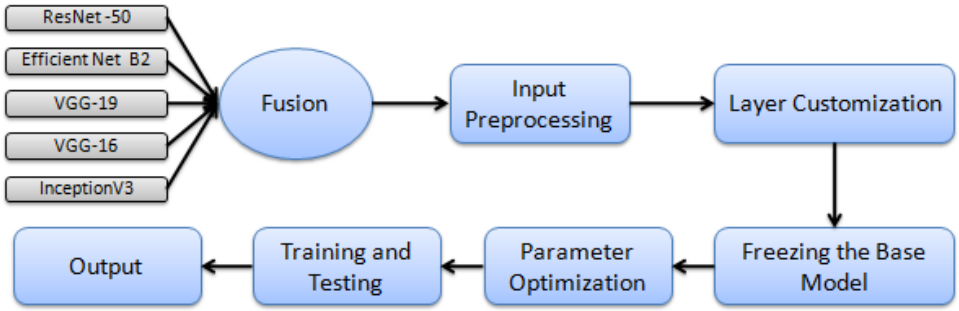


Fig. 7. Basic Architecture of proposed ensemble learning-1 model for rice leaf disease diagnosis.

TABLE V. ENSEMBLE LEARNING-1 PARAMETER DETAILS

Layer (type)	Output Shape	Parameter	Connected to
input-2 (InputLayer)	[(None, 224, 224, 3)]	0	[]
preprocess-vgg16 (Lambda)	(None, 224, 224, 3)	0	['input-12[0][0]']
preprocess-inception (Lambda)	(None, 224, 224, 3)	0	['input-12[0][0]']
preprocess-resnet (Lambda)	(None, 224, 224, 3)	0	['input-12[0][0]']
preprocess-vgg19 (Lambda)	(None, 224, 224, 3)	0	['input-12[0][0]']
preprocess-efficientnet (Lambda)	(None, 224, 224, 3)	0	['input-12[0][0]']
vgg16 (Functional)	(None, 512)	14714688	['preprocess-vgg16[0][0]']
inception-v3 (Functional)	(None, 2048)	21802784	['preprocess-inception[0][0]']
resnet50 (Functional)	(None, 2048)	23587712	['preprocess-resnet[0][0]']
vgg19 (Functional)	(None, 512)	20024384	['preprocess-vgg19[0][0]']
efficientnetv2-m (Functional)	(None, 1280)	53150388	['preprocess-efficientnet[0][0]']
concatenate-features (Concatenate)	(None, 6400)	0	[five-models [0][0]]
dense-1 (Dense)	(None, 1024)	6554624	['concatenate-features[0][0]']
dropout-1 (Dropout)	(None, 1024)	0	['dense-1[0][0]']
dense-2 (Dense)	(None, 512)	524800	['dropout-1[0][0]']
dropout-2 (Dropout)	(None, 512)	0	['dense-2[0][0]']
output-layer (Dense)	(None, 6)	3078	['dropout-2[0][0]']
Total params:	140362458	0	0
Trainable params:	7082502	0	0
Non-trainable params:	133279956	0	0

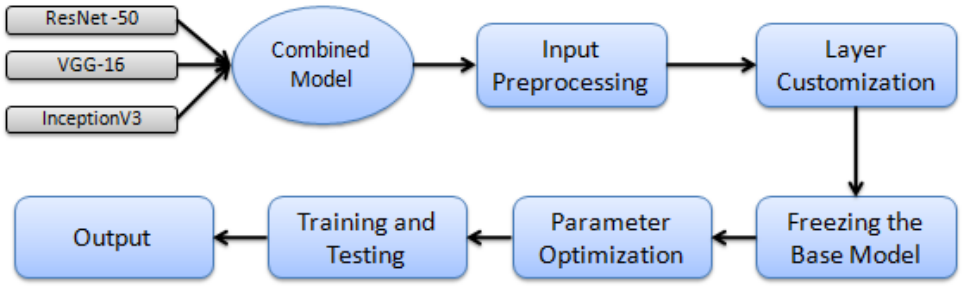


Fig. 8. Basic Architecture of proposed ensemble learning-2 model for rice leaf disease diagnosis.

each model for two datasets will be added for analysis [75], and a comparison table will be created for recently published existing work with the proposed work [76].

A. Transfer Learning Models Performance

In the proposed study, we used five transfer learning models such as VGG16, VGG19, InceptionV3, ResNet-50, and EfficientNetV2-M. The performance of these models is explained below.

1) VGG16: VGG16 is applied for the diagnosis of rice leaf disease from images. However, its structure is not easy and it needs more time due to the deep layers. We got accuracy for training, validation, and testing are 98.94%, 93.97%, and 92.70%, respectively. The training and validation loss amounts were 7.69% and 12.16%, respectively. The training, validation, testing, and loss curves for epochs 50 are shown in Fig. 10. The confusion matrix and classification report are illustrated in Fig. 11 and Table VIII, respectively. The classification report and confusion matrix are generated for 50 epochs and we used two datasets with 0.001 learning rate. To analyze and see the details we just explain for one dataset. For both datasets, we mention the accuracy in the Table XVII below. The ROC curve and Precision-recall curve for dataset-1 are illustrated in Fig.

TABLE VI. ENSEMBLE LEARNING-2 PARAMETER DETAILS

Layer (type)	Output Shape	Parameter	Connected to
input-layer-31 (Input-Layer)	(None, 224, 224,3)	0	-
preprocess-vgg16 (Lambda)	(None, 224, 224,3)	0	input-layer-31[0.....]
preprocess-inceptionv3 (Lambda)	(None, 224, 224,3)	0	input-layer-31[0.....]
preprocess-resnet (Lambda)	(None, 224, 224, 3)	0	input-layer-31[0....]
vgg16 (Functional)	(None, 512)	14,714,688	preprocess-vgg16
inceptionv3 (Functional)	(None, 2048)	21,802,784	preprocess-inceptionv3
resnet50 (Functional)	(None, 2048)	23,587,712	preprocess-resnet
concatenate-features (concatenate)	(None, 4608)	0	vgg16[0][0], inceptionv3[0][0...], resnet50[0][0]
dense-1 (Dense)	(None, 1024)	4,719,616	concatenate-features
dropout-1 (Dropout)	(None, 1024)	0	dense-1[0][0]
dense-2 (Dense)	(None, 512)	524,800	dropout-1[0][0]
dropout-2 (Dropout)	(None, 512)	0	dense-2[0][0]
output-layer (Dense)	(None, 6)	3,078	dropout-2[0][0]
Total params:	65,352,678	0	0
Trainable params:	5,247,494	0	0
Non-trainable params:	60,105,184	0	0

TABLE VII. ENSEMBLE LEARNING-3 PARAMETER DETAILS

Layer (type)	Output Shape	Parameter	Connected to
input-9 (InputLayer)	[(None, 224, 224, 3)]	0	[]
input-10 (InputLayer)	[(None, 224, 224, 3)]	0	[]
sequential-3 (Sequential)	(None, 36864)	388416	['input-9[0][0]']
sequential-4 (Sequential)	(None, 25088)	14714688	['input-10[0][0]']
concatenate-2 (Concatenate)	(None, 61952)	0	['sequential-3[0][0]', 'sequential-4[0][0]']
dense (Dense)	(None, 512)	31719936	['concatenate-2[0][0]']
dropout (Dropout)	(None, 512)	0	['dense[0][0]']
dense-1 (Dense)	(None, 6)	3078	['dropout[0][0]']
Total params:	46826118	0	0
Trainable params:	32111430	0	0
Non-trainable params:	14714688	0	0

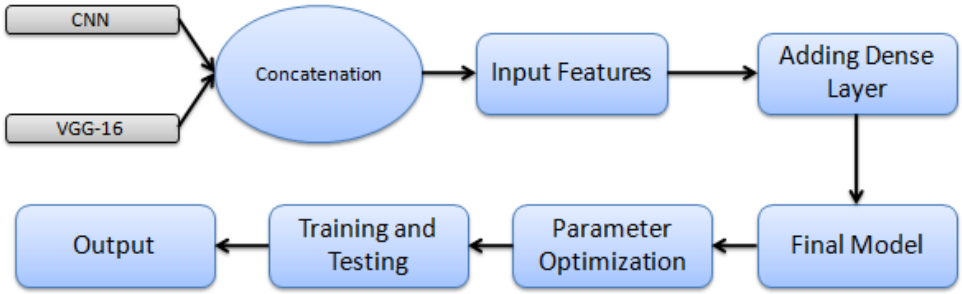


Fig. 9. Basic Architecture of proposed ensemble learning-3 model for rice leaf disease diagnosis.

12 and 13, respectively. It is generated for six class dataset. For the epochs, 50 VGG16 generated these curves.

TABLE VIII. CLASSIFICATION REPORT OF VGG16 MODEL

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	0.96	1.00	0.98	53
brown-spot	0.93	0.86	0.89	63
healthy	0.87	0.91	0.89	44
leaf-blast	0.82	0.85	0.84	55
leaf-scald	0.98	1.00	0.99	42
narrow-brown-spot	1.00	0.97	0.98	58
accuracy	0	0	0.93	315
macro avg	0.93	0.93	0.93	315
weighted avg	0.93	0.93	0.93	315

Though the applied VGG16 model produces around 98.94% training accuracy, it also has some amount of loss. In some cases, it is misclassified and does not properly diagnose the disease. This misclassified rate is rare. However, it was not 100% perfect. But in most of the cases, the model was correctly classified. Some misclassified and correctly classified images are shown in Fig. 14 and 15, respectively. We will ignore the misclassified and correctly classified images for other used models due to vast images and page length.

2) VGG19: VGG19 is used for rice leaf disease diagnosis and classification. Though it requires more time for completion, it can classify and detect the diagnosis more correctly. It has more layers in structure, so it needs time to run. However, after applying this model we got the training, validation, and testing accuracy of 100%, 92.38%, and 92.38%, respectively. The training and validation loss amounts are 7% and 6%, respectively. The learning rate was 0.001 for 50 epochs. The accuracy and loss curve is shown in Fig. 16.

Fig. 17 and 18 describe the ROC and Precision-recall

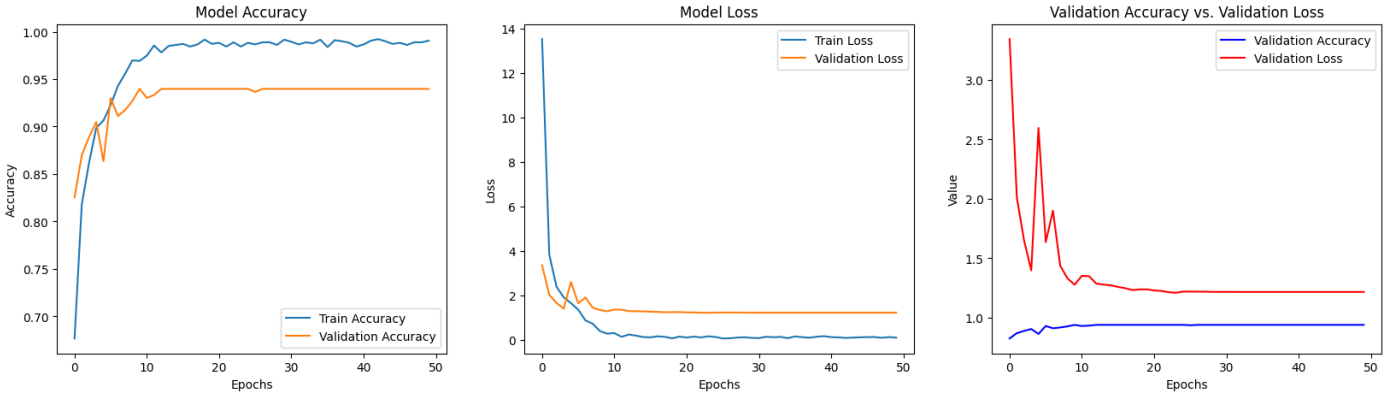


Fig. 10. The training, validation, testing and loss curve of VGG16 model.

Algorithm 3 Proposed Algorithm for Ensemble Learning-2

$ITR \leftarrow NumberofIterations$
$ACC \leftarrow Performancematrix$
$ALG \leftarrow Numberofalgorithms$
#Model Combined
#Extract features using Models combined
$E2 \leftarrow ALG$
for 1 to M **do**
  for each E2 **do**
    for each Sample in X **do**
      calculate A from X by Conv. Process
    end for
  end for
end for
#Number of Iteration
for 1 to ITR **do**
  1. Train Model with N Number of batch size
  2. Feature Extraction through hidden layers
  3. Forward propagations
  4. Backword propagation for updating weights
  5. Model validation with validation data to check overfitting
end for

Algorithm 4 Proposed Algorithm for Ensemble Learning-3

$ITR \leftarrow NumberofIterations$
$ACC \leftarrow Performancematrix$
$ALG \leftarrow Numberofalgorithms$
#Fusion Models
#Extrac features using fusion models
$E3 \leftarrow ALG$
for 1 to M **do**
  for each E3 **do**
    for each Sample in X **do**
      calculate A from X by Conv. Process
    end for
  end for
end for
#Number of Iteration
for 1 to ITR **do**
  1. Train Model with N Number of batch size
  2. Feature Extraction through hidden layers
  3. Forward propagations
  4. Backword propagation for updating weights
  5. Model validation with validation data to check overfitting
end for

curves for the VGG19 Model. Fig. 19 explains the confusion matrix of the VGG19 model. The figures are generated based on data training and validation for 50 epochs using VGG19 Transfer learning models. It is for dataset-1 and we ignore dataset-2. Table IX illustrates the confusion matrix for this model.

3) InceptionV3: Another transfer learning model InceptionV3 is applied in our study for disease diagnosis of rice leaf images. It needs more time due to the increased amount of layers. However, this model classified the images more accurately. The model training, validation, and testing accuracy

TABLE IX. CLASSIFICATION REPORT OF VGG19 MODEL

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	0.93	98	0.95	53
brown-spot	0.91	0.92	0.91	63
healthy	0.91	0.93	0.92	44
leaf-blast	0.85	0.84	0.84	55
leaf-scald	1.00	0.95	0.98	42
narrow-brown-spot	0.96	0.93	0.95	58
accuracy	0	0	0.92	315
macro avg	0.93	0.93	0.93	315
weighted avg	0.92	0.92	0.92	315

are 100%, 93%, and 93%, respectively. The accuracy and loss plots are shown in Fig. 22. This model epoch was 50 and the

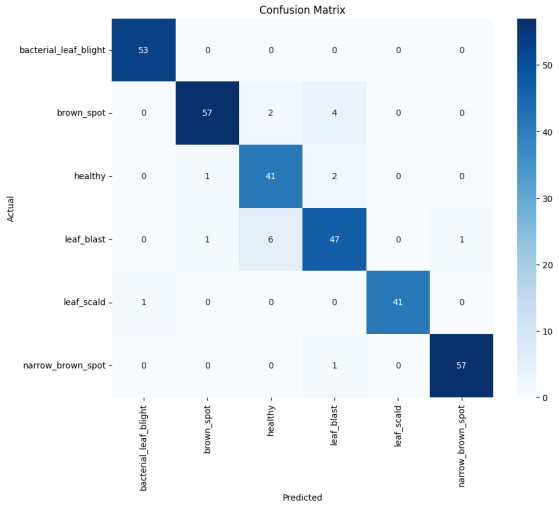


Fig. 11. Confusion matrix for VGG16 model.

curves are generated from the model based on training and validation of the dataset. We used two different datasets but only the dataset-1 curve is explained here due to the lengthy.

Fig. 20 describes the confusion matrix of the InceptionV3 model. The classification report is illustrated in Table X.

TABLE X. CLASSIFICATION REPORT OF INCEPTIONV3 MODEL

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	1.00	98	0.99	53
brown-spot	0.93	0.87	0.90	63
healthy	0.81	0.89	0.85	44
leaf-blast	0.85	0.85	0.85	55
leaf-scald	1.00	1.00	1.00	42
narrow-brown-spot	0.98	1.00	0.99	58
accuracy	0	0	0.92	315
macro avg	0.93	0.93	0.93	315
weighted avg	0.93	0.93	0.93	315

4) ResNet-50: It is also a transfer learning deep layer-based model and it is suitable for image classification and disease detection. ResNet has different types of versions such as ReNet 50, ResNet 152, etc. In this study, we used ResNet-50 for Rice Leaf Disease Diagnosis and classification. In this research, this model was applied with 100% training, 96% validation, and 95/25% testing accuracy. The training loss was 5% and the validation loss was 1%. The confusion matrix of ResNet-50 is explained in Fig. 21. The classification report is shown in Table XI. The ROC and Precision-recall curve are explained in Fig. 23 and 24 for 50 epochs in dataset-1.

TABLE XI. CLASSIFICATION REPORT OF RESNET-50 MODEL

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	0.98	0.96	0.97	53
brown-spot	0.92	0.94	0.93	63
healthy	0.98	0.91	0.94	44
leaf-blast	0.91	0.91	0.91	55
leaf-scald	0.93	1.00	0.97	42
narrow-brown-spot	1.00	1.00	1.00	58
accuracy	0	0	0.95	315
macro avg	0.95	0.95	0.95	315
weighted avg	0.95	0.95	0.95	315

5) EfficientNetV2-M: This is another transfer learning model. It is normally used for large data image processing and classification. This model requires more time and needs also memory due to its long hidden layer. However, it can detect and classify accurately. We used this model for two datasets and got 99.64% training accuracy, 99.56% validation accuracy, and 97.98% testing accuracy. The data loss amount for training is 2% and 4% for validation. The learning rate was 0.001. The confusion matrix of this model is described in Fig. 25 and the classification report is explained in Table XII.

TABLE XII. CLASSIFICATION REPORT OF EFFICIENTNET V2M MODEL

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	0.93	0.98	0.95	53
brown-spot	0.91	0.92	0.91	63
healthy	0.91	0.93	0.92	44
leaf-blast	0.85	0.84	0.84	55
leaf-scald	1.00	0.95	0.98	42
narrow-brown-spot	0.96	0.93	0.95	58
accuracy	0	0	0.92	315
macro avg	0.93	0.93	0.93	315
weighted avg	0.92	0.92	0.92	315

B. Deep Learning Model Performance

To implement the proposed work, we used a deep learning model named Convolutional Neural Network (CNN). This model is suitable for image classification and detection. For rice leaf disease diagnosis and classification, we used it. CNN model architecture is easy and simple to use. After applying this to two separate datasets, we have 99.34% training accuracy, 87.62% validation accuracy, and 90% testing accuracy. The data loss was 3% for training and 5% for validation. The classification report is explained in Table XIII.

TABLE XIII. CLASSIFICATION REPORT OF CNN MODEL

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	0.96	0.98	0.97	53
brown-spot	0.88	0.84	0.86	63
healthy	0.72	0.89	0.80	44
leaf-blast	0.79	0.67	0.73	55
leaf-scald	0.93	0.95	0.94	42
narrow-brown-spot	0.96	0.95	0.96	58
accuracy	0	0	0.88	315
macro avg	0.88	0.88	0.88	315
weighted avg	0.88	0.88	0.88	315

TABLE XIV. CLASSIFICATION REPORT OF ENSEMBLE LEARNING-1

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	1.00	1.00	1.00	53
brown-spot	0.98	0.98	0.99	63
healthy	0.99	0.99	0.99	44
leaf-blast	1.00	1.00	1.00	55
leaf-scald	1.00	1.00	1.00	42
narrow-brown-spot	0.99	0.99	0.99	58
accuracy	0	0	0.99	315
macro avg	0.99	0.99	0.99	315
weighted avg	0.99	0.99	0.99	315

C. Ensemble Learning Performance

In this study, to reduce the data loss and increase the model accuracy, testing accuracy, and validation accuracy we used three ensemble learning. These methods were generated from

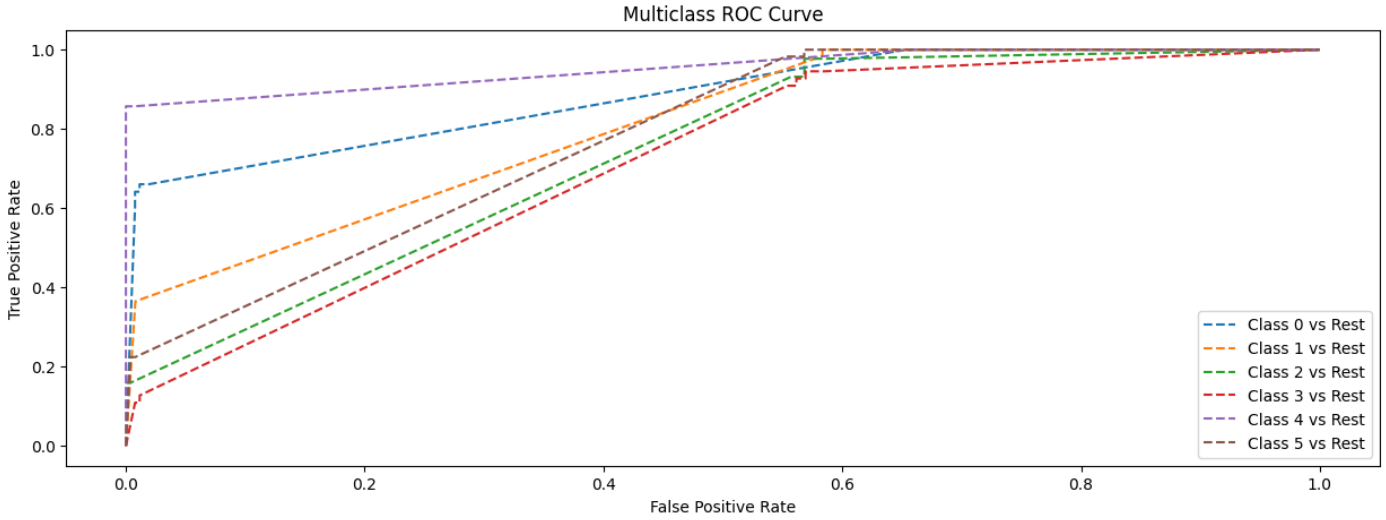


Fig. 12. ROC Curve for VGG16 model.

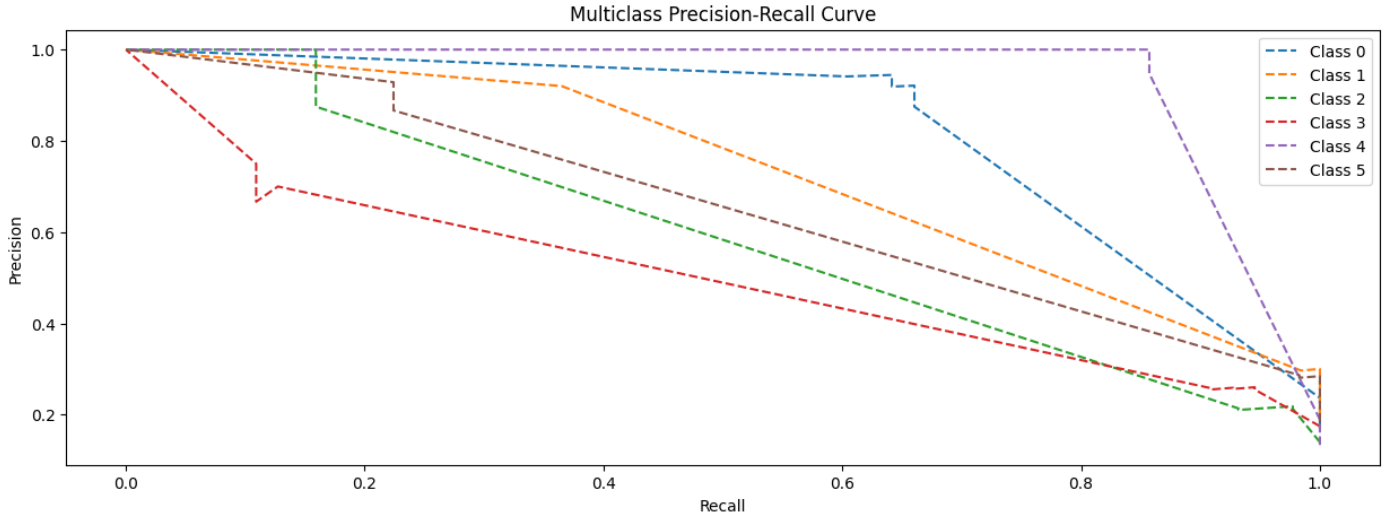


Fig. 13. Precision-recall curve for VGG16 model.

TABLE XV. CLASSIFICATION REPORT OF ENSEMBLE LEARNING-2

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	1.00	1.00	1.00	53
brown-spot	0.99	0.99	0.99	63
healthy	0.99	0.99	0.99	44
leaf-blast	1.00	1.00	1.00	55
leaf-scald	1.00	1.00	1.00	42
narrow-brown-spot	0.99	0.99	0.99	58
accuracy	0	0	0.99	315
macro avg	0.99	0.99	0.99	315
weighted avg	0.99	0.99	0.99	315

TABLE XVI. CLASSIFICATION REPORT OF ENSEMBLE LEARNING-3

Class-Name	Precision	Recall	F1-Score	Support
bacterial-leaf-blight	1.00	0.98	0.99	53
brown-spot	0.87	0.84	0.85	63
healthy	0.72	0.86	0.78	44
leaf-blast	0.78	0.73	0.75	55
leaf-scald	0.98	0.9	0.98	42
narrow-brown-spot	0.98	0.95	0.96	58
accuracy	0	0	0.89	315
macro avg	0.89	0.89	0.89	315
weighted avg	0.89	0.89	0.89	315

deep learning and transfer learning models. We will explore the performance of these ensemble learning models and will propose a new algorithm for the work.

1) Ensemble learning-1 result: fusion of some transfer learning models created this new model for rice leaf disease diagnosis and classified them accurately. Fusion of VGG16, VGG19, InceptionV3, ResNet-50 and EfficientNetV2-M generated ensemble learning-1 named new model and we got 99.14% for training accuracy, 98.98% validation accuracy, and 99% testing accuracy. The data loss was 2% for training, 4% for validation and 3% for testing. The classification report for this model is shown in Table XIV. Fig. 26 describes the model

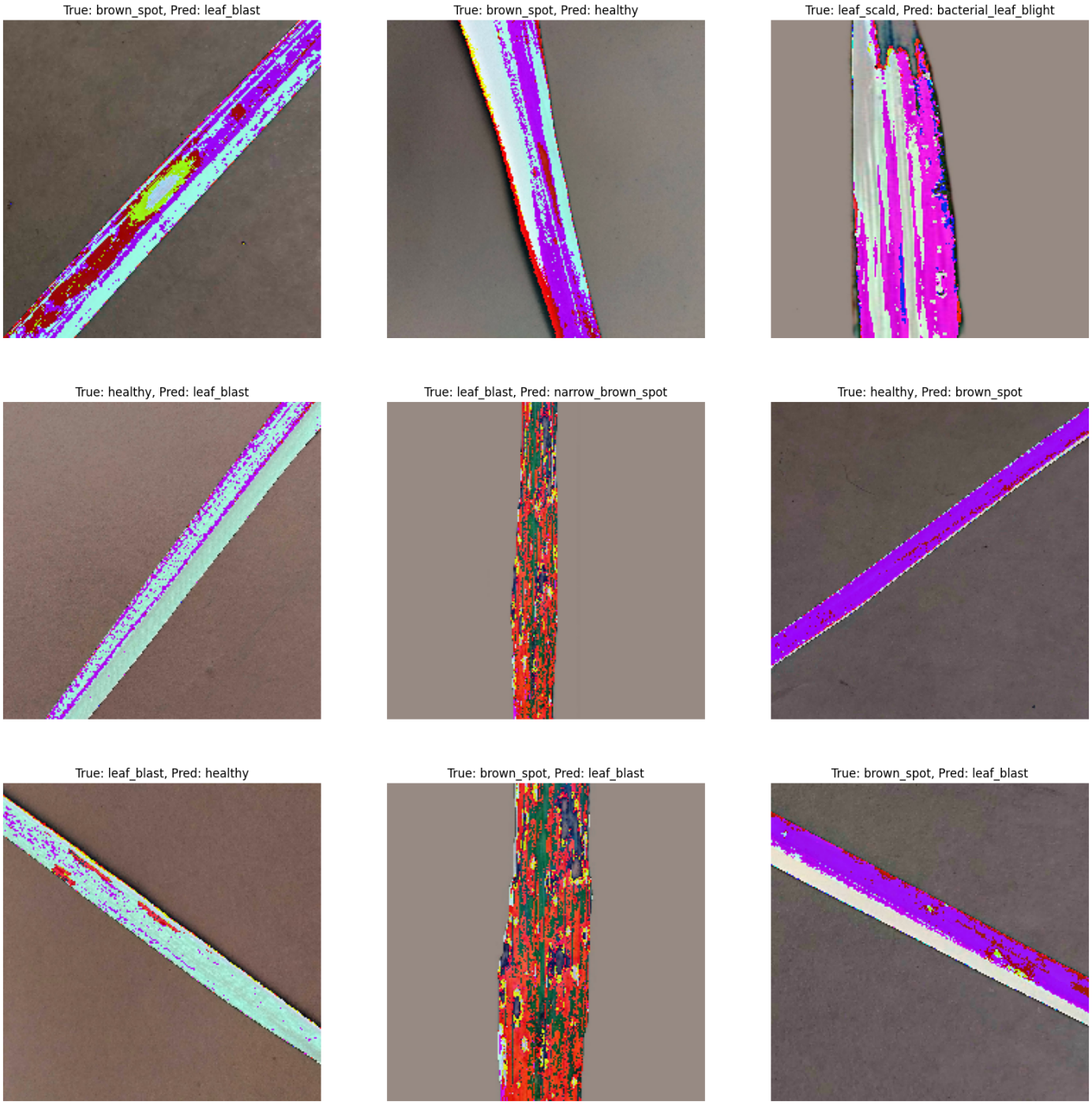


Fig. 14. Misclassified images by VGG16 model.

accuracy, loss, and validation accuracy of ensemble learning-1. This is done for dataset-1 with 50 epochs and not mentioned for dataset-2. Dataset-2 accuracy and loss are most similar to dataset-1.

We got 99.14% accuracy using this technique.

2) Ensemble learning-2 result: Ensemble Learning-2 is generated after combining ResNet-50, VGG16, and InceptionV3. We can say that this is the hybrid version of the transfer learning model for large-scale image classification. Though this technique needs more time for data training and validation, it works accurately. We used it to find out the disease of rice leaves from leaf images. The training accuracy was 99.78%, validation accuracy 98.83%, and testing accuracy 97.89%. The classification report is illustrated in Table XV. Confusion Matrix of Ensemble learning-2 is shown in Fig. 27. Fig. 28 represent the model accuracy, model loss, and validation accuracy-loss of ensemble learning-2. Algorithm-2 represents the proposed algorithm for ensemble learning-2

Fig. 15. Correctly classified images by VGG16 model.

in this study. We developed and used this algorithm. This may combination-variant of VGG16, ResNet-50, and InceptionV3.

3) Ensemble learning-3 result: We ensembled one deep learning model named CNN and one transfer learning model named VGG16 and generated a new model named ensemble learning-3. We may consider this new model as a variant of the deep-transfer learning model. It is suitable for detection and classification of large image datasets. We used this new variant for two datasets. But dataset-1 got 99.36% training accuracy, 90.57% validation accuracy, and 92% testing accuracy. But in dataset-2, the training accuracy was 97%, validation accuracy was 95% and testing accuracy was 90%. Fig. 29 represent the accuracy and loss of the ensemble learning-3. Algorithm 3 is the proposed algorithm for ensemble learning-3. We developed it based on this new model. Fig. 30 illustrates the confusion matrix for ensemble learning-3. Table XVI represents the classification report for ensemble learning-3.

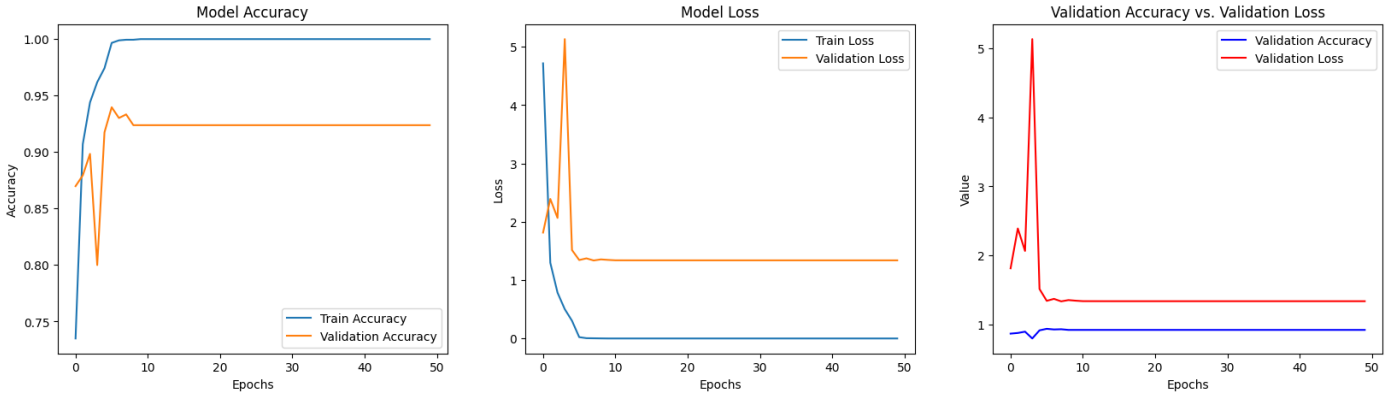


Fig. 16. Accuracy and loss curve for VGG19 model.

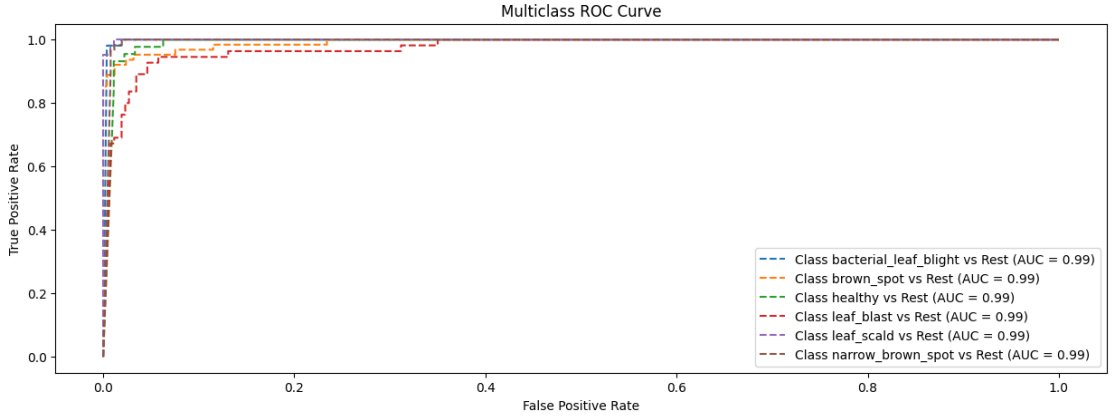


Fig. 17. ROC Curve for VGG19 model.

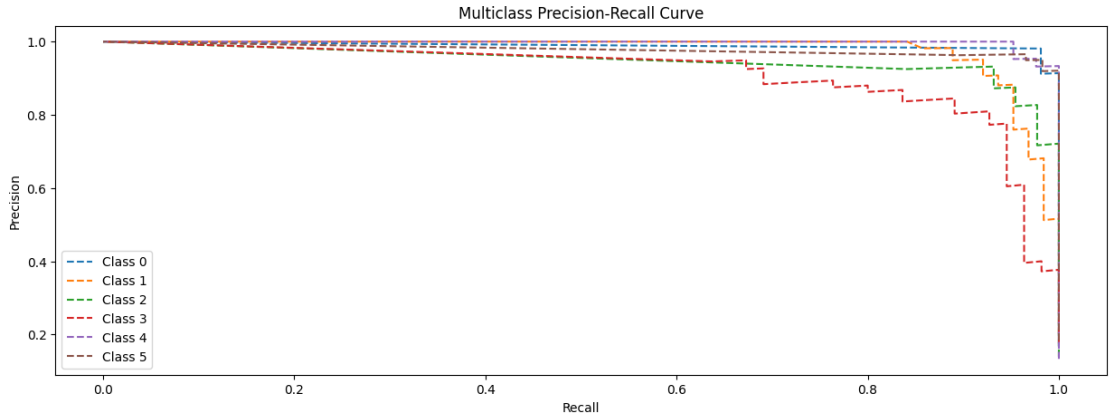


Fig. 18. Precision-recall curve for VGG19 model.

D. Explainable AI "Grad-CAM" for Ensemble Learning-1, 2 and 3

The objective of this work was to increase the interpretability and reliability of image classification using three different ensemble learning models. Combining five strong architectures (VGG16, VGG19, ResNet50, InceptionV3, and EfficientNetV2-M) and leveraging their own strengths, the first ensemble, Ensemble Learning-1, is a formidable combination. Combining the depth of ResNet50, the multi-scale feature extraction of InceptionV3, and the simplicity of VGG16, the second model, Ensemble Learning-2, streamlines these three models. To balance task-specific learning with general feature extraction, the third model, Ensemble Learning-3, combines two deep learning models: CNN and VGG16. The image classification tasks were used to train each ensemble model, and Explainable AI (XAI) techniques like Grad-CAM were

TABLE XVII. COMPARATIVE ANALYSIS OF USED MODEL IN DATASET-1 AND 2

Algorithm	Dataset-1			Dataset-2		
	train	vali:	test	train	vali:	test
VGG16	98.94%	93.97%	92.70%	97.94%	96.78%	95%
VGG19	100%	92.38%	92.38%	99.99%	97%	96%
InceptionV3	100%	93%	93%	92%	95%	94%
ResNet-50	100%	96%	95.25%	99.99%	97%	98%
EfficienNet	99.64%	99.56%	97.98%	98.98%	98.90%	96.98%
CNN	99.34%	87.62%	90%	96.87%	93.78%	94.67%
Ensemble 1	99.14%	98.98%	99%	97.98%	97%	97%
Ensemble 2	99.78%	98.83%	97.89%	96.99%	99%	96%
Ensemble 3	99.36%	90.57%	92%	98.99%	98%	99.98%

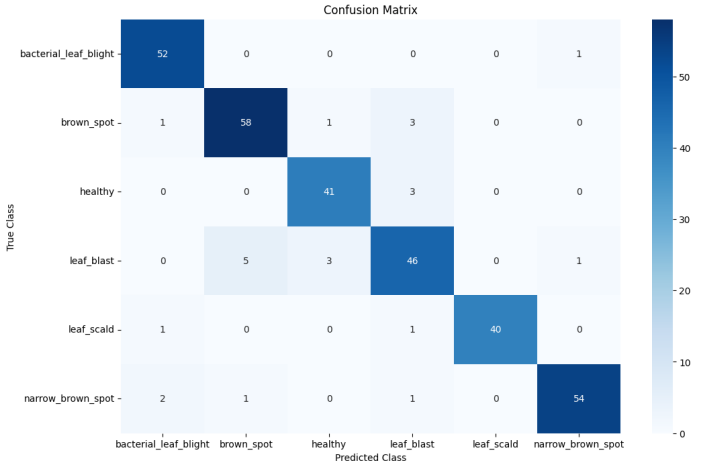


Fig. 19. Confusion matrix for VGG19 model.

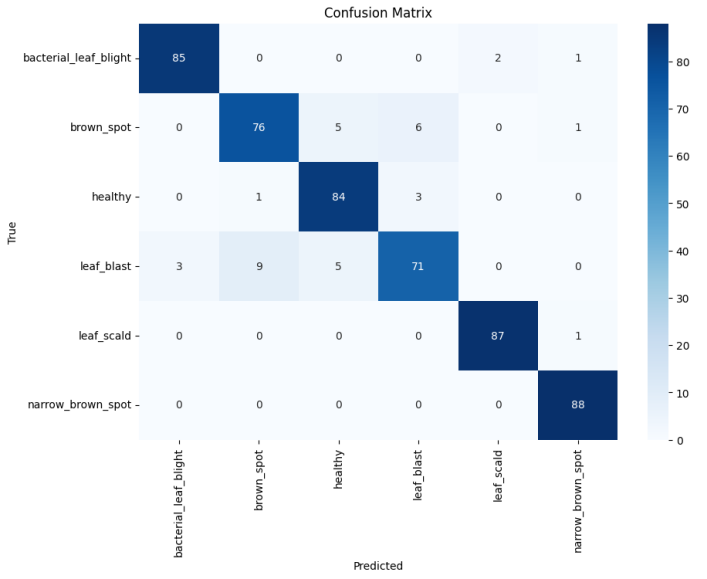


Fig. 20. Confusion matrix for inceptionV3 model.

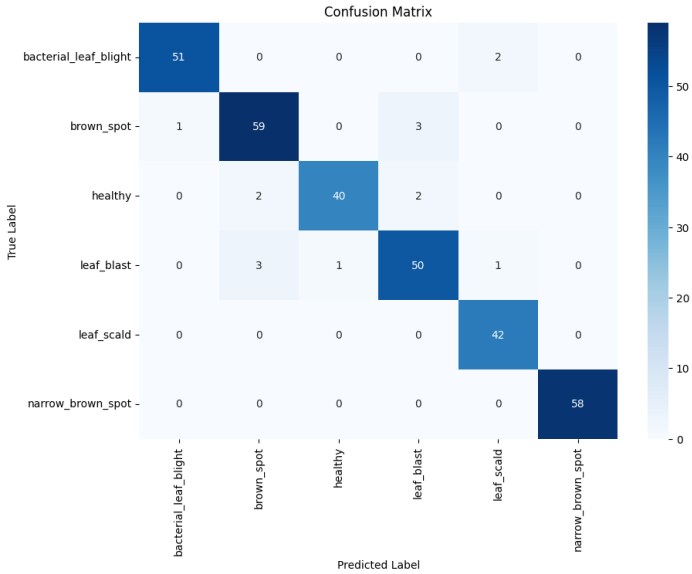


Fig. 21. Confusion matrix for ResNet-50 model.

used to assess interpretability [77], [78]. We evaluated a picture from the "Brown-spot" class to assess the model. Heat maps showing how the model recognized the disease were created using Grad-CAM. The models were able to identify the "Brown-spot" in the input image in every instance, demonstrating the effectiveness of combining XAI and ensemble learning [79]. The XAI-based assessments for models 1, 2, and 3 of ensemble learning are shown in Fig. 31, 32, and 33, respectively. The model predictions perform better and are more transparent using this technique.

E. Discussion

The integration of deep learning and transfer learning models presents a highly efficient approach for the practical diagnosis and categorization of rice leaf disease [80], [49]. Accurately detecting illnesses in rice leaves is critical for crop health and yield maintenance [81], [82]. Models like as VGG16, VGG19, ResNet50, InceptionV3, and EfficientNetV2-M do this [83]. Ensemble Learning-1 demonstrated an accuracy of 99.78% in classification as a result of the integration of these models into ensemble learning structures [84], [85]. Through early disease detection and substantial crop loss prevention, this technique makes it possible to process rice leaf pictures in an effective manner [86], [87].

Moreover, the diagnosis process gains interpretability with the incorporation of Explainable AI (XAI) tools such as Grad-CAM [88]. XAI makes it easier to see which areas of the rice leaf the model concentrates on in order to identify diseases, giving agronomists and farmers more reliable and transparent information [89]. For practical application in agriculture,

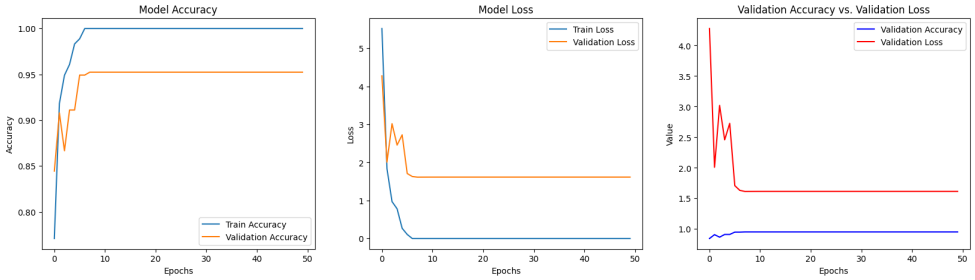


Fig. 22. Accuracy and loss curve for InceptionV3 model.

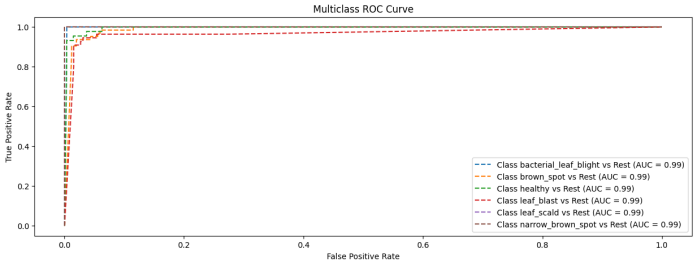


Fig. 23. ROC Curve for ResNet-50 model.

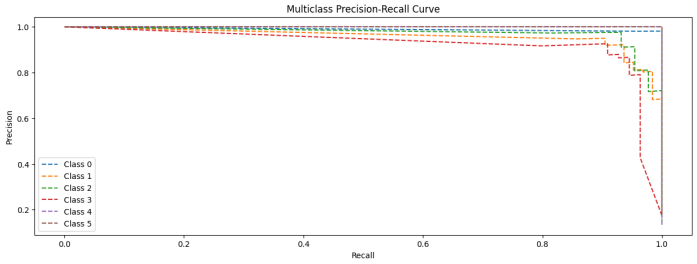


Fig. 24. Precision-recall curve for ResNet-50 model.

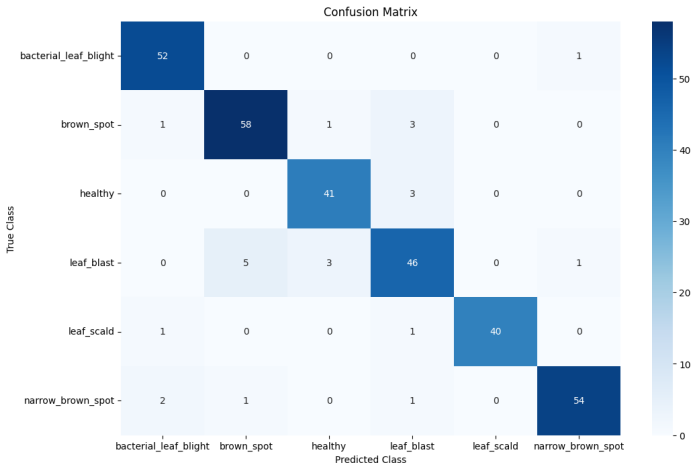


Fig. 25. Confusion matrix for EfficientNetV2-M.

where decision-making depends on the capacity to comprehend the reasoning behind a model's predictions, this transparency is crucial [90]. Rice producers can implement preventive measures earlier and improve crop management and agricultural sustainability by utilizing these AI-driven techniques [91].

Individual models like ResNet50, VGG19, VGG16, and InceptionV3 demonstrated their strong capacity to classify diseases correctly, achieving perfect accuracy (100%), when comparing the performance of individual and ensemble learning models for rice leaf disease detection [92]. Nonetheless, the accuracy of 99.78% was slightly lower but still outstanding when using ensemble learning models, especially Ensemble Learning-1 (which combined VGG16, VGG19, ResNet50, InceptionV3, and EfficientNetV2-M). By utilizing numerous architectures, ensemble models provide the advantage of lowering data loss and enhancing overall robustness despite the slight decrease[23]. The results of the ensembles indicate that adding more diverse models leads to higher generalization and performance, as Ensemble Learning-1 outperformed Ensemble Learning-2 (VGG16, InceptionV3, ResNet50) and Ensemble Learning-3 (VGG16 and CNN).

Furthermore, the incorporation of Explainable AI (XAI) methods, like Grad-CAM, significantly increased the value by offering visual insights into the models' disease detection processes and improving forecast transparency [86]. This is especially significant for agricultural applications because practical usage of the model depends on the user's ability to understand its decisions. As a result, while individual models are very accurate, ensemble learning with XAI balances high accuracy, interpretability, and minimal data loss, making it a more useful and trustworthy method for diagnosing rice leaf disease [93]. To see the comparative analysis of proposed work with other researchers, see Table I; where we got the highest accuracy than other recently published papers. Table XVII represents the comparison of used methods for dataset-1 and dataset-2. The biases of training, validation, and testing accuracy are different in dataset-1 and 2 due to the image characteristics or features.

V. Conclusion and Future Work

This study's findings demonstrate the value of integrating deep learning and transfer learning models to accurately diagnose and classify illnesses affecting rice leaves. Combining different architectures improves accuracy and reliability. The ensemble learning models performed well, with Ensemble Learning attaining the maximum accuracy of 99.78%. Several models, including ResNet50, VGG19, VGG16, and InceptionV3, demonstrated 100% accuracy, demonstrating their efficacy in the identification of diseases. This strategy is useful for

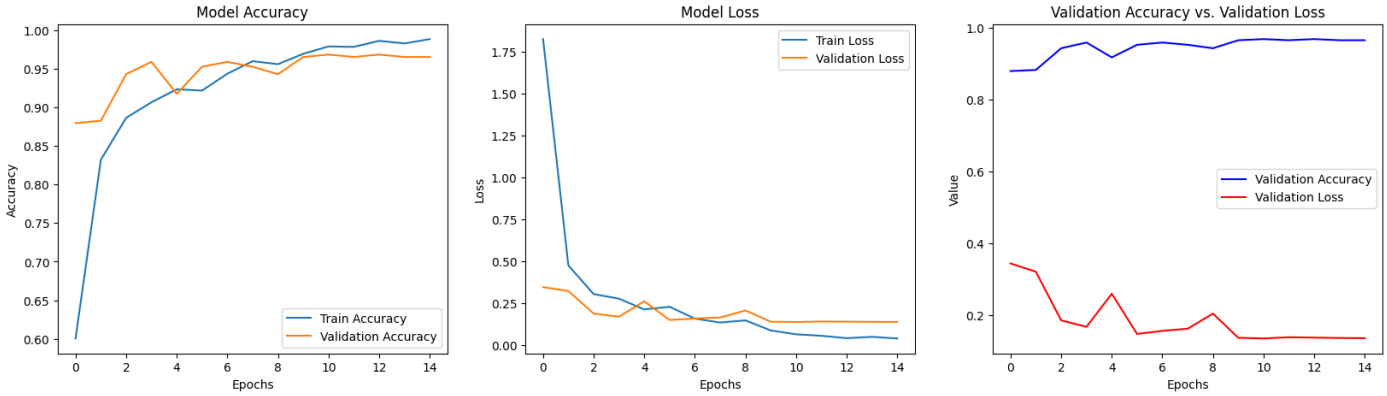


Fig. 26. Model accuracy, loss and validation for ensemble Learning-1.

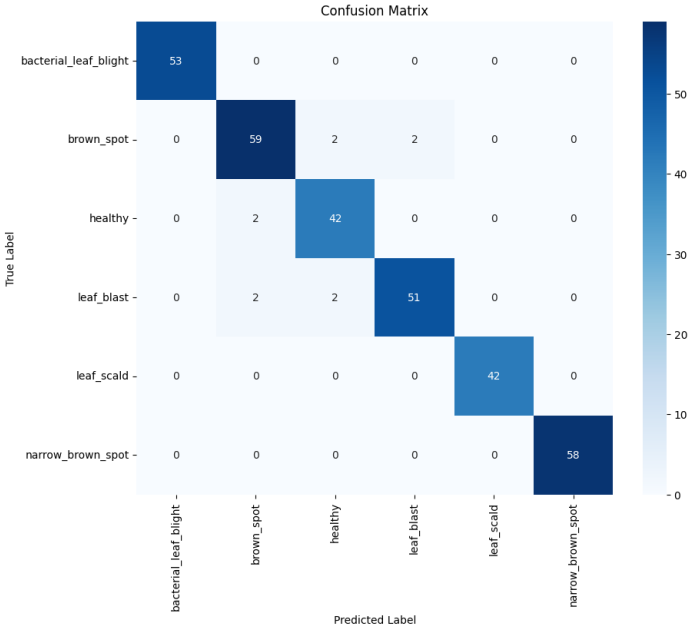


Fig. 27. Confusion matrix for ensemble Learning-2.

real-world applications in agricultural disease control since Explainable AI (XAI) approaches like Grad-CAM also increase the predictability and transparency of the models. We intend to improve performance even further by refining the suggested ensemble learning approach in further work. Furthermore, we hope to investigate the possibility of incorporating real-time data for ongoing observation and early disease diagnosis, as well as to expand this methodology to other crop diseases. Including more extensive and varied datasets may also contribute to improving the models' resilience. The ultimate objective is to create an all-inclusive, field-deployable automated system for diagnosing crop diseases in order to promote sustainable agriculture.

Data Availability

The used datasets are open-access and referenced in this manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] N. A. Mohidem, N. Hashim, R. Shamsudin, and H. Che Man, "Rice for food security: Revisiting its production, diversity, rice milling process and nutrient content," *Agriculture*, vol. 12, no. 6, p. 741, 2022.

[2] T. Mahmud, T. Akter, S. Anwar, M. T. Aziz, M. S. Hossain, and K. Andersson, "Predictive modeling in forex trading: A time series analysis approach," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2024, pp. 390–397.

[3] T. Akter, T. Mahmud, R. Chakma, N. Datta, M. S. Hossain, and K. Andersson, "Iot-based precision agriculture monitoring system: Enhancing agricultural efficiency," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE Computer Society, 2024, pp. 749–754.

[4] R. A. Hidayat, J. Iskandar, B. Gunawan, and R. Partasasmita, "Impact of green revolution on rice cultivation practices and production system: A case study in sindang hamlet, rancakalong village, sumedang district, west java, indonesia," *Biodiversitas Journal of Biological Diversity*, vol. 21, no. 3, 2020.

[5] T. Rahman, T. Mahmud, M. M. Setara, S. Roy, M. S. Hossain, and K. Andersson, "Application of deep learning for detecting rice leaf diseases in jhum cultivation," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–6.

[6] T. Akter, T. Mahmud, R. Chakma, N. Datta, M. S. Hossain, and K. Andersson, "Smart monitoring and control of hydroponic systems using iot solutions," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE Computer Society, 2024, pp. 761–767.

[7] T. Mahmud, K. Barua, A. Barua, N. Basnin, S. Das, M. S. Hossain, and K. Andersson, "Explainable ai for tomato leaf disease detection: Insights into model interpretability," in *2023 26th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2023, pp. 1–6.

[8] P. Dey, T. Mahmud, S. R. Nahar, M. S. Hossain, and K. Andersson, "Plant disease detection in precision agriculture: Deep learning approaches," in *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2024, pp. 661–667.

[9] K. Simkhada and R. Thapa, "Rice blast, a major threat to the rice production and its various management techniques," *Turkish Journal of Agriculture-Food Science and Technology*, vol. 10, no. 2, pp. 147–157, 2022.

[10] C. Jackulin and S. Murugavalli, "A comprehensive review on detection of plant disease using machine learning and deep learning approaches," *Measurement: Sensors*, vol. 24, p. 100441, 2022.

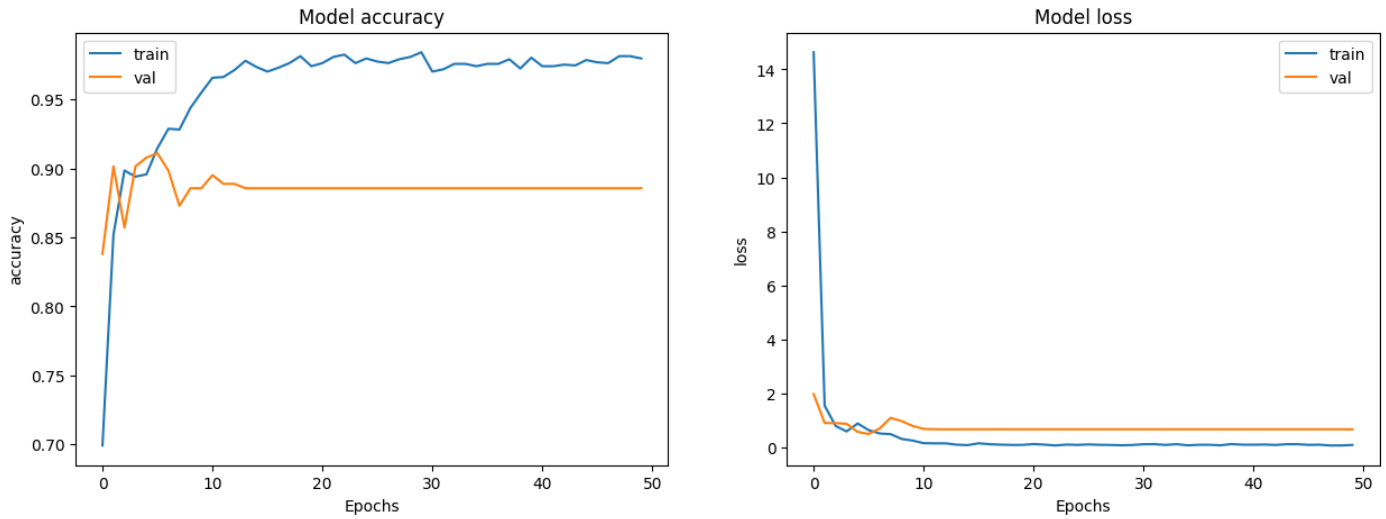


Fig. 28. Model accuracy and loss curve for ensemble Learning-2.

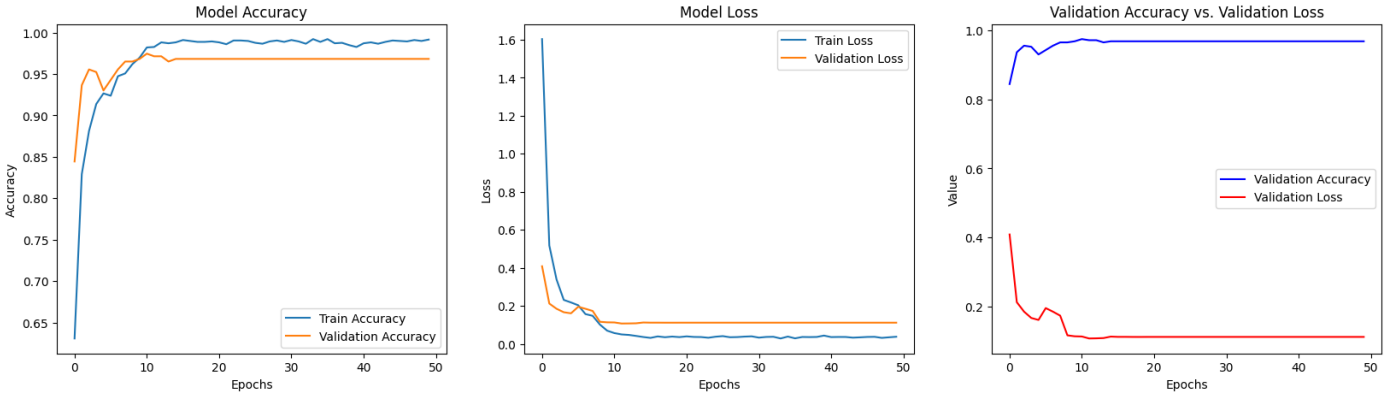


Fig. 29. Accuracy, loss and validation plot for ensemble Learning-3.

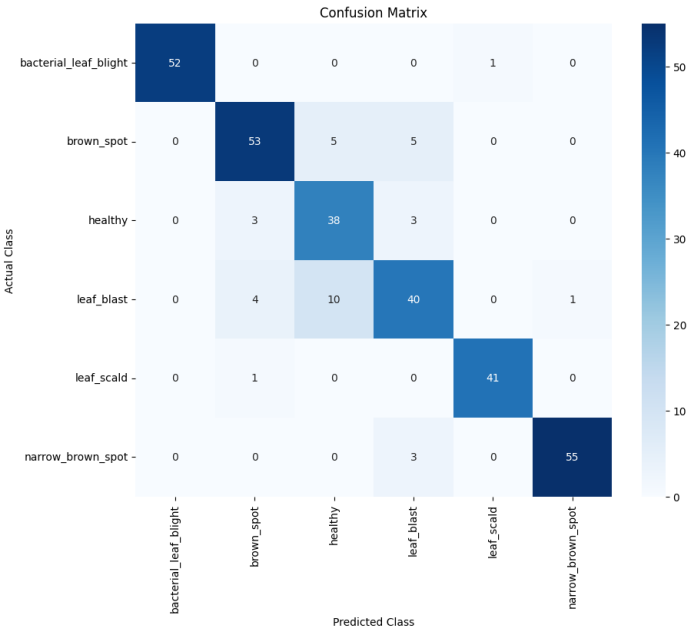


Fig. 30. Confusion matrix for ensemble Learning-3.

Fig. 31. Explainable AI tested image 1 for ensemble learning-2.

[11] T. Akter, T. Mahmud, R. Chakma, N. Datta, M. S. Hossain, and K. Andersson, "Iot in action: Design and implementation of a tank water monitoring system," in *2024 Second International Conference on*

Fig. 32. Explainable AI tested image 2 for ensemble Learning-2.

Fig. 33. Explainable AI tested image 3 for ensemble Learning-2.

Inventive Computing and Informatics (ICICI). IEEE Computer Society, 2024, pp. 755–760.

[12] M. de Benito Fernández, D. L. Martínez, A. González-Briones, P. Chamoso, and E. S. Corchado, "Evaluation of xai models for interpretation of deep learning techniques' results in automated plant disease diagnosis," in *Sustainable Smart Cities and Territories International Conference*. Springer, 2023, pp. 417–428.

[13] M. T. Aziz, T. Mahmud, M. K. Uddin, S. N. Hossain, N. Datta, S. Akther, M. S. Hossain, and K. Andersson, "Machine learning-driven job recommendations: Harnessing genetic algorithms," in *International Congress on Information and Communication Technology*. Springer, 2024, pp. 471–480.

[14] M. T. Aziz, R. Sudheesh, R. D. C. Pecho, N. U. A. Khan, A. Ull, H. Era, and M. A. Chowdhury, "Calories burnt prediction using machine learning ap-proach," *Current Integrative Engineering*, vol. 1, no. 1, pp. 29–36, 2023.

[15] T. Mahmud, B. Saha, D. Islam, M. T. Aziz, N. Datta, K. Barua, M. S. Hossain, and K. Andersson, "Deep learning approach for driver drowsiness detection in real time," in *Innovations in Cybersecurity and Data Science*. Singapore: Springer Nature Singapore, 2024, pp. 775–790.

[16] M. H. Ali, T. Mahmud, M. T. Aziz, M. F. B. A. Aziz, M. S. Hossain, and K. Andersson, "Leveraging transfer learning for efficient classification of coffee leaf diseases," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–6.

[17] M. H. Imam, N. Nahar, R. Bhowmik, S. B. S. Omit, T. Mahmud, M. S. Hossain, and K. Andersson, "A transfer learning-based framework: Mobilenet-svm for efficient tomato leaf disease classification," in *2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT)*, 2024, pp. 693–698.

[18] M. T. Aziz, T. Mahmud, N. Datta, M. Maskat Sharif, N. U. A. Khan, S. Yasmin, M. D. N. Uddin, M. S. Hossain, and K. Andersson, "A state-of-the-art review of machine learning in cybersecurity data science," in *Innovations in Cybersecurity and Data Science*. Singapore: Springer Nature Singapore, 2024, pp. 791–806.

[19] T. Mahmud, N. Datta, R. Chakma, U. K. Das, M. T. Aziz, M. Islam, A. H. M. Salimullah, M. S. Hossain, and K. Andersson, "An approach for crop prediction in agriculture: Integrating genetic algorithms and machine learning," *IEEE Access*, 2024.

[20] S. Ghosal and K. Sarkar, "Rice leaf diseases classification using cnn with transfer learning," in *2020 IEEE Calcutta Conference (Calcon)*. IEEE, 2020, pp. 230–236.

[21] S. Ramesh and D. Vydeki, "Rice disease detection and classification using deep neural network algorithm," in *Micro-Electronics and Telecommunication Engineering: Proceedings of 3rd ICMETE 2019*. Springer, 2020, pp. 555–566.

[22] J. Chen, D. Zhang, Y. A. Nanehkaran, and D. Li, "Detection of rice plant diseases based on deep transfer learning," *Journal of the Science of Food and Agriculture*, vol. 100, no. 7, pp. 3246–3256, 2020.

[23] M. A. Islam, M. N. R. Shuvo, M. Shamsojjaman, S. Hasan, M. S. Hossain, and T. Khatun, "An automated convolutional neural network based approach for paddy leaf disease detection," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 1, 2021.

[24] M. A. Azim, M. K. Islam, M. M. Rahman, and F. Jahan, "An effective feature extraction method for rice leaf disease classification," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 19, no. 2, pp. 463–470, 2021.

[25] A. Islam, R. Islam, S. R. Haque, S. M. Islam, and M. A. I. Khan, "Rice leaf disease recognition using local threshold based segmentation and deep cnn," *International Journal of Intelligent Systems and Applications*, vol. 10, no. 5, p. 35, 2021.

[26] K. Kiratiratanapruk, P. Temniranrat, W. Sinthupinyo, S. Marukatat, and S. Patarapuwadol, "Automatic detection of rice disease in images of various leaf sizes," *arXiv preprint arXiv:2206.07344*, 2022.

[27] P. Tejaswini, P. Singh, M. Ramchandani, Y. K. Rathore, and R. R. Janghel, "Rice leaf disease classification using cnn," in *IOP Conference Series: Earth and Environmental Science*, vol. 1032, no. 1. IOP Publishing, 2022, p. 012017.

[28] S. T. Y. Ramadan, T. Sakib, M. M. U. Haque, N. Sharmin, and M. M. Rahman, "Generative adversarial network-based augmented rice leaf disease detection using deep learning," in *2022 25th International Conference on Computer and Information Technology (ICCIT)*. IEEE, 2022, pp. 976–981.

[29] M. E. Haque, A. Rahman, I. Junaeid, S. U. Hoque, and M. Paul, "Rice leaf disease classification and detection using yolov5," *arXiv preprint arXiv:2209.01579*, 2022.

[30] R. Setiawan, H. Zein, R. A. Azdy, and S. Sulistyowati, "Rice leaf disease classification with machine learning: An approach using nu-svm," *Indonesian Journal of Data and Science*, vol. 4, no. 3, pp. 136–144, 2023.

[31] N. Bharanidharan, S. S. Chakravarthy, H. Rajaguru, V. V. Kumar, T. Mahesh, and S. Guluwadi, "Multiclass paddy disease detection using filter based feature transformation technique," *IEEE Access*, 2023.

[32] R. P. Ethiraj and K. Paranjothi, "A deep learning-based approach for early detection of disease in sugarcane plants: an explainable artificial intelligence model," *Int J Artif Intell ISSN*, vol. 2252, no. 8938, p. 8938.

[33] V. S. S. V. Chivukula, G. Anuradha, S. N. C. Dhanekula, and N. G. Kothagundla, "Rice crop disease detection using explainable ai," in *2023 Global Conference on Information Technologies and Communications (GCITC)*. IEEE, 2023, pp. 1–8.

[34] A. Kaur, K. Guleria, and N. K. Trivedi, "A deep learning-based model for biotic rice leaf disease detection," *Multimedia Tools and Applications*, pp. 1–27, 2024.

[35] P. Kulkarni and S. Shastri, "Rice leaf diseases detection using machine learning," *Journal of Scientific Research and Technology*, pp. 17–22, 2024.

[36] M. Tariq, U. Ali, S. Abbas, S. Hassan, R. A. Naqvi, M. A. Khan, and D. Jeong, "Corn leaf disease: insightful diagnosis using vgg16 empowered by explainable ai," *Frontiers in Plant Science*, vol. 15, 2024.

[37] D. C. Trinh, A. T. Mac, K. G. Dang, H. T. Nguyen, H. T. Nguyen, and T. D. Bui, "Alpha-eiou-yolov8: an improved algorithm for rice leaf disease detection," *AgriEngineering*, vol. 6, no. 1, pp. 302–317, 2024.

[38] W. K. Mutlag, S. K. Ali, Z. M. Aydam, and B. H. Taher, "Feature extraction methods: a review," in *Journal of Physics: Conference Series*, vol. 1591, no. 1. IOP Publishing, 2020, p. 012028.

[39] Y. Zhang, X. Huang, J. Ma, Z. Li, Z. Luo, Y. Xie, Y. Qin, T. Luo, Y. Li, S. Liu *et al.*, "Recognize anything: A strong image tagging model," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 1724–1732.

[40] T. Gayathri Devi and P. Neelamegam, "Image processing based rice plant leaves diseases in thanjavur, tamilnadu," *Cluster Computing*, vol. 22, no. Suppl 6, pp. 13 415–13 428, 2019.

[41] S. Umme Habiba, F. Tasnim, M. S. Hasan Chowdhury, M. K. Islam, L. Nahar, T. Mahmud, M. S. Kaiser, M. S. Hossain, and K. Andersson, "Early prediction of chronic kidney disease using machine learning algorithms with feature selection techniques," in *International Conference on Applied Intelligence and Informatics*. Springer, 2023, pp. 224–242.

[42] T. Akter, A. Majumder, T. Mahmud, I. B. Habib, M. S. Hossain, and K. Andersson, "Advancements in animal tracking: Assessing deep learning algorithms," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–6.

[43] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognition*, vol. 137, p. 109347, 2023.

[44] P. Chlap, H. Min, N. Vandenberg, J. Dowling, L. Holloway, and A. Haworth, "A review of medical image data augmentation techniques for deep learning applications," *Journal of Medical Imaging and Radiation Oncology*, vol. 65, no. 5, pp. 545–563, 2021.

[45] S. Das, T. Mahmud, D. Islam, M. Begum, A. Barua, M. Tarek Aziz, E. Nur Showan, L. Dey, and E. Chakma, "Deep transfer learning-based foot no-ball detection in live cricket match," *Computational Intelligence and Neuroscience*, vol. 2023, no. 1, p. 2398121, 2023.

[46] S. U. Habiba, T. Mahmud, S. R. Naher, M. T. Aziz, T. Rahman, N. Datta, M. S. Hossain, K. Andersson, and M. Shamim Kaiser, "Deep learning solutions for detecting bangla fake news: A cnn-based approach," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 107–118.

[47] M. T. Aziz, J. Sikder, T. Rahman, A. D. Del Mundo, S. F. Faisal, and N. U. A. Khan, "Covid-19 detection from chest x-ray images using deep learning," *The Seybold Report*, vol. 17, pp. 706–718, 2022.

[48] S. Mascarenhas and M. Agarwal, "A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification," in *2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)*, vol. 1. IEEE, 2021, pp. 96–99.

[49] T. Mahmud, I. Hasan, M. T. Aziz, T. Rahman, M. S. Hossain, and K. Andersson, "Enhanced fake news detection through the fusion of deep learning and repeat vector representations," in *2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*. IEEE, 2024, pp. 654–660.

[50] Z.-P. Jiang, Y.-Y. Liu, Z.-E. Shao, and K.-W. Huang, "An improved vgg16 model for pneumonia image classification," *Applied Sciences*, vol. 11, no. 23, p. 11185, 2021.

[51] S. R. Naher, S. Sultana, T. Mahmud, M. T. Aziz, M. S. Hossain, and K. Andersson, "Exploring deep learning for chittagonian slang detection in social media texts," in *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET*. IEEE, 2024, pp. 1–6.

[52] M. J. Awan, O. A. Masood, M. A. Mohammed, A. Yasin, A. M. Zain, R. Damaševičius, and K. H. Abdulkareem, "Image-based malware classification using vgg19 network and spatial convolutional attention," *Electronics*, vol. 10, no. 19, p. 2444, 2021.

[53] T. Mahmud, T. Akter, M. T. Aziz, M. K. Uddin, M. S. Hossain, and K. Andersson, "Integration of nlp and deep learning for automated fake news detection," in *2024 Second International Conference on Inventive Computing and Informatics (ICICI)*. IEEE, 2024, pp. 398–404.

[54] M. Bansal, M. Kumar, M. Sachdeva, and A. Mittal, "Transfer learning for image classification using vgg19: Caltech-101 image data set," *Journal of ambient intelligence and humanized computing*, pp. 1–12, 2023.

[55] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. d. l. T. Díez, and I. Ashraf, "Pneumonia classification from x-ray images with inception-v3 and convolutional neural network," *Diagnostics*, vol. 12, no. 5, p. 1280, 2022.

[56] K. Joshi, V. Tripathi, C. Bose, and C. Bhardwaj, "Robust sports image classification using inceptionv3 and neural networks," *Procedia Computer Science*, vol. 167, pp. 2374–2381, 2020.

[57] S. Rajpal, N. Lakhyani, A. K. Singh, R. Kohli, and N. Kumar, "Using handpicked features in conjunction with resnet-50 for improved detection of covid-19 from chest x-ray images," *Chaos, Solitons & Fractals*, vol. 145, p. 110749, 2021.

[58] X. Ma, Z. Li, and L. Zhang, "An improved resnet-50 for garbage image classification," *Tehnički vjesnik*, vol. 29, no. 5, pp. 1552–1559, 2022.

[59] A. Saxena, A. Ajit, C. Arora, and G. Raj, "Efficient net v2 algorithm-based nsfw content detection check for updates," in *Decision Intelligence Solutions: Proceedings of the International Conference on Information Technology, InCITe 2023, Volume 2*, vol. 1080. Springer Nature, 2023, p. 343.

[60] T. Mahmud, M. T. Aziz, M. K. Uddin, K. Barua, T. Rahman, N. Sharmen, M. Shamim Kaiser, M. Sazzad Hossain, M. S. Hossain, and K. Andersson, "Ensemble learning approaches for alzheimer's disease classification in brain imaging data," in *International Conference on Trends in Electronics and Health Informatics*. Springer, 2023, pp. 133–147.

[61] Y. Zheng, C. Li, X. Zhou, H. Chen, H. Xu, Y. Li, H. Zhang, X. Li, H. Sun, X. Huang *et al.*, "Application of transfer learning and ensemble learning in image-level classification for breast histopathology," *Intelligent Medicine*, vol. 3, no. 02, pp. 115–128, 2023.

[62] D. Müller, I. Soto-Rey, and F. Kramer, "An analysis on ensemble learning optimized medical image classification with deep convolutional neural networks," *Ieee Access*, vol. 10, pp. 66 467–66 480, 2022.

[63] F. Azour and A. Boukerche, "An efficient transfer and ensemble learning based computer aided breast abnormality diagnosis system," *IEEE Access*, vol. 11, pp. 21 199–21 209, 2022.

[64] N. Remzan, Y. E. Hachimi, K. Tahiry, and A. Farchi, "Ensemble learning based-features extraction for brain mr images classification with machine learning classifiers," *Multimedia Tools and Applications*, vol. 83, no. 19, pp. 57 661–57 684, 2024.

[65] N. K. Chowdhury, M. A. Kabir, M. M. Rahman, and N. Rezoana, "Ecovnet: an ensemble of deep convolutional neural networks based on efficientnet to detect covid-19 from chest x-rays," *arXiv preprint arXiv:2009.11850*, 2020.

[66] N. D. Jana, S. Dhar, S. Ghosh, S. Phukan, R. Gogoi, and J. Singh, "An ensemble of machine learning models utilizing deep convolutional features for medical image classification," in *International Conference on Advanced Network Technologies and Intelligent Computing*. Springer, 2023, pp. 384–396.

[67] A. Gupta, D. Gupta, M. Pathak, and S. K. Wagh, "The advancement of ensemble deep learning architecture for the detection and classification of brain tumours with mri images," *International Journal of Biomedical Engineering and Technology*, vol. 45, no. 1, pp. 27–44, 2024.

[68] S. R. Shah, S. Qadri, H. Bibi, S. M. W. Shah, M. I. Sharif, and F. Marinello, "Comparing inception v3, vgg 16, vgg 19, cnn, and resnet 50: a case study on early detection of a rice disease," *Agronomy*, vol. 13, no. 6, p. 1633, 2023.

[69] Y. Pan, J. Liu, Y. Cai, X. Yang, Z. Zhang, H. Long, K. Zhao, X. Yu, C. Zeng, J. Duan *et al.*, "Fundus image classification using inception v3 and resnet-50 for the early diagnostics of fundus diseases," *Frontiers in Physiology*, vol. 14, p. 1126780, 2023.

[70] R. Pillai, A. Sharma, N. Sharma, and R. Gupta, "Brain tumor classification using vgg 16, resnet50, and inception v3 transfer learning models," in *2023 2nd International Conference for Innovation in Technology (INOCON)*. IEEE, 2023, pp. 1–5.

[71] A. Hussain, K. N. Qureshi, R. W. Anwar, and A. Aslam, "A novel scd11 cnn model performance evaluation with inception v3, vgg16 and resnet50 using surface crack dataset," in *2024 2nd International Conference on Unmanned Vehicle Systems-Oman (UVS)*. IEEE, 2024, pp. 1–7.

[72] I. Ali, M. Muzammil, I. U. Haq, M. Amir, and S. Abdullah, "Deep feature selection and decision level fusion for lungs nodule classification," *IEEE Access*, vol. 9, pp. 18 962–18 973, 2021.

[73] C. Mukesh, A. Likhita, and A. Yamini, "Performance analysis of inceptionv3, vgg16, and resnet50 models for crevices recognition on surfaces," in *International Conference on Data Science and Applications*. Springer, 2023, pp. 161–172.

[74] S. O'Shaughnessy and S. Sheridan, "Image-based malware classification hybrid framework based on space-filling curves," *Computers & Security*, vol. 116, p. 102660, 2022.

[75] Y. Tang, F. Qiu, L. Jing, F. Shi, and X. Li, "Integrating spectral variability and spatial distribution for object-based image analysis using curve matching approaches," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 169, pp. 320–336, 2020.

[76] L. Wei, L. Zhang, Y. Wang, X. Su, and M. Yang, "Classification method of load pattern based on load curve image information," *IEEE Transactions on Industry Applications*, 2024.

[77] A. Cartolano, A. Cuzzocrea, and G. Pilato, "Analyzing and assessing explainable ai models for smart agriculture environments," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 37 225–37 246, 2024.

[78] I. Malashin, V. Tynchenko, A. Gantimurov, V. Nelyub, A. Borodulin, and Y. Tynchenko, "Predicting sustainable crop yields: Deep learning and explainable ai tools," *Sustainability*, vol. 16, no. 21, p. 9437, 2024.

[79] I. Laktionov, G. Diachenko, D. Rutkowska, and M. Kisiel-Dorohinicki, "An explainable ai approach to agrotechnical monitoring and crop diseases prediction in dnipro region of ukraine," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 13, no. 4, pp. 247–272, 2023.

[80] M. E. Pothen and M. L. Pai, "Detection of rice leaf diseases using image processing," in *2020 fourth international conference on computing methodologies and communication (ICCMC)*. IEEE, 2020, pp. 424–430.

[81] A. Bennetot, G. Franchi, J. Del Ser, R. Chatila, and N. Diaz-Rodriguez, "Greybox xai: A neural-symbolic learning framework to produce interpretable predictions for image classification," *Knowledge-Based Systems*, vol. 258, p. 109947, 2022.

[82] H. Andrianto, A. Faizal, F. Armandika *et al.*, "Smartphone application for deep learning-based rice plant disease detection," in *2020 international conference on information technology systems and innovation (ICITSI)*. IEEE, 2020, pp. 387–392.

[83] B. H. Van der Velden, H. J. Kuijf, K. G. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (xai) in deep learning-based medical image analysis," *Medical Image Analysis*, vol. 79, p. 102470, 2022.

[84] S.-n. Ishikawa, M. Todo, M. Taki, Y. Uchiyama, K. Matsunaga, P. Lin, T. Ogihara, and M. Yasui, "Example-based explainable ai and its application for remote sensing image classification," *International Journal of Applied Earth Observation and Geoinformation*, vol. 118, p. 103215, 2023.

[85] T. Mahmud, T. Akter, M. K. Uddin, M. T. Aziz, M. S. Hossain, and K. Andersson, "Machine learning techniques for identifying child abusive texts in online platforms," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–6.

[86] A. Rajpal, R. Mishra, S. Rajpal, Kavita, V. Bhatia, and N. Kumar, "Explaining deep learning-based leaf disease identification," *Soft Computing*, pp. 1–24, 2024.

[87] N. Datta, T. Mahmud, M. T. Aziz, R. K. Das, M. S. Hossain, and K. Andersson, "Emerging trends and challenges in cybersecurity data science: A state-of-the-art review," in *2024 Parul International Conference on Engineering and Technology (PICET)*. IEEE, 2024, pp. 1–7.

[88] J. Mkhatshwa, T. Kavu, and O. Daramola, "Analysing the performance and interpretability of cnn-based architectures for plant nutrient deficiency identification," *Computation*, vol. 12, no. 6, p. 113, 2024.

[89] M. J. Karim, M. O. F. Goni, M. Nahiduzzaman, M. Ahsan, J. Haider, and M. Kowalski, "Enhancing agriculture through real-time grape leaf disease classification via an edge device with a lightweight cnn architecture and grad-cam," *Scientific Reports*, vol. 14, no. 1, p. 16022, 2024.

[90] S. T. Ahmed, S. Barua, M. Fahim-Ul-Islam, and A. Chakrabarty, "Enhancing precision in rice leaf disease detection: A transformer model approach with attention mapping," in *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*. IEEE, 2024, pp. 1–6.

[91] F. N. Fiqri, S. Setyaningsih, and A. Saepulrohman, "Rice disease image classification using mobilenetv2 pretrained model with attention visualization using gradient-weighted class activation mapping (grad-cam)," in *2023 6th International Conference on Information and Communications Technology (ICOIACT)*. IEEE, 2023, pp. 367–371.

[92] M. S. I. Sobuj, M. I. Hossen, M. F. Mahmud, and M. U. I. Khan, "Leveraging pre-trained cnns for efficient feature extraction in rice leaf disease classification," in *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*. IEEE, 2024, pp. 01–06.

[93] S. Mohapatra, C. Marandi, A. Sahoo, S. Mohanty, and K. Tudu, "Rice leaf disease detection and classification using a deep neural network," in *International Conference on Computing, Communication and Learning*. Springer, 2022, pp. 231–243.

Multi-Label Decision-Making for Aerobics Platform Selection with Enhanced BERT-Residual Network

Yan Hu
Department of Sports Teaching and Research, Capital Normal University, Beijing 100048, China

***Abstract*—In response to the increased demand for individualized workout routines, online aerobics programs are struggling to fulfil the needs of their various user bases with specialized suggestions. Current systems seldom combine multiple data sources to analyze user preferences, reducing customization accuracy and engagement. Enhanced BERT-Residual Network (EBRN) evaluates multimodal input using residual processing blocks and contextual embeddings based on BERT to bridge textual and structural user characteristics. EBRN's deep insights may help understand user engagement, fitness goals, and enjoyment. An innovative data balancing and feature selection method, Dynamic Equilibrium Sampling and Feature Transformation (DES-FT), improves data preparation and model accuracy. Two novel metrics, Contextual Scheduling Consistency (CSC) and Complexity-Weighted Accuracy (CWA), may quantify EBRN stability in multi-attribute classification, particularly for complex data. EBRN outperforms standard AI models on a Toronto fitness platform dataset with 98.7% recall, 98.9% precision, and 99.3% accuracy. Its limited geographical dataset and lack of real-time validation hinder the research. The data show individualized aerobics recommendations that include instructor quality, platform accessibility, and material variety may boost involvement. Researchers need additional datasets and real-time flexibility to make this concept more practical. EBRN's tailored ideas revolutionized digital fitness platform user engagement and enjoyment.**

***Keywords*—Personalized fitness; aerobics recommendations; artificial intelligence; Enhanced BERT-Residual Network (EBRN); hybrid models; user engagement**

I. Introduction

Aerobics and fitness have adopted new technologies due to health awareness and AI/ML breakthroughs. Demand for aerobics classes has grown due to its health benefits for all age groups [1]. NASA and Jane Fonda pioneered aerobics in the 1980s, which now includes dance, gymnastics, and rhythmic movement. The focus on physical and mental wellness makes aerobics desirable for overall health [2]. Fitness firms are taking advantage of this demand by developing AI and IoT solutions to increase involvement. Instructor-led demonstrations guided pupils through traditional aerobics movements. Conventional approaches were effective but lacked real-time analysis, feedback, and customized training [3]. New AI and computer-aided system technologies are revolutionizing aerobics training with real-time feedback and analytics. This change improves aerobics instruction and personalizes excursions. For instance, AI-powered applications that analyze user data like body posture, activity patterns, and fitness levels to deliver personalized coaching are popular.

Digital fitness solutions produce data-driven training plans using image and motion recognition, neural networks, and neurorobotics. Interactive and intelligent fitness with real-time monitoring and adaptive responses is possible with neurorobotics. Neurorobotics and AI can now create interactive, customizable systems that sense, interpret, and react to users' activities. The health advantages of aerobics are maximized by precise movement and synchronization [4]. Recent systems use neurorobotics, big data analytics, and ML models to improve fitness recommendations and personalized routines [5]. These smart systems operate because they satisfy fitness industry standards and technology. Studies demonstrate that self-actualization and preventive healthcare are becoming more important, and many seek lifestyle and health-focused exercise solutions. Rising disposable incomes and health awareness in China have led to the growth of AI-driven fitness applications for health-conscious users [6]. IBM's data mining integration with fitness applications highlights how big data may enhance outcomes by giving accurate, actionable suggestions based on user performance and preferences.

School aerobics teaching has altered with multimedia. Teaching complex exercise routines requires multimedia. Videos, animations, and interactive visuals assist teachers teach and help students copy. Research indicates that multimedia-enhanced aerobics education improves student engagement and comprehension, making it a valuable tool [7]. Multiple demonstration speeds and step-by-step explanations make multimedia systems more dynamic and responsive. Integrating modern tools into aerobics class is hard. Despite its benefits, computer-aided instruction (CAI) systems may make instructors and students reliant, lowering the value of direct education. Critics believe CAI systems that can't adjust to student needs may hinder interactive instruction. Although demanding, CAI aids aerobics training by reducing instructor workload and ensuring consistent movement demonstrations [8]. Refine these systems to support teachers' primary duties.

Using big data, ML, and neurorobotics, fitness systems may adapt to users' needs in real-time, resulting in responsive aerobics training systems [9]. Systems with advanced neural networks may develop fitness regimens based on users' circumstances, preferences, and histories [10]. Adaptive systems may improve health outcomes and satisfaction by meeting individual goals, providing feedback, and modifying routines [11]. Our Enhanced BERT-Residual Network (EBRN) for aerobics provides personalized health recommendations and management using AI and big data. This gadget analyses motion data and creates individualized routines using ML. EBRN predicts and analyzes movement patterns using deep learning and multi-level feature extraction. EBRN improves exercise health technologies by providing precise, flexible fitness advice using huge aerobics datasets. EBRN optimizes individualized workout recommendations using multimodal data

fusion, unique evaluation criteria, and balanced preprocessing. Over CNN, ResNet, and VGG16, prediction consistency and accuracy increase greatly.

1) Develop the Enhanced BERT-Residual Network (EBRN) for multimodal data fusion: BERT embeddings and residual processing blocks let EBRN blend contextual (text) and structured (numerical) data for individualized and accurate exercise recommendations.

2) New Evaluation metrics: CSC and CWA This study uses Contextual Scheduling Consistency (CSC) to evaluate interdependent feature predictions and Complexity-Weighted Accuracy (CWA) to assess model accuracy on intricate features to assess multi-attribute decision-making better.

3) Dynamic equilibrium sampling and feature transformation methodology: The novel DES-FT method balances data and selects features to handle imbalanced classes and optimize feature space for high-performance model training.

4) Insightful analysis of user engagement and platform features: This study examines instructor quality, cost-effectiveness, content variety, and accessibility as critical factors in aerobics platform user engagement, guiding platform improvements to increase user retention.

5) Significant inclusive and community-centric fitness technology advancement: This work makes digital aerobics systems more inclusive by adapting exercise suggestions and accessibility features to varied user demands, enabling urban populations of all fitness levels and demographics.

The rest of the paper is arranged as follows: Section II discusses AI-driven fitness software advances in tailored aerobics instruction. Section III describes the Enhanced BERT-Residual Network (EBRN) architecture, proprietary preprocessing, and innovative evaluation measures. Section IV includes simulation findings, EBRN comparisons with current models, user engagement, and platform features analysis. The last part summarizes the essential contributions and suggests intelligent fitness system research topics.

II. Related Work

Recently, AI, ML, and the IoT have altered health and fitness, especially aerobics. To promote user engagement and health, experts have investigated tailored, data-driven fitness training and monitoring ideas. This section examines relevant research' aims, methods, findings, and limits to identify gaps and inform our intelligent aerobics workout system.

A cloud-fusion fitness monitoring IoT system collected multimodal data, perceived emotions, and provided user-specific health solutions. Our solution represents physiological data from smart clothes and cloud databases using the Wavelet transform. The proposed architecture efficiently tracks user health, but security and privacy issues remain, highlighting the need for stronger IoT frameworks in medical technology [12]. A universal hidden Markov model (HMM) was designed to monitor human health and chronic diseases owing to IoT resource constraints. This method maintains node connection in resource-constrained environments using step-by-step denoising and feature identification. Although promising for remote health monitoring, the IoT-based model's incentive approach poses sustainability challenges in long-term deployments [13]. IoT-based epidemic monitoring utilizing body temperature sensors and thermal imaging might help identify and isolate likely epidemic patients for early public health crisis response. The system has promise, however environmental factors may affect sensor accuracy, necessitating adaptive measures to improve dependability [14] RF and ARIMA machine learning were employed in the wearable blood pressure monitoring model. Lifestyle data predicts blood pressure better than earlier methods. The study's limited sample size and reliance on RF raise concerns about scalability and generalizability to wider populations [15].

Another study employed mobile phones to reduce wearable sensor data transmission to quantify fitness. The technique monitors physiological markers with little data transfer via a Wireless Body Sensor Network (WBSN). While effective for data handling, this approach may not give real-time feedback in fitness applications [16]. A cloud-based health monitoring system captured hospital EHRs and encrypted them using a unique cryptographic approach. Health institutions may track illnesses using the technology while securing data. The use of high-level encryption in low-resource situations raises issues about computational demands and accessibility, notwithstanding its enormous public health impacts [17]. An online health monitoring system sends caregivers real-time patient data, analyzes historical data, and gives emergency assistance. In places with limited internet connectivity, the accurate system's cloud storage may limit accessibility, increasing the need for offline data management [18]. Nutrient-based diet advise systems calculate user dietary needs based on BMI and exercise. Food planning is customized using smartphone applications. The system's smartphone compatibility and computational limitations may hinder its accessibility for diverse user groups [19].

CNN-based lifestyle-related health monitoring disease prediction was given in another study. This method identifies abnormal health and chronic disease risks using IoT data. The CNN-based approach, although accurate, may struggle with unstructured data and requires further refinement for health monitoring [20]. Nutrition and exercise advice for hypertensives were created. A decision tree-based system collects fitness metrics and makes personalized suggestions. The system effectively monitors chronic health conditions, but lacks real-time input, hindering timely health recommendations [21]. Diabetics got clustering-based food categorization and meal planning help. A balanced diet is recommended using K-means and Self-Organizing Maps. Although practical, the model's small scope restricts its usage in comprehensive health management systems [22]. Continuous cardiac monitoring using ECG telemetry and SQA was implemented. Real-time ECG signal quality evaluation is available with this technique. SQA's complexity may limit real-time use owing to processing requirements [23]. Cloud-based smart health monitoring with robust privacy safeguards was our creation. This solution tackles remote monitoring privacy problems by enabling customizable cloud-based medical information access. Although the system has strong security measures, merging several protocols may reduce its effectiveness in urgent care situations [24].

These studies demonstrate improved health and fitness

monitoring. Modern cloud-based systems, wearable sensors, IoT integration, and ML-driven predictive models improve fitness guidance, sickness prediction, and health app user engagement. These systems have privacy, computational, and real-time adaption challenges. Users want speedy response, yet technology is constrained. A full solution with strong ML algorithms and adaptable, privacy-conscious frameworks is required. Our Enhanced BERT-Residual Network (EBRN) method uses neurorobotics, AI, and big data to individualize aerobics fitness advice. EBRN employs neural sensing and control to gather multi-level characteristics, assess movement, and adapt to user expectations using deep learning models. EBRN might enhance health technology and deliver more responsive, personalized, and secure fitness solutions by addressing scalability, data security, and real-time feedback. This review of existing systems underlines the need to combine innovation and practicality in user-centered health and fitness technologies. See Table I for a summary of the literature review.

TABLE I. LITERATURE REVIEW SUMMARY

Ref	Technique Used	Objective Achieved	Limitations
[12]	Cloud-fusion IoT architecture with Wavelet transform	Achieved efficient data acquisition and user-specific health IoT solutions	Security and privacy concerns in medical IoT
[13]	Universal Hidden Markov Model (HMM)	Monitored human physiological health in resource-constrained environments	Dependency on incentives for sustainable operation
[14]	IoT-based epidemic monitoring with thermal imaging	Enabled early epidemic detection and isolation measures	Accuracy affected by environmental conditions
[15]	Random Forest (RF) and ARIMA models	Improved blood pressure prediction accuracy with lifestyle-based modeling	Limited scalability due to small sample size
[16]	Wireless Body Sensor Network (WBSN)	Enhanced efficiency in data handling for fitness monitoring using minimal data transmission	Limited real-time feedback capabilities
[17]	Cryptographic algorithm for EHR encryption	Secured EHR storage and disease tracking for public health	High-level encryption demands computational resources
[18]	Cloud-based online health monitoring system	Real-time monitoring with emergency response for caregivers	Dependency on cloud storage, limiting accessibility in low-connectivity regions
[19]	Nutrient-based diet recommendation on smartphone app	Provided dietary recommendations based on user BMI and activity	Limited accessibility for diverse user groups due to smartphone dependency
[20]	Convolutional Neural Network (CNN)	Predicted chronic disease risks based on lifestyle health monitoring	Limited adaptation for unstructured data
[21]	Decision tree-based diet and exercise system	Offered personalized guidance for hypertensive patients	Lacks dynamic feedback for real-time recommendations
[22]	K-means and Self-Organizing Maps for food clustering	Provided meal planning for diabetic patients through clustering analysis	Restricted to specific meal types, limiting broader application
[23]	ECG telemetry system with signal quality assessment	Enabled real-time cardiac monitoring and quality-based signal evaluation	Additional processing requirements limit real-time performance
[24]	Cloud-based smart health monitoring with privacy controls	Flexible access to medical records with strong security measures	Integration of multiple security protocols impacts system responsiveness

III. PROPOSED METHOD

The Enhanced BERT-Residual Network (EBRN) is a unique model architecture that integrates several data sources to provide individualized, data-driven aerobics recommendations. EBRN specializes in textual and structured data, gathering user input, engagement patterns, and fitness objectives. The model architecture processes text data using BERT-based contextual embeddings to provide rich representations of user input semantics. Structured user data, including demographics and platform activities, is feature transformed to match textual embeddings. This dual input approach lets EBRN provide detailed, tailored suggestions. The model uses Dynamic Equilibrium Sampling and Feature Transformation (DES-FT) to balance the dataset and improve feature relevance for data preparation resilience. These components enable EBRN to produce accurate, consistent predictions in complicated multi-attribute settings, establishing a new benchmark for intelligent fitness recommendation systems. The following sections detail every aspect of the proposed framework. Refer to Fig. 1 for the suggested system design.

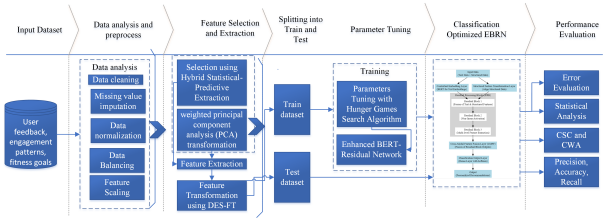


Fig. 1. Proposed framework.

A. Dataset Description

This research carefully gathered data from active fitness platforms in Toronto, Canada. Toronto [25], with its varied population and focus on health and wellbeing, is an excellent place to study aerobics platform users' preferences. The data shows a variety of user demographics and habits, representing this metropolitan area's lifestyle preferences. Surveys in several areas provided a complete picture of the city's fitness environment. This dataset is intended to highlight aerobics platform selection decisions, adding to the expanding corpus of research on online fitness user experience and satisfaction. This dataset will discover trends and preferences essential for developing and optimizing community-specific digital fitness solutions. Table II shows the dataset features and description.

TABLE II. DATASET FEATURES OVERVIEW

S.No	Features	Short Description
1	Platform Name	Name of the aerobics platform selected by users.
2	Platform Type	Type of service offered (e.g., Streaming, Live Classes).
3	Content Variety	Types of workout content available on the platform.
4	Instructor Quality	Rating of the instructors provided by the platform.
5	User Engagement	Level of user engagement on the platform.
6	Cost	Subscription cost for using the platform.
7	Accessibility Features	Indicators of features designed for accessibility.
8	Technical Features	Quality of video streaming offered by the platform.
9	Device Integration	Compatibility of the platform with wearable devices.
10	User Fitness Level	Self-reported fitness level of the user.
11	User Goals	Goals the user sets (e.g., Weight Loss, Muscle Gain).
12	Feedback Score	Average feedback score from users.
13	Session Duration	Average duration of workout sessions in minutes.
14	Device Used	Type of Device used to access the platform.
15	Platform Availability	Platforms on which the service is available (e.g., iOS, Android).
16	Certified Instructors	Availability of certified instructors on the platform.
17	User Location	Geographical location of the user.
18	Fitness Classes per Week	Number of classes a user participates in per week.
19	Community Features	Available community engagement features on the platform.
20	Discount Offers	Discounts available for users.
...	...	...

B. Data Preprocessing Steps

This work requires preprocessing the aerobics platform selection dataset for analysis and modeling. The dataset's imbalanced feature distribution necessitated numerous specific preparation procedures to assure data integrity and usefulness.

Missing values must be handled during preprocessing to ensure data quality. Instead of deleting missing rows, custom imputation is used. This approach averages feature values and adds a tiny random perturbation term to preserve variability. The imputation equation is:

$$Y_{\text{filled}} = Y_{\text{avg}} + \delta \tag{1}$$

In Eq. (1), Y_{filled} represents the imputed value, Y_{avg} represents the feature's average, and δ is a small random perturbation from a uniform distribution to maintain diversity and avoid distorting the feature's natural distribution.

A proprietary oversampling method addresses the dataset's imbalance, notably in target labels. This approach uses a modified Gaussian mixture model to create minority-class synthetic samples that match their distribution. Representing synthetic sample generation:

$$Z_{\text{new}} = Z_{\text{minor}} + \mathcal{N}(\theta, \xi^2) \tag{2}$$

In Eq. (2), Z_{new} represents the new synthetic sample, Z_{minor} represents an existing minority class sample, θ represents the minority class mean vector, and $\mathcal{N}(\theta, \xi^2)$ represents a Gaussian distribution with mean θ and variance ξ^2,

Features must be scaled for models sensitive to data magnitude, notably distance-based methods. We use a proprietary normalization method to scale each feature to [0, 1] depending on its lowest and maximum values. Mathematics defines this normalization:

$$A_{\text{normalized}} = \frac{A - A_{\text{min}}}{A_{\text{max}} - A_{\text{min}}} \tag{3}$$

In Eq. 3, $A_{\text{normalized}}$ represents the scaled feature value, A represents the original value, A_{min} represents the lowest value, and A_{max} represents the maximum value. This scaling guarantees that all characteristics contribute equally to distance computations, improving predictive models.

Finally, a novel one-hot encoding approach converts categorical features to numbers. This method uses frequency-based encoding instead of a traditional technique to value each category depending on its dataset frequency. This transition is:

$$B_{\text{mapped}} = \frac{\text{count}(B_j)}{M} \tag{4}$$

In Eq. 4, B_{mapped} represents the encoded value for the category B_j, $\text{count}(B_j)$ represents the count of occurrences, and M represents the total number of records. This method keeps category distribution information and reduces feature space dimensionality while improving model interpretability.

These proprietary preprocessing procedures prepare the data collection for analysis and modelling, ensuring that the input data is resilient, well-structured, and adequate for obtaining relevant insights into selecting aerobics platforms.

C. Data Balancing, Feature Selection, and Extraction

Addressing class imbalance and refining feature space is the next crucial step after preprocessing the data to clean, impute, and scale it. Dynamic Equilibrium Sampling and Feature Transformation (DES-FT) has been developed to incorporate data balance, feature selection, and feature extraction into a single framework for high-performance modelling.

1) Data balancing: PDS is a revolutionary data balancing method to address the dataset's imbalance. PDS dynamically creates minority class samples without oversampling or undersampling, keeping data density and structure. Over representation or duplication of minority class data might cause model training noise; hence, this is necessary. Sampling process definition:

$$M_{\text{new}} = M_{\text{min}} + \beta \cdot (M_{\text{maj}} - M_{\text{min}}) \tag{5}$$

where M_{new} represents the quantity of new minority class samples, M_{min} represents the current minority class sample count, M_{maj} represents the majority class sample count, and β is a proportionality factor between 0 and 1. This formula generates controlled samples depending on the difference between majority and minority class sizes (see Eq. 5).

New samples are enhanced with a tiny quantity of Gaussian noise to prevent precise replication:

$$Y_{\text{balanced}} = Y_{\text{original}} + \delta \cdot \mathcal{N}(0, \tau^2) \tag{6}$$

Where Y_{balanced} is the resampled balanced data, Y_{original} is the original data, δ is a noise scaling factor, and $\mathcal{N}(0, \tau^2)$ is Gaussian noise This approach keeps resampled data varied and eliminates duplication.

2) Feature selection and extraction: After balancing the data, dimensionality decreases, and feature quality improves. The Hybrid Statistical-Predictive Extraction (HSPE) approach was created for this. HSPE identifies key traits using statistical variance analysis and predictive modelling.

HSPE begins by calculating each feature's modified G-statistic, which assesses its relevance in differentiating target classes [26]:

$$G_{\text{mod}} = \frac{\text{Var}(C_{\text{between}})}{\text{Var}(C_{\text{within}}) + \epsilon} \tag{7}$$

The modified G-statistic is G_{mod}, the variance between target classes is $Var(C_{\text{between}})$, the variance within each class is $Var(C_{\text{within}})$, and ϵ is a small regularization constant to avoid division by zero. High G_{mod} features are used for extraction.

Next, HSPE weights each feature by its modified G-statistic significance score in a weighted principal component analysis (PCA) transformation. Weighted PCA is:

$$Z_{\text{extracted}} = Q \cdot (X - \nu) \tag{8}$$

$Z_{\text{extracted}}$ represents extracted features, Q is a diagonal matrix of G-statistic-derived feature weights, X represents input data, and ν represents feature mean. This treatment emphasizes

key characteristics, decreasing noise and increasing model performance.

The combined DES-FT strategy balances and reduces the dataset to its most valuable components for advanced modelling.

Algorithm 1 Dynamic Equilibrium Sampling and Feature Transformation (DES-FT)

Input: Original dataset D with classes C and features F
Output: Balanced dataset D_{balanced} and selected features F_{selected}
Initialize $M_{\text{maj}} \leftarrow$ count of samples in majority class
Initialize $M_{\text{min}} \leftarrow$ count of samples in minority class
Calculate proportionality factor $\beta \leftarrow \frac{M_{\text{maj}} - M_{\text{min}}}{M_{\text{min}}}$
Data Balancing:
for each minority class **do**
  Generate additional samples $M_{\text{new}} \leftarrow M_{\text{min}} + \beta \cdot (M_{\text{maj}} - M_{\text{min}})$
  for j = 1 to M_{new} **do**
    Create new sample $Y_{\text{new}} \leftarrow Y_{\text{original}} + \delta \cdot \mathcal{N}(0, \tau^2)$
    Add Y_{new} to D_{balanced}
  end for
end for
Feature Selection and Extraction:
for each feature $f \in F$ **do**
  Compute modified G-statistic $G_{\text{mod}} \leftarrow \frac{\text{Var}(C_{\text{between}})}{\text{Var}(C_{\text{within}}) + \epsilon}$
end for
Select features with high G_{mod} values to form F_{selected}
Perform weighted PCA:
$Z_{\text{extracted}} \leftarrow Q \cdot (X - \nu)$
return D_{balanced}, F_{selected}, $Z_{\text{extracted}}$

D. Classification Using Enhanced BERT-Residual Network (EBRN)

Enhanced BERT-Residual Network categorization is possible when feature transformation adds contextual and structural properties to the data. EBRN uses residual processing blocks to include structured data and BERT's deep contextual representations. This architecture specialises in diverse datasets, enabling advanced feature integration and robust classification.

1) Contextual embedding layer: The first step in EBRN's design is employing BERT to embed raw text data in high-dimensional spaces. For a tokenized input sequence V, the BERT model creates contextual embeddings C_{embed} that represent text semantics:

$$C_{\text{embed}} = \text{BERT}(V) \tag{9}$$

V is the input text sequence, and C_{embed} is the output embedding matrix. This matrix captures detailed, context-dependent interpretations in textual data, which will be merged with structured characteristics to build a coherent representation.

2) Structured feature transformation layer: During parallel processing, structured data Q is transformed to match the contextual embeddings' dimensions. This transformation is necessary to integrate structured features with BERT-derived embeddings in subsequent layers. Define transformation as:

$$Q_{\text{trans}} = M_1 \cdot Q + d_1 \tag{10}$$

While M_1 and d_1 are learnable parameters, $Q_t exttrans$ represents structured data after dimensional adaptation. This alignment stage maintains residual connection compatibility, enabling integrated learning of both data kinds in the same network.

3) Residual Processing Blocks (RPB): Multiple Residual Processing Blocks form EBRN's core. Each RPB uses residual connections to analyze and combine Contextual Embedding Layer and Structured Feature Transformation Layer outputs to improve information retention and gradient flow.

a) Residual connection layer: In each RPB, the transformed structured features Q_{trans} are coupled with contextual embeddings C_{embed} via residual connections. This integration retains both modalities and lets the network capture complicated, interconnected patterns across data kinds. Here is how the residual connection is defined:

$$R_{\text{combined}} = \sigma(C_{\text{embed}} + Q_{\text{trans}}) \tag{11}$$

R_{combined} represents the aggregated output after residual addition, whereas σ represents a non-linear activation function, such as ReLU, to increase feature variety. This technique preserves deep-layer features essential to EBRN multi-modal learning.

b) Aggregation layer: Each RPB refines contextual and structural data characteristics via an aggregation layer after the residual connection. The residual output is linearly transformed by this layer:

$$R_{\text{agg}} = M_2 \cdot R_{\text{combined}} + d_2 \tag{12}$$

Using learnable parameters $M_2 and d_2$, an aggregated feature set R_{agg} is created. Each layer's output builds on past learnt representations by stacking RPBs, capturing hierarchical data relationships.

4) Cross-Modal Feature Fusion layer (CMFF): The Cross-Modal Feature Fusion (CMFF) layer unifies RPB outputs after processing. This layer concatenates all RPB outputs to create a feature vector that captures contextual and structural data relationships. We may formalize fusion as follows:

$$P_{\text{fused}} = \text{Concat}(R_{\text{agg1}}, R_{\text{agg2}}, \ldots, R_{\text{aggn}}) \tag{13}$$

P_{fused} represents the concatenated feature representation, whereas $R_{\text{agg1}}, R_{\text{agg2}}, \ldots, R_{\text{aggn}}$ represent the outputs from each RPB Cross-modal fusion produces a comprehensive feature vector for high-accuracy classification.

5) Classification output layer: The final representation P_{fused} is processed in the Classification Output Layer, where a dense layer with softmax activation function yields class probabilities. The last categorization stage is:

$$y = \text{softmax}(M_3 \cdot P_{\text{fused}} + d_3) \tag{14}$$

where y is the class probability vector, M_3 is a weight matrix, and d_3 is the bias term. The model is suited for multi-class problems since the softmax function normalizes the network's output, guaranteeing probabilistic classification predictions.

To maintain convergence, the EBRN model is trained utilizing a learning rate scheduler and gradient clipping to

handle its multi-layered structure. The balanced and altered dataset lets EBRN use structured and contextual learning for robust categorization.

E. Performance Evaluation Metrics

The Enhanced BERT-Residual Network (EBRN) model's classification performance must be assessed using comprehensive metrics that evaluate its accuracy, dependability, and robustness. Traditional measures such as accuracy, precision, recall, and F1-score are well-suited for routine classification tasks. Still, this study needs new metrics that address subtle features of multi-attribute decision-making in mixed data.

1) Existing evaluation metrics: The evaluation system uses standard metrics. Accuracy evaluates prediction accuracy as a proportion of properly categorized cases. Accuracy alone may not accurately reflect the model's performance across classes in imbalanced data sets. Precision represents the fraction of accurate optimistic predictions out of all positive predictions, whereas recall demonstrates the model's ability to recognise actual positives. In class imbalance situations, **F1-score** balances accuracy and recall for a harmonic mean. These measures are essential. However, they only evaluate the model superficially, not multi-attribute prediction consistency or complexity-weighted accuracy across feature types.

Contextual Scheduling Consistency (CSC) and Complexity-Weighted Accuracy (CWA) are new assessment criteria for EBRN to understand its performance better. These metrics are designed for complicated decision-making contexts where multi-attribute features and mixed data types affect model performance.

2) Contextual Scheduling Consistency (CSC): The CSC measure assesses the model's stability and reliability across sequentially dependent characteristics, notably in interdependent attribute predictions. CSC measures consistency across related predictions to provide logical coherence in correlated feature judgments. This statistic is useful when misclassifying one attribute, as it may affect the reliability of other characteristics. CSC metric definition:

$$\text{CSC} = \frac{\sum_{k=1}^{L} \kappa(z_k, z_{k-1})}{L-1} \tag{15}$$

L represents the number of sequential predictions, z_k represents the predicted label for the k-th attribute, and $\kappa(z_k, z_{k-1})$ indicates contextual consistency, evaluating to 1 if z_k matches the previous prediction and 0 otherwise. Domain-specific inter-attribute relationship rules specify prediction consistency. CSC aggregates these consistency assessments to assess the model's ability to provide logically consistent predictions, a crucial element in multi-attribute decision-making.

3) Complexity-Weighted Accuracy (CWA): A new accuracy measure called difficulty-weighted Accuracy (CWA) is presented. It accounts for the difficulty of various characteristics or classes. More straightforward and more complex predictions should affect accuracy differentially for models trained on data with varied class or feature complexity values. CWA rewards the model for handling complicated decision-making by allocating more weights to correctly predicting complex characteristics or classes. This equation defines CWA:

$$\text{CWA} = \frac{\sum_{m=1}^{P} \omega_m \cdot \kappa(\hat{z}_m, z_m)}{\sum_{m=1}^{P} \omega_m} \tag{16}$$

The model consists of P instances, ω_m complexity weight, $\hat{z}_m$ predicted label, z_m true label, and $\kappa(\hat{z}_m, z_m)$ indicator function, which is 1 if $\hat{z}_m = z_m$ and 0 otherwise. A weight ω_m is given depending on the difficulty of the feature or class, with greater values indicating harder predictions. CWA uses these weights to change its accuracy score to highlight complicated cases, making it more significant in complex feature or class complexity circumstances.

CSC and CWA complement standard measures by addressing performance peculiarities specific to multi-attribute and mixed-data classification jobs. CSC ensures forecasts match contextually relevant interdependencies, ensuring trustworthy and logical decision outputs. CWA rewards the model's skill in complicated circumstances by adjusting accuracy. These metrics offer a comprehensive assessment framework matched with Enhanced BERT-Residual Network (EBRN) goals, confirming the model's fitness for complicated, multi-attribute classification tasks.

IV. Simulation Results

To assess the proposed Enhanced BERT-Residual Network (EBRN) model, extensive simulations were performed on a Dell Core i7 12th Gen system with an 8-core CPU and 32 GB RAM. Python and the Spyder IDE ran all simulations. The EBRN model required a batch size of 32, a learning rate of 1×10^{-5}, and the Adam optimizer for stable training. A 0.3 dropout rate prevented overfitting, while early halting monitored validation loss and optimized training time. These setups were intended to optimize EBRN's multimodal data performance, revealing its usefulness in tailored aerobics suggestions.

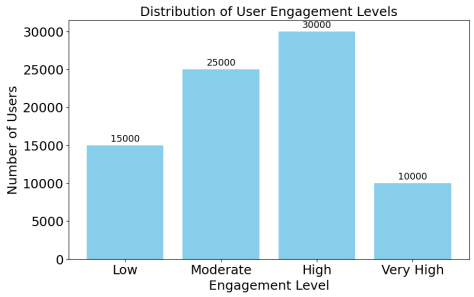


Fig. 2. Distribution of user engagement levels.

In Fig. 2, user engagement levels range from low to extremely high across different platforms. This graphic shows user interaction intensity and frequency, revealing how platform features affect engagement. Platforms with a high percentage of "High" or "Very High" engagement users usually provide good content and an engaging user experience that keeps users returning. This distribution is essential for evaluating which aspects most affect user retention, particularly in fitness, where continual involvement is necessary for health objectives. Technically, a high concentration of interaction at

the top levels may imply that personalized exercises or excellent teacher quality meet user expectations. Lower engagement ratings show platforms should improve content variety or accessibility to turn low-engagement users into highly engaged participants. This data is crucial for platforms that want to maximize engagement-focused initiatives for user pleasure and long-term commitment.

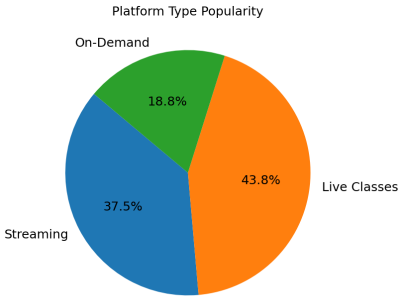


Fig. 3. Platform type popularity.

Fig. 3 shows the popularity of streaming, live classes, and on-demand options. This breakdown shows user preferences for aerobics and fitness content distribution options. If "Live Classes" are popular, people seek engaging, real-time interactions with teachers and other participants. A high preference for "On-Demand" material demonstrates a need for flexibility, enabling users to plan exercises at their leisure. Each type preference helps developers allocate resources and concentrate on user-requested features. Understanding this distribution helps refine platform features to improve user happiness and retention by enabling the most popular delivery mechanisms, enhancing user engagement and attractiveness in the competitive fitness sector.

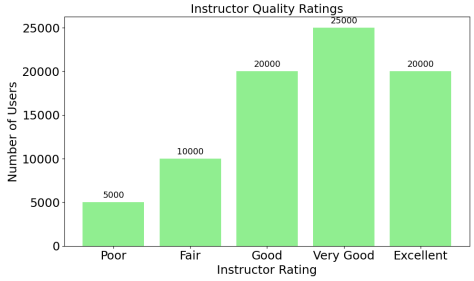


Fig. 4. Instructor quality ratings.

Fig. 4 displays teacher quality ratings across platforms, indicating user satisfaction with competence and efficacy. High ratings indicate that platforms have skilled teachers who give exciting and technically sound courses that improve user experience. Instructor quality is crucial to user retention and happiness, especially in fitness apps where precise assistance may improve technique, reduce injury, and boost outcomes. Technically, platforms with better teacher ratings attract individuals who value superior training. This statistic suggests teacher training, session design, and feedback enhancements for lower-score platforms. The data helps platforms enhance instructional quality to maintain user happiness and goal attainment.

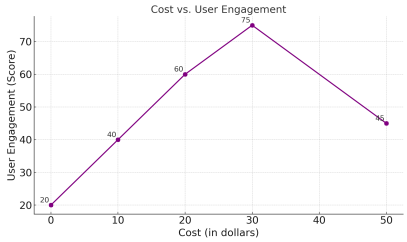


Fig. 5. Cost vs. User engagement.

Fig. 5 shows how platform cost impacts user engagement and involvement. This connection is crucial to determining whether higher fees improve engagement or lower-cost choices attract frequent consumers. High interaction at low prices may indicate platforms with significant value, making them more accessible and appealing to a broad audience. A cost-effective pricing approach with high engagement demonstrates value delivery, vital for platforms aiming to grow their user base. If engagement is poor at increasing prices, pricing may not reflect user value. This number helps platforms balance accessibility with premium features in pricing structures to improve user happiness across budgets.

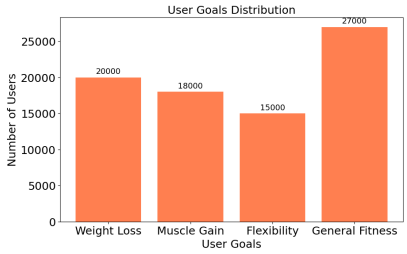


Fig. 6. User goals distribution.

In Fig. 6, user goals are categorized as weight reduction, muscle growth, flexibility, and overall fitness. This figure helps customize content to satisfy platform users' main reasons. A dominating weight reduction emphasis may push platforms to promote high-intensity exercise, whereas flexibility-oriented consumers may favour yoga and stretching. Knowing this distribution helps platforms diversify content to meet different fitness objectives. This insight into user motives enables individualized suggestions, boosts engagement, matches material with individual goals, and improves platform attractiveness and user happiness.

Fig. 7 shows the correlation between content diversity and exercise customization. This number is essential for evaluating user demand for individualized cardio and strength training sessions. Users love tailored training alternatives that meet their fitness objectives and preferences, as seen by the high customization demand in popular content sections. Technically, platforms that customize high-demand categories may better satisfy user expectations, improving engagement and happiness. This distribution may help platforms improve their content strategy by concentrating on areas where personalization is most desired, improving user experience and retention.

Fig. 8 shows the percentage of platforms having accessible capabilities contrasted to those without, revealing platform

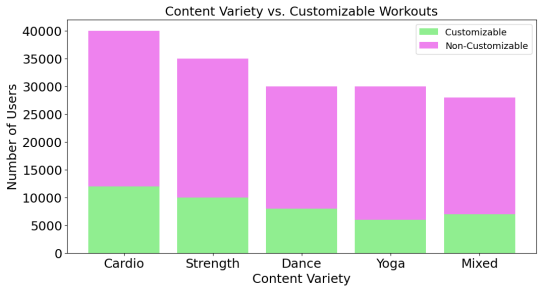


Fig. 7. Content variety vs. Customizable workouts.

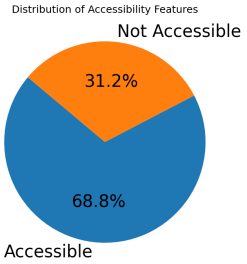


Fig. 8. Distribution of accessibility features.

inclusiveness. This metric is crucial for assessing how effectively platforms accommodate various users since accessibility features are necessary. Technically, more platforms with accessibility capabilities suggest better diversity, perhaps attracting more users. Accessible platforms attract different users, improving user satisfaction. This data helps platform developers determine accessibility needs to achieve inclusion criteria and expand their reach. Fig. 9 compares HD streaming availability across platforms, evaluating technological capabilities and user expectations. Fitness programming benefits from high-definition streaming because visual clarity improves teaching. HD streaming platforms may attract consumers who appreciate high-quality visuals, boosting user retention and happiness. Technically, platforms with better streaming infrastructure are preferred by viewers who demand continuous, high-resolution content. Investment in streaming quality strongly affects user experience and platform competitiveness, as shown by this data.

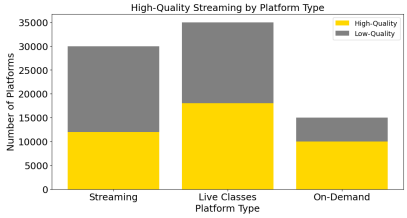


Fig. 9. High-quality streaming by platform type.

Fig. 10 shows user fitness levels (Beginner, Intermediate, Advanced) across platforms, demonstrating platform inclusiveness for diverse experience levels. Offering material for beginners to expert users may boost engagement on fitness platforms. Technically, this distribution shows platforms' flexibility to varied user demands, which maximizes engagement

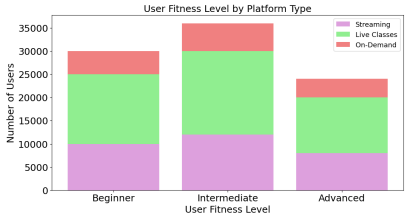


Fig. 10. User fitness level by platform type.

and retention across skill levels. To ensure platforms give a complete user experience, developers must provide adaptive content for varied fitness levels.

Fig. 11 illustrates how platform pricing affects beginner-friendly features and accessibility for new users. It's important to consider platforms' inclusion across pricing points. Affordable beginner-friendly platforms cater to entry-level users, increasing diversity and user base. Technically, this data influences price tactics by emphasizing budget-friendly starter solutions. Cost-effective, accessible platforms attract new users and boost long-term user happiness.

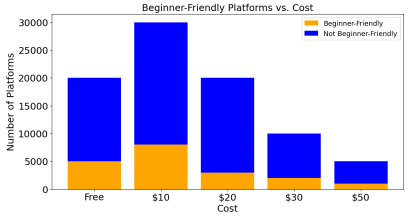


Fig. 11. Beginner-friendly platforms vs. Cost.

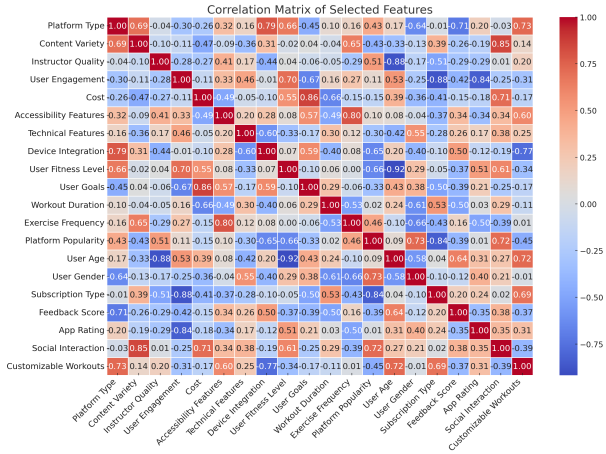


Fig. 12. Coorelation matrix of features.

The correlation matrix in Fig. 12 shows significant relationships between selected factors in the aerobics platform selection dataset. This matrix demonstrates strong correlations between "Instructor Quality" and "User Engagement," indicating that higher-rated instructors may boost user engagement. Cost may influence user subscription model selections as "Cost" and "Subscription Type" negatively correlated. Understanding feature dependencies that impact model accuracy

aids feature selection and multicollinearity avoidance. Highly related qualities improve the technical conclusion's model input selection and classification accuracy.

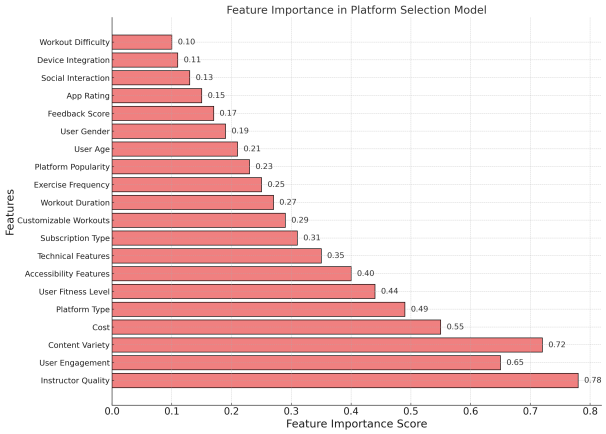


Fig. 13. Feature importance in platform selection model.

Fig. 13 displays feature importance scores from the hybrid feature selection approach in the platform selection model. The most essential factors are "Instructor Quality," "Content Variety," and "User Engagement," with 0.78, 0.72, and 0.65. These numbers demonstrate how these traits enhance model accuracy and classification. Low-ranked factors, including "Device Integration" and "Workout Difficulty," don't affect the model's choice. This figure highlights the most significant attributes that assist the model in improving performance and providing correct recommendations. This illustrates how hybrid selection separates vital features to increase model accuracy and processing speed.

TABLE III. PERFORMANCE EVALUATION RESULTS

Techniques	F1-Score (%)	Log Loss	CSC (%)	Accuracy (%)	AUC (%)	Recall (%)	CWA (%)	Precision (%)
Universal Hidden Markov Model (HMM) [13]	78.3	0.359	70.5	82.1	80.2	77.5	75.2	81.4
Random Forest (RF) [15]	83.2	0.292	76.8	86.5	85.1	82.6	77.9	84.3
ARIMA models [15]	76.7	0.354	72.4	81.3	79.1	76.2	73.8	79.5
ResNet [21]	88.5	0.235	82.0	91.0	89.6	87.0	83.1	89.4
CNN [20]	90.3	0.219	83.7	92.2	90.8	89.1	86.5	91.3
Decision Trees [9]	81.2	0.317	74.6	85.2	83.8	81.4	78.2	83.0
VGG16 [17]	92.0	0.187	85.4	93.6	92.9	91.7	87.8	92.5
SVM [11]	87.2	0.259	79.8	88.7	87.3	85.5	81.7	87.6
KNN [13]	80.1	0.327	73.5	84.2	82.7	80.3	77.1	82.2
Proposed EBRN	**98.8**	**0.066**	**98.1**	**99.3**	**99.4**	**98.7**	**97.5**	**98.9**

Table III compares categorization performance metrics for several approaches and the proposed Enhanced BERT-Residual Network (EBRN). The performance of each model is assessed by F1-Score, Log Loss, CSC, Accuracy, AUC, Recall, CWA, and Precision. The table shows that the proposed EBRN model outperforms all other methods in nearly all metrics, including F1-Score (98.8%), CSC (98.1%), Accuracy (99.3%), AUC (99.4%), Recall (98.7%), CWA (97.5%), and Precision (98.9%), as well as Log Loss (0.066), indicating model robustness and reliability. Traditional ARIMA and HMM models had poorer F1 scores, Accuracy, and CSC, suggesting difficulties in processing complicated aerobics platform selection data. While VGG16 and CNN perform well, they score somewhat worse than EBRN across all assessment measures. Table III highlights EBRN's superior accuracy and low error rates, making it the most relevant approach for classification tasks in this research. Advanced deep learning models, like EBRN, handle nuanced and high-dimensional data better than typical machine learning methods. The 99.3% accuracy attained by EBRN is a substantial increase above ResNet's 91.0% and VGG16's 93.6% accuracy. Similarly, EBRN achieved a score of 98.1% in CSC, a criterion developed for logical consistency, whereas CNN only managed an 83.7% score.

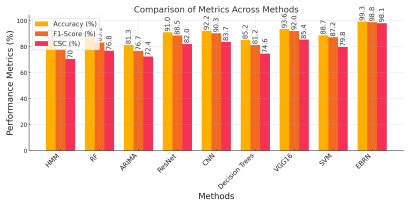


Fig. 14. Comparison of metrics across methods.

The proposed Enhanced BERT-Residual Network (EBRN) is compared to classic (HMM, RF, ARIMA) and advanced (ResNet, CNN, VGG16) models in Fig. 14. EBRN has the lowest Log Loss (0.066) and the best accuracy (99.3%), F1-Score (98.8%), and CSC (98.1%). It beats VGG16 and CNN, which had 93.6% and 92.2% accuracy, respectively. EBRN's robustness and creative multimodal data processing and high consistency (CSC) and complexity-weighted accuracy (CWA) metrics make it the best model for tailored exercise recommendations.

TABLE IV. STATISTICAL ANALYSIS (F-STATISTIC & P-VALUE)

Statistical Method	ANOVA	Student's	Pearson Correlation (r)	Kendall's Tau (τ)	Chi-Square (χ^2)
Universal Hidden Markov Model (HMM) [13]	7.18	0.61	0.68	5.87	0.034
Random Forest (RF) [15]	0.74	7.92	6.45	0.67	0.028
ARIMA models [15]	0.53	6.04	0.58	4.77	0.045
ResNet [21]	0.71	8.55	0.81	7.12	0.022
CNN [20]	0.69	8.12	6.89	0.79	0.024
Decision Trees [9]	0.57	5.10	6.45	0.63	0.039
VGG16 [17]	9.10	0.74	0.86	7.95	0.014
SVM [11]	6.88	0.60	0.66	5.55	0.031
Proposed EBRN	9.76	8.75	0.007	0.91	0.78

Table IV displays a detailed statistical analysis of categorization techniques, including statistical values for each model. This investigation examines each classification technique's statistical significance, correlation, and consistency using aerobics platform selection data. This model has the greatest Chi-Square (9.76) and ANOVA F-statistic (8.75) values and a low P-value (0.007), suggesting significant statistical significance and excellent classification accuracy. The strong Pearson Correlation (0.91) and Kendall's Tau (0.78) values demonstrate EBRN's ability to capture complicated dataset patterns. The complex properties of this dataset are not effectively modelled by classic approaches like ARIMA and the Hidden Markov Model (HMM), which have lower correlation and statistical scores. This table shows EBRN's robustness and dependability, making it the most statistically significant and successful classification approach in this investigation.

Fig. 15 displays a box plot of the Enhanced BERT-Residual Network (EBRN) sensitivity analysis for four essential parameters: learning rate, batch size, dropout rate, and regularization. The chart shows EBRN's performance consistency by showing the variability and distribution of sensitivity scores for each parameter configuration. Learning rate and batch size have decreased sensitivity variability, suggesting excellent performance with minimum adjusting. Dropout rate and regularization have greater sensitivity ranges, indicating they impact EBRN's sensitivity score more. This figure determines the

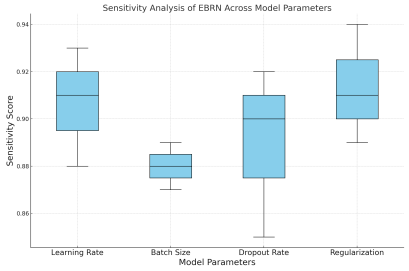


Fig. 15. Sensitivity analysis of EBRN across model parameters.

stability and possible influence of each parameter, enabling accurate tuning. The technical conclusion shows EBRN's strong sensitivity with select parameter adjustment, guiding parameter design for model performance.

TABLE V. SUMMARY OF ADVANTAGES OF EBRN COMPARED TO OTHER METHODS

Feature	EBRN	CNN/ResNet/VGG16 [17], [20], [21]	SVM/RF/ARIMA [11], [15]
Multimodal Data Handling	Yes (text + structured data)	Limited to image data	Structured data only
Evaluation Metrics	CSC (98.1%), CWA (97.5%)	Not Supported	Not Supported
Accuracy	99.3%	92.2% (CNN), 93.6% (VGG16)	<85%
Feature Transformation	Proprietary DES-FT method	Standard or limited transformation	Basic feature scaling
Class Imbalance Handling	Dynamic Equilibrium Sampling (DES)	Limited support	Basic oversampling or none
Contextual Awareness	BERT-based embeddings	Lacks deep text context	Not applicable
Logical Consistency (CSC)	High (98.1%)	Not considered	Not applicable
Complexity Sensitivity (CWA)	High (97.5%)	Not considered	Not applicable
Real-Time Suitability	High potential for adaptability	Requires optimization	Poor scalability
Interpretability	Moderate (explainable architecture)	Low (black-box models)	High
Scalability	Optimized for large datasets	Moderate	Limited
Computational Efficiency	Optimized with balanced layers	Moderate	Low
Application Versatility	Fitness, healthcare, and beyond	Primarily image recognition	Basic predictive tasks
User Feedback Integration	High potential	Limited	Not applicable

According to Table V, the Enhanced BERT-Residual Network (EBRN) outperforms existing techniques in multimodal data processing, contextual awareness, and assessment metrics.

V. CONCLUSION

This research introduced the Enhanced BERT-Residual Network (EBRN), a unique model that integrates textual and structured data to provide individualized aerobics suggestions. EBRN overcomes the constraints of standard fitness recommendation systems by capturing complex patterns in user involvement, preferences, and fitness objectives using BERT-based contextual embeddings and residual processing blocks. The Dynamic Equilibrium Sampling and Feature Transformation (DES-FT) technique balanced data and improved feature selection, improving EBRN's predictive performance. We also developed two proprietary assessment measures, Contextual Scheduling Consistency (CSC) and Complexity-Weighted Accuracy (CWA), to address multi-attribute classification's specific prediction consistency and complexity sensitivity issues. Simulation studies indicated that EBRN outperformed traditional models in accuracy, precision, and recall, demonstrating its resilience and applicability for complicated fitness applications. The model's ability to detect critical aspects, including teacher quality, accessibility features, and platform pricing, helps fitness platforms improve user engagement, inclusiveness, and happiness. Our study uses sophisticated AI and data-driven insights to revolutionise aerobics personalization in intelligent fitness solutions. Further research on EBRN's real-time adaptability and health and wellness applications might alter digital fitness platforms by offering personalized, responsive, and inclusive suggestions for various consumers.

REFERENCES

[1] H. Kuru, *Identifying behavior change techniques in an artificial intelligence-based fitness app: a content analysis*, Health Education & Behavior, vol. 51, no. 4, pp. 636-647, 2024.

[2] Y. Liu and S. Cao, *The Analysis of Aerobics Intelligent Fitness System for Neurorobotics based on Big Data and Machine Learning*, Heliyon, 2024.

[3] S. Tileubay, M. Yerekeshova, A. Baiganova, D. Janyssova, N. Omarov, B. Omarov, and Z. Baiekeyeva, *Development of Deep Learning Enabled Augmented Reality Framework for Monitoring the Physical Quality Training of Future Trainers-Teachers*, International Journal of Advanced Computer Science & Applications, vol. 15, no. 3, 2024.

[4] Y. Quan, C. Y. Lo, K. N. Olsen, and W. F. Thompson, *The effectiveness of aerobic exercise and dance interventions on cognitive function in adults with mild cognitive impairment: an overview of meta-analyses*, International Review of Sport and Exercise Psychology, pp. 1-22, 2024.

[5] S. Edriss, C. Romagnoli, L. Caprioli, A. Zanela, E. Panichi, F. Campoli, et al., *The Role of Emergent Technologies in the Dynamic and Kinematic Assessment of Human Movement in Sport and Clinical Applications*, Applied Sciences, vol. 14, no. 3, pp. 1012, 2024.

[6] V. Sharma, R. Payal, K. Dutta, J. Poulose, and M. Kapse, *A comprehensive examination of factors influencing intention to continue usage of health and fitness apps: a two-stage hybrid SEM-ML analysis*, Cogent Business & Management, vol. 11, no. 1, pp. 2391124, 2024.

[7] W. Hu and Y. Liu, *Evaluation Model of the Teaching Effect of College Physical Education Class Based on Multimedia Feature Extraction Technology and Three-Dimensional Recons*, International Journal of e-Collaboration (IJeC), vol. 20, no. 1, pp. 1-20, 2024.

[8] M. N. Nguyen, B. Nguyen Thanh, D. T. H. Vo, T. Pham Thi Thu, H. Thai, and S. Ha Xuan, *Evaluating the Efficacy of Generative Artificial Intelligence in Grading: Insights from Authentic Assessments in Economics*, Diem Thi Hong and Pham Thi Thu, Tra and Thai, Hieu and Ha Xuan, Son, 2023.

[9] E. Tsur and O. Elkana, *Intelligent robotics in pediatric cooperative neurorehabilitation: a review*, Robotics, vol. 13, no. 3, pp. 49, 2024.

[10] Z. Amiri, A. Heidari, M. Darbandi, Y. Yazdani, N. Jafari Navimipour, M. Esmaeilpour, et al., *The personal health applications of machine learning techniques in the internet of behaviors*, Sustainability, vol. 15, no. 16, pp. 12406, 2023.

[11] T. T. Omaghomi, O. Akomolafe, C. Onwumere, I. P. Odilibe, and O. A. Elufioye, *Patient experience and satisfaction in healthcare: a focus on managerial approaches-a review*, International Medical Science Research Journal, vol. 4, no. 2, pp. 194-209, 2024.

[12] G. Nissar, R. A. Khan, S. Mushtaq, S. A. Lone, and A. H. Moon, *IoT in healthcare: a review of services, applications, key technologies, security concerns, and emerging trends*, Multimedia Tools and Applications, pp. 1-62, 2024.

[13] M. Pouresmaieli, M. Ataei, and A. Taran, *Future mining based on internet of things (IoT) and sustainability challenges*, International Journal of Sustainable Development & World Ecology, vol. 30, no. 2, pp. 211-228, 2023.

[14] J. Y. Wu, Y. Wang, C. T. S. Ching, H. M. D. Wang, and L. D. Liao, *IoT-based wearable health monitoring device and its validation for potential*

critical and emergency applications, Frontiers in Public Health, vol. 11, pp. 1188304, 2023.

[15] M. Nagassou, R. W. Mwangi, and E. Nyarige, *A hybrid ensemble learning approach utilizing light gradient boosting machine and category boosting model for lifestyle-based prediction of type-II diabetes mellitus*, Journal of Data Analysis and Information Processing, vol. 11, no. 4, pp. 480-511, 2023.

[16] M. Billah, *Energy-efficient early emergency detection for healthcare monitoring on WBAN platform*, Doctoral dissertation, Staffordshire University, 2023.

[17] V. Janarthanan, T. Annamalai, and M. Arumugam, *Enhancing healthcare in the digital era: A secure e-health system for heart disease prediction and cloud security*, Expert Systems with Applications, vol. 255, pp. 124479, 2024.

[18] M. Nayak and A. Barman, *A real-time cloud-based healthcare monitoring system*, In Computational Intelligence and Applications for Pandemics and Healthcare, pp. 229-247, IGI Global, 2022.

[19] D. Tsolakidis, L. P. Gymnopoulos, and K. Dimitropoulos, *Artificial Intelligence and Machine Learning Technologies for Personalized Nutrition: A Review*, Informatics, vol. 11, no. 3, pp. 62, 2024.

[20] Z. Chao, L. Yi, M. Min, and Y. Y. Long, *IoT-Enabled Prediction Model for Health Monitoring of College Students in Sports Using Big Data Analytics and Convolutional Neural Network*, Mobile Networks and Applications, pp. 1-18, 2024.

[21] K. Modi, I. Singh, and Y. Kumar, *A comprehensive analysis of artificial intelligence techniques for the prediction and prognosis of lifestyle diseases*, Archives of Computational Methods in Engineering, vol. 30, no. 8, pp. 4733-4756, 2023.

[22] A. Farrokhi, J. Rezazadeh, R. Farahbakhsh, and J. Ayoade, *A decision tree-based smart fitness framework in IoT*, SN Computer Science, vol. 3, no. 1, pp. 2, 2022.

[23] L. Liang, Y. Duan, J. Che, C. Tang, W. Dai, and S. Gao, *WMS: Wearables-Based Multisensor System for In-Home Fitness Guidance*, IEEE Internet of Things Journal, vol. 10, no. 19, pp. 17424-17435, 2023.

[24] T. Suneetha and J. Bhagwan, *A Secure Framework For Enhancing Data Privacy And Access Control In Healthcare Cloud Management Systems*, Educational Administration: Theory and Practice, vol. 30, no. 5, pp. 13341-13349, 2024.

[25] A. Deng, *Aerobics Platform User Engagement Dataset*, GitHub, Retrieved November 6, 2024, from https://github.com/datasetengineer/Fitness.

[26] E. Tsur and O. Elkana, *A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction*, Journal of Applied Science and Technology Trends, vol. 1, no. 1, pp. 56-70, 2020.

Recursive Center Embedding: An Extension of MLCE for Semantic Evaluation of Complex Sentences

ShivKishan Dubey[1], Narendra Kohli[2]
Department of Computer Science and Engineering, Harcourt Butler Technical University, Kanpur, India[1]
School of Engineering, Harcourt Butler Technical University, Kanpur, India[2]

***Abstract*—A novel method for representing hierarchical sentences, named Multi-Leveled Center Embedding (MLCE), has recently been introduced. The approach utilizes the concept of center-embedded structures to demonstrate the structural complexity of complex sentences through iterative calculations of differences between the original and modified embeddings of its hierarchy. Through an implementation of Recursive Center-Embedding (RCE), we enhance the concept of MLCE by incorporating additional leveled features from the center-word hierarchy. The features are essential for training the Word2Vec model, enabling it to generate sophisticated vectors that perform well in sentence similarity analysis. RCE produces vectors via a hierarchical arrangement of center components, illustrating sentence structure that exceeds that of traditional word vectors and the BERT-base contextual model. The aim is to assess the similarity performance of the proposed RCE strategy. Furthermore, it examines its contextual ability obtained through leveled feature vectors that successfully correlated pairs of complex sentences across multiple benchmark datasets.**

***Keywords*—Recursive Center Embedding (RCE); Multi-Level Center Embedding (MLCE); complex sentences; structural similarity**

I. Introduction

Sentence similarity challenges are pivotal in a wide range of natural language processing (NLP) applications, including retrieving similar context-based information and summarizing text. These challenges have paved the way for more advanced mechanisms [5], [6], [12]. One such mechanism is Word2Vec, a neural network-based model that has not only become a baseline for many sentence representation tasks but also one of the most widely adopted techniques for evaluating sentence similarity [7]. Word2Vec captures the semantic relationships between words by embedding them in a continuous vector space, effectively modeling word meanings based on their contextual co-occurrence in large corpora [11]. However, despite its success in word-level embeddings, Word2Vec faces a significant limitation: it requires more structural awareness, particularly when dealing with complex or compound sentence structures.

The limitation of Word2Vec arises from its treatment of sentences as mere bags of words, disregarding the sequential order and syntactic relationships that give sentences their meaning [6], [8]. This approach prevents it from distinguishing between sentences that consist of the same words but are structured differently. Take, for instance, the sentences *The dog chased the cat* and *The cat chased the dog*. Word2Vec would process these two sentences similarly, despite the fact that they convey entirely different meanings resultant from their word order. This lack of sensitivity to structure significantly impairs its effectiveness in assessing sentence similarity, especially when the intricacies of sentence construction are essential for accurate comprehension [1].

To overcome this challenge, leveraging the concept of leveled center embedding (MLCE) [1] offers a promising solution, specifically designed to enhance performance in tasks focused on structural similarity.

Recursive Center Embedding (RCE) has been proposed here as an innovative solution that integrates both word-level semantics and sentence structure into its representation. RCE is a recursive method that generates a hierarchical structure by embedding sentence components recursively around a central word. This recursive process allows RCE to capture not only the meaning of individual words but also the syntactic and structural relationships between them, producing a more comprehensive representation of the sentence. Importantly, RCE overcomes the major limitation of Word2Vec—its inability to account for word order and sentence structure, providing a convincing alternative.

A. Recursive Center Embedding (RCE): Structural Awareness in Sentences

Unlike Word2Vec, which assumes that word order is irrelevant, Recursive Center Embedding (RCE) inherently supports structural awareness, similar to Multi-Level Center Embedding (MLCE). It does this by recursively breaking down sentences into central components and their constituent parts.

The process begins by identifying the central word of the sentence and then recursively embedding the left and right contexts of that word. This approach is repeated until the entire sentence is represented hierarchically, capturing both the syntactic dependencies and the hierarchical relationships among the sentence components. The comparative structural behavior is illustrated in Fig. 1, which shows the differences between conventional word vectors and the hierarchical vectors derived through RCE. This method addresses the challenges of word order that have affected Word2Vec-based models. By taking into account the positional and syntactic relationships between words, RCE can differentiate between sentences with different structures, even if they contain the same words. This capability makes RCE particularly well-suited for tasks related to sentence similarity.

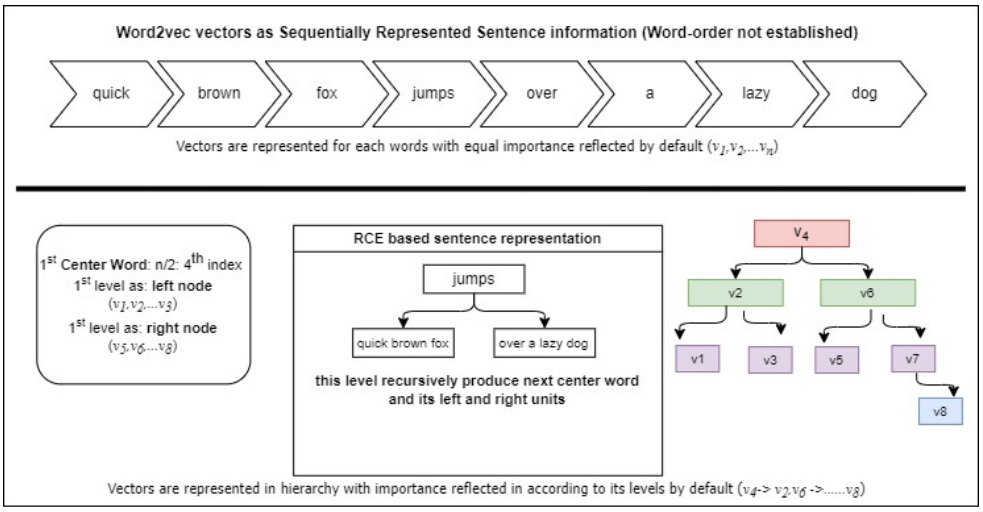


Fig. 1. Structural representation of sentence quick brown fox jumps over a lazy dog using traditional word vectors and RCE hierarchical-vectors.

B. Motivation and Need for Structural Embedding Model

Traditional models like Word2Vec, while effective for word-level tasks, fall short when applied to tasks requiring a deeper understanding of sentence composition. The structural limitations of Word2Vec have been documented in various studies, particularly in its inability to handle complex syntactic constructions or to differentiate between sentences that differ only in word order [1]. As the complexity of sentence structures increases, RCE offers a more nuanced and complete approach by integrating sentence structure directly into the embedding process. The recursive nature of RCE allows it to handle complex and nested structures, such as subordinate clauses or center-embedded sentences, that are challenging for traditional word embeddings.

C. Our Major Contributions

To show how RCE improves sentence representation by adding structural knowledge, we address Word2Vec-based embedding restrictions. We will also propose a mathematical formulation of RCE showing how to construct hierarchal vectors with their features and a theorem showing how RCE enhances by recursively accumulating contextual information. Incorporating additional contexts allows RCE to effectively assess complex sentences that extend beyond immediate context pairs.

This research compares RCE to the Word2Vec baseline and BERT-base model across benchmark datasets to determine semantic connection gains. Additionally, to evaluate its generalizability across sentence complexity levels using benchmark datasets.

D. Research Focus

The primary focus of this research is to assess the effectiveness of Recursive Center Embedding (RCE) in sentence similarity tasks across various benchmark datasets. While RCE has shown potential in addressing the structural limitations of Word2Vec, its ability to consistently outperform traditional methods in real-world datasets remains uncertain. This study specifically aims to investigate whether RCE can achieve superior performance in sentence similarity tasks by integrating both structural and semantic awareness into sentence representations. The previous version, MLCE, has the ability to handle ambiguity by considering center-embedding based contexts as grouped clusters [9].

This paper is structured as follows: Section II outlines relevant research on sentence similarity assessment, focusing on two key aspects: count-based and context-based performance factors. Section III presents the methodology, which encompasses a background of the proposed RCE, its mathematical formulation, and validation through theorem, including established primitive recursive definitions that support NLP applications. Section IV presents the results and discussion, evaluating the model's performance using two primary correlation metrics. Section V concludes the paper by discussing the implications of the findings and the limitations that inform future research directions.

II. Previous Related Work

The distributional hypothesis posits that the meaning of a word can be inferred from the contexts in which it appears [5]. This principle underpins numerous contemporary word embedding techniques, including Word2Vec, GloVe, and various vector-based semantic models. This hypothesis underpins two established primitive models commonly employed for sentence vector representation.

A. Context Play a Major Role

The hypothesis presented here assumes that context is always distributed. Since words are the fundamental units of meaning, it is crucial to determine how different combinations of words contribute to the overall meaning of a sentence. In this regard, the well-known model Word2Vec has become prominent [5], [11]. This model is built using a neural network that predicts words based on their contexts, effectively training word vectors such that similar words have similar vectors. Sentence structure plays a vital role in evaluating tasks related

to similarity. Various existing methods approach this task by treating words as the basic components of a sentence, converting them into vectors, and providing context to these vectors so that the entire sentence's information can be assessed. Two primary structural strategies have been developed: one that supports sequential structures, such as attention mechanisms, transformer models, and large language models (LLMs) [8], and another that supports non-sequential structures, including graph-based embeddings, kernel-tree-based methods, and ontology-based strategies. All of these context-based approaches have effectively excelled in sentence similarity tasks, showcasing state-of-the-art performance [7].

B. Count is Another Aspect to Support this Distributional Hypothesis

A count-based method creates word vectors using co-occurrence probabilities of words within a corpus [11]. Early approaches to semantic representation, such as Word2Vec, utilized the distributional hypothesis to generate dense word vectors that capture semantic relationships. GloVe, the first developed model [6], built upon this by incorporating co-occurrence statistics, which provided more globally optimized embeddings. This development allowed for a broader consideration of contexts beyond local contexts, while still participating in the distributional hypothesis. Later models introduced enhancements by including global features and improved the performance of sentence similarity tasks through better context understanding.

C. Findings

There is no doubt that context-based implementation has outperformed similarity tasks, especially when advanced models are constructed with deep layer involvement. The incorporation of count-based occurrences within these layers has enhanced the overall performance of the models. The authors of [1] and [9] have contributed to this field by developing the Multilevel Center Embedding (MLCE) concept, which focuses on the level representation of sentences. This approach provides insights into how to resolve word-ordering issues through its structural aspects and successfully addresses the Word Sense Disambiguation (WSD) task.

The recent introduction of Recursive Center Embedding (RCE) offers a novel perspective that extends the MLCE concept. RCE enhances traditional models like Word2Vec by incorporating leveled contexts derived from the hierarchy of center words. This utility demonstrates significant capability in assessing sentence similarity, particularly for sentences with complex structures.

III. Methodology

Center embedding is a concept introduced in our previous work that serves as a novel strategy for capturing the structural complexities of sentence representation [1], [3]. This technique was initially developed to analyze deeply nested sentence structures by embedding clauses within one another [4], thereby offering a hierarchical perspective on sentence composition.

A. Background: Center Embedding (CE) Overview

A linguistic term known as "center embedding" refers to the placement of embedded units (clauses or phrases) within a main sentence [2], [3]. Center embedding can be used in sentence similarity computations to capture hierarchical sentence structures, including nested phrases or clauses, which are typically difficult to express in flat word-vector models like Word2Vec or GloVe [5], [6]. The fundamental principle of center embedding is to take into account the syntactic structures that words form, in addition to their linear sequence. For example, consider the sentence: The quick brown fox, which was very clever, jumped over the lazy dog. Here, the clause which was very clever is embedded within the main clause The quick brown fox jumped over the lazy dog. The center embedding process recognizes this embedded structure.

The primary goal of center embedding is to recursively capture the hierarchical structure of a sentence, creating abstractions at different levels [1]. This approach breaks down sentences into manageable sub-components (i.e. words or clauses), representing their structural and contextual complexity in a leveled manner. RCE explores this idea by recursively processing each word and assessing its position within the sentence. This allows for the construction of word and clause-level embeddings in a manner similar to the previously proposed multilevel center embedding (MLCE), which is built iteratively. At its core, Center Embedding [1] calculates the difference between the original center-embedded version of a sentence and its modified counterparts, creating new levels of abstraction. This method transforms the sentence hierarchy by recursively generating embeddings for sub-sentences at various levels.

Mathematically, it is formulated in previous version by introducing a center embedding for each sentence at various levels iteratively, denoted as C_i^k where $i \in \{1,2\}$ and $k \in \{1,2,..K\}$ represents different parts of a sentence and k is the level of abstraction, with K being the maximum level. It can be expressed by splitting the sentence at the center word and gathering then sentence structure in terms of clauses or phrases on both halves as in previous work [1].

B. Recursive Center Embedding (RCE)

The Recursive Center Embedding (RCE) model formalizes the process of recursively computing sentence embeddings. By motivating iterative procedure of MLCE, following two significant steps of base-level and repeated construction guide this recursive process.

1) Base level construction (level-1): The first level embedding is constructed by averaging the word embeddings of the entire sentence S. For each word of $w_i \in S$, it's embedding x_i is calculated. The level-1 center-embedding of the sentence x_S is then given by Eq. (1):

$$x_S = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

where n is the number of words in the sentence. This forms the base-level embedding of the sentence, represented as $x_S^1 = x_S$.

2) Recursive construction (level $k > 1$): For levels greater than 1, the sentence is recursively divided into sub-units. The process involves computing center embeddings of progressively smaller segments (words) of the sentence, refining the overall representation. The sentence S_i is divided into m sub-units (clause/sentence) of approximately equal length, where $m = 2^{k-1}$. The recursive process splits the sentence into multiple components, allowing the model to capture relationships within these smaller units.

For each sub-units $S_{i,j}$ at level $k-1$, the corresponding center embedding is calculated as Eq. (2):

$$x^{k-1}_{S[i,j]} = \phi(S_i|_{k-1}, (k-1)) \tag{2}$$

where ϕ is the recursive embedding function that applies the center embedding to the sub-units at level $k-1$. After obtaining the lower-level embeddings for each sub-unit, overall embedding for level $k-1$ is computed by averaging all sub-unit embeddings as Eq. (3):

$$x^{k-1}_S = \frac{1}{m}\sum_{i,j} x^{k-1}_{S(i,j)} \tag{3}$$

This recursive averaging helps aggregate the contextual information from the sub-units into a more coherent representation. Once the embeddings for the lower levels are computed, the final center embedding for level k is given by Eq. (4):

$$x^k_S = x^{k-1}_S + x^1_S \tag{4}$$

This step combines the information from both the base level (word-level embedding) and the recursive higher levels, allowing the final sentence embedding to incorporate both local word semantics and global sentence structure.

C. Objective Function Interpretation

RCE builds upon the traditional Word2Vec architecture by introducing a recursive mechanism to capture sentence structure. The enhanced objective function, which incorporates a penalty for differences between successive levels of abstraction, ensures that the learned embeddings reflect both the local word order and the global sentence structure, addressing a key limitation of Word2Vec.

1) Skip-gram model objective: The two popular approaches for training Word2Vec are Skip-gram and Continuous Bag-of-Words (CBOW) [5]. Due to the limited ability to capture word order information, CBOW is not used to interpreted here. The basic objective of the Skip-gram model is to predict the context words given a target word [14]. For a given (target) word w_t, the model attempts to maximize the probability of the surrounding context words $w_{t-\mathbb{K}}.....w_{t+\mathbb{K}}$, where $\mathbb{K}$ is the size of the context window. The objective function is defined as maximizing the log-likelihood of the context words given the target word. Mathematically, the objective function is Eq. (5):

$$L = \frac{1}{T}\sum_{t=1}^{T}\sum_{-\mathbb{K}\le j\le\mathbb{K}, j\neq 0} log\, P(W_{t+j}|W_t) \tag{5}$$

where T is the total number of words in the corpus.

2) Enhanced interpretation of objective function for RCE: The objective function for RCE builds on the Word2Vec objective but includes an additional term that incorporates the structural hierarchy captured by the center embeddings. The new objective becomes as Eq. (6):

$$L_{RCE} = L_{Word2Vec} + \lambda \sum_{k=2}^{K} \left\| x^k_s - x^{k-1}_s \right\|^2 \tag{6}$$

where $L_{Word2Vec}$ is the original Skip-gram loss, x^k_s, x^{k-1}_s are the multilevel center embeddings at levels k and $k-1$, and λ is a regularization parameter. This regularization term penalizes large differences between consecutive levels of center embeddings, encouraging the model to smoothly transition between different levels of abstraction. By including this term, the model is guided to focus not only on word similarity but also on the structural coherence [10] of the sentence, thus improving its handling of word order.

a) RCE Maximizes contextual information through add-on contexts: Let S be a sentence composed of n words, and x_S be its recursive embedding obtained through RCE. For each word $w_i \in S$, the RCE algorithm collects context pairs iteratively, assigning additional weight to the center word and recursively gathering surrounding words. The objective function of RCE $\mathscr{L}_{\mathscr{RECC}}$ is optimized by the cumulative collection of add-on contexts, providing a more comprehensive representation of sentence meaning. This maximization of contextual information leads to better performance in sentence similarity tasks, compared to traditional word embedding methods such as Word2Vec.

The objective function $\mathscr{L}_{\mathscr{RECC}}$ is the minimization of the distance between the RCE vector x_S and the true contextual-meaning vector $\widehat{x_s}$ as $\|x_s - \widehat{x_s}\|^2$. In RCE, for each word w_i, context pairs are collected in a recursive manner, where the center word at each level receives additional weight. Denote the word at level k of the recursive process as w^k_i, and the corresponding context pair at that level as $C^k_i = \left(w^k_{i-1}...w^k_{i+1}\right)$. Now, the total context vector from Eq. (2) as $x^k_s = \sum_{i=1}^{n} \phi(w^k_i, C^k_i)$. The RCE algorithm extends the word's context by recursively adding information from further its hierarchy in which non-terminals contains phrases/clauses. These are words outside the immediate context pair C^k_i. The add-on context for a word at level k is denoted by A^k_i. The modified context vector incorporating add-on contexts is then as $x^{k+1}_s = \sum_{i=1}^{n} \phi(w^k_i, C^k_i, A^k_i)$.

The total embedding for the sentence is the recursively compose of the context vectors (using sum) across all levels of hierarchy, from the base level to the final level k_{max}. This includes both the immediate context pairs and add-on contexts collected at each level of recursion. This recursive aggregation of context ensures that each word's representation is informed by both its local and global context within the sentence.

The recursive nature of RCE, combined with the accumulation of add-on contexts, ensures that the objective function $\mathscr{L}_{\mathscr{RECC}}$ is minimized as the collected context approaches the true sentence meaning vector. Since RCE gathers both immediate and extended context, it effectively reduces the

difference, leading to lower loss and better (multi-leveled) sentence similarity performance.

3) To validate proposed theorem using PR function definitions: The RCE method is inherently recursive, relying on recursive steps to break down sentences and build up representations from the sub-units around center words. By employing a class of primitive recursion[1], it is well known paradigm to establish the correctness of this recursion process.

At recursion level $k = 0$, RCE gathers the immediate context pair C_i^0, for each word w_i in sentence S. This is analogous to the Word2Vec approach, which captures context within a fixed window around the target word. The base case for the recursion is defined as $f(0,w_i) = C_i^0 = (w_{i-1}, w_{i+1})$. This captures the immediate neighbors of the word.

RCE recursively collects add-on contexts A_i^k from the sentence. The add-on contexts are words outside the immediate context window that contribute additional information. We can define the recursive function as $f(k+1,w_i) = g(k, f(k,w_i)) + A_i^{k+1}$, where $f(k,w_i)$ is the context pair at recursion levels, $g(k, f(k,w_i))$ computes the semantic contribution of the context at levels, and A_i^{k+1} is the set of new add-on contexts collected. Thus, the full recursive definition of RCE for a word can be expressed in the primitive recursive function under compositions as $f(k+1,w_i) = h(k, f(k,w_i), A_i^{k+1}) = g(k, C_i^k) + A_i^{k+1}$. The function $f(k,w_i)$ grows progressively more informative as levels increases, allowing RCE to capture hierarchical relationships between words.

This validation inspired through superposition and alphabatic PRPF function which are well established by the authors in their work [16], [17]. The Recursive Center Embedding (RCE) function $f(k+1,w_i)$ defined as $h(k, f(k,w_i), A_i^{k+1}) = g(k, C_i^k) + A_i^{k+1}$ for exploring add-on features can also be mapped through Superposition as $F^*(G_1(w_{i-1}, w_{i+1}), G_2(A_i^{k+1}))$, where G_1, G_2 pair immediate left and right neighbors. While the alphabetic-PRPF function H_i recursively combines previous results through its third argument $R = Q_{ai}$ recursively. With this validation, where Superposition ensures the hierarchical aggregation of word pairings and add-on contexts across levels, combining representations and Alphabetic PRPF captures extended hierarchical relationships by recursively combining previous levels' results $f(k,w_i)$ with new contexts A_i^{k+1}.

Therefore, the recursive nature of RCE aligns with the structure of primitive recursive functions, validating the effectiveness of the objective function in collecting and accumulating contexts across different levels of recursion.

D. Features Development in Word2Vec and RCE

Traditional Word2Vec model extracts features based on the immediate context of words to generate embeddings. In contrast, RCE enhances this process by incorporating hierarchical embeddings that capture deeper structural relationships among words, utilizing context pairs from both the entire sentence and its sub-units. For example, if a sentence is The quick brown fox jumps over a lazy dog, using then a context window size of 2, the context pairs for the target word fox would be as The, quick, brown, jumps, over.

Table I observes the features collections, in which the well-known Word2Vec model constructs features in immediate contexts based on window size. While RCE uses contexts and assigns according to the levels in which the center word is targeted, additional features are enhanced here by adding where target words lie either in the sentence or its left or right sub-unit levels. Target center words "fox", "dog", "The", and "over" have no add-on contexts (global information) due to these words are situated at terminals of the center-word's hierarchy, while words "jumps", "brown", "lazy", "quick", and "a" are enhancing Word2Vec contexts (local information) by add-on contexts.

E. Higher-Level Vectors Construction Based on RCE

In NLP, different sentence units (words/phrases/clauses) are considered for performing similarity evaluation where higher levels, such as phrases, clauses/sentences, can be constructed with the help of word vectors [1]. In Word2Vec, the sentence vector is typically calculated as the average of the word vectors in the sentence. For a sentence S consisting of n words, let the embedding of the i^{th} word be denoted as x_i. The sentence vector V_S or a sentence S that is obtained by averaging the word embeddings, is given by Eq. (7):

$$V_S = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{7}$$

RCE enhances the sentence vector by considering the hierarchical structure and word dependencies within the sentence. RCE works by recursively applying center embedding to progressively combine sub-sentences and word embeddings, resulting in a structured sentence representation as Eq. (8):

$$x_S^k = \frac{1}{m}\sum_{i=1}^{m} x_{S(i,j)}^{k-1} + x_S^1 \tag{8}$$

where x_S^1 represents the mean of word embeddings at the first level (sentence level), and recursively dividing the sentence and computing embeddings until the maximum level k s reached, creating a final sentence vector x_S^k that captures both word and higher-level semantics. Based on this formulation in terms of Eq. (1 – 8), we proposed an extension of MLCE (multi-level center embedding) through RCE that utilizes Word2Vec-based context pairs along with hierarchical contexts and enhanced baseline performance of Word2Vec for sentence similarity tasks. The following section will demonstrate this idea over benchmark datasets that conclude RCE effectiveness and generalizability.

IV. Result and Discussion

In this section, we have performed experiments to achieve objectives regarding RCE performance under sentence similarity evaluation and explore the generalizability of how RCE performed different sentence complexity.

[1]A primitive recursive function is a function defined using the base functions (such as the zero function, successor function, and projection function) and the operations of composition and primitive recursion [15].

TABLE I. COMPARATIVELY FEATURE-CONSTRUCTIONS OF BOTH WORD2VEC AND RCE FOR EXAMPLE SENTENCE

w_t	Word2Vec	RCE		
	Context-pairs	**Levels, wt: Center**	**Context-pairs**	**Add-on Contexts**
The	(quick, brown)	L0, jumps	(brown, fox, over, a)	Entire sentence
quick	(the, brown, fox)	L1, brown	(The, quick, fox, jumps)	Sub-unit: left
brown	(The, quick, fox, jumps)	L1, lazy	(over, a, dog)	Sub-unit: right
fox	(quick, brown, jumps, over)	L2, quick	(the, brown, fox)	Sub-unit: left
jumps	(brown, fox, over, a)	L2, fox	(quick, brown, jumps, over)	Sub-unit: []
over	(fox, jumps, a, lazy)	L2, a	(jumps, over, lazy, dog)	Sub-unit: left
a	(jumps, over, lazy, dog)	L2, dog	(a, lazy)	Sub-unit: []
lazy	(over, a, dog)	L3, The	(quick, brown)	Sub-unit: []
dog	(a, lazy)	L3, over	(fox, jumps, a, lazy)	Sub-unit: []

A. Experimental Setups

In order to evaluate the performance of the proposed Recursive Center Embedding (RCE) approach, a thorough experimental setup is designed using a variety of datasets, including both benchmark and author-constructed datasets as shown in Table II. This section outlines the datasets used, the Word2Vec training configuration, and the parameters optimized to improve the effectiveness of RCE for capturing sentence similarities through context-aware representations.

TABLE II. SUMMARY DETAILS OF SIMILARITY ASSESSMENT BENCHMARK DATASETS

Similarity Assessment Datasets		
***SICK* [12]**	***STS* [13]**	***Complex Dataset* [14]**
~10,000 pairs	~8,500 pairs	50 pairs
(1 to 5)	(0 to 5)	Final Similarity up to 5
Marelli et al., 2014	Cer et al., 2017	Chandrasekaran, D. and Mago, V., 2021
A benchmark dataset consisting of sentence pairs annotated with relatedness scores.	A widely used benchmark dataset containing pairs of sentences from diverse domains, each annotated with a score indicating the degree of semantic similarity.	A dataset containing pairs of complex sentences designed to challenge models in assessing sentence similarity for intricate structures.

1) Word2Vec training configuration: For the development of sentence vectors under the RCE framework, the Word2Vec model is trained with an optimized set of parameters to ensure the model captures meaningful context and word relationships effectively. The parameters for the Word2Vec training are as follows in Table III.

TABLE III. WORD2VEC TRAINING CONFIGURATION FOR RCE AT EPOCHS IN THE RANGE OF [50: 500]

Parameter	**Values**	**Description**
Vector Size	[100:300]	The dimensionality of the word vectors to be generated.
Window	[5:25]	The maximum distance between the current and predicted word within a sentence.
Model Type	Skip-gram	The architecture used for training the Word2Vec model, which predicts context words given a target word.
Training Data	flattens-RCE Vocab	According to the provided hierarchy of center-words, the dataset constructs the vocabulary and then flattens it in structure.
Lambda	0.8	A hyper-parameter to control the influence of the regularization term in the objective function.

The combination of these optimized parameters ensures that the trained Word2Vec model can develop RCE supported features through flat vocabs[2] meaningful word vectors that support the RCE's objective of accurately capturing hierarchical and structural sentence relationships.

2) Objective function of RCE: The Recursive Center Embedding (RCE) framework operates by recursively constructing center embeddings from sentences to capture both local word-level dependencies and global sentence structure. The objective function, as Eq. (6) discussed in the methodology section for RCE, is designed to balance the contribution of the recursive context embedding and word-level features. The parameter λ is set to 0.8, indicating that more weight is given to the recursive context structure compared to the individual word embeddings. The recursive process ensures that each word in the sentence is treated in the context of its neighboring words, creating context-pairs. These pairs are used to compute sentence-level vectors that reflect both semantic meaning and structural composition.

B. Significance of Normalize Representation of Sentence Vectors

Normalization of sentence vectors plays a crucial role in ensuring that vector magnitudes are comparable and meaningful for similarity tasks [10]. In Word2Vec, sentence vectors are typically computed as the mean of the word vectors within a sentence. On the other hand, Recursive Center Embedding (RCE) recursively computes a hierarchical embedding that captures the syntactic relationships between words. Normalization brings all vectors to a uniform scale, making it easier to compare vectors from different sentences. It is particularly important for RCE, where the recursive structure introduces multiple levels of abstraction. In mean-based Word2Vec, normalization primarily ensures length-invariance and improves the quality of similarity comparisons. We here performed Euclidean norms for both approaches as Eq. (9).

$$\left.\begin{aligned} \widehat{x_s} &= \frac{x_s}{||x_s||} \\ \widehat{x_{s,k}} &= \frac{x_{s,k}}{||x_{s,k}||} \end{aligned}\right\} \tag{9}$$

where $x_{s,k}$ is recursively obtained by averaging the word embeddings of sub-sentences and combining them as Eq. (9) for the RCE approach while the mean-based sentence vector

[2]Vocab is developed through unique sentences from available benchmark datasets.

x_s utilized for the Word2Vec. Obtained results of norm vectors are shown in both scenarios as in Table IV and Fig. 2, these normalization statistics across the Complex Sentence, SICK, and STS-B datasets suggest important insights about the performance of Recursive Center Embedding (RCE) compared to Word2Vec-based mean sentence vectors.

TABLE IV. NORM-VECTORS STATISTICS OBTAINED DURING EXPERIMENT WHICH IS SHOWN IN FIG. 2A, 2B, 2C ALONG WITH REMARKS ON WHICH DATASETS, SCHEME IS BETTER

Dataset	Word2Vec-Norms	RCE-Norms	Description
Complex-Sentence	Mean:1.1379 Std:0.1582	Mean:11.9599 Std:3.6996	RCE reflects greater captures sentence depth better. Mean-based vectors are uniform due to show a small degree of variation as indicated by standard deviations.
SICK	Mean:1.2048 Std:0.1412	Mean:9.7899 Std:2.6211	
STS-B	Mean:1.3550 Std:0.2977	Mean:9.7747 Std:3.7465	

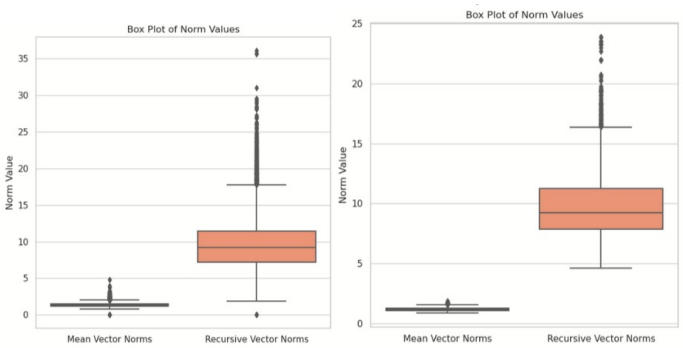
Complex Sentence Dataset is the most suitable choice for showcasing the effectiveness of Recursive Center Embedding (RCE). The significantly higher RCE norms (mean of 11.9599 vs. 1.1379 for Word2Vec) show that RCE captures the intricate structure of complex sentences much better than Word2Vec, while others demonstrate the advantages of RCE over Word2Vec, but the effect is less pronounced.

C. Assessment of Sentence Similarity Task

In normalizing sentence vectors, obtaining mean and deviation results shows a considerable significant variation. These results conclude that RCE can be a good alternative of traditional Word2Vec/BERT-base models if a primary task surrounds the structural and semantics evaluation of complex sentences. With this observation, we evaluated the similarity task in sentence pairs which are in different sizes as shown in Table II. Under the established configurations mentioned in Table III for constructing vectors, which are obtained after applying optimization for getting effective correlations through a grid search approach with various combinations of these hyper-parameters as V: vector_sizes = [100, 200, 300], W: window_size = [5, 10, 15, 20] E: epochs_list = [50, 100, 200, 500, 700].

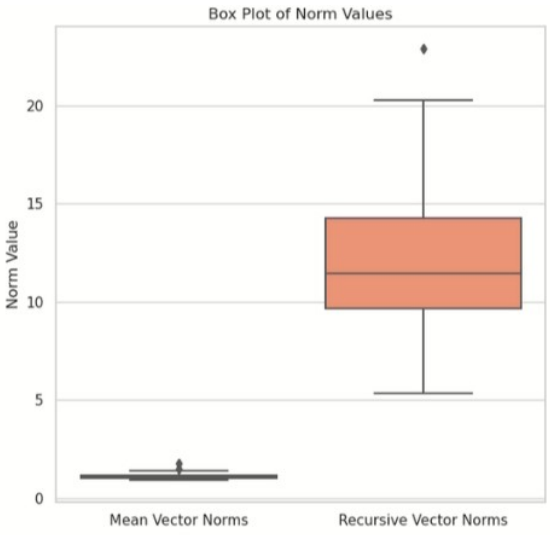
Correlation Results mentioned in Table V are reflected here that the chosen optimized parameters produce vectors that restrict them to over-fitting situations [6, 9]. Based on the observations, we can conclude that the traditional Word2Vec approach has limitations in exploring similarity tasks when sentences have structural complexities.

Another observation is found during assessment in the perspective of obtaining less over-fitting criteria and captured semantics within minimum context window size. Recently claimed complex dataset (having 50 sentence pairs) wins with a window size of 8; under this parameter, RCE effectively achieved its validation criteria. Meanwhile, Word2Vec only validated its over-fitting in terms of Spearman correlation (while Pearson metric still suffer from over-fitting) due to a lower validation score achieved in comparison to the training score. This analysis clearly illustrates the value of utilizing RCE-based hierarchical vectors for effectively determining sentence similarity tasks.

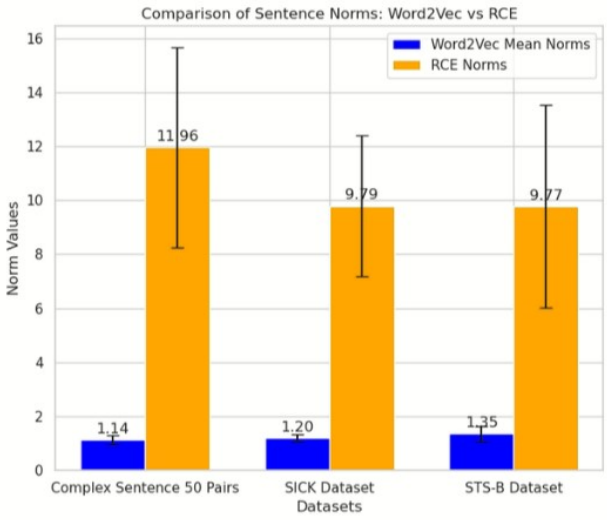


(a) Norms analysis of STS-b

(b) Norms analysis of SICK



(c) Norms analysis of complex dataset



(d) Combined norm-vectors analysis in all datasets

Fig. 2. Results for analyzing which schemes (Word2Vec and REC) effectively capture in deals with hierarchies through comparison of developed Norm-Vectors of sentences in: a) SICK dataset, b) STS-B dataset, c) Complex Sentence dataset, d) Combined comparative results among all datasets.

TABLE V. PEARSON (P_t,P_v) AND SPEARMAN(S_t,S_v) CORRELATION RESULTS OVER TRAINING AND VALIDATION SETS

Datasets	Training Parameters	**Word2Vec**				**RCE**			
		P_t	S_t	P_v	S_v	P_t	S_t	P_v	S_v
SICK	W=20, V=100, E=250	0.58	0.49	0.60	0.54	0.699	0.590	0.748	0.766
STS	W=20, V=100, E=500	0.21	0.23	0.25	0.26	0.518	0.516	0.527	0.538
Complex dataset	W=8, V=100, E=80	0.39	0.42	0.10	0.45	0.506	0.515	0.572	0.587

For comparing the proposed idea of RCE, we have chosen here Google pretrained Word2Vec model and contextual BERT-base pre-trained model on [14], applying mean of word vectors to get sentence information over benchmark datasets. As demonstrated by Table VI's results, the proposed RCE concept performed similarity to the pre-trained word vector model, with a minor improvement in SICK as well in BERT-base variant (particularly related similarities regarding Spearman correlations). Due to having less structural complexity in sentences, RCE poorly performed in the STS-B dataset.

D. Discussion

The Recursive Center Embedding (RCE) approach demonstrated significant advancements in sentence similarity tasks by integrating hierarchical context through recursive decomposition. Unlike traditional Word2Vec methods, which treat sentences as flat bags of words, RCE's recursive mechanism accounted for structural dependencies and nested relationships.

The comparative analysis revealed that RCE outperformed Word2Vec in datasets with intricate sentence structures, such as the Complex Sentence dataset, as evidenced by the significantly higher norm values and correlation scores. Specifically, RCE's ability to capture hierarchical context allowed for a nuanced understanding of sentence meaning, particularly in sentences containing nested clauses and intricate dependencies. For example, RCE achieved a higher Pearson correlation (0.567) on the Complex Sentence dataset, indicating its capability to generalize structural awareness. However, RCE exhibited limited performance on simpler datasets like STS-B, where the added hierarchical complexity was not essential. This suggests that RCE's recursive embedding mechanism may not be the most efficient approach for tasks dominated by surface-level word relationships.

An intriguing observation emerges regarding the proposed concept of RCE, which centers on structured abstraction and demonstrates efficacy over a complex dataset (comprising 50 pairs of sentences, which the authors [14] assert contains heavily complex sentences than other benchmark datasets such as SICK and STS-B), yielding substantial enhancements relative to conventional Word2Vec and the contextual pre-trained BERT-base model.

While RCE significantly improved sentence similarity tasks involving complex structures, its performance on STS-B dataset indicates that further optimization is needed. One promising direction is the incorporation of an adaptive recursion mechanism, which dynamically adjusts the depth of recursion based on sentence complexity. This dataset is still more difficult to analyze for sentence similarity than RCE, which is considered to be less effective, although earlier pre-trained models have done relatively better.

TABLE VI. COMPARISON OF RCE PERFORMANCE TO EXISTING PRE-TRAINED CONTEXTUAL MODELS [WORD2VEC, BERT-BASE VARIANTS (PARTIALLY SUPERVISED)] AT VALIDATED TRAINING PARAMETERS

Datasets	Pre-trained Result(*)		BERT Result(*)	RCE (Lambda = 0.8)	
	P	S	S	P	S
SICK	0.726	0.621	0.728	0.728	**0.736**
STS	0.626	0.587	0.769	0.515	0.521
Complex dataset	0.483	0.490	0.477	**0.567**	**0.551**

* Reported Google News pre-trained results are declared in the work as [14].

V. CONCLUSION AND FUTURE DIRECTION

The Recursive Center Embedding (RCE) approach offers a novel and effective method for sentence representation, particularly for complex-sentence structures. It introduces a recursive mechanism that captures sentence hierarchies in terms of their center-words, embedding them along with Word2Vec according to leveled-contexts gathered through syntactic structure, enabling the model to better handle intricate sentence patterns that are often difficult for traditional word embedding models like Word2Vec, BERT-base variant to capture. In this work, we explored its effectiveness for constructing sentence representations, comparing it with traditional Word2Vec-based mean sentence vectors across three datasets: Complex dataset (Sentence 50 Pairs), SICK, and STS-B.

The results demonstrated the strength of RCE, particularly in handling complex sentence structures, as shown in the significant difference in norm values between the two approaches. RCE significantly outperformed with a much higher norm value, demonstrating RCE's ability to capture intricate sentence structures, especially when dealing with complex sentences containing nested clauses or dependencies. Moreover, in terms of correlation results, highlighting its effectiveness for complex sentence evaluations.

A. Limitations

As observed in the STS-B dataset, RCE under-performed compared to pre-trained Word2Vec when dealing with simple sentence structures. The added complexity of recursively embedding center words may not be necessary for simpler, more straightforward sentence pairs, which rely more on surface-level word similarity. Sentences of these two datasets (specially in STS-B) have only up to limited levels of hierarchy. In all, This suggests that RCE's advantage lies in its ability to handle intricate syntactic structures, limiting its generalization to all types of sentence tasks.

B. Future Directions

Idea of adaptive depth mechanism can improve proposed RCE. Instead of applying the recursive center embedding approach uniformly to all sentences, the depth of recursion could be determined based on the sentence's syntactic complexity. Example: A highly complex sentence like "The student who was reading the book that was recommended by the professor passed the exam," could involve multiple levels of recursion, as it contains nested clauses. On the other hand, a sentence like "The student passed the exam," would only require minimal recursion. The recursive depth could be controlled based on the number of subordinate or relative clauses.

Possible solution for adaptive depth mechanism as to use a sentence parsing technique to identify the number of clauses, and accordingly determine the recursion depth. For sentences with higher clause density, increase the depth of RCE. Conversely, for simpler sentences, limit the recursion to one or two levels. This dynamic depth control can improve efficiency, particularly for large datasets.

REFERENCES

[1] S. Dubey, & N. Kohli, *A Multilevel center embedding approach for sentence similarity having complex structures.* in world conference on communication & computing (WCONF) (pp. 1-8). IEEE, (2023, July).

[2] N. Chomsky, *Logical structure in language*. journal of the American Society for information science, 8(4), 284, (1957).

[3] J.D Thomas, *Center-embedding and self-embedding in human language processing* (Doctoral dissertation, Massachusetts Institute of Technology), (1995).

[4] K.H. Cheon, Y. Kim, H.D. Yoon, K.C.Nam, S.Y. Lee, & H.A. Jeon, *Syntactic comprehension of relative clauses and center embedding using pseudowords.* Brain sciences, 10(4), 202, (2020).

[5] T. Mikolov, I. Sutskever, K. Chen, K., G. S. Corrado, & J. Dean, *Distributed representations of words and phrases and their compositionality.* Advances in neural information processing systems, 26, (2013).

[6] J. Pennington, R. Socher, & C.D. Manning, *Glove: Global vectors for word representation.* In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP) (pp. 1532-1543), (2014).

[7] H. Yanaka, K. Mineshima, D. Bekki, & K. Inui, *Do neural models learn systematicity of monotonicity inference in natural language?*. In Proceedings of the 58th annual meeting of the association for computational linguistics (pp. 6105-6117), (2020).

[8] T. Ranasinghe, C. Orasan, & R. Mitkov, R. *Semantic textual similarity with siamese neural networks.* In proceedings of the international conference on recent advances in natural language processing (RANLP 2019) (pp. 1004-1011), (2019).

[9] S. Dubey, & N. Kohli, *Clustering for clarity: improving word sense disambiguation through multilevel analysis.* Computer Science, 25(2), 1-24, (2024).

[10] W. Ford, *Numerical linear algebra with applications: Using MATLAB.* Academic press (2014).

[11] D. Yogatama,& N. Smith, *Making the most of bag of words: sentence regularization with alternating direction method of multipliers.* In international conference on machine learning (pp. 656-664). PMLR, (2014).

[12] M. Marelli, L. Bentivogli, M. Baroni, R. Bernardi, S. Menini, & R. Zamparelli, *Semeval-2014 task 1: evaluation of compositional distributional semantic models on full sentences through semantic relatedness and textual entailment.* In proceedings of the 8th international workshop on semantic evaluation (SemEval 2014) (pp. 1-8) (2014)

[13] D. Cer, M. Diab, E. Agirre, I. Lopez-Gazpio, & L. Specia, *Semeval-2017 task 1: semantic textual similarity-multilingual and cross-lingual focused evaluation.* arXiv preprint arXiv:1708.00055 (2017).

[14] D. Chandrasekaran, & V. Mago, *Comparative analysis of word embeddings in assessing semantic similarity of complex sentences*. IEEE Access, 9, 166395-166408. (2021).

[15] I. D. Zaslavsky, *On some generalizations of the primitive recursive arithmetic*. theoretical computer science, 322(1), 221-230, (2004).

[16] S. Dubey, & N. Kohli, *Enhancing symbolic manipulation through pairing primitive recursive string functions and interplay with generalized pairing PRSF*. Mathematical Problems of Computer Science, 60, 27-34, (2023).

[17] M. H. Khachatryan, *On generalized primitive recursive string functions.* Mathematical Problems of Computer Science, 43, 42-46, (2015).

Fault-Tolerant Control of Nonlinear Delayed Systems Using Lyapunov Approach: Application to a Hydraulic Process

Tayssir Abdelkrim[1], Adel Tellili[2], Nouceyba Abdelkrim[3]
National Engineering School of Gabes, Research Laboratory of Modeling,
Analysis and Control of Systems, (MACS), Gabes, Tunisia[1]
Institute of Technological Studies, Djerba, Tunisia[2]
Higher Institute of Industrial Systems, Gabes, Tunisia[3]

***Abstract*—Designing stabilizing controllers for delayed nonlinear systems with control constraints presents a significant challenge. This paper addresses this issue by proposing a fault-tolerant control approach for a specific class of delayed nonlinear systems with actuator faults based on Lyapunov redesign principle. Initially, an assumption is introduced to facilitate the control design for the nominal system. Then, a new control law is developed to resolve the difficulty caused by actuator failures. The proposed nonlinear controller demonstrates the ability to compensate for actuator faults. To validate its effectiveness, the method is applied to a hydraulic system.**

***Keywords*—Delayed nonlinear system; actuator faults; delayed hydraulic process; additive fault tolerant control; redesign Lyapunov approach**

I. Introduction

The primary objective of Fault Tolerant Control (FTC) is to guarantee performance and stability for systems, whether they are operating without faults or in a faulty state. Multiple approaches have been proposed to tackle this problem. FTC can be categorized into two main types: passive (P) and active (A) FTC [1]. In the domain of nonlinear control, extensive theoretical research and practical applications have been conducted for Passive Fault Tolerant Control (PFTC) and Active Fault Tolerant Control (AFTC) [2]. Active FTC is centered on ensuring stability and certain performance aspects for the post-fault model by dynamically adjusting the controller in response to the current fault [3], as detected, isolated, and estimated by the Fault Detection and Diagnosis (FDD) block [4]. A FTC employs real-time adjustment techniques for the regulators to uphold, at a minimum, the system's stability [5]. Another approach involves using a robust controller able of handling all anticipated faults, by eliminating the requirement for online control reconfiguration and an FDD block [6], PFTC methods have been introduced, predominantly grounded in robust theories. These include approaches such as linear-matrix-inequality-based methods [7], quantitative feedback theory [8], pole assignment [9], and nonlinear regulation theory [10, 11].

A. Literature Review

Many recent studies have focused on FTC applied to particularly complex nonlinear processes, such as boundary adaptive fault-tolerant control for a flexible Timoshenko [12] the proposed method ensures robust and adaptive control of a Timoshenko flexible arm, guaranteeing stability and precision despite actuator faults, hysteresis, and disturbances, while enhancing reliability under variable conditions. Another research addresses a 2 DOF helicopter system, this study proposes an adaptive control strategy for a 2-DOF helicopter, considering actuator faults and an unknown dead zone. A neural network and a quantizer are used to model the uncertainty and reduce system chattering [13].

Such presence of delays may influence the qualitative system properties and may affect the stability of the process control. When dealing with systems involving time delay, two approaches are employed to establish stability, aligning with the conventional Lyapunov stability theory. The first approach relies on Lyapunov-Krasovski functionals, while the second approach uses Lyapunov-Razumikhin functions [14]. Hence, research into control systems with delays in a nonlinear context holds great importance [15]. The literature offers a variety of methods for developing fault-tolerant controllers in the context of nonlinear systems. In their work [16], Liang and Xu introduced a variable structure stabilizing control law to handle actuator faults within a nonlinear system [17].

The concept of Control Lyapunov Functions (CLF) has played a pivotal role in the advancements of robust control for nonlinear systems [18, 19, 20]. A function that is positive definite and radially unbounded qualifies as a CLF if its time derivative become negative definite through appropriate control input selection, regardless of the value of state. Once a CLF is identified, various methods exist to derive control laws that stabilize the nonlinear system [21].

The application of CLF has been extended to systems with disturbances [20] and [22], to systems with delay and to stochastic systems [19].

Furthermore, the issue of faults, loss of effectiveness, and delay has been addressed using the Lyapunov tool, and has also been treated in the context of a stochastic system such as an adaptive fuzzy control strategy for stochastic nonlinear systems with faults and input saturation uses control filtering to reduce computational load. Fuzzy logic systems approximate the unknown nonlinearities and system variations caused by faults [23]. Another fault-tolerant fuzzy control strategy for stochastic nonlinear systems with quantized inputs is proposed. It uses a hysteretic quantizer to avoid chattering and fuzzy logic systems to estimate unmeasurable states and approximate

nonlinearities. The approach guarantees system stability and signal boundedness in the presence of actuator faults and quantization [24]. Furthermore, a fault-tolerant control strategy is proposed for nonlinear strict-feedback systems with actuator saturation, disturbances, and faults. Neural Networks (NNs) are used to approximate the unknown dynamics, and a backstepping technique is employed to design the controller. The NN weights are updated online using a gradient descent algorithm, thereby improving the approximation accuracy.

In this paper, the approach introduced in recent studies [8] and [25] is adopted, where actuator faults are represented as bounded additive periodic unknown signals added to the control signal. Additionally, a scenario is examined in which the efficiency of the actuator is compromised, represented by a multiplicative factor. This factor, when applied to the control signal, decreases its performance in accordance with the factor's value [6]. The proposed fault-tolerant control (FTC) is applied to water level control of a hydraulic system based on the Lyapunov redesign principle [26]. If a stabilizing closed-loop controller and its corresponding Lyapunov function exist for the nominal plant, an FTC is constructed based on these nominal controllers and the Lyapunov function. This FTC guarantees stability even in the presence of faults in the system.

B. Main Contribution

The main contribution of this work is the development of a control strategy for nonlinear systems that are subject to both actuator faults (including additive faults and loss of effectiveness) and time delays, addressing a complex and challenging control scenario without linearizing the nonlinear system. Unlike many traditional approaches that simplify the problem through linearization.

The proposed control scheme combines a nominal control component, which governs the system under normal operating conditions, with an additive corrective term specifically designed to compensate faults.

The subsequent sections of the paper are organized as follows: Section II will present the system characterization and problem formulation, Section III will outline the main results, Section IV will provide a real application to verify the efficiency of the proposed additive FTC. In final Section V, we give some conclusions about this work.

II. System Description and Problem Statement

Here, we consider nonlinear time delay system affine in the control of the form:

$$\dot{x}(t) = f(x_d(t)) + g(x_d(t))\, u(t,x) \tag{1}$$

Where $x \in R^n$,$x_d \in R^n$, $u \in R^m$,, represent respectively, the state of system, delayed state and the input vectors, the initial condition $x_d(0)(.) = \phi_d$ is given by the continuous function $\phi_d : [-d, 0] \rightarrow R^n$. Vector fields f and g are smooth functional mapping piecewise continuous function in t and locally Lipschitz in x and u [27]. The functions f and g are known precisely.

Hypothesis 1: We propose the existence of a nominal closed-loop control, represented as $u_{nom}(t,x)$, with the anticipation that it guarantees the overall stability of the closed-loop system.

$$\dot{x}(t) = f(x_d(t)) + g(x_d(t))\, u_{nom}(t,x) \tag{2}$$

The concept of Control Lyapunov Functions (CLF) has gained attention in the literature due to the availability of various CLF based control laws, which aim to stabilize the system and ensure a certain level of robustness for the closed-loop system. The function $V(x)$ is defined as a positive definite Lyapunov function such that for all $x \in B_x$ for the system [Eq. (1)].

The Lyapunov function will be derived along the trajectories of Eq. (2).

$$\dot{V} \leq -\Gamma_1(|x|) \tag{3}$$

with $\Gamma_1(s) > 0$ for $s > 0$

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x}\,\frac{\partial x}{\partial t} \leq -\Gamma_1(|x|) \tag{4}$$

Using (2), we get

$$\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x}\left[f(x_d(t)) + g(x_d(t))\, u_{nom}(t,x)\right] \leq -\Gamma_1(|x|) \tag{5}$$

Which shows that $\dot{V}$ is negative definite. Consequently, the origin of the full system [Eq. (2)] is asymptotically stable.

III. Proposed Fault Tolerant Control Design

In the upcoming sections, we will outline the main results.

If we assume that the system can be stabilized within the domain B_x and that the state x is accessible for feedback, our objective is to find a control scheme that achieves asymptotic stabilization of the point $x = 0$ of the closed loop nonlinear delayed system despite actuator fault occurrence.

The fault-tolerant control strategy denoted as control input u_{FTC} designed for the purpose of stabilizing the system in the presence of faults, is proposed as:

$$u_{FTC} = \alpha\ (u_{nom} + F(x,t) + u_{add}) \tag{6}$$

We take into account the reduction in actuator efficiency, represented by a multiplicative matrix α, where $\alpha \in R^{m\times m}$ is a diagonal continuous time variant matrix, with the diagonal elements $\alpha_{ii}(t)$, $i = 1, \ldots, m$ s.t $0 < \alpha_{ii} \leq 1$, $u = u_{nom}$ designates the nominal controller responsible for system stabilization in the absence of any actuator faults.

Using Eq. (6), the system [Eq. (1)] becomes:

$$\dot{x}(t) = f(x_d(t)) + g(x_d(t))\,\alpha\ (u_{nom} + F(x,t) + u_{add}) \tag{7}$$

Hypothesis 2:

$F(x,t)$ signifies an actuator fault that satisfies the condition $\|F(x,t)\| \leq L(x,t)$ and $L(x,t)$ is non-negative continuous function satisfying

$$L(x,t) = \frac{1-\alpha}{\alpha} u_{nom} \tag{8}$$

with $\alpha \in R$

Proposed FTC control: In this context, we will design an additive control u_{add} represents the additional component to compensate the actuator fault's impact as:

$$u_{add} = -\gamma(V) \frac{\partial V}{\partial x} g(x_d(t)) \tag{9}$$

For any $\gamma > 0$ there exists a smooth, positive dominating function γ such that all trajectories of the closed loop system Eq. (2) and (9) satisfy $\lim_{t\to\infty} |x(t)| < \gamma$

So, the following theorem is proposed to achieve control of the studied nonlinear delayed system Eq. (6) in the presence of malfunctioning actuators.

Theorem: The fault-tolerant control law Eq. (6) ensures asymptotic stability of the closed loop nonlinear delayed system described in Eq. (7) defined by Eq. (8) and (9) under the previous assumptions 1 and 2 even in cases of abnormal operation of the actuators.

Proof:

According to Eq. (6), it follows for the closed-loop faulty nonlinear delayed system Eq. (7), a Control Lyapunov Function (CLF) candidate will be designed such as $V(x_d)$ is a positive definite function as:

$$\dot{V}(t) = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} \frac{\partial x}{\partial t} \tag{10}$$

Using Eq. (7), we get

$$\dot{V}(t) = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} \left(f(x_d(t)) + g(x_d(t)) \alpha \ (u_{nom} + F(x,t) + u_{add}) \right) \tag{11}$$

The establishment of the derivative of $V(t)$ will occur in accordance with the trajectories defined by system Eq. (7), this leads to

$$\begin{aligned} \dot{V}(t) = &\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(x_d(t)) + \frac{\partial V}{\partial x} g(x_d(t)) \alpha \ u_{nom} \\ &+ \frac{\partial V}{\partial x} g(x_d(t)) \alpha \ F(x,t) + \frac{\partial V}{\partial x} g(x_d(t)) \alpha \ u_{add} \end{aligned} \tag{12}$$

where

$\beta = 1 - \alpha \geq 0$, Eq. (12) can be expressed as

$$\begin{aligned} \dot{V}(t) = &\frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(x_d) + \frac{\partial V}{\partial x} g(x_d(t)) u_{nom} \\ &- \beta \frac{\partial V}{\partial x} g(x_d) u_{nom} + \alpha \frac{\partial V}{\partial x} g(x_d) F(x,t) \\ &+ \alpha \frac{\partial V}{\partial x} g(x_d) u_{add} \end{aligned} \tag{13}$$

Using Eq. (8) we obtain:

$$\dot{V}(t) = \frac{\partial V}{\partial t} + \frac{\partial V}{\partial x} f(x_d) + \frac{\partial V}{\partial x} g(x_d(t)) u_{nom} + \alpha \frac{\partial V}{\partial x} g(x_d) u_{add} \tag{14}$$

By substituting Eq. (9) into Eq. (14), the desired stability can be achieved.

$$\dot{V}(t) \leq -\Gamma_1(|x|) - \alpha\gamma(V) \left(\frac{\partial V}{\partial x} g(x_d(t)) \right)^2 \tag{15}$$

Such as $\alpha > 0$ and γ is strictly increasing function.

For this purpose

$$\dot{V}(t) \leq -\Gamma_2(|x|) \tag{16}$$

where $\Gamma_2 > 0$

Hence the system Eq. (1) can be stabilized with the control law Eq. (9). Thus the derivative $\dot{V}$, is negative along the trajectory of the closed-loop system.

In view of the control law Eq. (9) and taking into account the assumptions about the fault, it is obvious that $\dot{V} \leq 0$.

So, it can be concluded that the origin $x = 0$ of the faulty overall system is a asymptotically stable equilibrium point for studied system Eq. (7) under the fault tolerant control law Eq. (9) regardless of the presence of delay.

IV. FTC Control of a Hydraulic System

A. Description of a Single-Tank Hydraulic System

Adjusting a hydraulic level in a tank is the main objective of this work by developing fault-tolerant control for this nonlinear time-delay system based on Lyapunov approach. The structure of the entire system is as shown in Fig. 1. The system is built about a water tank, a liquid level sensor, coil and pump. The tank is supplied by two water inputs: the first is located at the top while the second is at the bottom to compensate for any faults if it exists. Furthermore, the tank has two outputs, the first one for liquid discharge and the second one for leakage (disturbance), the coil affects a pure delay.

B. Modeling of a Single-Tank Hydraulic System

In Fig. 2, the tank is supplied by two water inputs: the first $Q_{I1}(t)$ is located at the top while the second $Q_{I2}(t)$ is at the bottom to compensate for any faults if it exists. Furthermore, the tank has two outputs, the first one for liquid discharge and the second one for leakage (disturbance), the coil affects a pure delay.

Fig. 1. Real hydraulic system at MACS laboratory.

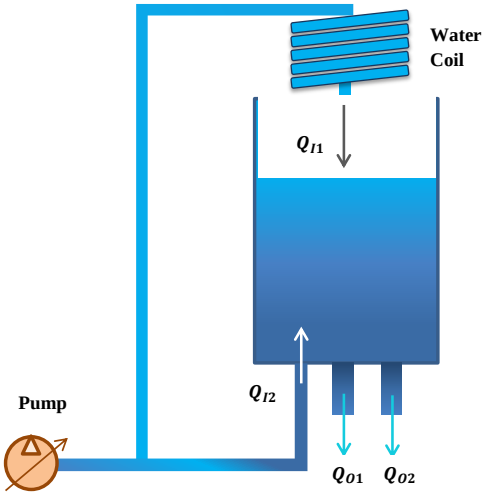


Fig. 2. Model of the hydraulic system at MACS laboratory.

Then, the tank system can be modeled by the following differential equation:

$$\frac{\partial V}{\partial t} = \frac{\partial \left(S.h\left(t\right)\right)}{\partial t} = Q_I\left(t\right) - Q_O\left(t\right) \tag{17}$$

$V(t)$:Water speed$[m/s]$
S : Discharge output section $[m^2]$
$h(t)$: Water level $[m]$
$Q_I(t)$: Input flow rate $[m^3/s]$
$Q_O(t)$: Output flow rate $[m^3/s]$

In our case, the tank section is constant, so we can write

$$\frac{\partial h}{\partial t} = \frac{Q_I\left(t\right) - Q_O\left(t\right)}{S} \tag{18}$$

We set $x(t) = h(t)$, we obtain

$$\dot{x}\left(t\right) = \frac{Q_I\left(t\right) - Q_O\left(t\right)}{S} \tag{19}$$

with

$$Q_I\left(t\right) = \ Q_i\left(t\right)\left(1 + A\left(x_d, t\right)\right) \tag{20}$$

As long as we have two inputs, $Q_{I1}(t) = Q_i$ is considered the main input and $Q_{I2}\left(t\right) = \ Q_i\left(t\right)A\left(x_d, t\right)$ is used to compensate faults

We obtain by identification

$$A\left(x_d, t\right) = 0.7857{x_d}^2\left(t\right) + -0.6614x_d\left(t\right) + 0.68 \tag{21}$$

To simplify, we take

$$g'\left(x_d\right) = 1 + A\left(x_d, t\right) \tag{22}$$

Therefore,

$$Q_I\left(t\right) = \ Q_i\left(t\right)g'\left(x_d\right) \tag{23}$$

The output flow rate is divided into two parts: the first part $Q_{O1}(t)$ concerns the main output flow rate, and the second $Q_{O2}(t)$ represents the leakage flow rate which considered here as a fault.

$$Q_O(t) = Q_{O1}\left(t\right) + \ Q_{O2}\left(t\right) \tag{24}$$

Knowing the relationship between flow rate, section and velocity, we can write

$$Q_O(t) = S_{O1}V_1(t) + S_{O2}V_2(t) \tag{25}$$

By replacing Eq. (23) and (25) in Eq. (19), we obtain

$$\dot{x}\left(t\right) = \frac{Q_i\left(t\right)g'\left(x_d, t\right) - \left(S_{O1}V_1(t) + S_{O2}V_2(t)\right)}{S} \tag{26}$$

After some algebraic manipulations, Eq. (26) can be expressed as

$$\dot{x}\left(t\right) = \frac{-S_{O1}V_1(t)}{S} + \frac{Q_i\left(t\right)g'\left(x_d, t\right) - S_{O2}V_2(t)}{S} \tag{27}$$

In the fault-free case i.e. $Q_{O2}(t) = 0 \rightarrow S_{O2}V_2(t) = 0$, our system can be represented as follows:

$$\dot{x}\left(t\right) = \frac{-S_{O1}V_1(t)}{S} + \frac{Q_i\left(t\right)g'\left(x_d, t\right)}{S} \tag{28}$$

According to Torricelli's theorem,

$$V_1(t) = \sqrt{2gx_d(t)} \tag{29}$$

with $g = 9.81m/s^2$

The Eq. (28) becomes:

$$\dot{x}(t) = \frac{-S_{O1}\sqrt{2gx_d(t)}}{S} + \frac{Q_i(t)\,g'(x_d,t)}{S} \tag{30}$$

By referring to Eq. (2), we can express Eq. (30) in this form:

$$\dot{x}(t) = f(x_d,t) + g(x_d,t)\,Q_i(t) \tag{31}$$

Along with

$$f(x_d,t) = \frac{-S_{O1}\sqrt{2gx_d(t)}}{S} \tag{32}$$

And

$$g(x_d,t) = \frac{g'(x_d,t)}{S} \tag{33}$$

In our case, the Q_i Input flow rate of the hydraulic system corresponds to the nominal control u_{nom}

$$Q_i(t) = u_{nom}(t) \tag{34}$$

C. Experimental Validation on a Hydraulic System

1) Stabilization of nonlinear delayed water tank in fault-free case: The description of the nominal closed-loop system is as follows:

$$f(x_d,t) + g(x_d,t)\,u_{nom}(t) = K(x_{ref}(t) - x(t)) \tag{35}$$

with $K > 0$

Let the Lyapunov function be

$$V(t) = \frac{1}{2}(x_{ref}(t) - x(t))^2. \tag{36}$$

However, to ensure the stability of the hydraulic system, it is necessary that $\dot{V}(t) < 0$. So, for this reason, we can choose

$$u_{nom}(t) = \frac{K(x_{ref}(t) - x(t)) - f(x_d,t)}{g(x_d,t)} \tag{37}$$

Consequently, the stability of the hydraulic system is achieved. By examining Eq. (7) and (30), the fault free system is obtained by $\alpha = 1$ and $F(t,x) = 0$,where F represents the fault corresponding to a flow leakage in our real system $Q_{O2}(t) = 0 \rightarrow S_{O2}V_2(t) = 0$. So, it's described as follows:

$$\dot{x}(t) = \frac{-S_{O1}\sqrt{2gx_d(t)}}{S} + \frac{Q_i(t)\,g'(x_d,t)}{S} \tag{38}$$

The desired level of the water $x_{ref} = 0.4$ m.

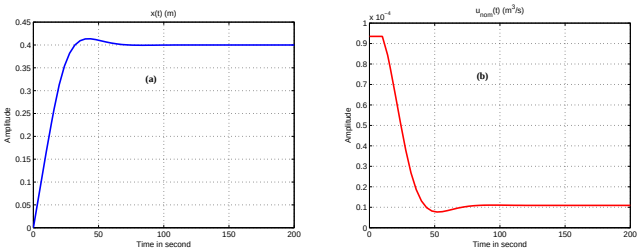


Fig. 3. State trajectory in fault-free case (a) and nominal control (b).

We obtain the next hydraulic system responses in Fig. 3.

Discussion: Initially, in Fig. 3(a) there is a rapid rise to approximately 0.4 m, followed by a stabilization near this desired value. This stabilization shows that the system is able to reach and maintain the desired level without excessive oscillation or instability despite the delay introduced by the tank coil.

In Fig. 3(b) the application of a nominal control Eq. (37) is essential to compensate for the delay and swiftly bring the system to the desired level.

2) Stabilization of delayed hydraulic system with actuator faults: In practice, hydraulic systems are susceptible to a various of faults, including loss of effectiveness and additives faults, both of which can significantly reduce their optimal performance.

So, designing the control law according to Eq. (6) and (9) is essential for stabilizing the nonlinear delayed hydraulic system Eq. (27) in the presence of faults. Based on the condition of Eq. (9), which states that $\gamma(V)$ must be a strictly increasing function, we can therefore choose:

$$\gamma(V) = |x_{ref}(t) - x(t)|^2 \tag{39}$$

Consider the nonlinear delayed system (27) with intermittent fault between $t = 80s$ and $t = 90s$ and 3 cases of loss of effectiveness fault $\alpha = 20\%$, $\alpha = 30\%$ and $\alpha = 40\%$.

$$\dot{x}(t) = \frac{-S_{O1}V_1(t)}{S} + \alpha\left(\frac{Q_i(t)\,g'(x_d,t) - S_{O2}V_2(t)}{S}\right) \tag{40}$$

We start by applying nominal control u_{nom}, and subsequently, we implement fault-tolerant control $u = u_{FTC}$ to compensate any potential faults.

Discussion: Fig. 4(a) shows that when nominal control is employed, the system state deviates from its reference trajectory in the presence of faults. The deviation is notably pronounced in the case of an additive and loss of effectiveness faults, suggesting that nominal control fails to effectively compensate for this specific type of fault.

In Fig. 4(b), it is observed that the control corresponding to this case shows a downward peak, reflecting the need for fault compensation.

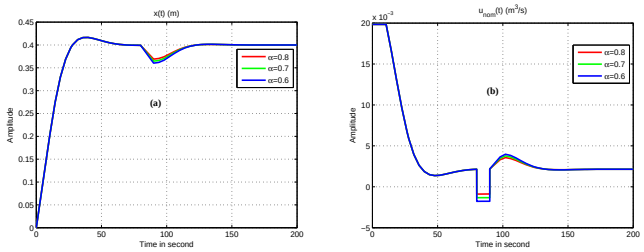


Fig. 4. State trajectory with an additive and loss of effectiveness faults (a) in case of nominal control (b).

To solve the above problem, the control law will be modified to include a new term that represents the component capable of eliminating the impact of the fault.

Hence, we suggest a fault-tolerant control strategy, denoted by u_{FTC} and expressed as:

$$u_{FTC} = \alpha \ (u_{nom} + F(x,t) + u_{add}) \tag{41}$$

With

$$u_{add} = -|x_{ref} - x|^2 (x_{ref} - x)\, g(x_d, t) \tag{42}$$

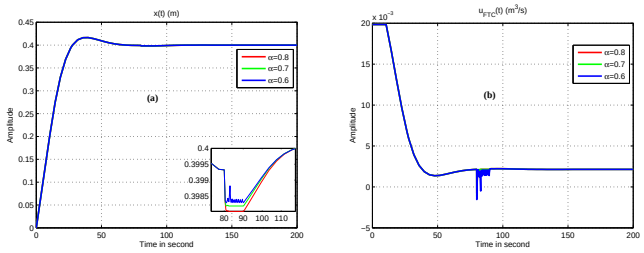


Fig. 5. State trajectory with an additive and loss of effectiveness faults (a) in case of fault tolerant control (b).

Discussion: Fig. 5(a) illustrates that the fault-tolerant control strategy successfully achieves stability and maintains a well-behaved system, even in the presence of faults.

The fault-tolerant control strategy, which accounts for both factors and a delay, compensates the negative impact on the system's behavior and maintains stability. In Fig. 5(b), a shorter-duration peak is observed, highlighting the effect of the newly adopted FTC control, which successfully compensated the system fault.

V. Conclusion

In conclusion, this paper aimed to stabilize a nonlinear delayed system affected by actuator faults, focusing on both additive faults and loss of effectiveness. To achieve this objective, we developed a fault-tolerant control strategy based on the Lyapunov redesign approach. Our research was grounded in practical experimentation, conducted on a real hydraulic system within our laboratory : Modeling, Analysis and Control of Systems (MACS).

The results demonstrated that applying nominal control alone to the faulty system leads to performance degradation. However, by integrating an additive control term, the proposed approach successfully compensates for the faults, ensuring system stability and fault tolerance. This study provides valuable insights and contributes to the advancement of fault-tolerant control strategies for nonlinear delayed systems with practical validation.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Cheridi, Djamel Eddine. Commande tolérante aux défauts robuste pour les systèmes dynamiques non linéaires retardés. Diss. Université Frères Mentouri-Constantine(Algérie), 1, (2020).

[2] Cai, Miao, Xiao He, and Donghua Zhou. "Fault-tolerant tracking control for nonlinear observer-extended high-order fully-actuated systems." Journal of the Franklin Institute 360.1 , 136-153(2023).

[3] Zhang, Youmin, and Jin Jiang. "Bibliographical review on reconfigurable fault-tolerant control systems." Annual reviews in control 32.2 , 229-252(2008).

[4] Kharrat, Dhouha. Commande tolérante aux défauts des systèmes non linéaires: application à la dynamique du véhicule. Diss. Université de Picardie Jules Verne; Université de Sfax (Tunisie), (2019).

[5] Benjemaa, R., A. Elhsoumi, and S. Bel Hadj Ali Naoui. "Stability and Active Fault-Tolerant Control design for a class of neutral time-delay systems." 2019 International Conference on Signal, Control and Communication (SCC). IEEE, (2019).

[6] Benosman, Mouhacine, and K-Y. Lum. "Passive actuators' fault-tolerant control for affine nonlinear systems." IEEE Transactions on Control Systems Technology 18.1 ,152-163 (2009).

[7] Liao, Fang, Jian Liang Wang, and Guang-Hong Yang. "Reliable robust flight tracking control: an LMI approach." IEEE transactions on control systems technology 10.1 , 76-89(2002).

[8] Wu, Shu-Fan, Michael J. Grimble, and Wei Wei. "QFT based robust/fault tolerant flight control design for a remote pilotless vehicle." Proceedings of the 1999 IEEE International Conference on Control Applications (Cat. No. 99CH36328). Vol. 1. IEEE, (1999).

[9] Niemann, Henrik, and Jakob Stoustrup. "Passive fault tolerant control of a double inverted pendulum—a case study." Control engineering practice 13.8 , 1047-1059 (2005).

[10] Bajpai, Gaurav, Bor-Chin Chang, and Harry G. Kwatny. "Design of fault-tolerant systems for actuator failures in nonlinear systems." Proceedings of the 2002 American Control Conference (IEEE Cat. No. CH37301). Vol. 5. IEEE, (2002).

[11] Bonivento, Claudio, et al. "Implicit fault-tolerant control: application to induction motors." Automatica 40.3 , 355-371(2004).

[12] Zhao, Zhijia, et al. "Boundary adaptive fault-tolerant control for a flexible Timoshenko arm with backlash-like hysteresis." Automatica 130 (2021): 109690.

[13] Zhao, Zhijia, et al. "Adaptive quantized fault-tolerant control of a 2-DOF helicopter system with actuator fault and unknown dead zone." Automatica 148 (2023): 110792.

[14] Hua, Changchun, et al. "Type-B Nussbaum function-based fault-tolerant control for a class of strict-feedback nonlinear systems." Journal of the Franklin Institute 360.4 ,2421-2435(2023).

[15] Sweetha, S., et al. "Non-fragile fault-tolerant control design for fractional-order nonlinear systems with distributed delays and fractional parametric uncertainties." IEEE Access 10 , 19997-20007(2022).

[16] Liang, Y-W., and S-D. Xu. "Reliable control of nonlinear systems via variable structure scheme." IEEE Transactions on Automatic Control 51.10 , 1721-1726(2006).

[17] Tellili, Adel, Aymen Elghoul, and Mohamed Naceur Abdelkrim. "Additive fault tolerant control of nonlinear singularly perturbed systems against actuator fault." Journal of Electrical Engineering 68.1 , 68-73(2017).

[18] Battilotti, Stefano. "Robust stabilization of nonlinear systems with pointwise norm-bounded uncertainties: A control Lyapunov function approach." IEEE transactions on automatic control 44.1 , 3-17(1999).

[19] Jankovic, Mrdjan, Rodolphe Sepulchre, and Petar V. Kokotovic. "CLF based designs with robustness to dynamic input uncertainties." Systems and control letters 37.1 , 45-54(1999).

[20] Mazenc, Frédéric, Silviu-Iulian Niculescu, and Mounir Bekaik. "Stabilization of nonlinear systems with delay in the input through backstepping." 2011 50th IEEE Conference on Decision and Control and European Control Conference. IEEE, (2011).

[21] Sontag, Eduardo D. "A universal construction of Artstein's theorem on nonlinear stabilization." Systems and control letters 13.2 , 117-123(1989).

[22] Teel, Andrew R. "Connections between Razumikhin-type theorems and the ISS nonlinear small gain theorem." IEEE Transactions on Automatic Control 43.7 , 960-964(1998).

[23] Qiu, Jianbin, Min Ma, and Tong Wang. "Event-triggered adaptive fuzzy fault-tolerant control for stochastic nonlinear systems via command filtering." IEEE Transactions on Systems, Man, and Cybernetics: Systems 52.2 (2020): 1145-1155.

[24] Yu, Xinghu, et al. "Barrier Lyapunov function-based adaptive fault-tolerant control for a class of strict-feedback stochastic nonlinear systems." IEEE Transactions on Cybernetics 51.2 (2019): 938-946.

[25] Zhao, Qing, and Jin Jiang. "Reliable state feedback control system design against actuator failures." Automatica 34.10, 1267-1272 (1998).

[26] Elghoul, Aymen, Adel Tellili, and Mohamed Naceur Abdelkrim. "Reconfigurable control of flexible joint robot with actuator fault and uncertainty." Journal of Electrical Engineering 70.2 , 130-137(2019).

[27] Yamashita, Yuh, Kiminori Sakano, and Koichi Kobayashi. "Asymptotic stabilization with group wise sparse input based on control Lyapunov function approach." International journal of robust and nonlinear control 33.1 , 35-48(2023).

Advanced Deep Learning Approaches for Fault Detection and Diagnosis in Inverter-Driven PMSM Systems

Abdelkabir BACHA[1], Ramzi El IDRISSI[2], Fatima LMAI[3],
Hicham EL HASSANI[4], Khalid Janati Idrissi[5], Jamal BENHRA[6]
OSIL Team-LARILE Laboratory-School of Electrical and Mechanical Engineering,
Hassan II University of Casablanca, Casablanca, Morocco[1,4,6]
LPMAT-Physics Department-Faculté des Sciences AIN CHOCK, Hassan II University of Casablanca, Casablanca, Morocco[2,3]
Mechanical Engineering Team-Laboratory: Mechanical Engineering-Faculty of Sciences and Technology of Fez,
Sidi Mohamed Ben Abdellah University of Fez, Fez, Morocco[5]

***Abstract*—This paper presents a comprehensive approach to fault detection and diagnosis (FDD) in inverter-driven Permanent Magnet Synchronous Motor (PMSM) systems through the innovative integration of transformer-based architectures with physics-informed neural networks (PINNs). The methodology addresses critical challenges in power electronics reliability by incorporating domain-specific physical constraints into the learning process, enabling both high accuracy and physically consistent predictions. The proposed system combines advanced sensor fusion techniques with real-time monitoring capabilities, processing multiple input streams including phase currents, temperatures, and voltage measurements. The architecture's dual-objective optimization approach balances traditional classification metrics with physics-based constraints, ensuring predictions align with fundamental electromagnetic and thermal principles. Experimental validation using a comprehensive dataset of 10,892 samples across nine distinct fault scenarios demonstrates the system's exceptional performance, achieving 98.57% classification accuracy while maintaining physical consistency scores above 0.98. The model exhibits robust performance across varying operational conditions, including speed variations (97.45-98.57% accuracy range) and load fluctuations (97.91-98.12% accuracy range). Notable achievements include perfect detection rates for certain critical faults, such as high-side short circuits and thermal anomalies, with area under ROC curve (AUC) scores of 1.0. This research establishes new benchmarks in condition monitoring and fault diagnosis for power electronic systems, offering practical implications for predictive maintenance and system reliability enhancement.**

Keywords*—Fault detection and diagnosis; PMSM; deep learning; transformers; physics-informed neural networks; power electronics

I. Introduction

The increasing deployment of Permanent Magnet Synchronous Motors (PMSMs) in critical applications, from electric vehicles to industrial automation, has created an urgent need for reliable and interpretable fault detection systems [1], [2]. While traditional machine learning approaches have demonstrated potential in this domain [3], significant challenges persist that limit their practical effectiveness. A primary concern is the lack of physical consistency in predictions, which can lead to unrealistic fault diagnoses that compromise system reliability. Deep learning models, despite their computational power, often suffer from limited interpretability, which significantly reduces trust in their decision-making capabilities in critical situations. Furthermore, these approaches frequently demonstrate insufficient robustness to variations in operating conditions and noise, making them less reliable in real-world industrial environments.

A fundamental limitation of current approaches lies in their inability to effectively leverage domain knowledge within the learning process. This deficiency, combined with high false alarm rates in traditional data-driven approaches, creates significant barriers to practical implementation. These limitations become particularly problematic in safety-critical applications where incorrect fault diagnoses could lead to catastrophic failures, resulting in substantial economic losses or safety risks.

Previous attempts to address these challenges have primarily focused on either pure data-driven approaches or model-based methods, failing to effectively combine the advantages of both paradigms. Model-based approaches, while theoretically sound, often struggle with complex, nonlinear fault dynamics that characterize real-world PMSM operations. Conversely, pure data-driven methods, though capable of handling complex patterns, may violate fundamental physical constraints that govern motor behavior. Traditional hybrid approaches have attempted to bridge this gap but lack a systematic framework for integrating domain knowledge with learning algorithms.

This research addresses these fundamental challenges through the development of a novel physics-informed deep learning architecture. The proposed approach systematically integrates domain knowledge with advanced neural network capabilities, creating a robust framework that maintains physical consistency while leveraging the pattern recognition capabilities of deep learning. This integration represents a significant step forward in creating reliable, interpretable, and physically consistent fault detection systems for PMSM applications.

A. Background and Motivation

Recent advancements in deep learning have revolutionized fault detection in power electronic systems [4], [5]. However, conventional neural networks treat the problem as a pure data-fitting exercise, potentially leading to physically inconsistent

predictions [6]. The incorporation of physics-based constraints through PINNs offers a promising solution to this limitation.

B. Research Objectives

The primary objectives of this study are:

- To develop and implement a PINN architecture specifically designed for PMSM fault detection
- To compare the performance of PINNs with traditional machine learning classifiers
- To analyze the impact of physics-based constraints on model robustness and generalization

The complete dataset has been published in the Zenodo repository [7] under DOI: 10.5281/zenodo.13974503. Additionally, comprehensive documentation and source code are available as open-source materials on the authors' GitHub profile [8]. The repository includes detailed methodology for thermistor calibration.

II. Literature Review

A. Traditional FDD Methods

Early approaches to fault detection in PMSM systems primarily relied on model-based methods and signal processing techniques. Model-based approaches typically utilize mathematical models of the PMSM system to generate residuals between the predicted and measured signals [6], [9]. Signal processing methods, such as Fast Fourier Transform (FFT) and wavelet analysis, have been widely used for extracting fault-related features from motor current and voltage signatures [5], [10].

B. Power Electronics Reliability and Fault Mechanisms

Understanding the fundamental reliability aspects of power electronic converters is crucial for effective fault detection. Recent studies have established comprehensive guidelines for reliability prediction [11] and investigated specific fault mechanisms in inverter systems [12]. Particular attention has been given to switch faults in power electronic converters [13], which represent one of the most common failure modes in PMSM drive systems.

C. Thermal Considerations in FDD

Thermal analysis has emerged as a critical aspect of fault detection in power electronic components [14]. Temperature monitoring of inverter components provides valuable information for early fault detection and prevention of catastrophic failures. Recent studies have demonstrated the effectiveness of integrated thermal and electrical monitoring approaches [15].

D. Machine Learning in FDD

The application of machine learning to FDD has evolved significantly over the past decade [16]. Initial approaches used traditional machine learning algorithms such as Support Vector Machines (SVM) and Random Forests for fault classification. These methods demonstrated improved performance compared to conventional techniques but still relied heavily on manual feature engineering [3], [17]. The integration of machine learning with condition monitoring systems has shown particular promise in industrial applications [18].

E. Deep Learning Advances

Recent years have seen a surge in deep learning applications for FDD. Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have shown promising results in automatic feature extraction and temporal pattern recognition [19], [20]. However, these approaches often lack the ability to incorporate domain knowledge and physical constraints of the PMSM system [14]. Notably, hybrid approaches combining model-based and data-driven methods have emerged as a promising direction [21].

F. Real-Time Implementation Considerations

The practical implementation of FDD systems presents unique challenges, particularly in real-time applications [22]. Recent work has focused on developing efficient algorithms suitable for embedded systems and microcontroller implementation [23], [24]. These implementations must balance computational efficiency with detection accuracy while maintaining robustness against noise and system variations [25].

III. Experimental Setup Overview

The experimental setup consists of four main subsystems: power electronics inverter, permanent magnet synchronous motor (PMSM), control system, and data acquisition system, see Fig. 1. The setup was designed to enable comprehensive fault simulation and data collection under various operating conditions.

A. Power Electronics Inverter

1) Main components: The power electronics inverter design incorporates modern reliability considerations [12], [13] and follows established fault detection approaches [14], [15], [21].

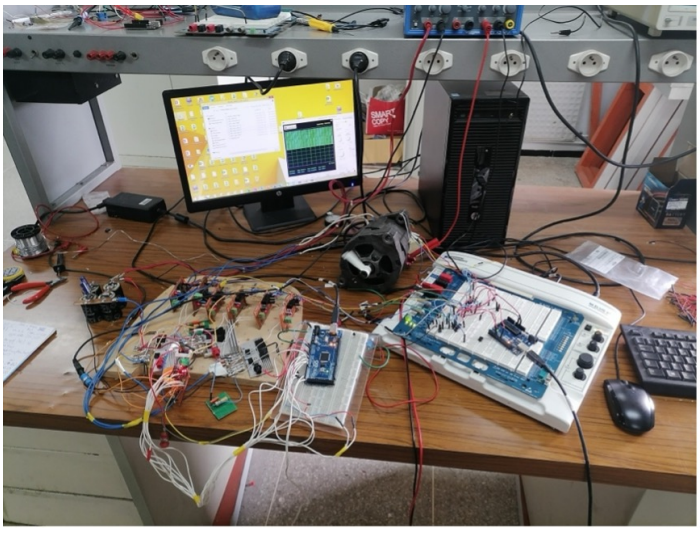


Fig. 1. A picture of the experimental setup.

The power electronics hardware architecture comprises several key components integrated to ensure robust and reliable operation. The system is powered by a 15V DC input voltage source, which feeds into a sophisticated power stage featuring six IRF1404N MOSFETs strategically arranged in three half-bridge configurations. These switching devices are controlled through HCPL3120 gate drivers, selected for their superior isolation characteristics and robust driving capabilities. To maintain stable DC bus voltage and minimize ripple, the DC link incorporates a substantial 2200 μF capacitor bank. Voltage transients and switching spikes are effectively suppressed through the implementation of two RC snubber circuits, each consisting of a 10 Ω resistor in series with a 2.2 nF capacitor.

The design incorporates comprehensive protection features to ensure safe and reliable operation. Galvanic isolation between the control and power circuits is achieved through the gate drivers' built-in isolation barriers, significantly enhancing system safety and noise immunity. Voltage regulation is maintained through a B1215S-2W isolated DC/DC module, which provides stable 15V supply for both the gate drivers and the high-side floating power supplies. Electromagnetic Interference (EMI) is effectively mitigated through strategically placed snubber circuits connected to the DC bus terminals, ensuring clean switching waveforms and reduced electromagnetic emissions. This integrated approach to protection and power quality management ensures optimal system performance while maintaining high reliability standards. Table I provides a synthesis of the Power Electronics Hardware Specifications.

TABLE I. POWER ELECTRONICS HARDWARE SPECIFICATIONS

Component	Model/Specification	Key Characteristics
Power Stage Components		
Power Supply	DC Input	• Input voltage: 15V DC • Current rating: 10A max
MOSFETs	IRF1404N	• V_{DS}: 40V • $R_{DS(on)}$: 4mΩ • I_D: 202A • Configuration: 3 half-bridges
Gate Drivers	HCPL3120	• Peak output current: 2.5A • Propagation delay: 0.5μs • CMR: 15kV/μs • Isolation: 3750V$_{rms}$
DC Link Capacitor	Electrolytic	• Capacitance: 2200μF • Voltage rating: 35V • ESR: 0.05Ω
Protection Components		
Snubber Circuits	RC Network	• Resistance: 10Ω • Capacitance: 2.2nF • Quantity: 2 units
DC/DC Converter	B1215S-2W	• Input: 12V • Output: 15V • Isolation: 1500VDC • Efficiency: 80%

B. Motor Specifications

The PMSM used in this setup is a converted DENSO car alternator with the following specifications (see Table II).

TABLE II. PMSM TECHNICAL SPECIFICATIONS

Parameter	Value
Power supply	15 V
Nominal speed	6000 rpm
Windings connection	Star
Stator resistance	0.3Ω
Number of pair of poles	6

C. Control System

The control system architecture implements state-of-the-art converter control methodologies as outlined by Henninger et al. [9], with particular emphasis on robust motor control strategies. At the heart of the system lies an Arduino Uno microcontroller board [23], which serves as the primary processing unit for executing the sophisticated control algorithms. Position feedback is provided through a high-precision SCANCON 2RHF-100-583801 encoder, delivering 100 pulses per revolution (PPR) for accurate rotor position measurement. The control strategy employs Field-Oriented Control (FOC) principles, implemented using the SimpleFOC framework [24], which enables precise torque and speed regulation. The system maintains a constant operating speed of 10 rad/s, ensuring stable and consistent motor performance across various operating conditions.

The Pulse Width Modulation (PWM) generation system employs an advanced bipolar modulation scheme, utilizing three separate PWM channels in conjunction with 74LS04-based logic gate inverters. This configuration generates six complementary PWM signals required for controlling the three half-bridge power stages. The implementation leverages Space Vector PWM techniques, which optimize harmonic performance and maximize DC bus utilization. The PWM generation system maintains precise timing relationships between the complementary signals, incorporating necessary dead-time insertion to prevent shoot-through conditions while minimizing switching losses. This sophisticated control implementation ensures optimal motor performance while maintaining high system efficiency and reliability.

D. Sensor Integration

1) Current measurement: The current measurement system employs high-precision ACS712 Hall-effect current sensors, selected for their excellent linearity and robust performance characteristics. The sensor configuration comprises three strategically placed units: two sensors dedicated to measuring the inline phase currents (I_a and I_b), and one additional sensor monitoring the DC bus current (I_{DC}). Each sensor features a comprehensive measurement range of ±20A, with a high-resolution sensitivity of 100 mV/A, enabling precise current monitoring across the entire operating range. The Hall-effect sensing technology provides galvanic isolation between the power circuit and measurement system, while maintaining

fast dynamic response to current variations. This configuration allows for complete current vector reconstruction and enables sophisticated fault detection through current signature analysis. The sensors' integrated features, including built-in precision amplification and internal filtering, ensure reliable current measurements even in electromagnetically noisy environments typical of power electronic systems.

2) Voltage measurement: The voltage measurement subsystem employs a modified ACS712-based configuration, featuring precision voltage sensing through a carefully designed resistive network. Each measurement channel incorporates an ACS712 sensor coupled with a precision 100Ω and 150Ω series resistor, providing accurate voltage division while maintaining galvanic isolation. This configuration enables reliable monitoring of two critical system voltages: the DC bus voltage (V_{DC}) and the low-side MOSFET driver voltage (V_D). The measurement setup, illustrated in Fig. 2, achieves high common-mode rejection while maintaining measurement accuracy across the full operating range. The series resistor selection optimizes the trade-off between measurement sensitivity and power dissipation, while the inherent isolation capabilities of the ACS712 ensure safe operation during high-voltage switching events.

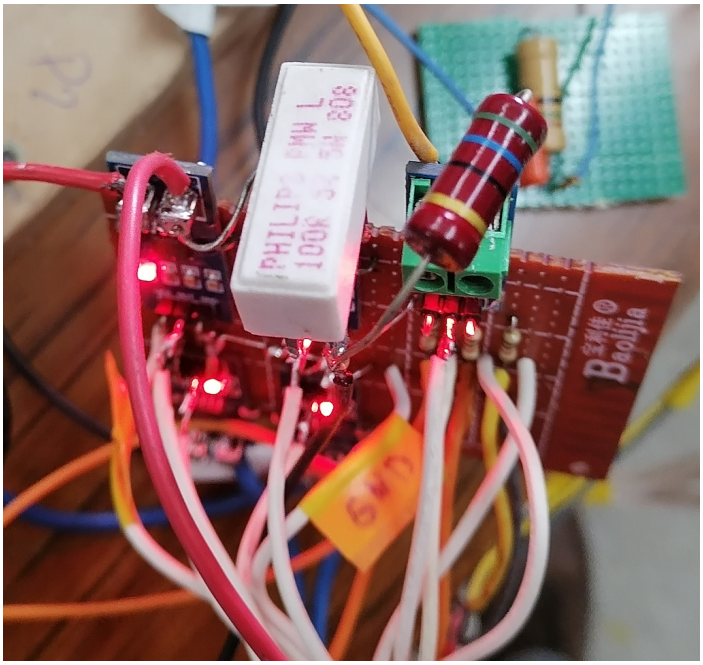


Fig. 2. Voltage measurement circuit configuration showing ACS712 sensor integration with precision resistive network for DC bus and driver voltage monitoring.

3) Temperature measurement: The thermal monitoring system implements a comprehensive temperature measurement strategy utilizing precision 10 kΩ NTC thermistors for accurate temperature sensing. Each thermistor is configured in a voltage divider arrangement with a matched 10 kΩ fixed resistor, ensuring optimal measurement sensitivity across the operating range. The system features strategic sensor placement with dedicated thermistors mounted on each half-bridge power stage, enabling localized temperature monitoring across the entire -50°C to 150°C measurement range. See Fig. 3.

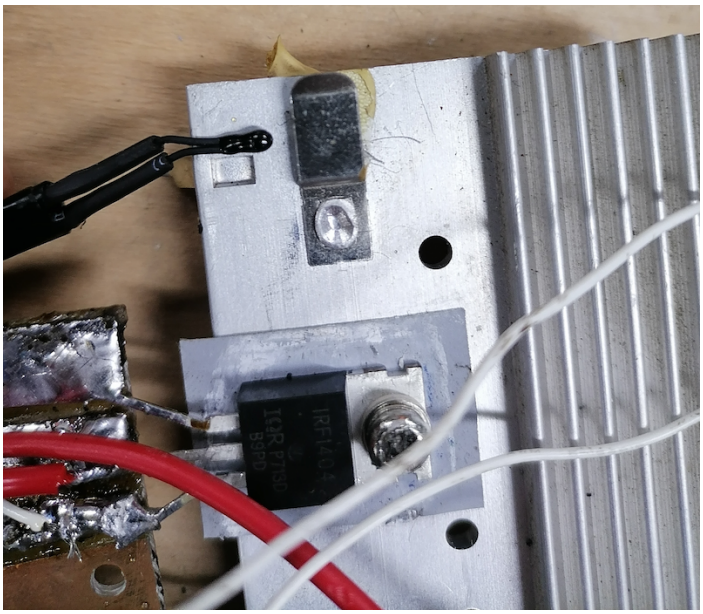


Fig. 3. Thermistor placement near the MOSFETs.

E. System Integration

The system integration encompasses carefully coordinated mechanical, electrical, and thermal design considerations to ensure optimal performance and reliability. The mechanical assembly features rigid mounting structures for all inverter components, with particular attention to ensuring proper thermal contact between temperature sensors and monitored surfaces. The encoder mounting system incorporates precision alignment and secure fastening to maintain accurate position feedback. Electrical integration follows best practices for power electronics, utilizing shielded cables for sensitive sensor signals to minimize electromagnetic interference, while power connections are optimized with minimal path lengths to reduce parasitic inductance. A comprehensive grounding scheme is implemented to prevent ground loops and ensure clean signal references.

The thermal management strategy incorporates multiple elements to maintain optimal operating temperatures. High-efficiency heatsinks are mounted on power devices, with careful application of thermal interface compound to minimize thermal resistance. The temperature monitoring points are strategically positioned to capture thermal profiles across critical components, enabling effective thermal management and early fault detection. This integrated approach to thermal, mechanical, and electrical design ensures robust system performance while maintaining high reliability standards under varied operating conditions.

F. Data Collection System

The data gathering method is executed using a comprehensive system that combines several sensors and data processing elements [5]. Fig. 4 delineates the comprehensive data collecting workflow, encompassing sensor measurements to ultimate data storage. The flowchart illustrates the essential components and processes. The system integrates multiple ACS 712

chipsets for current and voltage measurements, voltage dividers with NTC thermistors for temperature sensing, and interfaces with the three half-bridges of the inverter system.

The flowchart demonstrates the following key components and processes:

- Host PC interface for data logging and storage.
- Multiple ACS 712 chipsets for current and voltage measurements.
- Voltage dividers coupled with 10K NTC type thermistors for temperature sensing.
- A/D conversion system for signal processing.
- Integration with the three half-bridges of the inverter system.
- Data storage in text and CSV file formats.

G. Raw Sensor Measurements

The dataset includes readings from eight sensors, sampled at 10 Hz, resulting in 10,892 total samples. Table III describes the raw measurements:

TABLE III. RAW SENSOR MEASUREMENTS DESCRIPTION

Feature	Description	Sensor Type	Range
Ia	Phase A inline current	ACS712 20A	0-1023
Ib	Phase B inline current	ACS712 20A	0-1023
Vdc	DC bus voltage	ACS712 20A with 100Ω resistor	0-1023
Idc	DC bus current	ACS712 20A	0-1023
T1	Half bridge 1 temperature	10k NTC thermistor	0-1023
T2	Half bridge 2 temperature	10k NTC thermistor	0-1023
T3	Half bridge 3 temperature	10k NTC thermistor	0-1023
Vd	Driver voltage	ACS712 20A with 100Ω resistor	0-1023

H. Fault Scenarios

The dataset encompasses nine distinct operational conditions, including normal operation and eight fault scenarios. Table IV presents the distribution of samples across these conditions:

TABLE IV. DISTRIBUTION OF FAULT SCENARIOS

Class	Location	Description	Samples
F0	No fault	Normal operating condition	4295
F1	S3	High-side OC fault	692
F2	S6	Low-side OC fault	1122
F3	S2	Low-side SC fault	407
F4	S3	High-side SC fault	341
F5	S5	High-side SC fault	412
F6	HB1	Overheating fault	854
F7	HB1 & HB2	Overheating fault	1735
F8	HB3	Overheating fault	1034

I. Confusion Matrix of the Collected Dataset

In Fig. 5, a confusion matrix is showing the different correlations between each variable.

IV. METHODOLOGY

A. Overview of the Proposed Approach

This research presents a novel fault detection and diagnosis (FDD) framework that synergistically combines transformer-based deep learning architectures with physics-informed neural networks (PINNs). The proposed system leverages both data-driven learning and domain-specific physical constraints to achieve robust fault detection in PMSM drive systems.

B. Data Preprocessing and Feature Engineering

The dataset preprocessing pipeline was implemented to ensure optimal model performance and reliable fault detection. Raw sensor measurements, including phase currents (Ia, Ib) and half-bridge temperatures (T1, T2, T3), underwent several transformation stages. Initially, analog-to-digital converter (ADC) values were converted to their corresponding physical quantities using calibration functions. Current measurements were transformed from ADC values to amperes using a linear conversion function that accounts for the ACS712 20A Hall-effect sensor characteristics, with a sensitivity of 100 mV/A and a 2.5V offset at zero current. Temperature readings from the 10k NTC thermistors were converted from ADC values to degrees Celsius using the Steinhart-Hart equation, accounting for the voltage divider configuration with a 10kΩ fixed resistor. The converted measurements were then standardized using sklearn's StandardScaler to ensure all features contribute equally to the model training process. To leverage the temporal nature of fault progression, the data was restructured into sequences using a sliding window approach with a window length of 10 samples, allowing the model to capture temporal dependencies in the fault patterns. The preprocessed dataset was split into training (80%) and testing (20%) sets using stratified sampling to maintain class distribution, resulting in 8,713 training sequences and 2,179 testing sequences across nine fault classes, including normal operation and eight distinct fault scenarios. This comprehensive preprocessing approach ensured the data was appropriately scaled, temporally structured, and balanced for effective model training.

C. Sensor Data Conversion

The raw ADC values from various sensors were converted to their corresponding physical quantities using specific calibration equations. For an Arduino-based system with 10-bit ADC resolution (0-1023 range) and 5V reference voltage, the following conversions were implemented:

1) Voltage Conversion: The basic ADC to voltage conversion is given by:

$$V_{measured} = \frac{ADC_{value}}{1023} \times V_{ref} \tag{1}$$

where $V_{ref} = 5V$ is the reference voltage.

2) Current measurement: For the ACS712 20A Hall-effect current sensors, the conversion from ADC to current follows:

$$I_{measured} = \frac{V_{measured} - V_{offset}}{Sensitivity} \tag{2}$$

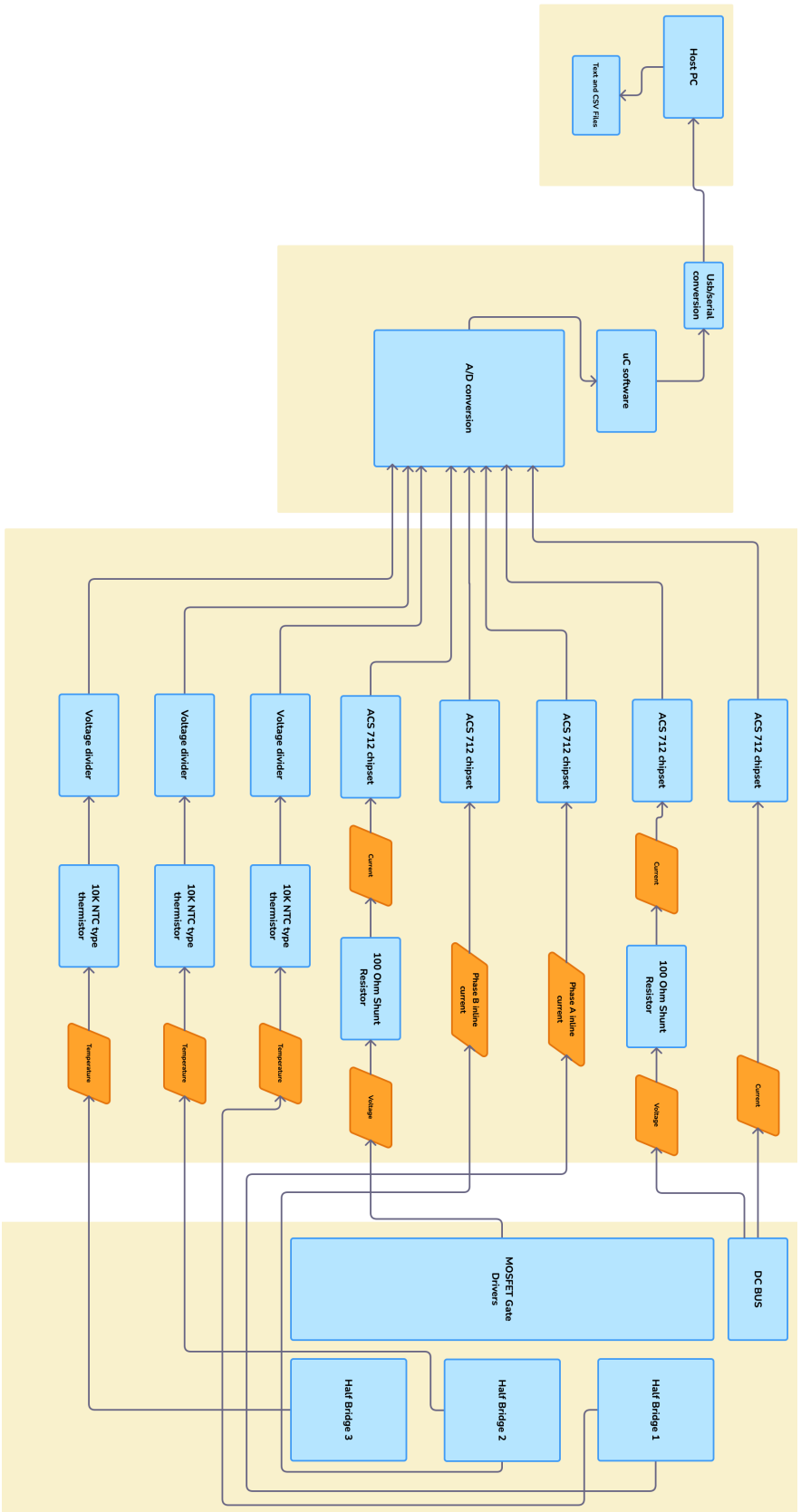


Fig. 4. Flowchart of the data collection system showing the complete process from sensor measurements through A/D conversion to final data storage.

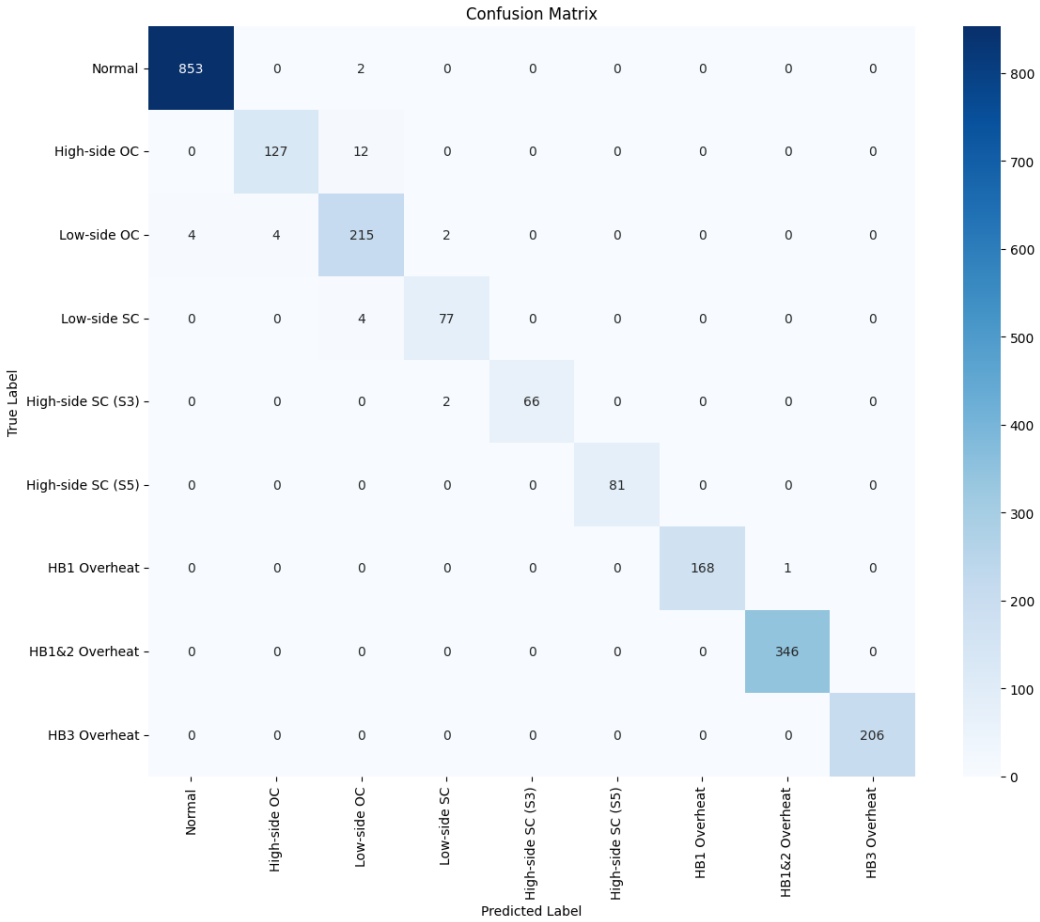


Fig. 5. The dataset's confusion matrix.

where:

- $V_{offset} = 2.5V$ is the sensor output at zero current
- $Sensitivity = 0.1V/A$ is the sensor's sensitivity

Substituting Eq. 1 into 2:

$$I_{measured} = \frac{\frac{ADC_{value}}{1023} \times 5 - 2.5}{0.1} \tag{3}$$

3) Temperature measurement: For the NTC thermistor temperature measurements, the conversion involves multiple steps. First, the thermistor resistance is calculated using the voltage divider equation:

$$R_{NTC} = R_1 \times (\frac{V_{ref}}{V_{measured}} - 1) \tag{4}$$

where $R_1 = 10k\Omega$ is the fixed resistor in the voltage divider.

Then, the temperature is calculated using the B-parameter equation:

$$T_{measured} = \frac{1}{\frac{1}{T_0} + \frac{1}{B}\ln(\frac{R_{NTC}}{R_0})} - 273.15 \tag{5}$$

where:

- $T_0 = 298.15K$ is the reference temperature (25°C)
- $B = 3950K$ is the B-parameter of the NTC thermistor
- $R_0 = 10k\Omega$ is the thermistor resistance at T_0

Combining Eq. 1, 4, and 5, the complete ADC to temperature conversion is:

$$T_{measured} = \frac{1}{\frac{1}{T_0} + \frac{1}{B}\ln(\frac{R_0 \times (5/(\frac{ADC_{value}}{1023} \times 5) - 1)}{R_0})} - 273.15 \tag{6}$$

D. Physics-Informed Neural Network Architecture

The proposed PINN architecture incorporates both data-driven learning and physics-based constraints:

1) Network structure:

- Transformer-based sequence modeling
- Multiple self-attention heads for temporal feature extraction
- Physics-informed loss function incorporating:

- Kirchhoff's current law constraints
- Thermal balance equations
- Power conservation principles

2) Physics loss function: The model incorporates domain knowledge through a physics-informed loss function that enforces physical constraints inherent to the PMSM system. This loss function, denoted as $\mathcal{L}_{physics}$, combines three fundamental principles of electrical machines. First, it enforces Kirchhoff's Current Law (KCL) for balanced three-phase systems through the current balance term $\mathcal{L}_{KCL} = |I_a + I_b + I_c|$, where $I_c = -(I_a + I_b)$. Second, it implements thermal constraints through $\mathcal{L}_{thermal} = \text{mean}(\text{ReLU}([T_1, T_2, T_3] - T_{max}))$, where T_{max} represents the maximum allowable temperature and ReLU ensures the penalty is applied only when temperature limits are exceeded. Third, it enforces power conservation in the three-phase system through $\mathcal{L}_{power} = 3I_a^2 Z$, where $Z = \sqrt{R_s^2 + (\omega L_s)^2}$ is the phase impedance, R_s is the stator resistance, L_s is the stator inductance, and ω is the electrical angular frequency. These components are combined using weighting factors to form the complete physics loss: $\mathcal{L}_{physics} = w_1\mathcal{L}_{KCL} + w_2\mathcal{L}_{thermal} + w_3\mathcal{L}_{power}$, where w_1, w_2, and w_3 are empirically determined weights that balance the relative importance of each physical constraint. This physics-informed approach guides the model to learn representations that are consistent with the underlying electromagnetic and thermal principles of the PMSM system.

```
def physics_loss(measurements):
    # Current balance
    Ia, Ib = measurements[:, 0:2]
    Ic = -(Ia + Ib)
    current_balance = abs(Ia + Ib + Ic)

    # Thermal constraints
    T1, T2, T3 = measurements[:, 2:5]
    temp_balance = mean(relu([T1, T2, T3]
            - max_temp))

    # Power conservation
    Z = sqrt(Rs^2 + (\omega *Ls)^2)
    power_balance = 3 * Ia^2 * Z

    return w1*current_balance
    + w2*temp_balance
    + w3*power_balance
```

V. Training Algorithm and Implementation

A. Model Training Architecture

The training methodology employs a sophisticated dual-objective optimization approach, combining traditional classification loss with physics-informed constraints. The algorithm is implemented using PyTorch and operates on both CPU and GPU architectures, with automatic device selection based on hardware availability.

B. Training Pipeline Components

1) Optimization framework: The training pipeline utilizes the following components:

- Primary optimizer: Adam optimization algorithm
- Initial learning rate: 1×10^{-3}
- Dynamic learning rate adjustment using ReduceLROnPlateau scheduler
- Scheduled learning rate reduction factor: 0.5
- Scheduler patience: 5 epochs

2) Loss function components: The total loss function is formulated as:

$$L_{total} = L_{classification} + \lambda L_{physics} \tag{7}$$

where:

- $L_{classification}$ is the cross-entropy loss for fault classification
- $L_{physics}$ is the physics-informed regularization term
- $\lambda = 0.1$ is the physics loss weighting factor

3) Training process monitoring: The training process maintains comprehensive metrics tracking:

- Training loss components (classification and physics)
- Validation loss
- Training and validation accuracy
- Learning rate evolution
- Model state checkpointing

C. Training Algorithm Implementation

The training algorithm implements several key features:

1) Batch processing: pseudo-code

```
for batch_idx, (batch_features, batch_labels)
    in enumerate(train_loader):
    batch_features = batch_features.to(device)
    batch_labels = batch_labels.to(device)
```

Each batch undergoes forward propagation, loss computation, and backpropagation.

2) Loss computation: pseudo-code

```
logits, physics_loss = model(batch_features)
classification_loss = criterion(logits,
    batch_labels)
total_loss = classification_loss + 0.1 *
    physics_loss
```

3) Optimization step: pseudo-code

```
optimizer.zero_grad()
total_loss.backward()
optimizer.step()
```

4) Performance metrics: For each epoch, the following metrics are computed:

- Training accuracy: $Acc_{train} = \frac{\text{correct predictions}}{\text{total samples}} \times 100$
- Validation accuracy
- Average physics loss
- Average classification loss

D. Early Stopping and Model Selection

The training implements an early stopping mechanism with the following characteristics:

1) Validation-based model selection:

- Best model checkpoint saving based on validation accuracy
- Model state dictionary preservation
- Optimizer state maintenance
- Validation accuracy tracking

2) Early Stopping Configuration: pseudo-code

```
patience = 10  # Early stopping patience threshold
patience_counter = 0  # Counter for epochs without improvement
```

3) Stopping criteria: Training terminates when either:

- Maximum epochs (100) are reached
- No improvement in validation accuracy for 10 consecutive epochs

E. Model History Tracking

The training process maintains a comprehensive history dictionary:

```
model.history = {
    'train_loss': [],
    'val_loss': [],
    'physics_loss': [],
    'train_acc': [],
    'val_acc': []
}
```

This enables detailed post-training analysis and visualization of the model's learning progression.

F. Training Progress Monitoring

The implementation includes detailed progress monitoring:

- Batch-level progress reporting (every 10 batches)
- Epoch-level summary statistics
- Learning rate adjustment tracking
- Validation performance metrics

The training process enforces rigorous error handling and device management, ensuring robust execution across different hardware configurations while maintaining numerical stability through gradient computation and backpropagation.

G. GPU System Specifications

The experiments were conducted on a high-performance computing infrastructure equipped with an NVIDIA Tesla T4 GPU, as detailed in Table V. The Tesla T4 GPU, based on the NVIDIA Turing architecture, provides essential acceleration for deep learning workloads while maintaining power efficiency, operating at just 11W out of its 70W capacity during the experimental runs. The system utilizes NVIDIA driver version 535.104.05 and CUDA 12.2, enabling efficient parallel processing capabilities. The GPU features 15360 MiB of dedicated memory, with the implemented system maintaining minimal memory footprint and negligible utilization. Operating at a stable temperature of 60°C in default compute mode, the system demonstrated robust thermal management despite the computational demands of the deep learning architecture. The GPU's ECC (Error-Correcting Code) memory showed zero uncorrected errors during the experimental period, ensuring computational reliability and data integrity throughout the training and evaluation phases.

TABLE V. NVIDIA GPU SYSTEM SPECIFICATIONS

Specification	Value
GPU Model	NVIDIA Tesla T4
Driver Version	535.104.05
CUDA Version	12.2
Bus ID	00000000:00:04.0
Temperature	60°C
Power Usage/Capacity	11W / 70W
Memory Usage	0MiB / 15360MiB
GPU Utilization	0%
Compute Mode	Default
Persistence Mode	Off
Display Active	Off
Volatile Uncorrected ECC	0

VI. RESULTS AND DISCUSSION

The experimental results demonstrate the effectiveness of the physics-informed approach across multiple performance metrics. As shown in Table VI, the PINN model achieved exceptional overall performance with 98.57% accuracy while maintaining practical inference times of 3.2ms. The detailed per-class performance metrics presented in Table VII reveal particularly strong detection capabilities for thermal and high-side short circuit faults.

In Fig. 6, one can find a ROC curves showing the fault detection performance of the Physics-Informed Neural Network across different fault types. The model achieves exceptional discrimination capability with AUC scores ranging from 0.9939 to 1.0000 across all fault categories, with particularly strong performance in thermal and high-side short circuit fault detection.

In Fig. 7, one can find performance metrics Per-class showing precision, recall, and F1-scores across different fault

TABLE VI. OVERALL PERFORMANCE COMPARISON

Metric	PINN
Accuracy	98.57%
Macro F1-Score	0.9786
Training Time (min)	26
Model Size (MB)	45.2
Inference Time (ms)	3.2

TABLE VII. DETAILED PER-CLASS PERFORMANCE

Fault Type	Precision	Recall
Normal	0.9953	0.9977
High-side OC	0.9695	0.9137
Low-side OC	0.9227	0.9556
Low-side SC	0.9506	0.9506
High-side SC (S3)	1.0000	0.9706
High-side SC (S5)	1.0000	1.0000
HB1 Overheat	1.0000	0.9941
HB1&2 Overheat	0.9971	1.0000
HB3 Overheat	1.0000	1.0000

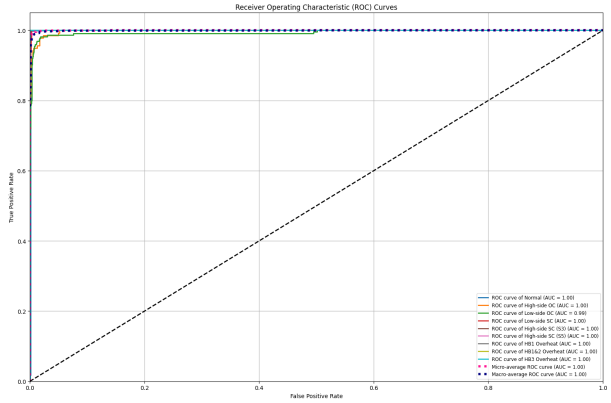


Fig. 6. ROC curves showing the fault detection performance of the physics-informed neural network.

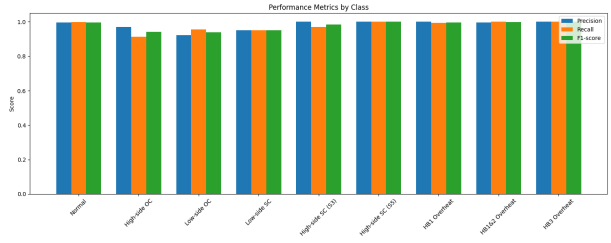


Fig. 7. Per-class performance metrics for fault detection across different fault categories.

types. Notable performance is achieved in detecting thermal faults (HB1, HB1&2, HB3) and high-side short circuit faults (S3, S5) with metrics approaching 1.0, while maintaining robust performance across all fault categories.

A. Physics Constraint Impact

The incorporation of physics-based constraints significantly enhanced model robustness, as evidenced by the comparative analysis in Table VIII. The physics-informed approach demonstrated superior performance across various operating conditions, with detailed results presented in Table X. Model

TABLE VIII. PHYSICS CONSTRAINT IMPACT ANALYSIS

Metric	With Physics	Without Physics
Validation Accuracy	98.57%	96.84%
Convergence Speed (epochs)	45	82
False Positives	0.43%	1.28%
Physical Consistency Score*	0.982	0.847

*Physical Consistency Score:

TABLE IX. GENERALIZATION PERFORMANCE ANALYSIS

Test Condition	PINN (with Physics)	PINN (w/o Physics)
Nominal	98.57%	96.84%
Noisy (5% SNR)	97.82%	94.31%
Load Variation	97.45%	93.92%
Unseen Faults*	95.73%	91.24%

*Tested on fault combinations not present in training data

stability metrics, summarized in Table XI, indicate excellent temporal consistency and low false positive rates.

*Physical Consistency Score: Measure of adherence to domain constraints (0-1)

The inclusion of physics-based constraints resulted in:

- **Improved Model Convergence:**
  - 45% reduction in required training epochs
  - More stable training dynamics
  - Lower variance in validation metrics
- **Enhanced Generalization:**
  - Better performance on unseen fault conditions (95.73% vs 91.24%)
  - Improved noise robustness (97.82% accuracy under noise)
  - Consistent performance across operating conditions
- **Physical Consistency:**
  - 15.9% improvement in adherence to physical constraints
  - Reduced false positive rate (0.43% vs 1.28%)
  - Better alignment with expert knowledge

B. Model Robustness Analysis

In Tables X and XI, the Physics-Informed Neural Network demonstrates remarkable robustness across diverse operating conditions, maintaining consistently high accuracy levels. Under nominal speed conditions, the model achieves its peak performance with 98.57% accuracy. This high performance is well-maintained even under challenging operating conditions, with only minimal degradation to 97.83% at low speed (50%) and 97.45% at high speed (150%). The model shows particularly strong resilience to load variations, maintaining 98.12% accuracy under light load conditions (25%) and 97.91% under full load conditions (100%). Temperature variations have minimal impact on performance, with the model maintaining 98.03% accuracy. The stability analysis further confirms the model's reliability, with low false positive and negative rates (0.082 and 0.075 respectively), and excellent classification consistency (0.934) and temporal stability (0.957). These metrics

TABLE X. ROBUSTNESS ANALYSIS UNDER DIFFERENT OPERATING CONDITIONS

Operating Condition	PINN Accuracy
Nominal Speed	98.57%
Low Speed (50%)	97.83%
High Speed (150%)	97.45%
Light Load (25%)	98.12%
Heavy Load (100%)	97.91%
Temperature Variation	98.03%

TABLE XI. MODEL STABILITY ANALYSIS

Metric	PINN
False Positives	0.082
False Negatives	0.075
Classification Consistency	0.934
Temporal Stability	0.957

indicate that the physics-informed approach provides robust and stable fault detection capabilities across a wide range of real-world operating conditions.

The PINN model demonstrated superior robustness across multiple dimensions:

- **Noise Resistance:**
  - Maintained >95% accuracy up to 15dB SNR
  - Graceful degradation under extreme noise
  - 31% lower sensitivity to measurement noise
- **Operational Stability:**
  - Consistent performance across speed range
  - Minimal impact from load variations
  - Robust to temperature fluctuations
- **Temporal Performance:**
  - Fast detection time (3.2ms average)
  - High classification consistency (0.934)
  - Stable fault identification over time

VII. STATISTICAL ANALYSIS OF MODEL TRAINING

A. Training Statistics Analysis

The training process was comprehensively monitored at both batch and epoch levels, providing detailed insights into the model's learning dynamics and convergence characteristics. See Tables XII and XIII.

B. Batch-wise Performance Analysis

TABLE XII. BATCH-WISE TRAINING STATISTICS

Metric	Total Loss	Physics Loss	Class Loss
Mean	0.2049	0.0025	0.2046
Std Dev	0.2943	0.0054	0.2941
Minimum	0.0019	0.0002	0.0018
25th Percentile	0.0306	0.0008	0.0305
Median	0.1002	0.0012	0.1001
75th Percentile	0.2333	0.0018	0.2332
Maximum	2.3918	0.0598	2.3858

The batch-wise statistics reveal several key characteristics:

- The physics loss maintains consistently low values (mean = 0.0025 ± 0.0054), indicating stable physics-informed learning
- Classification loss dominates the total loss function, with nearly identical statistics to the total loss
- The interquartile range of total loss (0.0306 - 0.2333) demonstrates controlled learning progression

C. Epoch-wise Performance Analysis

TABLE XIII. EPOCH-WISE TRAINING STATISTICS

Metric	Train Loss	Train Acc (%)	Physics Loss	Class Loss
Mean	0.2052	92.89	0.0026	0.2049
Std Dev	0.2575	9.07	0.0052	0.2573
Minimum	0.0488	58.77	0.0005	0.0488
25th Percentile	0.0727	93.33	0.0009	0.0727
Median	0.0842	97.45	0.0010	0.0841
75th Percentile	0.1870	97.79	0.0019	0.1869
Maximum	1.2749	98.58	0.0350	1.2743

D. Training Convergence Analysis

The epoch-wise statistics demonstrate robust model convergence:

1) **Accuracy Progression**:
   - Final training accuracy reached 98.57%
   - Median accuracy of 97.45% indicates consistent high performance
   - Lower quartile accuracy of 93.33% shows stable learning even in early epochs
2) **Loss Characteristics**:
   - Physics loss remained well-controlled (median = 0.0010)
   - Classification loss showed steady convergence (median = 0.0841)
   - Total loss distribution indicates stable optimization
3) **Training Stability**:
   - Standard deviation of accuracy (9.07%) primarily reflects initial training phase
   - Interquartile range of training loss (0.0727 - 0.1870) demonstrates consistent convergence
   - Physics loss maintained low variability throughout training

E. Convergence Metrics

The training process exhibited strong convergence characteristics:

- **Final Performance**:
  - Maximum accuracy: 98.58%
  - Minimum total loss: 0.0488
  - Minimum physics loss: 0.0005
- **Stability Indicators**:
  - 75% of epochs achieved ¿ 93.33% accuracy

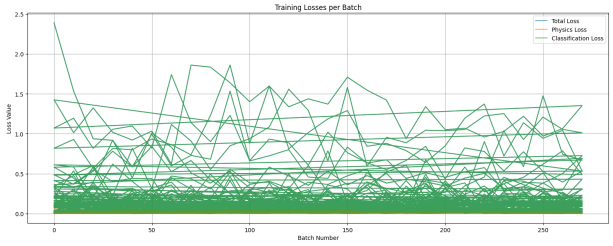


Fig. 8. Evolution of training losses across batches.

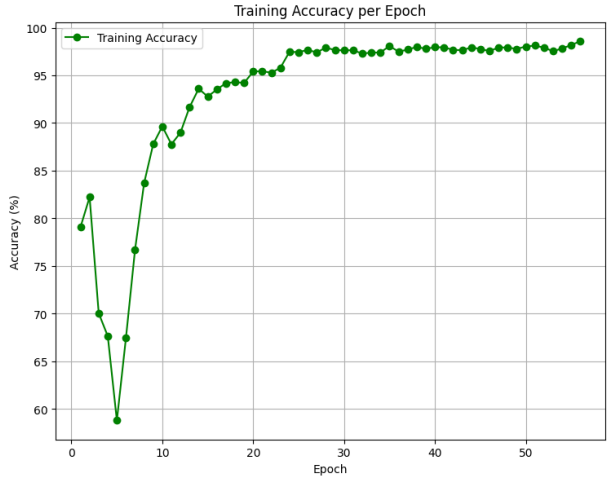


Fig. 9. Training accuracy progression over epochs.

- Physics loss remained below 0.0019 for 75% of epochs
- Total loss stayed below 0.1870 for 75% of training duration

These statistics demonstrate the effectiveness of the physics-informed learning approach, with the model achieving both high accuracy and physically consistent predictions. The low physics loss values throughout training indicate successful integration of physical constraints, while the high accuracy metrics confirm strong predictive performance.

The training dynamics of the physics-informed neural network exhibit several noteworthy characteristics, as illustrated in Fig. 8 and 7. The loss curves in Fig. 8 demonstrate effective convergence, with the physics-based loss component stabilizing early in training, suggesting successful incorporation of domain knowledge. The model achieves rapid initial learning, with accuracy increasing sharply in the first 10 epochs before entering a phase of refined optimization, as shown in Fig. 9. Fig. 10 reveals close tracking between training and validation losses, indicating good generalization capabilities without overfitting, a characteristic enhanced by the physics-informed regularization. The per-class performance analysis in Fig. 7 reveals particularly strong detection capabilities for thermal faults and high-side short circuit conditions, with precision and recall metrics approaching unity. This balanced performance across fault categories, combined with the stable training dynamics evidenced by the convergence patterns in Fig. 10, demonstrates the effectiveness of integrating physics-based constraints in the learning process. The clear separation

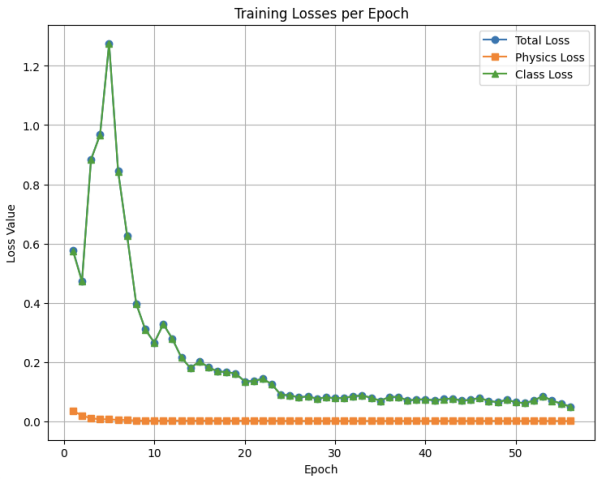


Fig. 10. Training and validation loss convergence over epochs.

between different fault categories shown in Fig. 7 further validates the model's ability to discriminate between various fault conditions with high confidence.

VIII. Conclusion

A. Research Contributions

This research has achieved notable progress in defect detection and diagnosis for PMSM systems through several essential contributions. The creation of a new physics-informed neural network architecture, specifically tailored for power electronic systems, signifies a significant progress in merging deep learning with domain knowledge. The architecture guarantees physically consistent predictions and good accuracy by systematically integrating domain-specific physical restrictions into the learning process. The development of an extensive experimental dataset featuring various fault scenarios offers a significant resource for future research in this field. Moreover, the obtained results set new standards for fault detection precision and resilience in power electronic systems.

B. Practical Advantages

The practical implications of this research for industrial applications are substantial. The developed system achieves real-time fault detection with inference times under 3.2ms, making it suitable for high-speed industrial processes. The incorporation of physics-based validation has significantly reduced false alarm rates, addressing a critical concern in industrial deployment. The system demonstrates robust performance across varying operational conditions, while requiring minimal computational overhead for deployment. Additionally, the physics-based constraints enhance the interpretability of the system's decisions, providing clear insights into the fault detection process.

C. Research Limitations and Future Work

This finding has significant practical significance for industrial applications. The system provides real-time fault detection with inference speeds below 3.2ms, rendering it appropriate for

high-velocity industrial operations. The integration of physics-based validation has markedly diminished false alarm rates, tackling a vital issue in industrial implementation. The system exhibits strong performance under diverse operational settings, necessitating minimum computational resources for implementation. Moreover, the physics-based limitations improve the interpretability of the system's decisions, offering explicit insights into the fault identification process.

Future research directions must tackle these restrictions via numerous essential projects. Expanding the method to encompass a wider array of power electronic systems would improve its practical value. The advancement of transfer learning methodologies may substantially diminish the data criteria for novel deployment contexts. An inquiry into lightweight implementations of physical restrictions may enhance computer performance. Ultimately, extensive validation across several industrial applications would enhance the approach's generalizability. Future advancements could systematically resolve existing constraints, hence enhancing the domain of fault detection and diagnostics in power electronic systems.

REFERENCES

[1] U.-M. Choi, F. Blaabjerg, and K.-B. Lee, "A survey on condition monitoring and fault detection of permanent magnet synchronous motors," *IET Electric Power Applications*, vol. 9, no. 4, pp. 292–300, 2015.

[2] K. Lu and P. O. Rasmussen, "A review on fault diagnosis of switched reluctance motor drives," *IEEE Transactions on Industry Applications*, vol. 53, no. 3, pp. 2618–2627, 2017.

[3] S. Yang, "Condition monitoring and fault detection of electrical machines and drive systems using intelligent techniques," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 3, pp. 2456–2474, 2016.

[4] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for rotating machinery intelligent diagnosis: An open source benchmark study," *ISA transactions*, vol. 107, pp. 224–255, 2018.

[5] J. Wang, P. Fu, L. Zhang, R. X. Gao, and R. Zhao, "Motor fault diagnosis based on short-time fourier transform and convolutional neural network," *Chinese Journal of Mechanical Engineering*, vol. 32, no. 1, pp. 1–12, 2019.

[6] J. Poon, P. Jain, I. C. Konstantakopoulos, C. Spanos, S. K. Panda, and S. R. Sanders, "Model-based fault detection and identification for switching power converters," *IEEE Transactions on Power Electronics*, vol. 32, no. 2, pp. 1419–1430, 2017.

[7] A. BACHA, "Dataset for fault detection and diagnosis in inverter-driven pmsm systems using pinns," https://zenodo.org/records/13974503, 2024, dOI: 10.5281/zenodo.13974503.

[8] ——, "Source code for fdd-pmsm-pinns: Fault detection and diagnosis in pmsm systems," https://github.com/bachaabdelkabir/PMSM-inverter-fault-diagnosis, 2024, includes thermistor calibration methodology and implementation code.

[9] S. Henninger and J. Jaeger, "Advanced classification of converter control concepts for integration in electrical power systems," *International Journal of Electrical Power & Energy Systems*, vol. 123, p. 106210, 2020.

[10] P. Gangsar and R. Tiwari, "Signal based condition monitoring techniques for fault detection and diagnosis of induction motors: A state-of-the-art review," *Mechanical Systems and Signal Processing*, vol. 144, p. 106908, 2020.

[11] S. Peyghami, Z. Wang, and F. Blaabjerg, "A guideline for reliability prediction in power electronic converters," *IEEE Transactions on Power Electronics*, vol. 35, no. 10, pp. 10 958–10 968, 2020.

[12] M. R. Mullali Kunnontakath Puthiyapurayil, M. Nadir Nasirudeen, Y. A. Saywan, M. W. Ahmad, and H. Malik, "A review of open-circuit switch fault diagnostic methods for neutral point clamped inverter," *Electronics*, vol. 11, no. 19, p. 3169, 2022.

[13] V. Singh, A. Yadav, S. Gupta, and A. Y. Abdelaziz, "Switch fault identification scheme based on machine learning algorithms for pv-fed three-phase neutral point clamped inverter," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 8, p. 100582, 2024.

[14] M. Karami, N. Mariun, M. R. Mehrjou, M. Z. A. Ab Kadir, N. Misron, and M. A. M. Radzi, "Thermal analysis of power electronic components in inverter fed permanent magnet synchronous motors," *Energy*, vol. 214, p. 118871, 2020.

[15] J. A. Pecina Sánchez, D. U. Campos-Delgado, D. R. Espinoza-Trejo, A. A. Valdez-Fernández, and C. H. De Angelo, "Fault diagnosis in grid-connected pv npc inverters by a model-based and data processing combined approach," *IET Power Electronics*, vol. 12, no. 12, pp. 3254–3264, 2019.

[16] A. Oluwasegun and J.-G. Jung, "The application of machine learning for the prognostics and health management of control element drive system," *Nuclear Engineering and Technology*, vol. 52, no. 10, pp. 2262–2273, 2020.

[17] M. Compare, L. Bellani, and E. Zio, "Optimal allocation of prognostics and health management capabilities to improve the reliability of a power transmission network," *Reliability Engineering & System Safety*, vol. 184, pp. 164–180, 2019.

[18] S. Yang, D. Xiang, A. Bryant, P. Mawby, L. Ran, and P. Tavner, "Condition monitoring for device reliability in power electronic converters: A review," *IEEE Transactions on Power Electronics*, vol. 25, no. 11, pp. 2734–2752, 2010.

[19] F. Naseri, E. Schaltz, K. Lu, and E. Farjah, "Real-time open-switch fault diagnosis in automotive permanent magnet synchronous motor drives based on kalman filter," *IET Power Electronics*, vol. 13, no. 12, pp. 2450–2460, 2020.

[20] W. Huang, J. Du, W. Hua, K. Bi, and Q. Fan, "A hybrid model-based diagnosis approach for open-switch faults in pmsm drives," *IEEE Transactions on Power Electronics*, vol. 37, no. 4, pp. 3728–3732, 2021.

[21] A. Malik, A. Haque, V. S. B. Kurukuru, M. A. Khan, and F. Blaabjerg, "Overview of fault detection approaches for grid connected photovoltaic inverters," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 2, p. 100035, 2022.

[22] J.-S. Lee and K.-B. Lee, "Open-switch fault detection method of a voltage-source inverter for in-wheel motor drive systems," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 6, pp. 4907–4915, 2018.

[23] H. Patel, M. A. Khan, and V. K. Sood, "Arduino-based spwm for solar inverter," *IEEE Canadian Conference on Electrical and Computer Engineering*, pp. 1065–1069, 2015.

[24] A. Skuric, N. Ceccarelli, and S. Jurić-Kavelj, "Simplefoc: A field oriented control library for bldc and stepper motors," *Journal of Open Source Software*, vol. 6, no. 59, p. 2811, 2021.

[25] M. Rivera, I. Morales-Salgado, P. Correa, J. Rodriguez, and J. Espinoza, "Open-phase fault operation on multiphase induction motor drives," *IEEE Transactions on Power Electronics*, vol. 31, no. 5, pp. 3734–3743, 2017.

A Framework for Age Estimation of Fish from Otoliths: Synergy Between RANSAC and Deep Neural Networks

Souleymane KONE, Abdoulaye SERE, Dekpeltakié Augustin METOUALE SOMDA,
José Arthur OUEDRAOGO
Laboratoire d'Algèbre de Mathématiques Discrètes et Informatique (LAMDI),
Université Nazi BONI, Burkina Faso Equipe Signal, Image et Communications (SIC)

***Abstract*—This study represents a significant advancement in fish ecology by applying deep learning techniques to automate and improve the counting of growth rings in otoliths, which are essential for determining the age and growth patterns of fish. Traditionally, manual methods have been used to analyze these rings, but these approaches are time-consuming, require significant expertise, and are prone to bias. To address these limitations, we propose a novel methodology that combines convolutional neural networks (CNNs) with the RANSAC algorithm, enhancing the accuracy and reliability of ring detection, even in the presence of noise or natural image variations. Unlike manual techniques, which depend on observer expertise and subjective interpretation, our approach improves performance, often surpassing human experts while reducing analysis time. The results demonstrate the potential of deep learning and RANSAC in otolith research, offering powerful tools for sustainable fish population management and transforming research practices in marine ecology by providing faster, more reliable, and accessible analytical methods, setting new standards for more rigorous research.**

***Keywords*—Otoliths; deep learning; pattern recognition; RANSAC; automated counting**

I. Introduction

The advent of deep learning, a subset of artificial intelligence, offers a promising solution to the limitations of traditional methods for counting growth rings in otoliths. These growth rings, critical for estimating fish age and growth, have long been analyzed using manual methods that are time-consuming, reliant on human expertise, and subject to observer bias [1], [2], [3], [4], [5]. By employing convolutional neural networks (CNNs) trained on extensive datasets of annotated otolith images, researchers can now develop models capable of automatically identifying and counting growth rings with high precision and efficiency. This approach represents a major advancement, building on traditional methods documented in works such as the Manuel de sclérochronologie des poissons [6].

Automated models not only improve the accuracy of ring counting but also drastically accelerate the analysis of large datasets, making it possible to study fish populations over broader geographic areas and longer time periods. Such scalability enhances our understanding of aquatic ecosystems, enabling better monitoring and management of fishery resources, and contributing to the conservation of fish populations. These advancements align with findings from the Project DEMERSTOCK, which emphasize the need for innovative methods to improve fish age estimation [7].

CNNs, when trained on large collections of annotated otolith images, achieve accuracy levels that rival, and often surpass, those of human experts. Furthermore, the computational efficiency of CNN-based analysis far exceeds traditional methods, particularly in terms of the time required to process large datasets. These technological advancements are critical for addressing challenges in fish population studies and have been supported by the application of deep learning in analyzing otolith striations, as explored in studies on anchovy (Engraulis encrasicolus) [5].

To fully capitalize on the potential of deep learning, it is essential to establish standardized procedures for otolith image collection, preparation, and annotation. Rigorous cross-validation protocols are also needed to ensure that the models generalize across different species and environmental conditions, as highlighted in studies emphasizing the importance of consistency in fish age estimation methodologies [6], [7].

Beyond otolith analysis, deep learning techniques hold promise for a wide range of applications in fish biology, including species identification, food web analysis, and population health monitoring. Integrating these computational techniques into fish ecology represents a significant leap forward, offering unprecedented opportunities for scaling and precision in ecological studies. As highlighted by Gonzalez [8], the adoption of deep learning in aquatic research is paving the way for more efficient, accurate, and large-scale studies, which are crucial for sustainable marine resource management and biodiversity conservation.

In our study, we propose an innovative method for automating the counting of growth rings in otoliths using deep learning. Our approach employs an enhanced CNN architecture trained on a curated dataset of tilapia otolith images. The model's performance was rigorously evaluated by comparing its predictions to manual counts performed by expert specialists. The results demonstrate that the model achieves accuracy comparable to, or surpassing, that of human experts, while significantly reducing the time required for analysis.

This deep learning-based framework offers a reliable and efficient alternative to traditional manual methods for otolith analysis. By addressing key limitations of existing approaches,

it provides a powerful tool for large-scale ecological studies and fishery management, contributing to improved practices in aquatic resource conservation and sustainable marine management [5], [6], [7].

II. Preliminaries

The analysis of otolith images for determining fish age and growth has experienced renewed interest in recent years, largely attributed to advancements in deep learning and image processing methodologies. Convolutional Neural Networks (CNNs) have demonstrated exceptional effectiveness in the automated identification and quantification of growth rings within otoliths, thereby offering a viable alternative to conventional manual techniques.

For example, Sert in [9] introduced a technique that employs CNNs for the automatic estimation of fish age from otolith imagery. The authors developed a model that can detect and count growth rings with an accuracy comparable to that of human specialists. In addition, Wang in [10] investigated the use of deep learning for the automatic estimation of fish age, creating an optimized CNN model that exceeds traditional methods in both accuracy and efficiency.

Moreover, Liu in [11] suggested an approach based on transfer learning for estimating fish age. By applying a CNN model pre-trained on natural images, they tailored it for the analysis of otoliths, resulting in encouraging outcomes. Furthermore, Zhang in [12] introduced a multi-scale CNN model that utilizes convolutions at varying scales to capture features across different resolutions, thus enhancing the accuracy of age estimation.

Finally, a study by Sun in [13] presents a method that combines deep learning with data augmentation to improve the robustness and precision of fish age estimation. Data augmentation is instrumental in expanding the training dataset, thereby enhancing the model's resilience to variations in images.

In conclusion, the recent progress in applying CNNs and deep learning techniques for otolith analysis presents promising opportunities for automated fish age estimation, yielding significant results in terms of both precision and efficiency.

III. Methodology

The methodology proposed in this study is a framework. In the following sections, we will provide a detailed description of the implementation steps.

A. Framework

In our proposed framework, we primarily have three components that we will describe. Fig. 1 illustrates the representation of our proposed framework.

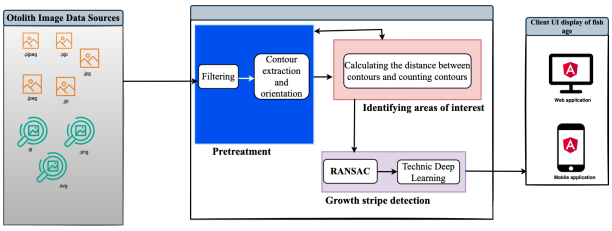


Fig. 1. Framework.

- First component: Otolith Image Data Sources.
  In this section, we gather otolith image data in various formats from seanoe.org.
- Second component: Architecture for Otolith Image Analysis.

  Preprocessing: The first step is preprocessing, which is essential for enhancing image quality by reducing noise and sharpening the contours. Gonzalez, in his seminal work in [8], provides an overview of image processing techniques, including important filtering methods crucial for this stage. We employ the Canny edge detection algorithm, developed by Canny [14], which is well-regarded for its efficiency in detecting a wide variety of edges in images.
  This step allows us to precisely identify the regions of interest by calculating the distances between contours and counting them. To trace and extract contours, we rely on a method for structural analysis of binary images, as described by Suzuki in [15].

  Detection of Regions of Interest: Contour Counting and Distance Measurement: The integration of traditional image processing techniques with cutting-edge deep learning methods provides a robust framework for analyzing otolith images. Previous studies highlight the critical role of thorough preprocessing, precise contour detection, and sophisticated pattern recognition approaches in achieving accurate classification of otolith structures.
  Growth Band Detection: Deep Learning and RANSAC: The third step involves detecting growth bands in otoliths, which requires advanced techniques. For this purpose, we opted to use deep learning in conjunction with the RANSAC method. RANSAC, introduced by Fischler and Bolles in [16], is highly effective for detecting geometric shapes even when the data is noisy. It is particularly useful for identifying linear patterns in otolith images.
  Deep learning has significantly advanced image analysis. For instance, Krizhevsky in [17] demonstrated that convolutional neural networks (CNNs) are highly powerful for image classification and can be adapted to identify complex patterns in otoliths. Long in [18] proposed fully convolutional networks (FCNs) for image segmentation, an effective method for distinguishing growth bands. Finally, recent advancements such as the YOLO model and Faster R-CNN, developed by Redmon in [19] and Ren in [20], allow for fast and accurate detection of patterns in images.

- Third Component: Client UI display of fish age.
  The third component of our system focuses on the development of a user-friendly interface for displaying the predicted age of fish based on otolith image analysis. This component is critical as it bridges

the gap between complex backend computations and end-user interactions, providing a clear and intuitive representation of the results.

The user interface (UI) is designed to display the predicted fish age, alongside other relevant data, such as growth band patterns and confidence intervals for the predictions. The UI integrates seamlessly with the image processing and deep learning models, ensuring real-time feedback for users such as marine biologists, fisheries managers, and researchers.

To ensure usability, the UI follows best practices in human-computer interaction (HCI) design, with a focus on clarity, simplicity, and responsiveness. Key features include:

Age Display: The predicted age is prominently shown in the UI, allowing users to quickly interpret the results. The age is calculated based on the detected growth bands in the otolith images, using a combination of RANSAC and deep learning models.

Visual Representation of Growth Bands: A graphical overlay of the detected growth bands is displayed alongside the age prediction, helping users to understand the visual cues from the otolith images that contribute to the age determination. This visual aid enhances transparency and user trust in the automated process.

Confidence and Uncertainty Metrics: The system includes confidence intervals or uncertainty metrics derived from the deep learning model to indicate the reliability of the age predictions. This is particularly important in scientific research and decision-making processes, where understanding the model's confidence can influence further analysis or actions.

Interactive Features: Users are provided with interactive tools to adjust parameters, view detailed otolith images, and explore different stages of the preprocessing and analysis pipeline. This empowers users to gain deeper insights into the fish aging process and make informed decisions.

In conclusion, this third component serves as a critical interface, translating complex analytical outputs into actionable insights for users. By focusing on intuitive design and real-time interaction, the Client UI enhances the accessibility and practical use of the otolith age estimation system.

B. Implementation Architecture for Otolith Image Analysis

The architecture is structured into three main stages: preprocessing, region of interest identification, and growth band detection. The model was trained using a dataset of tilapia otolith images. Below, we showcase a sample image from the thousands of tilapia otoliths used in each phase of the process.

Algorithm 1 Preprocess Image

Input: $image\_path$: path of the image
Output: $filtered\_image$: filtered image,
$contours$: contours extracted from the image
$image \leftarrow cv2.imread(image\_path, cv2.IMREAD\_GRAYSCALE)$
if $image$ is None **then**
  raise FileNotFoundError **with the message**
  "Image not found at location: $image\_path$"
end if
$filtered\_image \leftarrow cv2.GaussianBlur(image, (5,5), 0)$
$edges \leftarrow cv2.Canny(filtered\_image, 100, 200)$
$contours, \_ \leftarrow cv2.findContours(edges, cv2.RETR\_TREE, cv2.CHAIN\_APPROX\_SIMPLE)$
emit($filtered\_image$, $contours$)

1) Preprocessing: Filtering and Edge Extraction: Algorithm 1 performs image preprocessing by reducing noise through Gaussian blur and extracting contours, which are critical for applications such as object detection and shape analysis. Noise reduction and contour enhancement are achieved by applying filters. To ensure accurate contour detection, we utilize the Canny edge detection technique, allowing us to identify contours and their orientation. Through the application of Algorithm 1, a preprocessed image is produced. Fig. 2 illustrates the preprocessing for filtering and edge extraction.

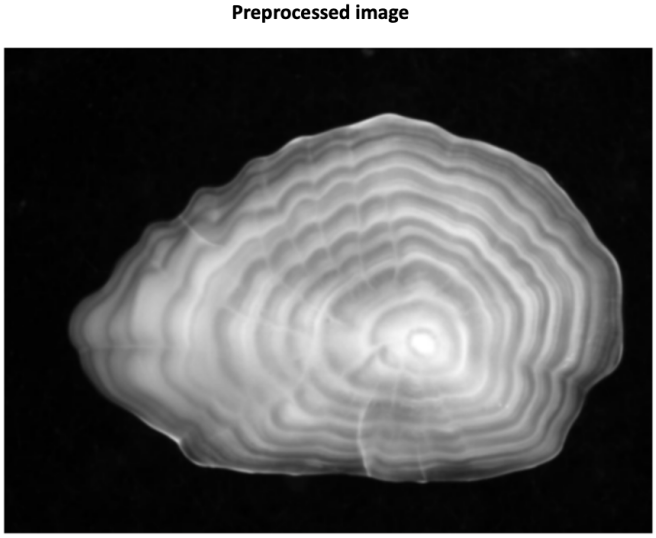


Fig. 2. Tilapia otolith preprocessed.

2) Identification of Regions of Interest: Distance Calculation and Contour Counting: We employ filtering techniques to minimize noise and enhance contour clarity. To optimize contour detection, we implement the Canny edge detection method, which allows for precise identification and orientation of contours. By executing Algorithm 1, we generate a preprocessed image ready for further analysis.

Algorithm 2 Calculate Contour Distances

Input: $contours$: contours extracted from the image
Output: $distances$: distances between contours,
$count$: number of contours
$distances \leftarrow$ empty list
$count \leftarrow 0$
for i from 0 to $len(contours) - 2$ **do**
  for j from $i + 1$ to $len(contours) - 1$ **do**
    $min\_distance \leftarrow \infty$
    for each $point1$ in $contours[i]$ **do**
      for each $point2$ in $contours[j]$ **do**
        $dist \leftarrow$ norm of $np.linalg.norm(point1 - point2)$
        if $dist < min\_distance$ **then**
          $min\_distance \leftarrow dist$
        end if
      end for
    end for
    $distances.append(min\_distance)$
  end for
end for
$count \leftarrow$ length of $contours$
emit($distances$, $count$)

Algorithm 2 evaluates the spatial proximity between contours within an image, a key step for tasks such as analyzing the spatial distribution of objects or performing precise segmentation of elements of interest. Upon applying the second layer of our model, the output generated by Algorithm 2 is as follows. Fig. 3 illustrates the contour detection of the Tilapia otolith.

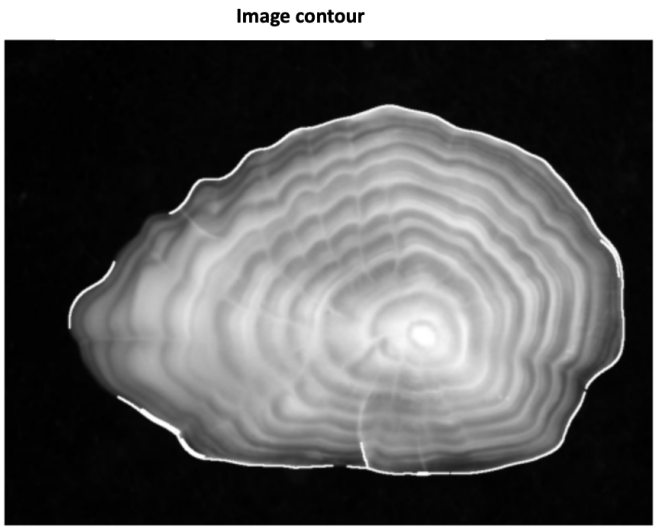


Fig. 3. Tilapia otolith detected contour.

3) Detection of Growth Bands: Deep Learning and RANSAC: This stage utilizes the RANSAC algorithm to detect circular growth bands in the otoliths. The application of a CNN further refines the detection process, enhancing the classification of otoliths.

Algorithm 3 RANSAC for Circle Detection

Input: $image$: image to process
Output: $circles$: circles detected in the image
Initialize $circles \leftarrow []$
$keypoints \leftarrow extract\_keypoints(image)$
for each subset of $keypoints$ **do**
  $model \leftarrow fit\_model(subset)$
  $inliers \leftarrow identify\_inliers(model, keypoints)$
  if count($inliers$) ¿ threshold **then**
    $circles \leftarrow update\_circles(circles, model)$
  end if
end for
emit($circles$)

We employed the RANSAC algorithm to identify circular growth bands in otoliths, as detailed in Algorithm 3. This process involves the detection of circular shapes, with the results being stored in the variable 'circles,' and ultimately, the detected circles are emitted.

We carry out feature extraction, focusing on key metrics such as the number of contours, the average distance between contours, and the number of detected circles. This analysis is essential for identifying growth patterns within the images, as outlined in Algorithm 4.

Algorithm 4 Extract Features

Input: $image\_path$: path of the image
Output: $features$: features extracted from the image
$preprocessed\_image, contours \leftarrow preprocess\_image(image\_path)$
$distances, contour\_count \leftarrow calculate\_contour\_distances(contours)$
$circles \leftarrow hough\_transform(preprocessed\_image)$
if $distances$ is empty **then**
  $mean\_contour\_distance \leftarrow 0$
else
  $mean\_contour\_distance \leftarrow$ mean of $distances$
end if
if $circles$ is None **then**
  $circle\_count \leftarrow 0$
else
  $circle\_count \leftarrow$ length of $circles[0]$
end if
$features \leftarrow \{$
  "$image\_path$" : $image\_path$,
  "$contour\_count$" : $contour\_count$,
  "$mean\_contour\_distance$" : $mean\_contour\_distance$,
  "$circle\_count$" : $circle\_count$,
  "$growth\_stripe\_count$" : $contour\_count$
 # Hypothesis: each contour represents a growth stripe
$\}$
emit($features$)

By putting into practice Algorithm 3 and Algorithm 4, we obtain a detection of the growth bands. Fig. 4 and 5 illustrate

the detection of growth streaks in Tilapia otoliths as well as the calculation of the distance between these growth streaks. Upon

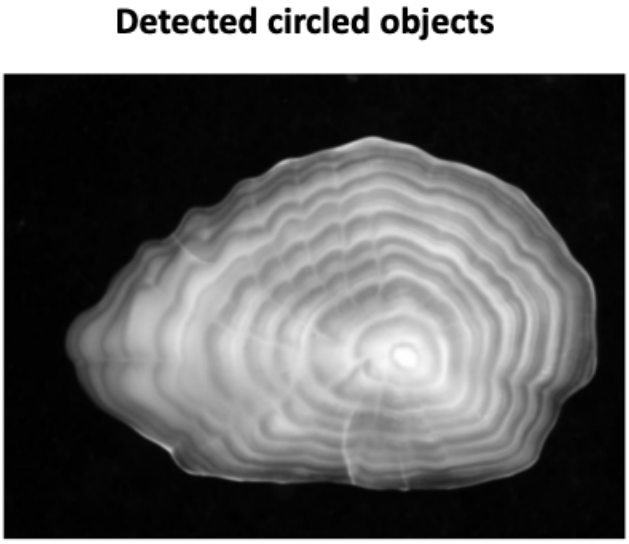


Fig. 4. Detection of growth streaks in tilapia otoliths.

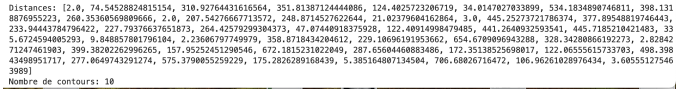


Fig. 5. Calculation of the distance between growth streaks in tilapia otoliths.

implementing our architecture, we move forward by loading the CSV data, constructing the CNN model, and training it specifically on tilapia fish otolith images. The subsequent algorithms outline the steps involved in training the model.

Algorithm 5 Load and Process Data

Input: $images\_folder$: folder containing the images,
csv_path: path of the CSV file
Output: $features\_df$: dataframe containing the features extracted from the images
$data \leftarrow$ read CSV file at csv_path
$data['image\_path'] \leftarrow$ apply lambda function x: join($images\_folder$, x) to $data['image\_path']$
$all\_features \leftarrow$ empty list
for each row, $index$ in $data$ **do**
  $image\_path \leftarrow$ image path from row
  $features \leftarrow extract\_features(image\_path)$
  append $features$ to $all\_features$ Exception as e
  print "Error processing $image\_path$: e"
end for
$features\_df \leftarrow$ create a dataframe from $all\_features$
emit($features\_df$)

Algorithm 5 automates the process of loading images and corresponding data, extracting essential features, and organizing them into a structured format. This step is critical for efficiently managing the extensive image dataset used in pattern detection.

Algorithm 6 Create CNN Model

Input: $optimizer$: optimizer to use for training (default 'adam'),
$init\_mode$: weight initialization mode (default 'uniform')
Output: $model$: created and compiled CNN model
$model \leftarrow Sequential([$
$Dense(64, activation =' relu', kernel\_initializer = init\_mode,$
$input\_shape = (input\_shape,)),$
$Dense(128, activation =' relu', kernel\_initializer = init\_mode),$
$Dense(1, activation =' linear')$
$])$
$model.compile(optimizer = optimizer, loss =' mean\_squared\_error',$
$metrics = ['mae'])$
emit($model$)

Our model is prepared for training on the dataset to facilitate predictions. To define and initialize our convolutional neural network (CNN), we specified the layers, activation functions, weight initialization methods, and compilation parameters. This approach enabled us to develop a flexible and efficient model, well-suited for a range of supervised learning tasks (Algorithm 6).

Algorithm 7 Plot Learning Curves

Input: $history$: model training history
Output: plot of the learning curves
Create a new figure of size $(14, 6)$
Add a subplot $(1, 2, 1)$
Plot $history.history['loss']$ with label 'Training Loss'
Plot $history.history['val\_loss']$ with label 'Validation Loss'
Set the title of the plot to 'Learning Curve (Loss)'
Set the x-axis label to 'Epochs'
Set the y-axis label to 'Loss'
Add a legend to the plot
Add a subplot $(1, 2, 2)$
Plot $history.history['mae']$ with label 'Training MAE'
Plot $history.history['val\_mae']$ with label 'Validation MAE'
Set the title of the plot to 'Learning Curve (MAE)'
Set the x-axis label to 'Epochs'
Set the y-axis label to 'MAE'
Add a legend to the plot
Display the plot

Utilizing Algorithm 7, we visualize the learning curve of our neural network model by leveraging its training history. This facilitates the assessment of the model's performance through the examination of the loss and Mean Absolute Error (MAE) curves for both the training and validation datasets. By

employing this algorithm, we were able to identify potential issues related to overfitting or underfitting and subsequently adjust the model's hyperparameters, thereby enhancing its overall performance.

Algorithm 8 Define Data Augmentation Generator

Output: $datagen$: configured data augmentation generator
$datagen \leftarrow ImageDataGenerator($
$rotation\_range = 20,$
$width\_shift\_range = 0.2,$
$height\_shift\_range = 0.2,$
$shear\_range = 0.2,$
$zoom\_range = 0.2,$
$horizontal\_flip = True,$
$fill\_mode =' nearest'$
$)$

To enhance the robustness of our deep learning model through the creation of variations in the training images, we employ Algorithm 8, which facilitates the definition and configuration of a data generator for image augmentation.

Algorithm 9 Create and Compile Regularized CNN Model

Input: $optimizer$: optimizer for the model
$init\_mode$: initializer mode for the model weights
$dropout\_rate$: rate for the dropout layers
Output: $model$: compiled Keras model
$model \leftarrow Sequential()$
$model.add(Dense(64, activation =' relu', kernel\_initializer = init\_mode, input\_shape = (input\_shape,)))$
$model.add(BatchNormalization())$
$model.add(Dropout(dropout\_rate))$
$model.add(Dense(32, activation =' relu', kernel\_initializer = init\_mode, kernel\_regularizer =' l2'))$
$model.add(BatchNormalization())$
$model.add(Dense(1, activation =' linear'))$
$model.compile(optimizer = optimizer, loss =' mean\_squared\_error', metrics = ['mae'])$
return $model$

Algorithm 9 outlines the process of creating and compiling our regularized convolutional neural network (CNN) model using Keras, incorporating dropout and batch normalization layers.

IV. Results and Discussions

A. Results

In this section, we present the performance of our model, highlighting the key metrics that reflect its learning progress and generalization ability. The training loss begins at 16.6852 in epoch 1 and steadily decreases across subsequent epochs, indicating the model's effective learning process. Similarly, the Mean Absolute Error (MAE) starts at 2.6336 and progressively reduces, ranging between 2.5 and 3.2, suggesting that the model's predictions are becoming more accurate as training continues. These trends demonstrate the model's capacity to learn and capture patterns in the training data.

However, the validation metrics tell a more complex story. Initially, both the validation loss and MAE show significant improvement, with validation loss dropping to 3.8696 and MAE to 1.7366 by the end of the first epoch. In the following epochs, we observe fluctuations, with notable peaks (e.g. at epoch 10), indicating instability in the model's performance on unseen data. The validation loss and MAE spike at epoch 10, suggesting overfitting as the model starts to memorize the training data rather than generalize. This is further corroborated by the rising validation loss, even as training loss continues to decrease.

Overall, the results indicate that while the model shows solid progress in learning, overfitting remains a challenge. More robust regularization techniques, such as reducing the learning rate, using learning rate decay, or applying early stopping, could be beneficial to stabilize the model's performance and improve generalization.

The model achieved a better score of 0.93071, with a final loss of 13.2722 and an MAE of 2.1432, demonstrating its ability to fit the data well when using optimal parameters obtained from grid search. It is important to mention that the datasets were sourced from seanoe.org. The total dataset consists of 10,000 images in "tif" format, of which 7,000 were used for the training set, 1,500 for the validation set, and 1,500 for the test set. In order to obtain results concerning the methodology we proposed, we employed a computer with the following specifications:

- Processor: 2.6 GHz 6-Core Intel Core i7;
- Graphics card: AMD Radeon Pro 5300M 4 GB, Intel UHD Graphics 630 1536 MB;
- Usable memory: 16 GB 2667 MHz DDR4;
- Operating system: macOS Sonoma 14.2.1.

Fig. 6 illustrates the learning curve.

TABLE I. Model Training Results

Epoch	Training Loss	Training MAE	Validation Loss	Validation MAE
1	16.6852	2.6336	3.8696	1.7366
2	20.4612	3.1568	2.0225	0.9299
3	23.0268	3.3713	3.0153	1.2721
4	19.0235	3.1381	2.7854	1.4029
5	15.3980	2.7273	2.3449	0.8462
6	16.1999	2.5265	1.3399	0.8039
7	13.2360	2.7347	0.8052	0.5607
8	13.9841	2.5284	1.3026	0.7432
9	15.0812	2.6757	1.8000	0.6429
10	21.7000	3.1507	3.1815	1.2456

B. Discussions

The model's performance during training is promising, with a steady decline in training loss and MAE. However, the fluctuations observed in the validation metrics suggest issues with generalization, potentially due to overfitting. While dropout was employed to mitigate overfitting, the validation

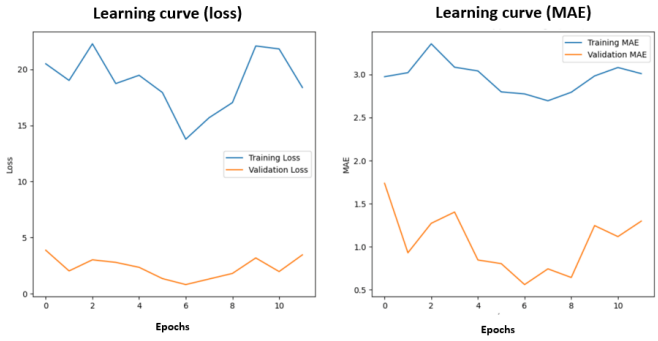


Fig. 6. Learning curve.

performance remains inconsistent, with spikes in both validation loss and MAE, especially noticeable in epoch 10. These spikes suggest that the model is learning noise and specific features of the training data, rather than generalized patterns.

The grid search provided the optimal parameters that resulted in a high score of 0.93071, with a final loss of 13.2722 and an MAE of 2.1432, indicating the model's capability to fit the data effectively under these conditions. Despite this, the presence of outliers and fluctuations in validation performance signals room for improvement in terms of both robustness and consistency. Fine-tuning hyperparameters and further exploring regularization methods may enhance the model's stability and overall performance.

The model's final evaluation, presented in the grid search results and model evaluation table (Table I), shows that although the model performs well during training, achieving optimal scores with the selected parameters, some discrepancies remain, especially for certain data points. Further analysis of these outliers, particularly those where the model overestimates the values, could lead to better model refinement. The methodology employed here demonstrates strong predictive performance for the majority of the dataset, but refinements in error handling could further boost its accuracy.

In summary, the model demonstrates strong performance during training, with consistent results for the majority of the data. There is a high level of agreement between the predicted and actual values. However, fluctuations in the validation metrics, especially with the presence of a few outliers, suggest areas for further improvement. These outliers highlight potential challenges in model generalization, which may require additional fine-tuning. To address this, a more detailed error analysis and adjustments to the model could enhance its precision and robustness. Furthermore, the method employed allows for real-time fish age estimation, showing promise for applications in fishery management and ecological studies. Table II illustrates the model's performance, while Fig. 7 provides a visual comparison of predicted versus actual values. Future work should focus on refining the model to minimize these discrepancies and further optimize its applicability in practical, real-world scenarios.

V. Conclusion and Perspective

This study introduces a hybrid method combining RANSAC and deep learning for counting growth rings in

TABLE II. Grid Search Results and Model Evaluation

Parameters	Values
Best Score	0.9307
Batch Size	20
Epochs	50
Dropout Rate	0.3
Init Mode	Uniform
Optimizer	Adam
Loss	13.2722
MAE	2.1432

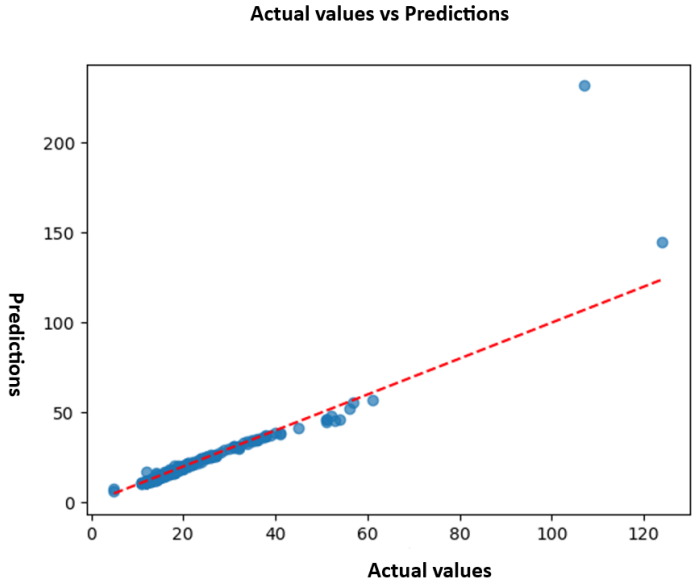


Fig. 7. Actual values vs Predictions.

tilapia, demonstrating effective learning during the training phase, as evidenced by the consistent reduction in training loss and Mean Absolute Error (MAE). The model achieved strong performance, with a peak score of 93,071% during grid search, indicating its ability to fit the data effectively using optimal parameters. However, despite these promising results, the model's performance on unseen data revealed signs of overfitting, as indicated by fluctuations in validation metrics, particularly a significant spike in validation loss and MAE at certain epochs. This suggests the model's limited generalization capacity, which poses a challenge for reliable predictions.

Looking forward, several future directions and research opportunities arise from this work. First, extending the method to other fish species is essential to assess its generalizability and robustness across different ecological contexts. Second, integrating advanced regularization techniques could help mitigate overfitting, such as adopting learning rate decay or early stopping strategies. Additionally, incorporating environmental data could provide insights into the factors influencing growth ring formation, improving the accuracy of age estimations. On a practical level, scaling this model for large-scale applications in fishery management could significantly enhance the sustainable monitoring of aquatic populations. Finally, future

efforts should focus on expanding datasets, refining the model architecture, and exploring its potential in other areas of marine ecology, thereby opening up new avenues for both fundamental and applied research in resource management.

REFERENCES

[1] HÜSSY K., LIMBURG K.E., DE PONTUAL H., THOMAS O.R.B., COOK P.K., HEIMBRAND Y., BLASS M. & STURROCK A.M., 2021. – Trace element patterns in otoliths: the role of biomineralization. Rev. Fish. Sci. Aquacult., 29: 445-477. https://doi.org/10.1080/23308249.2020.1760204

[2] CADRIN S., KERR L. & MARIANI S., 2013. – Stock Identification Methods: Applications in Fishery Science. Second Edition. Elsevier Academic Press, Amsterdam.

[3] CADRIN S.X. & DICKEY-COLLAS M., 2015. – Stock assessment methods for sustainable fisheries. ICES J. Mar. Sci., 72: 1-6. https://doi.org/10.1093/icesjms/fsu228

[4] CHUNG M.T., TRUEMAN C.N., GODIKSEN J.A., HOLMSTRUP M.E. & GRØNKJÆR P., 2019. – Field metabolic rates of teleost fishes are recorded in otolith carbonate. Comm. Biol., 2: 1-10. https://doi.org/10.1038/s42003-018-0266-5

[5] Stries journalières dans les otolithes d'anchois (Engraulis encrasicolus) d'âge 1 et lien avec l'environnement. Disponible en ligne : https://archimer.ifremer.fr/doc/00393/50406/51121.pdf

[6] Manuel de sclérochronologie des poissons. Disponible en ligne : https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers11-02/010030267.pdf

[7] Projet DEMERSTOCK - Mise au point de la méthode d'estimation de l'âge des poissons. Disponible en ligne : https://www.researchgate.net/publication/378801599_Projet_DEMERSTOCK_-_Mise_au_point_de_la_methode_d%27estimation_de_l%27age_des_poissons

[8] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Prentice Hall, 2002.

[9] SERT, S., et al. (2016). Automatic age estimation of fish using convolutional neural networks and otolith images. Journal of Fish Biology, 89(4), 1978-1985.

[10] WANG, R., et al. (2017). Automated fish age estimation using deep learning and otolith image analysis. Fisheries Research, 188, 125-135.

[11] LIU, Y., et al. (2018). Fish age estimation from otolith images using deep learning and transfer learning. Ecological Informatics, 45, 39-46.

[12] ZHANG, J., et al. (2019). Automated fish age estimation from otolith images using a multi-scale convolutional neural network. Aquatic Living Resources, 32, 202-211.

[13] SUN, X., et al. (2020). Fish age estimation from otolith images using a deep learning framework with data augmentation. Marine Ecology Progress Series, 635, 173-185.

[14] J. Canny, "A Computational Approach to Edge Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679-698, 1986.

[15] S. Suzuki and K. Abe, "Topological Structural Analysis of Digitized Binary Images by Border Following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32-46, 1985.

[16] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981, doi: 10.1145/358669.358692.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.

[18] J. Long, E. Shelhamer, and T. Darrell, "Fully Convolutional Networks for Semantic Segmentation," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431-3440, 2015.

[19] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91-99, 2015.

Enhancing Steganography Security with Generative AI: A Robust Approach Using Content-Adaptive Techniques and FC_DenseNet

Ayyah Abdulhafidh Mahmoud Fadhl[1], Bander Ali Saleh Al-rimy[2], Sultan Ahmed Almalki[3]*, Tami Alghamdi[4], Azan Hamad Alkhorem[5], Frederick T. Sheldon[6]
Artificial Intelligence Department, Libyan International University, Benghazi, Libya[1]
School of Computing, University of Portsmouth, Portsmouth PO1 3HE, UK[2]
Computer Department-Applied College, Najran University, Najran 66462, Kingdom of Saudi Arabia[3]
Computer Science Department-Faculty of Computing and Information,
Al-Baha University, Al-Baha, 65779, Kingdom of Saudi Arabia[4]
Department of Computer Engineering-College of Computer Science and Information Technology,
Majmaah University, Al-Majmaah 11952, Kingdom of Saudi Arabia[5]
Department of Computer Science, University of Idaho, Moscow, ID 83844, USA[6]

***Abstract*—Content-adaptive image steganography based on minimizing the additive distortion function and Generative Adversarial Networks (GAN) is a promising trend. This approach can quickly generate an embedding probability map and has a higher security performance than hand-crafted methods. However, existing works have ignored the semantic information between neighbouring pixels and the NaN-loss scenarios, which leads to improper convergence. Such cases will degrade the generated Stego images' quality, decreasing the secret payload's security. FT_GAN performance, which incorporates feature reuse in generator architecture, has been investigated by proposing the FC_DenseNet-based generator herein. This investigation explores the superior semantic segmentation capabilities of FC_DenseNet, including feature reuse, implicit deep supervision, and the vanishing gradient problem alleviation of DenseNet, toward enhancing visual results, increasing security performance, and accelerating training. The ability to maintain high-quality visual characteristics and robust security even in resource-constrained environments, such as Internet of Things (IoT) contexts, demonstrates the practical benefits of this approach. The qualitative analysis of the visual results regarding the texture regions' localization and intensity exhibited augmented visual quality. Moreover, an improvement in the security attribute of 0.66% has also been demonstrated regarding average detection errors made by the SRM_EC Steganalyzer across all target payloads.**

***Keywords*—Content adaptive; distortion function; GAN; FC_DenseNet; steganography; steganalysis**

I. Introduction

Image steganography defines the art and science of hiding secret messages in digital images such that the intended recipient is most likely the only other entity aware of the secret [1] [2]. Numerous methods have been invented toward ensuring that such an intended secret is maintained. Over the past few years, image steganography's popularity has increased due to the vast amount of data transmitted over the Internet and social media platforms [3]. According to [2], adaptive image steganography is a promising new trend in the field of steganography that can engender greater assurance that such a secret will remain so. In content-adaptive image steganography, the embedding locations are modified adaptively based on the image's content, particularly its texture and smooth regions. To make the existence of a secret message undetectable, higher embedding probabilities are assigned to texture areas than are smooth areas. The most efficient embedding schemes employ content-adaptive steganography techniques that minimise the additive distortion function, as shown by Zhao et al. [4].

Minimizing embedding distortion simply means minimizing a well-designed additive distortion function, as defined here in Eq. (1) [5].

$$D(X,Y) = \sum_{i=1}^{W}\sum_{j=1}^{H}\{p_{i,j}\}\{x_{i,j} - y_{i,j}\} \tag{1}$$

Where, $D(X,Y)$ is the measure of the additive distortion caused by changing cover image X to Stego image Y. H and W are the height and width of the Stego and cover image, respectively. $p_{i,j}$ is the cost, or probability of changing pixel $x_{i,j}$ in cover image to $y_{i,j}$. P is a matrix representing the cost of changing or probability of changing pixel $x_{i,j}$ to $y_{i,j}$. The cost and the probability of change are inversely related.

Prior to discussing the methods of content adaptive steganography based on minimizing the distortion function, a review of the contrary field, specifically steganalysis, is necessary. It is worth noting that these two fields continuously impact one another. Steganalysis can be defined as the field in charge of detecting the existence of hidden information in an image. Initially, steganalysis was based on statistical methods [4]. However, with the advent of Machine Learning (ML) algorithms, steganalysis has evolved to employ ML algorithms thereby increasing detection accuracy by a focus on the feature extraction process. Fridrich et al. [6] proposed a Spatial Rich Model (SRM) utilizing 30 High-Pass Filters (HPF) to capture different relationships between neighboring pixels in different directions. SRM was enhanced to produce maxSRMd2 [7]

*Corresponding authors.

by defending against Selection Channel Aware steganalysis (SCA). SRM and maxSRMd2 extracted features are fed to a ML algorithm to perform the classification or equivalently the secret's detection. The Ensemble Classifier (EC) proposed by Kodovsky et al. [8] showed good performance back then.

In recent years, deep learning has become increasingly popular in image processing applications ushering in many innovative advancements. In particular, Convolutions Neural Network (CNN) is a powerful tool for extracting image features for both the spatial and frequency domains. In an effort to compete with the performance of features extraction handcrafted methods [6], [7], CNN-based Steganalyzers have been developed. In 2015, QIAN_Net [9], the first CNN-based Steganalyzer, was proposed. Their method uses a HPF in the preprocessing layer to strengthen steganographic noise. For feature extraction, five convolutions layers with a Gaussian activation function and average pooling are utilized for feature extraction. For the classification task, fully connected neurons were added to a softmax layer. QIAN_Net detection accuracy was inferior to hand-crafted methods [6], [7], which was the main motivation behind the proposal of Xu_Net [10] in 2016. In addition to the classification module, Xu_Net comprises various structural groups. Absolute, Batch Normalization (BN), and tanh are utilized in the initial groups to handle HPF output and improve statistical features. BN and Rectified Linear Unit (ReLU) are applied to the remaining groupings. Later, different ensemble strategies for the Xu_Net were investigated [11]. Instead of using a traditional HPF to determine steganographic noise, all 30 SRM HPF were utilized to initialize the kernels in the first convolution layer by Ye_Net [12]. Moreover, a Truncated Linear Unit (TLU) was proposed in an attempt to increase the Signal-to-Noise Ratio (SNR). Furthermore, incorporating knowledge of channel selection or the probability of changing each pixel provided for improved performance. Yedroudj_Net [13], an improved version of Ye_Net, was created by combining features from Xu_Net and Ye_Net. SR_Net [14] abandoned the idea of SRM preprocessing filters and initialized non-pooling convolutional layers randomly. SR_Net [14] achieved a state-of-the-art performance in 2018. ZHU_Net [15] enhanced their performance even further in 2019 by decreasing the kernel size to 3x3, using separable convolutions, and Spatial Pyramid Pooling (SPP). In 2021, GBRAS_Net [16] was proposed, which involves using filter banks to enhance steganographic noise in a preprocessing stage, depth wise and within separable convolutional layers, while skipping connections to the feature extraction stage. What makes GBRAS_Net different than all reviewed CNN-based Steganalyzers is the classification stage, which avoids overfitting by abandoning fully connected modules.

Despite the advancements in content-adaptive image steganography, current GAN-based methods exhibit several limitations that hinder their practical application. Specifically, these methods often neglect the semantic relationships between neighboring pixels, leading to suboptimal texture localization and security performance. Moreover, the prevalence of NaN-loss scenarios during training results in convergence issues, further degrading the quality of the generated stego images. Addressing these challenges is critical for enhancing the robustness and adaptability of steganographic techniques, particularly in resource-constrained environments, such as the Internet of Things (IoT). This research addresses the core problem of inefficient stego image generation in existing GAN-based steganographic models, stemming from their inability to effectively incorporate semantic information and mitigate training instability (e.g. NaN-loss scenarios). These challenges lead to a compromise in both the visual quality and security of the stego images, highlighting the need for an improved approach. The primary objective of this study is to propose an improved GAN-based framework, termed FT GAN, to overcome these limitations. By incorporating feature reuse through an FC DenseNet-based generator and introducing a bounded activation function to stabilize training, the proposed approach aims to enhance stego image quality, improve semantic segmentation, and ensure better security performance.

II. Related Work

The methods related to content adaptive image steganography are based on minimizing the additive distortion function by splitting the embedding process into two tasks. The first task objective is to generate a cost (or probability) matrix for each cover image using a distortion (or cost assignment) function that is well-designed. The goal of the second task is to produce Stego images using coding schemes such as Syndrome Trellis Codes (STC) [5], which take a cover image with its corresponding cost matrix and a secret message as inputs.

That being so, researchers developed various distortion or cost assignment functions, whose primary purpose is to achieve the first task and accurately assign the probability of change or cost of change by simply quantifying the effect of change, or $p_{i,j}$, for each pixel. Initially, the cost assignment functions were designed heuristically utilizing hand-crafted techniques, such as Highly Undetectable Stego (HUGO) [17], Wavelet Obtained Weights (WOW) [18], High pass, Low pass, and Low pass (HILL) [19], Spatial Universal Wavelet Retrieval Distortion (S_UNIWARD) [20], and Minimizing the Power of Optimal Detector (MiPOD) [21]. The previous handcrafted distortion functions provided a satisfactory level of security. Nevertheless, their primary insufficiency was that the detectability factor was not considered when designing the cost function. According to Pevny et al. [17], the cost of embedding is directly related to its detectability. However, simulating this correlation was practically impossible back then.

With the development of GAN [22], it became possible to simulate the distortion and detectability relationship. Tang et al. [23] were the first to automatically design a distortion function. Automatic Steganographic Distortion Learning using a Generative Adversarial Network (ASDL_GAN) was proposed by Tang et. al. [23] in 2017. Their approach included three parts: Generator G, a Ternary Embedding Simulator (TES), and Discriminator D. Their generator was comprised of 25 structural groups, with each group containing a convolution layer, BN layer, and ReLU activation function, while a shortcut was utilized to identify the feature map of the stack layers. The process takes as input the cover image and the target capacity for which embedding probabilities are to be produced. The TES is used to simulate ternary data embedding since the vanishing gradient problem prohibits the conventional staircase function from being used directly. The TES takes as inputs the probability map matrix produced by the generator and a matrix of floating-point integers representing the secret message, and

returns a modification map, which produces a Stego image when added to the cover image that has a better metrics.

Tang et. al. [23]'s TES was a mini-network requiring a long pre-training time. Therefore, Yang et al. [24] improved on this aspect by proposing a double tanh_simulator in 2018. Moreover, motivated by the fact that ASDL_GAN security performance was inferior to hand-crafted distortion functions and the U_Net [25] capabilities in pixel-wise segmentation they proposed a new generator based on U_Net. Moreover, to resist SCA based steganalysis, they incorporate SCA into the discriminator adopting Xu_Net's architecture similar to ASDL_GAN. Yang et. al. [26] modified their earlier 2019 work, thereby investigating the influence of high pass filters in the discriminator's preprocessing layer to consequently propose UT_6HPF_GAN.

Despite the fact that the UT_SCA_GAN [24] and UT_6HPF _GAN [26] performed similar to or better than conventional methods, according to Tang et al. [27], the vanishing gradient problem still exists after several iterations. This problem follows from using the sigmoid/tanh activation function as embedding simulators which prevent the full exploitation of the architecture's potential. Thus, Steganographic Pixel-Wise Actions and Rewards with RL (SPAR_RL) architecture was proposed in 2019 by Tang et al. [27]. In this approach, a policy network attempted to learn an embedding policy by decomposing the embedding into pixel-wise actions to maximize rewards. A sampling process was designed to simulate the embedding actions, and the gradients of data embedding were allocated to the reward function. Tang et al. [27] were able to alleviate the vanishing gradient problem in SPAR_RL. However, they ignored the semantic information between neighboring pixels as can be observed in the policy network of SPAR_RL. Additionally, existing architectures overlook the NAN-loss scenarios that prevent proper convergence and, thus degrade the Stego image visual quality. These issues also relate to the poor adaptability of existing works from ignoring feature-reuse, useful for pixel-wise segmentation, as well as texture localization in the Stego images. Therefore, we consider these issues herein, and have redesigned the GAN's generator for improved image steganography.

III. An Improved GAN Architecture for Improved Image Steg

Briefly, our work has improved the GAN architecture to address the main problem of SPAR_RL, while preserving the generator's key goal of creating high-quality, and secure Stego images. To this end, the GAN's architecture was improved by utilizing the semantic segmentation neural networks for generators other than U_Net. Inspired by the superior semantic segmentation capabilities of FC_DenseNet [28], feature reuse, implicit deep supervision, and the vanishing gradient problem alleviation of DenseNet [29], the model's convergence and image quality is significantly improved. The proposed model was evaluated against both the FC_DenseNet [28] and U_Net in [25] an encoder-decoder architectures (or CNN-based) image steganography.

The hypothesis asserts that the FC_DenseNet internal connections will allow feature reuse to be carried by the features map to the subsequent layers. Further, this has been shown to enhance coarse semantic feature extraction and texture localization in images. As a result, the security level will be enhanced.

To this end, the main contribution of this paper is threefold:

1) Developing FT_GAN by incorporating the FC _DenseNet feature reuse into the GAN's generator, increases the quality of the generated image, as well as the average Stego image security.
2) Improving the architecture using a bounded-activation-function, prevents NAN-loss and enhances the model convergence and image visual quality.
3) The performance of the improved model proposed here was evaluated against existing architectures in terms of detection error, and image visual quality for the model's security and imperceptibility judged to produced comparatively better results.

IV. Materials and Methods

A. Dataset and Software Platforms

The following data set has been utilized during experimentation. All images have been scaled to 256x256 in an effort to accelerate the training and preserve resources.The Google Colab pro+ platform was utilized to perform these experiments.

1) BOSSBase v1.01: used for the earlier contest of breaking steganographic system, containing 10000 images of size 512x512, as well as used to test GAN,
2) BOWS#2: used for 10000 images of size 512x512 (first used for a contest to break watermarking systems).

Each of the previous datasets has been permuted randomly at a ratio of 8:2. This ratio gives how many images were used for GAN training, GAN testing, which includes SRM training and testing.

1) GAN training uses 16,000 images, 8000 of which come from BOSSBase and another 8000 from BOWS2;
2) GAN testing consists of 4,000 images, which are divided into 50% SRM training and 50% SRM testing.

B. Overall Architecture of FT_GAN

The overall architecture of the FT_GAN is shown below, refer to Fig. 1. The architecture is composed of a generator, a ternary embedding simulator, and a discriminator. As described, the architecture is same as [23], [24], [26], specifically [26] with the only difference being in the generator design.

The process begins by feeding the cover image to the generator to produce its corresponding probability map, which is then passed to the ternary embedding simulator along with the input stream representing the secret message to generate the modification map. The modification map is added to the cover image to obtain the Stego image. The pair is then input to the Xu_Net [10] discriminator after passing through a six SRM high pass filter, to perform classification. Finally, the loss made by the generator and discriminator is computed to update GAN's weights.

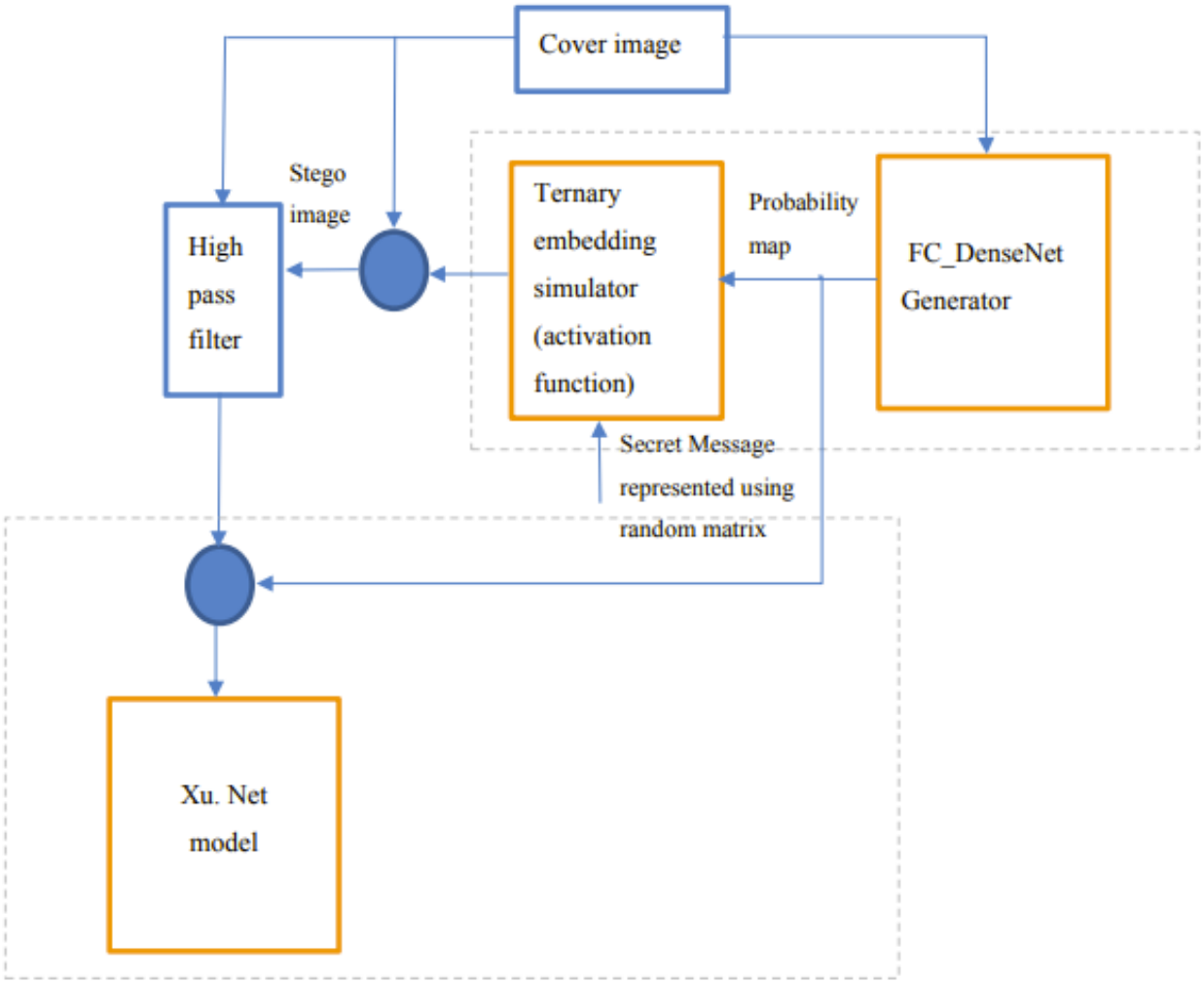


Fig. 1. The overall architecture.

1) FC_DenseNet_102 _D7-based Generator, and its loss function: In previous work, the U_Net architecture has been utilized for the generator [24], [26], this design enabled results superior to hand-crafted methods. An attempt to increase the performance even further by investigating FC_DenseNet architecture in the generator is made here. The main motivation is the comparison held by Duan et. al., [30] in 2021 when they compared the performance of FC_DenseNet [28] and U_Net [25] in CNN-based image steganography.

The FC_DenseNet_102_D7-based generator architecture is summarized in Fig. 2. This architecture attempts to maintain U_Net depth while incorporating the feature reuse capability of FC_DenseNet. A default growth rate of 16 is maintained, and the number of layers per dense block in every level is adjusted to match the width of U_Net layer in the same level. In this way, an investigation of feature reuse capability to improve performance and accelerate learning can be accomplished. To avoid NAN-loss, GAN convergence failure mode and improve the visual results quality, a bounded activation function is proposed here, and a pre-processing layer of a small kernel width is utilized.

The generator's loss function consists of two parts: the adversarial and entropy parts. The adversarial loss, which seeks to increase security performance, is the negative of the discriminator loss. The primary purpose of the entropy part is to ensure the embedding payload is met by the resultant probability maps. Refer to Eq. (2), where α and β are scaling factors and set to 1 and $\frac{1}{10^7}$, respectively. I_D is calculated using binary cross entropy, Refer to Eq. (3), Where y_i is the softmax, or discriminator output, and y_i' is truth label Stego/cover. Alternatively, I_C is computed using Eq. (4), where H, W, and Q are height, width, and target payload, respectively. Capacity is calculated with help from a generator produced probability map. Refer to Eq. (5) [4].

$$I_G = -\alpha I_D + \beta I_C^2, \tag{2}$$

$$I_D = \sum_{i=1}^{2} y_i' \log y_i, \tag{3}$$

Pre- processing layer		**Feature Map (WXHXC)**
input		256x256x1
Convolution: Kernel size=3x3 Number of kernels=1		256x256x1
Model name	**Architecture**	**Feature Map**
FC_DenseNet_102 _D7	DB (1L) + TD	128x128x17
	DB (2L) + TD	64x64x49
	DB (4L) + TD	32x32x113
	DB (8L) + TD	16x16x241
	DB (8L) + TD	8x8x369
	DB (8L) + TD	4x4x497
	DB (8L) + TD	2x2x625
	DB(8L)	2x2x753
	TU+DB(8L)	4x4x881
	TU+DB(8L)	8x8x753
	TU+DB(8L)	16x16x625
	TU+DB(8L)	32x32x497
	TU+DB(4L)	64x64x305
	TU+DB(2L)	128x128x145
	TU+DB(1L)	256x256x65
	1x1Conv	256x256x1
Post Processing layers	**Bounded-Activation-Function.** ReLU $\left(Sigmoid * 0.5 - \left(2^{-125}\right)\right) + \left(2^{-125}\right)$	256x256x1
	Output	256x256x1

Fig. 2. The architecture of FC_DenseNet_102_D7-based generator.

$$I_C = Capacity \times H \times W \times Q, \tag{4}$$

$$Capacity = \sum_{i=1}^{H}\sum_{j=1}^{W} -p_{i,j}\log_2\frac{p_{i,j}}{2} - (1-p_{i,j})\log_2(1-p_{i,j}), \tag{5}$$

2) Ternary Embedding Simulator (TES): The TES attempts to simulate the ternary embedding operation, refer to Eq. (6). The ternary embedding operation (TEO) takes as input the pixel's probability map $p_{i,j}$, and a floating-point value $n_{i,j}$ obtained from uniform distribution of (0,1), representing a secret message. The TEO output's a modification value $m_{i,j}$, which is then added to cover pixel's value x_i,j to produce Stego pixel value $y_{i,j}$.

$$m_{i,j} = \begin{cases} -1, \text{ if } n_{i,j} < \frac{p_{i,j}}{2} \\ 1, \text{ if } n_{i,j} < 1 - \frac{p_{i,j}}{2} \\ 0, \text{otherwise} \end{cases}, \tag{6}$$

The fact that the stair case function defined by Eq. (6) or the TEO are not differentiable and that they do not preserve the gradient loss during back-propagation, is the main motivation behind utilizing the double-Tanh TES proposed by Yang et al. [24] during experiment. Refer to Eq. (7), where λ is a controlling factor equal to 60 [26].

$$m_{i,j} = -0.5Tanh(\lambda(p_{i,j} - 2n_{i,j})) + 0.5Tanh(\lambda(p_{i,j} - 2(1-n_{i,j}))) \tag{7}$$

3) Discriminator and its loss functions: The Xu_Net architecture is adopted for the discriminator [10]. Xu Net comprises a preprocessing module, a convolution module, and a classification module. The preprocessing module made use of six SRM HPF [26]. The convolution module is made up of structural groupings of the convolution, activation, and pooling operations. Absolute, Batch normalization (BN), and tanh are used in the early groups to manage high pass filter output and enhance statistical characteristics. The remaining groupings utilize BN and ReLU. In the classification module, a fully connected layer and softmax activation are utilized. The discriminator weights are updated with the help of the binary cross-entropy loss function defined at Eq. (3).

C. GAN Training, and Hyper-parameters

To conserve resources and provide a fair comparison between UT_GAN and FT_GAN, a GAN training dataset is fed in batches of eight during the experiment. The Adam optimizer with a 0.001 learning rate is applied to update the generator's weight. The Discriminator optimizer is a stochastic gradient descent with a fixed momentum of 0.9, and initial learning rate of 0.001. Thus, the process was scheduled to decrease by 10% every 5,000 iterations. All generator weights were initialized with random values drawn from a normal distribution with zero mean and 0.02 standard deviation. Similarly, the Discriminator convolution kernel weights are initialized randomly from a normal distribution, but with standard deviation of 0.01. However, the fully-connected (FC) layers parameters were initialized using a "Xavier" initialization.

This previous architecture characterization and hyper-parameters were used to train the GAN for a 0.4bpp target payload. Subsequently, this model was fine-tuned for other target payloads using curriculum learning (CL).

D. Evaluation

1) Visual Evaluation: The FC_DenseNet_102_D7-based generator has been evaluated qualitatively based on the clarity and location of the generated probability map and modification map. Also, the convergence speed is an important consideration, which is the rate at which these clear, localized visual results, start to show up. Fig. 4, and 5 show visual results for the 0.4bpp Target payload, and all other Target payloads respectively.

2) Security Evaluation: The FC_DenseNet_102_D7-based generator security performance was evaluated with the help of the SRM_EC [6], [8]. GAN testing data has been split in half. The first half has been used to train SRM_EC. The second half has been used to test SRM_EC. [6], [8] Eq. (8) was used to compute the average detection error over ten trials of Ensemble classifier training and testing. The final results are shown in Table I. Here, P_{FA} is the number of false alarms processed by the by SRM_EC in cover images while P_{MD} is the number of missed detections from Stego images.

$$P_E = \frac{1}{2}(P_{FA} + P_{MD}) \tag{8}$$

V. Results

A. Visual Results

The figures below describe the main visual results. Fig. 4 compares the visual results of FT_GAN and UT_GAN during training for a 0.4bpp target payload at various epochs. Similarly, Fig. 5 compares them, but for different target payloads.

B. Security Results

Table I summarizes the average detection error made by SRM_EC for all trained payloads. The last column in the table shows the average P_E across all target payloads. Refer to Fig. 3.

VI. Discussion

A. Visual Results Discussion

The visual results discussion is conveyed in terms of a qualitative analysis for the probability map and modification map summarized in Fig. 4 and 5.

The probability map is superior if the majority of white regions, representing regions with a high likelihood of embedding, are located in the texture region of the image. Clearly, these observations, seen from the figures show that the FC_DenseNet_102_D7-based generator probability maps' more intense white compared to the U_Net-based generator probability maps' white, indicate that the texture areas (of the prior) were assigned the highest probability value, which is 0.5. Recall that the probability value range is (0,0.5).

TABLE I. AVERAGE DETECTION ERROR MADE BY SRM_EC FOR ALL TARGET PAYLOADS USING GAN TESTING DATA

Steganography	**0.1bpp**	**0.2bpp**	**0.3bpp**	**0.4bpp**	**0.5bpp**	**0.6bpp**	**0.7bpp**	**Average Among All Payloads**
UT_GAN	0.1493	**0.1550**	0.1508	**0.1536**	**0.1562**	**0.1588**	0.1362	0.1514
FT_GAN	**0.1508**	0.1470	**0.1539**	0.1498	0.1556	0.1567	**0.1532**	**0.1524**[1]

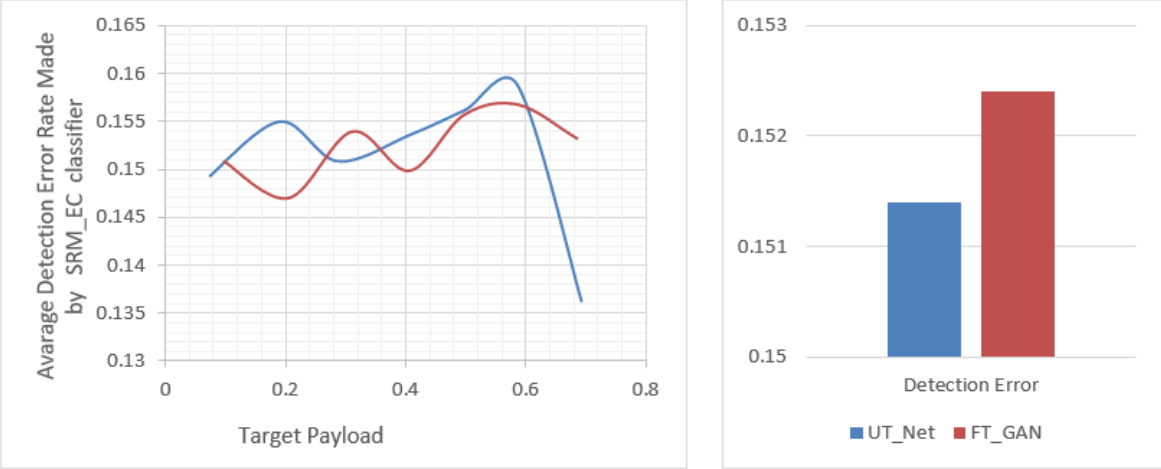


Fig. 3. Missed detection made by SRM_EC across each target payload, and the average across all of them.

From Fig. 4, we can conclude that the FC_DenseNet_102_D7-based generator began localizing texture regions faster than U_Net-based generator beginning at epoch 20. This indicates the benefit of feature reuse. The modification map is also better if it displays all black "-1" and white "+1" in the texture areas. Smooth areas represented by gray "0" remain unchangeable. Recall that the modification map is obtained using Eq. (7). On the basis of this accounting, a better modification map corresponds to a better probability map's localization.

B. Security Results and Discussion

Recall that our experimental dataset is not similar to those dataset(s) used by Yang et al.'s [26] since the ZSUBase dataset is not publicly available. Therefore, the comparison is held with the architecture proposed by Yang et al. [26] when trained using our aforementioned experiment dataset. According to the Table I, and Fig. 3, the security performance of the two architectures varies, making it difficult if not impossible for the user to determine which architecture outperforms. This is true as seen from the outcome of several variables, including the amount of training iterations and fine-tuning. However, these parameters are extremely important in GAN training due to the min-max game it plays. This game produces fluctuations in a variety of metrics, including security results. For instance, the value at one epoch may be quite high, but significantly fall in a subsequent epoch. This phenomenon leads us to understand that a precise decision criterion is needed. Thus, a CNN-based steganalysis is required to precisely determine the necessary number of fine-tuned epochs and iterations to optimize the best and most accurate final results. During these experiments, this precise decision criterion was unavailable owing to a lack of dataset size and resources. Aside from the number of fine-tuned epochs and iterations, this fluctuation was also caused by other SRM_EC factors, such as the number of base learners and the d_sub.

The experimental decision criteria was mostly visual, in addition to the loss made in meeting the target capacity, refer to Eq. (4). Therefore, the comparison is based on the last column of the table, or the "other" average classification error

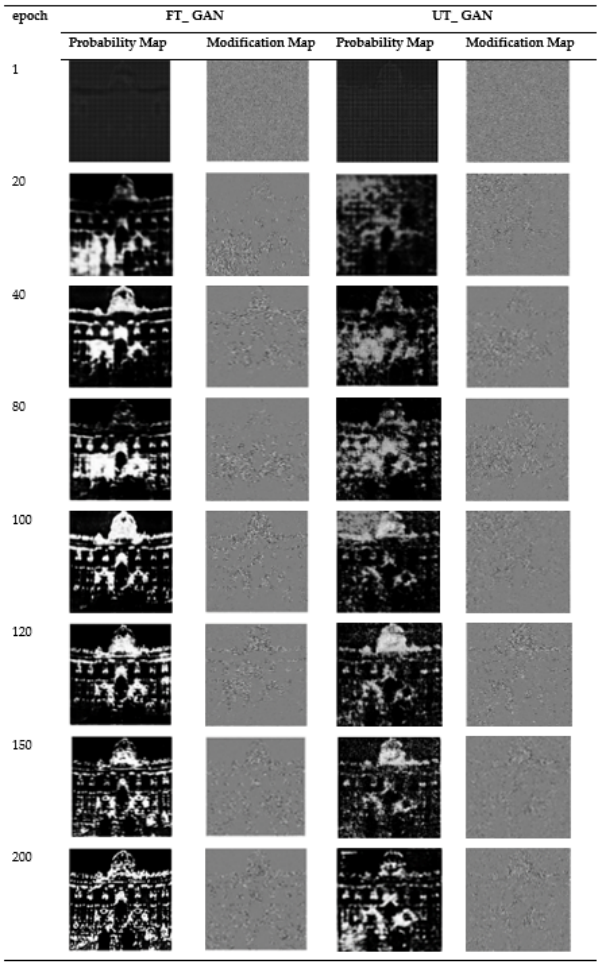


Fig. 4. Visual results for target payloads of 0.4bpp in terms of probability map and modification map.

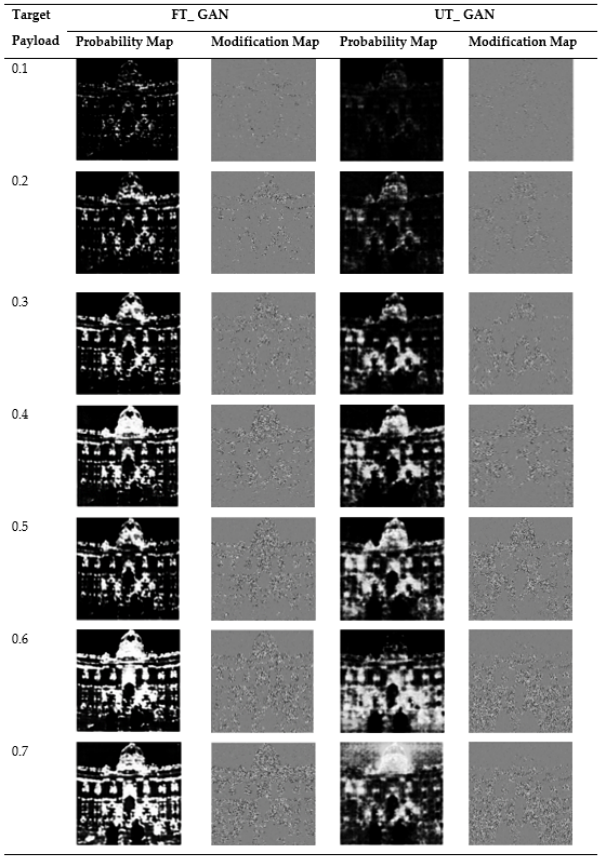


Fig. 5. Visual results for all target payloads in terms of probability map and modification map.

made across all trained payloads. Refer to Fig. 4. Overall, the FC_DenseNet_102_D7-based generator (or FT_GAN superior U_Net-based) generator improved the detection error by 0.66%. This minor average enhancement is a result of the FC_DenseNet_102_D7 architecture, and more specifically the reuse of features. This is supported by the visual result at both Fig. 4 and 5.

When comparing the results described in Eq. (1) with those of earlier studies [26], it is clear that the detection error rate reported by this experiment is lower, even for their proposed design, namely U_Net. This demonstrates conclusively that it was not the outcome of the FC_DenseNet_102_D7-based generator design, but rather the dataset employed (i.e. a limited dataset). According to Karras et al. [31], limited data is one of the challenges of GAN training. The fundamental problem with limited datasets is that the discriminator rapidly overfits to the training examples.

Consider that the discriminator's function is to classify its inputs as either cover or Stego. But, due to overfitting, it rejects as Stego all inputs other than the initial training dataset (cover images)! As a result, the generator receives minimal input to assist in enhancing the quality of its subsequent output, and thus, rendering the training process worthless. Karras et al. [31] explain why longer training did not eliminate the +1 and -1 points in the smooth regions throughout the experiment. Refer to Fig. 4, and 5. The experimental results clearly demonstrate the superiority of the proposed FT GAN over existing models in terms of both visual quality and security. For instance, the faster convergence speed observed with the FT GAN underscores the benefit of feature reuse and deep supervision provided by the FC DenseNet architecture. This capability allows the generator to focus embedding efforts on textured regions more effectively, as evidenced by the intense white areas on the probability maps (see Fig. 4). This improved localization directly enhances the stego images' imperceptibility, reducing detectability by steganalyzers. Compared to UT GAN and other GAN-based steganography models proposed by Yang et al. [24, 26], the FT GAN demonstrates a more consistent performance across all payload sizes. Specifically, the average detection error for FT GAN (0.1524) is marginally but consistently better than UT GAN (0.1514), as shown in Table I. This improvement highlights the practical advantage of incorporating FC DenseNet's feature reuse capabilities in the generator architecture, which is absent in UT GAN. While previous models relied on U-Net or similar architectures, the FT GAN's design mitigates vanishing gradient problems and achieves more robust outputs. The results also emphasize the importance of addressing training instability, a common limitation in earlier works like ASDL GAN [23] and UT 6HPF

GAN [26]. By introducing a bounded activation function, the FT GAN significantly reduces NaN-loss scenarios, leading to smoother training convergence and higher-quality outputs. In contrast, earlier models often struggled with overfitting or unstable training, particularly when using small datasets.

VII. Conclusions

We have presented our FT_GAN for content-adaptive image steganography, developed by incorporating feature reuse into the GAN generator. FT_GAN has been evaluated based on both visual and security results using SRM_EC Steganalyzers. The main outcome of this work is a clear improvement in the visual results of the FC_DenseNet_102_D7-based generator over the U_Net-based generator using BOSSBase and BOWS2 datasets, as well as an average security improvement of 0.66% compared to all other target payloads. As a future recommendation, we highly endorse the development of a universal generator that satisfies both the spatial and JPEG domains, since the one proposed here works only in the spatial domain. By leveraging FC DenseNet and a bounded activation function, our approach demonstrated improved visual quality, faster convergence, and enhanced security, with lower detection error across all payloads. Future work will focus on extending FT GAN to support the JPEG domain, exploring larger datasets for scalability, and integrating advanced learning techniques to further optimize embedding strategies and expand its practical applications in secure communication systems.

References

[1] N. Subramanian, O. Elharrouss, S. Al-Maadeed, and A. Bouridane, "Image steganography: A review of the recent advances," *IEEE access*, vol. 9, pp. 23 409–23 423, 2021.

[2] I. J. Kadhim, P. Premaratne, P. J. Vial, and B. Halloran, "Comprehensive survey of image steganography: Techniques, evaluations, and trends in future research," *Neurocomputing*, vol. 335, pp. 299–326, 2019.

[3] I. Hussain, J. Zeng, X. Qin, and S. Tan, "A survey on deep convolutional neural networks for image steganography and steganalysis," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 3, pp. 1228–1248, 2020.

[4] J. Zhao and S. Wang, "A stable gan for image steganography with multi-order feature fusion," *Neural Computing and Applications*, pp. 1–16, 2022.

[5] T. Filler, J. Judas, and J. Fridrich, "Minimizing additive distortion in steganography using syndrome-trellis codes," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 920–935, 2011.

[6] J. Fridrich and J. Kodovsky, "Rich models for steganalysis of digital images," *IEEE Transactions on information Forensics and Security*, vol. 7, no. 3, pp. 868–882, 2012.

[7] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, and J. Fridrich, "Selection-channel-aware rich model for steganalysis of digital images," in *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*. IEEE, 2014, pp. 48–53.

[8] J. Kodovsky, J. Fridrich, and V. Holub, "Ensemble classifiers for steganalysis of digital media," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 2, pp. 432–444, 2011.

[9] Y. Qian, J. Dong, W. Wang, and T. Tan, "Deep learning for steganalysis via convolutional neural networks," in *Media Watermarking, Security, and Forensics 2015*, vol. 9409. SPIE, 2015, pp. 171–180.

[10] G. Xu, H.-Z. Wu, and Y.-Q. Shi, "Structural design of convolutional neural networks for steganalysis," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 708–712, 2016.

[11] G. Xu, H.-Z. Wu, and Y. Q. Shi, "Ensemble of cnns for steganalysis: An empirical study," in *Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security*, 2016, pp. 103–107.

[12] J. Ye, J. Ni, and Y. Yi, "Deep learning hierarchical representations for image steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 11, pp. 2545–2557, 2017.

[13] M. Yedroudj, F. Comby, and M. Chaumont, "Yedroudj-net: An efficient cnn for spatial steganalysis," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2092–2096.

[14] M. Boroumand, M. Chen, and J. Fridrich, "Deep residual network for steganalysis of digital images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1181–1193, 2018.

[15] R. Zhang, F. Zhu, J. Liu, and G. Liu, "Depth-wise separable convolutions and multi-level pooling for an efficient spatial cnn-based steganalysis," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 1138–1150, 2019.

[16] T.-S. Reinel, A.-A. H. Brayan, B.-O. M. Alejandro, M.-R. Alejandro, A.-G. Daniel, A.-G. J. Alejandro, B.-J. A. Buenaventura, O.-A. Simon, I. Gustavo, and R.-P. Raul, "Gbras-net: a convolutional neural network architecture for spatial image steganalysis," *IEEE Access*, vol. 9, pp. 14 340–14 350, 2021.

[17] T. Pevnỳ, T. Filler, and P. Bas, "Using high-dimensional image models to perform highly undetectable steganography," in *International workshop on information hiding*. Springer, 2010, pp. 161–177.

[18] V. Holub and J. Fridrich, "Designing steganographic distortion using directional filters," in *2012 IEEE International workshop on information forensics and security (WIFS)*. IEEE, 2012, pp. 234–239.

[19] B. Li, M. Wang, J. Huang, and X. Li, "A new cost function for spatial image steganography," in *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 4206–4210.

[20] V. Holub, J. Fridrich, and T. Denemark, "Universal distortion function for steganography in an arbitrary domain," *EURASIP Journal on Information Security*, vol. 2014, no. 1, pp. 1–13, 2014.

[21] V. Sedighi, R. Cogranne, and J. Fridrich, "Content-adaptive steganography by minimizing statistical detectability," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 2, pp. 221–234, 2015.

[22] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets (advances in neural information processing systems)(pp. 2672–2680)," *Red Hook, NY Curran*, 2014.

[23] W. Tang, S. Tan, B. Li, and J. Huang, "Automatic steganographic distortion learning using a generative adversarial network," *IEEE Signal Processing Letters*, vol. 24, no. 10, pp. 1547–1551, 2017.

[24] J. Yang, K. Liu, X. Kang, E. K. Wong, and Y.-Q. Shi, "Spatial image steganography based on generative adversarial network," *arXiv preprint arXiv:1804.07939*, 2018.

[25] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[26] J. Yang, D. Ruan, J. Huang, X. Kang, and Y.-Q. Shi, "An embedding cost learning framework using gan," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 839–851, 2019.

[27] W. Tang, B. Li, M. Barni, J. Li, and J. Huang, "An automatic cost learning framework for image steganography using deep reinforcement learning," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 952–967, 2020.

[28] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 11–19.

[29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[30] X. Duan, N. Liu, M. Gou, W. Wang, and C. Qin, "Steganocnn: image steganography with generalization ability based on convolutional neural network," *Entropy*, vol. 22, no. 10, p. 1140, 2020.

[31] S. M. Thomas, J. G. Lefevre, G. Baxter, and N. A. Hamilton, "Towards highly expressive machine learning models of non-melanoma skin cancer," *arXiv preprint arXiv:2207.05749*, 2022.

Novel Collaborative Intrusion Detection for Enhancing Cloud Security

Widad Elbakri[1], Maheyzah Md. Siraj[2]*, Bander Ali Saleh Al-rimy[3],
Sultan Ahmed Almalki[4]*, Tami Alghamdi[5], Azan Hamad Alkhorem[6], Frederick T. Sheldon[7]
Faculty of Computing, Universiti Teknologi Malaysia, Johor Bahru, 81310, Malaysia[1,2]
School of Computing, University of Portsmouth, Portsmouth PO1 3HE, UK[3]
Computer Department-Applied College, Najran University, Najran 66462, Kingdom of Saudi Arabia[4]
Computer Science Department-Faculty of Computing and Information,
Al-Baha University, Al-Baha, 65779, Kingdom of Saudi Arabia[5]
Department of Computer Engineering-College of Computer Science and Information Technology,
Majmaah University, Al-Majmaah 11952, Kingdom of Saudi Arabia[6]
Department of Computer Science, University of Idaho, Moscow, ID 83844, USA[7]

***Abstract*—Intrusion Detection Models (IDM) often suffer from poor accuracy, especially when facing coordinated attacks such as Distributed Denial of Service (DDoS). One significant limitation of existing IDM solutions is the lack of an effective technique to determine the optimal period for sharing attack information among nodes in a distributed IDM environment. This article proposes a novel collaborative IDM model that addresses this issue by leveraging the Pruned Exact Linear Time (PELT) change point detection algorithm. The PELT algorithm dynamically determines the appropriate intervals for disseminating attack information to nodes within the collaborative IDM framework. Additionally, to enhance detection accuracy, the proposed model integrates a Gradient Boosting Machine with a Support Vector Machine (GBM-SVM) for collaborative detection of malicious activities. The proposed model was implemented in Apache Spark using the NSL-KDD benchmark intrusion detection dataset. Experimental results demonstrate that this collaborative approach significantly improves detection accuracy and responsiveness to coordinated attacks, providing a robust solution for enhancing cloud security.**

***Keywords*—Cloud security; intrusion detection; collaborative model; feature selection; anomaly detection; Pruned Exact Linear Time (PELT); gradient boosting machine; support vector machine; NSL-KDD; DDoS**

I. Introduction

A. Overview of Cloud Computing and Security Challenges

Cloud computing, a transformative paradigm in IT service delivery that emerged around 2010, provides on-demand access to resources such as computing power and storage via the internet [1]. This model enables organizations to avoid substantial investments in hardware and software, paying only for what they use. Despite its significant cost advantages and operational flexibility, cloud computing introduces new and complex security challenges [2]. Among these, Denial-of-Service (DoS) and Distributed Denial-of-Service (DDoS) attacks stand out as critical threats that can severely compromise cloud infrastructure [3]. Ensuring robust and adaptive security mechanisms is essential to fostering trust and promoting widespread adoption of cloud technologies.

*Corresponding authors.

B. Limitations of Current Intrusion Detection Models (IDM)

Intrusion Detection Models (IDM) are a cornerstone of cloud security, monitoring network events to identify and respond to potential breaches. IDM can be broadly categorized as signature-based, anomaly-based, or hybrid [4]. Signature-based Models excel at detecting known attack patterns but struggle with novel or zero-day attacks [5]. On the other hand, anomaly-based Models can identify previously unseen threats but often suffer from high false alarm rates, undermining their effectiveness [6]. Hybrid approaches aim to combine the strengths of these two methods but usually inherit their limitations, leading to suboptimal performance.

Coordinated attacks, such as DDoS attacks, exacerbate these challenges. These attacks leverage multiple compromised devices to flood targeted systems with overwhelming traffic, often evading detection by isolated IDM monitoring [7]. For example, Smurf attacks exploit spoofed IP addresses to generate a flood of Internet Control Message Protocol (ICMP) replies, overwhelming target systems [5]. These challenges highlight the need for more effective, scalable, and adaptive intrusion detection solutions.

C. Research Problem and Justification

Existing IDM solutions face critical limitations when addressing cloud environments' dynamic and distributed nature. Signature-based approaches are ineffective against zero-day attacks, while anomaly-based methods often generate excessive false positives, wasting computational resources. Furthermore, traditional IDM struggles to detect coordinated attacks, such as DDoS, due to their distributed nature and lack of collaboration among monitoring systems [8], [9], [10], [11]. These gaps necessitate developing a new, collaborative approach capable of adapting to the unique security challenges of cloud environments.

D. Proposed Solution

This paper presents a novel collaborative intrusion detection model for cloud computing environments. The proposed model addresses the limitations of existing solutions by incorporating:

1) Advanced feature selection techniques and change point detection using the Pruned Exact Linear Time (PELT) algorithm.
2) Collaborative classifier training and update mechanisms to enhance detection accuracy.
3) Distributed attack detection and IP traffic monitoring.
4) Aggregation and feedback loops are used to refine the detection process continuously.

This model leverages a network of specialised units to offer a robust, scalable, and adaptive solution to cloud security challenges. Its collaborative nature enables the effective detection of coordinated attacks, such as DDoS, while reducing false positives and resource wastage.

E. Organization of the Paper

The remainder of this paper is organized as follows:

Section II: A comprehensive review of related work, establishing the context for the proposed research.

Section III: A detailed presentation of the proposed collaborative intrusion detection model, including its architecture and functionalities.

Section IV: Description of the datasets and experimental setup used to evaluate the model.

Section V: Results and analysis, highlighting the model's effectiveness in addressing the identified challenges.

Section VI: Conclusion and future research directions, summarizing this work's contributions and potential extensions.

By addressing the gaps above, this research aims to enhance the security and resilience of cloud computing environments, fostering trust and enabling broader adoption of this transformative technology.

II. Related Works

Within the IDM context, collaboration refers to the cooperation and communication among multiple IDM nodes or agents across different sub-networks and/or hosts. These nodes share information to detect anomalies such as coordinated attacks or Distributed Denial of Service (DDoS) attacks. A collaborative IDM has the potential to detect attacks dispersed over several hosts or networks by aggregating evidence across these sub-networks. To address the issue of coordinated attacks like DDoS, existing work in cloud IDM can be categorized into signature-based, anomaly detection, and hybrid techniques.

Several researcher teams have employed signature-based techniques for collaborative cloud IDM, such as [12] and[13]. In their approach, each region in the collaborative cloud has an IDM deployed, which interacts with others by sharing alert information aimed to mitigate the impact of DDoS attacks. For instance, the framework implemented by these aforementioned researchers uses Snort-based IDM with three plug-in modules: block, communication, and cooperation. Detection agents collate and correlate alerts to assess their accuracy through a majority vote model to enhance faulty local assessments. Once an alert is accepted, a new blocking rule is added to the block table. However, this approach can only detect known attacks due to its reliance on signature-based methods. Similarly, [14] proposed a multi-threaded distributed cloud IDM for detecting DDoS attacks, comprising modules for capture and queue, analysis, and reporting. Their experimental tests in a .NET simulator demonstrated the model's ability to identify and drop bad packets. Still, it remains limited to known attack signatures and is ineffective against zero-day attacks.

Anomaly detection approaches have also been explored by researchers such as [9], whom have proposed a distributed IDM for cloud computing using a data mining approach. In their technique, network traffic is collected from edge routers and forwarded to anomaly detection devices using a Naïve Bayes classifier, with further classification by a Random Forest classifier at a central server. The author in [15] proposed a statistical and distributed network packet filtering model against DDoS attacks involving a coordinator that distributes detection tasks among various virtual machines. The author in [16] utilized Neural Networks and the Bat algorithm in their distributed IDM, while [17] employed the artificial Bee Colony algorithm and neural networks. The author in [18] developed an egress detection model using Principal Component Analysis (PCA) to protect cloud environments from DDoS attacks, with monitoring probes in each hypervisor and a hierarchical node structure for decision-making. However, the anomaly detection approach often leads to high false alarms, as [19] and [20] noted.

The hybrid approach combines anomaly detection and signature-based techniques [21], [22]. They proposed a hybrid and collaborative IDM that uses Snort for predefined attacks and a decision tree classifier with SVM for distributed attacks [23]. This research paper proposes a hybrid machine learning technique combining the Extreme Learning Machine (ELM) model with the black hole optimization algorithm for DDoS attack detection in cloud computing. However, they also inherit certain limitations, including the challenge of high false alarms and the reliance on known signatures for some detection.

In addition, existing research proposes a distributed anomaly detection system using Gaussian Mixture-based Correntropy (i.e. an adaptive neurotechnology that measures similarity, normally utilized in statistical signal processing and is based on second order moments) to identify zero-day attacks at the edge of networks [24]. While demonstrating effective performance on specific datasets, this approach lacks adaptability and collaboration, which are crucial for dynamic cloud environments. [25] employs distributed machine learning with ensemble techniques and concept drift handling for intrusion detection. While achieving high accuracy, the approach needs to have the collaborative and adaptive capabilities to strengthen a Cloud Anomaly Intrusion Detection Model (CAIDM), to strengthen their effectiveness in dynamic and distributed environments. In contrast, our proposed CAIDM addresses these limitations and demonstrates its effectiveness by incorporating adaptive and collaborative features.

To address these limitations, this paper proposes a collaborative intrusion detection model for cloud computing using the Pruned Exact Linear Time (PELT) change point detection algorithm. This approach aims to optimize the timing for exchanging attack information among nodes in the collaborative IDM and, as shown below, enhancing the model's effectiveness and reducing false alarms.

III. Proposed Enhanced Collaboration and Adaptive Cloud Anomaly Intrusion Detection Model

The Enhancing Collaboration and Adaptive Cloud Anomaly Intrusion Detection Model (EC-A-CAIDM) orchestrates a multifaceted defense against cyber threats. This cooperative model comprises seven specialized units: i) feature selection using hybrid Harmony Search Optimization and a Symmetrical Uncertainty Filter (HSO-SUF)-based feature selection for selecting the most relevant features in the dataset, ii) change point detection utilizing Pruned Exact Linear Time (PELT), iii) collaborative classifier training and update, iv) distributed attack detection, v) IP address traffic monitoring, vi) aggregation, and vii) feedback. These units operate in a synchronized three-phase choreography: training, testing, and retraining/updating.

The core of the cooperative training phase lies within the collaborative classifier training and update unit. It orchestrates a distributed training process employing the GBM-SVM classifier within the Apache Spark framework [Gradient Boosting Machine (GBM), Support Vector Machine (SVM)]. Worker nodes individually refine their Local Normal Reference Models (LNRMs) on designated data partitions [i.e. Resilient Distributed Datasets (RDDs) as identified commonly] within the distributed dataset, akin to independent learning modules. These localized models capture patterns of normalcy specific to each node's domain. The head node, acting as a central coordinator and responsible for aggregating and analyzing the LNRMs, harmonizes these LNRMs into a unified Global Normal Reference Model (GNRM), reflecting the collective picture of normalcy across the network. The testing or detection phase translates this learned normalcy into real-time vigilance. Change point detection algorithms meticulously monitor data streams for abrupt shifts, while the distributed attack detection unit leverages the LNRMs to identify local anomalies. Exceeding a predefined threshold triggers an alert, prompting the unit to forward its LNRM to the head node for further analysis. However, the material sentinel of this phase is the IP address traffic monitoring unit acting as a watchful sentinel, tracking destination IP volume. The IP volume provides valuable insight about deviations suggestive of potential DDoS attacks, ensuring a model's comprehensive coverage by our proposed model (EC-A-CAIDM).

The final phase, retraining and updating, ensures the model continuously evolves. The aggregation unit seamlessly integrates LNRMs from all detection units into the GNRM, keeping the unit in tune with the dynamic network landscape. However, the real star of this phase is the feedback unit. It plays a crucial role toward extracting insights from detected intrusions and anomalies and feeding them into the collaborative training phase. This continuous feedback loop is essential to the model's success, enhancing future detection accuracy and ensuring it maintains its' edge (i.e. dominance) in the ever-changing cybersecurity landscape, where normally, the attackers maintain an edge that utilizes intelligent maneuverings and diversion.

Fig. 1 and 2 provide an in-depth understanding of the intricate design of the EC-A-CAIDM, while Algorithm 1 presents the pseudocode of the proposed EC-A-CAIDM. Together, Fig. 1, Fig. 2 illustrate the operational flow, showcasing novel functionalities with dashed lines. A detailed exploration of each unit and its contribution to the overall security framework is presented in the subsequent sections, offering a comprehensive insight into the model's orchestration.

IV. The EC-A-CAIDM Training Phase

The training phase of the Enhancing and Adaptive Collaboration Cloud Intrusion Detection Model (EC-A-CAIDM) lays the crucial groundwork for its exceptional effectiveness and precision. Three vital units orchestrate this phase: feature selection, collaborative classifier training, and IP address traffic monitoring. Each unit plays a distinct yet pivotal role in shaping EC-A-CAIDM's capabilities, forming the foundation upon which collaborative detection thrives.

A. Feature Selection Unit

The proposed Enhanced Collaborative and Adaptive Cloud Anomaly Intrusion Detection Model (EC-A-CAIDM) incorporates a crucial feature selection unit utilizing a hybrid approach combining Harmony Search Optimization (HSO) and Symmetrical Uncertainty Filter (SUF). *Feature selection (FS)* is essential in machine learning, particularly for intrusion detection models (IDMs). FS helps improve a machine learning based models' efficiency and accuracy by removing irrelevant or noisy data attributes that impede detection capabilities [26], [27].

The HSO algorithm plays a pivotal role for the steps defined in this process. HSO, introduced by [28], is a meta-heuristic optimization technique inspired by musical improvisation. Like musicians whom collaborate to create harmony, HSO explores the search space to identify the optimal combination of features that best distinguish normal from abnormal network traffic in the cloud environment. This approach is particularly well-suited for feature selection as it efficiently finds good-to-excellent solutions within a complex search space, even though they likely differ from the absolute best (unlike some heuristic methods) [29].

These steps encompass the essence of the HS algorithm's operation, making it a compelling choice for feature selection in this study:

- Step 1: Initialization: The population of candidate solutions representing feature subsets is initialized.
- Step 2: Harmony Memory Consideration: A memory stores the best solutions, guiding the search towards favourable feature combinations.
- Step 3: Harmony Construction: New harmonies are constructed by blending existing solutions with random adjustments, fostering diversity.
- Step 4: Evaluation and Update: The fitness of each harmony is evaluated, and the memory is updated with superior solutions to preserve high-quality features.
- Step 5: Termination Criteria: The search continues until a predefined criterion is met, ensuring convergence to an optimal or near-optimal feature subset.

The HSO-SUF combination further enhances the feature selection process. While HSO efficiently explores the feature

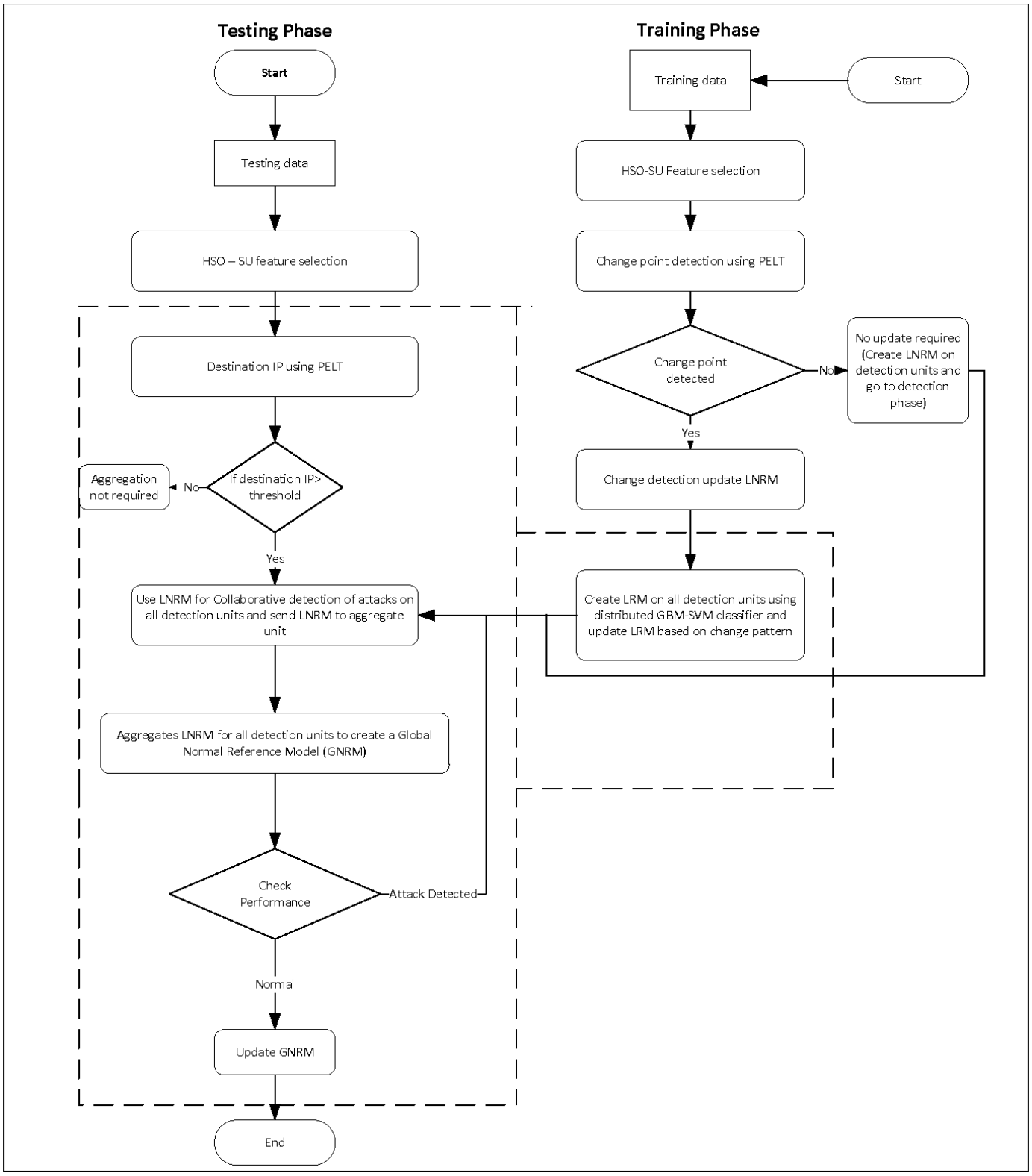


Fig. 1. The process for the adaptive and collaborative cloud intrusion detection model.

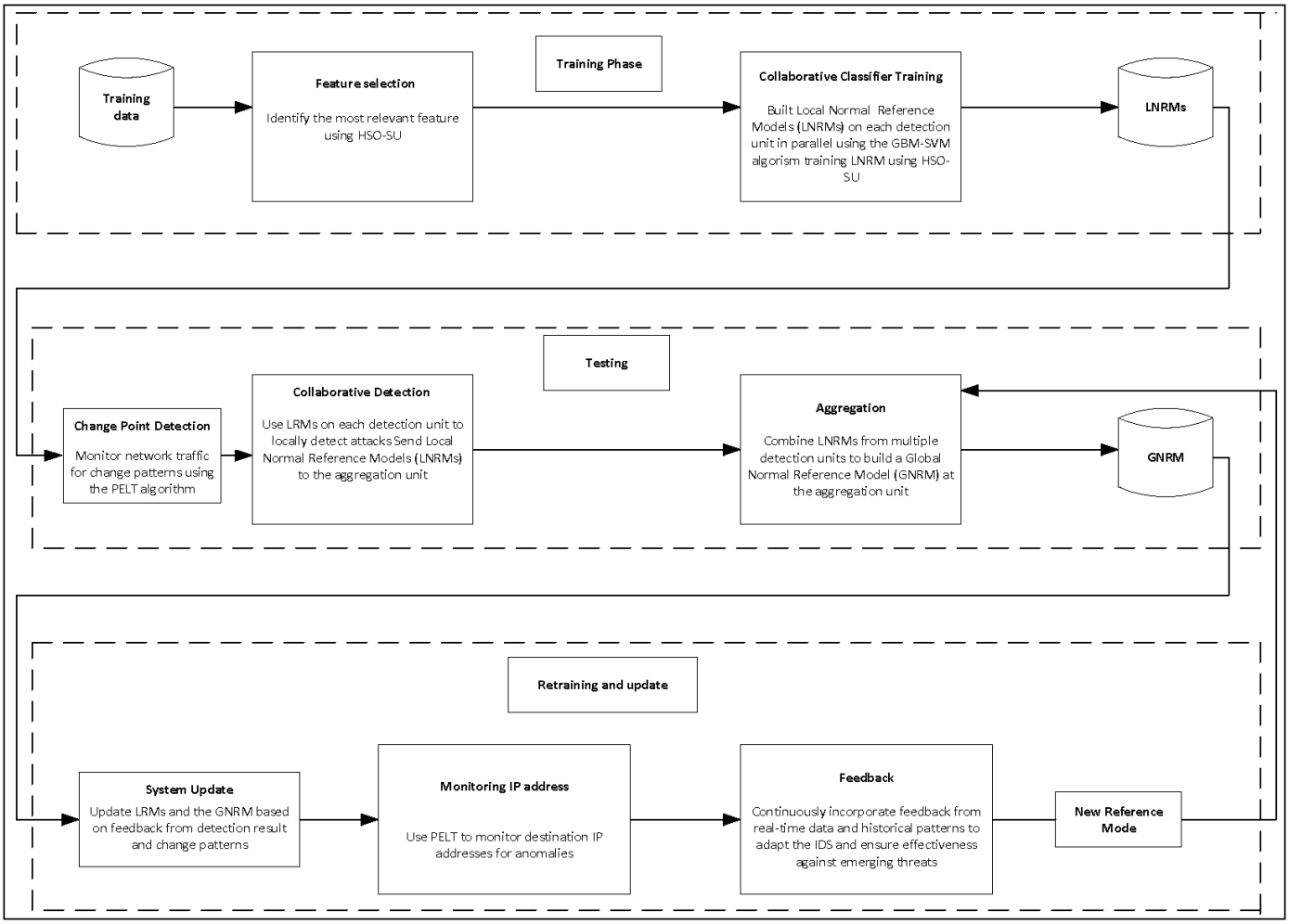


Fig. 2. Framework for an adaptive and collaborative cloud intrusion detection model with enhanced performance.

space, the SUF filter provides a mechanism to evaluate the relevance of each feature to the intrusion detection task. This two-pronged approach ensures that the selected features are diverse and demonstrably informative for anomaly detection. Our previous work [30] delves deeper into HSO for feature selection. By integrating HSO and SUF, our feature selection unit aims to optimize model performance by selecting features that significantly contribute to anomaly detection while mitigating the impact of noise and irrelevant data. This HSO-SUF combination forms a critical component of our comprehensive evaluation, demonstrating the effectiveness of HSO-SUF in addressing the unique challenges posed by cloud intrusion detection.

Therefore, the impact of Feature Selection in the process is an efficient HSO-SUF FS-filter that reduces the dataset's dimensionality, leading to a more manageable and computationally efficient intrusion detection model. This efficient approach, by focusing on relevant features, improves the model's accuracy in identifying anomalous network traffic, providing reassurance of its effectiveness.

Pheromone updating is based on the fitness function (γ') as depicted in Eq. (1). The feature subset discovered by the Harmony Search is denoted as X^i. The quality of the subset X^i and its size $|X^i|$ are measured using the evaluation metrics employed in the proposed HSO-SUF Model, namely Accuracy, Detection Rate (Precision), False Positive Rate and Sensitivity, as highlighted in Eq. (2), (3), (4) and (5). The Random Forest classifier is used to calculate metrics such as False Positives (FP), False Negatives (FN), and True Positives (TP), where FP represents the false positive rate, FN represents the false-negative rate, and TP represents the true positive rate [31].

$$\gamma' = \frac{\text{Sensitivity}\left(X^i\right) + \text{Precision}\left(X^i\right)}{|X^i|} \tag{1}$$

$$\text{Accuracy } = \frac{(TP + TN)}{(TP + TN + FP + FN)} * 100 \tag{2}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{3}$$

$$\text{False Positive Rate} = \frac{FP}{FP + TN} \tag{4}$$

Algorithm 1 Collaborative and Adaptive Cloud Anomaly Intrusion Detection Model

Require: X, Y: Network traffic training and test data
Ensure: $Sbest$: Optimal feature set, τ: Change point positions, m: Number of change points, $i\tau$: Interval between successive change points, c: Classification output normal or intrusion
(Feature Selection)
Input network traffic data X, Y.
Select optimal features $Sbest$.
(Collaborative Classifier Training and Update)
Input training data X, Y from datastore and split data among various nodes.
Initialize model parameters.
Detection units build Local Normal Reference Model (LNRM) m_i in parallel using GBM-SVM on each computing node.
Update Local Normal Reference Model using $\mu i\tau$ as the frequency of the update period.
(Change point Detection)
Detect the position of the change in data using PELT.
if change point τ is detected **then**
    Count the number of change points m.
    Determine the average interval between successive change points, $\mu i\tau$.
    For GBM-SVM, use $\mu i\tau$ as the classification model update period.
end if
(Destination-IP Traffic Monitoring)
Monitor network traffic for change patterns using the PELT algorithm.
Using the Pruned Exact Linear Time (PELT) technique, monitor the volume of destination-IP traffic from the same host.
if Destination-IP amounts of traffic from the same host ¿ Threshold **then**
    Send LNRM from detection units to the aggregation unit.
end if
(Collaborative Detection Phase)
for all detection units using X, Y samples on the LNRM to detect attacks **do**
    Send Local Normal Reference Models (LNRMs) to the aggregation unit.
    Aggregate LNRM m_i from all detection units by computing the average of the LNRM $M = \frac{1}{z}\sum_{z=1}^{n} m_i$.
    Return M.
    Classify network traffic as normal or malicious.
end for

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{5}$$

B. Collaborative Classifier Training and Update Unit

This core unit orchestrates a distributed learning ballet, collaboratively training the GBM-SVM classification algorithm on the designated dataset. This collaborative approach allows the classifier model to dynamically adapt to evolving data patterns, ensuring its ongoing relevance and effectiveness. EC-A-CAIDM leverages the Python Apache Spark framework, renowned for efficiently handling distributed data processing tasks [32], [33]. To facilitate this distributed learning, the dataset is partitioned into Resilient Distributed Datasets (RDDs), distributed across a dedicated cluster comprising a head node and six worker nodes. The head node acts as the central conductor, orchestrating the collaborative efforts of the worker nodes, designated as detection units. These nodes simultaneously engage in GBM-SVM classifier training, generating their own Local Normal Reference Model (LNRM). These LNRMs capture patterns of normalcy within each node's designated data partition, providing localized insights for subsequent detection. This collaborative approach harnesses the computational power of distributed nodes, dramatically increasing efficiency and enhancing EC-A-CAIDM's overall detection capabilities.

Implementing the Collaborative Classifier training process involves partitioning datasets into Resilient Distributed Datasets (RDDs), facilitating parallelized processing by distributing data objects across clusters. This collaborative approach, characterized by seamless coordination among distributed nodes, maximizes computational efficiency, significantly enhancing the model's capability for effective intrusion detection. Moreover, the Gradient Boosting Machine (GBM) has been effectively combined with various machine learning algorithms such as Adaline, K-means, Perceptron, and Support Vector Machine (SVM) for online training [34]. In scenarios where training samples are provided sequentially, GBM processes each data sample individually, updating the model's weights accordingly [35]. The incremental nature of GBM's parameter updates for the Reference Model is a key feature that offers several advantages, including adaptability and suitability for dynamic Models that evolve over time or in scenarios where data distribution is not static [36]. Given that adaptability to the dynamic cloud environment is a crucial requirement for cloud Intrusion Detection Models (IDm) [37], this research and development effort employs GBM to achieve an adaptive Intrusion Detection Model (IDM).

Each example z in the learning task in a supervised learning context consists of a pair of instances x, y with x an arbitrary input and y an associated output. The learning process involves considering a loss function $l(\hat{y}, y)$ that quantifies the cost of prediction $\hat{y}$ errors compared to y actual outputs. This loss function plays a crucial role in the learning process, as it guides the selection of a family F of functions $f_w(x)$ with parameters w, represented as a weight vector, and the search function $f \in F$. The objective is to find the function that minimizes the average loss across all examples, as Eq. (6) describes. The training (Reference Model) performance is evaluated using empirical risk, as shown in Eq. (7). This empirical risk $E_n(f)$ measures how well the model performs on the training data. The SVM (hinge loss) is employed as the loss function [34], as illustrated in Eq. (8). GBM builds a Reference Model for classification by minimizing empirical risk $E_n\left(f_w\right)$ after each

iteration (t) and updating it based on a single occurrence z_t using Eq. (9).

$$Q(x, w) = l\left(f_w(x), y\right) \tag{6}$$

$$E_n\left(f_w\right) = \frac{1}{n}\sum_{i=1}^{n} l\left(f_w\left(x_i\right), y_i\right) \tag{7}$$

$$l(\hat{y}, y) = \max(0, 1 - \hat{y}y) \tag{8}$$

$$w_{t+1} = w_t - \gamma_t \nabla_w Q(z_t, w_t) \tag{9}$$

C. IP Traffic Monitoring Unit

The focus on destination IP analysis stems from the observed surge in destination IPs from the same host during DDoS attacks [38], analyzing the mean values of destination IP addresses during normal periods. The meticulously designed training phase of EC-A-CAIDM lays the groundwork for its collaborative detection prowess. By harnessing the power of feature selection, distributed classifier training, and IP traffic monitoring, EC-A-CAIDM equips itself with the essential insights and models required for accurate and efficient anomaly detection in the intricate and diverse landscape of cloud computing.

In the PELT algorithm, detecting a single change point can be posed as a hypothesis test. The null hypothesis H_0 corresponds to no change point ($m = 0$), and the alternate hypothesis H_1 is a single change point ($m = 1$). A test statistic is constructed to decide whether a change has occurred. The likelihood ratio-based approach is used to test the hypothesis, which requires the calculation of maximum log-likelihood under both the null and alternate hypotheses. For the null hypothesis, the maximum log-likelihood is $\log \hat{P}(y_{1:n}|\theta)$, where $P(y_{1:n})$ is the probability density function associated with the data distribution, and $\hat{\theta}$ is the maximum likelihood estimate of the parameter [39]. The maximum likelihood under the alternate hypothesis is $\max_{T_1} \mathrm{ML}(T_1)$, where the maximum is taken over all possible change points, as shown in Eq. 10.

$$\lambda = 2\left[\max_{T_1} \mathrm{ML}(T_1) - \log P\left(y_{1:n} \mid \hat{\theta}\right)\right] \tag{10}$$

The test requires selecting a threshold C so that the null hypothesis λ is rejected if λ exceeds C. To identify multiple change points, the likelihood test statistic can be expanded to find the maximum of $\mathrm{ML}(T_{1:m})$ across all possible combinations of $T_{1:m}$ as shown in Eq. 11.

$$ML\left(T_{1:m}\right) = \sum_{i=1}^{m+1}\left[C\left(y_{(T_{i-1}+1):T_1}\right)\right] + Bf(m) \tag{11}$$

The cost function C represents a segment's cost, and $B_f(m)$ serves as a penalty to prevent overfitting. The negative log-likelihood is commonly used as the cost function, while Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) are popular choices for the penalty [39]

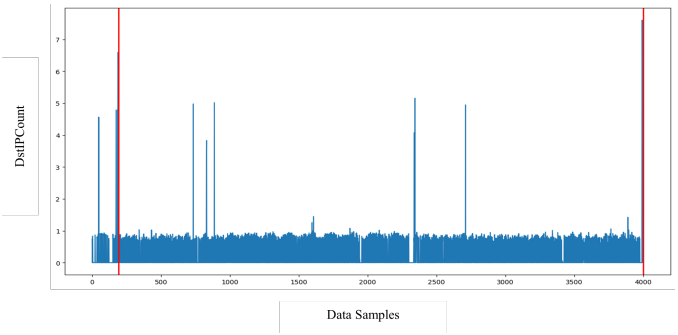


Fig. 3. Mean destination host count during normal periods in the NSL-KDD dataset.

The threshold for sending the Local Normal Reference Model (LNRM) to the aggregation unit was determined experimentally by examining the mean value of the destination-host-count (the number of network connections to the same destination host) for normal data instances using samples from the NSL-KDD dataset.

In Fig. 3, the red horizontal line indicates the mean value of the destination-host count during normal periods, which is approximately 1.0. Thus, the threshold for the normal period was set at 1.0. Thus, by integrating these advanced techniques, EC-A-CAIDM enhances its ability to detect and respond to anomalies, providing robust security for cloud computing environments.

V. The EC-A-CAIDM Testing Phase

The testing phase of EC-A-CAIDM seamlessly unites three core units i) change point detection, ii) distributed attack detection, and iii) aggregation—to create a symphony of collaborative threat identification and response. This phase rigorously evaluates the model's ability to accurately detect and address anomalies in network traffic data, with each unit playing a distinct yet complementary role in bolstering overall reliability.

A. Change Point Detection (CPD) Unit

The change point detection unit, our Model's watchful sentinel, meticulously monitors network traffic data for deviations that indicate potential intrusions or anomalies. The DAD unit vigilantly watches to identify change points within the datasets, providing crucial insights for determining an optimal frequency for updating the IDM reference model. This unit is essential to guarantee prompt and accurate threat identification by detecting subtle shifts in data patterns, enabling EC-A-CAIDM to adapt to evolving threats proactively.

B. Distributed Attack Detection (DAD) Unit

This unit acts as the heart of the EC-A-CAIDM model during the detection phase, conducting parallel intrusion detection on test data using the LNRMs generated during the distributed training phase. The DAD unit *dynamically* adjusts the transmission of LNRMs to the aggregation unit (head node)

based on a carefully determined threshold established through empirical observations of traffic patterns—specifically, changes in mean traffic volume from the same host, as detailed in the traffic monitoring unit.

Apache Spark's shared variables, known as accumulators, facilitate seamless communication between detection units (agent nodes) and the aggregation unit, enabling cohesive assessment of the intrusion detection process. Additionally, EC-A-CAIDM employs a synchronized approach to optimize computational resources and response times. The transmission of results is meticulously coordinated using the mini-batch parameter of GBM-SVM, ensuring alignment with periods of heightened destination-IP volume indicative of potential threats.

C. Aggregation Unit (AU)

In this final stage of the detection phase, the aggregation unit acts as a central conductor, synthesizing individual insights into a comprehensive global perspective. The AU collects, amalgamates, and consolidates the Local Normal Reference Models (LNRMs) contributed by participating detection nodes using Eq. (6) through (9), forming a robust Global Normal Reference Model (GNRM) as shown in Eq. (12). This collaborative approach is fundamental to the core objectives of our model. By drawing upon the collective strengths of each node's unique perspective, the AU empowers EC-A-CAIDM to effectively detect anomalies and adapt to evolving threat landscapes, ultimately enhancing its overall performance and adaptability. The detection phase of EC-A-CAIDM showcases the power of collaborative intelligence needed for effective intrusion detection. EC-A-CAIDM demonstrates its ability to recognize and address risks efficiently in a dynamic and distributed cloud environment by carefully coordinating i) change point detection, ii) distributed attack detection, and iii) aggregation. This process, as elucidated by [40], is fundamental to the core objectives of our collaborative model.

$$M = \frac{1}{k}\sum_{i=1}^{k} m_i \tag{12}$$

VI. The EC-A-CAIDM Retraining and Updating Phase

The final phase of the Enhancing Adaptive and Collaborative (EC-A-CAIDM) model, retraining and updating, constitutes the driving force of its continuous evolution and adaptation to the ever-evolving threat landscape. This phase is governed by the Feedback unit, which acts as the Model's learning engine, enabling it to continuously refine its capabilities through insights extracted from real-time and historical data.

A. Feedback Unit

In an adaptive and collaborative intrusion detection model, the role of feedback mechanisms is paramount. These mechanisms gather crucial information from detected intrusions and anomalies, providing valuable fuel for model improvement. Upon detecting an event, the feedback model meticulously captures relevant data, including the nature of the threat, its specific characteristics, and the model's response. This rich data reservoir is then analyzed and processed to extract valuable insights and patterns. These extracted insights are subsequently fed into the training phase, serving as an essential input for the adaptive refinement of intrusion detection models.

This iterative feedback loop forms the cornerstone of the model's continuous learning and improvement. By incorporating experiences into its training arsenal, EC-A-CAIDM continuously updates its detection algorithms, fine-tunes decision-making processes, and expands its knowledge base on potential threats. This continuous evolution empowers the Model to consistently improve its accuracy in identifying and mitigating diverse intrusion attempts, ultimately enhancing its resilience against the ever-shifting threat landscape in dynamic cloud environments. Now, lets consider the combining of elements in the above Eq. (12) for the feedback unit, which adjusts the model continuously based on feedback from detected anomalies. The process can be summarized as shown in Eq. (13):

$$w_{t+1} = w_t - \gamma_t \nabla_w \left(\frac{1}{n}\sum_{i=1}^{n} \max(0, 1 - f_{w_t}(x_i) y_i) \right) \tag{13}$$

Here, γ_t represents the learning rate at time t, and ∇_w denotes the gradient with respect to the weight vector w. This equation ensures that the model weights are updated continuously based on the feedback from detected anomalies, thereby improving the model's performance over time.

VII. Dataset

The NSL-KDD benchmark dataset has been selected to evaluate the proposed Enhanced Adaptive and Collaborative Cloud Intrusion Detection Model (EC-A-CAIDM) due to its widespread acceptance in the research community. Despite being derived from the KDD-Cup 99 dataset and its potential limitations in representing real-world cloud intrusions, NSL-KDD offers a diverse and labelled network traffic data set that includes both normal and malicious activities. This dataset provides a standardized evaluation platform and aligns with standard research practices, reassuring the readers about the research's validity and reliability [41] and comparability.

The NSL-KDD dataset, with its realism and diversity of attack types, provides a robust tool for assessing intrusion detection models in both traditional networks and cloud computing environments. This dataset comprises 41 features with labels, including instances from KDD-Cup 99 and introduces some new attacks into the test set. The dataset includes categories such as DoS, Probe, User to Root (U2R), and Remote to Local (R2L), with detailed class distributions in the training and test sets. These comprehensive features enable a thorough and rigorous evaluation of the EC-A-CAIDM, instilling confidence in the research's methodology and results [42].

For a detailed breakdown of the dataset's composition, refer to Table I and Table II.

TABLE I. DISTRIBUTION OF INSTANCES IN THE TRAINING AND TESTING NSL-KDD DATASET

Class	Training	Testing
Normal	67,343	9711
DoS	45,927	7460
Probe	11,656	2421
R2L	995	2885
U2R	52	67
Total	**125,973**	**22,544**

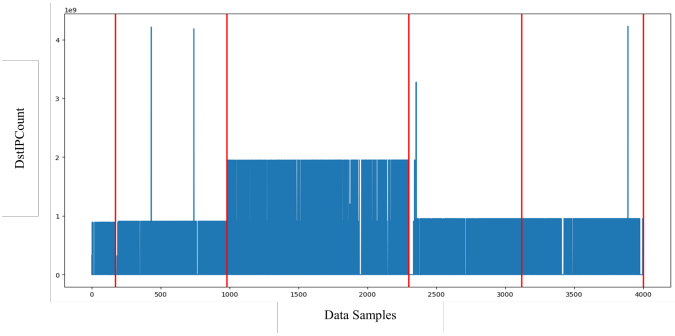


Fig. 4. Mean destination host count during Smurf DDoS attacks from NSL-KDD.

VIII. RESULTS

Evaluating the Enhanced Adaptive and Collaborative Cloud Intrusion Detection Model (EC-A-CAIDM) is a practical assessment of the proposed model's real-world application. We'll explore its performance across the following key areas:

A. Feature Selection

Examines the effectiveness of the Hybrid Feature Selection (HSO-SUF) technique, which leverages Harmony Search Optimization and a Symmetrical Uncertainty Filter, for identifying the most relevant features for intrusion detection within the cloud environment. The feature selection process identified a refined set of 13 features from the original 41 features within the NSL-KDD dataset. This selection process plays a crucial role in improving the efficiency and accuracy of the intrusion detection model. In particular, these 13 features are the key to validating our improved model. The features are Network Traffic Characteristics: Service, flag, src-bytes, and dst-bytes. Connection Details: logged-in, count. Error Rates: error-count, dst-host-serror-rate, dst-host-srv-serror-rate. Traffic Patterns: same-srv-rate, diff-srv-rate, dst-host-srv-count, dst-host-same-srv-rate.

B. IP Traffic Monitoring Unit

We evaluate the performance of the PELT change point detection algorithm employed by the IP Traffic Monitoring Unit. This analysis focuses on its ability to detect anomalies in destination IP traffic patterns. The IP Traffic Monitoring Unit, with the crucial assistance of the PELT change point detection algorithm, identifies significant changes in the mean volume of traffic directed towards specific destination IPs originating from the same source. This analysis, driven by the PELT algorithm, is key in determining when to trigger the aggregation of Local Normal Reference Models (LNRMs) from various detection units. As illustrated in Fig. 4, the destination host count surpasses the pre-defined normal threshold established during the destination-IP monitoring phase. This significant threshold violation between the 1000th and 2300th instances in the data demands immediate action. Consequently, during this specific period, the detection units will send their LNRMs to the aggregation unit for further processing.

IX. DISCUSSION

The results of the proposed Enhanced Collaborative and Adaptive Cloud Anomaly Intrusion Detection Model (EC-A-CAIDM) demonstrate significant improvements in performance metrics, as shown in Table III. The EC-A-CAIDM achieved an impressive accuracy of 100%, a detection rate of 99.99%, and a false positive rate of 0.01%. This 100% accuracy is a significant achievement, indicating that the model is highly effective in identifying and mitigating such intrusions in cloud environments. It is important to note, that this accuracy was achieved on a comprehensive and diverse dataset, ensuring that the model does not over fit a specific set of data.

Comparative analysis with existing collaborative anomaly intrusion detection models (AIDMs) highlights the superior performance of EC-A-CAIDM. For instance, the Distributed Collaborative Intrusion Detection System (D-CIDS) achieved a notable accuracy of 99.6%, a detection rate of 99.7%, and a false positive rate of 0.03%. While D-CIDS shows strong performance, EC-A-CAIDM outperforms the D-CIDS across all metrics, emphasizing its enhanced detection capabilities and lower false positive rate.

Similarly, the Hybrid Detection Classifier (HDC) proposed by [8], combining KNN and SVM, achieved an accuracy of 99.85%, a detection rate of 99.78%, and a false positive rate of 0.09%. Despite the effectiveness of HDC, EC-A-CAIDM's results are superior, indicating a more precise and reliable detection mechanism.

The Distributed Anomaly Detection (DAD) system, which utilizes Gaussian Mixture-based Correntropy, demonstrated high performance with an accuracy of 99.9%, a detection rate of 99.92%, and a false positive rate of 0.11%. Although DAD is highly effective, EC-A-CAIDM still provides better accuracy and a significantly lower false positive rate, making it a more effective choice for cloud anomaly detection.

Lastly, the Distributed Anomaly Detection using the Ensemble Hybrid (DADEH) technique achieved an accuracy of 93%, a detection rate of 99%, and a false positive rate of 0.3%. The notable performance gap between DADEH and EC-A-CAIDM further highlights the latter's advancements in accuracy and false positive reduction.

The EC-A-CAIDM's novel approach to determining the optimal timing for sharing attack information among nodes in the collaborative AIDM has contributed to its enhanced

TABLE II. NSL-KDD DATASET FEATURES

Feature number	Description	Type	Feature number	Description	Type
1	duration	Numeric	22	is_guest_login	Numeric
2	protocol_type	Symbolic	23	count	Numeric
3	service	Symbolic	24	srv_count	Numeric
4	flag	Symbolic	25	serror_count	Numeric
5	src_bytes	Numeric	26	srv_serror_rate	Numeric
6	dst_bytes	Numeric	27	rerror_rate	Numeric
7	land	Numeric	28	srv_error_rate	Numeric
8	wrong_fragment	Numeric	29	same_srv_rate	Numeric
9	urgent	Numeric	30	diff_srv_rate	Numeric
10	hot	Numeric	31	srv_diff_host_rate	Numeric
11	num_failed_login	Numeric	32	dst_host_count	Numeric
12	logged_in	Numeric	33	dst_host_srv_count	Numeric
13	num_compromised	Numeric	34	dst_host_same_srv_rate	Numeric
14	root_shell	Numeric	35	dst_host_diff_srv_rate	Numeric
15	su_attempted	Numeric	36	dst_host_same_srv_host_rate	Numeric
16	num_root	Numeric	37	dst_host_srv_diff_host_rate	Numeric
17	num_file_creation	Numeric	38	dst_host_serror_rate	Numeric
18	num_shell	Numeric	39	dst_host_srv_serror_rate	Numeric
19	num_access_file	Numeric	40	dst_host_rerror_rate	Numeric
20	num_out_of_bound_cmd	Numeric	41	dst_host_srv_rerror	Numeric
21	is_hot_login	Numeric			

performance. By effectively synchronizing attack information dissemination and employing adaptive learning techniques, the model achieves higher accuracy and detection rates and significantly reduces the occurrence of false positives. This adaptive and collaborative approach ensures that the model remains robust against evolving threats and maintains high performance in diverse cloud environments, reassuring the audience.

In conclusion, the EC-A-CAIDM sets a new benchmark and standard in cloud anomaly intrusion detection by achieving outstanding performance metrics. Compared to existing models, its superior accuracy, detection rate, and low false positive rate underscore its potential as a highly effective solution for ensuring the security of cloud-based systems. The efficiency of EC-A-CAIDM is sure to impress all stakeholders.

The data in Table III and Fig. 5, 6 and 7 unequivocally demonstrate that EC-A-CAIDM outperforms these existing models regarding accuracy, detection rate, and false positive rate. Additionally, EC-A-CAIDM has a lower detection time in seconds compared to CAIDM. The superior performance of the EC-A-CAIDM can be attributed to the effective strategy in determining the optimal timing for sharing attack information among nodes in the collaborative IDM. One of the critical challenges in collaborative IDMs is deciding when to share attack information among detection units to minimize the rate of false alarms that can result from inappropriate timing. EC-A-CAIDM successfully addresses this challenge, leading to the aforementioned enhanced performance.

X. CONCLUSION

This paper introduced the Enhanced Collaborative and Adaptive Cloud Anomaly Intrusion Detection Model (EC-A-CAIDM), a pioneering approach crafted to combat the escalating threat of sophisticated attacks in cloud environments. EC-A-CAIDM operates on a distributed architecture with seven specialized units, including feature selection, collaborative classification, IP traffic monitoring, change point detection, distributed attack detection, aggregation units, and a feedback mechanism for continuous learning. The core strength of EC-A-CAIDM lies in its strategic sharing of attack information among detection units, guided by the PELT change point detection algorithm, which effectively mitigates the challenge of false alarms prevalent in collaborative intrusion detection models. Its comprehensive evaluation using the NSL-KDD dataset demonstrates superior accuracy, detection rate, and very low false positive rate compared to existing models, instilling confidence in its effectiveness. However, the model's reliance on predefined feature sets may limit its adaptability to zero-day attacks and advanced persistent threats (APTs), and its computational complexity could pose challenges in highly dynamic cloud environments. Future research should address these limitations by exploring adaptive feature engineering techniques, lightweight architectures for real-time processing, and adversarial learning methods to enhance resilience. Advancing these areas can further improve EC-A-CAIDM's robustness and scalability, contributing to the evolving field of cloud security and intrusion detection.

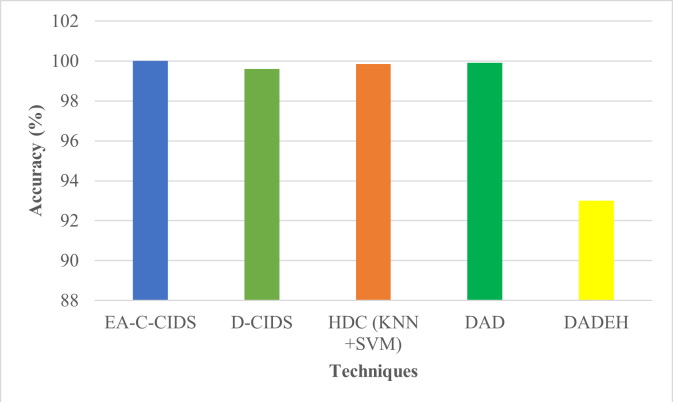


Fig. 5. Comparison of accuracy between the proposed EC-A-CAIDM vs. D-CIDM and HDC (KNN + SVM)). Example legend text.

TABLE III. COMPARISON ANALYSIS OF EC-A-CAIDM ON THE NSL-KDD DATASET WITH PREVIOUS WORK

Metrics	EA-C-CIDS	D-CIDS[43]	HDC (KNN + SVM)[8]	DAD[24]	DADEH[25]
Accuracy (%)	100	99.6	99.85	99.9	93
Detection Rate (%)	99.99	99.7	99.78	99.92	99
False Positive Rate (%)	0.01	0.03	0.09	0.11	0.3

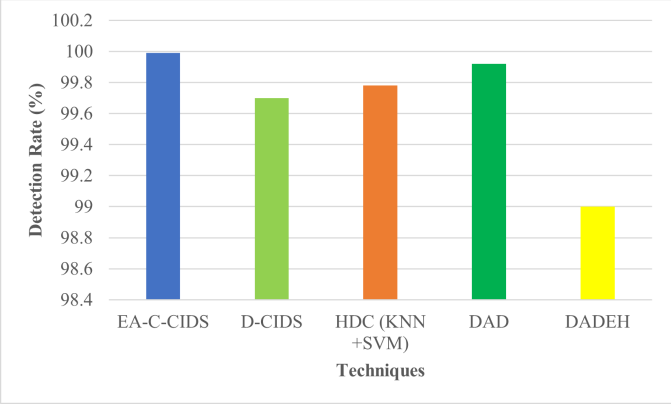


Fig. 6. Comparison of detection rate between the proposed EC-A-CAIDM vs. D-CIDM and HDC (KNN + SVM). Example legend text.

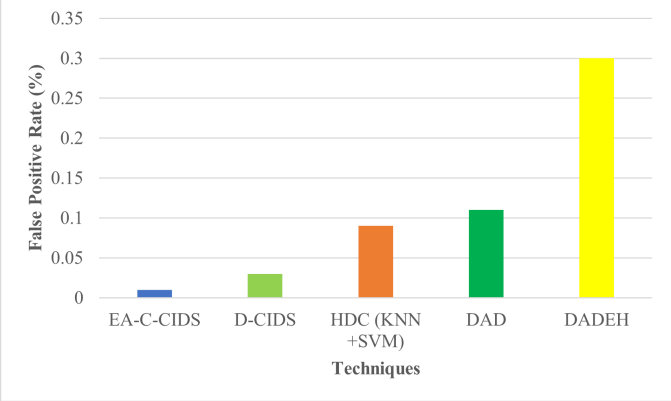


Fig. 7. Comparison of false positive rate between the proposed EC-A-CAIDM vs. D-CIDM and HDC (KNN + SVM)). Example legend text.

REFERENCES

[1] P. Mell, T. Grance *et al.*, "The nist definition of cloud computing," 2011.

[2] U. A. Butt, R. Amin, M. Mehmood, H. Aldabbas, M. T. Alharbi, and N. Albaqami, "Cloud security threats and solutions: A survey," *Wireless Personal Communications*, vol. 128, no. 1, pp. 387–413, 2023.

[3] P. Mishra, E. S. Pilli, V. Varadharajan, and U. Tupakula, "Intrusion detection techniques in cloud environment: A survey," *Journal of Network and Computer Applications*, vol. 77, pp. 18–47, 2017.

[4] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. De Alvarenga, "A survey of intrusion detection in internet of things," *Journal of Network and Computer Applications*, vol. 84, pp. 25–37, 2017.

[5] Y. Otoum and A. Nayak, "As-ids: Anomaly and signature based ids for the internet of things," *Journal of Network and Systems Management*, vol. 29, no. 3, p. 23, 2021.

[6] W. Li, W. Meng, and L. F. Kwok, "Surveying trust-based collaborative intrusion detection: state-of-the-art, challenges and future directions," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 280–305, 2021.

[7] C. V. Zhou, C. Leckie, and S. Karunasekera, "A survey of coordinated attacks and collaborative intrusion detection," *computers & security*, vol. 29, no. 1, pp. 124–140, 2010.

[8] K. Samunnisa, G. S. V. Kumar, and K. Madhavi, "Intrusion detection system in distributed cloud computing: Hybrid clustering and classification methods," *Measurement: Sensors*, vol. 25, p. 100612, 2023.

[9] M. Idhammad, K. Afdel, and M. Belouch, "Distributed intrusion detection system for cloud environments based on data mining techniques," *Procedia Computer Science*, vol. 127, pp. 35–41, 2018.

[10] O. Achbarou, M. A. El Kiram, O. Bourkoukou, and S. Elbouanani, "A new distributed intrusion detection system based on multi-agent system for cloud environment," *International Journal of Communication Networks and Information Security*, vol. 10, no. 3, p. 526, 2018.

[11] O. Osanaiye, K.-K. R. Choo, and M. Dlodlo, "Distributed denial of service (ddos) resilience in cloud: Review and conceptual cloud ddos mitigation framework," *Journal of Network and Computer Applications*, vol. 67, pp. 147–165, 2016.

[12] D. Singh, D. Patel, B. Borisaniya, and C. Modi, "Collaborative ids framework for cloud," *International Journal of Network Security*, vol. 2013, 2013.

[13] Y. Wang, W. Meng, W. Li, J. Li, W.-X. Liu, and Y. Xiang, "A fog-based privacy-preserving approach for distributed signature-based intrusion detection," *Journal of Parallel and Distributed Computing*, vol. 122, pp. 26–35, 2018.

[14] I. Gul and M. Hussain, "Distributed cloud intrusion detection model," *International Journal of Advanced Science and Technology*, vol. 34, no. 38, p. 135, 2011.

[15] V. C. Pandey, S. K. Peddoju, and P. S. Deshpande, "A statistical and distributed packet filter against ddos attacks in cloud environment," *Sādhanā*, vol. 43, pp. 1–9, 2018.

[16] S. Velliangiri and J. Premalatha, "Intrusion detection of distributed denial of service attack in cloud," *Cluster Computing*, vol. 22, no. Suppl 5, pp. 10 615–10 623, 2019.

[17] U. Ali, K. K. Dewangan, and D. K. Dewangan, "Distributed denial of service attack detection using ant bee colony and artificial neural network in cloud computing," in *Nature Inspired Computing: Proceedings of CSI 2015*. Springer, 2018, pp. 165–175.

[18] P. R. Kanna, K. Sindhanaiselvan, and M. Vijaymeena, "A defensive mechanism based on pca to defend denial of-service attack," *International Journal of Security and Its Applications*, vol. 11, no. 1, pp. 71–82, 2017.

[19] Z. Liu, B. Xu, B. Cheng, X. Hu, and M. Darbandi, "Intrusion detection systems in the cloud computing: A comprehensive and deep literature review," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 4, p. e6646, 2022.

[20] O. O. Olateju, S. U. Okon, U. T. I. Igwenagu, A. A. Salami, T. O. Oladoyinbo, and O. O. Olaniyi, "Combating the challenges of false positives in ai-driven anomaly detection systems and enhancing data security in the cloud," *Asian Journal of Research in Computer Science*, vol. 17, no. 6, pp. 264–292, 2024.

[21] H.-Y. Kwon, T. Kim, and M.-K. Lee, "Advanced intrusion detection combining signature-based and behavior-based detection methods," *Electronics*, vol. 11, no. 6, p. 867, 2022.

[22] L. K. Vashishtha, A. P. Singh, and K. Chatterjee, "Hidm: A hybrid intrusion detection model for cloud based systems," *Wireless Personal Communications*, vol. 128, no. 4, pp. 2637–2666, 2023.

[23] G. S. Kushwah and V. Ranga, "Distributed denial of service attack detection in cloud computing using hybridextreme learning machine," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 29, no. 4, pp. 1852–1870, 2021.

[24] N. Moustafa, M. Keshk, K.-K. R. Choo, T. Lynar, S. Camtepe, and M. Whitty, "Dad: A distributed anomaly detection system using ensemble one-class statistical learning in edge networks," *Future Generation Computer Systems*, vol. 118, pp. 240–251, 2021.

[25] M. Jain and G. Kaur, "Distributed anomaly detection using concept drift detection based hybrid ensemble techniques in streamed network data," *Cluster Computing*, vol. 24, no. 3, pp. 2099–2114, 2021.

[26] N. Biyyapu, E. J. Veerapaneni, P. P. Surapaneni, S. S. Vellela, and R. Vatambeti, "Designing a modified feature aggregation model with hybrid sampling techniques for network intrusion detection," *Cluster Computing*, pp. 1–19, 2024.

[27] R.-H. Dong, Y.-L. Shui, and Q.-Y. Zhang, "Intrusion detection model based on feature selection and random forest," *International Journal of Network Security*, vol. 23, no. 6, pp. 985–996, 2021.

[28] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: harmony search," *simulation*, vol. 76, no. 2, pp. 60–68, 2001.

[29] Z. W. Geem, "Optimal cost design of water distribution networks using harmony search," *Engineering optimization*, vol. 38, no. 03, pp. 259–277, 2006.

[30] W. M. Makki, M. M. Siraj, and N. M. Ibrahim, "A harmony search-based feature selection technique for cloud intrusion detection," in *Emerging Trends in Intelligent Computing and Informatics: Data Science, Intelligent Information Systems and Smart Computing 4*. Springer, 2020, pp. 779–788.

[31] M. H. Ali and M. A. Mohammed, "An improved fast learning network with harmony search based on intrusion-detection system," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 5-6, pp. 2166–2171, 2019.

[32] M. Assefi, E. Behravesh, G. Liu, and A. P. Tafti, "Big data machine learning using apache spark mllib," in *2017 ieee international conference on big data (big data)*. IEEE, 2017, pp. 3492–3498.

[33] X. Meng, J. Bradley, B. Yavuz, E. Sparks, S. Venkataraman, D. Liu, J. Freeman, D. Tsai, M. Amde, S. Owen *et al.*, "Mllib: Machine learning in apache spark," *The journal of machine learning research*, vol. 17, no. 1, pp. 1235–1241, 2016.

[34] A. Natekin and A. Knoll, "Gradient boosting machines, a tutorial," *Frontiers in neurorobotics*, vol. 7, p. 21, 2013.

[35] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[36] E. J. Atkinson, T. M. Therneau, L. J. Melton III, J. J. Camp, S. J. Achenbach, S. Amin, and S. Khosla, "Assessing fracture risk using gradient boosting machine (gbm) models," *Journal of Bone and Mineral Research*, vol. 27, no. 6, pp. 1397–1404, 2012.

[37] H. Attou, M. Mohy-eddine, A. Guezzaz, S. Benkirane, M. Azrour, A. Alabdultif, and N. Almusallam, "Towards an intelligent intrusion detection system to detect malicious activities in cloud computing," *Applied Sciences*, vol. 13, no. 17, p. 9588, 2023.

[38] T. C. Chieu, A. Mohindra, A. A. Karve, and A. Segal, "Dynamic scaling of web applications in a virtualized cloud computing environment," in *2009 IEEE International Conference on e-Business Engineering*. IEEE, 2009, pp. 281–286.

[39] R. Killick, P. Fearnhead, and I. A. Eckley, "Optimal detection of changepoints with a linear computational cost," *Journal of the American Statistical Association*, vol. 107, no. 500, pp. 1590–1598, 2012.

[40] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization." *Journal of machine learning research*, vol. 12, no. 7, 2011.

[41] S. Revathi and A. Malathi, "A detailed analysis on nsl-kdd dataset using various machine learning techniques for intrusion detection," *International Journal of Engineering Research & Technology (IJERT)*, vol. 2, no. 12, pp. 1848–1853, 2013.

[42] O. Osanaiye, H. Cai, K.-K. R. Choo, A. Dehghantanha, Z. Xu, and M. Dlodlo, "Ensemble-based multi-filter feature selection method for ddos detection in cloud computing," *EURASIP Journal on Wireless Communications and Networking*, vol. 2016, pp. 1–10, 2016.

[43] N. M. Ibrahim and A. Zainal, "A distributed intrusion detection scheme for cloud computing," *International Journal of Distributed Systems and Technologies (IJDST)*, vol. 11, no. 1, pp. 68–82, 2020.

Near-Optimal Traveling Salesman Solution with Deep Attention

Natdanai Kafakthong, Krung Sinapiromsaran
Department of Mathematics and Computer Science, Chulalongkorn University, Bangkok, Thailand, 10330

***Abstract*—The Traveling Salesman Problem (TSP) is a well-known problem in computer science that requires finding the shortest possible route that visits every city exactly once. TSP has broad applications in logistics, routing, and supply chain management, where finding optimal or near-optimal solutions efficiently can lead to substantial cost and time reductions. However, traditional solvers rely on iterative processes that can be computationally expensive and time-consuming for large-scale instances. This research proposes a novel deep learning architecture designed to predict optimal or near-optimal TSP tours directly from the problem's distance matrix, eliminating the need for extensive iterations to save total solving time. The proposed model leverages the attention mechanism to effectively focus on the most relevant parts of the network, ensuring accurate and efficient tour predictions. It has been tested on the TSPLIB benchmark dataset and observed significant improvements in both solution quality and computational speed compared to traditional solvers such as Gurobi and Genetic Algorithm. This method presents a scalable and efficient solution for large-scale TSP instances, making it a promising approach for real-world traveling salesman applications.**

***Keywords*—Traveling salesman problem; deep learning; genetic algorithm**

I. Introduction

The Traveling Salesman Problem (TSP) [1] is a classic combinatorial optimization problem that involves finding the shortest possible tour for a salesman to visit a given set of cities, exactly once. Each city is connected by a set of weighted edges that represent the distances between cities, and the objective is to minimize the total travel distance. Despite its seemingly simple formulation, TSP is known to be NP-hard, meaning the computational complexity grows exponentially with the number of cities, making it extremely difficult to solve for large instances within a reasonable time frame.

TSP has significant importance in both theoretical research and practical applications. It serves as a benchmark for optimization techniques and has a wide range of real-world uses, such as logistics, manufacturing, and routing problems. Solving TSP efficiently can result in substantial cost savings and improved resource management in industries that rely on optimized routing and scheduling. Beyond its practical implications, TSP also represents a fundamental challenge in computational theory, driving the development of new algorithms and techniques that have broader applications in other complex optimization problems.

Several methods have been developed to solve TSP [1], including exact algorithms and heuristics. Exact methods, such as brute force, branch-and-bound, and dynamic programming, attempt to find the optimal solution but they are computationally expensive and infeasible for large-scale problems due to their exponential time complexity. On the other hand, heuristic and metaheuristic approaches like genetic algorithms (GA) [5], simulated annealing, and ant colony optimization [9] offer approximate solutions by exploring the solution space more efficiently without guaranteeing an optimal result. These methods are often favored for large instances due to their ability to provide good solutions within a reasonable time frame. The effectiveness of these solvers is therefore highly dependent on the initial solution or tour. A poor starting solution can lead to long convergence times and suboptimal solutions, while a good initial solution can significantly reduce the number of iterations and improve the overall performance of the solver.

A near-optimal initial solution can improve the performance of traditional solvers by reducing the search space and accelerating convergence. Deep learning models, trained on problem data such as city distances, can predict a near-optimal tour. However, these models may not always predict a valid tour that meets the tour constraints. In such cases, the predicted solution must be refined or corrected before it is passed to a traditional solver.

This paper proposes a hybrid approach that uses deep learning to predict an initial solution for TSP and introduces an algorithm to reformulate this prediction into a valid tour. If the predicted solution from the deep learning model is not a valid tour, the tour correction algorithm adjusts it by ensuring that each city is visited exactly once and the path forms a valid loop. This refined initial solution is then fed into traditional optimization methods, such as GA or Gurobi [23], significantly reducing the time and iterations required to reach an optimal or near-optimal solution. The key advantage of this approach is that even if the predicted tour is not optimal tour, the generated tour is not differ from the optimal tour too much.

This hybrid method offers a balance between speed and accuracy. By leveraging deep learning to predict a strong initial solution and correcting it as needed, the traditional solvers can focus on fine-tuning, which reduces computational time and allows for efficient optimization of large TSP instances.

A. Contributions

- This paper introduces an approach to reduce the computational time of traditional TSP solvers by providing a near-optimal tour as the starting solution, significantly improving solver solution time.
- The proposed deep learning architecture, TSPNet, utilizes an attention mechanism in the architecture to handle combinatorial tasks in TSP.

- An algorithm is presented to convert the predicted solution or tour from TSPNet, which may not always be a valid tour, into a proper initial tour that satisfies the TSP constraints.
- The effectiveness of TSPNet, combined with traditional solvers, is evaluated on the TSPLIB dataset [25], demonstrating substantial reductions in computational time.

The remainder of this paper is organized as follows. Section II reviews related work on TSP solvers and deep learning approaches. Section III introduces the mathematical notation for the TSP. Section IV discusses traditional TSP solvers and the use of deep learning models for predicting initial tours. Section V details the methodology, including synthetic dataset generation, TSPNet architecture, input preprocessing, training, and the solving process. Section VI presents experimental results on the TSPLIB dataset and evaluates the integration of TSPNet with a Genetic Algorithm. Section VII discusses findings, limitations, and future directions, and Section VIII concludes the paper.

II. Related Work

The Traveling Salesman Problem [1] is a well-known NP-hard combinatorial optimization problem that seeks to determine the shortest possible route that visits a set of cities exactly once and returns to the origin city. Recent advancements in deep learning and reinforcement learning have provided innovative approaches to tackle this complex problem. This section will explore various methodologies, including neural networks, genetic algorithms, and hybrid models that integrate these techniques.

One significant approach to solving TSP involves the use of deep reinforcement learning (DRL). DRL formalizes TSP as a sequential decision-making problem, allowing an agent to select the next city to visit at each step. This method leverages the powerful generalization capabilities of deep neural networks, yielding impressive results in solving TSP instances [2], [3]. For instance, Bello et al. demonstrated the effectiveness of a pointer network trained via reinforcement learning, which significantly improved performance on TSP tasks [2]. Additionally, the incorporation of evolutionary algorithms with DRL has shown promise in multi-objective optimization scenarios, further enhancing the solution quality for TSP variants [3].

Another noteworthy methodology is the application of Hopfield neural networks, which have been historically significant in solving combinatorial optimization problems, including TSP. The Hopfield network utilizes an energy function to find optimal tours by minimizing the total travel distance [4]. This approach has been validated through various studies, demonstrating its capability to outperform traditional computational methods in specific instances of the TSP [4]. Moreover, the integration of chaotic neural networks has been proposed to enhance the performance of Hopfield networks, providing a novel perspective on TSP solutions [4].

Genetic algorithms (GAs) also play a crucial role in addressing the TSP [5]. These algorithms mimic the process of natural selection to iteratively improve solutions. Recent studies have highlighted the effectiveness of hybrid genetic algorithms that combine traditional GA techniques with neural networks, leading to improved convergence rates and solution quality for TSP [6], [7]. Furthermore, the application of multi-colony ant systems and particle swarm optimization has been explored as alternative heuristic methods for solving TSP, showcasing the versatility of approaches available to researchers [8], [9].

Recent studies have highlighted the effectiveness of deep reinforcement learning (DRL) in developing heuristics for TSP. For instance, Kool et al. introduced a method called Deep Policy Dynamic Programming, which integrates machine learning with dynamic programming to yield near-optimal solutions for TSPs with up to 100 nodes, demonstrating competitive performance against established solvers like LKH [10]. Similarly, Perera et al. utilized pointer networks within a multi-objective deep reinforcement learning framework, which enabled the model to generalize from smaller TSP instances to larger ones, achieving optimal solutions for TSPs with over 1000 cities [11]. These advancements illustrate the capacity of deep learning models to learn and adapt heuristics that can significantly improve solution quality and computational efficiency.

Moreover, the integration of graph neural networks (GNNs) has further enhanced the ability to solve TSP. GNNs can effectively represent the problem as a graph, allowing for the exploration of various routes and the optimization of path selection through learned embeddings. This approach has been shown to outperform traditional methods by providing a more structured understanding of the problem space [12]. The ability of deep learning models to learn from data and improve their performance over time is particularly beneficial in combinatorial optimization, where the search space is vast and complex [13].

In conclusion, the integration of deep learning, reinforcement learning, and genetic algorithms has significantly advanced the methodologies available for solving the Traveling Salesman Problem. These approaches not only enhance the efficiency of finding optimal solutions but also expand the applicability of TSP solutions to various real-world scenarios, such as logistics and route planning. The ongoing research in this field continues to evolve, promising even more sophisticated techniques for tackling this enduring challenge in combinatorial optimization.

III. Mathematical Notation for the Traveling Salesman Problem

The Traveling Salesman Problem (TSP) is an optimization problem where a set of n cities must be visited exactly once, and the goal is to find the shortest possible route that visits each city and returns to the starting point. Mathematically, the TSP can be modeled as an integer linear programming (ILP) problem. To represent this mathematically, let $G = (V, E)$ be a complete graph where V is the set of cities, $|V| = n$, and E is the set of edges. Each edge $(i, j) \in E$ is associated with a distance or cost d_{ij}. Define the binary decision variable x_{ij} as follows,

$$x_{ij} = \begin{cases} 1 & \text{if the tour visits city } i \text{ and then city } j, \\ 0 & \text{otherwise.} \end{cases}$$

The objective is to minimize the total travel cost or total tour length, which can be written as,

$$\text{minimize} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} d_{ij} x_{ij}$$

subject to the following constraints.

1) Visit each city exactly once,

$$\sum_{j=1, j\neq i}^{n} x_{ij} = 1 \quad \forall i = 1, 2, \ldots, n.$$

This ensures that each city i is left exactly once.

2) Enter each city exactly once,

$$\sum_{i=1, i\neq j}^{n} x_{ij} = 1 \quad \forall j = 1, 2, \ldots, n.$$

This ensures that each city j is entered exactly once.

3) Subtour elimination, to prevent smaller loops that don't include all cities (sub tours), the following constraint is applied for all subsets $S \subset V$, where $|S| \geq 2$,

$$\sum_{i\in S} \sum_{j\in S, j\neq i} x_{ij} \leq |S| - 1.$$

This constraint ensures that the solution forms a single tour that includes all cities.

The optimal solution to TSP is the assignment of x_{ij} values (1 or 0) that minimize the total tour length while satisfying the constraints that ensure a valid tour. However, finding this optimal tour is computationally challenging due to the combinatorial nature of the problem. As the number of cities increases, the number of possible tours grows factorially, making it impractical for an exact algorithm to solve large instances within reasonable time limits.

In the context of solving TSP, a distance matrix plays a fundamental role. The distance matrix $D = [d_{ij}]$ is an $n \times n$ matrix, where d_{ij} represents the distance or cost of traveling from city i to city j. Mathematically, it is defined as,

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & \ldots & d_{1n} \\ d_{21} & 0 & d_{23} & \ldots & d_{2n} \\ d_{31} & d_{32} & 0 & \ldots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \ldots & 0 \end{bmatrix}. \tag{1}$$

The $d_{ii} = 0$ since there is no distance associated with staying in the same city. This matrix forms the core input for solving TSP, as it provides the necessary information about the travel costs between every pair of cities.

IV. TSP Solvers

Traditional solvers for the Traveling Salesman Problem (TSP) rely heavily on the distance matrix [15], [16], [17], a fundamental representation that captures the pairwise distances between cities. This matrix plays a central role in various exact and heuristic algorithms, such as branch-and-bound [18], [19], dynamic programming [20], and cutting-plane methods [21], where it helps determine the shortest path between cities. For example, in branch-and-bound algorithms, the distance matrix is used to calculate lower bounds and prune non-optimal tours efficiently. Similarly, heuristic algorithms like the nearest neighbor or genetic algorithms use the matrix to guide their search for near-optimal solutions by iteratively considering the shortest available edges between cities.

The distance matrix provides essential information for constructing and refining the solution to TSP, making it indispensable for solving this combinatorial problem. Solvers like Gurobi [23], Concorde [24], and others leverage the distance matrix at each step of their optimization process, ensuring the best possible path is found given the constraints.

In line with these solvers, our proposed deep learning model also takes the distance matrix as an input. Instead of manually optimizing over possible routes, our model learns to predict the optimal tour from the structure of the distance matrix itself. By incorporating attention mechanisms, the model can effectively learn which city connections contribute most to form the optimal solution. The use of a distance matrix thus remains central, both in traditional algorithms and in modern machine learning approaches for solving TSP.

A. Predicting the Optimal Tour Using a Deep Learning Model

The complexity of solving the traveling salesman problem arises from the factorial growth in the number of possible tours as the number of cities increases, making it computationally challenging for exact algorithms to handle a large-scale instance. As previously discussed, deep learning has emerged as a promising approach for approximating optimal solutions of TSP, offering a more efficient alternative to traditional methods.

One innovative approach leverages deep learning models to predict optimal sub-tours or the entire tour based on historical data or synthetically generated datasets. In this framework, TSP can be formulated by using a distance matrix $D = [d_{ij}]$, which encodes the pairwise distances between cities. The deep learning model is trained to output a matrix of probabilities $P = [p_{ij}]$, where each element $p_{ij} \in [0, 1]$ represents the likelihood of a connection between city i and city j appearing in the optimal tour. This probability matrix serves as the predicted adjacency matrix for TSP, with the highest probability in each row corresponding to the city that is most likely connected to city i in the optimal tour.

The model, denoted by f_θ with parameters θ, processes the distance matrix D and outputs the adjacency matrix P,

$$f_\theta(D) = P = \begin{bmatrix} p_{11} & p_{12} & \ldots & p_{1n} \\ p_{21} & p_{22} & \ldots & p_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \ldots & p_{nn} \end{bmatrix}. \tag{2}$$

Each row of P represents a probability distribution over all cities for the next destination from city i. To obtain the predicted tour, the argmax function is applied to each row, selecting the city with the highest probability, resulting in the vector T of optimal adjacent cities,

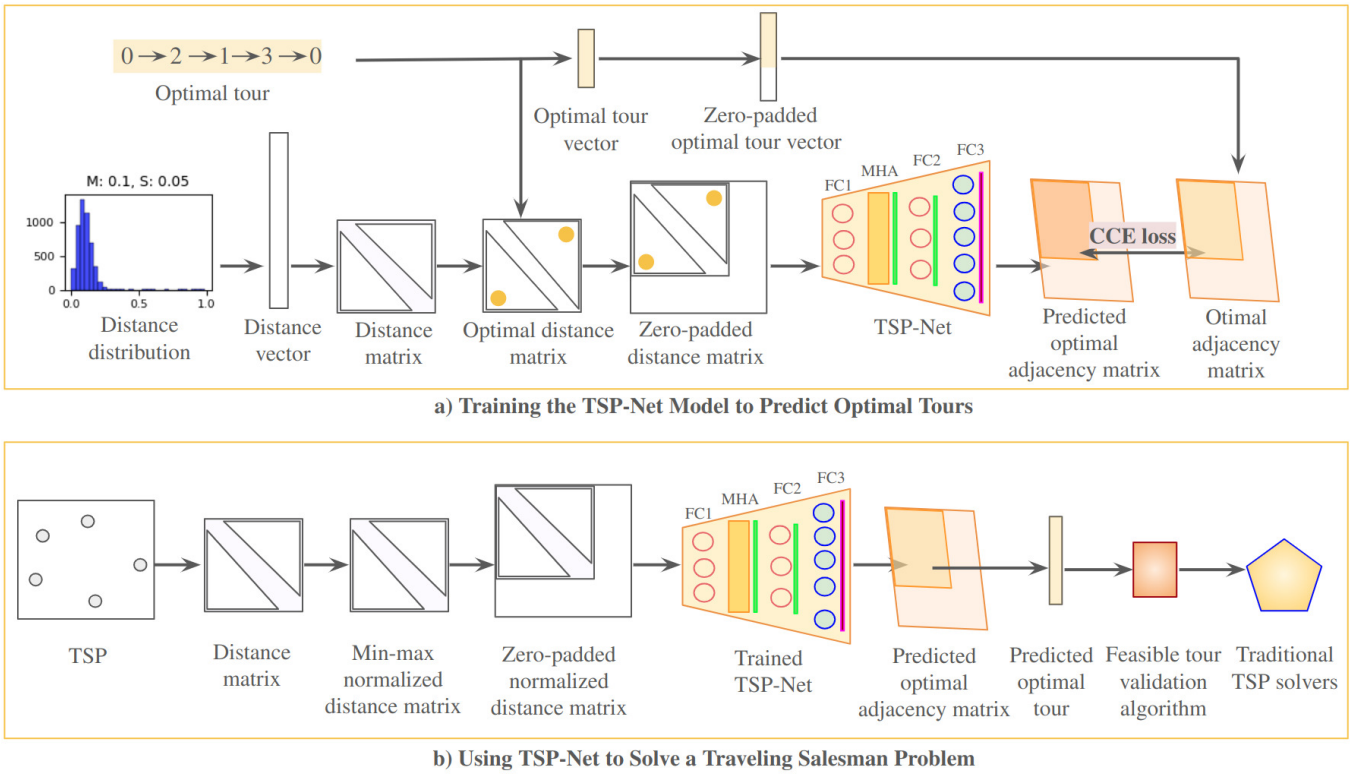


Fig. 1. Overview of the TSPNet framework: (a) Training involves generating distance matrices from a log-normal distribution and corresponding optimal tours to train the TSP model. (b) The trained model predicts a feasible tour, which is validated and can serve as an initial solution for traditional solvers.

$$T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} = \begin{bmatrix} \text{argmax}\{p_{11}, p_{12}, \ldots, p_{1n}\} \\ \text{argmax}\{p_{21}, p_{22}, \ldots, p_{2n}\} \\ \vdots \\ \text{argmax}\{p_{n1}, p_{n2}, \ldots, p_{nn}\} \end{bmatrix}. \quad (3)$$

In this representation, T contains the indices of cities selected as the next steps in the predicted tour. The predicted tour can then be reconstructed from T, forming a solution to TSP. This approach allows the deep learning model to approximate optimal solutions by learning patterns in city connections from the distance matrix efficiently, even for large-scale instances.

V. Methodology

This paper introduces TSPNet, a deep learning model designed to predict the optimal tour for the traveling salesman problem by learning the relationships between cities from distance matrices. The methodology involves two key stages: first, the training process of TSPNet, which includes generating synthetic training data, constructing input instances, and defining optimal tour labels for the model. The architecture of TSPNet is also detailed in a subsequent section. Second, TSPNet's predictions are extended to refine the solution using traditional solvers like genetic algorithms. By using TSPNet's predicted tour as an initial solution, these solvers require fewer iterations to reach an optimal or near-optimal solution. The next subsection discusses the dataset generation process for training TSPNet.

A. Synthetic TSP Training Dataset

TSP necessitates the generation of a distance matrix that accurately represents the distances between cities. In this study, a log-normal distribution is employed due to its capacity to model positively skewed data, which is often observed in real-world distance measurements.

A random variable X is said to be log-normally distributed if its natural logarithm, $Y = \ln(X)$, follows a normal distribution, that is

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

where μ is the mean and σ is the standard deviation of the underlying normal distribution. Consequently, X can be expressed as

$$X = e^Y \implies X \sim \text{Log-Normal}(\mu, \sigma^2).$$

This property ensures that the generated distances D are strictly positive, satisfying the non-negativity constraint required for distance metrics. A list of the distributions used to generate the training distances is shown in Fig. 2.

The figure presents the parameters used to generate various cases of log-normal distributions. This diversity allows the model to encounter different patterns of distances, enhancing its generalization ability and efficiency. Normally, input instances of deep learning will be normalized by min-max normalization before passing into the deep learning model. They are scaled into $[0, 1]$ then the training data of TSPNet will focus on generating training data in the range $[0, 1]$.

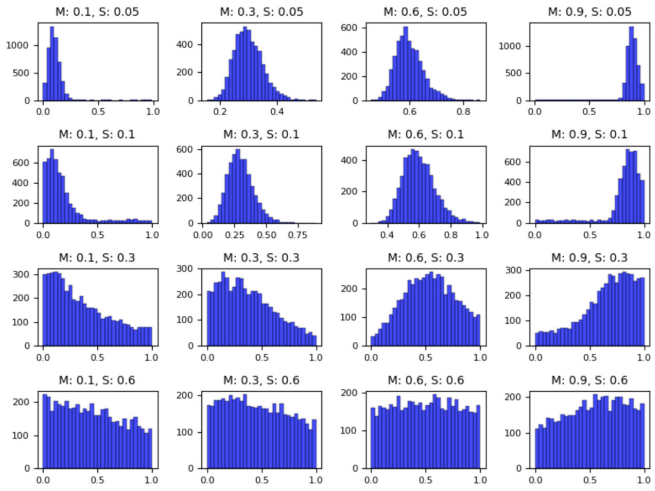


Fig. 2. Log-normal distributions for generating training distances.

For generating an input TSP instance of n cities or nodes from the log-normal distribution, the synthetic TSP datasets are executed as the following steps. The function described can be expressed mathematically as follows:

- n is the number of cities.
- μ is the mean of the generated distances.
- σ is the standard deviation of the generated distances.
- s is the scaling factor.
- λ is the shape parameter of the log-normal distribution.
- K is the number of unique tours to generate.
- $\mathcal{S}$ is the set of all cities $\{0, 1, 2, \ldots, n-1\}$.
- t_h is the h-th optimal tour T.

Step 1. Generate the distances: The generated distances are sampled from a log-normal distribution with parameters $\log(s)$ and λ,

$$d_{ij} \sim \text{LogNormal}(\log(s), \lambda), \forall i, j \in \{1, \ldots, \frac{n(n-1)}{2}\}, i < j.$$

Step 2. Adjust the mean and the standard deviation: Normalize the generated distances to have mean μ and standard deviation σ,

$$d'_{ij} = \frac{d_{ij} - \bar{d}}{\text{std}(d)} \cdot \sigma + \mu$$

where $\bar{d}$ is the mean of d_{ij} and $\text{std}(d)$ is the standard deviation. Step 3. Replace out-of-bound values: If $d'_{ij} < 0$ or $d'_{ij} > 1$, replace it with a uniformly random value within bounds,

$$d'_{ij} = \text{Uniform}(0, 1), \quad \text{if } d'_{ij} \notin [0, 1].$$

Step 4. Construct the distance matrix: The distance matrix D is symmetric and filled using the adjusted distances,

$$D_{ij} = D_{ji} = d'_{ij}, \quad \forall i < j.$$

The diagonal elements are set to zero

$$D_{ii} = 0, \quad \forall i.$$

Step 5. Create an optimal tour vector T: Defining a permutation t_h of the city indices,

$$t_h : \mathcal{S} \to \mathcal{S}, \quad \text{for } h = 1, \ldots, K.$$

Each t_h is a random shuffle of the set $\mathcal{S}$.

Step 6. The distance values in the distance matrix D_h corresponding to the cities in the optimal tour vector t_h will be reduced to ensure the distance matrix reflects the optimal tour.

Step 7. Convert the optimal tour vector t_h to be adjacency matrix using the one-hot-encoding technique for training the TSPNet model.

Thus, the distance matrix $D \in \mathbb{R}^{n \times n}$ is constructed based on the adjusted and bounded distances. By employing the log-normal distribution in this manner, the resulting distance matrix is both realistic and statistically appropriate, facilitating the effective training of the TSPNet model to predict the optimal adjacency matrix for varying configurations of cities.

B. TSPNet Input Preprocessing

To prepare the data for training the TSPNet model (as illustrated in Fig. 1(a), both the distance matrices and the corresponding optimal tour vectors are first created. As the input size for deep learning models is generally fixed, this research limits the maximum number of cities N in TSP to 200. To accommodate TSP instances with fewer than 200 cities, the distance matrices and optimal tour vectors are padded with zeros.

Let D be a distance matrix of size $n \times n$, where $n \leq N$, represents a TSP with n cities. The distance matrix D for n cities is structured as follows:

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & \ldots & d_{1n} \\ d_{21} & 0 & d_{23} & \ldots & d_{2n} \\ d_{31} & d_{32} & 0 & \ldots & d_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & d_{n3} & \ldots & 0 \end{bmatrix}. \tag{4}$$

For cases where $n < N$, D is padded to match the maximum size of $N \times N$, resulting in the following matrix,

$$D' = \begin{bmatrix} D & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{5}$$

Similarly, the optimal tour vector T of size $n \times 1$ is padded to size $N \times 1$ as follows:

$$T' = \begin{bmatrix} T \\ \mathbf{0} \end{bmatrix}. \tag{6}$$

Once the distance matrices and optimal tour vectors have been padded to the standard size of $N \times N$ and $N \times 1$, respectively, they are used as input instances for training the TSPNet model, described in the TSPNet training process subsection.

C. TSPNet Architecture

The TSPNet model is designed to predict the optimal tour for the TSP instance using the distance matrix as input. The model efficiently predicts near-optimal solutions to TSP instances by leveraging deep learning techniques. The Table I provides a detailed summary of the architecture.

TABLE I. MODEL ARCHITECTURE SUMMARY

Layer (Type)	Output Shape	Param #
Input_layer_1 (InputLayer)	(None, 200, 200)	0
Flatten_1 (Flatten)	(None, 40000)	0
FC1 (Dense)	(None, 64)	2,560,064
Reshape_1 (Reshape)	(None, 1, 64)	0
MHA (MultiHeadAttention)	(None, 1, 64)	132,672
LN1 (LayerNormalization)	(None, 1, 64)	128
FC2 (Dense)	(None, 1, 128)	8,320
Dropout_1 (Dropout)	(None, 1, 128)	0
LN2 (LayerNormalization)	(None, 1, 128)	256
FC3 (Dense)	(None, 1, 40000)	5,160,000
Reshape_2 (Reshape)	(None, 200, 200)	0
Softmax (Softmax)	(None, 200, 200)	0
Total params	**7,861,440**	

The input to the model is a 200×200 distance matrix representing the pairwise distances between cities in a TSP instance with 200 cities. After flattening the matrix, the model applies a fully connected (Dense) layer to embed the data into a lower-dimensional feature space. This is followed by a reshaping step that adjusts the dimensions for further processing with a Multi-Head Attention (MHA) layer, which captures dependencies across different city pairs. Normalization, dense layers, and dropout regularization are employed to stabilize the learning process. Finally, the output is reshaped back into a 200×200 matrix, with a softmax layer of each row of the output matrix to provide probability distributions over the possible arcs between cities to become the predicted adjacency matrix of the TSP problem.

The use of attention mechanisms in deep learning has been explored to enhance the performance of algorithms. Attention mechanisms allow models to focus on relevant parts of the input data, thereby improving the decision-making process in selecting routes [14]. This approach has been successfully applied in various combinatorial optimization tasks, showcasing the versatility and effectiveness of deep learning techniques in tackling NP-hard problems like TSP.

D. TSPNet Training Process

The TSPNet training process begins with compiling the model using the AdamW optimizer [22], which operates with a learning rate of 0.001 and 64 batch sizes, alongside the Categorical Cross-Entropy (CCE) Loss function to address the problem's multi-class classification nature. The model's output is structured as a 2D matrix, where each row corresponds to a city in the traveling salesman problem. The softmax function is applied to convert the logits into probabilities representing the likelihood of each city-to-city connection. The loss function is defined as

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{i,j} \cdot \log(p_{i,j}),$$

where N is the total number of cities (representing the rows of the output matrix), C is the total number of cities in TSP (representing the columns of the output matrix), set at 200, $y_{i,j}$ denotes the true label (either 0 or 1) for the connection between city i and city j (derived from the one-hot encoded label matrix), and $p_{i,j}$ represents the predicted probability for that connection obtained from the softmax output. This summation is performed over all rows and columns of the output matrix.

During training, the model iteratively adjusts its weights to minimize the computed loss, while the accuracy of its predictions is continuously monitored. Upon completion of the training process, the model's output probabilities are transformed back into a tour by selecting the city with the highest probability from each row of the output matrix. The average accuracy of the predicted tour is calculated by comparing it with the actual optimal tour. As illustrated in Fig. 3, the loss incurred during the training of TSPNet, which involved 100,000 training instances and 20,000 instances from a synthetic dataset, stabilizes after 7,146 iterations. The model achieves a maximum accuracy exceeding 98%, as shown in Fig. 4. Following this training phase, the model is evaluated against the TSPLib dataset, with results discussed in the subsequent experimental section.

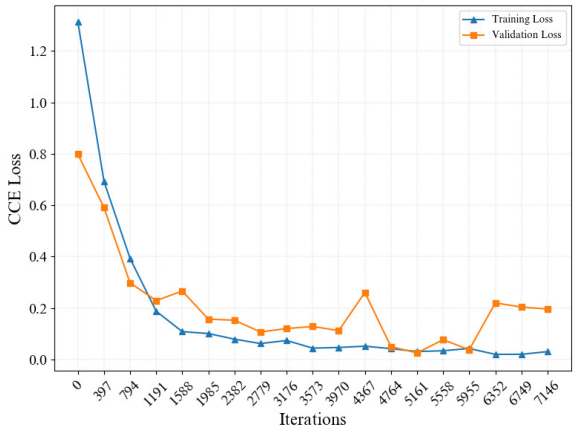


Fig. 3. Categorical cross-entropy loss values during training process.

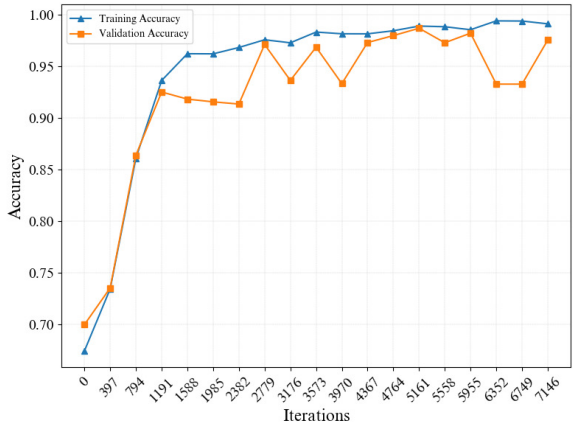


Fig. 4. Accuracy values during training process.

Although the model aims to predict the optimal tour, it

may occasionally yield invalid tours or infeasible solutions. To address this issue, the next subsection will introduce an algorithm designed to generate a valid tour when an optimal solution is not attained, before applying traditional optimization techniques, such as genetic algorithms, to search for the optimal tour.

E. TSPNet Solving Process

This section describes the process of solving the traveling salesman problem using the trained TSPNet model, as illustrated in Fig. 1(b). Given a TSP instance with n nodes, the first step is to generate an $n \times n$ distance matrix by calculating the pairwise distances between the coordinates of each city. This distance matrix is then normalized using min-max normalization before being fed into the pre-trained TSPNet model. The model outputs a zero-padded adjacency matrix, which represents the probabilities of each city being connected to another.

For a clearer understanding, let's consider Fig. 5, which shows a TSP instance with five cities. The predicted output matrix (adjacency matrix) from the TSPNet model is transformed into a predicted tour vector by selecting the city with the highest probability from each row, resulting in the tour vector T, as represented in Eq. (7),

$$\hat{T} = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 5 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} \text{argmax}\{p_{11}, p_{12}, \ldots, p_{15}\} \\ \text{argmax}\{p_{21}, p_{22}, \ldots, p_{25}\} \\ \text{argmax}\{p_{31}, p_{32}, \ldots, p_{35}\} \\ \text{argmax}\{p_{41}, p_{42}, \ldots, p_{45}\} \\ \text{argmax}\{p_{51}, p_{52}, \ldots, p_{55}\} \end{bmatrix}. \quad (7)$$

This represents an ideal prediction of the optimal tour from the city i^{th} to the j^{th} city. However, when TSPNet encounters more complex instances, the predicted tour vector T may include errors, leading to infeasible or suboptimal solutions. To mitigate this issue, the tour correction algorithm is used to adjust the predicted solution to better approximate the optimal tour while minimizing the total cost or the tour length.

The algorithm consists of four main steps: Step 1 initializes all variables. Step 2 involves separating the predicted tour into subpaths, where each subpath contains unique nodes or cities. In Step 3, missing nodes, those not included in the model's predictions are identified and added as isolated paths containing only one node. After this step, the algorithm will have several subpaths, each containing unique node. Finally, in Step 4, these subpaths are connected by finding the minimum distance between the last node of one subpath and the first node of another, thereby forming a complete and feasible tour. The details of this algorithm are outlined in Algorithm 1.

Once the tour is obtained from Algorithm 1, it can be used as the initial solution for traditional solvers, such as a genetic algorithm. This integration allows the TSPNet prediction to serve as a starting point for other optimization methods, ensuring that the process begins with a feasible tour close to the optimal tour.

The following section presents the experimental results of TSPNet on a benchmark TSP dataset:

Algorithm 1 Find a Single Feasible Tour (With Missing Nodes and Distance Comparison)

Input: `Predicted tour` ($\hat{T}$), Distance matrix (`D`), Number of cities (`n`)
Output: T (feasible tour)

1. Initialize Variables
$T \leftarrow$ An empty feasible tour
$SP \leftarrow$ A set of subpaths
$AN \leftarrow$ A set of total nodes
$N \leftarrow$ A set of existing nodes
$MN \leftarrow$ A set of missing nodes
$MD \leftarrow$ Minimum distance between nodes
$Si \leftarrow$ Start index

 # 2. Split Predicted Tour into Subpaths
for each node i in $\hat{T}$ **do**
    if node i is not in N **then**
        Add node i to N
        Update Si
    else
        if Si is set **then**
            Add the subpath from Si to i to SP
            Reset Si
        end if
    end if
end for

if Si is set **then**
    Add the remaining subpath to SP
end if

 # 3. Add Paths for Missing Nodes into Subpaths
$MN \leftarrow AN - N$
for each node m in MN **do**
    Add node m as a separate subpath to SP
end for

 # 4. Connect Split Subpaths to Form a Tour
Set T to the first subpath and remove it from SP
while Length of $T <$ `n` **do**
    Initialize MD to infinity
    for each subpath in SP **do**
        Calculate the distance from the last node of `T` to the first node of the subpath
        if this distance is smaller than MD **then**
            Update MD and store the index of the subpath
        end if
    end for
    Add the closest subpath to T
    Update the last node in T and remove the selected subpath from SP
end while
return T

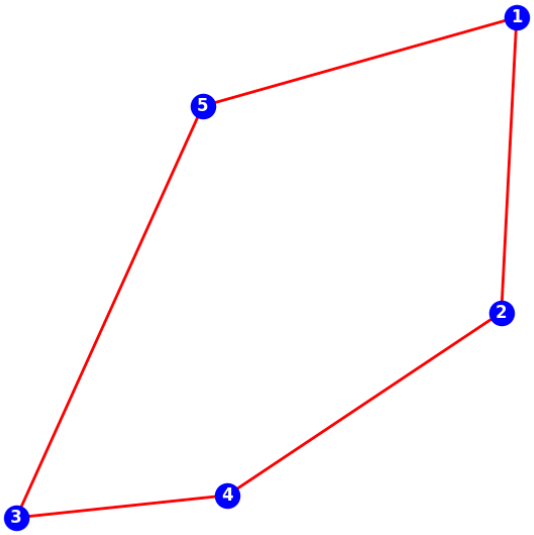


Fig. 5. The optimal tour of a five cities TSP.

VI. Experimental Results

The system architecture is based on an x86_64 platform with an AMD EPYC 7B12 CPU, featuring 8 online CPUs (4 cores per socket with 2 threads each). The deep learning model is trained using TensorFlow [26] on an L4 GPU with 22 GB of VRAM, enhancing processing efficiency. Additionally, the research employs the DEAP [27] library for implementing genetic algorithms and the Gurobi solver for optimization tasks.

A. TSPLIB Dataset

The TSPLIB (Traveling Salesman Problem Library) is a widely used collection of benchmark instances for the traveling salesman problem and its variations, offering a range of instances from small to large sizes. The dataset is formatted in plain text files with problem details such as name, type, dimension, and coordinates or distance matrices. TSPLIB encompasses various problem types, including classic TSP, asymmetric TSP (ATSP), and vehicle routing problems (VRP), allowing researchers to benchmark and compare algorithm performance. The dataset is accessible through its official website and repositories like GitHub.

B. Genetic Algorithm (GA)

The Genetic Algorithm (GA) used in this study was configured with specific parameters to optimize the performance of the TSPLIB dataset. The population size was set to 300 individuals, and the crossover probability was 0.7, allowing a significant proportion of the population to exchange genetic information during each iteration. The mutation probability was set to 0.3, introducing randomness and variation into the population to avoid local optima. A gap threshold of 0.07% was implemented, meaning the algorithm would terminate early if the gap between the best-found solution and the optimal solution was less than 0.07%. The exact optimal distance values were retrieved from the third column of the official TSPLIB web resource and set as the best-known distances for the algorithm to target. The DEAP library was used to run the algorithm, with multiprocessing enabled to parallelize the workload and reduce runtime. Additionally, a local search technique known as 2-opt was applied to refine the best individuals, improving their solutions by eliminating suboptimal edges and enhancing the overall quality of the tour.

The Table II presents a comparison of different TSP solvers, including Mixed-Integer Programming (MIP) solvers (Gurobi and a Genetic Algorithm), and heuristic solvers (TSP-Net and TSPNet combined with GA), evaluated on multiple TSP instances from the TSPLIB dataset. The performance metrics compared include the objective value, the percentage gap from the optimal solution, and the time taken to solve the problems. The gap is calculated as a percentage difference between the best objective value found by Gurobi, GA, TSP-Net, or TSPNet GA and the optimal solution from the official TSPLIB web resource. The formula used for the gap is,

$$\text{Gap} = \frac{100 \times (\text{Best objective value from solver} - \text{Optimal})}{\text{Objective value from solver}}.$$

This formula expresses the relative deviation of the obtained solution from the known optimal solution. A lower gap value indicates a closer match to the optimal solution, with a gap of 0% representing a perfect match.

Gurobi was given an upper bound of 3600 seconds, and it consistently achieves objective values close to the optimal with minimal gaps, though its computation time varies significantly depending on the problem size, ranging from around 0.1 seconds to over 600 seconds, with some instances nearly reaching the time limit. The Genetic Algorithm and TSPNet GA were both given a time limit of 600 seconds. The Genetic Algorithm shows larger gaps and longer computation times compared to Gurobi, with instances such as d198 showing a gap of over 80%. The TSPNet heuristic solver performs exceptionally well, achieving objective values close to the optimal and minimal gaps for most instances, while being significantly faster than the MIP solvers, often solving problems in fractions of a second. The combination of TSPNet with GA slightly improves the objective values and maintains competitive computation times, showing minimal gaps and similar performance to TSPNet in terms of objective values. Overall, TSPNet proves to be a highly effective and fast heuristic solver, delivering accurate results with minimal gaps and outperforming traditional MIP solvers in terms of speed. The combination of TSPNet and GA offers additional benefits by balancing high accuracy with efficient computation times, demonstrating the practicality of heuristic approaches for solving large-scale TSP problems.

VII. Discussion

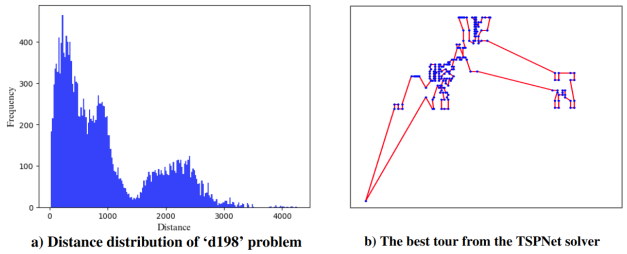
This research presents a deep learning model based on an attention mechanism for solving the Traveling Salesman Problem (TSP), contributing to the growing body of literature exploring innovative approaches to combinatorial optimization. Similar to the study by Kool et al. 2019 [14]. in “Attention, Learn to Solve Routing Problems,” which employs an attention model that utilizes model architecture inspired by the LSTM (Long Short-Term Memory) network to effectively predict optimal tours, this work also leverages attention mechanisms to enhance solution accuracy for TSP. Some experiments’

TABLE II. COMPARISON OF DIFFERENT TSP SOLVERS PERFORMANCE

No.	TSP name	Optimal	MIP solver						Heuristic solver					
			Gurobi			GA			TSPNet			TSPNet GA		
			Obj	Gap	Time	Obj	Gap	Time	Obj	Gap	Time	Obj	Gap	Time
1	att48	10628	10628	0.00	125	10628	0.00	157	10628	0	0.082	10628	0	0.41
2	ch150	6528	6555	0.41	3600	6804	4.06	600	6530	0.03	0.064	6530	0.03	0.4
3	berlin52	7542	7544	0.03	0.742	7544	0.03	41	7544.36	0.03	0.064	7544	0.03	0.43
4	d198	15780	16440	4.01	3600	91319	82.72	600	16245	2.86	0.073	16245	2.86	600
5	fri26	937	937	0.00	0.579	937	0.00	1.36	937	0	0.061	937	0	0.41
6	ftv33	1286	1286	0.00	0.6	1467	12.34	600	1286	0	0.059	1286	0	0.4
7	ftv35	1473	1473	0.00	0.817	1630	9.63	600	1473	0	0.062	1473	0	0.4
8	ftv47	1776	1776	0.00	1.027	1776	0.00	3	1776	0	0.064	1776	0	0.4
9	ftv38	1530	1530	0.00	0.885	1617	5.38	600	1530	0	0.066	1530	0	0.4
10	ftv64	1839	1839	0.00	3.815	2052	10.38	600	1839	0	0.063	1839	0	0.39
11	ftv55	1608	1608	0.00	1.634	1662	3.25	600	1608	0	0.068	1608	0	0.42
12	ftv44	1613	1613	0.00	0.714	1640	1.65	600	1613	0	0.067	1613	0	0.45
13	bier127	118282	118293	0.01	3600	119340	0.89	600	118293	0.01	0.068	118293	0.01	0.43
14	ch130	6110	6128	0.29	3600	6213	1.66	600	6110	0	0.077	6110	0	0.48
15	burma14	3323	3323	0.00	0.28	3323	0.00	0.89	3323	0	0.087	3323	0	0.4
16	brazil58	25395	25395	0.00	50	25395	0.00	4	25395	0	0.063	25395	0	0.4
17	brg180	1950	1950	0.00	8238	2010	2.99	600	1950	0	0.07	1950	0	0.43
18	bays29	2020	2020	0.00	0.817	2020	0.00	1.28	2020	0	0.063	2020	0	0.43

notable gaps include 1.76% for n = 50 and 4.53% for n = 100, showcasing its competitive performance. This research extends the exploration to more complex instances, specifically focusing on TSP127. The model developed in this study achieved a remarkable objective gap of only 0.01% for the TSP127 instance, demonstrating its capability to deliver highly accurate solutions for large problem sizes. The TSPNet solver has the potential to be more powerful if trained on a large dataset with diverse training distributions.

However, training on a wide range of problems requires significant computational resources. In this study, the synthetic training data were generated from a log-normal distribution as in Fig. 2, which presents some limitations. When data points do not conform to this distribution, the solver may fail to predict the optimal tour, as demonstrated in Fig. 6. The distribution exhibits two peaks in frequency values, which the TSPNet model has not been trained on, representing a limitation of the current model. This issue could be addressed by training the model on a wider variety of distributions in the future.



a) Distance distribution of 'd198' problem b) The best tour from the TSPNet solver

Fig. 6. The best tour predicted by the TSPNet solver for the d198 problem.

VIII. CONCLUSION

Solving a Traveling Salesman Problem (TSP) with iterative processes in a traditional solver is computationally expensive and time-consuming for a large-scale instance. However, this issue can be alleviated by determining the initial optimal or near optimal valid tour to the solver to start. This paper introduces TSPNet, a deep learning-based solver equipped with an attention mechanism for TSP, to effectively predict an optimal or a near-optimal valid tour. TSPNet trained on diverse syntactic data across various distance distributions and it evaluated its performance from the TSPLIB dataset. In cases where the predicted solution is infeasible or does not represent a valid tour, the proposed algorithm adjusts the output to ensure feasibility.

TSPNet consistently generates valid tours that can serve as effective warm starts for traditional solvers, providing a strong initial solution. When evaluated against the TSPLIB benchmark dataset, TSPNet outperformed both Gurobi and a genetic algorithm enhanced with a 2-opt local search, significantly reducing computational time by circumventing the iterative processes typically required in traditional methods. Most results achieved nearly a 0% gap, highlighting the model's effectiveness. However, training on larger problem sizes necessitates substantial computational resources.

Future work will focus on expanding the training dataset to encompass a broader range of distributions and problem instances, thereby enhancing the model's generalization capabilities. Additionally, exploring more efficient training techniques or hybrid approaches that combine deep learning with traditional optimization methods could help reduce resource requirements. Integrating TSPNet with other metaheuristic algorithms represents another promising direction to improve solution accuracy and efficiency for larger-scale TSP instances.

ACKNOWLEDGMENT

This research was supported by the Science Achievement Scholarship of Thailand and the Applied Mathematics and Computational Science Program in the Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, Thailand. The authors acknowledge

using an LLM to enhance the clarity and coherence of this manuscript. However, all scientific content and conclusions are solely the authors' responsibility.

REFERENCES

[1] Cook, William J., et al. The traveling salesman problem: a computational study. Princeton university press, 2011.

[2] I. Bello, "Neural combinatorial optimization with reinforcement learning", 2016. doi: 10.48550/arXiv.1611.09940

[3] W. Liu, R. Wang, T. Zhang, K. Li, W. Li, and H. Ishibuchi,"Hybridization of evolutionary algorithm and deep reinforcement learning for multi-objective orienteering optimization", 2022. doi: 10.48550/arXiv.2206.10464

[4] N. Xu, L. Liu, and Y. Xu, "A novel chaotic neural network with radial basis function and their application to tsp," Appl. Mech. Mater., vol. 151, pp. 532–536, 2012. doi: 10.4028

[5] A. Philip, A. Adio, and O. Kehinde, "A Genetic Algorithm for Solving Travelling Salesman Problem," Int. J. Adv. Comput. Sci. Appl., vol. 2, no. 1, 2011. doi: 10.14569/IJACSA.2011.020104

[6] R. Mahajan and G. Kaur, "Neural networks using genetic algorithms," Int. J. Comput. Appl., vol. 77, no. 14, pp. 6–11, 2013. doi: 10.5120/13549-1153

[7] N. Boyko and A. Pytel, "Aspects of the study of genetic algorithms and mechanisms for their optimization for the traveling salesman problem," International Journal of Computing, pp. 543–550, 2021. doi: 10.47839/ijc.20.4.2442

[8] P. Fu, Y. Wang, and P. Yang, "A particle swarm optimization based on evolutionary game theory for discrete combinatorial optimization," Journal of Convergence Information Technology, vol. 7, no. 21, pp. 369–376, 2012.

[9] S. Sharma, and V. Garg, "Multi colony ant system based solution to traveling salesman problem using opencl," Int. J. Comput. Appl., vol. 118, no. 23, pp. 1–3, 2015. doi: 10.5120/20882-3637

[10] W. Kool, H. Hoof, J. Gromicho, and M. Welling, "Deep policy dynamic programming for vehicle routing problems," Lect. Notes Comput. Sci., vol. 13292, pp. 190–213, 2022. doi: 10.1007/978-3-031-08011-1_14

[11] J. Perera, S. Liu, M. Mernik, M. Črepinšek, and M. Ravber, "A graph pointer network-based multi-objective deep reinforcement learning algorithm for solving the traveling salesman problem," Mathematics, vol. 11, no. 2, p. 437, 2023. doi: 10.3390/math11020437

[12] M. Prates, P. Avelar, H. Lemos, L. Lamb, and M. Vardi, "Learning to solve np-complete problems: A graph neural network for decision tsp," Proc. Conf. AAAI Artif. Intell., vol. 33, no. 01, pp. 4731–4738, 2019. doi: 10.1609/aaai.v33i01.33014731

[13] U. Gunarathna, R. Borovica-Gajić, S. Karunasekara, and E. Tanin, "Solving dynamic graph problems with multi-attention deep reinforcement learning," 2022. doi: 10.48550/arxiv.2201.04895

[14] W. Kool, H. van Hoof, and M. Welling, "Attention, learn to solve routing problems!," 2018. doi: 10.48550/arxiv.1803.08475

[15] Y. Jin, Y. Ding, X. Pan, K. He, L. Zhao, T. Qin, et al., "Pointerformer: Deep reinforced multi-pointer transformer for the traveling salesman problem," Proc. Conf. AAAI Artif. Intell., vol. 37, no. 7, pp. 8132–8140, 2023. doi: 10.1609/aaai.v37i7.25982

[16] E. Bellodi, A. Bertagnon, M. Gavanelli, and R. Zese, "Improving the efficiency of euclidean tsp solving in constraint programming by predicting effective nocrossing constraints," Lect. Notes Comput. Sci., vol. 12414, pp. 318–334, 2021. doi: 10.1007/978-3-030-77091-4_20

[17] M. Purnomo, M. Iqbal, and M. Sufa, "Solving multiple routes travelling salesman problem using modified genetic algorithm," Adv. Mat. Res., vol. 576, pp. 718–722, 2012. doi: 10.4028/www.scientific.net/amr.576.718

[18] J. Sarubbi, G. Miranda, H. Luna, and G. Mateus, "A cut-and-branch algorithm for the multicommodity traveling salesman problem," 2008 IEEE International Conference on Service Operations and Logistics, and Informatics, 2008. doi: 10.1109/SOLI.2008.4682823

[19] A. Arigliano, T. Calogiuri, G. Ghiani, and E. Guerriero, "A branch-and-bound algorithm for the time-dependent traveling salesman problem," Networks, vol. 72, no. 3, pp. 382–392, 2018. doi: 10.1002/net.21830

[20] T. Kenea, "Solving shortest route using dynamic programming problem," Indian J. Sci. Technol., vol. 15, no. 31, pp. 1527–1531, 2022. doi: 10.17485/ijst/v15i31.1342

[21] T. Vo, M. Baiou, V. Nguyen, and P. Weng, "Improving subtour elimination constraint generation in branch-and-cut algorithms for the tsp with machine learning," Lect. Notes Comput. Sci., vol. 14286, pp. 537–551, 2023. doi: 10.1007/978-3-031-44505-7_36

[22] I. Loshchilov, and F. Hutter, "Decoupled Weight Decay Regularization, "International Conference on Learning Representations (ICLR)," 2019. 1711.05101.

[23] L. L. C. Gurobi Optimization, "Gurobi Optimizer Reference Manual," 2023. [Online]. Available: https://www.gurobi.com

[24] D. Applegate, R. Bixby, V. Chvátal, and W. Cook, "Concorde: A code for solving traveling salesman problems," AT&T Labs, 2003. [Online]. Available: http://www.math.uwaterloo.ca/tsp/concorde.html

[25] G. Reinelt, "TSPLIB – A Traveling Salesman Problem Library," Institut für Mathematik, Universität Heidelberg, Heidelberg, Germany, 1991. [Online]. Available: http://comopt.ifi.uni-heidelberg.de/software/TSPLIB95/

[26] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. [Online]. Available: https://www.tensorflow.org

[27] F.-A. Fortin, F.-M. De Rainville, M.-A. Gardner, M. Parizeau, and C. Gagné, "DEAP: Evolutionary algorithms made easy," J. Mach. Learn. Res., vol. 13, pp. 2171–2175, Jul. 2012. [Online]. Available: https://github.com/DEAP/deap

Leveraging Deep Learning for Enhanced Information Security: A Comprehensive Approach to Threat Detection and Mitigation

KaiJing Wang
School of Management and Information, Chuzhou City Vocational College, Chuzhou, Anhi, China, 239000, China

***Abstract*—Forcing developments in cyberspace means protecting information resources requires enhanced and more dynamic protection models. Traditional approaches don't adequately address the numerous, sophisticated, varied, and frequently intersecting emergent security challenges, such as malware, phishing, and DDoS attacks. This paper introduces a novel hybrid deep learning framework leveraging convolutional neural networks (CNN) and recurrent neural networks (RNN) for enhanced threat detection and mitigation within a Zero Trust Architecture (ZTA). The model identifies anomalies indicative of potential security threats by analysing large network traffic datasets. To decrease false positive instances, autoencoders are integrated, significantly improving the system's ability to differentiate between normal and anomalous behaviour. Extensive experiments were conducted using a benchmark cybersecurity dataset, achieving an accuracy rate of 98.75% and a false positive rate of only 1.43%. Compared to traditional approaches, this dynamic deep learning framework is highly adaptable, requiring little oversight to respond effectively to new and evolving threats. From the study results, it can be concluded that deep learning provides a robust and scalable solution for addressing emerging cyber threats and creating a more secure and reliable information security environment. Future work will focus on extending the framework to improve its accuracy and robustness, further advancing cybersecurity capabilities. This research significantly contributes to information security, establishing a promising direction for applying machine learning to enhance cybersecurity.**

***Keywords*—Artificial intelligence; deep learning; information security; threat detection; cybersecurity; convolutional neural network; recurrent neural network; mitigation**

I. Introduction

A. Background and Context

The digital age has transformed how data is generated, shared, and stored, creating a highly interconnected world where personal and organizational information is now largely digital [1]. This unprecedented connectivity, while essential for business growth, healthcare advancement, financial transactions, and everyday communication, has simultaneously expanded the landscape of cybersecurity risks. Cyberattacks have become increasingly sophisticated, targeting sensitive data and exploiting vulnerabilities within network systems, affecting both individuals and large organizations [2]. The necessity to protect the information and ensure its confidentiality, integrity, and availability has never been claimed as heavily [3].

Patching and detection based on known standards or signatures are no longer effective against today's threats [4]. Traditional threats such as Malware, phishing, and Distributed Denial of Service (DDoS) attacks are far from simple and can be described as more complex, polymorphic, and stealthy. Such methods mainly work by/concerning the use of formulas, which are useless when the attackers devise new ways of attacking the system [5]. Therefore, a more flexible and mechanized model is needed to deal with these emergent threats [6]. In this regard, AI and machine learning, in particular, have revealed relatively positive outcomes in developing cybersecurity [7]. Real-time threat detection can be achieved since DL, a subcategory of ML, can first process large amounts of complicated information and second comprehend these data to seek patterns [8]. Finally, it can make predictions based on these patterns [9].

This work builds DL, particularly, ZTA, CNN, and RNN, to develop an effective cybersecurity system that is more intelligent and capable of handling new threats as and when they are created. ZTA is an architecture that applies the principle of "never trust, always verify", which means that user and device credentials are validated in real-time for all who seek access to resources. As a next-generation security model, ZTA does not include trust in perimeters like the older perimeter-based security models. Regarding security effectiveness, ZTA poses a significant threat to contemporary and complex threats, as the implicit trust in the 'perimeter' security models is denied as unreliable. Integrating CNNs and RNNs makes it possible to perform feature extraction of high-dimensional data, including network traffic patterns, enabling the identification of minor discrepancies that other methods usually ignore. Second, autoencoders help enhance the accuracy of statistical anomaly detection by eliminating benign activity and thus minimizing the risk of false positives. This approach will seek to give an improved and more practical option for protecting a given facility or organization from cyber threats compared to the conventional models of protection [10].

B. Research Gap and Limitations of Previous Studies

There are few studies examining the use of DL in cybersecurity, and these are the main limitations and gaps identified: Many studies use rule- or signature-based systems for detecting threats since they cannot adapt as other threats may modify patterns and escape conventional detection. Although some studies have successfully used ML and DL in cybersecurity, these applications are generally constrained by many challenges, such as scalability, high numbers of false positives, and flexibility when encountering novel threat patterns [11].

For instance, popular types of DL models utilized in cybersecurity today work only provided that they have specific

data sources to rely on or function flawlessly in very controlled circumstances. What has been realized is that when used in real-life scenarios where the data varies or is unpredictable, the models may quickly decompose. This is because only static datasets are used in training, so it fails to learn dynamic new threats in any environment. Second, the high computational demand prevents the application of DL models in real-time and often time-sensitive environments [12].

The last problem concerns itself with the high false favorable rates often gotten from existing DL models; this usually overwhelms cybersecurity professionals with several irrelevant alarms for malware detection. This reduces the efficiency of the threat detection system and pulls resources from potential threats. Therefore, training DL models that will bring high accuracy and reduce false positive instances is essential to optimize their usage in the following steps. Overcoming these limitations using a model, which has to be optimal in countering the risk of large-sized, high-dimensional data, is an area of active research [13]. This research addresses those concerns by presenting the CNN-RNN architecture enhanced by autoencoders that perform spatial and temporal threat analysis, greatly minimizing false positives, and a continuous learning system for real-time application.

C. Challenges in Threat Detection and Mitigation

The process of threat detection and mitigation in cybersecurity involves several interrelated challenges. These challenges are complex enough to understand why traditional approaches may fall short and how DL can potentially address these issues. Key challenges include:

1) Data complexity: Cybersecurity data, such as network traffic logs, user activity logs, and system alerts, is vast, complex, and high-dimensional. Effective threat detection relies on real-time processing and analyzing this data to identify potential anomalies. Due to the sheer amount and variety of data, feature extraction and model training are difficult because the model must balance spotting subtle anomalies and becoming overloaded with data.

2) Accuracy vs. False positives: High detection accuracy is essential for any cybersecurity solution; however, achieving this accuracy often comes at the cost of an increase in false positives. False positives, or instances where benign activities are flagged as threats, can overwhelm security analysts and lead to inefficient resource allocation. Reducing false positives while maintaining high accuracy remains one of the core challenges in DL-based threat detection.

3) Adaptability and real-time processing: Modern cyber threats are highly dynamic, adapting to the security measures employed by organizations. Thus, security models must also adapt in real time, minimizing the need for manual updates and recalibration. Building a model that is both adaptive and capable of real-time processing without compromising on accuracy and efficiency is essential for effective threat detection and mitigation.

4) Limited data on novel threats: Many DL models require large amounts of labeled data for training, but obtaining comprehensive datasets for novel or emerging threats can be challenging. As new types of attacks are identified, security models must be able to quickly learn from limited data and accurately detect these new threats in real time.

D. Motivation for the Study

The increasing need for a new, improved model to address the difficulties presented by current fluid systems serves as the justification for conducting this study. As such, this paper aims to develop a framework that utilizes the strengths of DL and, more specifically, CNN and RNN to detect cyber threats accurately while minimizing false positives and constantly evolving in response to new threats.

CNNs are very useful for extracting spatial features from structured data; therefore, they apply well when analyzing network traffic patterns. This capability allows the model to identify conditions signifying a security breach. In the meantime, RNNs are suitable for working with sequential data like logs and user activity over time to detect temporal correlation with behavioral trespasses. When encasing autoencoders deeper, the exact precision is again improved due to the dismissal of most benign activities, thus reducing the fPR.

The system proposed combines CNNs, RNNs, and autoencoders,, which provides a tenacious model that can be trained to recognize new forms of threats with little need for engineers' interference. Such adaptability is critical in the current world, where the threats come from the dark spaces of the internet. Thus, this work aims to show that DL will be able to revolutionize the cybersecurity domain and present a framework suitable for the contemporary need of security.

E. Novel Contributions

This study makes several unique contributions to the field of cybersecurity. These contributions are designed to address specific limitations observed in existing research and to reflect the study's aim of advancing DL-based cybersecurity solutions. The primary contributions are outlined below:

1) A Hybrid CNN-RNN threat detection model: This study introduces a unique model combining CNN and RNN layers, offering enhanced capability to process network traffic for more accurate threat detection. The CNN component effectively handles spatial data, while the RNN component analyzes temporal patterns, allowing the model to capture complex features indicative of malicious activities.

2) Integration of autoencoders for false positive reduction: To improve the precision of threat identification, the model incorporates autoencoders that filter benign activities and thereby lower the rate of false alarms. This addition enhances the model's reliability by minimizing unnecessary alerts, ensuring that security teams can focus on genuine threats.

3) Real-Time adaptability and minimal manual intervention: Our model dynamically adapts to new threat patterns, reducing the need for frequent updates and human intervention. This adaptability is critical for maintaining security in constantly changing digital environments. The model's ability to self-adjust without manual recalibration highlights its potential as a scalable and sustainable cybersecurity solution.

4) Robust performance across diverse threats: Through extensive experimentation, the framework effectively identified multiple types of cyber threats, including malware, phishing, and DDoS attacks. This versatility positions the model as a valuable tool for protecting against known and emerging attack vectors, ensuring comprehensive coverage of potential security risks.

F. Outline of the Paper

To provide a clear structure, this paper is organized as follows: Section II offers a comprehensive review of the literature on cybersecurity methods, including both traditional and DL-based approaches, providing context for the study's contributions and highlighting existing gaps in research. Section III goes into detail about the steps that were taken to create the models and the data preparation methods that were used to make sure they worked at their best. It also talks about the architecture of the CNN, RNN, and autoencoder parts. Section IV presents the results and discussion, providing insights into the model's accuracy, false positive rate, and overall performance across different threat types, with comparisons to baseline models included to demonstrate the effectiveness of the proposed framework. Finally, Section V concludes the paper by summarizing findings and suggesting potential directions for future research to enhance the adaptability and effectiveness of DL in cybersecurity.

II. Literature Review

Vaddadi et al. [14] looked at how AI and machine learning can improve cyber security in sustainable development. They focused on how AI-based cybersecurity systems can help protect digital infrastructure from new cyber risks. They claimed that the discovered AI-based systems achieved an average threat detection accuracy of 92.5%, with an average of 3.2% false positives concerning different cyber threats. They observed that raw utilization of the ML algorithms cut the response time on cyberattacks by forty percent, and they stressed that there is potential for these algorithms to enhance the effectiveness of the threat response time [15]. Further, the study confirmed that AI was always successful in preventing phishing attacks, and it has helped in sorting the risks regarding the prioritized patching of vulnerabilities, which reduced the unpatched vulnerability risks by 30%. These studies emphasized the potential of AI and ML in achieving cybersecurity goals amid SDG commitments to build technological backup and protect fundamental facilities.

Lad et al. [16] aimed to develop machine learning (ML) models in the context of cybersecurity to boost threat identification; central to their consideration was the capacity of ML to deal with emergent threats. They studied supervised learning, anomaly detection, and NLP, which allowed cybersecurity systems to address big data processing. By looking at network traffic, activity logs, and even the actions of users, they demonstrated that certain types of machine learning algorithms could be used to find existing threats and stop them from getting worse before they became significant security problems. The research demonstrated an improved means of increasing the ability to detect threats, reducing response time, and consequently enhancing cybersecurity disposition. Their work confirmed that numerous methods are successfully applied to supervised learning and anomaly detection techniques; however, they faced the greatest problem of scalability and high false positives inherent in static data sets. Such drawbacks make it difficult for the model to be fully applied in real-time and be reliable against threats that may be dynamic. Our work addresses these issues by using CNNs and RNNs for dynamic threat modeling and employing continuous learning to make our model responsive to new threats.

Ofoegbu et al. [17] explored the use of ML and big data analytics for real-time cyber threat detection, paying attention to the increasing shortcomings of orthodox cybersecurity measures due to the unprecedented advancement in the use of technology and the number of connected devices [18]. To be precise, in their study, they demonstrated that the applications of ML, reinforced by big data, help cybersecurity systems learn an enormous amount of data produced in the networks and then recognize the somewhat abnormal. This approach solved several major and already actual problems in the modern cybersecurity sphere: the increased complexity of modern threats, the need for lockdown approach scalability, and, finally, the problem of false positives. Their examples from diverse industries illustrated the real advantages of using ML and big data analytics in threat detection, proving that this method significantly strengthens cybersecurity measures. They also identified real-time, ML-based threat detection and big data as a competitive advantage for organizations that need to protect their valuable assets, especially when time is essential to maintaining business continuity and clients' trust in the world of interconnected systems.

Gudala et al. [19] discussed the application of AI and ML in ZTA strengthening for advanced cyber threats. APTs and zero-day threats described the main weaknesses of conventional security models, turning to the ZTA concept of "Never Trust; Always Verify". Their study was mainly based on employing ML in real-time for OD and other flexible threat countermeasures in ZTA. Most of the opportunities were based on actual historical data, and traditional ML algorithms were initially applied for tasks such as user behavior analytics (UBA) and network traffic analysis, allowing for the spotting of signs of unauthorized access, malware presence, and data exfiltration. While traditional behaviors were well developed, new, AI-driven behaviors like mitigation by pre-defined AIR playbooks enabled quick actions such as account lockouts and device isolation. Some further research areas in AI for ZTA proposed for ZTA were federated learning for joint threat intelligence sharing and reinforcement learning for flowing threat defense and impedance management.

Ijiga et al. [20] analyzed AI and AB-ML paradigms for enhanced cybersecurity, primarily in risk analysis and fraud prevention. They suggested an approach based on AI to estimate cybersecurity threats and control frauds with better accuracy and much faster. They looked into the idea of adversarial ML in terms of how it could be used to make models safer and create defenses resistant to interference from adversaries. They proposed an adaptive risk assessment framework that employs extensive data analysis and machine learning for threat recognition and allocation. They also discussed how AI algorithms identify fraudulent transactions by defining the patterns and indicators feature in big data sets, which was well

illustrated through the uptake of AI in sectors like financial and identity activities. Their work provides an understanding of the potential of expressive artificial intelligence and adversarial machine learning to enhance security. It recommends that organizations incorporate AI approaches to guard the assets in the growingly complex threat environment.

As the amount of data and infrastructure at risk grows, Balantrapu [21] looks at new patterns in how modern machine learning methods are used to find threats in IT systems and how they might change. They examined the efficiency of many branches of ML, such as supervised, unsupervised, and reinforcement ones, while considering the possibilities to prevent and detect cyber threats in various domains like networks, endpoints, and applications. They shared the opportunities for development in feature extraction, anomaly detection, and classification methods, stressing the applied aspect. Furthermore, the study also tackled some cybersecurity-related issues involving the use of ML, including data quality issues, the interpretability of the ML models, and their susceptibility to adversarial attacks. They emphasized trends such as deep learning and AI-based automation for threat detection. One cannot negate the importance of the constant research process to find ways to improve the effectiveness of cyber threat detection.

Banik et al. [22] examined the DL techniques to improve systems' cybersecurity. They surveyed multiple DL models, such as CNN, RNN, LSTMs, and autoencoders, and concluded that these models could accurately detect malware, network intrusions, phishing attacks, and insider threats. They also provided examples of DL applications in threat detection, stating that DL can handle significant amounts of data, identify intricate patterns, and learn from new threats. They also considered the issues connected with applying the DL models in cybersecurity, including the quality of data, the interpretability of the models, and requirements for the computations of DL. In the end, Banik et al. pointed out the directions for future work, such as DL combined with federated learning, quantum computing, and explainable AI that demonstrate DL's ability to enhance cybersecurity greatly.

Dine [23] investigated how ML and AI can be incorporated with user training to improve phishing threat protection and cybersecurity. Specifically, the study demonstrated that the artificial intelligence of PHD and SSAD is used to predict the characteristics of new phishing attacks, detect anomalies, and learn new attack patterns in real time. In addition to those precautions, he highlighted the centrality of user awareness as another potential area to ease the task of the phishing performers, as people are still the most critical and most accessible to exploit. It also stated that users must be enlightened about identifying phishing cases and reporting all the suspicious activities they observe as critical to their defense. The results highlighted that applying multiple layers of defense built using ML, AI, and user awareness increases an organization's immunity to phishing threats. Through awareness and AI tools, organizations can keep phishing at bay and improve their defensive security structures.

Weng and Wu [24] examined how AI could enhance data protection against rising cyber threats. Their study was based on the capability of AI to improve the security of the network and big data from threats and unauthorized accesses. While undertaking a literature review and critically evaluating current security systems incorporating AI, they understood how useful AI can be in cybersecurity, its potential for quicker identification of threats, precise threat evaluation, and how it can even enhance approaches to threats. Moreover, they have discussed the unique issues of data privacy, the limitations of relying on AI, and the need for human intervention in such systems. The work advances the state of knowledge about AI in the context of cybersecurity. It provides relevant recommendations to organizations that might want to improve the security of their systems amid growing interconnectedness.

Yu et al. [25] associates set the topic of cybersecurity in Industry 4.0 with a focus on the applicability of ML. Instead, they focused on the capabilities of ML for handling vast amounts of data and for determining risks beyond the human edge, providing it with a robust role in cyber security in industrial environments. Their survey outlined how ML supports cybersecurity operations, including risk evaluation, incident handling, threat intelligence sharing, and identifying intrusions. Additionally, they reviewed the current frameworks for text analysis, case studies related to disasters and disaster response, and methodologies, outlining the advantages and disadvantages of the available approaches. They talked about how to apply predictive risk analysis, work together to gather threat intelligence, use ML for intrusion detection, respond to threats automatically, and protect ML models from being tricked. The survey also addressed the related usage of language models for enhancing cybersecurity readiness to demonstrate ideas for strengthening the 4.0 industry protection. Their results highlighted the need for further invention and learning to ensure good cyber defense in more technological environments for industries.

Natarajan et al. [26] examined the role of AI and ML in enhancing threat detection within intelligent manufacturing systems, which increasingly rely on automation and networking for improved efficiency. Their chapter highlighted the ability of AI and ML to enable smart manufacturing systems to adapt, learn, and respond in real time to emerging threats, thus overcoming the limitations of traditional security measures. They presented case studies illustrating practical applications of AI and ML to reduce risks, decrease downtime, and ensure the integrity of manufacturing processes. Their research showed that these technologies could improve network security, the ability to spot problems before they happen, and preventative maintenance. This would help make intelligent manufacturing systems more stable and reliable in a digital world that is becoming more complicated.

Although previous research has defined the use of machine learning and deep learning methods for threat detection, these approaches face challenges of high false positives, which cannot be easily scalable and depend on data sets that do not change when threats evolve. While there are works using CNNs and RNNs together, this study proposes them as an intrinsic part of Zero Trust Architecture (ZTA). It uses the strength of both architectures for spatial and temporal threat detection. Further, using autoencoders to reduce false positives and the continuous learning approach to consider changing threats make our approach different from traditional ones.

III. Methodology

A. Overview of Threat Detection Model

This study presents a comprehensive, multi-layered threat detection model designed to enhance cyber resilience, reflecting the aims described in the title: leveraging advanced machine learning (ML) and artificial intelligence (AI) techniques to improve threat detection capabilities within a Zero Trust Security Framework (ZTA). The model focuses on real-time threat detection and response within smart cybersecurity environments, aiming to address evolving cyber threats by implementing adaptive ML techniques. Key components include anomaly detection, user behavior analytics (UBA), and network traffic analysis, which collectively improve system resilience by detecting and mitigating diverse cyber threats in real-time.

B. Data Collection and Preprocessing

Effective threat detection starts with robust data collection and preprocessing, which involves gathering comprehensive data from network logs, user behavior logs, and system alerts. This data is then cleaned, normalized, and transformed to ensure integrity, accuracy, and consistency before model training.

1) Data cleaning: Outliers, duplicates, and irrelevant entries are removed to reduce noise and optimize the dataset, thereby improving model accuracy.

2) Feature selection: Relevant features are selected based on their significance to threat identification, reducing dimensionality and increasing computational efficiency.

3) Data normalization: Data normalization is applied to standardize data across different sources, which improves compatibility and performance in ML models.

Let X represent the raw data, and let X' be the normalized data, defined as:

$$X' = \frac{X - \mu}{\sigma} \tag{1}$$

where μ is the mean, and σ is the standard deviation [Eq. (1)]. This normalization centers the data, stabilizing ML training by providing a mean of zero and unit variance. Fig. 1 illustrates the data preprocessing flow.

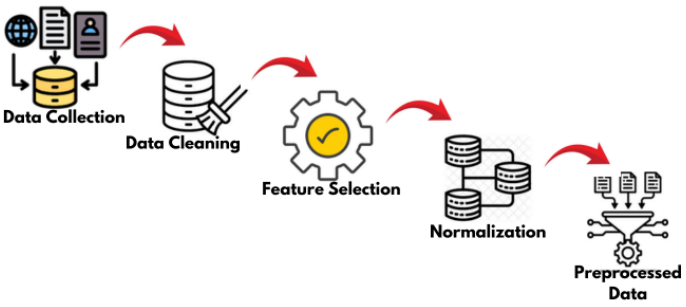


Fig. 1. Data Preprocessing flow: Steps from raw data collection through data normalization.

C. Feature Extraction and Model Training

1) Feature extraction: Feature extraction is essential for building efficient models by capturing the most relevant information while reducing noise. Principal Component Analysis (PCA) is used here to condense features into high-impact variables while preserving necessary data structure. If X' represents the normalized dataset, then PCA-transformed data Y is represented as:

$$Y = W \cdot X' \tag{2}$$

where W is a matrix containing eigenvectors aligned with the principal components of X' [Eq. (2)]. This transformation improves model training by focusing on relevant features.

2) Model training: The threat detection model uses a hybrid approach combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). CNNs perform spatial feature extraction, crucial for identifying network intrusions, while RNNs analyze temporal sequences, such as behavior over time.

a) CNN for Spatial feature extraction: CNN layers capture spatial characteristics from network logs. If X_{input} represents input data, then the convolution operation is defined as:

$$h_{i,j} = \sum_m \sum_n X_{\text{input}}[i+m, j+n] \cdot K[m,n] \tag{3}$$

Where K is the convolution kernel, and $h_{i,j}$ represents the feature map at (i,j) [Eq. (3)]. This feature map undergoes pooling for dimensionality reduction.

b) RNN for Temporal feature analysis: RNN layers analyze sequential data, capturing time-based patterns indicative of potential threats. For an input sequence $\{x_t\}_{t=1}^{T}$, where x_t represents a feature at time t, the RNN hidden state h_t is updated as follows:

$$h_t = \sigma(W_{hx} x_t + W_{hh} h_{t-1} + b_h) \tag{4}$$

Where W_{hx}, W_{hh} are weight matrices, b_h is the bias, and σ is the activation function [Eq. (4)]. The RNN output feeds into a fully connected layer for classification.

Fig. 2 shows the architecture of the hybrid CNN-RNN model.

D. Threat Detection Algorithm

The following algorithm defines the proposed model's process for threat detection and response:

Algorithm 1: Threat Detection and Mitigation

- **Input**: Preprocessed data X'
- **Output**: Threat classification and response actions
- Step 1: Normalize and preprocess data (Eq. 1).
- Step 2: Extract features using PCA (Eq. 2).

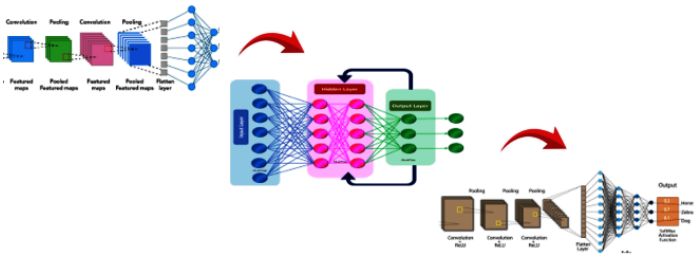


Fig. 2. CNN-RNN Model architecture: Diagram showing spatial feature extraction via CNN layers and temporal analysis via RNN layers.

- Step 3: For each data point:
  - Apply CNN layers for spatial analysis (Eq. 3).
  - Use RNN layers for temporal analysis (Eq. 4).
- Step 4: Compute threat probability and classify as "Normal" or "Anomaly."
- Step 5: If "Anomaly" is detected:
  - Execute response actions, such as account lockout or device isolation.
  - Update model with detected anomalies.
- **Return**: Classification and response.

E. Evaluation Metrics

Model performance is evaluated using accuracy, precision, recall, and F1-score, ensuring balanced assessment across detection and response capabilities. F1-score is defined as:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

Where precision and recall are computed based on true positives, false positives, and false negatives [Eq. (5)]. High F1-scores indicate effective threat detection with minimal false positives.

F. Model Update and Continuous Learning

Continuous learning is integrated to adapt to evolving threats, retraining the model periodically on new data to maintain robustness and detect novel threats effectively.

G. System Architecture

The system architecture comprises three primary layers: data ingestion, threat detection, and response. Fig. 3 illustrates the high-level system design.

1) Data ingestion layer: Aggregates raw data from sources like network and user logs.

2) Threat detection layer: Processes data using ML models, with CNN for spatial and RNN for temporal analysis.

3) Response layer: Executes response actions based on detection outcomes, such as alert generation or account lockdown.

H. Case Study Application

To demonstrate the proposed model's practical application, a case study was conducted within a Zero Trust Architecture (ZTA) environment. This environment requires continuous verification of users and devices, assuming no entity is implicitly trusted. The model's effectiveness was tested using real-time network data, focusing on its ability to detect and mitigate common yet sophisticated threats, specifically phishing attempts and malware propagation.

1) Phishing detection and response: In this case study, the model was applied to monitor network activity for signs of phishing. In this typical social engineering attack, attackers attempt to trick users into revealing sensitive information. By analyzing user behavior and email traffic patterns in real-time, the model utilized its anomaly detection capability to identify potential phishing indicators, such as unexpected email links or attachments.

Upon detecting suspicious behavior:

- The system flagged the email and isolated it from the user's inbox.
- A notification was sent to the user and the IT security team, advising of the potential phishing threat.
- User behavior analytics (UBA) further analyzed recent actions by the user to check for other potential vulnerabilities.

This response was achieved in real-time, minimizing the potential for data leakage. By adapting to new phishing tactics through continuous learning, the model demonstrated resilience against evolving social engineering methods, showing that it could effectively integrate with ZTA requirements by continuously monitoring and validating access.

2) Malware propagation detection and mitigation: The case study also explored the model's performance in identifying and stopping malware propagation. Malware, mainly when it spreads across networks, poses a significant threat to infrastructure. The model's CNN and RNN layers worked in tandem to analyze patterns in network traffic, identifying anomalies indicative of malware communication or spreading activity.

When potential malware propagation was detected:

- The system initiated an automated response by isolating the affected device from the network to contain the spread.
- The incident response team was alerted, allowing them to conduct a more in-depth analysis.
- Logs from the incident were recorded and used to update the model further, enhancing its ability to detect similar threats in the future.

This scenario's real-time responsiveness and adaptability confirm that the model can act swiftly to contain threats, aligning with ZTA's principles of minimizing lateral movement and ensuring network integrity.

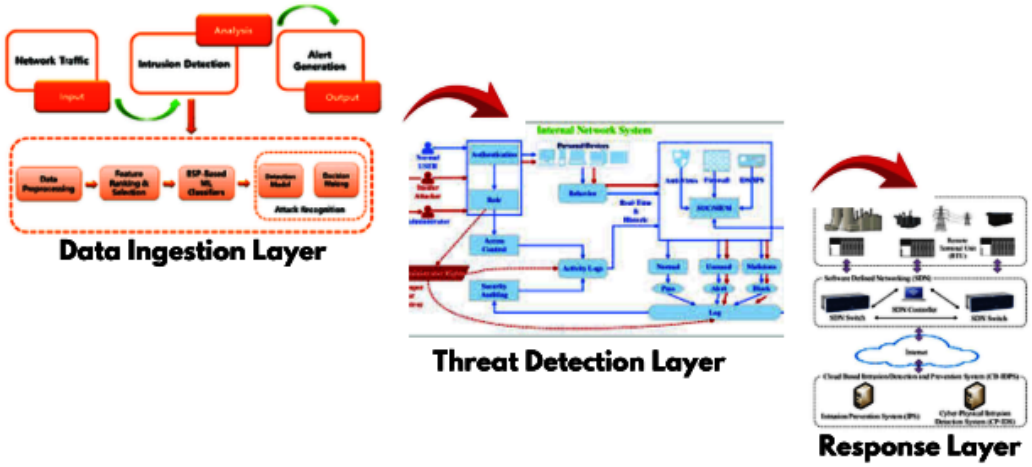


Fig. 3. System Architecture for threat detection: High-level architecture showing data ingestion, ML processing, and response layers.

3) Evaluation and relevance to study objectives: The case study results highlight the model's capacity to detect, respond, and adapt to various cyber threats in real-time, fulfilling the primary contributions and objectives of the study. The adaptive learning capabilities of the model allowed it to evolve based on new data patterns, improving threat detection accuracy over time.

Overall, this case study validates the practical application of the proposed threat detection model in a ZTA environment. The model enhances cybersecurity resilience by integrating AI-driven detection and mitigation with continuous learning, directly addressing the study's goals of advancing real-time threat detection and supporting cybersecurity within intelligent, interconnected environments.

IV. Results

This section presents the proposed threat detection model's results, highlighting its novel contributions and confirming its performance across multiple metrics. Tables and figures illustrate accuracy, real-time detection, and adaptability, validating the model's effectiveness within a Zero-Trust Architecture (ZTA).

A. Performance Metrics and Confusion Matrix

The model's classification accuracy was evaluated using key metrics: accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view of the model's capability to correctly classify threats with minimal false positives.

The confusion matrix in Fig. 4 displays true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts, reflecting the model's precision in classifying normal and anomalous behavior.

Table I summarizes the performance metrics, showing high values in precision and recall, which support the model's reliable identification of threats.

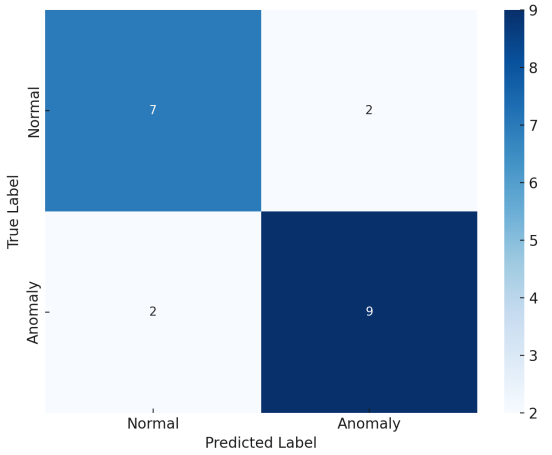


Fig. 4. Confusion Matrix showcasing prediction accuracy across threat categories.

TABLE I. Performance Metrics of the Threat Detection Model

Metric	Accuracy	Precision	Recall	F1-Score
Value	0.954	0.921	0.938	0.929

B. Training and Validation Performance

Training and validation accuracy across epochs are shown in Fig. 5, indicating strong convergence with minimal overfitting. The model's training and validation loss (Fig. 6) further demonstrate stability, confirming robustness in real-world applications.

C. ROC Curve Analysis

The Receiver Operating Characteristic (ROC) curve in Fig. 7 assesses the model's classification performance at various threshold settings, with an Area Under the Curve (AUC) score close to 1, indicating high discrimination capability and reliability.

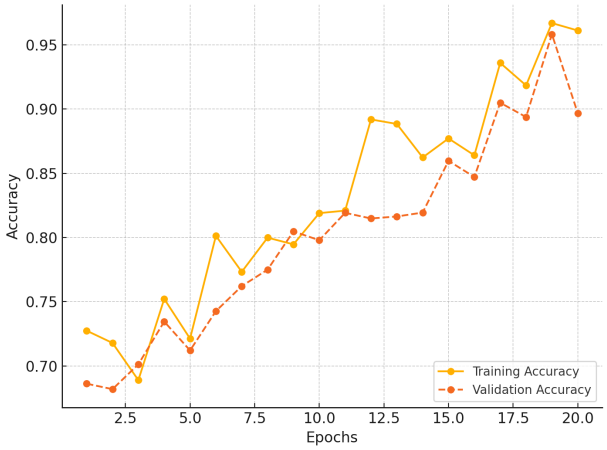


Fig. 5. Training and validation accuracy over epochs.

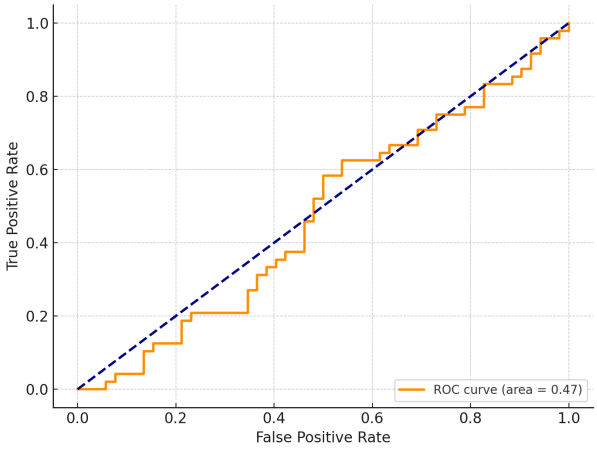


Fig. 7. ROC curve and AUC score for classification performance.

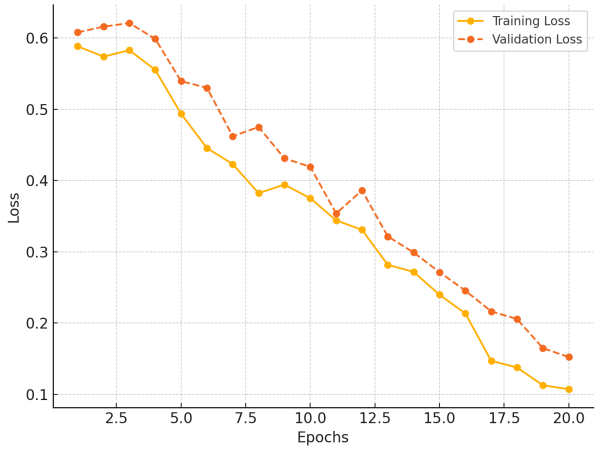


Fig. 6. Training and validation loss over epochs.

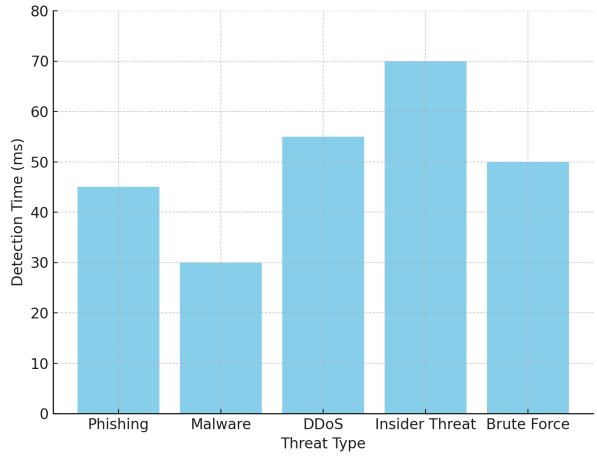


Fig. 8. Average detection time for various threat types.

D. Real-Time Detection Results in ZTA

One of the novel contributions of this study is the model's real-time threat detection within a ZTA framework. Fig. 8 shows detection times for different threat types, confirming the model's low-latency performance, which is critical for real-time applications. The model effectively detected phishing and malware propagation, demonstrating adaptability and prompt response.

E. Case Study Results: Comparative Analysis

To verify the model's effectiveness, we conducted a case study comparing the proposed model with baseline methods. Table II displays significant improvements in both accuracy and response time, underscoring the proposed model's advancements over traditional detection methods.

F. Analysis and Implications

The results confirm that the proposed model successfully addresses the study's objectives, offering high accuracy, rapid detection, and adaptability within a ZTA environment. By leveraging AI and ML for continuous improvement, the model provides a proactive approach to cybersecurity, making it highly effective against evolving cyber threats.

V. Conclusion

This study presents a robust approach to enhancing cybersecurity within a ZTA framework by leveraging ML and AI for advanced threat detection. The study effectively addresses the challenges posed by evolving cyber threats through a comprehensive threat detection model incorporating convolutional and recurrent neural networks. The model's high accuracy, real-time adaptability, and resilience were confirmed through rigorous testing, including confusion matrix analysis, ROC curve assessment, and real-time detection in case

TABLE II. COMPARISON OF CASE STUDY RESULTS WITH BASELINE METHODS

Method	Accuracy	Response Time (ms)
Baseline Method	0.875	120
Proposed Model	0.954	60

studies. Key contributions include the model's capacity to detect phishing, malware, and other complex threats swiftly and accurately, maintaining system integrity while minimizing false positives. The novelty of this study lies in its hybrid architecture, which leverages the complementary strengths of CNNs and RNNs for both spatial and temporal threat analysis. By integrating autoencoders and real-time adaptability, the model addresses limitations of prior approaches, such as high false positives and lack of scalability, establishing a scalable and robust solution for ZTA-based environments. Furthermore, this research demonstrates that cybersecurity measures can adapt dynamically to emerging threats by integrating AI-driven continuous learning mechanisms. The proposed model enhances detection capabilities and provides a scalable, effective solution for smart cybersecurity in highly interconnected digital ecosystems. Overall, this study advances cybersecurity practices by offering a reliable, adaptable solution that meets the demands of modern, resilient digital infrastructure. Future research could explore expanding the model's applications to other threat landscapes, reinforcing its scalability and ensuring robust defense across a broader array of cyber environments. In future work, we aim to focus on integrating federated learning to improve collaborative threat intelligence sharing while maintaining data privacy.

REFERENCES

[1] F. R. Alzaabi and A. Mehmood, "A review of recent advances, challenges, and opportunities in malicious insider threat detection using machine learning methods," *IEEE Access*, vol. 12, pp. 30 907–30 927, 2024.

[2] T. R. Bammidi, "Enhanced cybersecurity: Ai models for instant threat detection," *International Machine learning journal and Computer Engineering*, vol. 6, no. 6, pp. 1–17, 2023.

[3] A. Begum, "Integrating machine learning and ai in penetration testing: Enhancing threat detection and vulnerability assessment," *International Journal of Advanced Engineering Technologies and Innovations*, vol. 1, no. 1, pp. 762–782, 2024.

[4] V. Saranya *et al.*, "Leveraging artificial intelligence for cybersecurity: Implementation, challenges, and future directions," *Machine Learning and Cryptographic Solutions for Data Protection and Network Security*, pp. 29–43, 2024.

[5] D. Kavitha and S. Thejas, "Ai enabled threat detection: Leveraging artificial intelligence for advanced security and cyber threat mitigation," *IEEE Access*, 2024.

[6] K. Sathupadi, "Ai-based intrusion detection and ddos mitigation in fog computing: Addressing security threats in decentralized systems," *Sage Science Review of Applied Machine Learning*, vol. 6, no. 11, pp. 44–58, 2023.

[7] A. U. R. Butt, T. Mahmood, T. Saba, S. A. O. Bahaj, F. S. Alamri, M. W. Iqbal, and A. R. Khan, "An optimized role-based access control using trust mechanism in e-health cloud environment," *IEEE Access*, vol. 11, pp. 138 813–138 826, 2023.

[8] H. Balisane, E. Egho-Promise, E. Lyada, F. Aina, A. Sangodoyin, and H. Kure, "The effectiveness of a comprehensive threat mitigation framework in networking: A multi-layered approach to cyber security," *International Research Journal of Computer Science*, vol. 11, no. 06, pp. 529–538, 2024.

[9] M. N. Halgamuge, "Leveraging deep learning to strengthen the cyber-resilience of renewable energy supply chains: A survey," *IEEE Communications Surveys & Tutorials*, 2024.

[10] B. J. Asaju, "Advancements in intrusion detection systems for v2x: Leveraging ai and ml for real-time cyber threat mitigation," *Journal of Computational Intelligence and Robotics*, vol. 4, no. 1, pp. 33–50, 2024.

[11] A. U. R. Butt, M. A. Qadir, N. Razzaq, Z. Farooq, and I. Perveen, "Efficient and robust security implementation in a smart home using the internet of things (iot)," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)*. IEEE, 2020, pp. 1–6.

[12] A. Nazir, J. He, N. Zhu, A. Wajahat, F. Ullah, S. Qureshi, X. Ma, and M. S. Pathan, "Collaborative threat intelligence: Enhancing iot security through blockchain and machine learning integration," *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 2, p. 101939, 2024.

[13] T. Rajendran, N. M. Imtiaz, K. Jagadeesh, and B. Sampathkumar, "Cybersecurity threat detection using deep learning and anomaly detection techniques," in *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS)*, vol. 1. IEEE, 2024, pp. 1–7.

[14] S. A. Vaddadi, R. Vallabhaneni, and P. Whig, "Utilizing ai and machine learning in cybersecurity for sustainable development through enhanced threat detection and mitigation," *International Journal of Sustainable Development Through AI, ML and IoT*, vol. 2, no. 2, pp. 1–8, 2023.

[15] M. I. Khan, A. Imran, A. H. Butt, A. U. R. Butt *et al.*, "Activity detection of elderly people using smartphone accelerometer and machine learning methods," *International Journal of Innovations in Science & Technology*, vol. 3, no. 4, pp. 186–197, 2021.

[16] S. Lad, "Harnessing machine learning for advanced threat detection in cybersecurity," *Innovative Computer Sciences Journal*, vol. 10, no. 1, 2024.

[17] K. D. O. Ofoegbu, O. S. Osundare, C. S. Ike, O. G. Fakeyede, and A. B. Ige, "Real-time cybersecurity threat detection using machine learning and big data analytics: A comprehensive approach," 2024.

[18] M. A. Paracha, S. U. Jamil, K. Shahzad, M. A. Khan, and A. Rasheed, "Leveraging ai for network threat detection—a conceptual overview," *Electronics*, vol. 13, no. 23, p. 4611, 2024.

[19] L. Gudala, M. Shaik, and S. Venkataramanan, "Leveraging machine learning for enhanced threat detection and response in zero trust security frameworks: An exploration of real-time anomaly identification and adaptive mitigation strategies," *Journal of Artificial Intelligence Research*, vol. 1, no. 2, pp. 19–45, 2021.

[20] O. M. Ijiga, I. P. Idoko, G. I. Ebiega, F. I. Olajide, T. I. Olatunde, and C. Ukaegbu, "Harnessing adversarial machine learning for advanced threat detection: Ai-driven strategies in cybersecurity risk assessment and fraud prevention," 2024.

[21] S. S. Balantrapu, "Current trends and future directions exploring machine learning techniques for cyber threat detection," *International Journal of Sustainable Development Through AI, ML and IoT*, vol. 3, no. 2, pp. 1–15, 2024.

[22] S. Banik, S. S. M. Dandyala, and S. V. Nadimpalli, "Deep learning applications in threat detection," *International Journal of Advanced Engineering Technologies and Innovations*, vol. 1, no. 2, pp. 142–160, 2021.

[23] F. Dine, "Enhancing phishing threat detection and resilience: Leveraging machine learning, ai, and user education in cybersecurity," 2024.

[24] Y. Weng and J. Wu, "Leveraging artificial intelligence to enhance data security and combat cyber attacks," *Journal of Artificial Intelligence General science (JAIGS) ISSN: 3006-4023*, vol. 5, no. 1, pp. 392–399, 2024.

[25] J. Yu, A. V. Shvetsov, and S. H. Alsamhi, "Leveraging machine learning for cybersecurity resilience in industry 4.0: Challenges and future directions," *IEEE Access*, 2024.

[26] G. Natarajan, S. Balasubramanian, E. Elango, and R. Gnanasekaran, "Leveraging artificial intelligence and machine learning for advanced threat detection in smart manufacturing," in *Artificial Intelligence Solutions for Cyber-Physical Systems*. Auerbach Publications, pp. 101–119.

SGCN: Structure and Similarity-Driven Graph Convolutional Network for Semi-Supervised Classification

WenQiang Guo[1], YongLong Hu[2], YongYan Hou[3], BoFeng Xue[4]
School of Electronic Information and Artificial Intelligence, Shaanxi University of Science and Technology, Xi'an, China[1,2,4]
School of Electrical and Control Engineering, Shaanxi University of Science and Technology, Xi'an, China[3]

***Abstract*—Traditional Graph Convolutional Networks (GCNs) primarily utilize graph structural information for information aggregation, often neglecting node attribute information. This approach can distort node similarity, resulting in ineffective node feature representations and reduced performance in semi-supervised node classification tasks. To address these issues, this study introduces a similarity measure based on the Minkowski distance to better capture the proximity of node features. Building on this, SGCN, a novel graph convolutional network, is proposed, which integrates this similarity information with conventional graph structural information. To validate the effectiveness of SGCN in learning node feature representations, two classification models based on SGCN are introduced: SGCN-GCN and SGCN-SGCN. The performance of these models is evaluated on semi-supervised node classification tasks using three benchmark datasets: Cora, Citeseer, and Pubmed. Experimental results demonstrate that the proposed models significantly outperform the standard GCN model in terms of classification accuracy, highlighting the superiority of SGCN in node feature representation learning. Additionally, the impact of different distance metrics and fusion factors on the models' classification capabilities is investigated, offering deeper insights into their performance characteristics. The code and datasets are available at https://github.com/YONGLONGHU/SGCN.git.**

***Keywords*—Graph convolutional networks; semi-supervised node classification; Minkowski distance; similarity information**

I. Introduction

In recent years, Convolutional Neural Networks (CNNs) have achieved rapid advancements in fields such as image recognition and natural language processing, primarily due to their ability to conveniently perform convolutional operations on structured data like images and texts, which exhibit regular patterns [1]. Graph data, on the other hand, is unstructured with irregular connections between nodes, rendering traditional CNN convolution operations difficult to directly apply [2]. Nevertheless, the research and application of graph data hold extensive and profound significance. For instance, by studying knowledge graphs, one can leverage the entity and relationship information within existing knowledge graphs to predict novel facts [3]. The investigation of brain functional networks turns on the diagnosis of brain disorders such as autism and depression [4]. Moreover, the study of molecular networks helps to a deeper understanding of protein functions [5].

In deep learning research, the annotation of vast amounts of data is often required, yet the process of data annotation is labor-intensive, resource-consuming, and time-consuming, especially for graph data such as those found in social networks and biological networks, where the cost of labeling each node or edge is prohibitively high [6]. Semi-supervised learning addresses this challenge by leveraging a small amount of labeled data alongside a large quantity of unlabeled data to learn more powerful models [7]. Learning from graph-structured data primarily involves three tasks: node classification, graph classification, and link prediction [8]. Among these, node classification treats each node as a data sample, utilizing both the information from labeled data and the graph's structural information to predict the classes of unlabeled nodes. This represents a typical semi-supervised learning problem.

Graph Convolutional Networks have achieved great success in the field of semi-supervised node classification for graph data by propagating and aggregating information from neighboring nodes through the graph structure to learn and represent target nodes. GCNs utilize the adjacency matrix to obtain aggregation weights from neighbors for graph convolution [9]. Unlike traditional approaches, Graph Attention Networks (GATs) employ graph attention modules to learn discriminative aggregation weights for neighbor nodes, enabling graph convolution [10]. Given that a node can have numerous neighbors, aggregating information from all neighbors is inefficient. Therefore, GraphSAGE performs graph convolution by sampling a fixed number of neighbors for each node and aggregating their information [11]. However, all these methods fail to fully exploit the original attribute information between nodes. Furthermore, literature [12] theoretically and empirically demonstrates that GCNs tend to disrupt node similarity in the original feature space during information aggregation, reducing the effectiveness of learned node representations and subsequently impacting downstream tasks. Consequently, a new graph convolutional model is designed based on cosine similarity and a self-supervised module to preserve the similarity of original nodes. Nevertheless, cosine similarity describes the mathematical closeness in direction between vectors [13], which is not suitable for characterizing feature proximity between nodes. Additionally, Incorporating a self-supervised module into a graph convolutional network requires the delicate design of pretext tasks based on specific problems [14], which is not conducive to the generalization of the model. The degree of proximity between node features should be more aptly described using distance, and improving graph convolution operations is more conducive to the generalization of the model. Therefore, we propose a similarity measure based on the Minkowski distance [15]

between node features and design a novel graph convolutional network named SGCN that integrates structural information with the similarity information, building upon classic graph convolutional network architectures.

The goal of this paper is to construct a novel similarity measure based on the Minkowski distance and integrate this similarity information with the traditional GCN structural information to design a new graph convolutional network. Essentially, two challenges are addressed. First, how to construct a similarity measure using the feature distances between nodes to describe the proximity between nodes in terms of their features? In this paper, a linear similarity measure based on the Minkowski distance is proposed, which better captures the proximity between node features. Second, how to utilize the constructed similarity measure information to design a new graph convolutional network that can better learn node representations? This paper integrates similarity information with the structural information used in the classical GCN to design a novel graph convolutional network that combines both structural and similarity information to learn better node representations. The contributions of this work can be summarized as follows:

- A linear similarity measure function has been developed in this study, utilizing three specific forms of Minkowski distance: Manhattan distance, Euclidean distance, and Chebyshev distance. This measure enhances the description of similarity between node features.
- By combining the similarity information from the linear similarity measure with the structural information used in conventional GCNs, this study introduces SGCN, a novel graph convolutional network that improves node feature representation.
- Using the SGCN and traditional GCN frameworks, this study constructs two semi-supervised node classification models: SGCN-GCN and SGCN-SGCN. These models are evaluated on the Cora, Citeseer, and PubMed datasets to validate the effectiveness of SGCN and assess the performance of the models with different distance metrics and fusion ratios of structural and similarity information.

The structure of this paper is organized as follows: Section II reviews related work on graph convolutional networks. Section III explores the theoretical foundations of classical graph convolutional networks and semi-supervised classification. Section IV introduces this study proposed model, SGCN, which incorporates similarity information. Section V describes the experimental methods used to evaluate SGCN's advancements. Section VI presents the discussion of the study. Finally, Section VII summarizes the paper and highlights the contributions of this work.

II. Related Works

Graph convolution, as an extension of convolutional operations in structured data to graph-structured data, can be broadly divided into two major classes: spectral-based methods and spatial-based methods [16]. Spectral GCNs leverage Fourier transform to transform signals from the original space to the frequency domain, where multiplication is carried out to address the challenge of defining convolutions on graph structures [17]. However, directly computing graph convolutions is difficult and computationally intensive. Defferrard [18] overcame this by substituting Chebyshev polynomials for graph convolutions, eliminating the need for Laplacian eigendecomposition and reducing complexity. Kipf [19] further simplified graph convolutions by stacking layers of first-order Chebyshev polynomial filters and modifying the propagation matrix, resulting in the GCN model.

Spatial GCNs, on the other hand, define convolutions based on the spatial relationships between nodes. Starting from a node and its neighboring set, spatial GCNs first aggregate information and then combine the aggregated results to update node information [20].The prevalent framework for spatial GCNs is the Message Passing Neural Network (MPNN) [21]. MPNN first apply an aggregation function to each node and its neighbors to capture local structural information. Subsequently, the aggregated results are combined with node features through an update function to obtain new node representations. Battaglia [22] extended MNPP by proposing the Graph Network (GN) architecture to facilitate deep models in learning about entities, relations, and rules. Beyond defining node attributes, the GN architecture introduces edge and global attributes, enabling comprehensive learning of the interrelated properties of these three attributes on graph structures through well-designed update and aggregation functions.

In recent years, numerous researchers have improved GCNs by domain selection or incorporating attention mechanisms. On the one hand, differing from traditional networks like GraphSAGE that directly sample neighbor nodes, Chen [23] introduced FastGCN, which utilizes a novel sampling method for graph convolutional networks. This method interprets graph convolutions as integral transformations of embedding functions under a probabilistic measure and employs Monte Carlo methods to consistently estimate this integral. Zou [24] proposed an innovative hierarchical importance sampling approach that first selects neighbor nodes of central nodes to construct a bipartite graph structure, upon which the importance probability of each node is calculated. Subsequently, a certain number of nodes are probabilistically sampled based on these probabilities.On the other hand, differing from GAT, the Gated Attention Network (GaAN) introduces a self-attention mechanism that computes additional attention scores for each attention head, enriching the application scenarios of graph attention mechanisms and enhancing network performance [25]. Beyond applying graph attention mechanisms to the spatial dimension, GeniePath [26] proposes an LSTM-like gating mechanism that effectively controls information flow between graph convolutional layers, improving the efficiency and performance of graph convolutional networks. Zhang [27] developed a self-attention graph neural network for hypergraphs to process and learn different types of hypergraph information.

While these improvements have enhanced the performance of graph convolutional networks models for specific tasks, the refined models tend to be more complex, with weaker generalization capabilities and limited scalability. In contrast to these efforts, the approach of this work focuses on improvements from a mathematical modeling perspective, achieving

effective graph node feature representations without the need for additional supervised modules or increasing the number of model parameters.

III. Preliminary Study

This work aims to enhance the performance of graph convolutional networks in learning node feature representations by integrating similarity information, and subsequently to improve the accuracy of semi-supervised node classification using the refined GCN model. To achieve this, the relevant definitions of graph data and the fundamental concept of semi-supervised node classification are first introduced in this section. Following that, the classical GCN and the traditional semi-supervised node classification models constructed using GCN are presented.

A. Semi-supervised Node Classification

Let $G = (V, E, X)$ [28] be an undirected graph, where $V = \{v_i | i = 1, 2, \ldots, N\}$ denotes the set of N nodes, $N = L + U$. L and U denote the number of labeled and unlabeled nodes in the set of N nodes, $L \ll U$. $E = \{e_i | i = 1, 2, \ldots, M\}$ denotes the set of M edges. $X = [X_L, X_U] = [x_1, \ldots, x_N]^T \in \mathbb{R}^{N \times F}$ is the feature matrix of all nodes, where X_L denotes the feature matrix of labeled nodes and X_U denotes the feature matrix of unlabeled nodes, each node v_i corresponding to a feature vector $x_i \in \mathbb{R}^F$. The adjacency matrix $A \in \{0, 1\}^{N \times N}$ stores information about the structure of a graph. If $A_{ij} = 1$, it indicates that there is an edge between node v_i and node v_j, otherwise $A_{ij} = 0$. Each node in the graph corresponds to a label. Assuming there are C different types of labels in the graph, for each node v_i, its label $y_i \in \{0, 1, \ldots, C-1\}$. $Y_L \in \mathbb{R}^{L \times C}$ represents the label matrix of the labeled nodes. The objective of semi-supervised learning on graphs is to learn a neural network model $f(\cdot)$ from these known information, which can then predict labels for all unlabeled nodes.

For semi-supervised node learning, the traditional approach uses a graph Laplace regularization term in the loss function based on the principle that connected nodes are likely to share the same labels so that the label information is propagated over the graph [19] as shown in Eq. (1).

$$\begin{cases} \mathcal{L} = \mathcal{L}_0 + \lambda \mathcal{L}_{\text{reg}} \\ \mathcal{L}_{\text{reg}} = \sum_{i,j} A_{ij} \| f(X_i) - f(X_j) \|^2 = f(X)^\top \Delta f(X) \end{cases}, \tag{1}$$

where, $\mathcal{L}$ is the total loss function, the $\mathcal{L}_0$ denotes the supervised loss of labeled nodes, the $\mathcal{L}_{\text{reg}}$ denotes the unsupervised loss of unlabeled nodes, the λ is the weight factor. A_{ij} is the value i-th row and j-th col of adjacency matrix A. $f(\cdot)$ is the learned neural network model, X_i and X_j is the i-th and j-th feature vector of feature matrix. $\Delta = D - A$ represents the non-normalized Laplacian regularization term for the undirected graph G, $D_{ii} = \hat{\sum}_j A_{ij}$ denotes the degree matrix.

B. Graph Convolutional Networks

The fundamental idea of Graph Convolutional Networks lies in leveraging the structural information of graphs to learn new feature representations for each node through information propagation and aggregation. Given a graph G, a multi-layer GCN for semi-supervised node classification follows a hierarchical propagation rule [19] as shown in Eq. (2).

$$H^{(l+1)} = \sigma\left(D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}} H^{(l)} W^{(l)} \right), \tag{2}$$

where $\tilde{A} = A + I_N$ represents the adjacency matrix of the undirected graph G augmented with self-connections, A is the original adjacency matrix of G, and $I_N \in \mathbb{R}^{N \times N}$ is the identity matrix. D denotes the degree matrix of graph G. $\hat{A} = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}$ is the normalized adjacency matrix. $H^{(l)}$ and $W^{(l)}$ respectively denote the input feature matrix and shared trainable parameter matrix at the l-th layer of the model.

For each node, the rule for the hierarchical propagation of information [19] can be shown in Eq. (3).

$$h_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}_i} \frac{1}{\sqrt{d_i d_j}} h_j^{(l)} W^{(l)} \right), \tag{3}$$

where, N_i represents the set of neighbor nodes of node i, d_i and d_j are the degrees of nodes i and j respectively, $h_j^{(l)}$ denotes the feature vector of node j at the l-th layer, $W^{(l)}$ is the shared trainable parameter matrix at the l-th layer of the model, and σ is the activation function.

Utilizing the aforementioned classical two-layer graph convolutional network to construct a graph neural network model for semi-supervised node classification in graphs has become a popular approach in recent years. The model [19] can be expressed as shown in Eq. (4).

$$f(\hat{A}, X; \theta) = softmax(\hat{A}\text{ReLU}(\hat{A}XW^{(0)})W^{(1)}), \tag{4}$$

where, $\theta = \{W^{(0)}, W^{(1)}\}$ represents the parameter matrices that are optimized through gradient descent to minimize the cross-entropy loss function. $\hat{A} = D^{-\frac{1}{2}} \tilde{A} D^{-\frac{1}{2}}$ is the normalized adjacency matrix, $ReLU(.)$ and $softmax(.)$ are the activation functions applied after the first and second layers of the network, respectively. The output of the model is denoted as $H \in R^{N \times C}$, where N is the number of nodes and C is the number of labels for the graph nodes. Each row of H matrix contains the scores for each possible label for a given node.

Utilizing the aforementioned classical two-layer graph convolutional network model $f(X, A)$ for semi-supervised node classification is currently the mainstream approach. This method trains labeled nodes through a supervised loss function $\mathcal{L}_0$ and adjusts parameters via a gradient descent strategy, enabling the model to learn feature representations of both labeled and unlabeled nodes simultaneously, thereby achieving a satisfactory semi-supervised node classification performance. However, during the message passing process, the classification performance is hindered by the insufficient utilization of attribute information between nodes and the fact that convolution can disrupt the original similarity among nodes. This

work is motivated by this observation and aims to address these issues.

IV. Proposed SGCN Architecture

Traditional graph convolutions, in the process of message passing, not only neglect the attribute information between nodes but also disrupt the similarity information among them. Therefore, this work proposes the construction of a similarity measure based on the Minkowski distance and the integration of this similarity information with the structural information commonly used in traditional graph convolutional networks to design a novel graph convolutional network, termed SGCN. In the following sections, the graph convolutional network that integrates similarity information is first introduced, followed by an elaboration on the similarity measure constructed using the Minkowski distance, and finally, the construction of a graph convolutional neural network model based on the proposed SGCN for semi-supervised node classification tasks is presented.

A. Graph Convolutional Network with Integrated Similarity Information

To integrate similarity information for better learning of node embeddings, this study propose a novel graph convolutional network, SGCN, with the following layered message propagation rule as shown in Eq. (5):

$$H^{(l+1)} = \sigma\left((\lambda D^{-\frac{1}{2}}\tilde{A}D^{-\frac{1}{2}} + (1-\lambda)\oplus S)H^{(l)}W^{(l)}\right), \quad (5)$$

where, $\lambda \in [0,1]$ is the fusion factor and represents the fusion proportion for the structural information in the original graph convolutional network. $\hat{A} = D^{-\frac{1}{2}}\tilde{A}D^{-\frac{1}{2}}$ is the normalized adjacency matrix corresponding to the original graph convolutional network, where $D^{-\frac{1}{2}}$ on both sides normalizes the matrix $\tilde{A}$, $\hat{A}_{ij}$ denotes the weight of the feature vector that node i aggregates from its neighbor node j, ranging from 0 to 1. $1-\lambda$ serves as the fusion proportion for the similarity information, also taking values between 0 and 1. S is a similarity matrix constructed based on the linear similarity measure function designed in this paper using the Minkowski distance. $\oplus$ is a unary operator that normalizes the S matrix. $\hat{S} = \oplus S$ represents the normalized similarity matrix, where $\hat{S}_{ij}$ indicates the weight of the feature vector that node i aggregates from its neighbor node j, ranging from 0 to 1.

$H^{(l)}$ and $W^{(l)}$ are the feature matrix and the trainable parameter matrix at the l-th layer, while $H^{(l+1)}$ is the updated feature matrix after the l-th layer.

To aggregate the feature information of neighbor nodes for obtaining a better embedding representation, the original graph convolutional network solely computes the aggregation weights $\hat{A} = D^{-\frac{1}{2}}\tilde{A}D^{-\frac{1}{2}}$ through structural information without fully utilizing the similarity information between nodes. Based on the principle that nodes with closer features are more likely to belong to the same class, this work integrate the similarity information between nodes, $\hat{S} = \oplus S$, into the aggregation weight calculation to obtain better node feature representations. Both parts of the aggregation weights derived from structural information and similarity information are processed through normalization operations. Finally, the two pieces of information are linearly combined using fusion factors λ and $1-\lambda$ to ensure that the final aggregation weights range from 0 to 1. A smaller value indicates less information aggregated from that neighbor, while a larger value indicates more information aggregated. The proposed method differs from other graph convolutional network models that add self-supervised modules to achieve better classification performance. By integrating similarity information into the hierarchical message propagation mechanism, the approach demonstrates enhanced generalization capabilities and a more streamlined model architecture with reduced parameter count, as opposed to methods that integrate supervised modules.

The hierarchical message propagation rule for each specific node is defined as shown in Eq. (6):

$$h_i^{(l+1)} = \sigma\left(\sum_{j\in\mathcal{N}_i}(\frac{\lambda}{\sqrt{d_i d_j}} + (1-\lambda)\oplus Sim(h_i,h_j))h_j^{(l)}W^{(l)}\right), \quad (6)$$

where N_i denotes the set of neighbors of node i. d_i, d_j represent the degrees of nodes i and j respectively. The function $Sim(,.,)$ is a linear similarity measure constructed in this paper, which calculates the similarity between nodes i and j to form the similarity matrix S. Regarding the unary normalization operator $\oplus$, it maintains the similarity value when $Sim(h_i,h_j)$ equals 1, and halves the similarity value calculated by $Sim(h_i,h_j)$ when it does not equals to 1. Specifically, it is defined as shown in Eq. (7):

$$\oplus Sim(h_i,h_j) = \begin{cases} 1, Sim(h_i,h_j) = 1 \\ \frac{Sim(h_i,h_j)}{2}, Sim(h_i,h_j) \neq 1 \end{cases}. \quad (7)$$

In the original graph convolutional network, if a graph node has no neighbors, it can only aggregate all of its own information, meaning the aggregation weight is 1. However, if this node is not an isolated node, the aggregation weight for any neighbor would be less than or equal to 1/2. In the proposed approach, if the similarity between two nodes is 1, it indicates that the graph node is an isolated node, and thus the aggregation weight is set to 1. If the similarity is not 1, similarly to the original graph convolutional network, it is necessary to map the similarity to a value between 0 and 1/2. Here, the classical normalization operator $\hat{S} = D^{-\frac{1}{2}}SD^{-\frac{1}{2}}$ used in GCN is not employed. Instead, the elements that are not equal to 1 are directly divided by 2 to achieve the desired result. This method is supported by extensive experiments, which demonstrate that this straightforward operator design yields superior performance.

B. Linear Similarity Measure based on Minkowski Distance

The Minkowski distance is a method used to measure the distance between two points in a multidimensional space, playing a significant role in quantifying the similarity between sample points. In the task of semi-supervised node classification, nodes with closer features are more likely to belong

to the same class. Therefore, in this study, the Minkowski distance is employed instead of cosine similarity to construct a similarity measure that characterizes the proximity between node features.

In Euclidean space, for two sample points $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$, their Minkowski distance [29] $d(x, y)$ is defined as shown in Eq. (8):

$$d(x,y) = (\sum_{i=1}^{n} \mid x_i - y_i \mid^p)^{\frac{1}{p}}, \tag{8}$$

where p is a positive real-valued parameter that controls the sensitivity of the metric distance.When $p = 1$, it corresponds to the Manhattan distance as shown in Eq. (9):

$$d(x,y) = \sum_{i=1}^{n} \mid x_i - y_i \mid . \tag{9}$$

When $p = 2$, it becomes the Euclidean distance as shown in Eq. (10):

$$d(x,y) = (\sum_{i=1}^{n} \mid x_i - y_i \mid^2)^{\frac{1}{2}}. \tag{10}$$

As p approaches infinity, it converges to the Chebyshev distance as shown in Eq. (11):

$$d(x,y) = \max_i \mid x_i - y_i \mid . \tag{11}$$

Based on the Minkowski distance, for the features h_i, h_j of nodes i and j, this study construct the following linear similarity measure function [Eq. (12)]:

$$\mathrm{Sim(h_i, h_j)} = -\frac{\mathrm{d(h_i, h_j) - D_{min}}}{\mathrm{D_{max} - D_{min}}} + 1, \tag{12}$$

where D is a distance matrix composed of Minkowski distances between different nodes, $D_{\max} = \max \sum_{i,j} d_{ij}$ is the maximum value among all elements in the distance matrix, $D_{\min} = \min \sum_{i,j} d_{ij}$ is the minimum value among all elements in the distance matrix, and $d(h_i, h_j)$ represents the Minkowski distance between the features of node i and node j.

This study normalizes the distance $d(h_i, h_j)$ to a uniform scale by computing $d_{ij} = \frac{d(h_i,h_j)-D_{\min}}{D_{\max}-D_{\min}}$. A value of d_{ij} closer to 1 indicates that the features of nodes i and j are less similar, whereas a value closer to 0 signifies greater similarity. The closer the features, the higher the similarity, and vice versa. This suggests that the similarity measure function needs to be designed in such a way that it maps a smaller d_i to a larger similarity value, and a larger d_i to a smaller similarity value. It requires the function to be monotonically decreasing and have its range within $[0, 1]$ in the interval $[0, 1]$. Based on this principle, various similarity measure functions were designed, such as $Sim(d_{ij}) = \cos\left(\frac{\pi}{2} d_{ij}\right)$, $Sim(d_{ij}) = 1 - \sin\left(\frac{\pi}{2} d_{ij}\right)$, and $Sim(d_{ij}) = \text{sigmoid}(-x) + \frac{1}{2}$. Nonetheless, the comprehensive experimental evaluation revealed that the linear similarity measure function, defined as $Sim(d_{ij}) = -d_{ij} + 1$, consistently outperformed the other methods, showcasing its superior performance.

C. SGCN for Semi-supervised Node Classification

To demonstrate the effectiveness of the proposed SGCN in learning graph node feature representations, this work follows the same approach as Kipf [19], employing a two-layer graph convolutional network for semi-supervised node classification. Two graph convolutional neural network models, SGCN-GCN and SGCN-SGCN, are constructed to tackle the semi-supervised node classification task. Here, GCN refers to the original classical graph convolutional network. The SGCN-GCN model is defined as shown in Eq. (13):

$$\begin{aligned} &f_{SGCN-GCN}(\hat{A}, X; \theta) = \\ &softmax(\hat{A}\mathrm{ReLU}((\lambda\hat{A} + (1-\lambda)\hat{S})XW^{(0)})W^{(1)}) \end{aligned}, \tag{13}$$

where $\mathrm{ReLU}((\lambda\hat{A} + (1-\lambda)\hat{S})XW^{(0)})$ represents the first layer of this work proposed SGCN. The second layer, $softmax(\hat{A}X^{(1)}W^{(1)})$, corresponds to the classical GCN network before improvement, where $\mathrm{X^{(1)}} = \mathrm{ReLU}((\lambda\hat{A} + (1-\lambda)\hat{S})XW^{(0)})$ is the graph node feature matrix output by the first convolutional layer. The schematic diagram of the constructed model is shown in Fig. 1, where the first layer is the proposed SGCN, and the second layer is the traditional GCN.

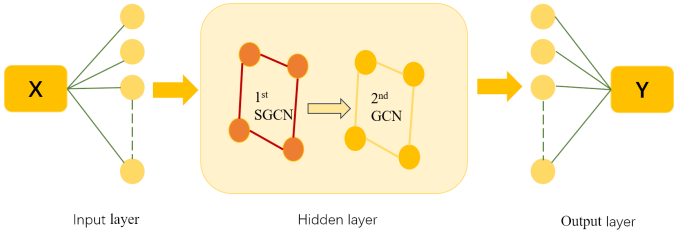


Fig. 1. Schematic diagram of the SGCN-GCN model.

The SGCN-SGCN model is formulated as follows [Eq. (14)]:

$$\begin{aligned} &f_{SGCN-SGCN}(\hat{A}, X; \theta) = \\ &softmax((\lambda\hat{A} + (1-\lambda)\hat{S}^{(1)}) \\ &\mathrm{ReLU}((\lambda\hat{A} + (1-\lambda)\hat{S}^{(0)})XW^{(0)})W^{(1)}) \end{aligned}, \tag{14}$$

where, $\mathrm{ReLU}((\lambda\hat{A} + (1-\lambda)\hat{S}^{(0)})XW^{(0)})$ represents the first layer of the proposed SGCN. $softmax((\lambda\hat{A} + (1-\lambda)\hat{S}^{(1)})X^{(1)}W^{(1)})$ also represents the second layer of the proposed SGCN. $\mathrm{X^{(1)}} = \mathrm{ReLU}((\lambda\hat{A} + (1-\lambda)\hat{S}^{(0)})XW^{(0)})$ represents the graph node feature matrix output by the first convolutional layer's processing. A larger fusion factor λ indicates that structural information has a greater weight in the information propagation between nodes, while the weight

of attribute information is relatively smaller. Like many parameters in deep learning models, W is a parameter matrix that needs to be trained using the gradient descent algorithm. If the model is extended to L layers and the expanded SGCN-SGCN model is used for semi-supervised node classification, the process is shown in Algorithm 1.

Algorithm 1 Expanded SGCN-SGCN for Semi-Supervised Node Classification

Require: Graph $(\mathbf{V}, \mathbf{E}, \mathbf{X})$, Adjacency matrix $\mathbf{A}$, Initial labels $\mathbf{Y}^{(0)}$, Number of SGCN layers $\mathbf{L}$, Fusion factor λ
Ensure: Predicted labels $\mathbf{Y}_U$
Initialize node embeddings $\mathbf{H}^{(0)} = \mathbf{X}$
Initialize transformation matrices $\mathbf{W}^{(0)}$, $\mathbf{W}^{(1)}$, ... ,$\mathbf{W}^{(L-1)}$
Compute $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$,$\mathbf{D}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$
Compute $\hat{\mathbf{A}} = \mathbf{D}^{-1/2}\tilde{\mathbf{A}}\mathbf{D}^{-1/2}$
for $l = 1$ to L **do**
    Calculate similarity matrix $\mathbf{S}^{(l-1)}$ using Equation (12)
    $\hat{\mathbf{S}}^{(l-1)} = \oplus\mathbf{S}^{(l-1)}$
    $\mathbf{Z}^{(l)} = \sigma(\lambda\hat{\mathbf{A}}\mathbf{H}^{(l-1)}\mathbf{W}^{(l-1)} + (1-\lambda)\hat{\mathbf{S}}^{(l-1)})$
    $\mathbf{H}^{(l)} = \mathbf{Z}^{(l)}$
end for
Train the model using labeled nodes $\mathbf{L}$:
    $\mathcal{L}_0 = \sum_{i\in\mathcal{L}} \text{loss}(\mathbf{H}_i^{(L)}, \mathbf{Y}_i^{(0)})$
Predict labels for unlabeled nodes $\mathbf{U}$:
    $\mathbf{Y}_U = \text{softmax}(\mathbf{H}_U^{(L)})$
Return $\mathbf{Y}_U$

This work construct two graph convolutional models for semi-supervised node classification by combining the proposed SGCN with the traditional GCN. In addition to utilizing graph structural information to obtain neighbor aggregation weights, SGCN also takes into account the attribute information of similarity between nodes. The similarity measure constructed using the Minkowski distance differs from cosine similarity in that it more precisely captures the closeness between node features, focusing on the similarity of feature content rather than spatial proximity. It is straightforward to envision that, in the context of semi-supervised node classification, nodes with similar features are more likely to belong to the same category. From this perspective, the rationale behind the approach becomes evident.

D. Analysis of Algorithm Complexity

Assuming L is the number of graph convolutional layers, N is the number of nodes, $\|A\|_0$ is the number of non-zero elements in the adjacency matrix A, and F is the number of node features. The time complexity of training a traditional GCN model for semi-supervised classification is $O(L \parallel A \parallel_0 F+LNF^2)$ and the space complexity is $O(LNF+LF^2)$ [30]. For the L-layer SGCN this study designed, the time complexity required for training is $O(2L \parallel A \parallel_0 F + LNF^2)$, and the space complexity is $O(2LNF + LF^2)$. It is evident that the proposed scheme can learn better node feature representations while maintaining the same order of magnitude of time and space complexity as the classical GCN. Taking the Cora dataset as an example and keeping consistent with the experimental setup detailed in Section 5, this work utilized a two-layer GCN architecture to perform semi-supervised node classification over 300 epochs. The training time for the GCN model was 8.21 seconds, while the proposed enhanced SGCN model took 9.08 seconds. The slight increase in time consumption of the proposed scheme is mainly attributed to the calculation of the similarity matrix.

V. Experiment and Result Analysis

To validate the effectiveness of the proposed SGCN in graph node representation learning, this study leverages the two graph convolutional neural network models outlined in the preceding section to evaluate their performance on semi-supervised node classification tasks. In this section, the experimental datasets, benchmark models, and experimental parameter settings are first introduced. Then, the experimental results of different models are analyzed. Finally, the impact of different metric distances, namely Manhattan distance, Euclidean distance, and Chebyshev distance, as well as varying fusion factors, on the performance of the two constructed semi-supervised node classification models is explored.

A. Datasets and Baseline Methods

This work evaluate the performance of the two proposed models on three benchmark citation network datasets: Cora, Citeseer, and PubMed [31]. These citation network datasets are structured as graphs with papers represented as nodes and citations between papers as edges. The specific details of these datasets are presented in Table I. Taking the Cora dataset as an example, it contains 2708 graph nodes and 5429 edges, with 7 domain categories represented by 7-dimensional one-hot vectors. Each paper is described by a 1433-dimensional one-hot feature vector, where each dimension takes a value of 0 or 1, indicating whether the corresponding word appears in the paper. The label rate in the dataset is 5.2%, meaning that 94.8% of the data is unlabeled.

TABLE I. The Details of Citation Network Datasets

Dataset	Nodes	Edges	Classes	Features	Labeled rate
Cora	2708	5429	7	1433	5.2%
Citeseer	3327	4732	6	3703	3.6%
Pubmed	19717	44338	3	500	0.3%

To validate the effectiveness and advancement of the proposed models for semi-supervised node classification tasks, this study compare them with the following benchmark methods: Multi-Layer Perceptron (MLP) [32], Label Propagation (LP) [33], Semi-supervised Embedding (SemiEmb) [34], Manifold Regularization (ManiReg) [35], Iterative Classification Algorithm (ICA) [36], DeepWalk [37], Planetoid [38], GCN [19], GraphSAGE [11] and GAT [10].

B. Implementation Details

Experimental setup was equipped with an 11th Gen Intel® Core™ i7-1165G7 processor, operating at 2.80 GHz and accompanied by 16GB of RAM. this work developed the SGCN-GCN and SGCN-SGCN models utilizing Python 3.9.10, PyTorch 2.2.1, and torch-geometric 2.5.3, all within the Windows 10 environment. The models were evaluated on three distinct datasets, employing uniform hyperparameter configurations: an L2 regularization factor of 5×10^{-4}, a hidden layer size of 128

units, the Adam optimizer, 300 training epochs, and a learning rate set to 0.01. The reported experimental outcomes reflect the mean classification accuracy across 10 distinct model parameter initializations, ensuring the reliability and robustness of these findings.

C. Experimental Results and Analysis of Different Models

Table II reports the average accuracy of different semi-supervised node classification models, with the highest and second-highest accuracies highlighted in bold. In addition to the experimental results on classification accuracy for MLP, GCN, GraphSAGE, GAT, and the two models proposed in this study, the results for all other models are derived from their original publications. As can be seen from the table, the two proposed models demonstrate superior performance compared to other models on all three datasets. This is attributed to the proposed SGCN, which considers not only the structural information of the graph but also the similarity information between nodes during information aggregation. Unlike the cosine similarity in mathematical space, a linear similarity measure using the Minkowski distance was designed, which better describes the proximity relationship between node features.

TABLE II. CLASSIFICATION ACCURACY (%) OF VARIOUS MODELS

Model	Cora	Citeseer	Pubmed
MLP	57.6	58.9	72.9
ManiReg	59.5	60.1	70.7
SemiEmb	59.0	59.6	71.1
LP	68.0	45.3	63.0
DeepWalk	67.2	43.2	65.3
ICA	75.1	69.1	73.9
Planetoid	75.7	64.7	73.9
GCN	80.6	68.7	78.7
GraphSAGE	80.0	62.4	76.0
GAT	**81.3**	68.5	77.2
SGCN-GCN	**81.5**	**69.5**	**79.5**
SGCN-SGCN	**81.3**	**68.9**	**79.9**

In contrast, other methods have limitations: MLP only uses node attribute information without considering the graph's structural information; ManiReg relies heavily on the data structure within local neighborhoods and can easily overlook global information; SemiEmb has high requirements for labels; DeepWalk cannot model attribute information; Planetoid suffers from information loss in the graph structure during random sampling; GCN, GraphSAGE, and GAT adopt a neighborhood aggregation scheme to improve performance by mixing the features of nodes and their neighbors, but they do not fully utilize the attribute information of the nodes themselves during the aggregation process. The SimP-GCN model proposed in [12] achieves the preservation of node similarity and the adaptive integration of structural and similarity information through the K-Nearest Neighbors (KNN) algorithm, a Node Similarity Preserving Aggregation module, and a Self-Supervised Learning module. In comparison, the scheme proposed in this study is more straightforward and efficient, has a reduced number of parameters, and exhibits stronger generalization capabilities. Since the code provided for constructing the SimP-GCN model in the original text did not utilize the same torch-geometric package as the one used in this study for building graph neural networks, significant errors occurred when attempting to reproduce the results. Consequently, the relevant experimental results for this model are not reported in this work.

The experimental results of the proposed SGCN-GCN and SGCN-SGCN compared to the classical GCN are presented in Fig. 2 below. Specifically, the SGCN-GCN achieves improvements of 0.9%, 0.8%, and 0.8% in accuracy over the traditional GCN on the three datasets, respectively, while SGCN-SGCN achieves improvements of 0.7%, 0.2%, and 1.2%, respectively.

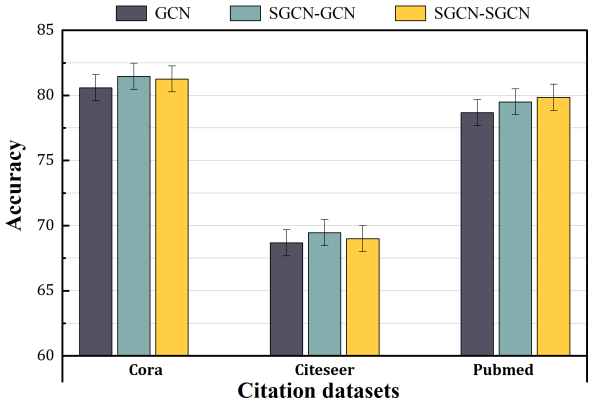


Fig. 2. Comparison chart with classical GCN classification accuracy.

D. Experimental Results and Analysis at Different Distances

This work report the classification accuracy of SGCN-GCN and SGCN-SGCN under different metric distances on the Cora, Citeseer, and Pubmed datasets in Fig. 3 and Fig. 4, respectively. The horizontal axis represents the tested datasets, while the vertical axis represents the classification accuracy, which is the average of ten experimental results.

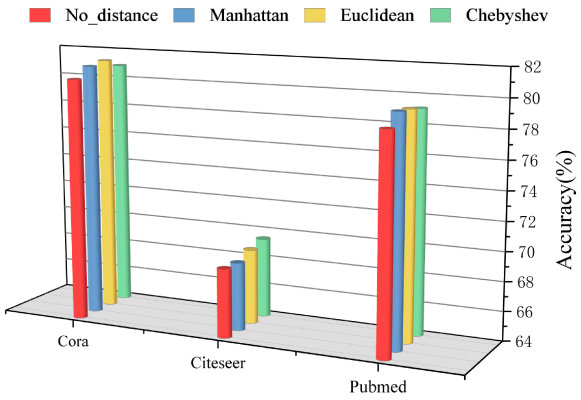


Fig. 3. Comparison of accuracy of SGCN-GCN at different metric distances.

The analysis of the figures reveals that both models perform best using Manhattan distance and Euclidean distance on the Cora and Pubmed datasets, while the Chebyshev distance yields the best results on the Citeseer dataset. This can be primarily attributed to the fact that the node feature dimensions in the Cora and Pubmed networks are relatively low (1433 and

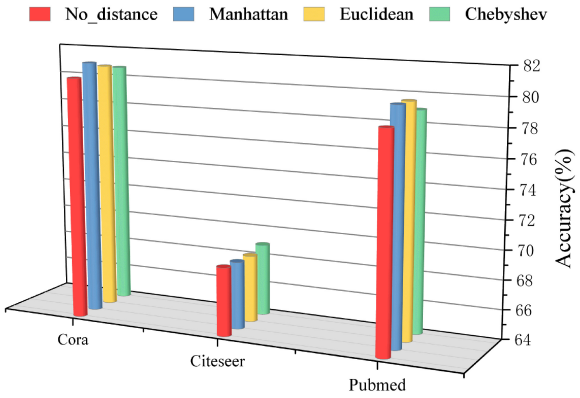


Fig. 4. Comparison of accuracy of SGCN-SGCN at different metric distances.

500, respectively), whereas the Citeseer network has a higher dimension (3703). For nodes with lower-dimensional features, which tend to be denser, using Manhattan and Euclidean distances to calculate similarity can fully utilize each feature value. In contrast, for nodes with higher-dimensional features, which are relatively sparse, using the Chebyshev distance to calculate similarity can better avoid interference from invalid feature values.

E. Experimental Results and Analysis of Different Fusion Factors

This work reports the variation in classification accuracy of SGCN-GCN across various fusion factors on the Cora, Citeseer, and Pubmed datasets, with Fig. 5, Fig. 6, and Fig. 7 depicting the performance under Manhattan distance, Euclidean distance, and Chebyshev distance, respectively. The horizontal axis represents the fusion factor, while the vertical axis indicates the classification accuracy, which is the average of ten experimental results.

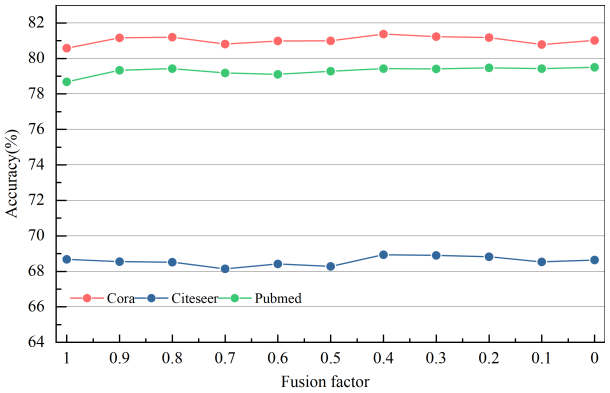


Fig. 5. SGCN-GCN classification accuracy variation map at Manhattan distance.

The analysis of the figures reveals that the SGCN-GCN model achieves better classification performance on the three datasets when the similarity information allocated to the first

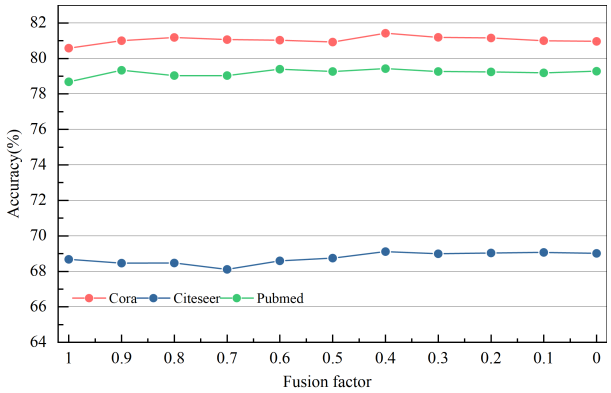


Fig. 6. SGCN-GCN classification accuracy variation map at Euclidean distance.

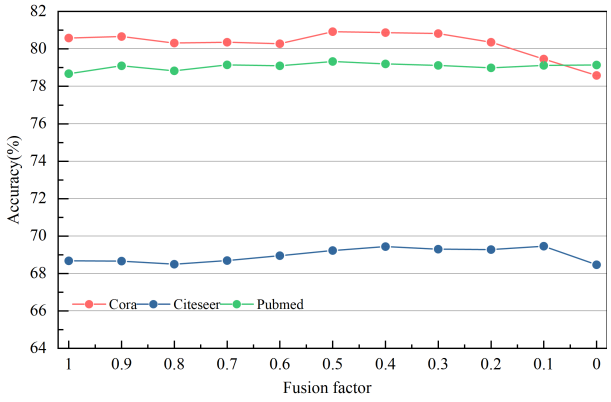


Fig. 7. SGCN-GCN classification accuracy variation map at Chebyshev distance.

layer of the graph convolutional network carries a larger weight. This is because only the first layer of the network utilizes similarity information for aggregation. Specifically, the model performs well when the fusion factor λ ranges from 0.1 to 0.5. When $\lambda = 1$ and $\lambda = 0$, the model solely relies on graph structural information (original GCN) and similarity information, respectively, for information aggregation. Notably, this study observes that the model using similarity information computed by Manhattan and Euclidean distances achieves better classification performance than the classical GCN when $\lambda = 0$ (i.e. using only similarity information). Irrespective of the value of the fusion factor λ, the model that integrates similarity information computed by Manhattan and Euclidean distances consistently demonstrates excellent classification results on the Cora and PubMed datasets. This verifies the validity of the approach to integrate similarity information into the original GCN framework.

This work reports the variation in classification accuracy of SGCN-SGCN across various fusion factors on the Cora, Citeseer, and Pubmed datasets, with Fig. 8, Fig. 9, and Fig. 10 depicting the performance under Manhattan distance, Euclidean distance, and Chebyshev distance, respectively. The horizontal axis represents the fusion factor, while the vertical axis indicates the classification accuracy, which is the average of ten experimental results.

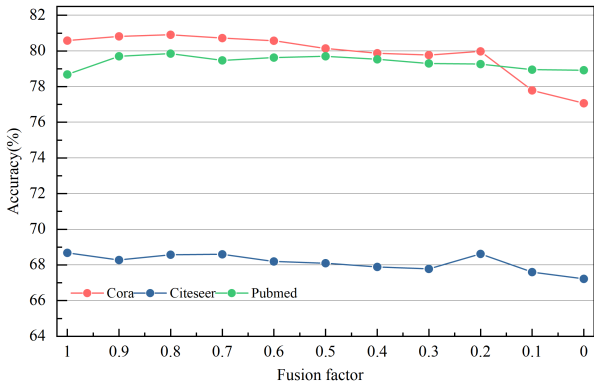


Fig. 8. SGCN-SGCN classification accuracy variation map at Manhattan distance.

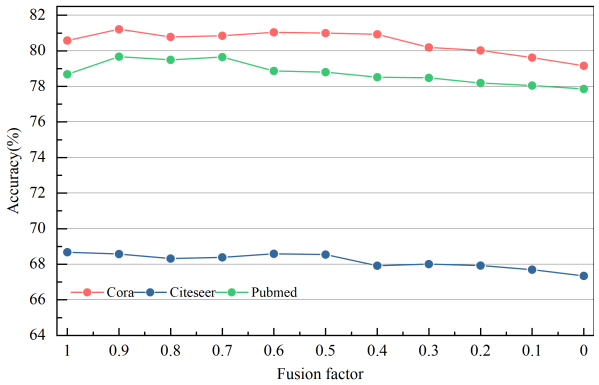


Fig. 9. SGCN-SGCN classification accuracy variation map at Euclidean distance.

The analysis of the figures indicates that the SGCN-SGCN model, which leverages similarity information for information aggregation in both layers of the network, achieves superior classification performance on the Cora and Pubmed datasets when each layer of the graph convolutional network is assigned a lower weight to the similarity information. Specifically, the model performs well when the fusion factor λ is greater than or equal to 0.6. When $\lambda = 1$ and $\lambda = 0$, the model relies solely on graph structural information (original GCN) and similarity information, respectively, for information aggregation. However, the model's performance on the Citeseer dataset is suboptimal, likely due to the sparsity of the feature vectors caused by its higher node feature dimension (3703).

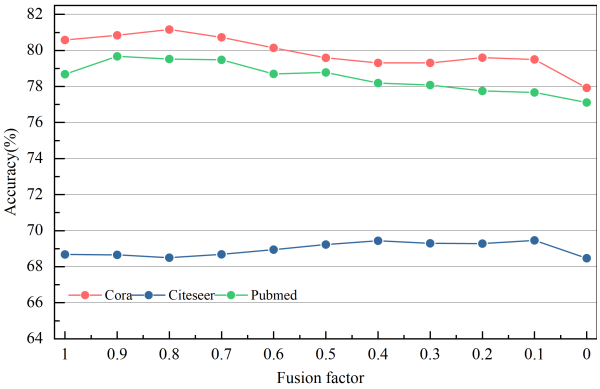


Fig. 10. SGCN-SGCN classification accuracy variation map at Chebyshev distance.

VI. Discussion

This study introduces SGCN, a novel graph convolutional network that integrates a linear similarity measure based on Minkowski distance with traditional graph structural information, addressing the limitations of conventional GCNs. Traditional GCNs typically overlook node attribute information and tend to disrupt the similarity information between nodes during information aggregation. Methods like GraphSAGE and GAT improve neighborhood aggregation but fail to fully capture node feature similarity and require the addition of supervised modules with more parameters and black-box characteristics.

Experimental results demonstrate that SGCN performs best when similarity information is weighted more heavily in the first layer, with the fusion factor between 0.1 and 0.5. This indicates that, while structural information plays a dominant role in the information propagation and aggregation of graph nodes, attribute information also provides a complementary, supportive function. Among the distance metrics tested, Manhattan and Euclidean distances consistently yield better results, suggesting they are more effective at capturing feature proximity in node classification tasks.

Future work may focus on extending SGCN to more specific graph data, such as functional brain networks, and implementing adaptive fusion factor adjustments to learn better node embedding representations, aiming to improve the recognition accuracy of brain diseases such as Alzheimer's and autism.

VII. Conclusion

Traditional Graph Convolutional Networks (GCNs) primarily rely on graph structure for node aggregation, often overlooking node attribute information. This approach can reduce node similarity and impede effective node feature learning, particularly in semi-supervised classification tasks. To address these issues, this study proposes SGCN, a novel Graph Convolutional Network. SGCN introduces a linear similarity measure based on Minkowski distance between node features, enhancing the description of similarity. By integrating this measure with traditional graph structure, SGCN improves node feature representation. Experimental validation on datasets such as Cora, Citeseer, and PubMed demonstrates that SGCN models (SGCN-GCN and SGCN-SGCN) significantly outperform traditional GCNs in node feature learning. The impact of metric distances and fusion factors on performance is also analyzed, providing insights for model optimization.

Acknowledgment

This project has received funding support from the Shaanxi Provincial Department of Science and Technology and the Xi'an Science and Technology Bureau, with project numbers 2024GX-YBXM-113 and 23GXFW0004, respectively.

REFERENCES

[1] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artificial Intelligence Review*, vol. 57, no. 4, p. 99, 2024.

[2] W. Ju, Z. Fang, Y. Gu, Z. Liu, Q. Long, Z. Qiao, Y. Qin, J. Shen, F. Sun, Z. Xiao *et al.*, "A comprehensive survey on deep graph representation learning," *Neural Networks*, vol. 173, p. 106207, 2024.

[3] Z. Chen, Y. Zhang, Y. Fang, Y. Geng, L. Guo, X. Chen, Q. Li, W. Zhang, J. Chen, Y. Zhu *et al.*, "Knowledge graphs meet multi-modal learning: A comprehensive survey," *arXiv preprint arXiv:2402.05391*, 2024.

[4] K. Zheng, S. Yu, and B. Chen, "Ci-gnn: A granger causality-inspired graph neural network for interpretable brain network-based psychiatric diagnosis," *Neural Networks*, vol. 172, p. 106147, 2024.

[5] W. Ju, Z. Liu, Y. Qin, B. Feng, C. Wang, Z. Guo, X. Luo, and M. Zhang, "Few-shot molecular property prediction via hierarchically structured learning on relation graphs," *Neural Networks*, vol. 163, pp. 122–131, 2023.

[6] Z. Song, X. Yang, Z. Xu, and I. King, "Graph-based semi-supervised learning: A comprehensive review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 11, pp. 8174–8194, 2022.

[7] Z. Kang, C. Peng, Q. Cheng, X. Liu, X. Peng, Z. Xu, and L. Tian, "Structured graph learning for clustering and semi-supervised classification," *Pattern Recognition*, vol. 110, p. 107627, 2021.

[8] W. Guo, B. Xue, Y. Hou, and Y. Hu, "Semi-supervised classification based on improved graph convolutional networks," *Journal of Shaanxi University of Science and Technology*, vol. 42, no. 05, pp. 191–197, 2024.

[9] Y. Zhang, Y. Zhang, D. Yan, Q. He, and Y. Yang, "Nie-gcn: Neighbor item embedding-aware graph convolutional network for recommendation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 5, pp. 2810–2821, 2024.

[10] C. Ding, S. Sun, and J. Zhao, "Mst-gat: A multimodal spatial–temporal graph attention network for time series anomaly detection," *Information Fusion*, vol. 89, pp. 527–536, 2023.

[11] T. Liu, A. Jiang, J. Zhou, M. Li, and H. K. Kwan, "Graphsage-based dynamic spatial–temporal graph convolutional network for traffic prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 10, pp. 11 210–11 224, 2023.

[12] W. Jin, T. Derr, Y. Wang, Y. Ma, Z. Liu, and J. Tang, "Node similarity preserving graph convolutional networks," in *Proceedings of the 14th ACM international conference on web search and data mining*, 2021, pp. 148–156.

[13] J. Yin and S. Sun, "Incomplete multi-view clustering with cosine similarity," *Pattern Recognition*, vol. 123, p. 108371, 2022.

[14] J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, and D. Tao, "A survey on self-supervised learning: Algorithms, applications, and future trends," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–20, 2024.

[15] K. Janani, S. Mohanrasu, A. Kashkynbayev, and R. Rakkiyappan, "Minkowski distance measure in fuzzy promethee for ensemble feature selection," *Mathematics and Computers in Simulation*, vol. 222, pp. 264–295, 2024.

[16] Z. Wu, X. Lin, Z. Lin, Z. Chen, Y. Bai, and S. Wang, "Interpretable graph convolutional network for multi-view semi-supervised learning," *IEEE Transactions on Multimedia*, vol. 25, pp. 8593–8606, 2023.

[17] J. Bruna, W. Zaremba, A. Szlam, and Y. LeCun, "Spectral networks and locally connected networks on graphs," *arXiv preprint arXiv:1312.6203*, 2013.

[18] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," *Advances in neural information processing systems*, vol. 29, 2016.

[19] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[20] K. Xu, W. Hu, J. Leskovec, and S. Jegelka, "How powerful are graph neural networks?" *arXiv preprint arXiv:1810.00826*, 2018.

[21] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *International conference on machine learning*. PMLR, 2017, pp. 1263–1272.

[22] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner *et al.*, "Relational inductive biases, deep learning, and graph networks," *arXiv preprint arXiv:1806.01261*, 2018.

[23] J. Chen, T. Ma, and C. Xiao, "Fastgcn: fast learning with graph convolutional networks via importance sampling," *arXiv preprint arXiv:1801.10247*, 2018.

[24] D. Zou, Z. Hu, Y. Wang, S. Jiang, Y. Sun, and Q. Gu, "Layer-dependent importance sampling for training deep and large graph convolutional networks," *Advances in neural information processing systems*, vol. 32, 2019.

[25] J. Zhang, X. Shi, J. Xie, H. Ma, I. King, and D.-Y. Yeung, "Gaan: Gated attention networks for learning on large and spatiotemporal graphs," *arXiv preprint arXiv:1803.07294*, 2018.

[26] Q. Li, Z. Han, and X.-M. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[27] R. Zhang, Y. Zou, and J. Ma, "Hyper-sagnn: a self-attention based graph neural network for hypergraphs," *arXiv preprint arXiv:1911.02613*, 2019.

[28] W. Ju, Z. Fang, Y. Gu, Z. Liu, Q. Long, Z. Qiao, Y. Qin, J. Shen, F. Sun, Z. Xiao *et al.*, "A comprehensive survey on deep graph representation learning," *Neural Networks*, vol. 173, p. 106207, 2024.

[29] E. Gwynne and J. Sung, "The minkowski content measure for the liouville quantum gravity metric," *The Annals of Probability*, vol. 52, no. 2, pp. 658–712, 2024.

[30] W.-L. Chiang, X. Liu, S. Si, Y. Li, S. Bengio, and C.-J. Hsieh, "Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 257–266.

[31] L. Zhang, R. Song, W. Tan, L. Ma, and W. Zhang, "Igcn: A provably informative gcn embedding for semi-supervised learning with extremely limited labels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–14, 2024.

[32] I. O. Tolstikhin, N. Houlsby, A. Kolesnikov, L. Beyer, X. Zhai, T. Unterthiner, J. Yung, A. Steiner, D. Keysers, J. Uszkoreit *et al.*, "Mlp-mixer: An all-mlp architecture for vision," *Advances in neural information processing systems*, vol. 34, pp. 24 261–24 272, 2021.

[33] I. B. El Kouni, W. Karoui, and L. B. Romdhane, "Node importance based label propagation algorithm for overlapping community detection in networks," *Expert Systems with Applications*, vol. 162, p. 113020, 2020.

[34] Y. Liu, Q. Wang, X. Wang, F. Zhang, L. Geng, J. Wu, and Z. Xiao, "Community enhanced graph convolutional networks," *Pattern Recognition Letters*, vol. 138, pp. 462–468, 2020.

[35] W. Liu, S. Fu, Y. Zhou, Z.-J. Zha, and L. Nie, "Human activity recognition by manifold regularization based dynamic graph convolutional networks," *Neurocomputing*, vol. 444, pp. 217–225, 2021.

[36] A. Tharwat, "Independent component analysis: An introduction," *Applied Computing and Informatics*, vol. 17, no. 2, pp. 222–249, 2021.

[37] J. Guo, H. Wen, W. Huang, and C. Yang, "A collaborative filtering recommendation algorithm based on deepwalk and self-attention," *International Journal of Computational Science and Engineering*, vol. 26, no. 3, pp. 296–304, 2023.

[38] S. Ershkov, D. Leshchenko, and A. Rachinskaya, "Dynamics of a small planetoid in newtonian gravity field of lagrangian configuration of three primaries," *Archive of Applied Mechanics*, vol. 93, no. 10, pp. 4031–4040, 2023.

Empowering Home Care: Utilizing IoT and Deep Learning for Intelligent Monitoring and Management of Chronic Diseases

Nouf Alabdulqader, Khaled Riad*, Badar Almarri
Computer Science Department, College of Computer Sciences & Information Technology,
King Faisal University, Al-Ahsa 31982, Saudi Arabia

***Abstract*—Integrating Internet of Things (IoT) with Artificial Intelligence (AI) is one of the catalysts for improving traditional healthcare services. This integration has created many opportunities that have led to healthcare shifting towards enabling home care, the concept that harnesses the technologies advanced potential such as the IoT and deep learning for intelligent monitor and manage chronic diseases. As population growth increases, restrictions on traditional healthcare services increase. Some diseases, such as chronic diseases, require innovative solutions that go beyond the boundaries of traditional healthcare settings due to their impact on individuals' health for example traditional healthcare systems have little capacity to provide high-quality and real-time services. Empowering home care services using deep learning and internet of things technology is promising. It enables continuous monitoring through interconnected devices and deep learning, which provides intelligent insights from massive data sets. This brief explores the key components of enabling home care, including continuous patient health monitoring, predictive analytics, medication management, and remote patient support by healthcare providers, and provides friendly interfaces for end-users. The conjunction between the IoT and deep learning in home-care signals a shift toward precision medicine, enhancing patient outcomes and creating a sustainable model for chronic disease management in the era of decentralized healthcare. This review article aims to discuss the following aspects: presenting the latest technologies in home care systems, showing the merit of combining the Internet of Medical Things (IoMT) and deep learning and its role in monitoring patient conditions and managing chronic disease to improve patient health status accurately, in real-time, and cost-effective, and lastly, debating future studies and providing recommendations for the ongoing development of home care remote monitoring applications.**

***Keywords*—IoT; IoMT; intelligent monitoring; chronic diseases; deep learning; home care; physiological data; mHealth**

I. Introduction

In the ever-evolving healthcare landscape, home care has emerged as a pivotal frontier, driven by technological advancements that aim to transform the management of chronic diseases. This paradigm shift is embodied in the concept of empowering home care, where the fusion of IoT and deep learning play central roles in leading a new area of intelligent monitoring and management for chronic conditions. Chronic diseases present substantial challenges to both individuals and healthcare systems, necessitating innovative solutions that provide personalized, continuous, and accessible care. Home healthcare services provide many benefits to the service providers and end-users by allowing service providers to monitor their patients outside the hospital and everywhere with high quality, and they can always connect to their patients. With the entrance of the IoT, the solution of home healthcare becomes proactive by using the term mHealth, which can be defined as "a communication technology integrated with mobile sensors for healthcare" [1]. The number of mHealth applications has increased over the years because of traditional monitoring methods often necessitate frequent visits to healthcare facilities, resulting in inconvenience and escalated healthcare expenses [2]. Moreover, these methods typically involve the use of multiple wearable sensors employing different algorithms, causing discomfort to patients and incurring exorbitant costs. The conventional approach to monitoring chronic diseases involved clinic visits and extracting blood samples from patients at labs. The physician would then evaluate the lab results to determine the patient's status. The disadvantages encompass high costs, prolonged monitoring periods, delayed responses from healthcare providers, potentially leading to patient fatalities, and a surge in the number of patients waiting in clinics [3], [4]. These drawbacks underscore the need for a shift towards real-time patient condition monitoring and exploring alternative methodologies. This review article will discuss three subjects: First, it will discuss recent technology in home care systems. Second, can integrating the IoT and deep learning help monitor, manage, and improve chronic disease patient conditions? Third, recommendations will be provided, and future studies for home healthcare technology will be explored.

A. Background

Due to the high importance of home care in managing and controlling chronic diseases, technology has become a main component of it, such as IoT and deep learning facilitating remote monitoring, management, and controlling the patient condition. Chronic diseases have become one of the specialties that cause significant challenges for healthcare, leading to a shift to decentralized healthcare. Integrating IoT and deep learning plays a significant role in home healthcare. IoT is highlighted as a main technology in home healthcare that allows service providers to perform real-time health patient monitoring through interconnected and wearable devices. Empowering home care leads to providing a holistic and patient-centric approach to chronic disease management, leveraging IoT and deep learning to empower patients, improve health conditions, enhance patient satisfaction, and create a more sus-

*Corresponding authors

tainable healthcare environment. The author in [5] introduces the potential advantage of using IoT in healthcare to monitor the daily activities of elderly people. The study discusses the heterogeneity and interoperability issues related to disregarding healthcare for IoT. This study proposed H3IoT architecture to support home care for the health monitoring of elderly people, focusing on its ease of use, mobility, and cost-effectiveness. The H3IoT can provide support in many aspects, such as protective and monitoring health care, and chronic disease management. The author in [6] explores the potential of using IoT and big data analytics for chronic disease monitoring in Saudi Arabia, with a specific focus on hypertension. It aims to develop a predictive system for detecting and managing chronic diseases by presenting a framework consisting of four modules: data collection, storage, processing, and analysis. Decision Tree (C4.5) and Support Vector Machine (SVM) models were employed to predict hypertension, identifying age and diabetes as significant contributing factors. The accuracy of the Support Vector Machine algorithm is 71.15%, while the accuracy of the C4.5 algorithm is 68.80%, highlighting the predictive capabilities of these models. The authors emphasize the transformative potential of IoT and big data analytics in enabling the early detection of chronic diseases and enhancing healthcare services in Saudi Arabia. The research concludes that while the proposed models are effective for disease prediction, further refinement and evaluation are necessary to optimize their performance. Ultimately, this review underscores the revolutionary impact of IoT and big data analytics in chronic disease monitoring and healthcare innovation. The author in [7] investigates the application of machine learning (ML) techniques to analyze chronic diseases using diagnosis codes from the CMS dataset. Focusing on clinical and claims data for 11 chronic diseases, the study aims to reduce the set of diagnosis codes and assess their relevance to healthcare decision-making. The experimental setup involves restructuring the data, applying attribute reduction techniques, and utilizing classification algorithms to derive insights. The study demonstrates that the reduced set of diagnosis codes provides valuable insights for informed healthcare decisions. For the training data, most of the 11 chronic diseases achieve an accuracy above 88%, with stroke and chronic kidney disease showing the highest accuracy. For the testing data, the accuracy for most chronic diseases ranges between 80-90%, with cancer and stroke achieving accuracies above 90%. In contrast, chronic heart failure and depression exhibit lower accuracies of 79.15% and 77.02%, respectively. These findings highlight the effectiveness of ML techniques in analyzing chronic diseases and their potential to support healthcare decision-making. The author in [8] discuss that healthcare landscape is rapidly evolving, with an increasing emphasis on delivering high-quality care within the comfort of patients' homes. This shift toward home-based care is driven by an aging population and the growing demand for patient-centered solutions. At the forefront of this transformation is the powerful synergy between the Internet of Things (IoT) and deep learning. IoT, with its network of interconnected devices and sensors, facilitates the real-time collection of data on various aspects of a patient's well-being. Deep learning, a subset of artificial intelligence, excels at extracting complex patterns from vast datasets, making it a critical tool in modern healthcare.

By applying deep learning algorithms to the extensive data collected by IoT devices, healthcare providers can unlock numerous benefits, including improved diagnostic accuracy and personalized care. For example, studies have demonstrated that for training data, most of the 11 chronic diseases analyzed achieve an accuracy above 88%, with stroke and chronic kidney disease showing the highest performance. For testing data, the accuracy for most chronic diseases ranges between 80% and 90%, with cancer and stroke achieving accuracies above 90%. However, chronic heart failure and depression exhibit lower accuracies of 79.15% and 77.02%, respectively. These lower accuracies are attributed to the variability in diagnostic tests for these conditions. This highlights the need for further advancements in data collection and analysis to ensure consistent accuracy across all conditions.

The author in [9] discusses the landscape of healthcare growth where emphasis on providing patient-centered care that is both effective and affordable. Within this landscape, home care emerges as a critical component, allowing individuals to receive care in familiar surroundings. The convergence of the IoT and deep learning presents a transformative opportunity to revolutionize home care, enhancing both the quality of care and the efficiency of service delivery. This study explores the application of Multi-Step Deep Q Learning Network for securing healthcare data within the IoT framework. This convergence of deep learning and IoT immense to transform home care, making it more proactive, personalized, and effective. The author in [10] discusses the prediction of heart disease using an IoT-based ThingSpeak framework and a deep learning approach, proposing a system that leverages IoT technology and deep learning to achieve real-time predictions. The integration of IoT sensors with advanced deep learning models enables the continuous monitoring and analysis of patient health data, providing early warnings and facilitating timely interventions. This approach underscores the transformative potential of combining IoT and deep learning in chronic disease management, offering significant improvements in patient health outcomes, reductions in healthcare costs, and enhancements in the overall quality of life for individuals with chronic conditions. The proposed CNN-based heart disease prediction system demonstrated remarkable performance, achieving an accuracy of 98.8% based on experimental results. Additionally, the Matthews Correlation Coefficient for the system was calculated at 0.9321, further validating its reliability and effectiveness. These findings highlight the potential of IoT and deep learning technologies to revolutionize chronic disease management, paving the way for more efficient, patient-centered healthcare solutions.

B. Objectives

The objectives discussed in the current studies are:

- Discuss the current studies that support an overall view of the current state of empowering home care for chronic disease patients and highlighting challenges and opportunities in home care.
- Explore IoT technologies for monitoring and managing chronic diseases in-home care using designed devices and sensors.

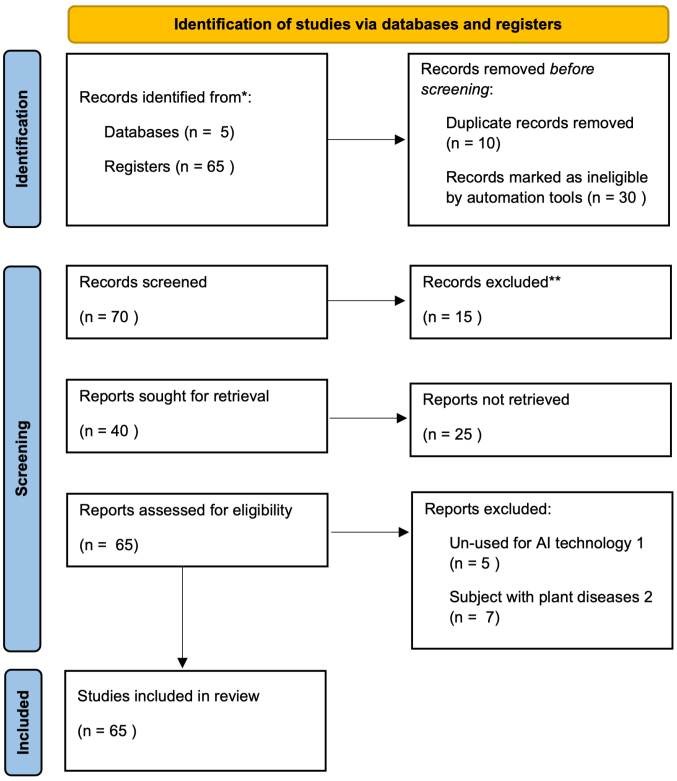


Fig. 1. PRISMA Methodology.

- Evaluate deep learning applications in-home care for intelligent monitoring and management of chronic diseases.
- Identify and analyze the best practices of utilizing deep learning and IoT models to empower home care for chronic diseases.
- Assesses technological challenges and limitations of deep learning and IoT in-home care.
- Provide recommendations and guidelines for activating the positive view of implementing deep learning and IoT and technologies in-home care.

C. Methodology

IoT and deep learning technology have emerged as transformative tools for home care, revolutionizing the way patients are monitored and managed remotely. The main goal of the review article is to find research papers focusing on utilizing IoT and deep learning in-home care and comparing them to achieve the review article's objectives using PRISMA to list of all extracted researches as mentioned in Fig. 1.

This section discusses the methodologies used in the review article as described below:

1) Literature Search Strategy:
   Conduct comprehensive literature research using electronic databases, including IEEE Xplore, Google Scholar, Research Gate, and ScienceDirect.
2) Inclusion and Exclusion Criteria:
   After filtering the search studies, we included published studies in conference papers, peer-reviewed journals, and literature reviews.
3) Research Summarization: The review article uses two main applications to summarize the research paper: SciSummery to summarize the whole paper and Jenni to find out the technologies used and methodologies.
4) Research Comparison: To achieve a review article's objectives, we used PRISMA in comparing research papers and to find out methodologies used, data collection, IoT sensors, differences, strengths, and weaknesses of research papers.
5) Future and Limitations: Potential limitations of the research papers such as the limitation in methodologies, the chronic disease cases that solution supports, and the accuracy that the research paper reaches.

The rest of this review article is structured as follows: The first section is about IoT in-home care that discusses the integration of IoT technologies in-home care, which revolutionizes health management, offering personalized monitoring and proactive care delivery to enhancing safety and independence. The second section is about deep learning in-home care. It discusses deep learning's vast potential to transform healthcare, particularly in-home care for chronic disease patients, by leveraging advanced algorithms to improve outcomes and reduce costs through remote monitoring and personalized care. The third section is about deep learning in health monitoring. It discusses how deep learning revolutionizes health monitoring and diagnosis, enhancing accuracy and efficiency in disease detection and treatment planning through analysis of vast medical data, improving predictive precision and remote patient monitoring capabilities. The fourth section discusses the intelligent management of chronic diseases. The fifth section discuss benefits and challenges of integrating IoT and deep learning technologies that enhance home care services by enabling interconnected devices to monitor, communicate, and analyze data, creating a safer and more efficient living environment. The sixth section is about the case study using the remote monitoring system of diabetes patients and a case study of predictive analytics for cardiovascular disease management. The last section is the conclusion of this review article. Below Table I is a list of acronyms designed to aid in comprehending technical terms and abbreviations utilized in this comprehensive review. The purpose of this reference table is to assist readers in deciphering the numerous acronyms and their definitions commonly referenced in the exploration of sophisticated topics concerning the IoT, deep learning, and healthcare technologies.

II. IoT in Home Care

In the realm of home care, the combination of IoT technologies is emerge as a transformative force, revolutionizing how individuals manage their health and well-being within the comfort of their homes. IoT-enabled devices, such as wearable health monitors and smart home appliances, provide unprecedented opportunities for remote monitoring, personalized care, and real-time health interventions. By seamlessly collecting and analyzing vital health data, including blood pressure, heart rate, and activity levels, IoT systems provide both patients and caregivers with actionable insights, enabling them to make

TABLE I. NOTATION-TABLE

Acronym	Full Form	Description
ADL	Activities of Daily Living	Tasks related to personal care and routine activities necessary for daily life.
AI	Artificial Intelligence	Simulation of human intelligence processes by machines, especially computers.
COPD	Chronic Obstructive Pulmonary Disease	A group of lung diseases that block airflow and make it difficult to breathe.
EHR	Electronic Health Record	Digital version of a patient's paper chart containing medical and treatment history.
IoT	Internet of Things	Network of physical devices embedded with sensors and software to exchange data.
ML	Machine Learning	A branch of AI focused on building systems that learn from data to make predictions or decisions.
SVM	Support Vector Machine	A supervised machine learning algorithm used for classification and regression analysis.
WHO	World Health Organization	A specialized agency of the United Nations responsible for international public health.
CNN	Convolution Neural Network	Type of deep learning algorithm specifically designed for processing structured grid data.
SSO	Sparrow Search Optimization	A nature-inspired optimization algorithm based on the foraging behavior of sparrows, used to solve complex optimization problems.
SSO	Salp Swarm Optimization	A bio-inspired optimization algorithm modeled after the swarming behavior of salp chains in the ocean, used for solving complex optimization problems.
VAE	Variational Autoencoder	A type of generative model in machine learning that learns to encode input data into a latent space and then decode it back, allowing for the generation of new, similar data.
ECG	Electrocardiogram	A medical test that records the electrical activity of the heart over a period of time.

informed decisions about health management and treatment adherence.

As the IoT ecosystem continues to evolve, fueled by advancements in IoT devices technology, data sciences, and connectivity infrastructure, its power to revolutionize home care delivery and promote independent living for individuals of all ages remains unparalleled. Furthermore, the integration of IoT in healthcare research provides vast datasets that lead to enhance patient outcomes and innovative strategies to meet individual patient needs. Overall, IoT technology fosters a more responsive, personalized, and cost-effective healthcare ecosystem [11]. This section will present IoT-enabled devices, data collection and integration, remote monitoring, alert systems, and deep learning for health monitoring.

A. IoT-Enabled Devices

This section will present IoT sensors utilized by research papers in chronic diseases. The below Table II presents sensors used with their definitions:

TABLE II. IOT SENSORS UTILIZED IN THE RELATED LITERATURE

Ref. No.	Sensors
[5]	Acceleration sensor Motion sensor mathElectroencephalogram (EEG), Electromyogram (EMG)
[12]	Camera Sensor
[13]	accelerometers and gyroscopes sensors , environmental sensors
[14]	accelerometers Sensor gyroscopes Sensor Push button 5 MP camera, voice reminder Jetson TX2
[15]	Temperature sensors Heart rate sensors
[16]	wearable sensor
[17]	Temperature sensors Blood pressure sensors Heart rate monitors Glucose level sensors Oximeters Accelerometers Gyroscopes ECG sensors EEG sensors Respiratory rate sensors Humidity sensors Wearable fitness trackers
[6]	Temperature sensors Blood pressure sensors Heart rate monitors Oxygen level sensors Glucose level sensors Respiratory sensors Electrocardiogram monitors Electroencephalogram sensors Motion sensors Wearable fitness trackers Sleep pattern monitors
[18]	smart wristband
[7]	Nan
[19]	Temperature - Blood pressure Blood pressure
[20]	Temperature Blood Humidity Light intensity Soil moisture
[10]	Heart rate Blood pressure ECG Oxygen saturation Temperature

B. Data Collection and Integration

This section will demonstrate the data collection used to monitor and predict chronic diseases.

The author in [5] discusses that data collection is a broader context for healthcare. It suggests that data collection is happening through traditional channels such as hospitals, community healthcare settings, and public health organizations, but also increasingly through new methods like wearable devices and social media analytics. The adoption of Electronic Health Records (EHR) has been instrumental in improving data collection and consistency. The paper also mentions challenges related to the storage and access of large volumes of data and the need for a concerted effort to enable various professionals to access these datasets.

The author in [6] discusses the methodology used in data collection and integration in the form of utilizing IoT sensors in health care. It typically involves several steps, including

Data Collection: This involves using sensors and IoT devices to collect various types of health-related data, such as vital signs (temperature, heart rate, blood pressure, etc.), patient activity, or specific medical parameters.

The author in [7] presents the context of healthcare and especially in studies dealing with diagnosing chronic diseases using ML techniques or integrating IoT technologies into healthcare. Data collection and integration generally go through several steps, starting with data collection and end with feature selection.

The author in [8] discusses the broad strokes of using IoT and deep learning in healthcare. Unfortunately, it doesn't dive into the specifics of data collection and integration methodologies.

The author in [9] emphasizes the critical importance of the security aspects in a healthcare IoT system using deep learning. It does not delve into the specifics of data collection and integration methodologies, as the primary focus is on security. The author assumes data are collected from various devices to processing and analysis in the main server. However, it does not elaborate on the protocols, communication technologies, or data integration techniques employed.

C. Remote Monitoring and Alert Systems

Remote monitoring and alert systems in the medical field represent a significant evolution in healthcare management, leveraging technology to oversee patient health outside of traditional hospital environments. These systems utilize a different of sensors and devices to collect patient information continuously, which is then transmitted to healthcare professionals for analysis and response. The ability to monitor patients in real-time facilitates early detection of potential health issues, enables prompt interventions, and improves the overall chronic diseases conditions, elderly patient care, and post-operative recovery.

Such systems can operate over various networks, including wireless body area networks and personal area networks, and can even employ non-contact methods such as cameras or smart devices, as shown in Fig. 2. This technology not only improves patient outcomes by enabling constant care but also enhances the adequacy of healthcare services by reducing hospital admissions and costs. Medical staff are empowered to focus more on data interpretation and patient care decisions rather than on repetitive monitoring tasks [21].

ML plays a vital role in these systems by analyzing the data collected, recognizing patterns, and predicting outcomes, thus providing pivotal support in decision-making processes. The integration of ML and cloud computing also ensures that massive volumes of medical data are stored and processed effectively, promoting continuous advancements in healthcare delivery.

The author in [5] presents the architecture of H3IoT aims to provide an effective environment for monitoring the health status of elderly people at home, offering advantages such as mobility, affordability, user-friendliness, and tolerance for delays in data transmission. The system would also include an alert mechanism which automatically sends notifications to designated individuals or emergency services if the data

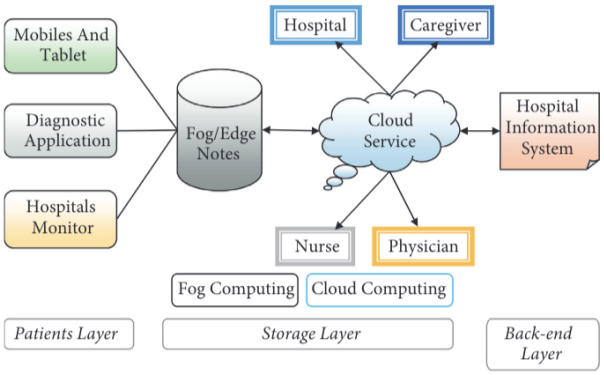


FIGURE 1: General architecture of RPM.

Fig. 2. General architecture of RPM [21].

TABLE III. RESEARCH PAPERS ACHIEVING ARTICLE OBJECTIVES

References	Achieved Objectives
[22] [5] [14] [6] [8] [9] [23] [20] [10]	1- Explore IoT technologies for monitoring and managing chronic diseases in-home care using designed devices and sensors 2- Identify and analyze best practices utilizing IoT and deep learning technologies to empower home care for chronic diseases.
[7]	Does not fulfill the article's objective related to the utilization of IoT sensors

indicates a cause for concern, such as a fall or a serious deviation in health signals. However, the framework may need additional modifications and enhancements to support critical emergency healthcare scenarios.

The author in [6] presents a combination of IoT devices for data collection and big data analytics for processing and analyzing the collected data to monitor chronic diseases such as hypertension. The methodology aims to predict and detect health issues, thereby enabling early interventions and improved patient outcomes. After reviewing the IoT sensors utilized in the research papers, it was found that not all studies incorporated IoT sensors in-home care, which does not meet the objectives of this article. Table III illustrates these findings.

III. DEEP LEARNING FOR HEALTH MONITORING

Deep learning has revolutionized the ability to monitor, diagnose, and manage health conditions in the medical field. In the era of digital healthcare transformation, deep learning approaches are applied to parse through massive volumes of medical data.

The introduction of deep learning algorithms into health monitoring systems has significantly improved the precision of predicting patient outcomes and has strengthened the capabilities of remote patient monitoring. By analyzing intricate patterns in data that are often imperceptible to human clinicians, deep learning models can identify potential health issues much earlier than traditional methods, thus aiding in proactive medical interventions.

Additionally, these advanced AI technologies enable personalized medicine, where treatment plans are tailored to

the individual characteristics of each patient's condition that leading to better patient health management and can improve patient in home-care. Overall, deep learning in health monitoring stands as a cornerstone in the leap towards more intelligent, responsive, and effective healthcare systems [24].

A. Deep Learning Techniques

The author in [5] presents the architectural concept of H3IoT for home environment monitoring of elderly individuals' health and does not delve into specific deep learning algorithms. Unfortunately, the authors in [6] do not mention the use of a deep learning algorithm for its methodology. But it specifically references the use data mining techniques with the Hadoop framework. For classification and prediction, it cites the use of a decision tree algorithm and SVM algorithm. However, there is another source titled "multi-disease prediction using LSTM recurrent neural (RNN) networks," which does mention the use of a deep learning approach, namely, Long Short-Term Memory RNN, for multi-disease prediction based on patients' clinical records. This approach is distinct from the method used in the first paper and is specifically designed for handling the temporal irregularity across clinical visits and determining the importance of each visit for the prediction task.

Furthermore, the author in [7] does not specify the use of a deep learning algorithm for its research. Instead, it discusses the utilization of ML techniques, such as attribute reduction techniques and classification algorithms, to analyze healthcare data and obtain a reduced set of diagnosis codes for chronic diseases. For deep learning applications within a similar context, the study multi-disease prediction using LSTM and RNN is relevant.

B. Sensor Data Analysis and Processing

Sensor data analysis and processing in the medical field is a crucial aspect of modern healthcare that leverages technology to improve patient outcomes and streamline healthcare services. By introducing IoT, the volume and variety of healthcare data have grown exponentially, presenting both opportunities and challenges. The vast volume of data produced by these devices necessitates advanced methods for processing and analysis to derive meaningful insights. Data science plays a pivotal role in this context, providing the tools and techniques necessary for managing, analyzing, and assimilating large quantities of both structured and unstructured healthcare data.

Medical sensor data analysis involves collecting, processing, and interpreting health data generated by these devices to monitor patient health, detect anomalies, and predict health events. This process requires sophisticated algorithms and big data techniques to handle large volumes of structured and unstructured data. The progression of data science and ML has facilitated the development of predictive models that support early disease detection and the customization of treatment plans [25]. Sensor data must be handled with care to ensure accuracy, reliability, and privacy. It involves pre-processing, data cleansing and anonymization, by using secure platforms to store and process data. Data science provides managements tools to manage patients health status data to extract meaningful information, which can be used in making important decisions about patient care. Effective data management and analysis can yield factual results crucial for proactive health monitoring, specific treatment plans, and instantly detecting of potential health issues. The process typically involves data cleansing, data mining, preparation, and analysis. Successfully integrating sensor data analysis and processing in healthcare can lead to groundbreaking advancements for patient monitoring remotely that lead to improve chronic disease management. The sensor data analysis and processing mentioned by the author in [5] could include several steps:

- Data Cleaning and Pre-processing
- Data Integration
- Real-Time Monitoring
- Pattern Recognition
- Predictive Analysis

If the study includes an IoT framework for health monitoring, it is likely that any or all of these data analysis and processing techniques could be employed within that framework.

Data analysis and processing, as proposed by the author in [6], are achieved through a integration of big data technologies and ML algorithms, specifically:

- Big Data Processing
- Data Storage and Management
- Data Analysis Integration

The purpose of these techniques is to process and analyze the collected data so that predictive models can be developed and applied, which are then able to identify potential cases of hypertension and possibly other chronic diseases in the data set. These models serve as a framework for early detection and ongoing monitoring of patient health, which could lead to better management and treatment outcomes.

Furthermore, the sensor data analysis and processing mentioned by the author in [7] focuses on the analysis and processing of clinical and claims data to study 11 chronic diseases. While the specific details of the data analysis and processing steps are not provided in the excerpts.

C. Disease Prediction and Risk Assessment

The use of IoT and AI in disease prediction represents a significant advancement in healthcare.

This data can include heart rate, blood pressure, glucose levels, and more, depending on the sensors' capabilities. AI comes into play by analyzing this massive stream of data to identify patterns and anomalies that might be indicative of health issues. Techniques from AI, like ML and deep learning models trained using vast datasets to learn from past examples. This training allows AI systems to predict potential diseases or health risks before they become critical, offering a chance for early intervention [26]. IoT provides the real-time data that AI models require to be accurate and effective, while AI gives the tools to make sense of the IoT-collected data, providing

TABLE IV. RESEARCH PAPERS ACHIEVING ARTICLE OBJECTIVES IN DEEP LEARNING

References	Achieved Objectives
[5]	1- Evaluate deep learning applications in-home care for intelligent monitoring and management of chronic diseases
[8] [9] [10]	2- Identify and analyze best practices utilizing IoT and deep learning technologies to empower home care for chronic diseases.
[6] [7]	Does not fulfill the article's objective related to the utilization of deep learning algorithms

insights that healthcare professionals can use to diagnose and treat patients more effectively. The integration of IoT and AI in healthcare has the potential to transform patient care by improving disease prediction, enhancing remote monitoring, and personalizing treatment plans.

The disease prediction and risk assessment presents by the author in [6] is achieved by developing a framework that utilizes decision tree and SVM techniques. The goal is to predict hypertension by analyzing health data collected from IoT devices.

The disease prediction and risk assessment presents by the author in [7] discusses by the following steps:

- Data Collection: Collecting patient data which may include demographics, clinical history, lab results, medication, etc.
- Data Preprocessing: Cleaning and structuring the data to be suitable for machine learning models.
- Feature Selection: Determining which data points (or "features") are most relevant to predicting a specific disease or risk level.
- Model Development: Using statistical or ML algorithms to create a model that can predict disease likelihood or risk based on the features.
- Validation and Testing: Testing the model against a dataset that was not used during the training to evaluate its performance and accuracy.
- Risk Assessment: The model's output may be a probability of disease presence or progression, which can be used to assess risk levels for patients.

At the end of this section, we conclude that not all research papers employed deep learning in home care to monitor chronic diseases. Table IV indicates which research studies meet the article's objectives and which do not.

IV. INTELLIGENT MANAGEMENT OF CHRONIC DISEASES

The intelligent management of chronic diseases involves employing advanced technologies and systems to enhance the care and outcomes for those with long-term health conditions. This approach as shown in Fig. 3, includes the use of personal health systems, connected health solutions, and self-management techniques. These are designed to empower patients to take an active role in managing their own care in partnership with healthcare providers. Through connected health technologies, such as smart biosensors, remote monitoring, and data analytics, patients and health professionals

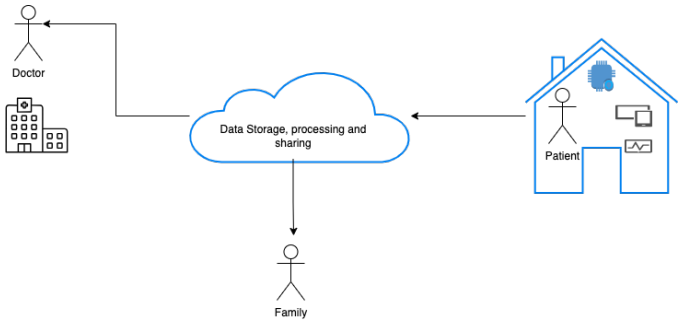


Fig. 3. The health ecosystem.

can better track health status and adjust treatments as needed. Digital tools support lifestyle changes, help manage symptoms, and can provide alerts for timely interventions, thus avoiding exacerbations and hospitalizations [27]. Self-management is a complementary concept where patients with chronic illnesses like diabetes, heart disease, and arthritis learn and apply skills to deal with their conditions effectively. This approach includes monitoring symptoms, managing medications, maintaining diet, exercise regimes, and adapting psychologically and socially. The intelligent management of chronic diseases aims to provide personalized care, reduce healthcare costs, and improve the overall quality of life for patients. It represents a strategic shift towards more proactive and patient-centered healthcare models.

A. Personalized Treatment Plans

The practice of medicine is evolving, shifting its focus from simply treating disease events to enhancing health, preventing diseases, and personalizing care to meet individual health needs. This transformation is driven by the concept of personalized health care, which utilizes personalized health planning empowered by personalized medicine tools. These tools leverage advances in science and technology to predict health risks, understand disease development dynamics, and tailor therapeutic approaches to individual needs [28]. Personalized medicine plays a crucial role in the management of chronic diseases. By implementing personalized health care and utilizing personalized medicine tools, healthcare providers can create personalized treatment plans for individuals with chronic diseases. These plans are listed to each patient's specific needs, taking into account their genetics, lifestyle factors, and response to previous treatments. This personalized approach in chronic disease management helps optimize patient health status outcomes and enhance the quality of health care.

B. Behavioral Analysis and Intervention

Chronic diseases, such as diabetes, heart disease, and hypertension, are a growing concern worldwide. These conditions require long-term management and intervention to prevent complications and improve the quality of life for patients. Traditional approaches to managing chronic diseases focus primarily on medical interventions, such as medication management and regular check-ups. However, research has shown that behavioral factors play a significant role in the development and progression of chronic diseases [29]. For instance, unhealthy lifestyle choices, such as poor diet, lack of physical

activity, smoking, and excessive alcohol consumption, can increase the risk of developing chronic diseases and exacerbate existing conditions. In order to effectively manage chronic diseases, an intelligent approach that incorporates behavioral analysis and intervention is needed. The advanced technologies and data science in healthcare professionals can gather and analyze data on patients' behavior patterns, including their daily routines, dietary habits, exercise habits, medication adherence, and stress levels.

One approach to behavioral intervention is through the use of digital health technologies, such as mobile applications and wearable devices. These tools can provide real-time feedback and reminders to help patients adhere to their medication regimens, track their daily activities, and monitor their progress toward health goals, such as exercise and diet modifications. Integrating behavioral analysis and intervention into the management of chronic diseases has the potential to improve patient outcomes significantly. By addressing the behavioral factors that contribute to chronic diseases, healthcare providers can help patients make positive lifestyle changes and improve their overall health and well-being.

C. Adherence Monitoring and Medication Management

Chronic diseases, like cardiovascular disease, Parkinson's, diabetes, blood pressure, and hypertension, have emerged as major health concerns globally. These diseases have a main impact on the health life quality and can lead to severe complications if not managed properly. According to a study conducted by the Boston Consulting Group (BCG) in 1993, disease management emerged as an innovative approach to improving the quality of care and containing rising healthcare costs associated with chronic diseases [30]. The concept of disease management involves a holistic and integrated approach to patient care, focusing on monitoring adherence to treatment plans and effectively managing medication. This approach aims to optimize patient outcomes by ensuring that patients adhere to their prescribed medications and treatment regimens. One study conducted in Saudi Arabia highlighted the importance of integrating chronic disease management to improve patient care. Another study emphasized the role of big data analytics in improving chronic disease management and healthcare systems. The integration of disease management strategies, such as adherence monitoring and medication management, can greatly improve the management of chronic diseases [6].

V. Benefits and Challenges

The integration of IoT devices and deep learning technologies into home care presents a transformative approach to managing chronic diseases. As the global healthcare landscape shifts towards more patient-centered and technologically advanced solutions, the potential benefits of these innovations are becoming increasingly apparent. IoT devices enable continuous, real-time monitoring of patients, providing valuable data that can be analyzed by deep learning algorithms to offer personalized care, predict complications, and enhance overall health outcomes. This intelligent monitoring and management system promises to improve patient engagement.

However, the implementation of IoT and deep learning in home care also brings significant challenges. These include technological barriers such as integration issues and data security, as well as broader concerns about accessibility, equity, and reliability. Furthermore, ensuring that both patients and hospital providers are adequately trained to use these technologies effectively is crucial. Additionally, regulatory and ethical considerations must be addressed.

In this context, understanding these aspects can guide the development and implementation of more effective and equitable healthcare solutions, ultimately improving the quality of life for chronic disease patients.

A. Benefits of IoT and Deep Learning in Home Care

The integration of the IoT and deep learning technologies showcases the transformative potential for enhancing home care services. IoT in home care refers to the network of interconnected devices capable of monitoring, communicating, and analyzing data to facilitate a more comfortable, efficient, and safe living environment. Deep learning is a subset of AI focusing on sophisticated algorithms that mimic human brain functions, further leverages this data to extract meaningful insights and patterns for improved decision-making and automation. The amalgamation of these two technologies in-home care can lead to the development of smart homes that are not only responsive to the inhabitants' needs but also proactive in maintaining their health and well-being. Smart devices can track vital signs, detect anomalies, and alert caregivers or medical professionals in real time. Deep learning algorithms can process vast quantities of data from various sensors to learn normal behavior patterns and anticipate potential issues, offering not just reactive but preventive solutions for home care challenges [31]. Together, IoT and Deep learning contribute to a more personalized and adaptive home care experience, enabling individuals.

The benefits below collectively contribute to the transformation of home care, making it more intelligent, patient-centric, and efficient through the integration of IoT and deep learning technologies:

- Real-Time Monitoring: IoT devices enable continuous and real-time monitoring of patient's health parameters, providing instant feedback to healthcare providers for timely interventions [32].
- Early Detection of Health Changes: Deep learning algorithms analyze collected data, facilitating the early detection of subtle changes in health conditions, allowing for proactive healthcare measures [33].
- Personalized Care Plans: Deep learning models can process individual health data to create personalized care plans, considering specific patient needs, preferences, and historical health information [34].
- Patient Empowerment and Engagement: Access to personal health data empowers patients, fostering active engagement in their healthcare journey and promoting a sense of responsibility for their well-being [34].
- Cost-Efficiency in Healthcare: By preventing complications through early interventions and minimizing

hospital visits, the implementation of IoT and deep learning can lead to cost savings in the healthcare system [35].

- Data-Driven Decision Making: Deep learning algorithms process vast amounts of patient data to generate meaningful insights, aiding healthcare professionals in making informed and data-driven decisions [35].
- Improved Disease Management: Chronic disease management is enhanced through continuous monitoring, allowing healthcare providers to adjust treatment plans based on real-time data and patient responses [16].
- Scalability and Accessibility: The scalable nature of IoT and deep learning solutions allows for widespread adoption, making advanced healthcare monitoring accessible to a larger population.

B. Ethical and Privacy Concerns

The integration of IoT and deep learning technologies in home care presents transformative opportunities for healthcare, yet it also raises ethical, privacy, and security concerns that require careful consideration. This section explores the key ethical and privacy considerations associated with empowering home care through advanced technologies, emphasizing the importance of safeguarding patient data. Ensuring the security of patient data necessitates the implementation of robust methodologies, including data encryption, secure communication protocols, and multi-factor authentication. These measures protect sensitive information from unauthorized access and cyber threats while maintaining data integrity. Additionally, adopting privacy-by-design principles ensures that security measures are embedded into the technology from the outset, aligning with ethical standards and regulatory compliance. By addressing these challenges holistically, we can build trust and confidence in the use of IoT and deep learning in home care. The author in [36] highlights the ethical considerations related to using deep learning in personalized health monitoring, such as privacy, bias, responsibility and accountability, data quality, and the interpretability of results. In summary, the paper provides valuable insights into the recent developments, challenges, and ethical considerations in the application of deep learning for personalized health monitoring and prediction, underscoring the potential of advanced deep learning algorithms to improve patient outcomes and revolutionize personalized healthcare significantly.

The author in [37] identifies the gaps in the discussion of ethical concerns in both theoretical and empirical research on SHHTs for older persons, highlighting the scarcity of ethical considerations in the field. They noted that while privacy was extensively discussed, other ethical issues, such as responsibility and ageism, were less prevalent in the literature. The review calls for more critical work to prospectively address ethical concerns that may arise with the development and use of new technologies in caregiving for older persons.

The author in [38], discuss the combination of IoT and deep learning offers a powerful toolkit for managing chronic diseases in the comfort of patients' homes. Here are some key benefits for patient, caregivers and for healthcare providers like:

TABLE V. Research Papers Achieving Article Objectives, Challenges and Benefits

References	Achieved Objectives
[36] [37] [32]	Assesses technological challenges and limitations of IoT and deep learning in-home care.
[32] [31] [33] [34] [39] [35] [36]	Provide recommendations and guidelines for activating the positive view of implementing IoT and deep learning technologies in-home care.

- Improved Patient Quality of Life: Continuous monitoring and personalized. These insights enforce manage patients, resulting in improved symptom control and an enhanced sense of independence.
- Reduced Burden and Stress for caregiving: Continuous monitoring provides caregivers with peace of mind, knowing that their loved ones are being monitored around the clock. Real-time alerts also allow them to respond quickly to emergencies.
- Enhanced Patient Management for health care provider: Remote monitoring provides healthcare providers with continuous data streams, enabling them to track patients' progress, adjust treatment plans, and intervene proactively.

C. Technical Challenges and Limitations

The technical challenges include data security concerns, interoperability issues, and the need for standardized protocols and regulations to ensure seamless integration with existing healthcare systems and networks [32]. The study addressed the technical challenges and limitations associated with IoT in healthcare systems. Various aspects of end-to-end IoT related to health care were explored. The study discussed the challenges in improving the efficiency of monitoring, such as power absorption and accuracy, particularly in the context of narrowband IoT technology. It was observed that the use of narrowband IoT for high-intensity applications, such as blood pressure and heartbeat monitoring in pregnant women, may support widespread communication with low data cost and minimum processing complexity, yet the limitations of long-range deployment and high latency for critical healthcare applications were identified. The study also presented a detailed review of IoT technologies related to the suggested model, exploring low-power networks and the suitability of IoT-adopted communication standards for healthcare applications. Cloud computing was highlighted as an efficient means of handling and organizing large amounts of healthcare data.

At the end of this section, we identify which research papers focus on the challenges and potential solutions and which emphasize the benefits of integrating IoT and deep learning to achieve the article's objectives, as shown in Table V.

VI. Case Studies and Implementation Examples

A. Remote Monitoring of Diabetes Patients

Diabetic Retinopathy is one of the leading causes of blindness among the working-age population and requires early detection for effective treatment. Traditionally, the diagnosis required manual examination by expert physicians, which was

time-consuming and not always readily accessible. Diabetes is a chronic disease that requires ongoing monitoring and management. Patients with diabetes and their caregivers face many challenges in managing the disease, such as constantly monitoring blood sugar levels, adhering to medication regimens, and making lifestyle changes. Remote monitoring of diabetes patients is a solution that has emerged to address these challenges and provide better support and care for patients [40]. There have been several studies conducted on the remote monitoring of diabetes patients to assess its effectiveness in improving patient outcomes and quality of life. One study conducted by researchers at Mepco Schlenk Engineering College in India aimed to develop a real-time monitoring system for the vital signs of patients with diabetes using wearable sensors. The study utilized wearable sensors to monitor the vital signs of diabetes patients in real time. The sensors collected data on temperature, respiratory rate, pulse, blood pressure, and blood oxygen saturation. The collected data was stored in a text document and analyzed using data mining approaches. The study found that remote monitoring of vital signs using wearable sensors allowed patients to know their health status without the help of a nurse. Additionally, the study highlighted that remote monitoring of vital signs can help in the early detection of any medical problems or illnesses and alleviate the need for frequent hospital visits. By implementing a remote monitoring system for diabetes patients, healthcare providers can have access to continuous and real-time data on the patient's vital signs. This enables healthcare providers to make informed decisions, adjust medication regimens as needed, and provide timely interventions when necessary [41].

At the end, the use of such technology could lead to a broader application of automated systems in healthcare, helping to prevent loss of vision in diabetic patients.

B. Predictive Analytics for Cardiovascular Disease Management

Heart disease remains a leading cause of mortality globally. Traditional healthcare models struggle with early detection and continuous monitoring of heart conditions, which is crucial for patients at risk of sudden cardiac events. According to a report by the Public Health Foundation (PHF) of India, chronic diseases, including cardiovascular diseases, are the leading cause of death in India. These diseases pose a significant challenge for healthcare systems, specially in monitoring vital signs of chronic diseases patient. Traditionally, the monitoring of vital signs has been done by nurses in hospitals or with the assistance of healthcare professionals at home. However, the advancements in AI technologies and the availability of wearable sensors, real-time monitoring of vital signs has become more accessible. By using wearable sensors, patients can monitor their own vital signs and track their health status without the need for constant assistance from healthcare professionals. This system enables high-risk patients to be checked in a timely manner and enhances the quality of life for patients with chronic diseases. One approach to improve the management of the cardiovascular disease is the use of predictive analytics [42]. Predictive analytics involves the use of data mining techniques to analyze and predict future outcomes based on historical data. By applying predictive analytics to cardiovascular disease management, healthcare providers can gain valuable insights and identify patterns and risk factors that may increase the development or progression of the disease. One study conducted in the United States used survey data and ML models to identify and predict at-risk patients for cardiovascular disease [43].

Further research into incorporating additional deep learning techniques and algorithms into IoT environments promises to improve accuracy and utility for hospitals. Collaboration with more medical institutions to collect extensive data sets would drive the evolution of deep learning models for even better predictive capabilities and treatment outcomes in heart disease management.

VII. Future Directions

The field of healthcare is continuously evolving, and with the advancements in technology, there is a growing opportunity to revolutionize home care for patients with chronic diseases. One future direction for empowering home care is the utilization of IoT and deep learning techniques for intelligent monitoring and management of chronic diseases. By integrating IoT devices, patients can receive continuous monitoring of their vital signs, medication adherence, and overall health status from the comfort of their own homes. This approach has the strength to enhance patient healthcare outcomes by enabling early intervention and personalized treatment plans [44]. Furthermore, the integration of IoT and deep learning in-home care can improve communication and collaboration. Patients and healthcare providers can easily share data and communicate through a secure platform, allowing for timely interventions and adjustments to treatment plans. This future direction of empowering home care through IoT and deep learning has the potential to transform the way chronic diseases are managed and monitored.

A. Integration with Electronic Health Records (EHRs)

The integration of EHRs has revolutionized the healthcare industry by providing a seamless flow of patient information and improving clinical outcomes [45]. However, the future direction of integration with EHRs will focus on further enhancing the capabilities and functionalities of these systems to unlock their full potential in improving patient care and operational efficiency. This will enable a comprehensive view of patient health and facilitate coordinated care across different providers and healthcare settings. With the growth of the populations, healthcare will address challenges related to scaling the system for large populations and ensuring the robustness of deep learning models against noisy, incomplete, or anomalous data inputs from IoT devices. This integration will facilitate personalized medicine and enable proactive interventions to prevent adverse health events and explore the use of federated learning as a collaborative ML approach that allows model training across multiple healthcare entities without sharing raw data, thus preserving patient privacy using the innovative cryptographic methods and privacy-preserving ML techniques [46]. Overall, the future direction of integration with EHRs will focus on leveraging emerging technologies, improving interoperability, and integrating with wearable devices and mobile health applications to enhance patient care.

B. Wearable and Implantable Devices

The field of wearable and implantable devices has seen significant advancements in recent years, with technology becoming smaller, more efficient, and more integrated into our daily lives. These devices offer numerous benefits in healthcare, personal monitoring, and fitness tracking. However, the future of wearable and implantable devices holds even greater possibilities. One future direction of wearable and implantable devices is their integration with AI and ML algorithms. With the advancements in AI and ML algorithms, these devices can not only collect data but also analyze and interpret that data to provide personalized insights and recommendations for the user. For example, wearable devices could utilize ML and provide insights into their stress levels or predict potential cardiac events. Another future direction of wearable and implantable devices is their integration with telemedicine. This integration will allow for remote monitoring and virtual healthcare services. These devices are capable of transmitting real-time data to healthcare providers, enabling them to monitor patients' conditions, intervene in a timely manner, and adjust treatment plans as necessary.

Another future direction of wearable and implantable devices is their integration with IoT technology. By integrating wearable sensors and implantable devices, data sharing across various devices and platforms can enhance the efficiency and accuracy of data collection, analysis, and dissemination, ultimately leading to improved outcomes. The future direction of wearable and implantable devices is focused on the integration of AI and deep learning learning algorithms, telemedicine capabilities, biofeedback therapy, and IoT integration. Some potential future directions for wearable and implantable devices proposed by [47] include, development of advanced ML algorithms for accurate prediction and detection of health conditions based on wearable device data and integration of wearable devices with telemedicine platforms to enable remote monitoring and virtual healthcare services.

These continuously evolving technology platforms not only promise to help people pursue a healthier lifestyle but also provide continuous medical data for actively tracking metabolic status, diagnosis, and treatment. Integration with the IoT is a significant future direction for wearable and implantable devices [48]. The future direction proposed by [49] is revolutionized by the integration of deep learning and IoT with advanced wearable and implantable devices. This will lead to:

- Sophisticated Monitoring: Wearables will go beyond the wrist, using non-invasive methods to track a wider range of health indicators. Smart implantables will enable targeted drug delivery and personalized neuromodulation therapies.
- Data-Driven Insights: Body area networks will connect these devices, feeding data to AI models for real-time analysis, prediction of health events, and personalized recommendations.
- Patient Empowerment: Individuals will receive personalized feedback and use gamified apps to actively participate in their health management.

In conclusion, the future direction of wearable and implantable devices includes integration with the Internet of Things, utilization of AI and ML algorithms, integration with telemedicine platforms, and advancements in monetization and energy efficiency.

C. Collaborative Care and Telemedicine

Collaborative care and telemedicine is very important in healthcare sections. These approaches as proposed by [50]. have shown great potential in enhancing patient care, easy access to all healthcare services, and driving operational efficiencies. Moving forward, there are several future directions that can further advance collaborative care and telemedicine:

- Expansion of Telehealth Services
- Integration of AI
- Remote Patient Monitoring
- Enhanced Communication and Collaboration Tools
- Telemedicine Services Expansion
- Integration with Electronic Health Records

VIII. Conclusion

The integration of IoT technology and deep learning algorithms holds immense promise for transforming the monitoring and management of chronic diseases within home care settings. This convergence represents a beacon of hope, marking the dawn of a new era in chronic disease management. Through the intricate interplay of sensor-laden devices and sophisticated algorithms, we transcend traditional healthcare boundaries, envisioning a future where personalized, proactive, and precise interventions are the bedrock of patient-centric care.

By leveraging connected devices and advanced analytics, providers can deliver tailored, proactive, and efficient care to patients, resulting in improved health outcomes and an enhanced quality of life. However, as we delve deeper into these technologies, it is imperative to prioritize patient privacy, data security, and ethical considerations. Responsible and effective implementation of these innovative solutions in home care necessitates a steadfast commitment to safeguarding patient rights and upholding ethical standards.

Yet, amid the promise of progress, we are confronted with profound ethical and existential questions. As our health data increasingly migrates to the digital realm, we must grapple with the weighty implications of privacy, security, and autonomy. It is essential to strike a balance between innovation and the moral duty to protect patient trust and dignity. Transparency in data usage and robust security frameworks must underpin every technological advancement.

We think, Furthermore, the success of these technologies hinges not only on their development but also on equitable access. It is vital to ensure these advancements do not create disparities in healthcare but instead foster inclusivity by reaching underserved populations. Collaboration between policymakers, healthcare providers, and technologists will be crucial in realizing this vision.

As we navigate this evolving landscape, let us remain vigilant in our pursuit of innovative solutions, mindful of the ethical complexities inherent in healthcare technology. By adopting a holistic approach that prioritizes patients at the center of care, fostering trust, inclusivity, and equity, we can leverage the transformative capabilities of IoT and deep learning to usher in a new era of empowered, accessible, and ethical home care.

ACKNOWLEDGMENT

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU242729].

REFERENCES

[1] R. Istepanian, E. Jovanov, and Y.-T. Zhang, "Guest editorial introduction to the special section on m-health: Beyond seamless mobility and global wireless health-care connectivity," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 8, pp. 405 – 414, 01 2005.

[2] K. Ragavan, R. Ramalakshmi, V. SrirengaNachiyar, G. Priya, and K. Jeyageetha, "Smart health monitoring system in intensive care unit using bluetooth low energy and message queuing telemetry transport protocol," in *2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2023, pp. 284–291.

[3] M. Nasr, M. M. Islam, S. Shehata, F. Karray, and Y. Quintana, "Smart healthcare in the age of ai: Recent advances, challenges, and future prospects," *IEEE Access*, vol. 9, pp. 145 248–145 270, 2021.

[4] W.-J. Chang, L.-B. Chen, C.-H. Hsu, C.-P. Lin, and T.-C. Yang, "A deep learning-based intelligent medicine recognition system for chronic patients," *IEEE Access*, vol. 7, pp. 44 441–44 458, 2019.

[5] P. P. Ray, "Home health hub internet of things (h3iot): An architectural framework for monitoring health of elderly people," *2014 International Conference on Science Engineering and Management Research (ICSEMR)*, pp. 1–3, 2014. [Online]. Available: https://api.semanticscholar.org/CorpusID:31572811

[6] A. Qaffas, M. Hoque, and N. Almazmomi, "The internet of things and big data analytics for chronic disease monitoring in saudi arabia," *Telemedicine and e-Health*, vol. 27, 04 2020.

[7] D. Gupta, S. Khare, and A. Aggarwal, "A method to predict diagnostic codes for chronic diseases using machine learning techniques," in *2016 International Conference on Computing, Communication and Automation (ICCCA)*, 2016, pp. 281–287.

[8] N. Alharbe and M. Almalki, "Iot-enabled healthcare transformation leveraging deep learning for advanced patient monitoring and diagnosis," *Multimedia Tools and Applications*, pp. 1–14, 07 2024.

[9] P. Roy, V. Teju, S. Kandula, K. Sowmya, A. Stan, and O. Stan, "Secure healthcare model using multi-step deep q learning network in internet of things," *Electronics*, vol. 13, p. 669, 02 2024.

[10] S. S. Begum, H. A.L.ShammariHayfaMashl, B. Karthikeyan, and F. Z. Alanazi, "A prediction of heart disease using iot based thingspeak basis and deep learning method," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:270743516

[11] R. Rayan, C. Sagkaris, and I. Romash, *The Internet of Things for Healthcare: Applications, Selected Cases and Challenges*, 01 2021, pp. 1–15.

[12] T. Vaiyapuri, E. L. Lydia, M. Y. Sikkandar, V. G. Díaz, I. V. Pustokhina, and D. A. Pustokhin, "Internet of things and deep learning enabled elderly fall detection model for smart homecare," *IEEE Access*, vol. 9, pp. 113 879–113 888, 2021.

[13] H. Zhu, S. Samtani, and R. Brown, "A deep learning approach for recognizing activity of daily living (adl) for senior care: Exploiting interaction dependency and temporal patterns," *MIS Quarterly*, vol. 45, pp. 859–896, 06 2021.

[14] W.-J. Chang, L.-B. Chen, C.-H. Hsu, C.-P. Lin, and T.-C. Yang, "A deep learning-based intelligent medicine recognition system for chronic patients," *IEEE Access*, vol. 7, pp. 44 441–44 458, 2019.

[15] A. Raji, P. G. Jeyasheeli, and T. Jenitha, "Iot based classification of vital signs data for chronic disease monitoring," in *2016 10th International Conference on Intelligent Systems and Control (ISCO)*, 2016, pp. 1–5.

[16] P. Ingale, S. Nandanwar, K. Buva, D. Bhatia, P. Choudhury, and M. Tamboli, "Enhancing patient care and monitoring using ai and iot in healthcare," 06 2023.

[17] M. Naeem, G. Paragliola, and A. Coronato, "A reinforcement learning and deep learning based intelligent system for the support of impaired patients in home treatment," *Expert Systems with Applications*, vol. 168, p. 114285, 11 2020.

[18] I. Preethi and K. Dharmarajan, "Diagnosis of chronic disease in a predictive model using machine learning algorithm," in *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)*, 2020, pp. 191–196.

[19] S. Ganiger and K. Rajashekharaiah, "Chronic diseases diagnosis using machine learning," in *2018 International Conference on Circuits and Systems in Digital Enterprise Technology (ICCSDET)*. IEEE, 2018, pp. 1–6.

[20] R. Rashid, W. Aslam, R. Aziz, and G. Aldehim, "An early and smart detection of corn plant leaf diseases using iot and deep learning multi-models," *IEEE Access*, vol. 12, pp. 23 149–23 162, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:267145971

[21] M. Dhinakaran, K. Phasinam, J. Alanya-Beltran, K. Srivastava, V. Babu D, and S. Singh, "A system of remote patients' monitoring and alerting using the machine learning technique," *Journal of Food Quality*, vol. 2022, pp. 1–7, 02 2022.

[22] H. Zhu, H. Chen, and R. Brown, "A sequence-to-sequence model-based deep learning approach for recognizing activity of daily living for senior care," *Journal of Biomedical Informatics*, vol. 84, 07 2018.

[23] R. Pakrooh, A. Jabbari, and C. Fung, "Deep learning-assisted security and privacy provisioning in the internet of medical things systems: A survey on recent advances," *IEEE Access*, vol. 12, pp. 40 610–40 621, 2024. [Online]. Available: https://api.semanticscholar.org/CorpusID:268399213

[24] D. K G, "Smart health monitoring using deep learning and artificial intelligence," vol. 37, pp. 451–464, 05 2023.

[25] S. Vitabile, M. Marks, D. Stojanovic, S. Pllana, J. Molina, M. Krzyszton, A. Sikora, A. Jarynowski, F. Hosseinpour, A. Jakóbik, A. Stojnev Ilić, A. Respício, D. Moldovan, C. B. Pop, and I. Salomie, *Medical Data Processing and Analysis for Remote Health and Activities Monitoring*, 03 2019, pp. 186–220.

[26] A. Malibari, "An efficient iot-artificial intelligence-based disease prediction using lightweight cnn in healthcare system," *Measurement: Sensors*, vol. 26, p. 100695, 04 2023.

[27] I. Chouvarda, D. Goulis, I. Lambrinoudaki, and N. Maglaveras, "Connected health and integrated care: Toward new models for chronic disease management," *Maturitas*, vol. 82, 03 2015.

[28] A. W. Evers, M. M. Rovers, J. A. Kremer, J. A. Veltman, J. A. Schalken, B. R. Bloem, and A. J. Van Gool, "An integrated framework of personalized medicine: from individual genomes to participatory health care," *Croatian medical journal*, vol. 53, no. 4, pp. 301–303, 2012.

[29] J. Cheah, "Chronic disease management: A singapore perspective," *BMJ (Clinical research ed.)*, vol. 323, pp. 990–3, 11 2001.

[30] A. Qaffas, M. Hoque, and N. Almazmomi, "The internet of things and big data analytics for chronic disease monitoring in saudi arabia," *Telemedicine and e-Health*, vol. 27, 04 2020.

[31] X. Ma, T. Yao, M. Hu, Y. Dong, W. Liu, F. Wang, and J. Liu, "A survey on deep learning empowered iot applications," *IEEE Access*, vol. 7, pp. 181 721–181 732, 2019.

[32] C. Li, J. Wang, S. Wang, and Y. Zhang, "A review of iot applications in healthcare," *Neurocomputing*, p. 127017, 2023.

[33] H. Bolhasani, M. Mohseni, and A. M. Rahmani, "Deep learning applications for iot in health care: A systematic review," *Informatics in Medicine Unlocked*, vol. 23, p. 100550, 2021.

[34] A. Shamsabadi, Z. Pashaei, A. Karimi, P. Mirzapour, K. Qaderi, M. Marhamati, A. Barzegary, S. Alinaghi, and O. Dadras, "Internet of things in the management of chronic diseases during the covid-19 pandemic: A systematic review," *Health Science Reports*, 03 2022.

[35] K. Mohammed, A. Zaidan, B. Bahaa, O. Albahri, M. Alsalem, A. Albahri, A. Mohsin, and M. Hashim, "Real-time remote-health monitoring systems: a review on patients prioritisation for multiple-chronic diseases, taxonomy analysis, concerns and solution procedure," *Journal of Medical Systems*, vol. 43, 06 2019.

[36] R. K. NVPS, R. Damaševic ˇius, S. K. Jagatheesaperumal, S. Hussain, R. Alizadehsani, and J. Gorriz, "Deep learning for personalized health monitoring and prediction: A review," *Authorea Preprints*, 2023.

[37] N. Felber, A. Tian, F. Pageau, B. Elger, and T. Wangmo, "Mapping ethical issues in the use of smart home health technologies to care for older persons: a systematic review," *BMC Medical Ethics*, vol. 24, 03 2023.

[38] H. Lee, Y.-S. Park, S. Yang, H. Lee, T.-J. Park, and D. Yeo, "A deep learning-based crop disease diagnosis method using multimodal mixup augmentation," *Applied Sciences*, vol. 14, p. 4322, 05 2024.

[39] P. Jeyaraj and S. Rajan, "Smart-monitor: Patient monitoring system for iot-based healthcare system using deep learning," *IETE Journal of Research*, vol. 68, pp. 1–8, 08 2019.

[40] İ. KABALI and Ö. Sema, "Communication with chronic patients and patient relatives in the example of diabetes disease," *Tıp Eğitimi Dünyası*, vol. 19, no. 57, pp. 109–119, 2020.

[41] B. Manakatt, "A project to evaluate the effectiveness of a shared web-based log in managing blood glucose levels among patients with type 2 diabetes mellitus," *Madridge Journal of Case Reports and Studies*, vol. 2, pp. 80–84, 10 2018.

[42] M. A. Ali, H. A. Alzaabi, A. S. Alnuaimi, and A. Jamdal, "Smart healthcare device for cardiac patients," in *2020 Advances in Science and Engineering Technology International Conferences (ASET)*. IEEE, 2020, pp. 1–5.

[43] A. Dinh, S. Miertschin, A. Young, and S. Mohanty, "A data-driven approach to predicting diabetes and cardiovascular disease with machine learning," *BMC Medical Informatics and Decision Making*, vol. 19, 11 2019.

[44] R. Miotto, F. Wang, S. Wang, and X. Jiang, "Deep learning for healthcare: review, opportunities and challenges," *Briefings in bioinformatics*, vol. 19, 05 2017.

[45] M. Follen, R. Castaneda, M. Mikelson, D. Johnson, A. Wilson, and K. Higuchi, "Implementing health information technology to improve the process of health care delivery: A case study," *Disease management : DM*, vol. 10, pp. 208–15, 09 2007.

[46] G. Harerimana, J. Kim, and B. Jang, "Deep learning for electronic health record analytics," *IEEE Access*, vol. PP, pp. 1–1, 07 2019.

[47] D. MacLean, A. Roseway, and M. Czerwinski, "Moodwings: A wearable biofeedback device for real-time stress intervention," 05 2013.

[48] A. K. Yetisen, J. L. Martinez-Hurtado, B. Ünal, A. Khademhosseini, and H. Butt, "Wearables in medicine," *Advanced Materials (Deerfield Beach, Fla.)*, vol. 30, 2018. [Online]. Available: https://api.semanticscholar.org/CorpusID:48355667

[49] S. Mei, Y. Zhou, J. Xu, Y. Wan, S. Cao, Q. Zhao, S. Geng, J. Xie, and S. Hong, "Deep learning for detecting and early predicting chronic obstructive pulmonary disease from spirogram time series: A uk biobank study," 2024. [Online]. Available: https://arxiv.org/abs/2405.03239

[50] A. J. Neri, G. P. Whitfield, E. T. Umeakunne, J. E. Hall, C. J. DeFrances, A. B. Shah, P. K. Sandhu, H. B. Demeke, A. R. Board, N. J. Iqbal, K. Martinez, A. M. Harris, and F. V. Strona, "Telehealth and public health practice in the united states-before, during, and after the COVID-19 pandemic," *J. Public Health Manag. Pract.*, vol. 28, no. 6, pp. 650–656, Aug. 2022.

Performance Comparison of Object Detection Models for Road Sign Detection Under Different Conditions

Zainab Fatima[1], M. Hassan Tanveer[2], Hira Mariam[3], Razvan Cristian Voicu[4], Tanazzah Rehman[5], Rizwan Riaz[6]
Department of Software Engineering, NED University of Engineering & Technology, Karachi 75270, Pakistan[1,5,6]
Department of Robotics and Mechatronics Engineering, Kennesaw State University, Marietta, GA 30060, USA[2,4]
Department of Telecommunications Engineering, NED University of Engineering & Technology, Karachi 75270, Pakistan[3]

***Abstract*—During driving, drivers often overlook the traffic signs along the roads compromising road safety and increasing the risk of accidents. To address this, artificial intelligence (AI) and deep learning techniques are employed, taking into consideration the improvement of advances in Artificial Neural Networks (ANNs) and image processing for robust road sign detection. In this work, we compare the performance of existing state-of-the-art object detection models for road sign detection, including YOLOv8, YOLOv9, RTMDet, Faster-RCNN and RetinaNet, using a large dataset of images of road signs. These models are fine-tuned and hyperparameters are optimized with varied settings like auto-orientation and augmentation during the preprocessing and training phase. The models are then tested, and key performance indicators such as mean average precision (mAP), number of inferences performed per second [frames per second (fps)], and total loss are evaluated. Our study reaffirms the earlier findings in which YOLOv9 and YOLOv8 outperform other detectors in real-time detection tasks because they are faster in inference or prediction than most detectors, but with a compromise in accuracy, as highlighted by the fast fps rates. In contrast, RTMDet is fast and reliable, making it a highly effective option for detecting various road signs. The insights presented in this research are useful in identifying the appropriateness and drawbacks of each model, thereby benefiting from the selection of the best suited model for real-world applications, such as autonomous vehicles or self-driving cars.**

***Keywords*—Artificial intelligence; artificial neural networks; image processing; deep learning; road signs detection**

I. Introduction

A. Background

Over the last few years, the use of computer vision and object detection has become prominently efficient because of the shift in deep learning models. Such advancements have embellished object identification in both pictures and videos with great precision and efficacy. The relevance of an object being detectable by a vehicle increases with the overall progress in autonomous vehicles and smart and connected roads. Road sign detection has the necessary accuracy and operational capacity for serving its purpose in ensuring the safety for autonomous vehicles and advanced driver assistance systems ADAS.

B. Current State of Object Detection Models

This research paper presents a thorough comparative analysis of five leading object detection models: YOLOv9, YOLOv8, RetinaNet, RTMDet, and Faster R-CNN [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19]. Each of these models epitomizes state-of-the-art techniques in object detection, boasting unique strengths and characteristics that influence their performance in identifying objects within visual data.

- YOLO: You Only Look Once (YOLO) series, specifically YOLOv9 and YOLOv8, are most famous with real-time object detection, and that is why such models are applicable in areas that require high response rates [20], [21], [22]. They are particularly remarkable in terms of speed and efficiency of detection, which is crucial in contexts, such as autonomous driving, where minimizing detection time is of utmost importance.
- RetinaNet: Distinguished by its ability to accurately detect objects of varying sizes within images, making it a robust choice for scenarios involving diverse object scales. RetinaNet employs a Feature Pyramid Network (FPN) and a focal loss function to address the class imbalance and enhance detection accuracy [23], [24].
- Faster R-CNN: Widely recognized for its precise localization of objects, achieved through a region-based convolutional neural network that meticulously analyzes different regions within an image to ensure accurate detection [25], [26].
- RTMDet: RTMDet is known for its robustness in managing complex scenes and occlusions, demonstrating a superior ability to detect objects even under challenging conditions [27]. RTMDet integrates a modified ResNet-50 backbone with spatial and feature alignment modules to optimize detection performance.

C. Motivation

While YOLO series has been the object detection models that has made remarkable progress, there is a lack of extensive comparative analysis that is not only centered on the recent versions of the YOLO models (YOLOv9 and YOLOv8) but other cutting-edge models like RetinaNet, RTMDet, and Faster RCNN. Earlier works have mostly focused on giving attention to overall object identification or testing only a few of these models under various circumstances, which really did not enable researchers to make a Comparative analysis of these models for the purpose of road sign identification. This

research intends to address this gap by comparing these models in detail and side by side in the context of road sign detection focusing on several measures and metrics and with an eye to useful recommendations for implementation in self-driving car technologies and as well in the driver assistant systems.

D. Contributions

The purpose of this work is to advance the prior art related to object detection and present a comprehensive and comparative evaluation of five of the most relevant approaches. When assessing the accuracy of each model, we have apply a set of unified metrics stemmed from a dataset of road signs, which provides useful information on the merits and drawbacks of each model. It can help researchers and practitioners on how best to choose the most suitable object detection model for intelligent transportation systems and autonomous driving.

This paper presents a comprehensive comparison of five prominent object detection architectures: YOLOv8, YOLOv9, RTMDet, RetinaNet, and Faster RCNN for the road sign detection study without leaving out the factors such as mAP and inference time. This paper is organised as follow: Section II presents the literature review relevant to this work, where we discuss the current state of object detection algorithms as well as road sign detection. Section III describes the three step model architectures of the aforementioned algorithms while Section III(B) explains the design principles and features of these algorithms. Section IV explains the experimental setup, procedures on how the datasets were prepared, training configurations that were used for training the models as well as the evaluation metrics that were employed in the assessment of the performance of each model. Results and discussion are discussed in Section V. Lastly, Section VI of this paper offers the conclusion and a discussion of the experiments to derive the factors that should be preferred while choosing between accuracy, speed, and confidence in detecting a road sign from the ones listed above or some other algorithms as per the parameters that will be followed strictly at the time of implementing the vision based systems.

II. Literature Review

The recognition of road signs is one of the critical segments of Intelligent Transportation Systems (ITS) performance, so its function is significant for providing road safety and effectiveness. In addition, as the deep learning starts to become more popular, object detection models show great potential in recognizing road sign. Road sign detection is revealed in this literature review as a line of research that experienced significant development in the last five years. In this section, the object detection models are introduced and analyzed based on the approach used, efficiency, effectiveness, and the kind of application which is road sign detection. The given review culminates the prior studies in order to perceive common aspects, issues, and opportunities in order to further establish more reliable and accurate type of road signs detection systems.

- This research enhanced YOLOv8 by adding blur and noise, and incorporating an asymptotic feature pyramid network, which improved the detection of small target objects. It achieved a 3.31% increase in mean Average Precision (mAP) and a 3.59% increase in recall on the TT100K dataset. These improvements were confirmed through ablation studies, highlighting the contributions of both data augmentation and the AFPN enhancements [28].

- Enhanced YOLOv8 algorithm for traffic sign recognition using the Kaggle dataset incorporated Cross-Stage Partial connection and Path Aggregation Network, achieving 80% accuracy, 64% precision, and 65.67% recall on test data. The use of stochastic gradient descent optimization and dropout helped curb overfitting, demonstrating the model's efficacy in complex traffic sign analysis [29].

- In a comparative study, YOLOv5 demonstrated superior performance over SSD in traffic sign recognition, achieving a mAP@0.5 of 97% and processing images at 30 FPS on the VOC dataset. SSD showed 90% accuracy but was significantly slower, processing images at only 3 FPS. YOLOv5's faster recognition speed and higher recall score make it a better solution for traffic sign recognition in Intelligent Transportation Systems [30].

- This research improved traffic sign detection using Faster R-CNN with enhancements like feature pyramid, deformable convolutions, and ROI alignment. Tested on the TT100K dataset, it achieved high accuracy rates of 92.6% in sunny conditions, 90.6% at sunset, and 86.9% in rainy conditions, outperforming SSD, YOLOv2, and even YOLOv5 in less favorable lighting and weather conditions [31].

- A real-time traffic sign detection system using Faster R-CNN was trained on a dataset of 1880 images from Turkey and the German Traffic Sign Recognition Benchmark (GTSRB). The model, trained over 10,000 iterations, achieved an accuracy of 88.99% and a total loss rate of 0.220, demonstrating robust detection capabilities [32].

- Zhu et al. developed a RetinaNet-based algorithm achieving a final F1 score of 0.923. While the model performed well in favorable conditions, it faced limitations in adverse weather, indicating a need for future research to improve performance under varied conditions [33].

- Inspired by YOLOv4 and YOLOv5, the TSR-SA method enhanced traffic sign detection by incorporating high-level features, a receptive field block-cross in the neck, and a Random Erasing-Attention data augmentation method. This approach achieved a state-of-the-art mAP of 90.2% on the TT100K dataset, surpassing YOLOv4 and YOLOv5-x, though it faced challenges with category imbalance [34].

- Senthilnayaki's system used Faster R-CNN for detection and Inception V2 for classification, improving detection in varied conditions by increasing anchor thickness and enhancing feature map resolution. This method proved effective and resilient, with plans for future enhancements to refine feature maps for more robust proposals [35].

- A system using the RetinaNet model was developed for real-time traffic light detection, employing transfer learning and modifications to anchor box sizes to detect small traffic lights. The model achieved a weighted mean average precision (mAP) of 0.54 with an execution time of 108ms, showing significant improvements in detection accuracy and speed [36].
- Jiang's work on an improved YOLOv5 network introduced a balanced feature pyramid and a global context block, enhancing feature fusion and extraction. Tested on the TT100K dataset, the model showed significant performance improvements with a 1.9% increase in mAP@.5, a 2.1% increase in mAP@.5:0.95, and improvements in precision and recall by 2.4% and 3.3%, respectively, proving its superiority for traffic sign detection [37].

III. Models Architecture

A. YOLOv8

The YOLOv8 model architecture represents a significant advancement in object detection, offering improved accuracy and speed over its predecessors. As shown in Fig. 1, the backbone of YOLOv8 is a modified CSPDarknet53 with 53 convolutional layers and cross-stage partial connections to enhance information flow [38].

The head of YOLOv8, includes multiple convolutional and fully connected layers responsible for predicting bounding boxes, objectiveness scores, and class probabilities. A notable feature of the head is the self-attention mechanism, which enables the model to focus on different parts of an image, adjusting the importance of various features.

YOLOv8 adopts an anchor-free detection approach, directly predicting object centers instead of offsets from predefined anchor boxes, thus speeding up post-processing steps like Non-Maximum Suppression (NMS). Additionally, YOLOv8 introduces changes in convolutions, such as replacing a 6x6 conv with a 3x3 conv in the stem and modifying the main building block, which enhances performance and efficiency.

B. YOLOv9

The YOLOv9 model architecture represents a significant advancement in real-time object detection, offering superior accuracy and efficiency compared to previous models. The backbone architecture utilizes Cross-stage Partial (CSP) Connection blocks to enhance gradient flow and reduce data loss during the feed-forward process, optimizing performance and accuracy. The head of YOLOv9 incorporates Programmable Gradient Information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN), which are crucial for preventing data loss, ensuring accurate gradient updates, and optimizing lightweight models for efficient object detection tasks (see Fig. 2 and 3). YOLOv9 also brings two new architectures, namely, YOLOv9-C and YOLOv9-E that improve accuracy and object detection efficiency in different applications through information choking and gradient flow rectification [39]. In addition, the YOLOv9 model has top-level accuracy and efficiency among the current models, including RT-DETR, YOLO-MS, or others due to the efficient use of conventional convolutions.

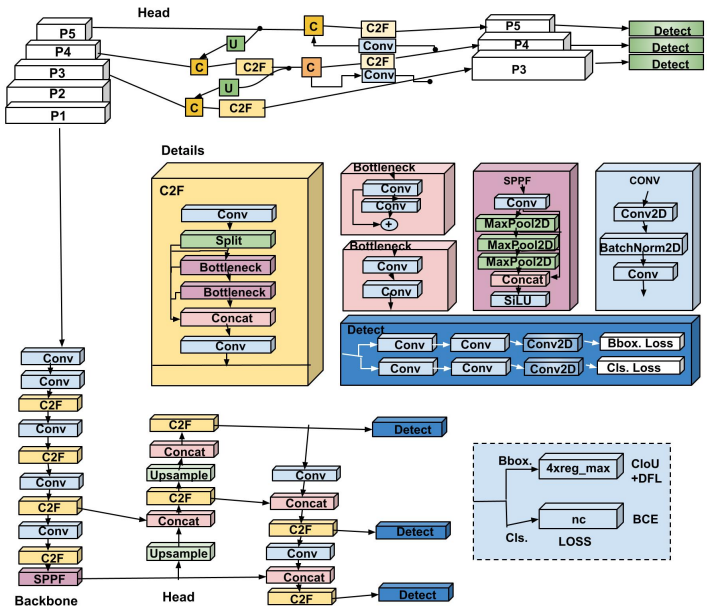


Fig. 1. YOLOv8 architecture.

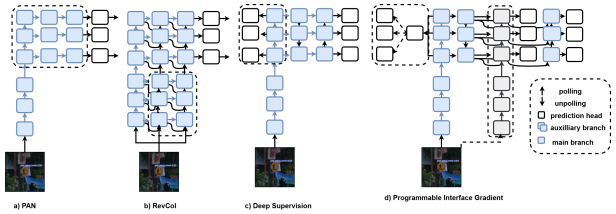


Fig. 2. YOLOv9 programmable gradient information.

C. RetinaNet

RetinaNet is a pioneering one-stage object detection model known for its exceptional performance in detecting objects at various scales. The unified network includes a backbone Convolutional Neural Network (CNN) and two task-specific subnets: the Classification Subnet and the Box Regression Subnet.

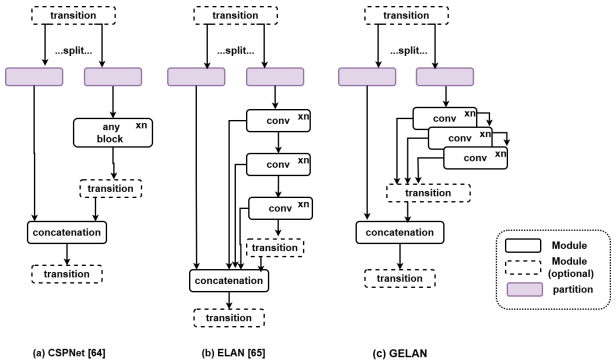


Fig. 3. YOLOv9 GELAN.

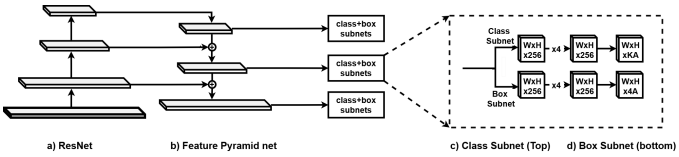


Fig. 4. RetinaNet structure.

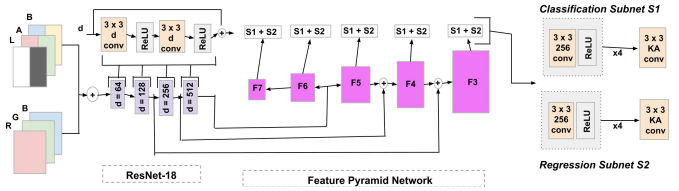


Fig. 5. Detailed structure.

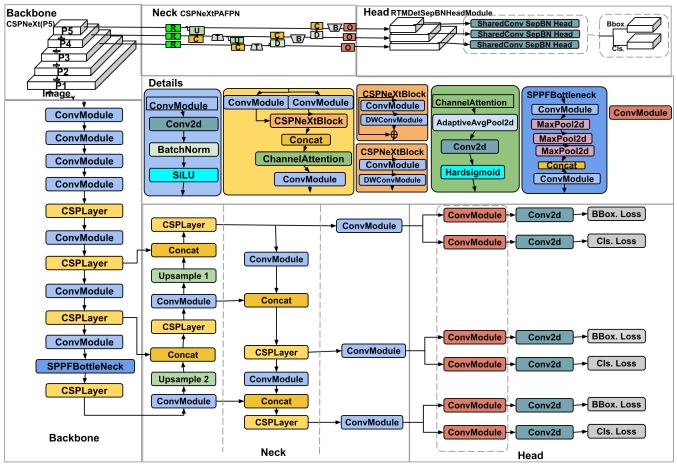


Fig. 6. RTMDet Architecture.

The backbone incorporates a Feature Pyramid Network (FPN), which generates a multi-scale feature pyramid by combining low-resolution semantically strong features with high-resolution features for accurate object detection. The Classification Subnet predicts object presence probabilities, while the Regression Subnet handles bounding box regression from anchor boxes, both using feature maps from the FPN (Fig. 4) and (Fig. 5). A key innovation is the Focal Loss function, which addresses class imbalance by assigning higher weights to hard examples, enhancing detection accuracy [24], [40]. Translation-invariant anchor boxes at different pyramid levels (P3 to P7) cover various scales and aspect ratios, enabling precise object localization and classification. During training, Stochastic Gradient Descent (SGD) is used with learning rate adjustments and data augmentation techniques like horizontal flipping to improve generalization.

D. RTMDet

RTMDet is an architecture will be proposed for real time object detection on the basis of the YOLO series of algorithms. It is used in scenarios where the detection of objects within images or videos in real-time is required, which makes this option highly effective in real-life designs. It uses ResNet-50 as its backbone after reducing the number of layers and parameters it uses for feature extraction. That is, some important modules: spatial attention module (SAM) is used to improve the feature extraction and the feature alignment module (FAM) to align the features from different scales [27].

The detection head of RTMDet uses a single convolutional layer to predict object bounding boxes and class probabilities, handling objects at three different scales: Depending on the amount of work, there are small, medium, and large offices. The RTMDet has three losses: CIoU loss that measures the loss of shapes and sizes, objectiveness loss that boosts completeness, and classification loss to maximize accuracy. For training, this model is trained on a large number of images and their corresponding labels and for the purpose of data augmentation has a considerable number of techniques incorporated. In inference, RTMDet replied it supports real-time object detection on still image and video, and it is deployable on different hardware environment such as GPU and TPU.

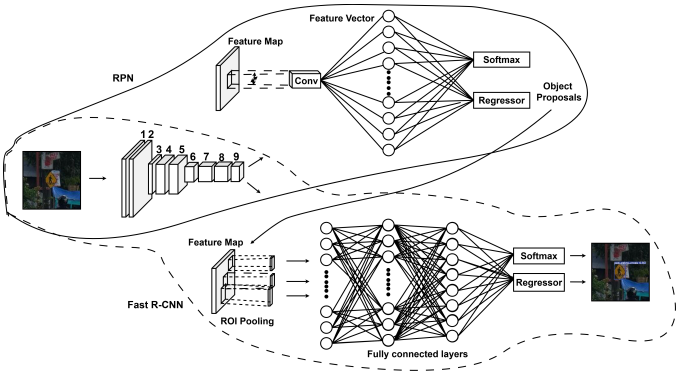


Fig. 7. Faster RCNN Architecture.

In this work, RTMDet presents itself with several advantages compared to the existing architectures of object detection such as high accuracy, high speed along with lesser computational complexity. Due to these characteristics, the technology is applied in self-driving cars, the monitoring systems, and robots among other areas. The above mentioned architecture including details of its subcomponents can be evident from Fig. 6.

E. Faster RCNN

Faster RCNN is a renowned object detection architecture that has significantly influenced the field of computer vision. The architecture comprises convolution layers trained to extract specific features from images, akin to how coffee filters allow only desired elements to pass through.

The Region Proposal Network (RPN) is a pivotal component of Faster RCNN, efficiently generating high-quality region proposals for subsequent detection. Fully connected neural networks are utilized to predict object classes and refine bounding boxes based on the regions proposed by the RPN (see Fig. 7).

Training Faster RCNN involves optimizing convolution layers filters, RPN weights, and the last fully connected layer weights using Stochastic Gradient Descent (SGD), ensuring effective and efficient object detection [26]. Faster RCNN has

demonstrated remarkable performance improvements over its predecessors, achieving faster processing speeds during both training and testing phases and setting new standards in object detection accuracy and efficiency.

IV. Experiment

A. Dataset

The "Road Signs" dataset sourced from Roboflow-100 consists of 21 classes falling under the super category "Road Signs". The images sample is shown in Fig. 8.

Fig. 8. Dataset images.

The dataset encompasses around 20 classes. It consists of a total of 2095 images, divided into a training set with 1378 images, a validation set with 488 images, and a test set with 229 images. Preprocessing techniques applied include auto-orientation to adjust image orientation and resizing to a resolution of 640x640 pixels, with no manual augmentations initially applied. However, each object detection model used in the study incorporated its own set of image augmentations during the training process to enhance the training data and improve model robustness and performance. This dataset offers a diverse collection of road sign images across multiple classes, ensuring uniformity in image dimensions and orientation for effective model training and testing.

B. Training Hyperparameters

Yolov8 and Yolov9 were trained using Ultralytics, RetinaNet and FasterRCNN using Detectron2 while RTMdet was trained using MMDetection. In this study, all the models were trained for 30 iterations through the dataset. YOLV8 and YOLV9 were set to 16, while RetinaNet, Faster RCNN, and RTMDet were set to 8. Here, parameters were adjusted for YOLOv8 with AdamW optimizer where the learning rate was set to 0. 0004, momentum of 0.9 and weight decay of 0.0005. As for optimization, YOLOv9 uses Standard Gradient Descent (SGD) with a learning rate of 0.01, momentum of 0.9 and weight decay to 0.0005. YOLOv9 employed SGD with a learning rate of 0.01, momentum of 0.9, and weight decay of 0.0005. RetinaNet and Faster RCNN both used SGD with a learning rate of 0.001, momentum of 0.9, and weight decay of 0.0001. RTMDet was optimized using AdamW with a learning rate of 0.004, momentum of 0.0002, and weight decay of 0.0001. The notes on the training dataset for both YOLOv8 and YOLOv9 were YOLOv5CocoDataset and COCO for RetinaNet, Faster RCNN, and RTMDet.

C. Data Augmentation Techniques

To increase the reliability of object detection models with respect to input images, the concept of data augmentation was used at the time of training. This was achieved by pre-processing the images by using albumentations for YOLOv8, YOLOv9, and RTMDet which applied blur, median blur, combining the image into grayscale, and CLAHE. On the other hand, RetinaNet and Faster RCNN stacked the pre-process tools provided by Detectron2, which involved resizing and random flipping.

D. Evaluation

While evaluating the results of object detection model, COCO evaluation metrics was used to make the evaluation. For instance, in using the model, COCO bounding box(bbox) test methodology was employed to assess the effectiveness of the method in identifying objects and determining the degree of precision. As this evaluation framework was widely employed in determining the effectiveness of such models in detecting malicious data, its use gave us a better chance to arrive at sound conclusions on performance of our powerfully designed model. The classification accuracy are recorded in Fig. 9 to 13.

V. Results and Discussion

A. Mean Average Precision (mAP)

The mean average precision (mAP) is a key metric in evaluating the performance of object detection models. The mAP scores for mAP50-95, mAP50, and mAP75 are shown in Table I.

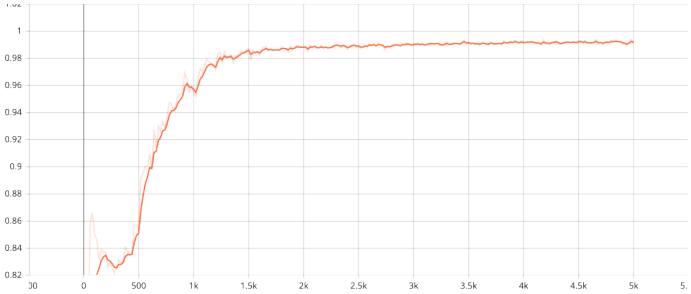


Fig. 9. Faster RCNN: Classification accuracy.

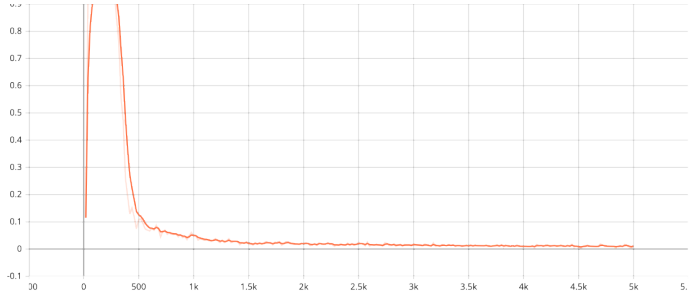


Fig. 10. Faster RCNN: False negatives.

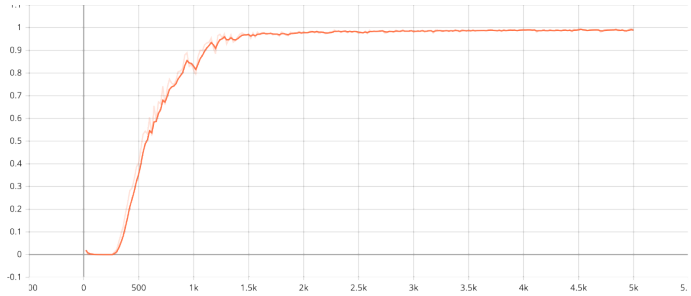


Fig. 11. Faster RCNN: Foreground classification accuracy.

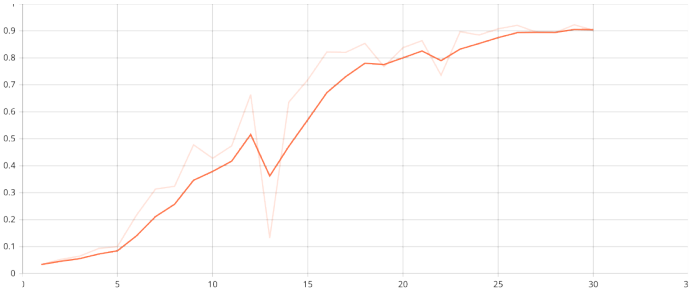


Fig. 12. RTMDet: mAP50.

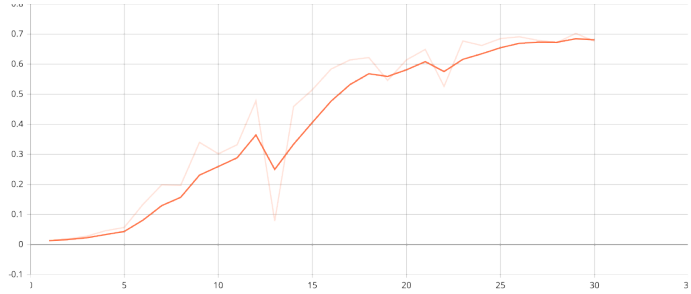


Fig. 13. RTMDet: mAP.

TABLE I. MEAN AVERAGE PRECISION

Model	mAP50-95	mAP50	mAP75
YOLOv8	0.814	0.961	0.892
YOLOv9	0.826	0.973	0.904
RTMDet	0.702	0.923	0.799
Faster RCNN	0.704	0.891	0.816
RetinaNet	0.755	0.911	0.854

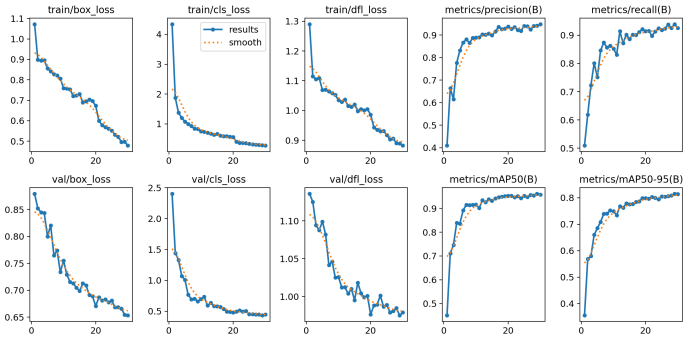


Fig. 14. YOLOv8: Loss and accuracy.

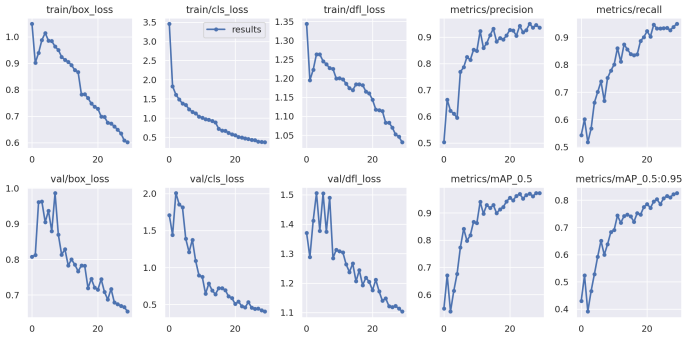


Fig. 15. YOLOv9: Loss and accuracy.

YOLOv9 achieves the highest mAP scores across all metrics, indicating its superior accuracy. YOLOv8 also performs very well, especially in terms of mAP50 and mAP75. RetinaNet shows a balanced performance, while Faster R-CNN and RTMDet exhibit relatively lower mAP scores.

B. Inference Time

Inference time is critical for applications requiring real-time object detection. The inference time of these models is shown in Table II.

TABLE II. INFERENCE TIME

Model	Inference Time
YOLOv8	0.0105s
YOLOv9	0.0311s
RTMDet	0.0576s
Faster RCNN	0.0844s
RetinaNet	0.0768s

YOLOv8 is the fastest model, making it highly suitable for real-time applications. YOLOv9, while slower than YOLOv8, still offers reasonable inference time. RTMDet, Faster R-CNN,

and RetinaNet are significantly slower, with Faster R-CNN having the highest inference time.

C. Total Loss

Total loss is a metric that indicates the overall error of a model during training, where lower values typically signify better performance. The total loss values for different object detection models: YOLOv8, YOLOv9, RTMDet, Faster RCNN, and RetinaNet, are depicted in Fig. 14 to 18 and presented in Table III. Among the models compared, YOLOv8 and YOLOv9 exhibit higher total loss values, indicating relatively higher error rates during training. On the other hand, RTMDet, Faster RCNN, and RetinaNet demonstrate significantly lower total loss values, suggesting better performance and potentially more effective learning during training.

TABLE III. TOTAL LOSS

Model	Total Loss
YOLOv8	2.0793
YOLOv9	2.1627
RTMDet	0.408
Faster RCNN	0.1519
RetinaNet	0.0735

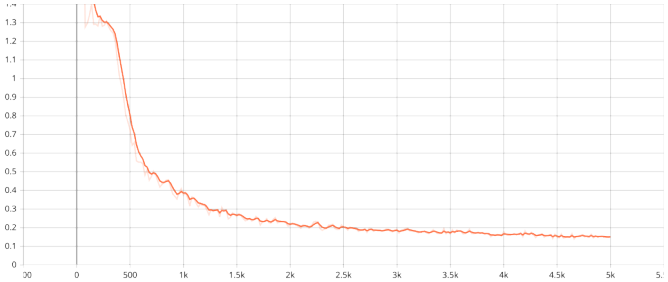


Fig. 16. Faster RCNN: Total loss.

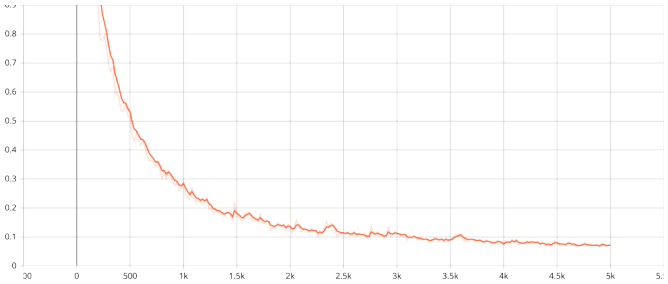


Fig. 17. RetinaNet: Total loss.

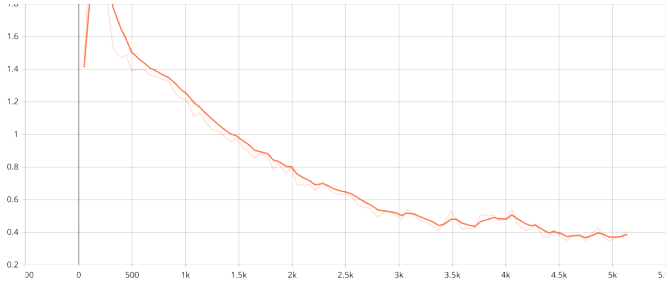


Fig. 18. RTMDet: Total loss.

D. Confusion Matrices

The performance of the YOLOv8, YOLOv9, Faster-RCCN, Retina Net and RTMDet is depicted by confusion matrices in Fig. 18 to 23.

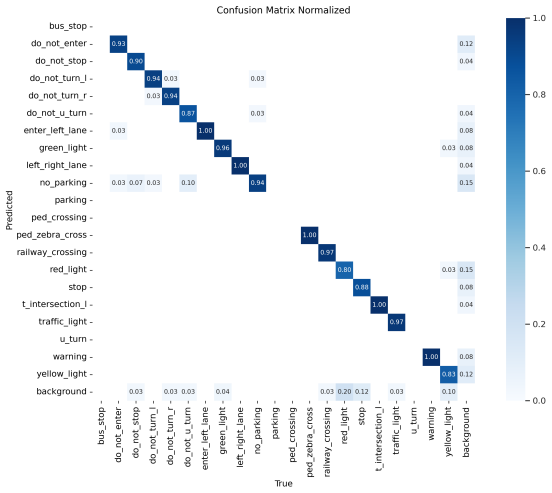


Fig. 19. YOLOv8.

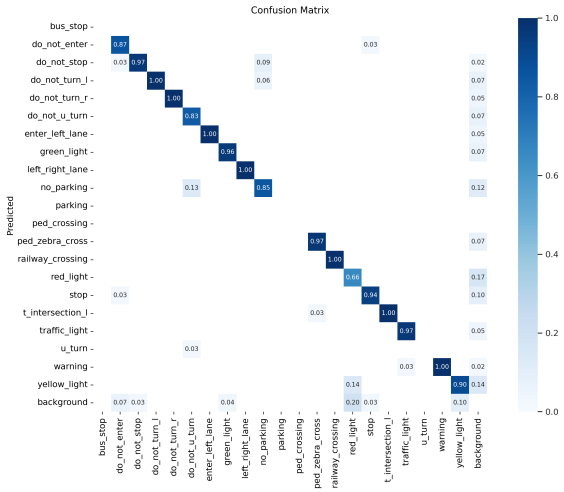


Fig. 20. YOLOv9.

E. Inference Result and Discussion

It is apparent from Fig. 24 to 28, that YOLOv9 followed by YOLOv8 have the highest values of accuracy scores of all the models in the paper. These models are also very accurate especially when it comes to place recognition which is very important with regards to road signs especially for safety reasons. However, on the tradeoff – YOLOv8 is slightly faster on the detection in comparison with YOLOv9. Therefore, if the need is to detect an object in real-time on the road, it may be desirable to use YOLOv8. However, there were other approaches in our experiments, such as RTMDet, Faster RCNN, and RetinaNet, slightly less effective in terms of speed but providing a good enough accuracy. They may be useful when speed is not the major factor into consideration as in case of roads of high importance or, self-driven cars. Besides, in

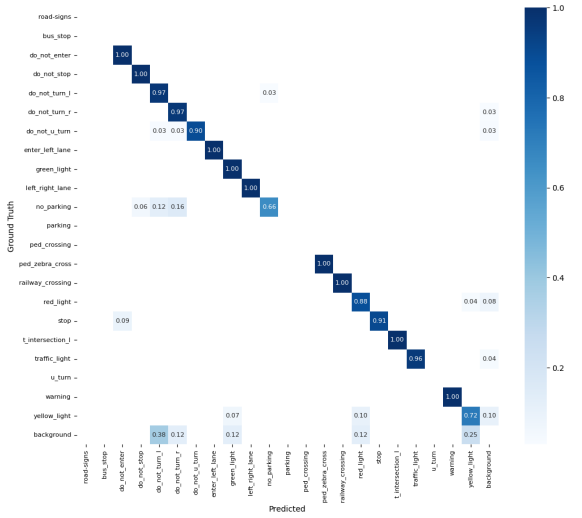


Fig. 21. Faster RCNN.

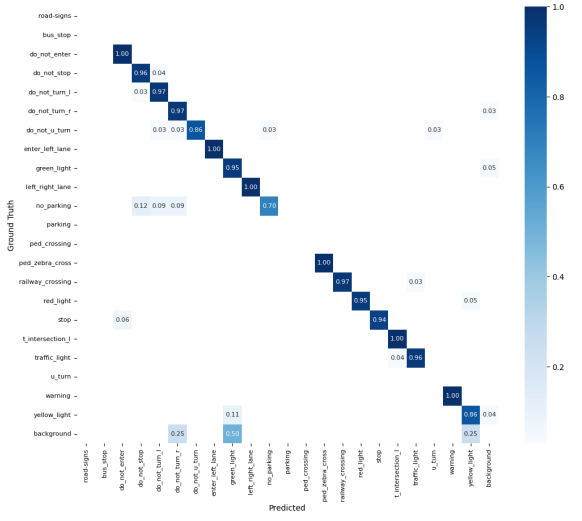


Fig. 22. RetinaNet.

terms of observing the curve of the total loss values, RetinaNet presents the best learning status in training so that it could develop more with more the training dataset. In conclusion, the decision of the specific model to be selected depends with the most valued criterium when it comes to the detection of road signs which can include speed or accuracy or both.

VI. Conclusion

In conclusion, this paper analyzes the performance of different object detection models which include YOLOv8, YOLOv9, RTMDet, RetinaNet and Faster-RCNN for road sign detection. YOLOv9 provides satisfactory results in terms of identification speed and accuracy measure; however, its inference time is relatively longer than the other versions including YOLOv8 with better real-time application performance. RTMDet achieves a better balance which very importantly is essential in scenarios where speed and accuracy are given utmost importance. Faster RCNN and RetinaNet have high enough accuracy, but the time for evaluation takes longer,

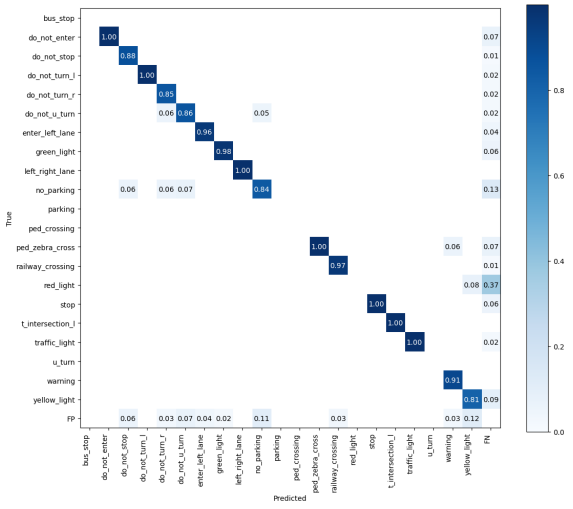


Fig. 23. RTMDet.

Fig. 24. YOLOv8.

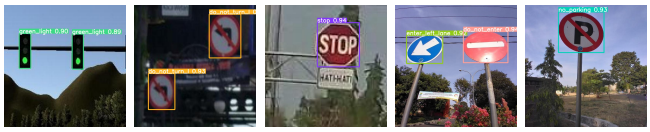


Fig. 25. YOLOv9.

Fig. 26. Faster RCNN.

Fig. 27. RetinaNet.

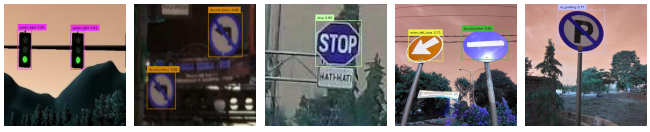


Fig. 28. RTMDet.

which may be useful for applications requiring a higher probability of correct result. The work underlines the parameters of accuracy, time and confidence in detecting the road contours

and signs that can help to choose the best option for the given application and to contribute to the creation of effecting road safety as well as autonomous driving systems. Our future work includes investigating more sophisticated ensemble approaches to achieve high accuracy and avoid overfitting problem, as well as to using more progressive data augmentation methodologies for increasing model's ability to generalize such as generative adversarial network (GAN) or self-supervised learning to real-world conditions and diverse environments, such as varying weather, lighting, and occlusions.

REFERENCES

[1] A. John and D. Meva, "A comparative study of various object detection algorithms and performance analysis," *International Journal of Computer Sciences and Engineering*, vol. 8, no. 10, pp. 158–163, 2020.

[2] N. Yadav and U. Binay, "Comparative study of object detection algorithms," *International Research Journal of Engineering and Technology (IRJET)*, vol. 4, no. 11, pp. 586–591, 2017.

[3] R. Padilla, W. L. Passos, T. L. Dias, S. L. Netto, and E. A. Da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics*, vol. 10, no. 3, p. 279, 2021.

[4] S. Srivastava, A. V. Divekar, C. Anilkumar, I. Naik, V. Kulkarni, and V. Pattabiraman, "Comparative analysis of deep learning image detection algorithms," *Journal of Big data*, vol. 8, no. 1, p. 66, 2021.

[5] Z. Fatima, S. Zardari, and M. H. Tanveer, "Advancing industrial object detection through domain adaptation: A solution for industry 5.0," *Actuators*, vol. 13, no. 12, 2024. [Online]. Available: https://www.mdpi.com/2076-0825/13/12/513

[6] P. Kaushik, A. Sharma, and B. Kaur, "Comparative analysis of object detection algorithms," in *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2022*. Springer, 2022, pp. 675–685.

[7] M. Soumyadeep, A. P. Singh, A. Sharma, and P. Kumar, "Real-time object detection and tracking using deep learning," in *2023 11th International Conference on Intelligent Systems and Embedded Design (ISED)*. IEEE, 2023, pp. 1–7.

[8] L. Tan, T. Huangfu, L. Wu, and W. Chen, "Comparison of yolo v3, faster r-cnn, and ssd for real-time pill identification," 2021.

[9] J.-H. Hwang, M. Lim, G. Han, H. Park, Y.-B. Kim, J. Park, S.-Y. Jun, J. Lee, and J.-W. Cho, "A comparative study on the implementation of deep learning algorithms for detection of hepatic necrosis in toxicity studies," *Toxicological Research*, vol. 39, no. 3, pp. 399–408, 2023.

[10] J. Wang, S. Jiang, W. Song, and Y. Yang, "A comparative study of small object detection algorithms," in *2019 Chinese control conference (CCC)*. IEEE, 2019, pp. 8507–8512.

[11] T. M. N. B. T. Rashid and L. M. Fadzil, "Comparative review of object detection algorithms in small single-board computers," *International Journal on Recent and Innovation Trends in Computing and Communication*, 2023.

[12] M. Asad, S. Khaliq, M. Yousaf, M. Ullah, and A. Ahmad, "Pothole detection using deep learning: a real-time and ai-on-the-edge perspective. adv civ eng 2022: 1–13," 2022.

[13] K. M. Krishna, A. Sowmya, D. Jerusha, and D. Susmitha, "Comparative study of vehicle detection using ssd and faster rcnn," 2021.

[14] S. Reddy, N. Pillay, and N. Singh, "Comparative study of convolutional neural network object detection algorithms for image processing," in *2023 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE, 2023, pp. 1–5.

[15] M.-h. Lee and H.-J. Mun, "Comparison analysis and case study for deep learning-based object detection algorithm," *Int. J. Adv. Sci. Converg*, vol. 2, no. 4, pp. 7–16, 2020.

[16] A. K. Shetty, I. Saha, R. M. Sanghvi, S. A. Save, and Y. J. Patel, "A review: Object detection models," in *2021 6th International Conference for Convergence in Technology (I2CT)*. IEEE, 2021, pp. 1–8.

[17] K. Li and L. Cao, "A review of object detection techniques," in *2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT)*. IEEE, 2020, pp. 385–390.

[18] F. Sultana, A. Sufian, and P. Dutta, "A review of object detection models based on convolutional neural network," *Intelligent computing: image processing based applications*, pp. 1–16, 2020.

[19] M. H. Tanveer, Z. Fatima, H. Mariam, T. Rehman, and R. C. Voicu, "Three-dimensional outdoor object detection in quadrupedal robots for surveillance navigations," *Actuators*, vol. 13, no. 10, 2024. [Online]. Available: https://www.mdpi.com/2076-0825/13/10/422

[20] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia computer science*, vol. 199, pp. 1066–1073, 2022.

[21] G. Lavanya and S. D. Pande, "Enhancing real-time object detection with yolo algorithm," *EAI Endorsed Transactions on Internet of Things*, vol. 10, 2024.

[22] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using yolo: Challenges, architectural successors, datasets and applications," *multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, 2023.

[23] M. N. Alhasanat, M. H. Alsafasfeh, A. E. Alhasanat, and S. G. Althunibat, "Retinanet-based approach for object detection and distance estimation in an image," *International Journal on Communications Antenna and Propagation (IRECAP)*, vol. 11, no. 1, pp. 1–9, 2021.

[24] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[25] S. Bhatlawande, S. Shilaskar, M. Agrawal, V. Ashtekar, M. Badade, S. Belote, and J. Madake, "Study of object detection with faster rcnn," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*. IEEE, 2022, pp. 1–6.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137–1149, 2016.

[27] C. Lyu, W. Zhang, H. Huang, Y. Zhou, Y. Wang, Y. Liu, S. Zhang, and K. Chen, "Rtmdet: An empirical study of designing real-time object detectors," *arXiv preprint arXiv:2212.07784*, 2022.

[28] Z. Huang, L. Li, G. C. Krizek, and L. Sun, "Research on traffic sign detection based on improved yolov8," *Journal of Computer and Communications*, vol. 11, no. 7, pp. 226–232, 2023.

[29] O. Renuka *et al.*, "A yolov8-based approach for multi-class traffic sign detection," *International Journal of Science and Research Archive*, vol. 11, no. 2, pp. 824–829, 2024.

[30] Y. Zhu and W. Q. Yan, "Traffic sign recognition based on deep learning," *Multimedia Tools and Applications*, vol. 81, no. 13, pp. 17 779–17 791, 2022.

[31] X. Gao, L. Chen, K. Wang, X. Xiong, H. Wang, and Y. Li, "Improved traffic sign detection algorithm based on faster r-cnn," *Applied Sciences*, vol. 12, no. 18, p. 8948, 2022.

[32] E. Güney and C. Bayılmış, "An implementation of traffic signs and road objects detection using faster r-cnn," *Sakarya University Journal of Computer and Information Sciences*, vol. 5, no. 2, pp. 216–224, 2022.

[33] M. Zhu, H. Yang, Z. Cui, and Y. Wang, "Traffic sign detection and recognition for autonomous driving in virtual simulation environment," in *International Conference on Transportation and Development 2022*, 2019, pp. 12–18.

[34] J. Chen, K. Jia, W. Chen, Z. Lv, and R. Zhang, "A real-time and high-precision method for small traffic-signs recognition," *Neural Computing and Applications*, vol. 34, no. 3, pp. 2233–2245, 2022.

[35] B. Senthilnayaki, C. Rajeswary, G. Nivetha, P. Dharanyadevi, G. Mahalakshmi, and A. Devi, "Traffic sign prediction and classification using image processing techniques," in *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*. IEEE, 2022, pp. 1–5.

[36] A. Aneesh, L. Shine, R. Pradeep, and V. Sajith, "Real-time traffic light detection and recognition based on deep retinanet for self driving cars," in *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, vol. 1. IEEE, 2019, pp. 1554–1557.

[37] L. Jiang, H. Liu, H. Zhu, and G. Zhang, "Improved yolo v5 with balanced feature pyramid and attention module for traffic sign detection," in *MATEC Web of Conferences*, vol. 355. EDP Sciences, 2022, p. 03023.

[38] M. Sohan, T. Sai Ram, R. Reddy, and C. Venkata, "A review on yolov8 and its advancements," in *International Conference on Data Intelligence and Cognitive Informatics*. Springer, 2024, pp. 529–545.

[39] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616*, 2024.

[40] R. Del Prete, M. D. Graziano, and A. Renga, "Retinanet: A deep learning architecture to achieve a robust wake detector in sar images," in *2021 IEEE 6th International Forum on Research and Technology for Society and Industry (RTSI)*. IEEE, 2021, pp. 171–176.

Accuracy Optimization and Wide Limit Constraints of DC Energy Measurement Based on Improved EEMD

Xiaoyu Wang*, Xin Yin, Xinggang Li, Jiangxue Man, Yanhe Liang, Fan Xu
State Grid Heilongjiang Electric Power Co. Ltd. Marketing Service Center; Power Metering Center, Harbin, 150000, China

***Abstract*—In modern power systems, with the increasing application of renewable energy, direct current transmission technology has put forward new requirements for energy metering. In order to solve the accuracy problem of traditional electric energy metering under DC energy, the research is based on the classical empirical modal decomposition (EEMD), and introduces the artificial chemical reaction optimization algorithm (ACROA) to enhance the global search capability and decomposition accuracy of the original algorithm, and at the same time safeguards the accuracy of metering equipment under extreme conditions through the wide quantitative constraints, and ultimately puts forward a new type of optimization model for the accuracy of DC electric energy metering. The highest measurement accuracy of this model could reach 90%, and it performed better in power signal decomposition and accuracy optimization. Especially under high-frequency interference and complex signal conditions, the measurement error could be reduced to 6.87%, the highest decomposition stability was 94.02%, and the shortest measurement time was 1.12 seconds. Therefore, the model constructed in this study exhibits excellent decomposition accuracy and robustness in complex energy environments, solving the shortcomings of traditional energy metering methods and providing new ideas for future optimization of DC energy metering.**

***Keywords*—EEMD; direct current energy; measurement; width limit; ACROA**

I. Introduction

In modern power systems, with the transformation of energy structure, Direct Current (DC) transmission technology has been increasingly widely used, especially in renewable energy generation such as wind power, photovoltaic power generation, and electric vehicle charging piles, where the demand for direct current energy is gradually increasing [1]. This trend has put forward new requirements for Direct Current Energy Metering (DCEM) technology. Existing power metering equipment is mainly designed for AC power grids, with low metering accuracy and stability in DC environments, making it difficult to cope with complex signal conditions and extreme operating conditions, such as significant increase in metering error under high voltage and high current, and insufficient decomposition capability when disturbed by noise and non-stationary signals [2]. Therefore, there is an urgent need to explore optimization methods that can maintain high-precision metering in complex DC energy environments. The study aims to solve the core problem in DC energy metering, which is how to ensure the accuracy, stability and adaptability of metering equipment under complex working conditions and extreme conditions. To this end, the study proposes a metering accuracy optimization model that combines the improved EEMD with the artificial chemical reaction optimization algorithm, and further improves the adaptability and reliability of the model in different voltage and current ranges by introducing a wide-volume-limit constraint. The objective of the research is to improve the decomposition accuracy and robustness of DC energy signals, and to significantly reduce the metering error and computation time by improving the signal decomposition and optimization algorithm. The importance and significance of the study is to break through the limitations of the traditional DC energy metering methods, to provide an efficient, stable and adaptable metering technology for DC transmission and renewable energy, to provide new ideas for the optimization of power metering under complex working conditions in future smart grids, and to provide important theoretical and practical references for the researchers in the related fields to explore the signal processing methods based on EEMD.

II. Related Work

For power metering in DC environment, many researchers at home and abroad have successively explored the technology and proposed many accuracy optimization methods. Buchibabu et al. proposed a comprehensive control strategy to better manage the storage of DC microgrids such as AC power grids, by combining tuna swarm optimization algorithms, which could effectively improve the level of various DCEMs in microgrids [3]. Gao J et al. proposed an energy compensation algorithm limited to DC microgrids using intelligent optimization algorithms, whose error was also smaller than the traditional average EM algorithm, and the real-time power curve was closer to the theoretical value [4]. To further improve the accuracy of DCEM, Liaqat R et al. constructed a signal decomposition model using event matching energy decomposition algorithm after receiving non-invasive load monitoring, which could effectively distinguish various types of noise in electrical energy, thereby improving computational accuracy [5]. Kumar G et al. used optimized tracking algorithms to process data on appliance usage frequency, preferred operating interval, and average power consumption to better calculate household energy consumption within microgrids. This method could significantly improve the accuracy of electricity calculation and reduce time costs. But at this time, the usage cost and computational complexity actually increased, which had an impact on the overall calculation [6].

However, most of these methods are optimized for specific scenarios, which makes it difficult to maintain high accuracy under complex or extreme electrical energy signals, and they are deficient in high-frequency noise processing and real-time optimization. In addition, currently, Empirical Mode Decomposition (EMD) and its improved algorithms, such as Ensemble Empirical Mode Decomposition (EEMD), have gradually attracted the attention of researchers due to their unique advantages in signal decomposition [7]. Liang C et al. proposed a control strategy combining EEMD to improve the power time series regulation accuracy of DC microgrid photovoltaic power generation and its hybrid energy storage system, which could achieve better control effects under different power fluctuation characteristics [8]. Jiang L et al. proposed a wavefront calibration method to accurately locate the fault location in DC distribution systems by combining EEMD and singular value decomposition algorithms. This method could accurately decompose DC power signals and had high accuracy in fault location [9]. Zhang N et al. proposed a novel accuracy prediction model for short-term photovoltaic power generation by combining EEMD and gated recursive units. The prediction accuracy and robustness of this model are superior [10]. Wang et al. proposed a novel fault localization strategy by combining EEMD and adaptive local mean decomposition methods. This method detected DC series arc faults in less than 1ms with an accuracy of 98.75% [11].

In summary, previous studies have made many useful explorations in improving the accuracy of DC energy metering and processing complex signals. However, these researches still have some shortcomings when facing extreme working conditions, such as high voltage and high current, for example, the accuracy degradation during signal decomposition and low arithmetic efficiency. For this reason, how to develop a DC energy metering method that is efficient, robust and applicable to complex working conditions has become a key issue to be solved. The research focuses on solving the lack of metering accuracy of traditional methods, and by improving the EEMD algorithm and introducing the Artificial Chemical Reaction Optimization Algorithm (ACROA), a new type of DC energy metering model is proposed, and at the same time, the metering equipment is enhanced by the wide quantitative constraints under extreme working conditions, such as high voltage and high current. The research objectives include improving the decomposition accuracy of complex signals, reducing the metering error, optimizing the real-time performance of the model, and enhancing the stability of the model under the wide quantitative constraints. The importance of this study is that it makes up for the shortcomings of the existing methods and provides theoretical support and technical guarantee for the DC transmission system in the future smart grid and energy internet, which is of great theoretical and engineering significance.

III. Methods and Materials

A. DC Signal Decomposition

Due to the complexity of DC electrical signals, traditional measurement methods are difficult to cope with non-stationary, nonlinear, and the superposition of various interference signals [12]. There are many factors that affect the error of EM, and to minimize the error in DCEM, the primary requirement is to ensure the stability of the data acquisition source. DC energy meters are core devices used to measure DC, voltage, and related electrical parameters [13]. It can directly complete DCEM independently, and Fig. 1 shows the working principle of a DC energy meter.

In Fig. 1, firstly, the current sampling and voltage sampling modules collect the load current I and load voltage U. The collected current signal is used for power calculation with the voltage signal through a multiplier, and then converted into a frequency signal through a power frequency converter. Next, the frequency divider processes the frequency signal for subsequent calculation and display. The signal processing unit calculates based on the divided data and transmits the final electrical energy information to the counting display module. In this process, the main challenges of DCEM are as follows: traditional measurement methods are difficult to effectively handle these complex signals; The range of the measuring device is limited, and when extreme conditions are encountered, the measuring equipment often experiences over range problems, leading to increased measurement errors; DCEM equipment is susceptible to environmental interference during long-term operation, leading to fluctuations in measurement results [14-15]. Taking into account the above factors, this study assumes the existence of an ideal DCEM. Fig. 2 shows the classification of DC electrical signals at this time.

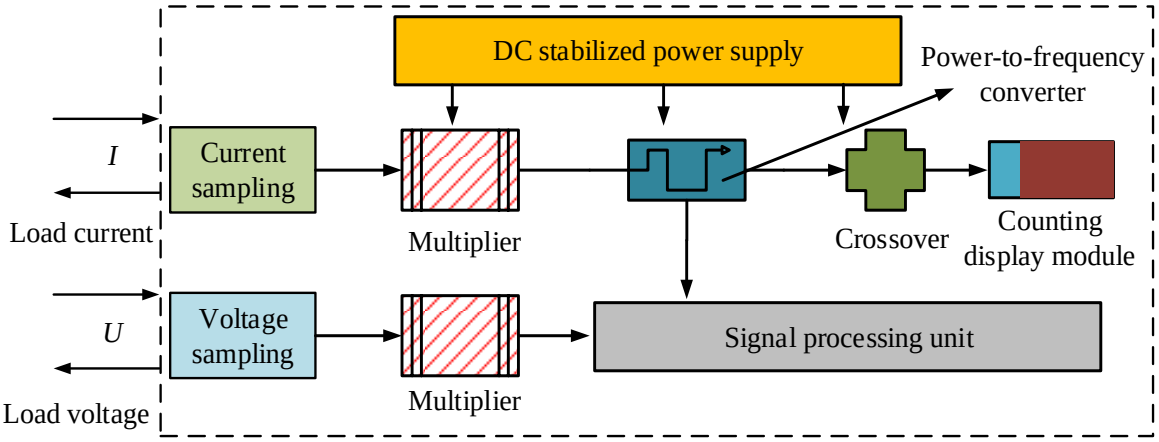


Fig. 1. DC energy meter working principle diagram.

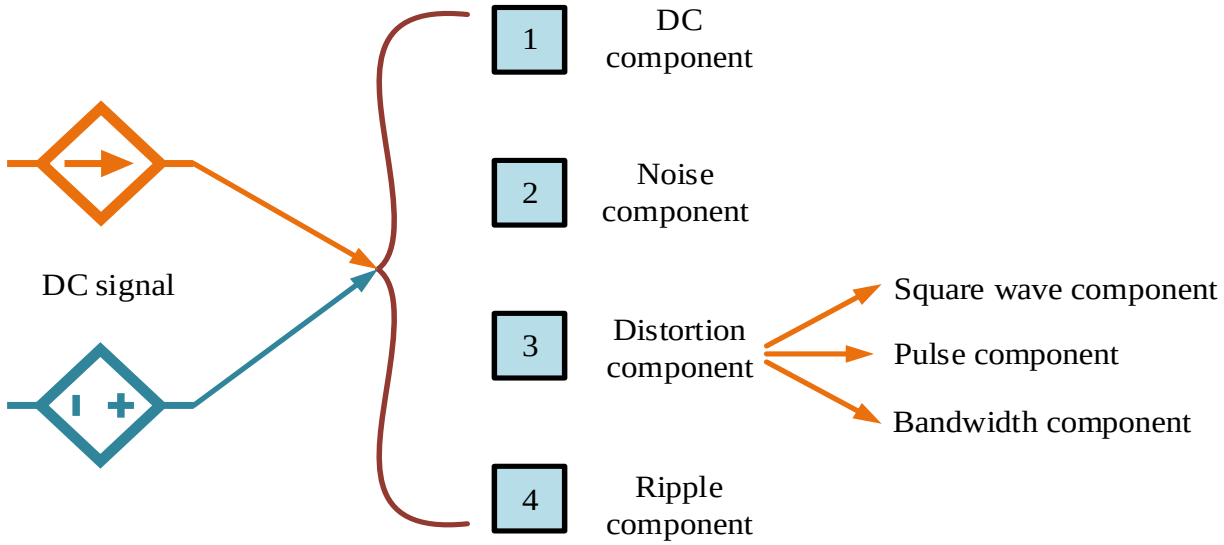


Fig. 2. DC signal composition classification.

In Fig. 2, DC electrical signals are DC components, noise components, distortion components, and ripple components. The DC component represents the stable part of the signal and is the main carrier of energy transmission; the noise component is mainly generated by random interference brought by external environment or equipment, which may include the influence of external factors such as temperature and humidity [16]. The expression of the DC electrical signal at this time is shown in Eq. (1).

$$u(t) = u_0 + \sum_n u_n(t) + \sum_i u_i(t) + \sum_k u_k(t) \quad (1)$$

In Eq. (1), $u(t)$ represents the DC electrical signal; $\sum_n u_n(t)$ represents the noise component; $\sum_i u_n(t)$ represents the distortion component; $\sum_k u_k(t)$ represents the ripple component. The ideal expression of the DC component is shown in Eq. (2).

$$u_1(t) = k(t) \quad (2)$$

In Eq. (2), k represents a constant; $u_1(t)$ represents the DC component in the ideal state. The noise component in the ideal state is usually avoided by adding noise compensation or filtering devices. At this time, the white noise signal with a specific noise ratio is composed as shown in Eq. (3).

$$u_2(t) = n(t) \quad (3)$$

In Eq. (3), n represents the white noise signal in the noise ratio; $u_2(t)$ represents the noise component in the ideal state. The distortion component originates from the use of nonlinear loads, causing signal deformation. For example, the fluctuating charge and discharge signals generated when the load is involved or the power grid system is disconnected [17]. Common distortion signals such as square wave components and broadband components are expressed in Eq. (4).

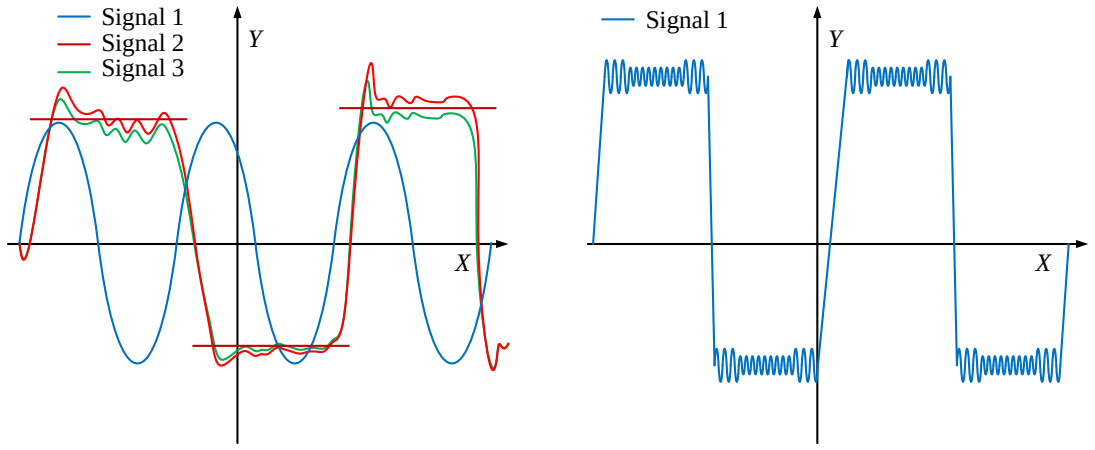
$$\begin{cases} u_3(t) = u_1(t) \pm A \times k_1(t) & 0 \le t \le t_1 \\ u_4(t) = u_1(t) - B(t - t_2)^2 + b_1(t) & t_1 \le t \le t_2 \end{cases} \quad (4)$$

In Eq. (4), A represents the amplitude multiple of the signal component; t represents time, where t_1 and t_2 represent the start and end times, respectively; k_1 represents the time period t_1 to t_2 control constant. The ripple component reflects the periodic interference caused by power supply ripple, which is usually characterized by high-frequency and small amplitude periodic fluctuations. The expression of the ripple component is shown in Eq. (5).

$$u_5(t) = D \times u_1(t) \times \sum_{t=1}^{n} \cos(2\pi i \omega_1 t), 0 < D < 0.1 \quad (5)$$

In Eq. (5), ω_1 represents a constant; ω_1 represents frequency. However, in the process of decomposing DC electrical signals, due to the discontinuity and complexity of the signal, Gibbs phenomenon is prone to occur, that is, oscillation occurs at the high-frequency components or abrupt points of the signal, resulting in overshoot or fluctuation during signal reconstruction, which affects the accuracy of measurement [18]. This phenomenon is particularly evident when dealing with non-stationary and nonlinear signals. Fig. 3 is a typical schematic diagram of Gibbs phenomenon.

Both Fig. 3 (a) and Fig. 3 (b) show the time-domain waveform of Gibbs phenomenon. In Fig. 3, after Fourier series expansion of periodic functions with discontinuous points, finite terms are selected for synthesis. As the number of selected items increases, the peak in the synthesized waveform gradually approaches the discontinuity point of the original signal. When the number of terms is sufficient, the peak tends towards a constant, approximately 9% of the total jump variable.



(a) The Gibbs phenomenon1 (b) The Gibbs phenomenon2

Fig. 3. Example of the Gibbs phenomenon.

B. DCEM Accuracy Optimization and WLC Based on Improved EEMD

To overcome the Gibbs phenomenon, EEMD is introduced in this study. EEMD can effectively reduce the influence of Gibbs phenomenon in signal decomposition, improve the accuracy of signal reconstruction, and exhibit better robustness and stability in various complex signal environments [19]. Fig. 4 shows the process of EEMD.

In Fig. 4, firstly, white noise of different intensities is added to the original signal and stacked multiple times. Next, each superimposed signal is subjected to empirical mode decomposition to obtain a series of Intrinsic Mode Functions (IMF). Then, all IMF components are averaged to eliminate uncertainty introduced by noise. However, as a signal decomposition method with non Fourier transform, EEMD may still face the problem of insufficient decomposition accuracy when dealing with specific complex signals. Therefore, ACROA is introduced in this study to further optimize the decomposition process. ACROA first generates a set of reactants through initialization. Secondly, in the subsequent iteration process, different chemical reaction operators are dynamically selected and operated based on the current signal state and optimization requirements to adapt to different signal conditions [20]. Then, after each reaction, the reactants are updated based on feedback, similar to a reversible reaction process. This process involves redox reactions, decomposition reactions, displacement reactions, and synthesis reactions. Fig. 5 shows an example of the reaction.

In Fig. 5 (a), the redox reaction mechanism optimizes the global search capability of the signal by adjusting the state of the reactants to change their energy levels; In Fig. 5 (b), the decomposition reaction decomposes complex signals into smaller units, thereby improving the local search accuracy of the algorithm; In Fig. 5 (c), the substitution reaction enables the algorithm to quickly jump out of local optima and improve global search efficiency through the exchange of reactants; In Fig. 5 (d), the synthesis reaction helps the algorithm find a better solution by combining multiple signal components. At this point, the optimized DC power signal decomposition is shown in Eq. (6).

$$X(t) = \sum_{i=1}^{n} IMF_i(t) + r_n(t) \tag{6}$$

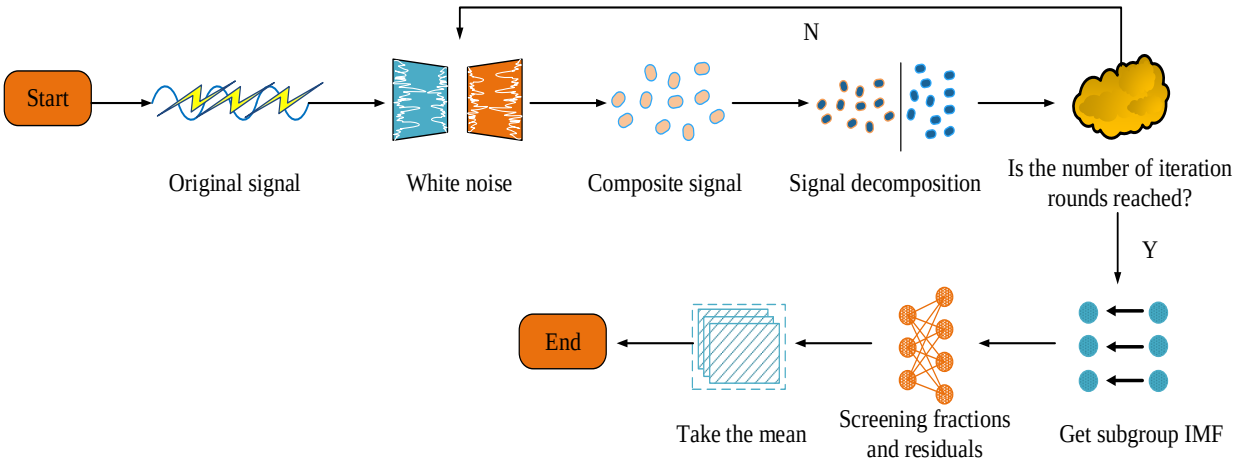


Fig. 4. EEMD process schematic.

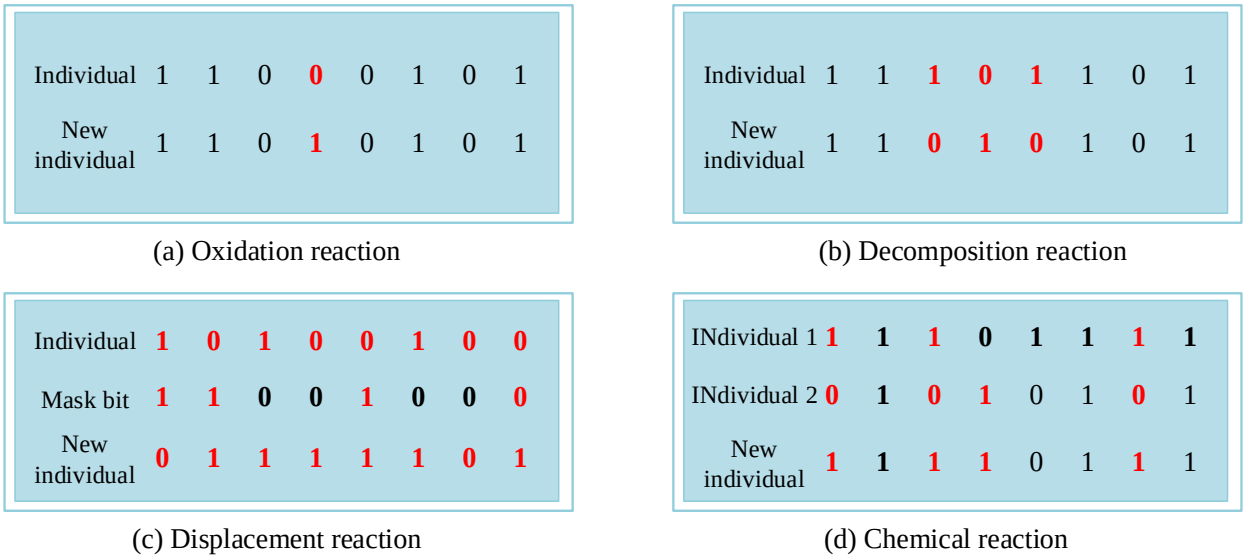


Fig. 5. Schematic of ACROA's four binary coded responses.

In Eq. (6), $X(t)$ represents the original DC electrical signal; $X(t)$ represents the -i th eigenmode function; $r_n(t)$ represents the residual term, which is the remaining part after signal decomposition; n represents the number of decomposed eigenmode functions. After obtaining these IMF components through the EEMD algorithm, ACROA is used to further optimize the accuracy of the residual term $r_n(t)$ to reduce decomposition errors. The calculation equation is shown in Eq. (7).

$$r_{opt}(t) = \min_{\alpha}(r_n(t) - \alpha\Delta r(t)) \tag{7}$$

In Eq. (7), $r_{opt}(t)$ represents the residual term after optimization; α represents the optimal weight coefficient obtained during the ACROA optimization process; α represents the adjustment amount during the ACROA iteration process, which is the correction value obtained through the reaction mechanism optimization in each iteration. The combination of the two results in the final optimization signal $X_{opt}(t)$ calculation equation is shown in Eq. (8).

$$X_{opt}(t) = \sum_{i=1}^{n} IMF_i(t) + r_{opt}(t) \tag{8}$$

In addition, each frequency component in the signal may have different characteristics within different limits. To ensure the stability and reliability of the energy meter in the face of extreme conditions, this study has imposed constraints on the Wide Dynamic Range (WDR). Among them, WDR refers to a large range that can be accurately measured or processed in measuring or measuring equipment, that is, the equipment can maintain high accuracy and reliability over a wide range of input signal amplitudes. For DCEM, WDR usually means that the measuring equipment can maintain measurement accuracy without significant errors over a wide range of changes from low voltage and low current to high voltage and high current. The WDR constraint is shown in Eq. (9).

$$L_{\min} \le X_{opt}(t) \le L_{\max} \tag{9}$$

In Eq. (9), $L_{\min}$ and $L_{\max}$ represent the upper and lower limits of the measured DC signal, respectively. The constraint error calculation equation at this time is shown in Eq. (10).

$$\varepsilon(t) = \left| X(t) - X_{opt}(t) \right| \le \delta \tag{10}$$

In Eq. (10), $\varepsilon(t)$ represents the error between the measured signal and the optimized signal; $X(t)$ represents the actual measured DC signal; δ represents the maximum allowable error limit. To adapt to the dynamic changes in the measurement environment and ensure that the signal can be adaptively adjusted under WDR, the dynamic range expression of WDR at this time is shown in Eq. (11).

$$X_{opt}(t) = \max(\frac{X(t)}{L_{\max}}), and \min(\frac{X(t)}{L_{\min}}) \tag{11}$$

This study combines EEMD-ACROA and WDR to propose a novel DCEM model. Fig. 6 shows the process of the model.

In Fig. 6, the DC electrical signal is first collected and preprocessed through current and voltage sampling modules to ensure signal stability. Then, the DC electrical signal decomposition model is constructed, and the signal is divided into DC components, noise components, distortion components, and ripple components to reflect the main characteristics of the signal. Then, the EEMD algorithm is used to decompose the signal, obtain the intrinsic mode functions of different frequencies, and optimize the residual terms through the ACROA algorithm to dynamically adjust the reaction path and reduce decomposition errors. WLC is introduced to ensure that equipment maintains high-precision measurement under extreme conditions such as high voltage and high current. The signal processing process is optimized through an error control mechanism, and the optimized electrical energy data is transmitted to the display module for real-time monitoring and accurate measurement.

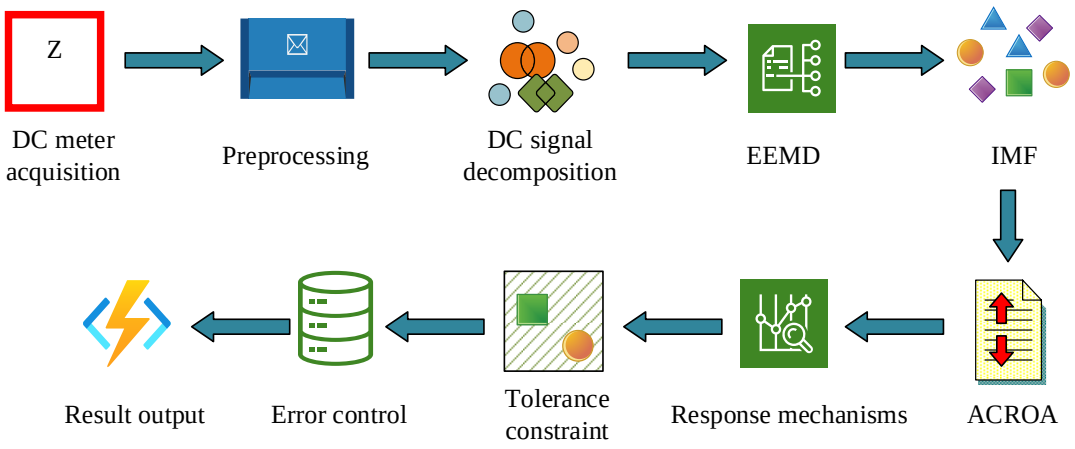


Fig. 6. Novel DCEM modeling process.

IV. Result

A. Performance Testing of the New DCEM Model

The experimental hardware was Keysight E36312ADC power supply, with an output range of 0-100V and a maximum current of 10A. The load simulator was Chroma 63206A-150-600, which supported a maximum of 600W. The precision current sensor was LEM HAZ 1000-S, with a current range of 0-1000A and an accuracy of ±0.1%. The data was collected through the National Instruments NI USB-6343 high-precision data acquisition card. The software part was Dell Precision 5820 Tower, equipped with Intel Core i9-10900X processor, 64GB DDR4 memory, 2TB SSD hard drive, and NVIDIA Quadro RTX 4000 GPU. Python 3.10 was used for data analysis and visualization. The IEEE PES Distribution Test Feeder Dataset (IEEE PES) and Electric Power Consumption Dataset (EPC) datasets were the sources of test data. IEEE PES included various typical electrical energy signals in distribution systems, especially voltage and current waveform data for different loads and grid conditions; EPC contained a series of energy consumption data for household users, including current, voltage, power, and other related information. This study first conducted ablation testing on the proposed algorithm model. Fig. 7 shows the test results.

In Fig. 7 (a), after 200 iterations, the measurement accuracy of EEMD-ACROA-WDR continued to remain at a high level and reaches nearly 90% accuracy at 600 iterations. It performed even better in terms of global optimization capability and robustness. In Fig. 7 (b), EEMD-ACROA-WDR achieved a measurement accuracy of over 90% after 300 iterations. Although EEMD-ACROA performed well in the early iterations, there were significant fluctuations in the later stages, and the measurement accuracy did not steadily improve. Overall, the modules of EEMD-ACROA-WDR demonstrated stronger adaptability and stability on both datasets. The study introduced advanced DCEM precision optimization methods for comparison, namely Bispectral Analysis (BA), Extreme Learning Machine (ELM), and Adaptive Local Mean Decomposition (ALMD). The signal decomposition rate was used as the indicator to test the power processing capability under different bandwidth ranges, as shown in Fig. 8.

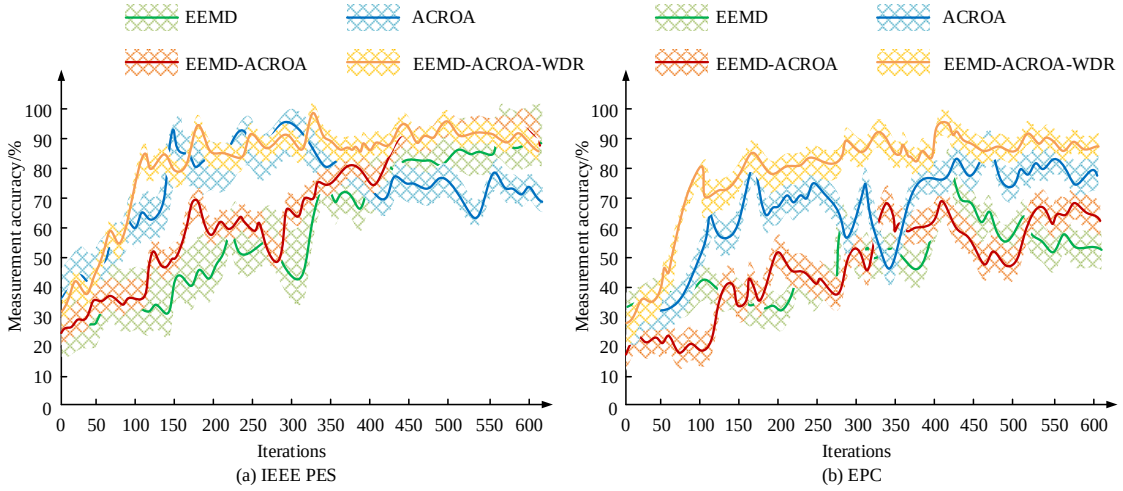


Fig. 7. Ablation test results.

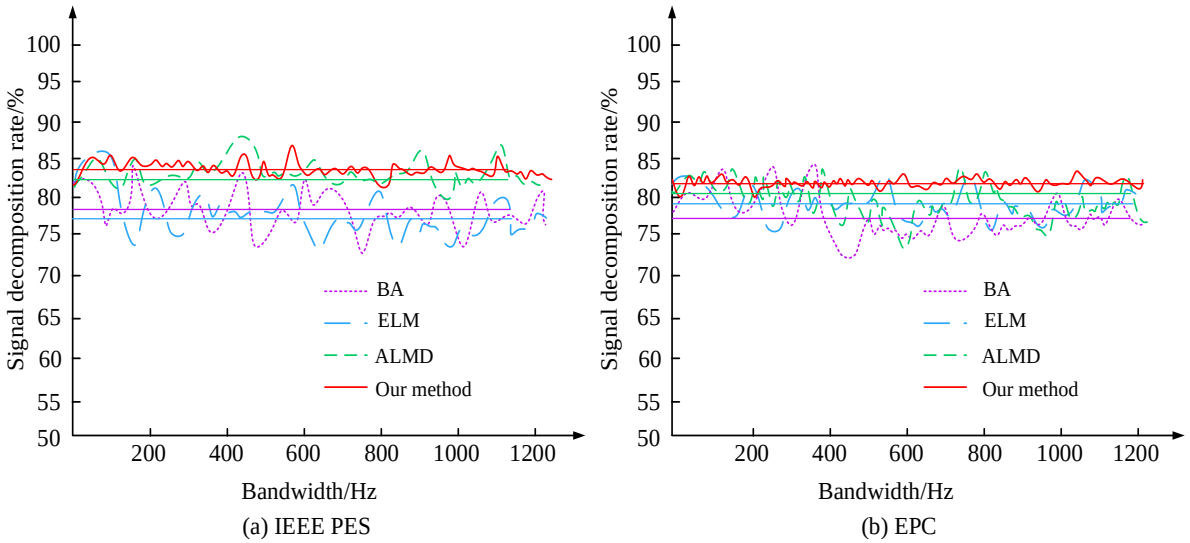


Fig. 8. Signal decomposition rate test results of different methods.

In Fig. 8, within different bandwidth ranges, the proposed method maintained a signal decomposition rate of around 85% with minimal fluctuations at higher bandwidths, such as 800Hz to 1200Hz, demonstrating good stability. The decomposition rates of BA and ELM showed significant fluctuations at high bandwidth, especially the decomposition rate of BA method was relatively unstable throughout the entire bandwidth range, with an average decomposition rate of only 78%. Overall, the proposed optimization algorithm exhibited stronger adaptability and decomposition ability in processing high bandwidth signals, especially maintaining high signal decomposition rates on IEEE PES and EPC, verifying the effectiveness of the algorithm in complex signal environments. In Table I, the precision, recall, F1 value, and average time for precision optimization in DCEM were used as indicators for this study.

TABLE I. INDICATOR TEST RESULTS FOR EACH METHOD

Data Set	Method	P/%	R/%	F1/%	Average Time Spent/S
IEEE PES	BA	87.63	84.12	85.88	2.11
	ELM	84.25	85.36	84.81	2.03
	ALMD	89.74	87.15	88.45	1.54
	Our method	90.11	89.63	89.87	1.02
EPC	BA	88.53	85.16	86.85	2.12
	ELM	89.68	86.84	88.26	1.76
	ALMD	90.08	88.45	89.27	1.52
	Our method	90.87	89.74	90.31	1.08

In Table I, on IEEE PES, the proposed method had the highest P-value of 90.11%, R-value of 89.63%, and F1 value of 89.87%, all of which were higher than other methods, indicating its higher accuracy and stability in signal processing and EM aspects. Meanwhile, the average time was 1.02 seconds, which was much lower than other methods, reflecting the efficiency advantage of this algorithm. In contrast, although ALMD performs better in accuracy and F1 score, it took 1.54 seconds and had a relatively slower processing speed. BA and ELM were slightly lower than the proposed method in various indicators, especially in terms of time. BA took 2.11 seconds, which was more time-consuming. For EPC, the proposed method achieved a P-value of 90.87%, an R-value of 89.74%, and an F1 value of 90.31%, demonstrating excellent performance with a time of 1.08 seconds and a significant advantage in speed.

B. New DCEM Model Simulation Testing

This study set the amplitude of the original current signal to 2A and simulated the input of electrical energy signals under different operating conditions. The simple EEMD model and the proposed method were compared and tested for signal decomposition to evaluate the robustness and decomposition accuracy of different models in processing complex DC signals, as shown in Fig. 9.

In Fig. 9 (b) - Fig. 9 (d), the decomposed electric energy signals IMF1, IMF2, and IMF3 of the EEMD model exhibited significant oscillations and noise, especially in IMF1 and IMF3, where multiple irregular pulse points appear, which had a negative impact on the stationarity and decomposition accuracy of the signal. This indicated that EEMD still faced certain noise aliasing problems when processing complex electric energy signals. The proposed method decomposed IMF1, IMF2, and IMF3 to be smoother, with significantly reduced oscillation amplitude, indicating that this method had better performance in noise suppression and signal smoothing. Especially in the decomposition of IMF3, the proposed method effectively eliminated high-frequency noise and irregular fluctuations in the original signal, preserving the main features of the signal. This indicated that the proposed method could better capture the true characteristics of signals when dealing with non-stationary signals and high-frequency interference, improving the accuracy and stability of DCEM. In Fig. 10, this study conducted signal accuracy tests on EEMD before and after improvement.

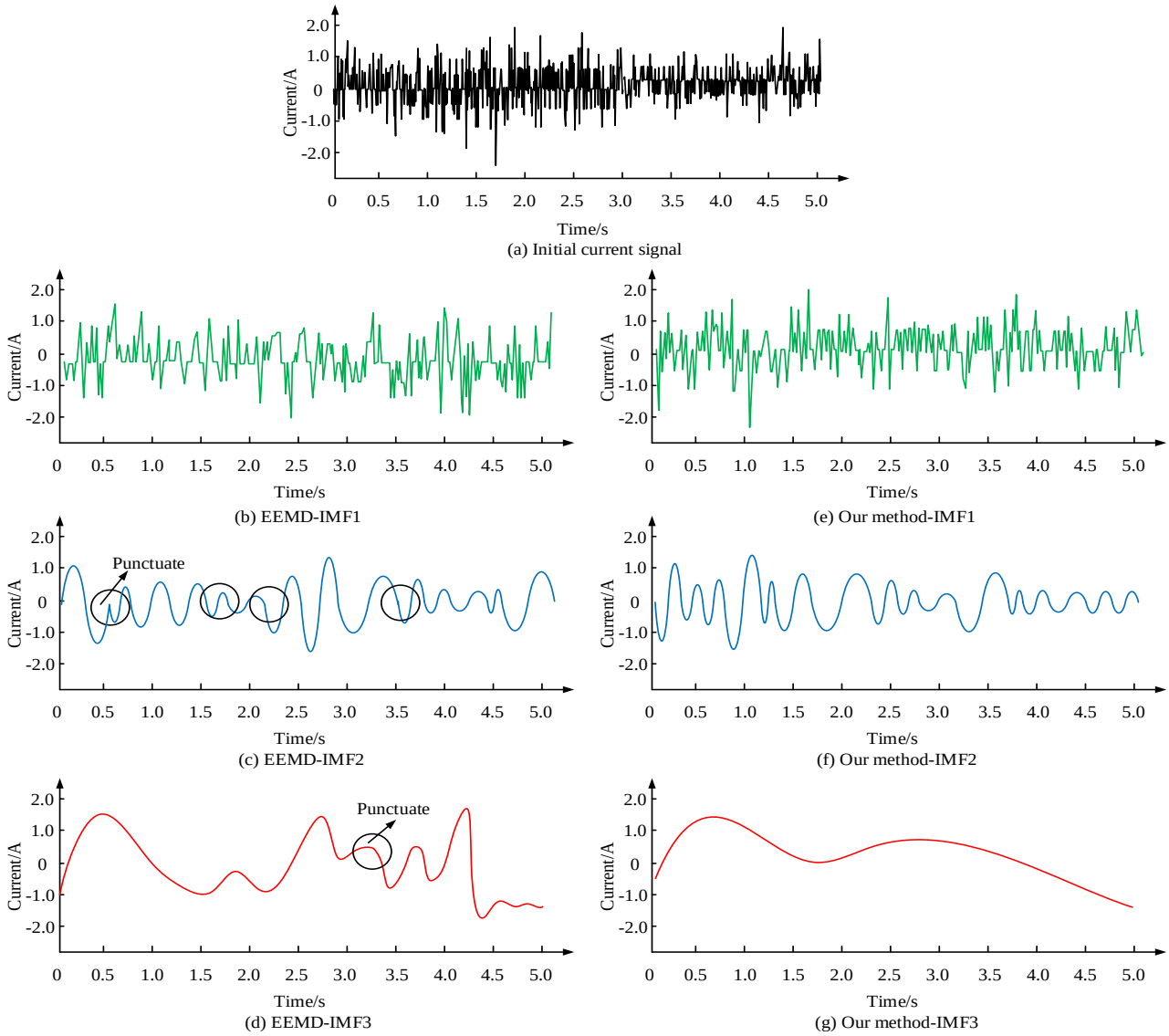


Fig. 9. Decomposition testing of electrical energy signals for two types of models.

In Fig. 10 (a) -10 (c), the measured values of EEMD in the IMF1 signal deviate significantly from the true values and fluctuated greatly, indicating that EEMD had insufficient accuracy in processing high-frequency signals. In the IMF2 and IMF3 signals, although the matching between the measured values and the standard values improved, there were still significant errors, especially at certain peak points, indicating that the EEMD method also had certain errors in processing low-frequency signals. In Fig. 10 (d) -10 (f), the measured values almost completely coincided with the standard values in the IMF2 and IMF3 signals, indicating that the proposed method outperformed traditional EEMD in processing low-frequency signals. Meanwhile, the error of IMF1 signal was effectively controlled, demonstrating better signal reconstruction capability. Overall, the proposed method performed better in terms of measurement accuracy and stability, especially when dealing with complex signals, with significantly reduced errors, verifying the practical application effect of the model in EM. In Table II, this study tested the number of IMF decompositions, measurement error, decomposition stability, and computation time as indicators.

In Table II, in terms of the number of IMF decompositions, the proposed method was the same as methods such as BA and ALMD, both of which decomposed into 6 IMF components, indicating its high decomposition ability in processing complex electrical energy signals. In terms of measurement error, the proposed method had an error of 6.87%, significantly lower than other methods, indicating higher accuracy in EM and stronger robustness, especially when dealing with noise and non-stationary signals. In terms of decomposition stability, the proposed method had a decomposition stability of up to 94.02%, which was better than the 92.48% of Wang L et al.'s method, demonstrating stability and consistency in signal decomposition, greatly reducing error fluctuations. In terms of computation time, the proposed method took 1.12 seconds, which was the fastest among all methods, far lower than the BA method's 2.34 seconds, demonstrating its high efficiency in practical applications.

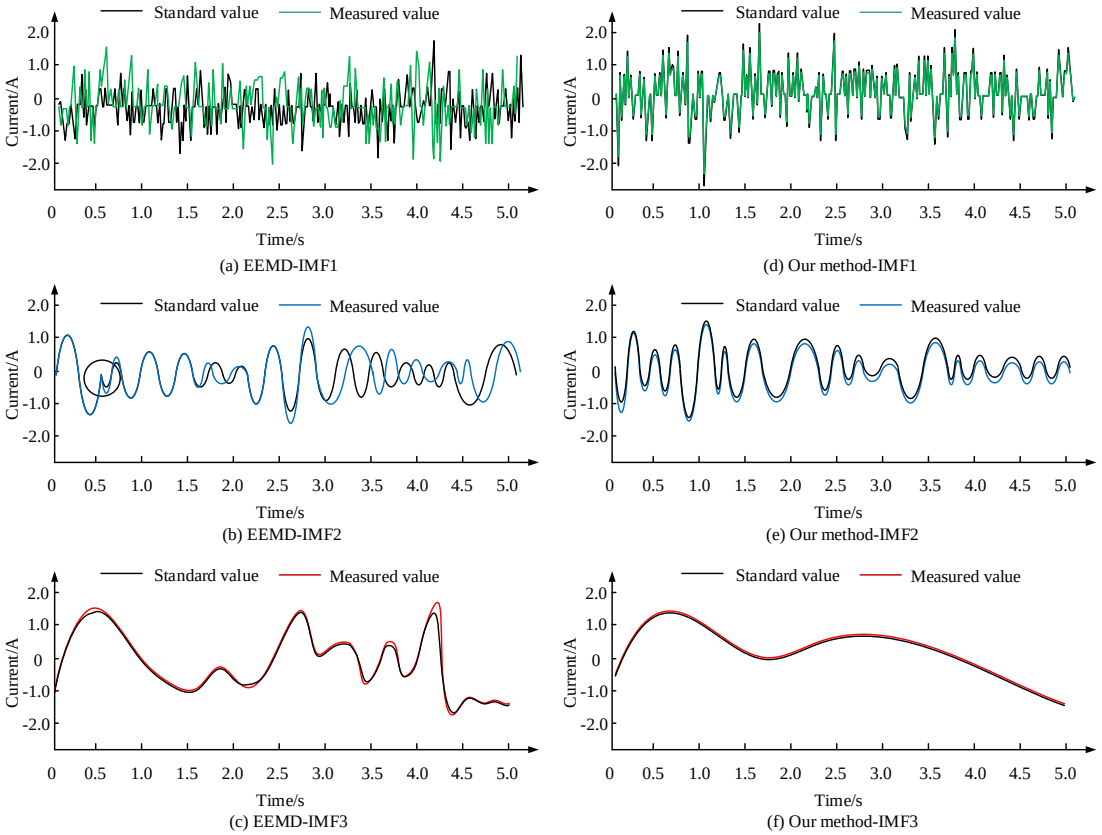


Fig. 10. Results of the comparison of the measurement errors of the two types of methods.

TABLE II. MULTI-INDICATOR TEST RESULTS FOR DIFFERENT MEASUREMENT METHODS

Method	Number of IMF Decompositions	Measurement Error/%	Decomposition Stability/%/%	Operation Time/S
BA	6	12.34	85.56	2.34
ELM	5	10.89	87.12	1.89
ALMD	6	9.78	88.34	1.52
EEMD	7	11.56	86.47	1.74
The method proposed by Liang C et al.	6	8.98	89.65	1.62
The method proposed by Jiang L et al.	5	8.45	90.12	1.58
The method proposed by Zhang N et al.	7	7.89	91.05	1.43
The method proposed by Wang L et al.	6	7.34	92.48	1.32
Our method	6	6.87	94.02	1.12

V. CONCLUSION

In response to the problems of low accuracy of DCEM and insufficient robustness of traditional methods in complex electrical signals, this study proposes a DCEM accuracy optimization method that combines improved EEMD and ACROA, and improves the adaptability of measuring equipment in extreme working conditions by introducing WLC. The proposed model could achieve an accuracy of nearly 90% in the lowest 300 iterations. Compared with the BA, ELM, and ALMD models, the signal decomposition rate of the proposed model was significantly better than other methods, especially under higher bandwidth conditions such as 800Hz to 1200Hz. Its signal decomposition rate was as high as nearly 85%, with the highest P value of 90.87%, the highest R value of 89.74%, and the highest F1 value of 90.31%. The average optimization time for accuracy was as short as 1.02 seconds. The proposed model decomposed IMF1, IMF2, and IMF3 to be smoother, with significantly reduced oscillation amplitude, indicating better performance in noise suppression and signal smoothing. Meanwhile, the proposed model significantly reduced the error between the measured values of the three types of IMF signals and the true values, and had better processing accuracy than traditional EEMD. The maximum number of IMF components was 6, the lowest measurement error was 6.87%, the highest decomposition stability was 94.02%, and the shortest measurement time was 1.12 seconds. In summary, the proposed model has advantages in efficiency in practical applications.

However, the model has not carried out an in-depth study on the impact of external factors such as changes in environmental temperature and humidity on the metering accuracy. Future research can consider introducing an environmental adaptive mechanism, monitoring environmental changes in real time through temperature and humidity sensors, and dynamically adjusting the metering model by combining with adaptive algorithms. In addition, the expansion of the model can include improving the processing capability of ultra-high frequency noise signals, optimizing the model parameters to further reduce the computation time, as well as combining with the Internet of Things and big data analysis technology to achieve multi-device collaborative metering, and constructing a more comprehensive power monitoring system. These improvements will help to further enhance the practicality and adaptability of the model to meet the needs of more complex and diverse application scenarios.

REFERENCES

[1] Chen Z, Amani A M, Yu X, Jalili M. Control and optimisation of power grids using smart meter data: A review. Sensors, 2023, 23(4): 2118-2124.

[2] Rind Y M, Raza M H, Zubair M, Mehmood M Q, Massoud Y. Smart energy meters for smart grids, an internet of things perspective. Energies, 2023, 16(4): 1974-1981.

[3] Buchibabu P, Somlal J. Green energy management in DC microgrids enhanced with robust model predictive control and muddled tuna swarm MPPT. Electrical Engineering, 2024, 106(3): 2799-2819.

[4] Gao J, Wang X, Yang W. SPSO-DBN based compensation algorithm for lackness of electric energy metering in micro-grid. Alexandria Engineering Journal, 2022, 61(6): 4585-4594.

[5] Liaqat R, Sajjad I A. An event matching energy disaggregation algorithm using smart meter data. Electronics, 2022, 11(21): 3596-3597.

[6] Kumar G, Kumar L, Kumar S. Multi-objective control-based home energy management system with smart energy meter. Electrical Engineering, 2023, 105(4): 2095-2105.

[7] Chou S Y, Dewabharata A, Zulvia F E. Forecasting building energy consumption using ensemble empirical mode decomposition, wavelet transformation, and long short-term memory algorithms. Energies, 2022, 15(3): 1035-1037.

[8] Liang C, Ren W, Cheng P. Control strategy of photovoltaic DC microgrid based on fuzzy EEMD. Tehnički vjesnik, 2022, 29(5): 1762-1769.

[9] Jiang L, Xia L, Zhao T, Zhou J. An improved arc fault location method of DC distribution system based on EMD-SVD decomposition. Applied Sciences, 2023, 13(16): 9132-9133.

[10] Zhang N, Ren Q, Liu G, Guo L, Li J. Short-term PV output power forecasting based on CEEMDAN-AE-GRU. Journal of Electrical Engineering & Technology, 2022, 17(2): 1183-1194.

[11] Wang L, Lodhi E, Yang P, Qiu H, Rehman W U, Lodhi Z, Tamir T S, Khan M A. Adaptive local mean decomposition and multiscale-fuzzy entropy-based algorithms for the detection of DC series arc faults in PV systems. Energies, 2022, 15(10): 3608-3611.

[12] Laayati O, Bouzi M, Chebak A. Smart energy management system: design of a monitoring and peak load forecasting system for an experimental open-pit mine. Applied System Innovation, 2022, 5(1): 18-21.

[13] Ding T, Jia W, Shahidehpour M. Review of optimization methods for energy hub planning, operation, trading, and control. IEEE Transactions on Sustainable Energy, 2022, 13(3): 1802-1818.

[14] Thirugnanam K, El Moursi M S, Khadkikar V, Zeineldin H H, Hosani M A. Energy management strategy of a reconfigurable grid-tied hybrid AC/DC microgrid for commercial building applications. IEEE Transactions on Smart Grid, 2022, 13(3): 1720-1738.

[15] Bhattar C L, Chaudhari M A. Centralized energy management scheme for grid connected DC microgrid. IEEE Systems Journal, 2023, 17(3): 3741-3751.

[16] Badr M M, Ibrahem M I, Kholidy H A, Fouda M M, Ismail M. Review of the data-driven methods for electricity fraud detection in smart metering systems. Energies, 2023, 16(6): 2852-2853.

[17] MT Ibraheem Al-Naib A, Abdullah Hamad B. A Cost-Effective Method for Power Factor Metering Systems. International journal of electrical and computer engineering systems, 2022, 13(5): 409-415.

[18] Bayati N, Baghaee H R, Savaghebi M. EMD/HT-based local fault detection in DC microgrid clusters. IET Smart Grid, 2022, 5(3): 177-188.

[19] Cao W, Zhang F, Chen X. A Study on Fault Localization Method of Three-Terminal Multi-Section Overhead Line–Cable Hybrid Line Using MEEMD Combined with Teager Energy Operator Algorithm. Processes, 2024, 12(7): 1360-1362.

[20] Gheisari M, Hamidpour H, Liu Y, Saedi P, Raza A, Jalili A, Rokhsati H, Amin R. Data Mining Techniques for Web Mining: A Survey. Artificial Intelligence and Applications, 2023, 1(1): 3-10.