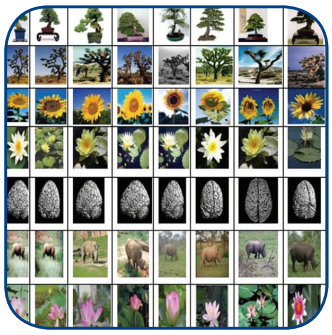
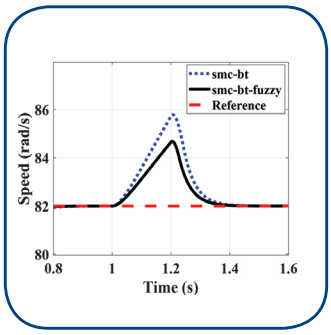
Volume 15, Number 5, 2024 ISSN 1847-6996

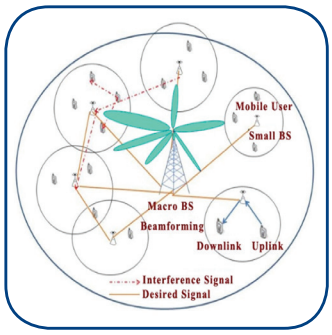
FERIT
FACULTY OF ELECTRICAL ENGINEERING, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY OSIJEK

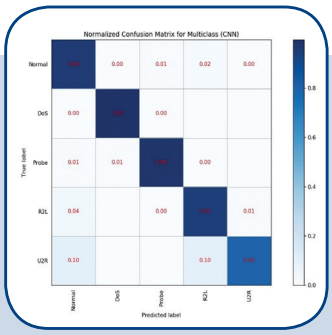
IJECES International Journal of Electrical and Computer Engineering Systems

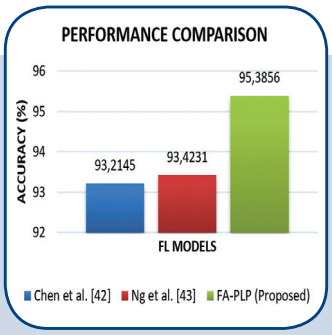
International Journal of Electrical and Computer Engineering Systems

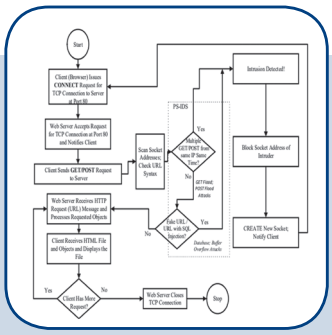












ISSN: 1847-6996

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, Croatia

Osijek, Croatia | Volume 15, Number 5, 2024 | Pages 387 - 467

The International Journal of Electrical and Computer Engineering Systems is published with the financial support of the Ministry of Science and Education of the Republic of Croatia

CONTACT

International Journal of Electrical and Computer Engineering Systems (IJECES)

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b, 31000 Osijek, Croatia
Phone: +38531224600, Fax: +38531224605
e-mail: ijeces@ferit.hr

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.
Giro account: 2390001 - 1100016777, Croatian Postal Bank

Proofreader

Ivanka Ferčec
J.J. Strossmayer University of Osijek, Croatia

Editing and technical assistence

Davor Vrandečić
J.J. Strossmayer University of Osijek, Croatia

Stephen Ward
J.J. Strossmayer University of Osijek, Croatia

Dražen Bajer
J.J. Strossmayer University of Osijek, Croatia

Journal is referred in:

- Scopus
- Web of Science Core Collection (Emerging Sources Citation Index - ESCI)
- Google Scholar
- CiteFactor
- Genamics
- Hrčak
- Ulrichweb
- Reaxys
- Embase
- Engineering Village

Bibliographic Information
Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003
Published: quarterly
Circulation: 300

IJECES online
https://ijeces.ferit.hr

TABLE OF CONTENTS

Security Assessment Framework for IOT via Glove Optimized CNN-BiLSTM

Original Scientific Paper

Arun V
Department of Computing Technologies, School of Computing,
SRM Institute of Science and Technology,
Kattankulathur, Chengalpattu 603203, India
arun.AR543@outlook.com

Ramesh S
Department of Computer Science Engineering,
Krishnasamy College of Engineering Technology,
Anand Nagar, Kumarapuram, Cuddalore, India
remesh765@gmail.com

Carmel Sobia M
Department of Electrical and Electronics Engineering,
PSR Engineering College, Sivakasi, Tamil Nadu
626140, India
sobia654@gmail.com

Geetha A
Department of Electrical and Electronics Engineering,
PSR Engineering College, Sivakasi, Tamil Nadu
626140, India
geetha32GA@gmail.com

Abstract *– The Internet of Things (IoT) is a vast network of real, tangible objects or "things" that can communicate and share data with other systems and gadgets over the Internet. A vital component of assuring the secure and dependable operation of IoT systems and devices is IoT security. Attackers may use IoT devices to get unauthorized access, change functionality, or compromise the data that the device collects and transmits. The risks of IoT security breaches grow as more devices connect and exchange sensitive data. To check the vulnerability in Iot devices, a novel Blockchain enabled Iot based Security Assessment for intrusion (Block-ISAI) technique has been proposed. Bag of Words (BoW) technique is used for feature extraction of API documents which helps to make the document simpler. Blockchain technology is utilized for secure data storage and IoT device registration. In order to detect intrusion, a deep learning architecture is designed using the verified data The attack is either detected or not detected when the vulnerability is found using the GloVe-CNN-BiLSTM Model. If the vulnerability is detected then alerts will be given. Utilizing evaluation measures like accuracy, time efficiency, precision, F1 score, detection rate, recall, false alarm rate, usability and reliability the efficacy of the suggested ISAI technique has been assessed. By the comparison analysis, the proposed ISAI technique's detection rate is 18.22%, 19.43%, and 3.13% higher than the existing HIDS, NIDS, and ML-DDoS techniques respectively. The accuracy of the proposed system is increased by 0.69%, 6.04%, and 36.15% as compared to the HIDS, NIDS, and ML-DDoS method using UNSW-NB 15 dataset and increases by 2.37%, 18.32%, and 5.95% using KDDCUP 19 dataset respectively.*

Keywords: *Internet of things, Security assessment, Vulnerabilities, Bag of words, deep learning*

Received: September 8, 2023; Received in revised form: January 23, 2024; Accepted: January 23, 2024

1. INTRODUCTION

Internet of Things (IoT) is to connect a collection of connected objects so that they may exchange data and communicate with one another online [1]. Its applications extend across numerous industries, enabling companies, boosting productivity, and raising people's quality of life all across the world [2]. IoT's fundamental idea is that by enabling communication between linked things and people, a massive network of interconnected devices can be built [3-5]. Smart homes, healthcare, transportation, agriculture, manufacturing, and many other sectors and businesses have the potential to undergo major transformations as a result of IoT technology [6].

Device security, which focuses on protecting specific devices from unwanted access, tampering, or exploitation, is a vital component of IoT security [7]. Making sure that only permitted parties can access and control IoT devices, entails designing secure hardware and firmware designs, enabling encryption methods, and utilizing authentication techniques [8,9]. IoT device security is used for the security procedures implemented to protect IoT devices and the data they gather, transport, and store [10]. IoT devices are real-life objects that have sensors, software, and connection built into them so they can communicate with other IoT devices and systems via the Internet [11, 12]. A block, which is a sort of digital information, and a chain, which is an open database, make up the first blockchain. As soon as in-

formation is embedded into the immutable sequence of blocks, it becomes impossible to change, providing protection against data poisoning attacks [12]. Decentralized architecture enables smart contracts to improve trust between the parties involved in data transfer. These smart contracts carry out and enforce the conditions of the contract on their own. Moreover, consensus processes provide an extra degree of security by securing the integrity of the distributed data stored in the blockchain [13].

IoT devices regularly capture and communicate sensitive data, such as private information, health data, or financial details. If vulnerabilities are present and not discovered, hackers may use them to intercept data or obtain unauthorized access to the device. IoT device adoption has increased worries regarding security, privacy, and dependability [14]. IoT devices could include security flaws that would be easy for bad actors to use if vulnerability detection wasn't present [15]. The need to address potential vulnerabilities and defend against malicious attacks is becoming more and more important as the number of IoT devices increases. In this paper, a novel Blockchain enabled IoT based Security Assessment for intrusion (Block-ISAI) technique has been proposed to detect the vulnerability attack in IoT devices. The following is a list of the paper's main contributions.

- Initially, API documents are collected from the IoT vendors and then the API document undergoes into feature extraction process.
- In the feature extraction process, the document is analyzed and the Bag of Words (BoW) technique is used for feature extraction and then the output is given to the input message creation module from the feature extraction module.
- A new input message is created and the text message is given to the IoT devices, it generates the response and it is verified by the verifier.
- Blockchain technology is utilized for secure data storage and IoT device registration. In order to detect intrusion, a deep learning architecture is designed using the verified data.
- The vulnerability is detected by using the GloVe-CNN-BiLSTM Model and the output is generated as attack detected and attack not detected.

The remainder of this study is explained in the manner that follows: Section II analyses the study based on the literature. Section III provides a detailed description of the proposed system. Section IV represents the result and discussion, and Section V represents the conclusion.

2. LITERATURE REVIEW

In 2019, Khraisat et al. [16] suggested a unique ensemble Hybrid Intrusion Detection System (HIDS) to safeguard Internet of Things devices. The findings indicate that, in comparison to SIDS and AIDS methods, the suggested hybrid IDS yields a higher detection rate and a smaller percentage of false positives. In 2021, Roy and Srirama [17] suggested a decentralized security system for the Internet of Things (IoT) mobile edge and fog computing. The trial results shows that it outperforms all other methods in its sector and can be used effectively and efficiently as a security feature.

In 2021, Kumar et al., [18] presented a fog-cloud architecture-driven framework for ensemble learning that is used to detect cyberattacks on Internet of medical devices. The experimental results show that the it can achieve 99.98% detection rates, an accuracy of 96.35, and limit false alarm rates up to 5.59%. In 2021,

In Qaddoura et al. [19] recommended a strong intrusion detection system that makes use of a thorough multi-layer categorization method. The proposed technique outperforms the alternatives in terms of the G-mean, which is 78% instead of KNN's 75%.

In 2021, Awotunde et al. [20] proposed several Network Intrusion Detection Systems (NIDSs) to defend and combat IIoT systems in terms of FPR, detection rate, and accuracy, the recommended technique outperforms other pertinent methods by 99.0%, 99.0%, and 1.0%, respectively. In 2022, Hamza et al. [21] suggested the HSAS-MD analyzer, a new hybrid (static and dynamic) SAS that highlights IoT programs from a thorough analytical perspective. The results of the test indicate that HSAS-MD provides 93%, 91%, 94%, and 95% F-measure, recall, precision, and accuracy, respectively.

In 2022, Hayat et al. [22] suggested a multilayer DDoS mitigation technique (ML-DDoS) that uses a blockchain-based infrastructure to protect devices. The findings show that, proposed framework offers up to 35% throughput improvement, up to 40% latency improvement, and up to 25% better CPU utilization.

The comparison table of existing methods are given in the Table 1.

Table 1. Comparison with existing Techniques

Authors	Methods	Evaluation Criterion	Results
Khraisat et al. [16]	HIDS	True positive rate, F-measure, false positive rate, and accuracy	The accuracy of malware detection is 94%.
Roy and Srirama [17]	Security system for the IoT mobile edge and fog computing with block chain	Mathew correlation coefficient (MCC), Positive Predictive Value (PPV), Identification Rate (IR), Accuracy, F-Score, Identification Time (IT)	It has the accuracy of 95.2%
Kumar et al. [18]	Ensemble learning and fog-cloud architecture-driven cyber-attack detection framework.	Accuracy, precision, detection rate, F1 score and false alarm rate	The experimental results show a 99.98% detection rate, a 96.35% accuracy rate.

Qaddoura et al. [19]	A deep multi-layer classification approach	Accuracy, Recall, and G-mean measures	G-mean's value of 78% is in contrast to KNN's 75%
Awotunde, et al. [20]	NIDS	F1-score, recall, specificity, accuracy, and precision	Accuracy, detection rate, and FPR by 99.0%, 99.0%, and 1.0%, respectively
Hamza et al. [21]	HSAS-MD analyzer	Assessed using the widely accepted metrics of recall, accuracy, precision, and F1 score	For accuracy, precision, recall, and F-measure, it offers 95%, 94%, 91%, and 93%, respectively
Hayat et al. [22]	ML-DDoS	Precision of detection, efficacy of mitigation, scalability, and resilience against hostile assaults	It improves throughput by up to 35%, latency by up to 40%, and CPU utilization by up to 25%

3. BLOCKCHAIN ENABLED IOT BASED SECURITY ASSESSMENT FOR INTRUSION (BLOCK-ISAI) TECHNIQUE

In this paper, a novel Blockchain enabled Iot based Security Assessment for intrusion (Block-ISAI) technique has been proposed to detect vulnerabilities in IoT devices. Initially, API documents are collected from the IoT vendors and then the API document undergoes into feature extraction process. These API docs provide details on the acceptable inputs for calling the API-based functionality of IoT devices. The Bag of Words (BoW) algorithm is used for feature extraction of API documents provided by the IoT vendors a new input message is created and the text message is given to the IoT devices, it generates the response and it is verified by the verifier. Blockchain technology is utilized for secure data storage and IoT device registration. In order to detect intrusion, a deep learning architecture is designed using the verified data. GloVe-CNN-BiLSTM Model is used to detect vulnerability in IoT devices. The proposed Block-ISAI method's whole framework is shown in Fig 1.

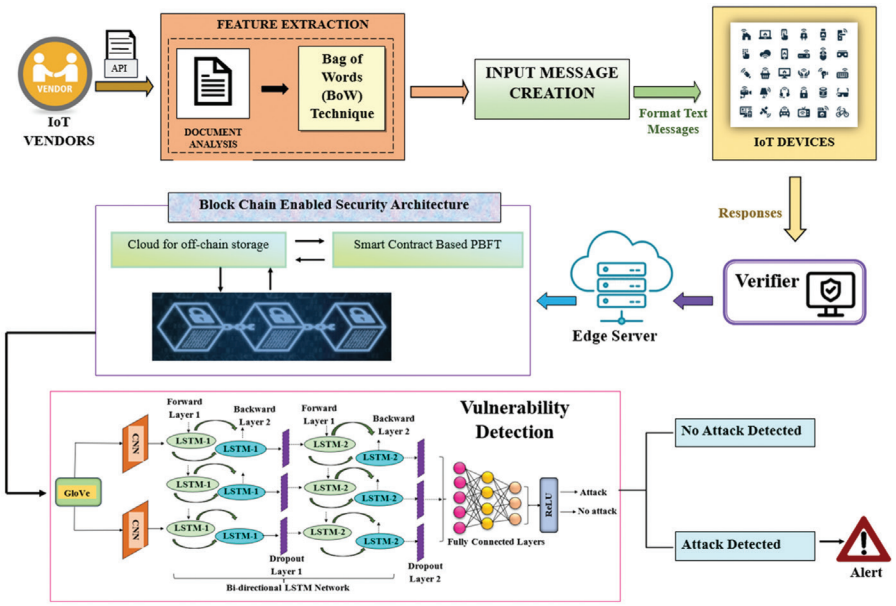


Fig. 1. Blockchain enabled IoT-based Security Assessment for Intrusion (Block-ISAI) Framework

3.1. API DOCUMENTS FOR IOT DEVICES

To assist developers in using their device APIs, the majority of IoT vendors publish an API document. The API document is semi-structured, published in HTML pages, and is available to the public on the Internet. The specifics of the API requirements are usually the first section of an API document. API specifications are information obligatory to construct an appeal note to use a certain device API.

3.2. FEATURE EXTRACTION

In the feature extraction process, the Bag of Words (BoW) technique is used for extracting features and analyzing the document is a key stage, especially when working with textual materials.

3.2.1 Bag of Words (BoW) technique

The Bag of Words is a simple and commonly used feature extraction technique. Text representation is the first step for a machine to comprehend the text. The formula for the bag of words representation of a document is given in (1).

$$bow(d_o)=[count(w_{o_1}, d_o), count(w_{o_2}, d_o)..., count(w_{o_n}, d_o)] \quad (1)$$

Where n represents vocabulary size and the document as d_o, $bow(d_o)$ represents the bag of words representation. Text tokenization is the process of segmenting text into words by utilizing white space and punctuation as delimiters. Using the BoW technique, every document is represented by a numerical vector,

resulting in a fixed feature set. Word frequency in the document is indicated by values in the vector. Formula (2) expresses the BoW design.

$$z=[z_1, z_2, z_3,.....z_n] \quad (2)$$

Where, $z_j = n_j$ if the j-th word appears in the text and $z_j = 0$ if the j-th word does not appear in the text. Two types of features—permission and API function calls—are extracted from the API specification using the BOW approach. The permissions may be collected from the manifest files, and the API function calls are taken from the Java source files. Then, they will include the two collections into the feature set, which serves as an input for the deep learning network's training and testing purposes. Two classes can be distinguished from the classification result based on the DL model. Table 2 provides some instances for the List of Permission Feature Groups from the API document.

Table 2. Permission Feature Groups from API document

Permission Group	Permissions
CALENDAR	android. permission. READ_CALENDAR android. permission. WRITE_CALENDAR
STORAGE	android.permission.READ_EXTERNAL_STORAGE android.permission.WRITE_EXTERNAL_STORAGE
SENSORS	android.permission.BODY_SENSORS android.permission.USE_FINGERPRINT

3. 3. INPUT MESSAGE CREATION

The feature extraction module generates numerical vectors for the input message creation, aligning with target IoT device APIs. User-configured values serve as templates, with default parameters in input vectors. In the advanced block, unnecessary parameters are randomly discarded, and missing ones are created. The module efficiently updates a parameter subset, ensuring a formatted message is sent to IoT devices. Responses are directed to a blockchain-enabled security architecture. Fig 2 shows the proposed ISAI technique's flow chart.

3.4. BLOCK CHAIN ENABLED SECURITY ARCHITECTURE

Six separate processes comprise the first degree of security: 1) Starting; 2) Registration and Authentication; 3) Encoding and Decoding; 4) Block Generation and Verification; 5) Data Creation and Updation of block; and 6) Consensus. Below is a full explanation of how each phase operates.

3.4.1. Starting Phase

In order to register the IoT device (ID), the trusted verifier (V_r) assesses this phase and bootstraps the framework parameters. Stage 1: The verifier (V_r) selects the largest prime value (BP_m) suitable for a non-singular elliptical arc. Random generator g is chosen for $g1$, and bilinear mapping b, is established from $g1 \times g1 \rightarrow g2$. Stage 2: The $Pr_{Vr}k$ (private key) is selected at random by the verifier.

Detailed explanation is as follow

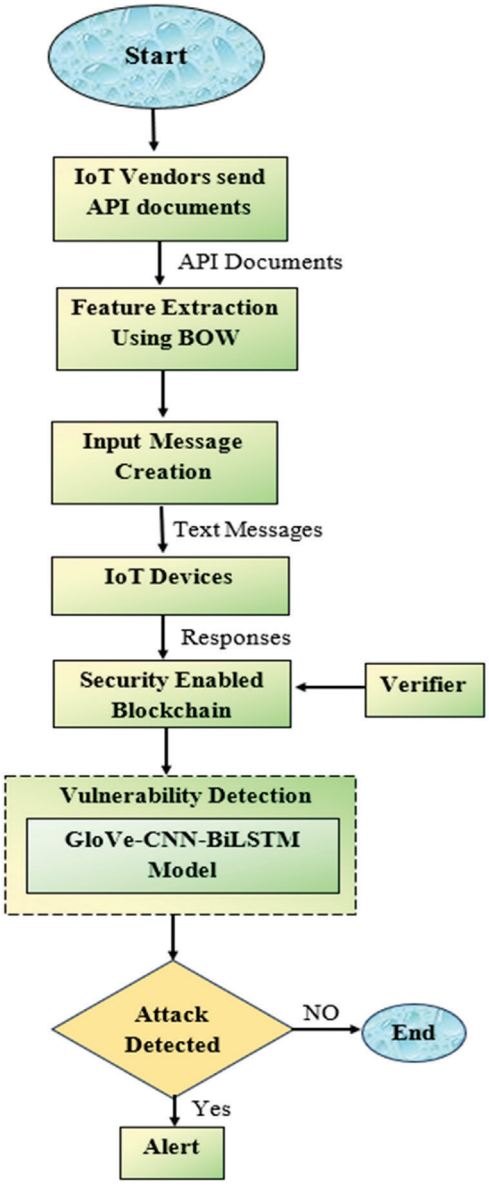


Fig. 2. Flowchart of the proposed Block-ISAI method

Next, $Pb_{Vr}k = Pr_{Vr}k$ is used to generate the public key, or $Pb_{Vr}k$ where $K.g$ stands for multiplication points on an elliptic curve. Stage 3: Vr then selects the one-way cryptographic hash function $Hh(.)$.

3.4.2. Registration and Authentication Phase

IoT device ID requests verifier Vr to join the blockchain (BC) network during the registration step. The IoT device's mac address (M_{ID}) and device identification (i_{ID}) are the two main components that ID uses to construct the provisional key PL_K. Timestamp (TS_j) is saved for ID registration verification when the PL_K is successfully produced. Both PL_K, along with matching i_{ID} and M_{ID}, are sent to the verifier.

3.4.3. IoT-generated data encryption and decryption Phase

Following the successful registration of the IoT device (ID) with the verifying authority, V_r is a public key Pb_{ID} and Pr_{ID} private key is produced. Next, the secret keySK_{ID} is computed over the infinite field ZBP_m and random picked point BP_m over the elliptic curve. Equations (3) and (4) illustrated the two distinct ciphertexts from which the encrypted data are separated.

$$C_a = (BP_m 1 \times BP_m) + SK_{ID} \quad (3)$$

$$C_b = M + (BP_m 1 + Pb_{ID}) + SK_{ID} \quad (4)$$

$$M = ((C_a - Pb_{ID}) \times C_a - SK_{ID} \quad (5)$$

The message created by an IoT device is represented by M, while C_a and C_b indicate the ciphertext. Equation 5 is finally used to decrypt the message.

3.4.4. Block Generation and Verification phase

The process of creating and validating blocks begins after a successful ID registration. Stage 1: The first step consists of key pairs for IoT devices (ID), such as Pb_{ID} and, where Pb_{ID} is a public key and Pr_{ID} is a private key. Stage 2: Ed generates Ed_{sg} and sends it to ID for verification. ID validates the signature, and submits a request for Ed_{sg} to join the BC network. Stage 3: A new block i_{ID}^{block} is created and sent for blockchain.

3.4.5. Data Creation and Updation of Block

This stage explains the process of creating data and updating the corresponding block. Stage 1: Initially, a new transaction (i_{ID}^{NTC}) is established along with Sig_{ID}, Pb_{ID}, and i_{ID} of ID. Stage 2: Furthermore, records are verified Pb_{ID} for the corresponding i_{ID}, in addition to i_{ID}^{TC} and Sig_{ID}. Stage 3: Further, i_{ID}^{block} is successfully appended to the BC network and updated.

3.4.6.Consensus Phase

The i_{ID} is generated, transmitted to IoT devices, and integrated into the BC following ZP verification. The PBFT consensus technique is employed for transaction authentication and addition to the blockchain network (i_{ID}^{TC} by i_{ID}). The SHA-512 algorithm computes the transaction hash and the block is added to the BC.

3.5. GLOVE-CNN-BILSTM MODEL

GloVe-CNN-BiLSTM Model is the combination of Global Vectors for Word Representation (GloVe) with Convolution Neural Networks-Bidirectional Long Short-Term Memory (CNN-BiLSTM) algorithm to detect the vulnerability in the IoT devices.

3.5.1. GloVe Model

A GloVe model is a useful tool for using data from the global corpus and adjusting the learning model based on the context window. The following equation (6) can be used to define the GloVe model:

$$K = \sum_{j,i}^{M} f(Y_{ji}) \left[W_j^T W_i + a_j + a_i - \ln(Y_{ji})\right]^2 \quad (6)$$

where Y is the cooccurrence matrix, Y_{ji} represents how many times the terms j and i appear together in a single window, W_j and W_i stand for the word vectors of j and i. M is the dimension of the cooccurrence matrix $M \times M$, a_j and a_i are the deviation terms, and f is the weight function. The following is the formula for $f(y)$:

$$f(y) = \begin{cases} (y/y_{max})^{\alpha}, & y < y_{max} \\ 1, & y \geq y_{max} \end{cases} \quad (7)$$

3.5.2. CNN-BiLSTM Model

The GloVe model output is fed into the CNN-BiLSTM Model for vulnerability detection. the CNN structure comprises input, pool, and convolution layers, followed by a classifier. For a comment message M={m(1),m(2),…..,m(n)}, each word $w_o(j)$ is transformed into the corresponding word vector $V_e(w_o(j))$ by GloVe, generating a sentence matrix SM_{ji} (8) from the word-by-word statement $w_o(j)$.

$$SM_{ji} = \{V_e(w_o(1)), V_e(w_o(2)), \ldots \ldots V_e(w_o(j))\} \; 1 \leq j \leq n \quad (8)$$

SM_{ji} is the convolution layer's input in the CNN model, and the convolution layer convolves SM_{ji} with a size filter $s \times t$ to derive the regional semantic traits of SM_{ji}. The calculation formula is given in (9)

$$b_{ji} = f(F_l \times V_e(w)(j: j + s - 1) + a) \quad (9)$$

Where, F_l represent the filter of $s \times t$, f indicates the ReLU nonlinear conversion Finally, all pooled attributes are integrated at the entire connection layer to produce the output vector.

$$i_s = \sigma(V^i y_s + X^i h_{s-1}) \quad (10)$$

Let, i_s be the input gate function of the BiLSTM network at time s and V^i, X^i are the weight matrices.

$$f_s = \sigma(V^f y_s + X^f h_{s-1}) \quad (11)$$

In (11), f_s denotes the forget gate function at time step s, σ is the activation function, V^f, X^f indicates the weight matrices of the forget gate function. h_s is the hidden state at time step s.

$$o_s = \sigma(V^o y_s + X^o h_{s-1}) \quad (12)$$

From equation (12), o_s is the output gate function, y_s is the input at the time step s. V^o, X^o are the weight matrices of the output gate function.

$$c'_s = tanh(V^c y_s + X^c h_{s-1}) \quad (13)$$

$$c_s = i_s \times c'_s \times f_s \times c'_{s-1} \quad (14)$$

In equations (13), (14), c_s and c_s' are the cell state during time step s, h_{s-1} is the concealed state and tanh is the hyperbolic tangent activation function.

$$h_s = o_s \times \tanh(c_s) \quad (15)$$

h_s represents the LSTM cell's final hidden state at the most recent time step (s), while o_s denotes the output gate activation at the last time step (s).

The BiLSTM model integrates past and future knowledge using feature data from time *t*. The CNN pooling layer's output, feeds into opposing LSTM networks. Both forward and backward LSTMs capture input sequence information. Vector splicing produces the final hidden layer representation. The GloVe-CNN-BiLSTM Model issues alerts upon detecting attacks.

4. RESULT AND DISCUSSION

The experimental results of the Block-ISAI method are analyzed, and performance is discussed using various evaluation metrics. KDDCUP 19 and UNSW 15 datasets are employed for assessment. Effectiveness is compared with HIDS [16], NIDS [20], and ML-DDoS [22] across F1-Score, accuracy, detection rate, precision, false alarm rate, usability, and reliability.

4.1. DESCRIPTION OF DATASETS

The KDDCUP 19 dataset, a subset of the 1998 DARPA IDS evaluation program, features 28 dimensions out of 41, totalling 31,279 instances. Additionally, the ISCX subset contributes 33,746 instances. The UNSW-NB15 dataset, with 42 features (39 numeric, 3 categorical), is split into UNSW-NB15-TRAIN for training and UNSW-NB15-TEST for testing, serving as a crucial evaluation resource [23-25].

4.2. COMPARATIVE ANALYSIS

This section includes simulations to evaluate the effectiveness of the proposed technique.

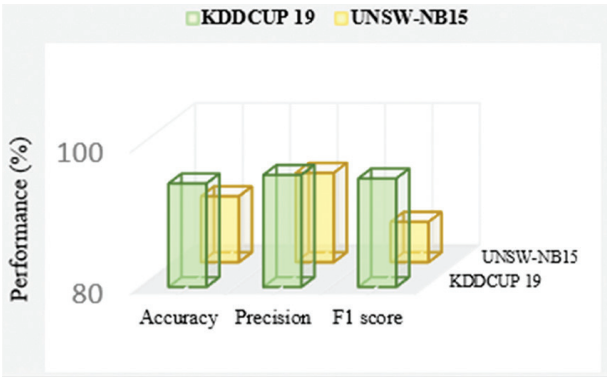


Fig. 3. Performance Comparison

Fig 3 evaluates model performance on KDDCUP 19 and UNSW-NB 15 datasets. For KDDCUP 19, the model achieves outstanding accuracy, precision, and F1 score of 94.6%, 95.8%, and 95.3%. On UNSW-NB 15, it demonstrates strong performance with scores of 89.3%, 92.6%, and 85.7% for accuracy, precision, and F1 score.

Fig 4 compares the accuracy of the proposed Block-ISAI strategy with other approaches (HIDS, NIDS, ML-DDoS) using KDDCUP 19 and UNSW-NB 15 datasets. Our method exhibits significant accuracy improvements of 0.69%, 6.04%, and 36.15%, showcasing superior vulnerability detection compared to existing techniques.

In Fig. 5, the performance comparison of the proposed ISAI technique and existing methods (HIDS, NIDS, ML-DDoS) is depicted, focusing on detection rates using datasets. The Block-ISAI technique exhibits a superior detection rate, surpassing HIDS, NIDS, and ML-DDoS by 18.22%, 19.43%, and 3.13% respectively

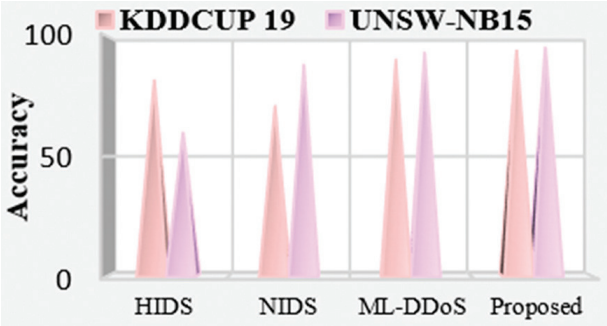


Fig. 4. Performance comparison in terms of accuracy

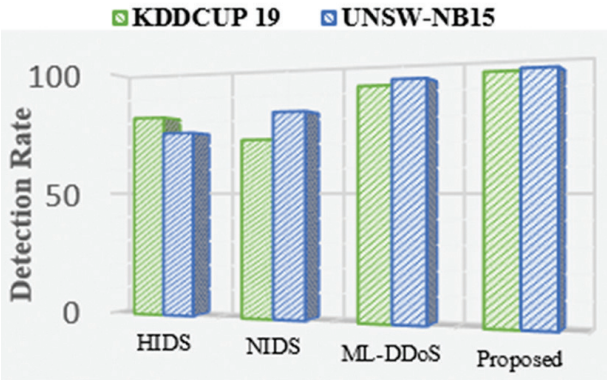


Fig. 5. Comparison in terms of detection rate

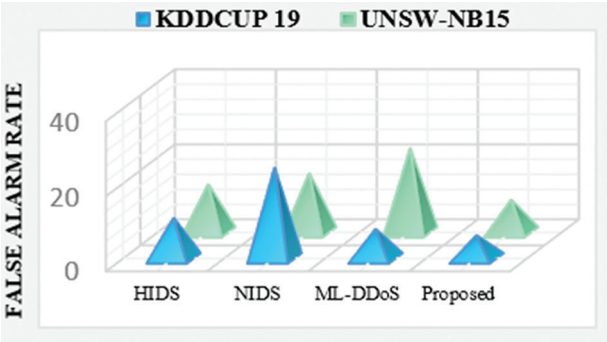


Fig. 6. Comparison in terms of False alarm rate

Fig. 6 compares false alarm rates of our ISAI technique with HIDS, NIDS, ML-DDoS using datasets. Block-ISAI exhibits a lower false alarm rate, demonstrating greater accuracy in threat identification compared to HIDS, NIDS, and ML-DDoS.

Fig. 7 displays results of a blockchain-driven security architecture examination. Block generation and access timings (Figs. 7a and 7b) show stability at 350 TC with up to 40 nodes. However, with 80 nodes, block creation and access take longer than with 60, highlighting scalability challenges in blockchain systems with increased nodes.

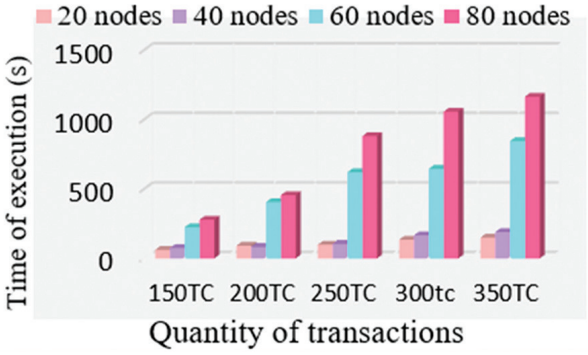


Fig. 7. (a) Block access time across various transaction sizes (TCs)

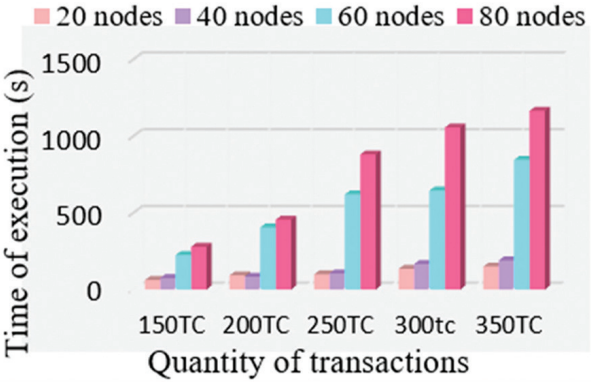


Fig. 7 (b). Block creation time across various transaction sizes (TCs)

Fig. 8 compares our blockchain-enabled IoT security assessment method with traditional approaches, highlighting superior usability and reliability. Enhanced usability comes from a user-friendly interface and robust data integrity procedures, while the decentralized blockchain foundation ensures heightened security for IoT ecosystems.

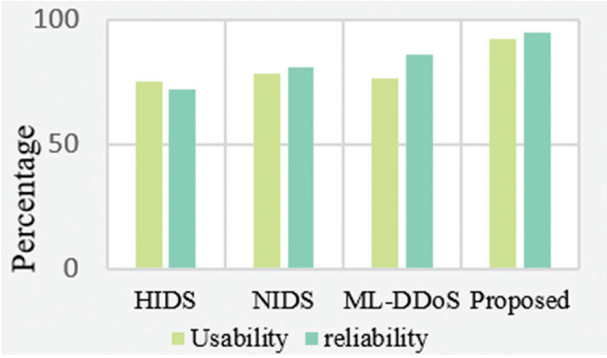


Fig. 8. Comparison in terms of usability and reliability

5. CONCLUSION

In this paper, a novel blockchain enabled IoT based Security Assessment Intrusion (Block-ISAI) technique has been proposed to detect the vulnerability in IoT devices. By extracting the most pertinent and important information, feature extraction helps to make the document simpler. Blockchain technology is utilized for secure data storage and IoT device registration. The vulnerability is detected by using GloVe-CNN-BiLSTM Model and the output is generated as attack detected and attack not detected. The effectiveness of the proposed Block-ISAI technique has been determined using evaluation metrics such as false alarm rate, accuracy, recall, precision, detection rate, F1 score, usability and reliability. According to the comparative analysis, the accuracy of the proposed system is increased by 0.69%, 6.04%, and 36.15% as compared to the HIDS, NIDS, and ML-DDoS method using UNSW-NB 15 dataset and increases by 2.37%, 18.32%, and 5.95% using KDDCUP 19 dataset. Future work will focus on developing user-friendly interfaces for simple configuration, management, and monitoring of security assessments.

6. REFERENCES

[1] A. M. Rahmani, S. Bayramov, B. K. Kalejahi, "Internet of Things Applications: Opportunities and Threats", Wireless Personal Communications, Vol. 122, No.1, 2022, pp. 451-476.

[2] R. Sissodia, M. S. Rauthan, V. Barthwal, "Challenges in Various Applications Using IoT", Handbook of Research on Machine Learning-Enabled IoT for Smart Applications Across Industries, IGI Global, 2023, pp. 1-17.

[3] B. Chander, S. Pal, D. De, R. Buyya, "Artificial Intelligence-based Internet of Things for Industry 5.0", Artificial intelligence-based Internet of things systems, Springer, 2022, pp. 3-45.

[4] K. Elgazzar, H. Khalil, T. Alghamdi, A. Badr, G. Abdelkader, A. Elewah, R. Buyya, "Revisiting the internet of things: New trends, opportunities and grand challenges", Frontiers in the Internet of Things, Vol. 1, 2022, p. 1073780.

[5] F. Firouzi, B. Farahani, M. Weinberger, G. DePace, F. S. Aliee, "IoT fundamentals: Definitions, architectures, challenges, and promises", Intelligent Internet of Things: From Device to Fog and Cloud, Springer, 2020, pp. 3-50.

[6] I. Ahmed, Y. Zhang, G. Jeon, W. Lin, M. R. Khosravi, L. Qi, "A blockchain-and artificial intelligence-enabled smart IoT framework for sustainable city", International Journal of Intelligent Systems, Vol. 37, No. 9, 2022, pp. 6493-6507.

[7] E. R. K. Sen, E. A. Dash, "Unveiling the Shadows: Exploring the Security Challenges of the Internet of Things (IoT)", International Journal of Scientific

Research and Management, Vol. 7, No. 7, 2023 pp. 1-12.

[8] F. Thabit, O. Can, A. O. Aljahdali, G. H. Al-Gaphari, H. A. Alkhzaimi, "Cryptography Algorithms for Enhancing IoT Security", Internet of Things, Vol. 22, 2023, p. 100759.

[9] K. Balasamy, N. Krishnaraj, J. Ramprasath, P. Ramprakash, "A Secure Framework for Protecting Clinical Data in Medical IoT Environment", Smart Healthcare System Design: Security and Privacy Aspects, Wiley, 2022, pp. 203-234.

[10] E. Schiller, A. Aidoo, J. Fuhrer, J. Stahl, M. Ziörjen, B. Stiller, "Landscape of IoT security", Computer Science Review, Vol. 44, 2022, p. 100467.

[11] M. A. Khan, I. Ahmad, A. N. Nordin, A. E. S. Ahmed, H. Mewada, Y. I. Daradkeh, S. Rasheed, E. T. Eldin, M. Shafiq, "Smart android-based home automation system using internet of things (IoT)", Sustainability, Vol. 14, No. 17, 2022, p. 10717.

[12] R. A. Mouha, "Internet of things (IoT)," Journal of Data Analysis and Information Processing, Vol. 9, No. 2, 2021, pp. 77-101.

[13] C. Komalavalli, D. Saxena, C. Laroiya, "Overview of blockchain technology concepts", Handbook of Research on Blockchain Technology, Academic Press, 2020, pp. 349-371.

[14] S. Wang, L. Ouyang, Y. Yuan, X. Ni, X. Han, F. Y. Wang, "Blockchain-Enabled Smart Contracts: Architecture, Applications, and Future Trends", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 49, No. 11, 2019, pp. 2266-2277.

[15] L. Babun, K. Denney, Z. B. Celik, P. McDaniel, A. S. Uluagac, "A survey on IoT platforms: Communication, security, and privacy perspectives", Computer Networks, Vol. 192, 2021, p. 108040.

[16] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, A. Alazab, "A Novel Ensemble of Hybrid Intrusion Detection System for Detecting Internet of Things Attacks", Electronics, Vol. 8, No. 11, 2019, p. 1210.

[17] D. Guha Roy, S. N. Srirama, "A Blockchain-based Cyber Attack Detection Scheme for Decentralized Internet of Things using Software-Defined Network", Software: Practice and Experience, Vol. 51, No. 7, 2021, pp. 1540-1556.

[18] P. Kumar, G. P. Gupta, R. Tripathi, "An ensemble learning and fog-cloud architecture-driven cyber-attack detection framework for IoMT networks", Computer Communications, Vol. 166, 2021, pp. 110-124.

[19] R. M. Qaddoura, A. Al-Zoubi, H. Faris, I. Almomani, "A Multi-Layer Classification Approach for Intrusion Detection in IoT Networks Based on Deep Learning", Sensors, Vol. 21, No. 9, 2021, p. 2987.

[20] J. B. Awotunde, C. Chakraborty, A. E. Adeniyi, "Intrusion Detection in Industrial Internet of Things Network-Based on Deep Learning Model with Rule-Based Feature Selection", Wireless Communications and Mobile Computing, Vol. 2021, 2021.

[21] A. A. Hamza, I. T. A. Halim, M. A. Sobh, A. M. Bahaa-Eldin, "HSAS-MD Analyzer: A Hybrid Security Analysis System Using Model-Checking Technique and Deep Learning for Malware Detection in IoT Apps", Sensors, Vol. 22, No.3, 2022, p. 1079.

[22] R. F. Hayat, S. Aurangzeb, M. Aleem, G. Srivastava, J. C. W. Lin, "ML-DDoS: A blockchain-based multilevel DDoS mitigation mechanism for IoT environments", IEEE Transactions on Engineering Management, 2022. (in press)

[23] R. R. Sathiya, S. Rajakumar, J. Sathiamoorthy, "Secure Blockchain Based Deep Learning Approach for Data Transmission in IOT-Enabled Healthcare System", International Journal of Computer and Engineering Optimization, Vol. 1, No. 1, 2023, 15-23.

[24] M. Dhipa, D. Anitha, "Detection of Violence in Football Stadium Through Big Data Framework and Deep Learning Approach", International Journal of Data Science and Artificial Intelligence, Vol. 1, No. 2, 2023, pp. 21-31.

[25] S. Zafar, N. Iftekhar, A. Yadav, A. Ahilan, S. N. Kumar, A. Jeyam, "An IoT Method for Telemedicine: Lossless Medical Image Compression Using Local Adaptive Blocks", IEEE Sensors Journal, Vol. 22, No. 15, 2022, pp. 15345-15352.

A Regeneration Model for Mitigation Against Attacks on HTTP Servers for Mobile Wireless Networks

Original Scientific Paper

Abiodun Akinwale
A. Akinwale Obafemi Awolowo University, Department of Computer Science & Engineering
Ile-Ife, Osun State, Nigeria, logitronics@yahoo.com

Emmanuel Olajubu
E. A. Olajubu Obafemi Awolowo University, Department of Computer Science & Engineering
Ile-Ife, Osun State, Nigeria, emmolajubu@oauife.edu.ng

Adesola Aderounmu
G. A. Aderounmu Obafemi Awolowo University, Department of Computer Science & Engineering
Ile-Ife, Osun State, Nigeria, gaderun@oauife.edu.ng

Abstract *– The widespread expansion of the internet has fueled a global surge in the utilization of various online transactions, with a significant portion of such services operating on mobile web platforms. Simultaneously, the deployment of innovative Mobile Ad Hoc Network (MANET) technologies and mobile applications has grown as solutions for diverse tasks. Unfortunately, this progress has attracted the attention of hackers, who continually devise new strategies to exploit the vulnerabilities inherent in mobile networks. This study aims to address the escalating challenges posed by cyber threats in the era of widespread internet expansion, particularly focusing on securing mobile web platforms against sophisticated attacks such as Structured Query Language Injection (SQLi) for web-based database solutions and Denial of Service (DoS/DDoS) for various applications. In response to the identified vulnerabilities, this paper proposes an HTTP regeneration (HReg) model that not only detects various cyber-attacks but also ensures the uninterrupted provision of critical services during such incidents. The model introduces an innovative regeneration algorithm capable of scanning both the connection channel and web application to detect attacks, creating survivable connections within the underlying TCP engine to replace compromised ones during an ongoing attack. In simulated environments using OMNeT++, where the server is subjected to attacks, the experimental results demonstrate the efficacy of the model. The response and performance metrics, including throughput (73%), delivery ratio (68.8%), delay (3s), and network load, showcase the model's ability to detect and neutralize attacks. A comparison with state-of-the-art approaches highlights the superior performance of the regeneration model, attributed to its additional survivability layer. While the regeneration model proves robust in simulated environments, its real-world application may encounter limitations. Future research should explore these limitations to enhance the practical applicability of the proposed model. The proposed HReg model's resilient performance under attack conditions ensures the survivability of web-based applications. This innovative approach offers practical implications for securing mobile web platforms, providing continuous delivery of critical services even in the face of persistent and evolving cyber threats. This research addresses a significant gap in existing efforts by not only focusing on attack detection but also emphasizing the development of a survivable TCP connection for HTTP servers during attacks. The introduced regeneration model stands out for its unique ability to maintain service continuity, showcasing originality in the approach to cybersecurity in the context of web-based applications.*

Keywords: *Cyberattacks, DoS, HTTP, MANET, Network Regeneration, OMNeT++, Protocol, Security, SQLi, TCP, TCP/IP*

Received: December 27, 2023; Received in revised form: March 22, 2024; Accepted: March 25, 2024

1. INTRODUCTION

The exponential growth of information and communication technology (ICT) dependence in sectors like business, education, and governance has led to a significant increase in global online transactions.

A significant portion of these transactions now take place over mobile networks, with Mobile Ad Hoc Networks (MANETs) emerging as a crucial representative of such distributed systems or networks. This makes MANETs attractive targets for cybercriminals due to perceived vulnerabilities [1].

Mobile Ad Hoc Networks (MANETs) are a type of mobile wireless network where mobile devices communicate with each other without the need for a fixed infrastructure or centralized administration. In Mobile Ad Hoc Networks (MANETs), the network topology undergoes constant fluctuations due to nodes' movements, additions, and departures. This dynamic characteristic renders MANETs well-suited for scenarios lacking pre-existing infrastructure, such as disaster areas, military operations, or temporary events. Additionally, MANETs operate without centralized control, empowering each node to manage its data routing and participate in the network's self-organization, enhancing adaptability and resilience.

Furthermore, mobile devices within MANETs often face resource constraints, including limited battery power, processing capabilities, and bandwidth. Hence, efficient resource utilization and energy-aware protocols become imperative in MANET design. Moreover, communication between nodes typically occurs through multi-hop routing in the absence of fixed infrastructure, wherein data is relayed through intermediate nodes to reach its intended destination. This multi-hop communication strategy enables effective data transmission despite the dynamic topology and resource limitations, underscoring the importance of protocols designed to efficiently manage changes in network topology, node mobility, link failures, and new node arrivals to ensure smooth MANET operation.

According to Kaspersky's 2023 attack statistics bulletin [2], the daily rate of attacks exceeded 400,000, marking a 3% rise compared to 2022 figures. Correspondingly, [3] highlighted economic losses of up to $100 billion annually due to heightened hacking activities in the United States, while [4] emphasized the importance of users embracing improved cybersecurity strategies, as per their survey on the web threats landscape.

Most of the reported activities happen at the application layer of the TCP/IP protocol stack since this layer supplies essential services for internet users [5]. Hypertext transfer protocol (HTTP) for example, runs on a client/server basis and defines the rules for applications running on diverse systems on how to pass messages to each other. It equally specifies the types, syntax, and semantics needed for processing and responding to requested messages. A node that needs some services is the client and the supply node is the server. Port-to-port connection for different services or processes is provided by the underlying TCP protocol that supplies persistent connectivity until all data segments are delivered to the client after which the TCP connection working in the background ends. Many protocols drive various internet services in this layer such as file transmission protocol (FTP), simple mail transmission protocol (SMTP), post office protocol (POP), domain name service (DNS) - a support service for other applications, Telnet, trivial file transmission protocol (TFTP) and HTTP [6]. HTTP is the most popular, most important and the most widely attacked protocol in the application layer and the subject of this study. It is almost synonymous with the World Wide Web.

The two common attacks on HTTP are DoS, with its variant called DDoS [7], and SQL injection attack [8]. The DoS and DDoS attacks on HTTP are very difficult to differentiate from valid traffic because they use standard URL requests that bear subtle similarities to legitimate requests. This makes them one of the most advanced non-vulnerability security challenges facing servers and applications today. Traditional rate-based detection is ineffective in detecting HTTP flood attacks since traffic volume in HTTP floods is often under detection thresholds or low sensitivity.

On the other hand, the Structure Query Language injection (SQLi) is an attack where a malicious query is posted to a database to manipulate the backend system to access unauthorized data. This attack manifests in various ways and this makes it very difficult to track and prevent. This has allowed hackers to gain access to sensitive unauthorized data in large databases. The common and simplest SQLi attacks are mostly based on the UNION operator. The UNION operator combines two query statements to fetch data from a database. Microsoft SQL servers are particularly vulnerable to error-based SQLi attacks. In this scenario, the attackers trick applications into flagging error messages that contain the data of interest to them. In a blind SQLi attack, the intruders post different types of SQL queries that make the database respond either TRUE or FALSE. The attackers gain information from the database response and use it to modify the SQL query before re-launching the attack. There are other ways of injecting malicious code into a database, such as using HTTP headers. Headers with random SQL queries will inject malicious queries into the database. The second degree of SQLi attacks are more complex and difficult to detect in that the malicious query can be dormant in the system for a long time which may even be considered harmless but when triggered can be very malicious [9].

Existing MANET Intrusion Detection Systems (IDSs) typically generate alerts or apply automated blocking in response to attacks. However, these approaches are insufficient against IP address spoofing, and designing efficient detectors with low computational overhead remains a concern. This work provides mitigation beyond blocking and introduces resilience for survivability with a simple low-power regeneration algorithm necessary to ensure service delivery continuity despite an attack incidence.

This paper introduces a protocol-based intrusion detection system with a regeneration algorithm to mitigate DoS and SQL injection attacks on HTTP servers in MANETs. The regeneration model, inspired by the biological concept of regeneration, replaces compromised TCP connections for HTTP services, ensuring resilience, recovery, adaptability, and survivability. This innovative model allows HTTP servers to continue delivering ser-

vices without succumbing to attacks. Notably, this work represents the first instance of a regeneration model developed for the HTTP protocol in MANET networks.

The rest of the paper is arranged as follows: the next section presents the review of literature on detection systems for DoS, DDoS, and SQL injection attacks on the operations of HTTP. The system methodology is presented in section 3 while the results and conclusions are in sections 4 and 5, respectively.

2. RELATED WORKS

This section reviews current models deployed against DoS and its variants and SQL injection.

In [10], a two-sided algorithm (static and dynamic) was proposed to resolve SQL injection attacks. The static side performs the analysis of SQL query statements to discover any malicious code inserted into the query on the web applications. This is to ensure that user input is not compromised; this prevents any form of SQL injection attack. The second side refers to the dynamic nature of the algorithm which expunges any malicious code detected by the static algorithm. The advantage of the system is that it does not require any web application alteration. Nevertheless, any malicious code found must be manually expunged by the developer/administrator. The model has the same performance with BEBSSARI [11]. The advantage of the model over BEBSSARI is that while BEBSSARI is partially automated, the proposed model is fully automated.

Another machine learning algorithms proposal authored by [12] demonstrated the use of a sliding window which optimizes the detection accuracy by dynamically obtaining the segment for intrusion detection. The paper applied the Burstiness statistical measure, Sharon entropy, and Quantile Cumulative probability threshold algorithms for the calculation and optimization of the smallest window. The method used in this model is consistent with [13] and [14]. The study employed a sliding window algorithm based on morphological fractal dimension, the result presented showed that there is improved detection of DDoS attacks with an accuracy of 99.30%. Ref. [15] proposed a security model with two ways of operation against cyberbioattacks. Cyberbiotattacks use biologically engineered bacteria to implement distributed denial-of-service. The first part of the model produces molecules that can block the traffic sent by cyberbioattacks, this is called quorum quenching, while the second called the amplification approach, radiates molecules with the capacity to increase the proportion required to build more biofilm structure. The performance of the model was measured by creating a dynamic scenario that allows DDoS traffic to be generated and sent as a cyberbioattack. The result presented showed that the model reduced the influence of cyberbioattack. The quorum quenching scheme performs better than the amplification approach concerning attack detection, but the amplification approach familiarizes better in attack configurations with the biofilm formation. One of the major threats to 5G communication network reliability and quality of service is DoS. A model was developed by [16] to mitigate the impact of DoS on the 5G communication network so that proper quality of service can be experienced by the subscribers. The model deployed a differential flow management scheme to tackle the menace of DoS on the network. The traffic was classified as continuous or discrete flow. The discrete flow was used as a sub-optimal differential problem, the model aimed at converging the anomalous detection time and reformulating the allocation of resources as a continuous flow based on the leftover. To retain the response time and transmission rate of the user equipment, the flow was modeled continuously on service and transmission intervals. The results presented by the authors were consistent as intrusion detection time was minimized, the delivery ratio was maximized, and the response time of the system was retained. Information security on the web is becoming a herculean task as hackers have devised various attack schemes to access unauthorized information and disrupt opponent networks to gain some reputation.

Cyber-attack through SQLi is one of the most difficult attacks to trace and the detection often takes a long period. Ref. [17] designed an SQLi scanning model using the MySQL injector tool that can easily conduct penetration tests on a PHP-based website. The tool contains four developmental phases in which phase one is called Inception. This gathers the vulnerabilities of the website while the second phase called elaboration translates all the vulnerability requirements gathered in the first phase into software design models. The construction phase or phase three is where a prototype is developed based on the requirements and design while the transition phase is the full development of the system to testing. In the same stride, [18] worked on the vulnerability of wireless ad-hoc networks to DoS attacks and identified that attack drains the battery power of nodes, making the resources of the network unavailable resulting in degrading the network performance. The work reviews various techniques such as packet leash, spread spectrum, energy weight monitoring system, and lightweight secured mechanism among several others employed to counter DoS attacks. These schemes have not been able to effectively checkmate DoS attacks. Refs. [19] and [20] opined that DDoS attack is the main obstruction in the success rate of Software Define Networks (SDN), despite several proposals, mitigation of this type of attack remains very difficult. A detection proposal was modeled and implemented. The system consists of a programmable network monitoring kit and a control structure that is very flexible for fast attack detection. This architecture of the attack detection system is based on a statistical function that can address the flooding problem. The simulation results presented show that the model was effective in addressing the DDoS attack.

Ref. [21] conceptualized the SQLi attack as a Markovian decision process with a reinforcement learning problem. The reinforcement learning agent was developed to perform SQL injection. The training of the system was enacted in such a way as to have a generic guideline that affects SQL injection attacks and not specific code injection to the web. In the analysis, the results outlined the agent's learning process and the complexity of the algorithm. The developed agent succeeded in deep penetration testing. The overall result of the study demonstrated the possibility and weakness of using reinforcement learning in the security environment. A machine learning algorithm was proposed by [22] which is to monitor the external aggressor and the VM against DDoS attack. The model comprises of data clustering technique through a feature selection algorithm by principal component analysis. The Agglomerative Clustering and K-means algorithm was used in the model. The results showed that the model had 0.9130 on the adjusted Rand Index metric. The model was able to effectively handle the external and internal traffic against DDoS on a cloud computing platform.

Many research outputs have attempted to proffer solutions to the susceptibility of HTTP protocol stack to SQLi and DDoS attacks. Ref. [23] in their work responded to the growing reliance on cyber infrastructures and the escalating frequency of cyber-attacks with the continuous demand for the development of advanced protection mechanisms. The study focuses on creating and implementing a Deep Neural Network model for the detection of intrusions in computer networks. Techniques such as SMOTE and Random Sampling were applied to address data imbalance in the CICIDS 2017 dataset. The entire experiment was conducted on a single Jupyter notebook in the Google Colaboratory environment, importing and implementing relevant software libraries like Seaborn, Pandas, Matplotlib, Keras, and TensorFlow as needed. The results revealed excellent performance, with a 99.68% accuracy score and a loss of 0.0102 in predicting attacks with the CICIDS 2017 dataset.

In [24], the researchers aim to enhance automated intrusion detection by developing a highly accurate classifier with minimal false alarms. Their motivation is to address the challenges of high dimensionality in intrusion detection and improve classifier performance for more accurate and efficient intrusion detection. Experiments were conducted using the NSL-KDD dataset, initially employing the entire feature set, where the J48 tree achieves the highest reported accuracy of 79.1%. The study explores Random Projection and PCA to enhance classifier performance, with Random Projection proving more time-efficient and achieving a notable accuracy improvement to 82.0%. The research concludes that random projection is effective in improving intrusion detection accuracy while reducing training time, contributing valuable insights to the cybersecurity field. The study in [25] focused on SQL injection (SQLi) attacks on websites, using tools like Whois, SSL Scan, Nmap, OWASP Zap, and SQL Map. Researchers identified and mitigated vulnerabilities on the web server, discovering 14 issues with OWASP Zap, categorized by severity. SQL Map successfully accessed the database and username. The research aims to recommend firewall installation for mitigating SQLi risks, providing a structured methodology for addressing vulnerabilities in web servers and enhancing data security. According to [26], DDoS attacks pose a significant threat to the web due to the rise in web-based transactions and Internet application services. Countering these attacks has become more challenging with attackers utilizing vast resources and techniques. Unlike traditional network-layer attacks, application-layer DDoS attacks, particularly slowloris attacks, can be more effective by inundating victim resources with legitimate HTTP requests. The paper proposes a slowloris attack detection method based on an adaptive timeout-based approach, comprising a suspect determination module and an attacker verification module. The experimental results demonstrate the algorithm's efficacy in detecting attackers with low false alarms and high accuracy before consuming all resources.

A frequently observed constraint noted across papers [10-26] is the lack of mechanisms ensuring uninterrupted service delivery during attacks despite the prevalent utilization of machine learning algorithms in the reviewed methods for highly accurate attack detection. This research enhances the functionality of the HTTP protocol through the introduction of a novel model termed HReg (HTTP Regeneration). HReg not only identifies but also withstands SQL injection and DoS attacks targeting HTTP servers. It achieves this through the incorporation of network-based attack scanners, protocol surveillance detection for protocol-specific application-based attacks, and a regeneration-based technique aimed at ensuring system survivability.

2.1. HTTP OPERATIONS AND TRANSACTIONS

HTTP defines the protocol for retrieving various types of messages, including text, video, audio, graphics, and other files, from World Wide Web servers through client browsers. Clients initiate web page requests, and servers respond with the requested information. When HTTP clients seek web pages from server hosts, they utilize resource identifiers, with the most common being the Uniform Resource Locator (URL), also known as the web address. The URL follows a standard format: protocol://host:port/path. Often, specifying protocol://host is sufficient to connect to a resource, allowing users to navigate to desired paths. For instance, http://www.yahoo.com features the host or domain name www.yahoo.com, and HTTP utilizes TCP port 80 for its connections. While HTTPS is a more secure protocol and industry preference is shifting towards it, the proposed model's operations are equally applicable to it, as it is not entirely immune to all attacks.

HTTP transactions follow a straightforward process:

- The client initiates a request using methods such as GET, POST, OPTION, HEAD, PUT, TRACE, etc. For instance, GET is used for static content requests, while POST is employed for dynamic content like databases. The request method is accompanied by a header specifying the desired response format from the server.
- The server responds with a reply typically commencing with a status code (e.g., 100, 101, 200, 400, 401, 501, etc.), followed by the header and the message in case of successful retrieval. The client browser then displays the response contents using Hypertext Markup Language (HTML).

2.2. ATTACKS ON HTTP PROTOCOL

HTTP attacks are usually carried out through code injection and result in snooping/sniffing, modification, masquerading, and denial of service attacks. The attack is most effective when it forces the server or application to distribute the maximum resources possible in response to each single request. Thus, the perpetrator will generally aim to inundate the server or application with multiple process-intensive requests. For this reason, HTTP flood attacks using POST requests tend to be the most resource-effective from the attacker's perspective; as POST requests may include parameters that trigger complex server-side processing. On the other hand, HTTP GET-based attacks are simpler to create, and can more effectively scale in a botnet scenario.

3. METHODOLOGY

In the context of this work, regeneration refers to the process through which networks under attack replace compromised TCP connections underlying HTTP with replicas from an available pool of sockets. The replacement socket is an exact copy of the compromised or lost one, and this substitution occurs nearly instantaneously.

Drawing inspiration from bio-defense mechanisms, cell regeneration in organisms serves as a captivating biological process allowing them to regenerate body parts as a defense against predators. For instance, octopuses exhibit an impressive ability to regenerate lost body parts, including arms and their central nervous system. Similarly, in the proposed model, the replication of new arms (connections) promptly replaces lost ones from an internal pool of cells. While regeneration is advantageous for defense, it comes at a cost to the prey (network resources). HTTP connections, akin to cells, require replenishment or promotion of regeneration to restore damaged connections during an attack. This capability ensures that networks continue delivering on their stated missions despite damaged links or nodes, maintaining service quality without stochastic fluctuations. This aspect is crucial in network design to ensure high-quality service (QoS) or secured socket connections in the HTTP protocol, mitigating the impact of intruders on link quality or connections. The ability to sustain good service delivery despite attacks determines the survivability of the network, which is the primary focus of this work.

It is worth noting that while this study focuses on HTTP servers, the transport engine is the TCP protocol, ensuring the connection between the server and the client. In socket programming, the process of setting up a TCP socket on the server side involves creating a socket, binding it to an address, listening for connections, accepting a connection, sending and receiving data, and closing the connection.

3.1. REGENERATION MODEL FORMULATION

In this work, the regeneration mechanism for the HTTP protocol is initially modeled to achieve optimal link or connection replacement from the network pool of TCP sockets providing connection service. The modeling of TCP socket regeneration in a network entails the consideration of factors including the generation of new sockets, the closure of existing sockets, and the potential influence of network events like the velocity of mobile nodes.

Let S(t) be the size of active TCP sockets at time t. The regeneration of TCP sockets can be modeled as follows:

$$\frac{dS}{dt} = \alpha - \beta S - \gamma E \tag{1}$$

Here:

- α represents the rate at which new TCP sockets are created or regenerated.
- β represents the rate at which existing TCP sockets close or become inactive.
- γ represents the impact of velocity (denoted by E) that might affect the TCP socket number.

This model assumes a constant regeneration rate α and a linear dependence on the current number or population of sockets. The term $-\beta S$ represents the closing or inactivation of existing sockets, and $-\gamma E$ accounts for external events affecting the regeneration process.

Next, consideration is given to the impact of attacks like SQL injection (SQLi) on HTTP server connections. The regeneration of underlying TCP sockets in the presence of SQLi attacks necessitates an examination of their impact on socket size.

The ensuing differential equation, which integrates the effect of SQLi attacks, is presented as follows:

If $S(t)$ is the number of active TCP sockets at time t, and $A(t)$ is the size of SQL injection attacks. The regeneration of TCP sockets in the presence of SQLi attacks is as follows:

$$\frac{dS}{dt} = \alpha - \beta S - \gamma ES - \delta A \tag{2}$$

$$\frac{dA}{dt} = \phi A - \theta SA \tag{3}$$

Where:

- α represents the rate at which new TCP sockets are created or regenerated.
- β represents the rate at which existing TCP sockets close or become inactive.
- γ represents the impact of external events (denoted by E) on the existing TCP sockets.
- δ represents the impact of SQL injection attacks (A) on the TCP socket population.
- ϕ represents the rate of SQL injection attacks.
- θ represents the interaction strength between the TCP sockets and the SQL injection attacks.
- The term $-\gamma ES$ in the equation for dS/dt models the effect of external events such as node movement on the existing sockets, and δA models the impact of SQL injection attacks on the socket population. The equation for dA/dt describes the dynamics of the SQL injection attacks, with ϕ being the rate of attacks and θS representing the interaction strength between the SQL injection attacks and the existing TCP sockets.

The regeneration of TCP sockets in the presence of DDoS attacks is as follows:

$$\frac{dS}{dt} = \alpha - \beta S - \gamma ES - \delta A^2 \quad (4)$$

$$\frac{dA}{dt} = \phi A - \theta SA \quad (5)$$

Where:

- α represents the rate at which new TCP sockets are created or regenerated.
- β represents the rate at which existing TCP sockets close or become inactive.
- γ represents the impact of external events (denoted by E) on the existing TCP sockets.
- δ represents the impact of SQL injection attacks (A) on the TCP socket population. The square term A^2 indicates a potential nonlinear impact distributed attack.
- ϕ represents the rate of DDoS attacks.
- θ represents the interaction strength between the TCP sockets and the DDoS attacks.

3.1.1. ATTACK PROBABILITIES

Given the stochastic nature of attack scenarios, probabilities of attack neutralization are now incorporated into the differential equations mentioned earlier.

Let $P_n(t)$ be the probability of neutralizing the SQL injection attack, and $P_o(t)$ be the probability of the attack overcoming the defense. The regeneration of TCP sockets in the presence of SQLi attacks with probabilities of attack neutralization can be modeled as follows:

$$\frac{dS}{dt} = \alpha - \beta S - \gamma ES - \delta A2Po \quad (5)$$

$$\frac{dA}{dt} = \phi A - \theta SAPn \quad (6)$$

$$\frac{dP_n}{dt} = \rho Pn(1 - Po) \quad (7)$$

$$\frac{dP_o}{dx} = \sigma Pn - \omega Po \quad (8)$$

Here:

- α represents the rate at which new TCP sockets are created or regenerated.
- β represents the rate at which existing TCP sockets close or become inactive.
- γ represents the impact of mobility (denoted by E) on the existing TCP sockets.
- δ represents the impact of SQL injection attacks (A) on the TCP socket population, with P_o indicating the probability of the attack overcoming the defense.
- ϕ represents the rate of SQL injection attacks.
- θ represents the interaction strength between the TCP sockets and the SQL injection attacks, with P_n indicating the probability of attack neutralization.
- ρ is the rate at which the probability of neutralizing the attack increases.
- σ is the rate at which the probability of the attack overcoming the defense increases.
- ω is the rate at which the probability of the attack overcoming the defense decreases.

The terms $-\delta A2P_o$ in the equation for dS/dt and θSAP_n in the equation for dA/dt introduce the probabilities P_o and P_n, respectively, influencing the dynamics of TCP socket regeneration in the presence of SQLi attacks. The term $-\gamma ES$ in the equation for dS/dt models the effect of external events such as node movement on the existing sockets, and $\delta A2$ models the impact of DDoS attacks on the socket population. The equation for dA/dt describes the dynamics of the DDoS attacks, with ϕ being the rate of attacks and θS representing the interaction strength between the DDoS attacks and the existing TCP sockets.

2.1.1. COST OF REGENERATION DEFENSE DEPLOYMENT

Deploying regeneration defense incurs a cost associated with actively neutralizing attacks, and it influences the decision-making process in the network dynamics. Incorporating the cost of attack neutralization into the equations involves introducing a cost term in the equation for TCP socket dynamics dS/dt.

The updated set of equations including cost becomes:

$$\frac{dS}{dt} = \alpha - \beta S - \gamma ES - \delta A2P_o - \chi P_n \quad (9)$$

$$\frac{dT}{dt} = \phi A - \theta SAP_n \quad (10)$$

$$\frac{dP_n}{dt} = \rho Pn(1 - Po) - \omega Pn \quad (11)$$

$$\frac{dP_o}{dx} = \sigma Pn - \omega Po \quad (12)$$

In Equations (9) – (11):

χ represents the cost of attack neutralization. The term $-\chi P_n$ is subtracted from the TCP socket regeneration rate, reflecting the cost incurred in neutralizing attacks. This is essentially the product of the cost parameter (χ) and the probability of attack neutralization (P_n). Mathematically, the cost of attack neutralization (C_n) can be expressed as:

$$Cn=\chi P_n \quad (13)$$

This equation reflects that the cost of attack neutralization is proportional to both the cost parameter (χ) and the probability of attack neutralization (P_n).

3.1.3. MODEL PARAMETRIZATION

Model parameters can be obtained by experiment, systems in production, or empirically. To ensure that attacks are always neutralized, parameters must be set to promote a high probability of attack neutralization and prevent the probability of the attack overcoming the defense (P_o) from becoming dominant. The following considerations are important to achieve this:

1. Set ρ to a high value: A high value for ρ in the equation $dP_n/dt = \rho P_n(1-P_o)$ will make the probability of attack neutralization (P_n) increases rapidly.
2. Set σ and ω to low values: Low values for σ and ω in the equation $dP_o/dt = \sigma P_n - \omega P_o$ will make the probability of the attack overcoming the defense (P_o) change slowly.
3. Ensure ϕ, the rate of SQL injection attacks is not too high: If the rate of attacks (ϕ) is very high, it might be challenging to neutralize all of them. A reasonable ϕ value, combined with a high ρ, will help maintain a high probability of neutralization.
4. Adjust the other parameters ($\alpha, \beta, \gamma, \delta, \theta$): These parameters govern the dynamics of TCP socket regeneration and attack interactions. Adjust them based on your understanding of the system and the desired behavior.

3.2. INTRUSION DETECTION FOR HTTP PROTOCOL

HReg detection methodology entails the deployment of a scanner at each susceptible point within the HTTP protocol. The detection engine consists of two main modules: the protocol scanning module and the dynamic database updater module. Through continuous analysis of all traffic, the scanner identifies new attacks and maintains communication with the database module to verify signatures of known attacks. Upon detection of a novel attack, the database is promptly updated. The IDS scanners vigilantly monitor socket addresses (port + IP addresses) to effectively detect protocol-specific threats.

Each socket is associated with a port number, facilitating the recognition of the intended application by the TCP layer. Intrusions are identified when multiple GET/POST requests originate from the same IP address, indicating suspicious activity. Furthermore, alerts are triggered when fake URL commands or syntax anomalies are detected in other requests. Upon identifying an attack, the IDS system issues an alert, initiating the activation of countermeasures for regeneration.

3.3. REGENERATION MODEL ALGORITHM

Algorithm 1 delineates the series of steps involved in overseeing HTTP requests, executing intrusion detection procedures, and deploying necessary actions upon intrusion detection. The procedure encompasses interactions between the HTTP client and server, detection of potential intrusions through scanning, and the restoration of connections following intrusion detection alerts. Upon activation of an intrusion alert, the socket linked with the intruder is blocked and placed on a blacklist. Subsequently, a new 'create socket' instruction is executed, as illustrated in Fig. 1.

The algorithm is designed based on the behavior of an HTTP connection to a server during an attack, such as a Denial of Service (DoS). Each available socket connection represents a potential route to withstand the attack. The client utilizes the discovered connection and explores other connections to the socket at the server end.

Algorithm 1: Regeneration Algorithm Pseudocode

Algorithm: RegenerationDefense

1. Start
2. ConnectToServer() // HTTP Client issues CONNECT Request for TCP connection at Port 80 to HTTP Server
3. NotifyServerConnection() // Server accepts request for TCP connection at port 80 and notifies client
4. SendRequestToServer() // Client sends GET/POST Request to server
5. ScanAndCheckURLSyntax() // IDS Scans Socket Addresses and checks URL syntax
6. **IF** MultipleRequestsFromSameIP() **OR** FakeURLorSQLiDetected() THEN
   - 6.1. IntrusionDetectionAlert()
   - 6.2. BlockIntruderSocketAddress()
   - 6.3. RegenerateNewSocketAndNotifyClient()
   - 6.4. **GO TO** 2
7. **ELSE**
   - 7.1. ReceiveHTTPRequestAndProcessObjects() // Web Server receives HTTP Request (URL) Message and processes requested objects
   - 7.2 SendHTMLAndObjectsToClient() // Client receives HTML file and objects and displays the file
   - 7.3 **IF** MoreRequestsFromClient() **THEN**
     - 7.3.1 GO TO 4
   - 7.4 ELSE
     - 7.4.1 CloseConnection()
8. Stop

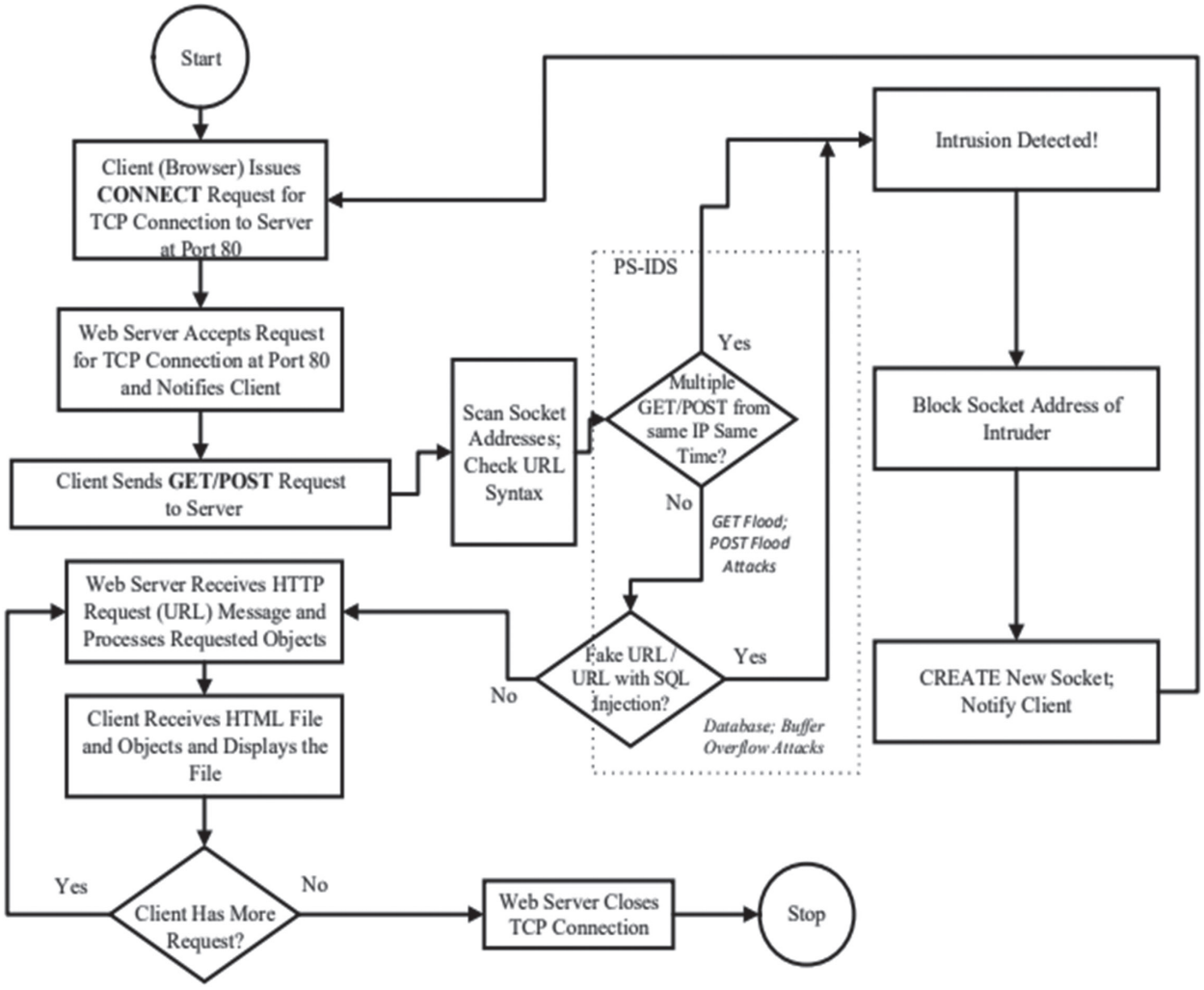


Fig. 1. Regeneration Defense Flowchart

OMNeT++ served as the primary tool for evaluating the performance of the proposed model. Widely recognized as one of the most effective network simulators, OMNeT++ enjoys broad support within the scientific community. Written in C++, it is a prevalent choice for simulating networks. The INET Framework module suite was employed in OMNeT++ because of its comprehensive assortment of models for mobile networks, TCP/IP protocol stack, mobility, and its flexibility in allowing users to customize output vector statistics according to their needs. The experiment's parameters are detailed in Table 1, while Fig. 2 illustrates the simulation scenario for the model.

In the context of the OMNeT++ discrete event simulator, the purpose of HTTP is to manage traffic between the browser (referred to as the client) and the server. Three key components are employed: the browser (simulated by HTTPBrowser), the server (simulated by HTTPServer), and the controller (managed by HTTPController).

To initialize the HttpTools components in a simulation, a single controller object needs to be instantiated at the scenario level. It is crucial to note that any number of nodes with browser or server components can be generated, subject to memory constraints, processing power, and practical considerations. Nodes can be linked through OMNeT++ communication links and the INET TCP/IP networking stack implementation. Alternatively, direct message passing can be employed to eliminate network effects. Simulated HTTP messages are employed by the browser and server components.

For the browser and server components, two types of hosts are available. *StandardHost* from the INET framework, suitable for a complete network simulation using the TCP/IP stack. *DirectHost*, a component of *HttpTools*, is designed for hosts employing direct message passing.

The server and browser modules leverage the TCP/IP modeling of the INET framework for transport and integrate with its StandardHost module. Additionally, these modules support OMNeT++ direct message passing when the network infrastructure effects are not considered. To provide a lightweight alternative to StandardHost, a simple container named DirectHost has been created for this purpose.

In the event of an attack, the regeneration defense is employed to reset the TCP connection, and the connection is altered.

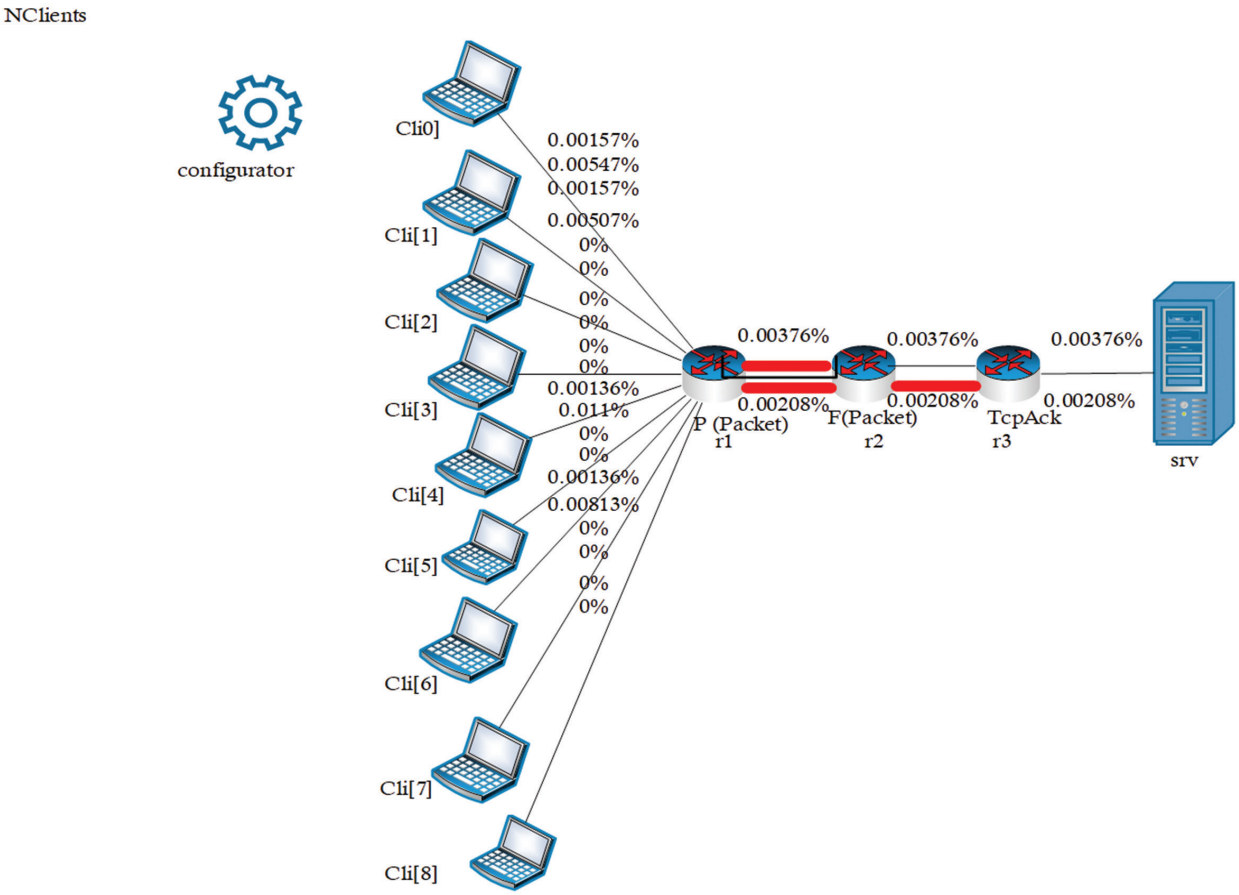


Fig. 2. OMNeT++ Simulation Setup

3.5. OMNET++ PERFORMANCE METRICS FOR EVALUATION

Performance metrics serve as a comprehensive characterization of an entire network or a specific node under examination. These metrics played a crucial role in evaluating the simulated regeneration model. The study employed four key performance metrics: network load, packet delivery ratio, network throughput, and end-to-end delay.

The concept of packet load is crucial in assessing the traffic volume generated during both the establishment and maintenance phases of the HTTP protocol, incorporating overhead for regeneration implementation. Notably, the proposed model integrates mobility rate as a significant factor, as increased mobility may result in intermittent connection disruptions during regeneration defense operations. This consideration underscores the importance of accounting for mobility's impact on connection stability and overall system performance.

Network throughput, a key metric reflecting the efficiency of data transmission, is essential for evaluating protocol performance. Despite being a desirable goal for any protocol, this study acknowledges the trade-offs involved, particularly when regeneration defense operations cause temporary connection resets. Additionally, heightened mobility exacerbates the situation by inducing frequent topology changes, affecting packet transmission across different sockets and potentially impeding throughput optimization. These insights highlight the necessity of balancing network throughput objectives with the disruptive effects of regeneration defense and mobility on connection stability and data transmission efficiency.

The packet delivery ratio, indicative of data loss and protocol efficiency, plays a critical role in assessing a protocol's performance. However, it's vital to recognize that a high mobility rate can diminish protocol throughput, underscoring the intricate relationship between mobility, data delivery, and protocol efficiency. Furthermore, end-to-end delay, encompassing various factors influencing packet transmission time, such as buffer queues and transmission delays, is observed to increase with higher mobility rates. These findings emphasize the need for comprehensive evaluations that consider mobility's impact on data delivery, protocol efficiency, and end-to-end transmission delays to ensure robust system performance.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section provides a detailed discussion of the simulation results based on the input parameters outlined in Table 1. The proposed regeneration defense model's performance was assessed using the four metrics when confronted with denial-of-service attacks.

Table 1. Parameter Settings for Layer 4 HTTP Protocol Simulation

Parameter	Value
Simulation	3600
Number of Nodes	100
Simulation Area (m)	500 X 500
Sending Interval) s)	1
Protocol	HTTP
Node Velocity (m/s)	10
Data Rate (MB/s)	10
Transmit Power	2.0mW
No. of Channels	10
Carrier Frequency	2.4/5Ghz
Traffic Model	TCP
MAC Protocol	IEEE 802.11g
Packet Size – CBR Bytes	100

As illustrated in Fig. 3a, during the attack, packet data delivery plummeted to 31% of the normal value, preventing nodes from transmitting data and causing a denial of service to the server for intended packet transmissions. Upon deploying the regeneration defense, channel access was swiftly restored to 69% of its initial value within the 10ms experimental timeframe, effectively neutralizing the attack. Once neutralized, resources were promptly redirected to data transmission, achieving a 69% data transmission rate. The remaining 31% of resources were utilized to assess the extent of damage caused by the attack and transmit attack information to the network administrator for policy enforcement.

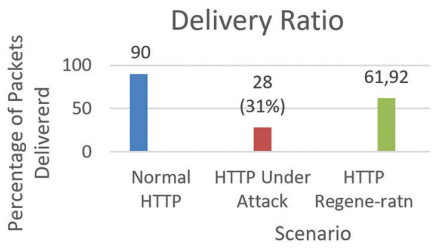


Fig. 3a. Delivery Ratio Results

Fig. 3b depicts the system throughput, revealing a reduction to 62 packets/s under attack conditions, signifying a 19-packet/s decrease during the investigation and neutralization period. As the system resumed normal data transmission, throughput increased to 73% of its value within the 10ms experimental time.

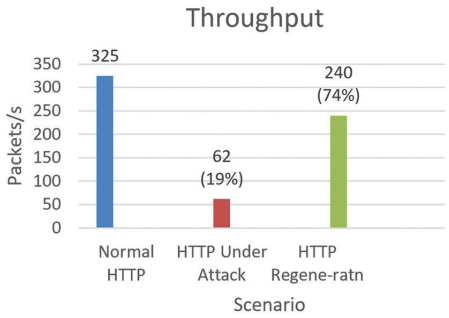


Fig. 3b. Throughput Results

The delay experienced during the attack is presented in Fig. 3c, indicating a delay of 3s. With the deployment of regeneration, the delay was reduced to 0.6s, despite the need for connection resets and resource reallocations. Notably, this delay is nearly equivalent to the delay observed in the absence of an attack (0.5s).

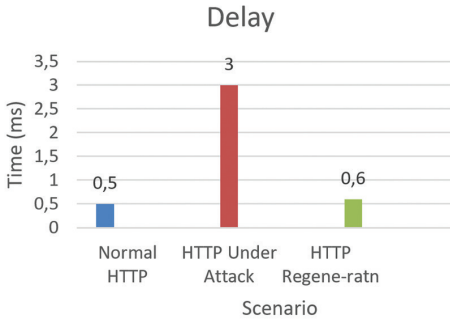


Fig. 3c. Delay Results

In Fig. 3d, the load is considerably lower (2%) during an attack compared to normal or defense operations. This reduction is attributed to the denial-of-service attack, which suspended all network control and maintenance packets during DDoS, rendering normal network load packets non-functional. The model efficiently restored the native load to 100% of its value, along with an additional overhead (136%) from the regeneration implementation, almost immediately.

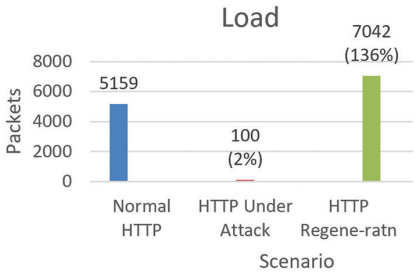


Fig. 3d. Network Load Results

4.1. COMPARISON RESULTS

This section presents a comprehensive comparison of the proposed HReg model with the state-of-the-art approach [26]. The evaluation encompasses the design, experimental methodology, metrics employed, results obtained, and outlook. Table 2 provides a succinct summary of this comparative analysis. For a detailed performance assessment, our proposed model evaluates throughput, delivery ratio, delay, and load individually. This approach allows us to discern the simulation's efficiency in terms of network performance, message delivery, delay characteristics, and load handling. Notably, this work focuses on survivability as its primary goal.

In contrast, the existing approach employs binary classification to measure results concerning recall, precision, specificity, and sensitivity. Their methodology is tailored towards detection only, aligning with the

primary objective of the previous work. Furthermore, while the datasets for the existing model are generated from their university network, the proposed model utilizes embedded datasets from OMNET++. This choice enhances the versatility and applicability of the HReg model in diverse scenarios.

Table 2. Comparative Analysis of Regeneration Model and State-of-the-art

Category	Existing Model	Regeneration Model
Experimental Platform	C and Java Programs	OMNET++ Simulation Environment
Method	Binary Classification	Discrete Events Simulation
Dataset	Tezpur University Web Server Dataset	OMNET++ INET Dataset
Metrics	Recall, Precision, Specificity, and Sensitivity	Throughput, Delivery Ratio, Delay, and Load
Attack Target(s)	Slowloris DDoS Attack	DoS, DDoS, SQLi Attacks
Model Objective(s)	Detection of Attacks Only	Detection and Survivability from Attacks
Deployment Platform(s)	No Specific Domain	Low-resource, Full-resource Mobile Wireless Domains

One significant distinction lies in the scope of attack detection. The existing model can only detect Slowloris DDoS attacks, whereas our model is proficient in identifying all variants of DoS and SQLi attacks, ensuring a broader range of security coverage and improved survivability. Moreover, our model is designed to operate seamlessly on both low-resource and full-resource devices within wireless networks. In contrast, the existing model lacks a specified target domain or platform, making our approach more adaptable to various environments and resource constraints.

This study developed a distinctive approach to safeguarding HTTP servers by merging survivability with intrusion detection. This method stands in contrast to the current state-of-the-art practices, which typically focus on either intrusion detection or intrusion prevention.

5. CONCLUSION

A thorough examination of existing Intrusion Detection Systems (IDS) reveals a notable emphasis on the detection of attacks, often overlooking the critical aspect of survivability during such incidents. The proposed HReg regeneration model for HTTP intrusion detection and defense mechanism breaks away from this trend by addressing both the connection channel and the web application's protection. This approach aims to detect attacks and establish survivable connections within the underlying TCP layer, replacing compromised connections during imminent or full-scale attacks. Specifically designed for Mobile Ad-Hoc Networks (MANETs) and other HTTP-based mobile wireless devices, the model is adaptable to diverse distributed networks in real-world scenarios. It incorporates a dynamic threat database for MANET networks to respond to evolving and known threats effectively. Continuous monitoring of its performance over time in various network environments will provide insights into the dynamic nature of network traffic and potential emerging threats.

Future research endeavors should aim to integrate simulation methodologies with machine learning techniques. This involves training the machine learning model on simulated datasets to optimize its performance in real-world applications.

Acknowledgments:

The authors express gratitude to TETFUND for the support received through the 2020 research grant intervention. Special thanks to ACE OAK-PARK, Obafemi Awolowo University, for providing additional funding, laboratory facilities, and necessary space for this research.

6. REFERENCES

[1] K. Kumar, "Denial of service attacks – an updated perspective", Systems Science & Control Engineering, Vol. 4, No. 1, 2016, pp. 285-294.

[2] Kaspersky, "Rising Threats: Cybercriminals Unleash 411,000 Malicious Files Daily in 2023", https://www.kaspersky.com/about/press-releases/2023_rising-threats-cybercriminals-unleash-411000-malicious-files-daily-in-2023 (accessed: 2024)

[3] Z. Cekerevac, Z. Dvorak, L. Prigoda, P. Cekerevac, "Hacking, Protection and the Consequences of Hacking", Komunikacie, Vol. 20, No. 2, 2018, pp. 83-87.

[4] B. A. Obotivere, A. O. Nwaezeigwe, "Cyber Security Threats on the Internet and Possible Solutions", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 9, No. 9, 2020, pp. 92-97.

[5] B. A. Forouzan, "Data Communications and Networking", The McGraw-Hill Companies Inc., 2007.

[6] V. J. Glowniak, "An Introduction to the Internet, Part 3, Internet Services", Journal of Nuclear Medicine Technology, Vol. 23, No. 4, 1995, pp. 231-248.

[7] R. F. Silva, R. Barbosa, J. Bernardino, "Intrusion Detection Systems for Mitigating SQL Injection Attacks: Review and State-of-Practice", International Journal of Information Security and Privacy, Vol. 14, No. 2, 2020, pp. 20-40.

[8] Patil, A. Laturkar, S. V. Athawale, R. Takale, P. Tathawade "A multilevel system to mitigate DDOS, brute force and SQL injection attack for cloud security", Proceedings of the International Conference on Information, Communication, Instrumentation and Control, Indore, India, 17-19 August 2017, pp. 1-7.

[9] A. Dizdar, "SQL Injection Attack: Real Life Attacks and Code Example", https://brightsec.com/blog/sql-injection-attack/ (accessed: 2023)

[10] I. Lee, S. Jeong, S. Yeo, J. Moon, "A novel method for SQL injection attack detection based on removing SQL query attribute values", Mathematical and Computer Modelling, Vol. 55, 2012. pp. 58-68.

[11] Y. Huang, F. Yu, C. Hang, C. H. Tsai, D. T. Lee, S. Y. Kuo, "Securing web application code by static analysis and runtime protection", Proceedings of the 13th International World Wide Web Conference, May 2004, pp. 40-52.

[12] G. Baldini, I. Amerini, "Online Distributed Denial of Service (DDoS) intrusion detection based on the adaptive sliding window and morphological fractal dimension", Computer Networks: The International Journal of Computer and Telecommunications Networking, Vol. 210, No. C, 2022.

[13] G. Baldini, "On the application of entropy measures with sliding window for intrusion detection in automotive in-vehicle networks", Entropy, Vol. 22 No. 9, 2022, p. 1044.

[14] S. D. Çakmakç, T. Kemmerich, T. Ahmed, N. Baykal, "Online DDoS attack detection using Mahalanobis distance and kernel-based learning algorithm", Journal of Network and Computing Applications, Vol. 168, 2020, p. 102756.

[15] S. L. Bernal, D. P. Martins, A. H. Celdrán, "Towards the mitigation of distributed denial-of-service cyberbioattacks in bacteria-based biosensing systems", Digital Signal Processing, Vol. 118, 2021, p. 103241.

[16] P. Sakthibalan, K. Devarajan, "DFMS: Differential flow management scheme for denial-of-service impact mitigation in 5G communications", Journal of King Saud University – Computer and Information Sciences, Vol. 34, 2022, pp. 5366–5374.

[17] A. B. M Ali, A. Y. I Shakhatreh, M. S. Abdullah, J. Alostad, "SQL-injection vulnerability scanning tool for automatic creation of SQL-injection attacks", Procedia Computer Science, Vol. 3, 2011, pp. 453-458.

[18] A. Patil, R. Gaikwad, "Comparative analysis of the Prevention Techniques of Denial-of-Service Attacks in Wireless Sensor Networks", Procedia Computer Science, Vol. 48, 2015, pp. 387-393.

[19] F. Bensalah, N. E. Kamoun, M. E. Houssaini, "Online detection of Denial-of-Service Attacks in Software Defined Networking using the Hotelling Chart", Procedia Computer Science, Vol. 160, 2019, pp. 785-790.

[20] Y. Tian, V. Tran, M. Kuerban, "DOS Attack Mitigation Strategies on SDN Controller", Proceedings of the IEEE 9th Annual Computing and Communication Workshop and Conference, Las Vegas, NV, USA, 7-9 January 2019, pp. 701-707.

[21] L. Erdődi, Å. Å. Sommervoll, F. M. Zennaro, "Simulating SQL injection vulnerability exploitation using Q-learning reinforcement learning agents", Journal of Information Security and Applications, Vol. 61, 2021, p. 102903.

[22] F. J. Abdullayeva, "Distributed denial of service attack detection in E-government cloud via data clustering", Array, Vol. 15, 2022, p. 100229.

[23] E. Osa, P. E. Orukpe, U. Iruansi, "Design and implementation of a deep neural network approach for intrusion detection systems", e-Prime - Advances in Electrical Engineering, Electronics and Energy, Vol. 7, 2024, p. 100434.

[24] F. Nabi, X. Zhou, "Enhancing intrusion detection systems through dimensionality reduction: A comparative study of machine learning techniques for cyber security", Cyber Security and Applications, Vol. 2, 2024, p. 100033.

[25] A. Fadlila, I. Riadib, M. A. Mu'min. "Mitigation from SQL Injection Attacks on Web Server using Open Web Application Security Project Framework", Internation Journal of Engineering, Vol. 37, No. 4, 2024, pp. 635-645.

[26] T. Kangkan, B. Debojit, "Slowloris Attack Detection Using Adaptive Timeout-Based Approach", ISeCure, Vol. 16, No. 1, 2024, p. 79.

Deep Learning-based DDoS Detection in Network Traffic Data

Original Scientific Paper

Teeb Hussein Hadi
Middle Technical University,
IT Department, Technical College of Management, Baghdad, Iraq
eng.teebhussien@mtu.edu.iq

Abstract *– In today's society, the cloud is essential for communication since it allows access to important information anytime and anywhere. However, cloud services also attract hackers who want to exploit online details. This has caused significant changes in the cyber-attack landscape. Distributed Denial of Service (DDoS) is the most common attack. Traditional tools like firewalls and encryption can mitigate these risks, but new models are needed to cope with the changing nature of cyber-attacks. Detecting DDoS attacks is particularly challenging since network traffic data is complex and often contains unnecessary features. To address this, a new approach is proposed using Denoising AutoEncoder (DAE) and a Convolutional Neural Network (CNN) for feature selection and classification. The NSL-KDD dataset is used to evaluate the performance of this new model with three main steps: Data Preprocessing, Hyper-parameter Optimization, and Classification. Our method performed better in all four metrics, such as Accuracy, Recall, Precision, and F1-score, with rates of 97.7, 98.1, 97.7, and 97.8, respectively. The multiclass classification detection rate for DOS was 100%. Similarly, the detection rates for Probe, R2L, and U2R were 98%, 95%, and 80%, respectively. Python version 3.6 with Keras 2.2.4 and TensorFlow Engine was used in this paper.*

Keywords: *Network security, DOS, DAE, CNN, Multiclass classification, Deep Learning*

Received: December 15, 2023; Received in revised form: March 4, 2024; Accepted: March 5, 2024

1. INTRODUCTION

Today's interconnected society has revolutionized communication through the advent of IoT services, making vast amounts of information readily accessible online, anytime and from anywhere. Regrettably, this accessibility also exposes the data to cyber-attacks, capitalizing on vulnerabilities that are either unknown or capable of circumventing existing security measures. An effective solution for safeguarding network integrity is the deployment of Intrusion Detection Systems (IDS) [1-3].

IDSs can be categorized in various ways, with one common classification based on their detection method. This categorization divides IDSs into two primary types: signature-based or misuse detection and anomaly-based detection. Signature-based IDSs compare data points with known signatures and trigger an alarm upon detection of a match [4,5]. Conversely, an anomaly-based IDS establishes a pattern from normal traffic and flags any deviation from this pattern as an abnormal transaction. Both methods possess distinct advantages and drawbacks. While signature-based IDSs excel at identifying known attacks, they necessitate frequent manual updates to their signature database. On the other hand, anomaly-based IDSs are adept at uncovering unknown attacks but often produce a plethora of false alarms. Contemporary techniques such as Deep Learning (DL) and Deep Neural Networks (DNN) are increasingly utilized to mitigate these limitations. DL can autonomously learn features and minimize false alarms [6-9].

In this study, we introduce a Convolutional Neural Network (CNN) architecture into our intrusion detection system to identify attacks. Our objective is to classify all four attack categories and subsequently prioritize the detection of Denial of Service (DOS) attacks. Before further processing to reduce data dimensions, we employed a Denoising AutoEncoder (DAE) to select an optimal feature set. We evaluated our model using the NSL-KDD dataset, a refined version of the KDDCup99 and one of the most commonly employed datasets in this domain. Developed by the Defense Advanced Research Projects Agency (DARPA), the KDDCup99 dataset is a benchmark for intrusion detection studies.

While prior research has primarily focused on distinguishing between different attack types, our study proposes a novel approach to binary and multiclass

classification. This approach integrates DAE and CNN for feature selection and classification, respectively. To demonstrate the efficacy of our methodology, we compared the multiclass classification results with those of three previous studies. Our method outperformed existing approaches across all four metrics, including Accuracy, Recall, Precision, and F1-score, as evaluated on the NSL-KDD dataset.

2. NETWORK SECURITY

Network security refers to the different mechanisms and techniques to prevent unauthorized access to digital assets in a network environment. Its main objective is to establish a set of practices that comply with the CIA triad, which stands for confidentiality, integrity, and availability and is the foundation of any security program in an organization [10-12].

This paper is organized as follows: Section 3 describes the dataset, Section 4 presents related work, Section 5 explains the research methodology, Section 6 covers the experimental results and analysis, and Section 7 concludes with future work recommendations.

3. DATASET DESCRIPTIONS

3.1. NSL-KDD

The KDDCup99 is older and has unnecessary data points, which leads to model performance in accuracy while detecting intrusions in an IDS. This issue has been resolved in the refined version of KDDCup99, NSL-KDD. The NSL-KDD is one of the most commonly used datasets in the domain of IDSs. In this work, KDDTrain+. TXT and KDDTest+. TXT files, which have 125,973 and 22,544 records, respectively, are considered. The total number of features in NSL-KDD is 41, with the data types nominal, binary, and numeric. It has four major categories of attacks, which are R2L, U2R, Probe, and DoS, in addition to the Normal class [13-16].

Table 1. Provides a list of features for the dataset

Feature and type	Feature and type	Feature and type
[Duration]=num	[Su Attempted]=bin	[Same Sry Rate]=num
[Protocol Type]=nom	[Num Root]=num	[Diff Sry Rate]=num
[Service]=nom	[Num File Creations]=num	[Sry Diff Host Rate]=nunn
[Flag]=nom	[Num Shells]=num	[Dst Host Count]=num
[Src Bytes]=num	[Num Access Files]=num	[Dst Host Sry Count]=num
[Dst Bytes]=num	[Num Outbound Cmds]=nunn	[Dst Host Same Sry Rate]=num
[Land]=bin	[Is Hot Logins]=bin	[Dst Host Diff Sry Rate]=num
[Wrong Fragment]=num	[Is Guest Login]=bin	[Dst Host Same Srv Rate]=num
[Urgent]=num	[Count]=num	[Dst Host Srv Diff Host Rate]=num
[Hot]=num	[Srv Count]=num	[Dst Host Serror Rate]=num
[Num Failed Logins]=num	[Serror Rate]=num	[Dst Host Srv Diff Host Rate]=num
[Logged In]=bin	[Srv Serror Rate]=num	[Dst Host Serror Rate]=num
[Num Compromised]=num	[Rerror Rate]=num	[Dst Host Srv Rerror Rate]=num
[Root Shell]=bin	[SR/ Rerror Rate]=num	[Label]=nom

4. RELATED WORKS

In [17], the author utilized the NSL-KDD dataset to assess the efficacy of various classification algorithms in detecting abnormalities in network traffic patterns. Their study has yielded valuable insights into the relationship between protocols and network attacks. Their model improves the accuracy of intrusion detection systems and introduces a new research direction in this field. In [18,19], the authors investigated Deep Learning (DL) algorithms to be highly effective in solving various problems across different domains, such as Long Short-Term Memory (LSTM) and Fully Connected Neural Networks (FCNN) that used to categorize benign and malicious connections in intrusion datasets. To achieve a more accurate classification of multi-class assault patterns, They proposed a deep learning model that produces more precise classifications when applied to five-class issues. The model achieves an accuracy of 99.99% when tested on the KDDCup99 dataset and 99.95% on the NSL-KDD dataset. Our model secures the maximum output on both datasets.

In [20], the authors combined two feature selection approaches using LDA and CCA with seven different classifiers: Naive Bayes, Random Tree, Rep-tree, Random Forest, Random Committee, Bagging Randomizable, and Filtered. They concluded that LDA feature selection with Random Tree performed best among the various combinations of feature selection and classifiers. Utilizing LDA and the Random Tree algorithm in anomaly detection was found to be faster and more effective than other methods. Moreover, the accuracy of the Random Tree algorithm surpasses that of different algorithms. This method accurately distinguishes between normal data and various types of attacks. The accuracy of the approach can be further enhanced by employing feature reduction techniques. Based on these findings, it can be inferred that this approach excels in speed, efficiency, and accuracy, especially when implemented on Apache Spark.

In [21], the authors proposed a scenario for backdoor attacks, focusing on the "AlertNet" intrusion detection model and utilizing the NSL-KDD dataset, widely used

in NIDS research. Their study used KL-divergence and OneClassSVM for distribution comparisons to demonstrate resilience against manual inspection by a human expert for outliers. Their experimental results indicated that utilizing decision trees significantly improves the attack's success rate and validated the anomaly regions through KL-divergence, OneClassSVM, and manual inspection.

Authors in [22] proposed a new method to enhance the performance of Intrusion Detection Systems (IDS) on the NSL-KDD dataset. They employed meta-heuristic algorithms and machine-learning techniques for this purpose. Multiple meta-heuristic algorithms were utilized to optimize the hyperparameters of machine learning models, including Random Forest (RF), Support Vector Machine (SVM), Classification and Regression Trees (CART), and Multilayer Perceptron (MLP). The performance of the IDS was evaluated using metrics such as precision, recall, F1-score, and accuracy. Their experimental results demonstrated that the proposed approach outperforms existing techniques in accurately and robustly detecting intrusions.

In [23], the author implemented an IDS framework using Machine Learning (ML) techniques that incorporated various types of Recurrent Neural Networks (RNNs), such as Gated Recurrent Unit (GRU), Long-Short Term Memory (LSTM), and Simple RNN. His results demonstrated that for binary classification tasks using NSL-KDD, XGBoost-LSTM achieved the best performance, with a test accuracy (TAC) of 88.13%, a validation accuracy (VAC) of 99.49%, and a training time of 225.46 seconds. On the other hand, for UNSW-NB15, XGBoost-Simple-RNN was the most efficient model, with a TAC of 87.07%.

In [24], the authors introduced a new approach to enhance the accuracy and efficiency of intrusion detection systems. Their approach utilized Long Short-Term Memory (LSTM) optimized with the Penguin Optimization Algorithm (EPO). Initially, the features underwent preprocessing, including normalization, cleaning, and formatting into numerical format. Subsequently, the Linear Discriminant Analysis (LDA) method was employed to reduce the dimensions of the processed features. Following this, the EPO algorithm was utilized to optimize the size of the hidden units in the LSTM network. Finally, the optimized network was evaluated using the NSL-KDD dataset, a widely recognized benchmark dataset in intrusion detection. Their training and test datasets results were 99.4% and 98.8%, respectively.

Authors in [25,26] decreased the number of features in data using PCA and AutoEncoder. Then, they used Lenet5 CNN for intrusion detection on the KDDCup99 dataset, concluding that the CNN performed better for detecting intrusion on the KDDCup99. According to experimental results, the CNN-IDS model outperforms traditional algorithms in AC, FAR, and timeliness.

5. METHODOLOGY

The general steps of our proposed model are shown in Fig.1. Broadly, it includes data pre-processing, Hyperparameter Optimization, and Classification.

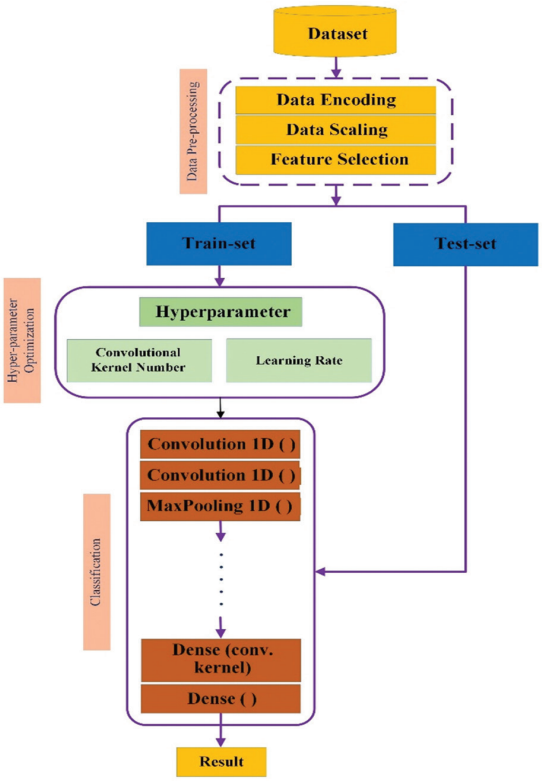


Fig. 1. General steps of the proposed model

5.1. DATA PRE-PROCESSING

The NLS-KDD has some nominal features with many values in each. Those features have to be encoded before any other operation. In this study, a one-hot encoding technique is applied, and then the scaling is performed using the min-max technique, which transforms each feature between 0 and 1. The formula for min-max is given in the equation 1 [27-29].

$$X_a' = \frac{X_a - Min(X)}{Max(X) - Min(X)} \quad (1)$$

Where X_a denotes the original value, X_a' represents the scaled value, $Min(X)$ stands for the minimum value of the feature, and $Max(X)$ gives the maximum value of the feature. The encoding generates many new features in the data, totaling 121 features. A feature selection technique is applied using DAE in the next data preprocessing phase to reduce the number of features. Out of 121 features, only 15 are selected.

5.2. HYPER-PARAMETER OPTIMIZATION

Traditionally, hyperparameters were rarely optimized due to their computational cost requirements. With the advancement of technology, this task is now carried out using modern technologies and powerful algorithms to enhance model performance. This study fo-

cuses on two parameters: convolutional kernel number and learning rate. We provide a range of learning rates, namely 0.03, 0.01, 0.008, 0.006, and 0.004. The convolutional kernel number ranges for optimization are 16-16-32-32, 16-16-64-64, and 32-32-64-64.

5.3. CLASSIFICATION

This study employed a one-dimensional convolutional neural network as a classification model. The classification results in a CNN-based model are directly influenced by the number of convolution kernels and the learning rate [30-33]. We conducted experiments on multiple convolution kernels with different learning rates to obtain the optimal set of parameters. This experiment was carried out on NSL-KDD for multiclass classification. Some significant configurations in the CNN model include loss function = categorical cross-entropy, optimizer = Nadam, pooling = Max Pooling, output activation = softmax, activation function for other layers = ReLU, and dropout parameter = 0.5. We tested three different convolution kernels with five learning rates, as mentioned in section 4.2. The classification metrics used in this paper are Accuracy, Precision, Recall, and F1-score. The calculations for each metric are given by equations 2, 3, 4, and 5, respectively [28-33].

$$Accuracy = \frac{(TN + TP)}{(TP + TN + FP + FN)} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$DR(Recall) = \frac{TP}{(FN + TP)} \quad (4)$$

$$F1 - score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)} \quad (5)$$

TP stands for True Positive value, TN represents True Negative, FP gives False Positive, and FN denotes False Negative.

6. EXPERIMENTAL RESULTS AND DISCUSSION

The experiment's workstation configuration and tools included a Windows 11 Pro 64-bit operating system with 32 GB RAM and an Intel CPU. The version of Python used was 3.6 with Keras 2.2.4 and Tensorflow Engine. The data was divided into a train set, test set, and validation set with a ratio of 70%, 20%, and 10%, respectively. Table 2 illustrates the appropriate train-test split for the dataset.

Table 2. The number of instances in each class of the NSL-KDD dataset

Class	Train-set (70%)	Test-set (20%)	Validation-set (10%)	Total
Normal	54,153	15,168	7,733	77,054
DoS	37,520	10,508	5,357	53,385
Probe	9,896	2,772	1,409	14,077
R2L	2,637	738	374	3,749
U2R	180	50	22	252
Total	104,386	29,236	14,895	148,517

Intensive comparative analysis has been conducted with the 1D CNN model through various learning rates and convolutional kernel numbers. Table 3 provides the close results.

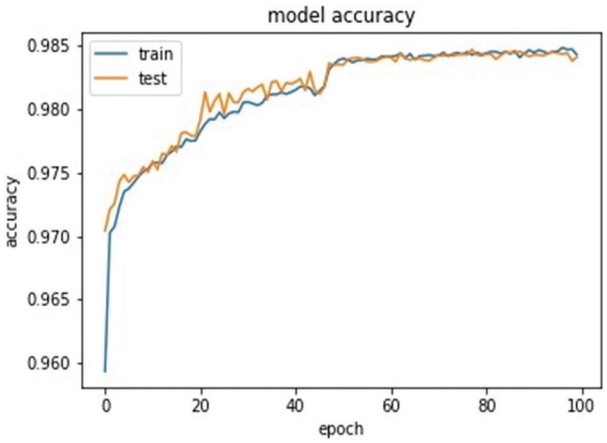
Table 3. Comparison of the proposed model with various numbers of convolution kernels at different learning rates in multi-class classification on the NSL-KDD dataset

Conv. Kernel #	LR	Accuracy %	Precision %	Recall %	F1-score %	Train time in sec.	Test-time in sec.
16-16-32-32	0.03	95.95	97.33	95.95	96.38	76.25	0.82
	0.01	96.49	97.46	96.49	96.80	85.58	0.94
	0.008	96.61	97.49	96.61	96.88	106.17	1.09
	0.006	96.12	97.31	96.12	96.49	104.23	1.18
	0.004	96.21	97.38	96.21	96.58	110.27	1.34
16-16-64-64	0.03	97.52	97.99	97.52	97.67	74.54	1.90
	0.01	97.14	97.77	97.14	97.33	67.24	2.03
	0.008	97.20	97.80	97.20	97.38	65.37	2.17
	0.006	97.48	97.95	97.48	97.62	98.78	2.25
	0.004	97.12	97.79	97.12	97.33	74.49	2.44
32-32-64-64	0.03	97.53	98.04	97.53	97.69	60.85	2.98
	0.01	97.34	97.91	97.34	97.51	67.26	3.20
	0.008	97.68	98.10	97.68	97.81	71.76	3.32
	0.006	97.48	97.96	97.48	97.63	74.33	3.44
	0.004	97.27	97.90	97.27	97.46	101.58	3.59

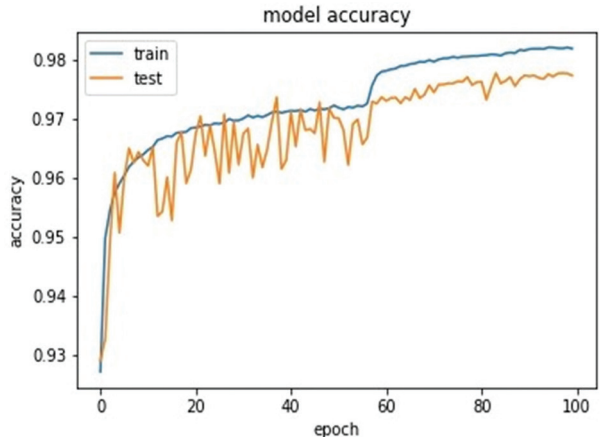
We have observed that the convolution kernel 32-32-64-64, with a learning rate of 0.008, outperforms other configurations in Accuracy, Precision, Recall, and F1-score. However, the training and testing time is minimized with a learning rate of 0.03 with 32-32-64-64 and 16-16-32-32.

Based on the comparison, we can conclude that the convolution kernel 32-32-64-64 with a learning rate of 0.008 is the best among the other configurations. This configuration has been selected as the proposed method for this work. As mentioned in Section 1, this study focuses on binary and multiclass classification.

Fig. 2: (a) depicts the model's accuracy for binary classification across 100 epochs. Similarly, Fig. 2 (b) illustrates the model's accuracy for multiclass classification over the specified epochs. These two figures show that the model's performance improves significantly around 50 epochs and then gradually stabilizes near 100 epochs.



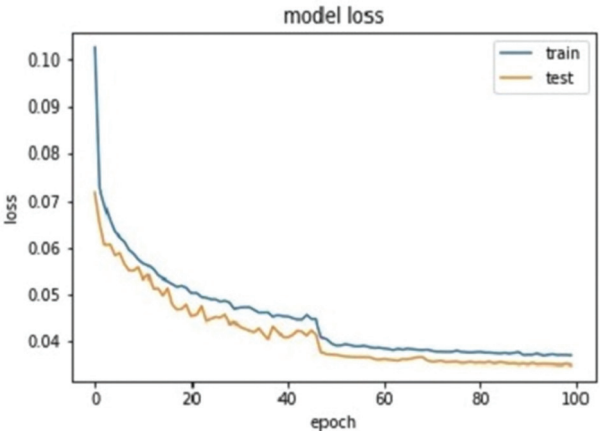
(a)



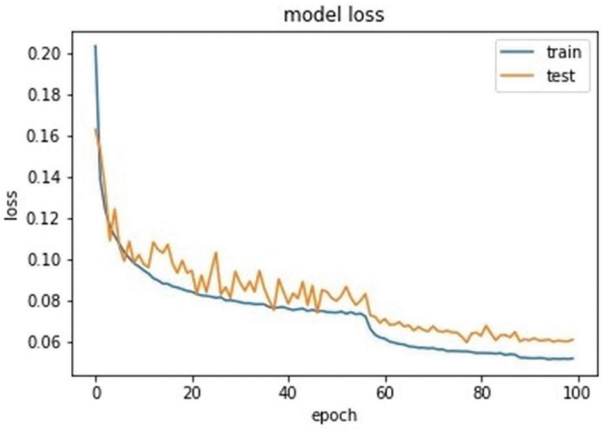
(b)

Fig. 2. Model classification accuracy: (a) Binary (b) Multiclass

Fig. 3: (a) provides a loss of the model for binary classification in the range of 100 epochs. Similarly, Fig. 3: (b) gives the model's loss for multiclass classification in the given epochs. From these two figures, we observe that the loss of the model has dropped around 50 epochs and then slowly stabilized near 100 epochs.



(a)

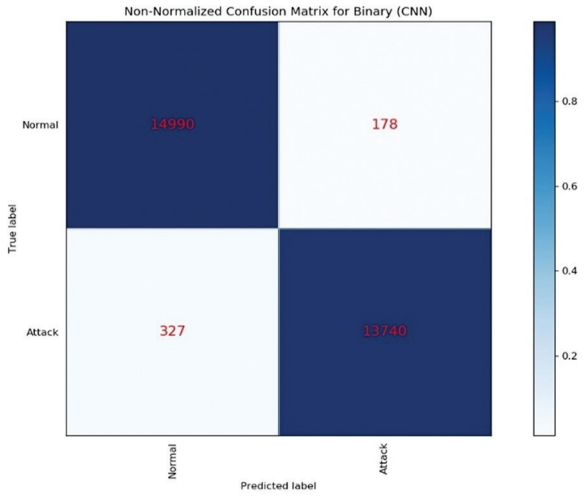


(b)

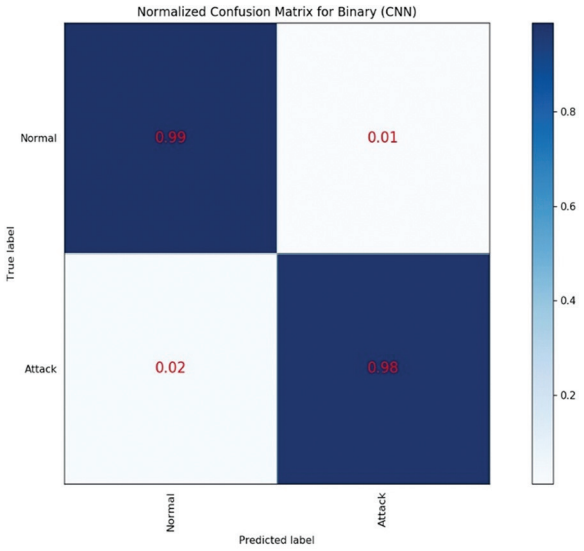
Fig. 3. Model classification loss: (a) Binary (b) Multiclass

Fig. 4: (a) provides a non-normalized confusion matrix of the model for binary classification. Similarly,

Fig. 4: (b) gives the normalized confusion matrix of the model for the same. From these two Figures, we observe that the accuracy performance for binary classification is 0.99 and 0.98 for the normal and attack classes, respectively.

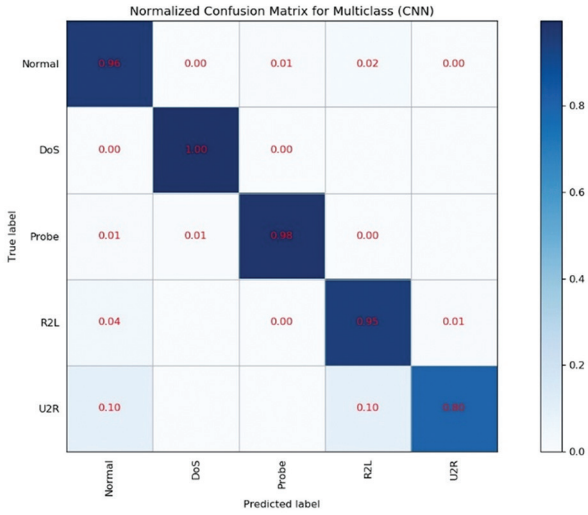


(a)

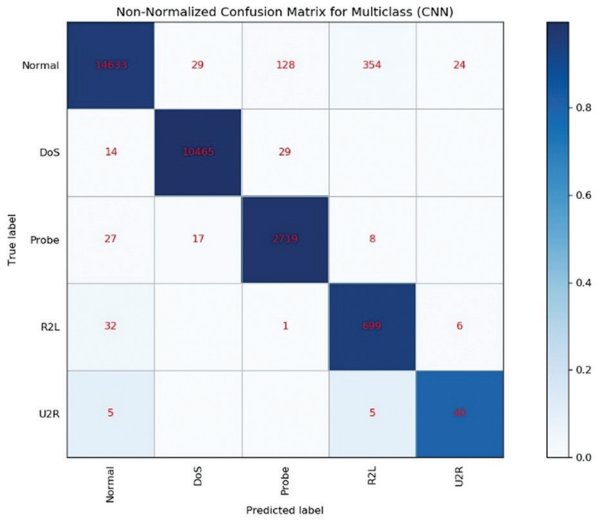


(b)

Fig. 4. Binary classification confusion matrices: (a) Non-normalized (b) Normalized



(a)



(b)

Fig. 5. confusion matrix: (a) Non-normalized (b) Normalized

Fig. 5: (a) presents the non-normalized confusion matrix of the model for multiclass classification. Similarly, Fig. 5 (b) illustrates the normalized confusion matrix of the model for the same task. From these two figures, it is evident that the detection rate performance for multiclass classification is satisfactory. Specifically, in Fig. 5(b), the detection rate for DOS is 100%, while for Probe, R2L, and U2R, the detection rates are 98%, 95%, and 80%, respectively. These results indicate that our approach outperforms three previous works in the domain.

To demonstrate the effectiveness of our method, we have compared our multiclass classification results with some of the previous works in Table 4.

Table 4. Comparison of our results with some of the state-of-the-art

Model	Acc. %	Precision %	DR %	F1-score %
Gaussian–Bernoulli RBM [25]	73.2	62.3	95.1	75.3
ICVAE-DNN [26]	86.0	97.4	77.4	86.3
ID-CVAE [27]	80.1	81.6	80.1	79.1
In this work	**97.7**	**98.1**	**97.7**	**97.8**

7. CONCLUSION AND FUTURE WORK

This research introduces a one-dimensional CNN-based model for intrusion detection. The proposed method comprises three main steps: Data Pre-processing, Hyper-parameter Optimization, and Classification. The number of convolutional kernels and learning rate are two crucial hyperparameters in CNN, so we conducted intensive tuning to identify the best-performing set of parameters. Our experiments showed that the configuration of 32-32-64-64 with a learning rate of 0.008 yielded the best results among all compared configurations. We tested the binary and multiclass classification model on the NSLKDD dataset using Python version 3.6, Keras 2.2.4, and the Tensorflow Engine. The multiclass classification detection rate for DOS was 100%. Similarly, the detection rates for Probe, R2L, and U2R were 98%, 95%, and 80%, respectively. To demonstrate the effectiveness of our method, we compared the multiclass classification results with three previous works. Our method outperformed all four metrics, such as Accuracy, Recall, Precision, and F1-score, with 97.7, 98.1, 97.7, and 97.8 rates, respectively. We plan to conduct further hyperparameter tuning and evaluate the model's performance on different datasets.

8. REFERENCES

[1] S. Mukkamala, G. Janoski, A. Sung, "Intrusion detection using neural networks and support vector machines", Proceedings of the IEEE International Conference on Service-Oriented System Engineering, Oxford, UK, 23-26 August 2021.

[2] M. K. Hooshmand, I. Gad, "Feature selection approach using ensemble learning for network anomaly detection", CAAI Transactions on Intelligent Technology, Vol. 5, No. 4, 2020, pp. 283-293.

[3] J. Kim, J. Kim, H. Kim, M. Shim, E. Choi, "CNN-Based Network Intrusion Detection against Denial-of-Service Attacks", Electronics, Vol. 9, No. 6, 2020, p. 916.

[4] F. Abdulaziz, A. Dahou, M. A. A. Al-cases, S. Lu, M. A. Elaziz, "Advanced Feature Extraction and Selection Approach Using Deep Learning and Aquila Optimizer for IoT Intrusion Detection System", Sensors, Vol. 22, No. 1, 2021, p. 140.

[5] L. Dhanabal, S. P. Shantharajah, "A Study on NSL-KDD Dataset for Intrusion Detection System Based on Classification Algorithms", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 4, No. 6, 2015, p. 6395.

[6] S. Mohammed, "A Machine Learning-Based Intrusion Detection of DDoS Attack on IoT Devices", International Journal of Advanced Trends in Computer Science and Engineering, Vol. 10, No. 4, 2021, pp. 12-45.

[7] M. K. Hooshmand, M. D. Huchaiah, "Network Intrusion Detection with 1D Convolutional Neural Networks", Digital Technologies Research and Applications, Vol. 1, No. 2, 2022, pp. 25-34.

[8] H. Gharaee, H. Hosseinvand, "A new feature selection IDS based on genetic algorithm and SVM", Proceedings of the 8th International Symposium on Telecommunications, Tehran, Iran, 27-28 September 2016, pp. 139-144.

[9] Sumeet Dua, Xian Du, "Data Mining and Machine Learning in Cybersecurity", 1st Edition, Auerbach Publications, 2016, p.256.

[10] S. Maitra, R. K. Ojha, K. Ghosh, Impact of Convolutional Neural Network Input Parameters on Classification Performance", Proceedings of the 4th International Conference for Convergence in Technology, Mangalore, India, 27-28 October 2018.

[11] H. Zhang, C. Q. Wu, S. Gao, Z. Wang, Y. Xu, Y. Liu, "An Effective Deep Learning Based Scheme for Network Intrusion Detection", Proceedings of the 24th International Conference on Pattern Recognition, Beijing, China, 20-24 August 2018, pp. 682-687.

[12] Z. Ahmad, A. S. Khan, K. Nisar, I. Haider, R. Hassan, M. R. Haque, S. Tarmizi, J. J. P. C. Rodrigues, "Anomaly Detection Using Deep Neural Network for IoT Architecture", Applied Sciences, Vol. 11, No. 15, 2021, pp. 7050-7055.

[13] D. Kumar, R. K. Pateriya, R. K. Gupta, V. Dehalwar, A. Sharma, "DDoS Detection using Deep Learning", Procedia Computer Science, Vol. 218, 2023, pp. 2420-2429.

[14] L. Wen, L. Gao, X. Li, B. Zeng, "Convolutional Neural Network With Automatic Learning Rate Scheduler for Fault Classification", IEEE Transactions on Instrumentation and Measurement, Vol. 70, 2021, pp. 1-12.

[15] A. Chakrabarti, S. Shrivastava, "Intrusion Detection System using Long Short-Term Memory and Fully Connected Neural Network on Kddcup99 and NSL-KDD Dataset", International Journal of Intelligent Systems and Applications in Engineering, Vol.11, No. 9s, 2023, pp. 621-635.

[16] D. Kwon, H. Kim, J. Kim, S. C. Suh, I. Kim, K. J. Kim, "A survey of deep learning-based network anomaly detection", Cluster Computing, Vol. 22, No. 1, 2019, pp. 949-961.

[17] R. Sonali, S. Amit, M. Manish, "Intrusion Detection System on KDDCup99 Dataset: A Survey", International Journal of Computer Science and Information Technologies, Vol. 6, No. 4, 2015, pp. 3345-3348.

[18] Y. A. Al-Khassawneh, "An investigation of the Intrusion detection system for the NSL-KDD dataset using machine-learning algorithms", Proceedings of the IEEE International Conference on Electro Information Technology, Romeoville, IL, USA, 18-20 May 2023, pp. 82-87.

[19] Y. Xiao, C. Xing, T. Zhang, and Z. Zhao, "An Intrusion Detection Model Based on Feature Reduction and Convolutional Neural Networks", IEEE Access, Vol. 7, 2019, pp. 42210-42219.

[20] S. W. A. Alsudani, A. Ghazikhani, "Enhancing Intrusion Detection with LSTM Recurrent Neural Network Optimized by Emperor Penguin Algorithm",

Wasit Journal of Computer and Mathematical Science, Vol. 2, No. 3, 2023, pp. 69-80.

[21] J. Jang, Y. An, D. Kim, D. Cho, "Feature Importance-Based Backdoor Attack in NSL-KDD", Electronics, Vol. 12, No. 24, 2023, p. 4953.

[22] H. G. Ahmed, I. E. Samir, E. K. Ayman, "A Proposed Model for Predicting Employee Turnover of Information Technology Specialists Using Data Mining Techniques", International Journal of Electrical and Computer Engineering Systems, Vol. 12, No. 2, 2021, pp. 113-121.

[23] S. M. Kasongo, "A deep learning technique for intrusion detection system using a Recurrent Neural Networks based framework", Computer Communications, Vol. 199, 2023, pp. 113-125.

[24] K. Dinesh, D. Kalaivani, "Enhancing Performance of Intrusion Detection System in the NSL-KDD Dataset using Meta-Heuristic and Machine Learning Algorithms-Design thinking approach", Proceedings of the International Conference on Sustainable Computing and Smart Systems, Coimbatore, India, 14-16 July 2023, pp. 139-144.

[25] Sowmya, T. M. Anita, "An Intelligent Hybrid GA-PI Feature Selection Technique for Network Intrusion Detection Systems", International Journal of Intelligent Systems and Applications in Engineering, Vol. 11, No. 7s, 2023, pp. 718-731.

[26] L. Y. Ahmed, M. M. Hamdy, H. Mahmoud, "Improved DDoS Detection Utilizing Deep Neural Networks and Feedforward Neural Networks as Auto Encoder", Future Internet, Vol. 14, No. 8, 2022, pp. 240-248.

[27] I. Rawaa, T. Abeer, F. Nidaa, "Hybrid Feature Selection Approach to Improve the Deep Neural Network on New Flow-Based Dataset for NIDS," Wasit Journal of Computer and Mathematics Science, Vol. 1, No. 1, 2021, pp. 49-61.

[28] J. Man, G. Sun, "A residual learning-based network intrusion detection system", Security and Communication Networks, Vol. 18, No. 1, 2021, pp. 56-89.

[29] P. Dahiya, D. K. Srivastava, "Network Intrusion Detection in Big Dataset Using Spark", Procedia Computer Science, Vol. 132, 2018, pp. 253-262.

[30] H. Esra'a, A. Hadeel, S. Rizik, A. Orieb, "Hybrid Feature Selection Method for Intrusion Detection Systems Based on an Improved Intelligent Water Drop Algorithm", Cybernetics and Information Technologies, Vol. 22, No. 4, 2022, pp. 73-90.

[31] M. Sheikhan, Z. Jadidi, A. Farrokhi, "Intrusion detection using reduced-size RNN based on feature grouping", Neural Computing and Applications, Vol. 21, No. 6, 2012, pp. 1185-1190.

[32] M. Idhammad, K. Afdel, M. Belouch, "DoS Detection Method based on Artificial Neural Networks", International Journal of Advanced Computer Science and Applications, Vol. 8, No. 4, 2017, pp. 465-471.

[33] Y. Fu, Y. Du, Z. Cao, Q. Li, W. Xiang, "A Deep Learning Model for Network Intrusion Detection with Imbalanced Data", Electronics, Vol. 11, No. 6, 2022, pp. 898-904.

Federated Learning Implementation with Privacy Leakage Prevention for Hand-Written Digit Recognition

Original Scientific Paper

N. Indira Priyadarsini
Department of Computer Science Engineering,
Koneru Lakshmaiah Education Foundation,
Vaddeswaram, AP, India
nethalapriya@gmail.com

Dr G. Raja
Department of Computer Science Engineering,
Koneru Lakshmaiah Education Foundation
Vaddeswaram, AP, India
rajajjcet06@kluniversity.in

Abstract *– Federated learning (FL) has brought significant advantages to applications where collaborative learning should occur at multiple participating devices to enhance user experience in specific tasks. However, FL results in privacy leakage when n-1 clients collude to infer the model of another client. In this paper, we not only implement an FL framework but propose a methodology for preventing privacy leakage while realizing machine learning-based automatic hand-written digit recognition. Our framework supports the FL of deep networks where models trained locally are averaged. Two machine learning models Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP) are implemented with FL. We proposed an algorithm, Federated Averaging with Privacy Leakage Prevention (FA-PLP), for model averaging to be done by the server. Our algorithm exploits differential privacy (DP) for realizing model averaging while getting rid of chances of privacy leakage. We evaluated our framework with two distributions of the MNIST dataset. Our empirical results revealed that FA-PLP with the CNN model could achieve the highest accuracy of 95.38%.*

Keywords*: Federated Learning, Machine Learning, Deep Learning, Privacy, Collaborative Machine Learning*

Received: December 16, 2023; Received in revised form: March 22, 2024; Accepted: April 1, 2024

1. INTRODUCTION

Federated learning (FL) is a novel phenomenon in which multiple distributed clients are involved in the machine learning process collaboratively while preserving the privacy of locally available training data. Though FL minimizes privacy risk, it still may cause leakage of information about local training data in terms of the model's parameters or weights. Therefore, it is indispensable to overcome this problem by proposing algorithms to realize ML models while preserving privacy. With the emergence of fog computing and edge computing, it is made possible for diversified computing devices can participate in the FL process. For instance, modern smartphones when involved in FL can result in a rich user experience [1]. FL enables ML models to be trained in remote clients while localizing training data. A real-world example for FL is that in the healthcare domain, many hospitals (clients) can collaboratively participate in training a model to leverage prediction accuracy for a given disease diagnosis. FL assumes significance when the clients are not willing to share their training data due to locally prevailing privacy policies.

Many research endeavours are found in the literature on FL. Tao et al. [2] address privacy concerns in Vehicular Edge Computing (VEC) with Federated Learning (FL) in autonomous driving, considering malicious parties. Kang et al. [3] introduced FedGRU, a federated learning-based traffic flow prediction algorithm that maintains privacy while achieving accurate predictions. Zhao et al. [4] proposed a smart home system using federated learning (FL) and a reputation mechanism to help home appliance manufacturers improve their products. Zhang et al. [5] introduced VFL, a privacy-preserving and verifiable federated learning method for big data in industrial IoT, enabling effective verification with constant overhead. Fang et al. [6] introduced an efficient, privacy-preserving federated learning

(HFWP) scheme for cloud computing. It is observed from the literature that FL has significant limitations such as probably insecure communication and privacy leakage. Privacy leakage occurs when n-1 clients collude to infer the model of another client. In this paper, we focus on proposing a framework which addresses privacy concerns in FL. Our contributions to this paper are as follows.

1. We proposed an FL framework along with a methodology for preventing privacy leakage while realizing machine learning-based automatic handwritten digit recognition.
2. We proposed an algorithm known as Federated Averaging with Privacy Leakage Prevention (FA-PLP) for model averaging to be done by the server. It addressed the problem of n-1 clients colluding to infer the model of another client (privacy leakage).
3. We built an application to evaluate our FL framework using machine learning techniques like CNN and MLP, for automatic handwritten digit recognition, on two data distributions.

The remainder of the paper is structured as follows. Section 2 reviews existing FL methods and their limitations. Section 3 presents the proposed FL framework with underlying mechanisms and algorithms. Section 4 presents the results of our experiments with two data distributions. Section 5 concludes our work and provides directions for the future scope of the research.

2. RELATED WORK

This section reviews existing methods on FL. Chunyi et al. [1] proposed a fog computing scheme to enhance federated learning, bolstering IoT data privacy and security against various attacks. Demonstrated efficiency and potential for further improvements. Li et al. [2] address privacy concerns in Vehicular Edge Computing (VEC) with Federated Learning (FL) in autonomous driving, considering malicious parties. FL improves training efficiency and privacy, reducing training loss by 73.7% and enhancing accuracy in simulations under different scenarios. The proposed system significantly reduces bandwidth requirements.

Yi et al. [3] introduced FedGRU, a federated learning-based traffic flow prediction algorithm that maintains privacy while achieving accurate predictions. It outperforms state-of-the-art methods in privacy preservation, demonstrating minimal accuracy loss. In Further the work is to enhance prediction accuracy using a Graph Convolutional Network (GCN).

Yang et al. [4] proposed a smart home system using federated learning (FL) and a reputation mechanism to help home appliance manufacturers improve their products. The system involves two stages: customers train an initial model provided by the manufacturer using mobile phones and edge computing. Differential privacy protects features and ensures privacy. The proposed approach guarantees accuracy and data privacy. Anmin et al. [5] introduced VFL, a privacy-preserving and verifiable federated learning method for big data in industrial IoT, enabling effective verification with constant overhead. Experimental results support its efficiency. Chen et al. [6] introduced an efficient, privacy-preserving federated learning (HFWP) scheme for cloud computing. It employs lightweight encryption and optimization strategies. The approach is secure, improves efficiency, and is suitable for cloud and fog computing applications, offering possibilities for further research, including combining SMC with DP and exploring alternative SMC techniques like Pallier. The private leakage prevention approach in FL in the proposed methodology in this paper is different from [6] in both client-side and server-side phenomena besides in the usage of differential privacy.

Zhao et al. [7] proposed a privacy-preserving federated learning approach for industrial big data. It minimizes parameter sharing, uses differential privacy with a Gaussian mechanism, a proxy server for anonymity, and a self-stop mechanism to enhance privacy while maintaining accuracy and performance and It is also related to the previous article.

Yu et al. [8] proposed a privacy-preserving federated learning scheme that ensures both privacy and integrity through a Trusted Execution Environment (TEE). This scheme addresses causative attacks, making collaborative deep learning more secure and practical. It aims to bring the benefits of deep learning to domains with privacy and availability concerns.

Elgabli et al. [9] proposed an analog-based federated learning framework, that addresses wireless channel challenges to improve privacy, bandwidth efficiency, and scalability. It uses analogue transmissions, preserving data privacy and demonstrating effectiveness under various conditions. Major contributions include theoretical advancements and algorithmic innovations. Yang et al. [10] introduced an asynchronous federated learning (AFL) framework for multi-UAV networks, allowing local model training without transmitting raw data. It employs device selection and an A3C-based algorithm to improve learning accuracy and speed. Simulations confirm its superior performance. Yunlong et al. [11] presented a blockchain-based secure data-sharing system for Industrial IoT, integrating federated learning into permissioned blockchain for data privacy and efficiency. Numerical results validate its effectiveness. Future work should explore further security threats, enhance data model utility, and address resource constraints in IIoT data sharing. Xiaoxiao et al. [12] introduced a privacy-preserving federated learning framework for multi-site fMRI analysis, overcoming privacy concerns and enhancing neuroimage analysis. It offers potential benefits in other medical data analysis fields. The approach allows data from various institutions to be utilized while safeguarding privacy, and fostering collaboration in medical research.

Xiaofeng et al. [13] explored real-time data sharing for smart cities must ensure privacy. An adaptive pseudonymization framework enhances privacy robustness in real-time information brokering, with early positive results. Future work includes comprehensive validation and consideration of potential multi-dimensional correlation attacks. The approach could be applied to various information sources beyond energy data. Islam [14] focused on enhancing Federated Learning (FL) for Electronic Health Records (EHRs) by ensuring privacy through techniques such as data generalization, feature selection, and noise minimization. A distributed framework is proposed where local models make predictions based on local features, with added privacy protection using differential privacy. Weighted feature functions ensure a balanced trade-off between privacy and utility. No raw data, features, or model parameters are shared. The method aims to maintain data localization and can be applied to healthcare data, with the potential for future comparisons and improvements.

Chamikara et al. [15] introduce a distributed perturbation algorithm called DISTPAB, addressing privacy concerns in distributed machine learning for geographically dispersed data, like healthcare and banking. DISTPAB shows minimal utility degradation and serves as a promising privacy preservation method for distributed machine learning. Future work will explore further efficiency improvements, particularly in the context of vertical federated learning with varying feature spaces.

Zhang et al. [16] discussed federated learning for privacy-preserving medical models in IoT-based healthcare. It uses cryptographic techniques and data quality weighting. The proposed scheme maintains privacy, and the experiments indicate promising accuracy for lesion cell type detection. In future, it includes optimizing for heterogeneous environments and addressing malicious server issues.

Jiang et al. [17] introduced PFLM, a privacy-preserving federated learning scheme with membership proof, addressing the dropout constraint while ensuring security and verifiability. Security analysis and experiments confirm its efficiency. Yuanhang et al. [18] found that a blockchain-based federated learning system ensures secure and privacy-preserving traffic flow prediction by decentralizing model updates and applying differential privacy. Yin et al. [19] proposed a novel hybrid privacy-preserving federated learning approach that uses advanced encryption, noise addition, and sparse differential gradients to enhance security and efficiency. Yunlong et al. [20] describe an intelligent, secure architecture and privacy-preserving federated learning in VCPS to combat data leakage effectively, ensuring accuracy and security. Ma et al. [21] A privacy-preserving Byzantine-robust federated learning scheme (PBFL) enhances robustness and privacy by using encryption and zero-knowledge proof, providing higher privacy protection. Xiaoyuan et al. [22] introduced an Adaptive Privacy-preserving Federated Learning framework with differential privacy. It uses relevance propagation and adjustment technology to optimize the trade-off between accuracy and privacy, demonstrated through formal analysis and experiments.

Shixiang et al. [23] presented CI-PPFL, a class-imbalance privacy-preserving federated learning framework for decentralized wind turbine fault diagnosis. Experiments on real-world data show its superiority and privacy preservation. Future work includes extending it to heterogeneous label subspaces and integrating vibration data and SCADA data for broader applications. Wei et al. [24] observed that UDP algorithm adds artificial noise to shared models in Federated Learning, ensuring user-level differential privacy. CRD method enhances learning efficiency and model quality for specified privacy levels. Future work aims to refine privacy budget allocation.

Ali et al. [25] explored privacy concerns in IoMT by introducing federated learning (FL) as a solution. It surveys privacy issues in IoMT, discusses existing privacy techniques, and emphasizes FL's collaborative, privacy-preserving nature. The survey further explores FL's advanced architectures with DRL, DNN, and GANs. Finally, it suggests real-time applications and future research directions for improving privacy in smart healthcare systems.

Kong et al. [26] focused on privacy-preserving, flexible model aggregation in federated learning-based automotive navigation called FedLoc. Extensive analysis demonstrates its privacy and security properties, along with improved computational efficiency during participant changes. Future work includes real-world testing and performance assessment. Han et al. [27] proposed a verifiable federated learning scheme is for deep neural networks. It addresses privacy, trust, and accuracy concerns using key exchange, double masking, and tag aggregation. Security and efficiency analyses confirm its effectiveness. Fang et al. [28] introduced PCFL, a privacy-preserving, communication-efficient federated learning approach for IoT. PCFL excels in communication efficiency and model accuracy. Future work targets multi-task learning and advanced cryptographic protocols for IoT security. Tian et al. [29] explored federated learning's unique attributes and challenges, highlighting its distinct nature compared to traditional machine learning. It provides an overview of current approaches and identifies areas for future interdisciplinary research. Cheng et al. [30] introduced SecureBoost, a privacy-preserving tree-boosting system in the context of federated learning, offering accuracy comparable to non-private methods. Information leakage is analysed, and solutions are suggested.

Huafei et al. [31] studied privacy-preserving weighted federated learning within a secret-sharing framework. It introduces weighted federated learning (wFL) and presents its implementation using random splitting and ElGamal encryption. The proposed solution is secure against honest-but-curious adversaries.

Wang et al. [32] introduced VANE, a secure and non-interactive federated learning scheme for regression training with gradient descent. VANE facilitates training global regression models while preserving data privacy. It features a secure data aggregation algorithm and improved training efficiency. Security analysis and experiments demonstrate its effectiveness. Li et al. [33] reviewed the evolution of Federated Learning (FL) in industrial engineering and computer science. It identifies research fronts, summarizes applications, and outlines FL's development prospects. This comprehensive analysis aims to guide future applications and address remaining challenges in FL. Lakhan et al. [34] discussed privacy and fraud issues in machine-learning-based Internet of Medical Things (IoMT) systems. It introduces the FL-BETS framework, focusing on healthcare applications with energy and delay constraints. FL-BETS outperforms existing models. Future work aims to address mobility fraud and extend security measures. Jie et al. [35] stated that the proliferation of healthcare data offers significant potential for improving care, but privacy challenges and data fragmentation persist. This survey reviews federated learning technologies, including their application in healthcare, addressing statistical, system, and privacy challenges. Challenges such as data quality and standardization in healthcare data are also discussed.

Liu et al. [36] observed that edge computing is a technology to extends cloud services to the network edge, raises privacy concerns with user data transmission. P2FEC integrates federated learning and edge computing to preserve privacy and build deep learning models without central data storage, outperforming standard edge computing in privacy protection. Future work is to enhance protection against privacy-sensitive data leakage. Zengpeng et al. [37] introduced a triple-band cylindrical dielectric resonator antenna (CDRA) with HEM11, TM01, and HEM12 modes excited simultaneously using a composite feeding structure. Diverse radiation patterns make it suitable for various wireless applications, including WiMAX and vehicular use. Wang et al. [38] discussed the privacy issues in federated learning, particularly in ternary federated learning (TernGrad). This innovative approach improves communication efficiency and accuracy, representing the first research combining ternary federated learning with privacy-preserving technologies. Future work includes enhancing efficiency and security. Wei et al. [39] proposed NbAFL, a differential privacy-based approach in federated learning to enhance privacy, involving noise, trade-offs, simulations, and future considerations. Aledhari et al. [40] provided a comprehensive study of Federated Learning (FL), highlighting its importance, enabling technologies, and challenges. It explores real-life applications and suggests directions for the future. FL holds the potential to improve data handling and privacy, but challenges such as fault tolerance, performance, and fairness need addressing in its implementation. From the review of literature issues like privacy and security in communications were still found possible. In this paper, we focus on proposing a framework which addresses privacy concerns.

3. PROPOSED FRAMEWORK

This section presents the system model, problem definition, our methodology for federated learning implementation with privacy leakage prevention for hand-written digit recognition and the proposed algorithm.

3.1. SYSTEM MODEL AND PROBLEM STATEMENT

Let us consider a distributed environment where multiple mobile devices participate in language modelling tasks to recognize hand-written digits. All participating mobile devices train an ML model in a collaborative fashion. Each device trains a model ΔWi locally instead of sending its training data to a remote server. Therefore, each mobile device is known as a client which needs to communicate with the server to send local model to it. The server is responsible for computing a global model send it back to each client. The training process is repeated until it reaches a stopping condition or convergence. The system model with the FL approach is illustrated in Fig. 1.

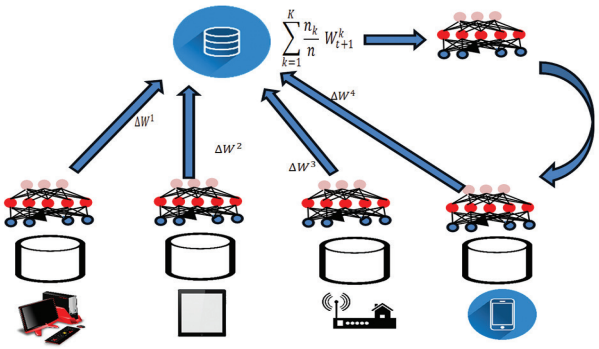


Fig. 1. Illustrates our system model for federated learning

An important advantage of FL is that it is able to decouple model training process and gets rid of direct access to training data. However, the server is essential to coordinate the training process. Therefore, it is essential to have trust in the server or assume it. Nevertheless, there is privacy achieved due to the non-sharing of locally available training data. Thus FL has the potential to minimize security and privacy risks as the attack surface is reduced to the device instead of the attack surface encompassing to entire environment, probably, including the cloud. FL is found ideal for solving many kinds of problems that share common qualities such as distributed availability of data in multiple devices, data is privacy-sensitive and supervised learning where labels can be interactively inferred. However, FL has significant limitations such as probably insecure communication and privacy leakage.

Privacy leakage occurs when n-1 clients collude to infer model of another client. The former (security problem) can be overcome by implementing a secure multiparty communication (MPC) system while the latter (privacy leakage) can be implemented with differential privacy (DP) at each client. In this paper, we focused on FL with privacy-preserving model training through DP implementation.

3.2. OUR METHODOLOGY

The proposed methodology for FL with privacy leakage prevention is based on the system model illustrated in Fig. 1. We considered the problem of privacy leakage which occurs when n-1 clients collude to infer the model of another client. Federated learning, due to its modus operandi, has specific privacy advantages. However, privacy leakage occurs when n-1 clients collude to infer the model of another client. In the FL task, there is a minimal update required to improve model. Privacy depends on the content that needs to be updated in the learning process. Nevertheless, the updates are generally minimal and the source of the update is not required by the aggregation process. Still there is the probability of n-1 clients colluding to cause privacy leakage. Our implementation overcomes this issue as it takes care of privacy-preserving model training. We combine FL with differential privacy to ensure the prevention of privacy leakage in FL. An asynchronous scheme is considered for updates while proceeding with federated communication. A number of clients involved in FL I fixed and their local dataset is also fixed. When each round starts, a fraction of clients are chosen randomly and a global state is obtained from the server. The notion of selecting a fraction of clients is to improve efficiency. Clients perform computation locally on the locally available dataset depending on the global state provided by the server. The result of local computation is sent to the server. Afterwards, the server uses the updates to modify the global state and this procedure is done repeatedly. We considered a finite-sum-based objective for FL as expressed in Eq. 1.

$$\min_{w\in R^d} f(w) \qquad \text{where} \qquad f(w)\stackrel{def}{=}\frac{1}{n}\sum_{i=1}^{n} f_i(w) \tag{1}$$

For given ML problem, $f_i(w)=l(x_i,y_i,w)$ is considered where w denotes model parameters and (x_i,y_i) is the given example on which loss is computed. Assuming that there are k number of clients and data is partitioned accordingly consisting of indexes P_k associated to data in client k and $n_k=|P_k|$ where P_k is the partition. Then the objective can be modified as expressed in Eq. 2.

$$f(w)=\sum_{k=1}^{K}\frac{n_k}{n}F_k(w) \text{where } F_k(w)=\frac{1}{n_k}\sum_{i\in P_k} f_i(w) \tag{2}$$

P_k is the partition associated with training examples for different clients distributed randomly, it forms the expression $E(P_k)[F_k(w)]=f(w)$. It was observed empirically that in FL communication costs are more than computational costs, unlike the data centre-based approach. In our implementation, each client is involved in less number of updates in FL.

To ensure the prevention of the possibility of privacy leakage in FL, we used differential privacy (DP) which helps in adding noise so as to address privacy attacks. DP is the mathematical model to ensure the privacy of data being exchanged among participants in FL. The DP in its simplest form can be expressed as in Eq. 3.

$$Pr[M(D_1)\in S]\leq e^{\in} Pr[M(D_2)\in S]+\delta \tag{3}$$

When the DP mechanism satisfies expression, it can be used to add noise to the data so as to ensure privacy-preserving communication among clients and servers in FL. It is known as ϵ-differential privacy as discussed in [6]. We consider the Laplacian mechanism that has the potential to preserve ϵ-differential privacy. Considering random noise X, concerning Laplacian distribution, the PFF is expressed in Eq. 4.

$$f(x)=\frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}} \tag{4}$$

where λ denotes scale parameter, X is the random noise and the scale value is expressed as in Eq. 5.

$$\lambda=\Delta/\epsilon \tag{5}$$

We also support a distributed approach in adding noise. In this approach, each client adds its portion of noise. Since DP is compatible with the Laplace mechanism, it is possible to generate a Laplace random variable as expressed in Eq. 6.

$$L(\mu,\lambda)=\frac{\mu}{n}+\sum_{p=1}^{n}\gamma_p-\gamma_p', \tag{6}$$

where γ_p and γ_p' are random variables as per Gamma distribution, μ and λ denote mean and scale parameters respectively in the Laplace mechanism. Now this leads to the expression in Eq. 7.

$$\frac{(1/s)^{1/n}}{\Gamma(1/n)}x^{\frac{1}{n}-1}e^{-x/s} \tag{7}$$

where s and $1/n$ denote scale and shape parameters respectively. A technique expressed in Eq. 8 is used for each client adding γ_p-γ_p' in the proposed algorithm which makes use of distributed privacy.

Algorithm 1: Federated Averaging with Privacy Leakage Prevention

Server Side:
Server initializes w_0
For each round r in R
    For each client k in K
$w^k_{r+1}\leftarrow$ClientSideProcess(k, w_r)
    End For

$$w_{r+1}\leftarrow\frac{1}{n}\sum_{i=1}^{n}w^i_{r+1}$$

End For
***ClientSideProcess* (k, w)**:
For each local update u in U
$w\leftarrow w-\eta\nabla g(w)$
End For
Return $w+\gamma-\gamma'$

As presented in Algorithm 1, there are a number founds in which communication takes place between servers and clients as part of FL. In the process, there is server-side functionality and also client-side functionality. In the local updates about weights, noise is added by each client leading to a distributed approach to noise addition. This has the potential to prevent n-1 clients from colluding to infer models of another client. Thus the proposed algorithm helps in preventing privacy leakage. Each client compares its weight with that of the previous round where the server sends global weights to the client. As each client is contributing to the noise addition, it has a more efficient privacy-preserving mechanism in FL. Moreover, on local convergence, each client can come out of the FL system. Each time a client receives federated weight from the server, it can subtract the DP noise it has contributed for those federated weights. With this modus operandi, the proposed FL achieved privacy by defeating any privacy attacks besides supporting the inherent privacy involved in FL.

4. EXPERIMENTAL RESULTS

We made experiments with our implemented prototype for realizing FL. The MNIST dataset used for the empirical study is collected from [41]. The dataset is partitioned as the number of clients involved in the FL. Two approaches are followed to partition data over clients. The first approach simply shuffles the dataset D and distributes it among clients. We call it as D1. The second approach sorts the dataset D based on the digit label, divides it into a number of shards of a given size and each client is provided with a specified number of shards. This is called D2. Experiments are made with both D1 and D2. Models such as CNN and MLP are used for realizing FL.

Table 2. Parameters of CNN along with their values

Parameter	Description	Value
rounds	Number of training rounds	100
C	Client fraction	0.1
K	Number of clients	100
E	Number of training passes on a local dataset for each round	5
batch_size	Batch size	10
LR	Learning rate	0.01

Table 2 shows the parameters used for the CNN model. It uses 100 clients and 100 training rounds with a learning rate of 0.001 and a batch size of 10.

Table 3. Parameters of MLP along with their values

Parameter	Description	Value
rounds	Number of training rounds	100
C	Client fraction	0.1
K	Number of clients	100
E	Number of training passes on a local dataset for each round	5
batch_size	Batch size	10
LR	Learning rate	0.03

Table 3 shows the parameters used for the MLP model. It uses 100 clients and 100 training rounds with a learning rate of 0.03 and batch size 10.

4.1. DATA VISUALIZATION

The dataset collected from [41] is used for experiments. It is related to hand-written digits. It is widely used in machine learning for language modelling and other related applications.

Fig. 2. An excerpt from training data

Fig. 3. An excerpt from test data

Fig. 2 shows an excerpt from training data while Figure 3 presents an excerpt from test data. The dataset is used for hand-written text recognition tasks with FL approach.

4.2. RESULTS OF THE CNN MODEL

Experimental results of FL with CNN model are presented in this section. It provides average loss dynamics and accuracy of the CNN model for two dataset distributions namely D1 and D2.

Table 4. Average loss exhibited by CNN for D1

# Rounds	Average Loss
Round 1	0.818
Round 10	0.04
Round 20	0.026
Round 30	0.02
Round 40	0.018
Round 50	0.014
Round 60	0.013
Round 70	0.013
Round 80	0.009
Round 90	0.008
Round 100	0.006

As presented in Table 4 the average loss exhibited by CNN in FL against different numbers of rounds is provided for D1.

As presented in Fig. 4, the average loss value exhibited by CNN in FL is gradually decreases as the number of rounds is increased. At round 1 the average loss is exhibited as 0.818. The observation at round 10 is reduced to 0.04. When the number of rounds is increased to 50, the average loss value is 0.014. When the number of rounds reaches 100, the average loss observed is the

least with 0.006. These observations are recorded when D1 is used for experiments. Less average loss indicates better performance.

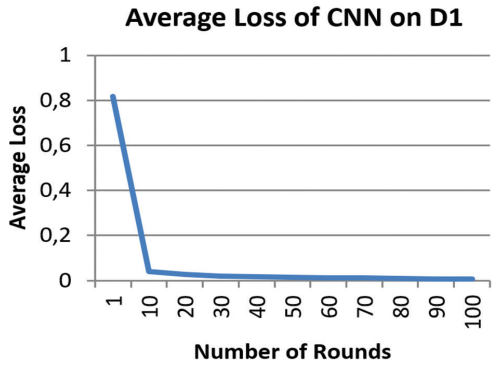


Fig. 4. Average loss of CNN in FL against number of rounds when D1 is used

Table 5. Average loss exhibited by CNN for D2

# Rounds	Average Loss
Round 1	0.097
Round 10	0.021
Round 20	0.017
Round 30	0.008
Round 40	0.012
Round 50	0.007
Round 60	0.006
Round 70	0.006
Round 80	0.006
Round 90	0.006
Round 100	0.004

As presented in Table 5 the average loss exhibited by CNN in FL against different numbers of rounds is provided for D2.

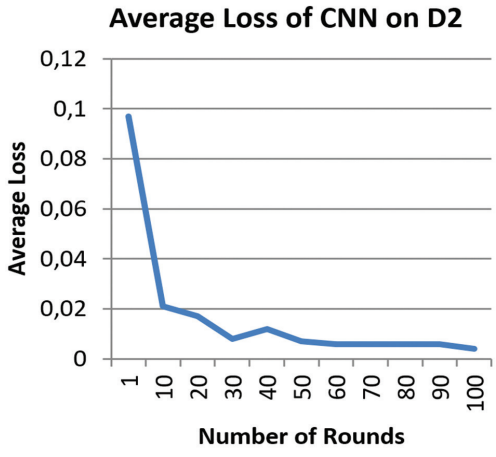


Fig. 5. Average loss of CNN in FL against number of rounds when D2 is used

As presented in Fig. 5, the average loss value exhibited by CNN in FL gradually decreased as the number of rounds is increased. At round 1 the average loss is exhibited as 0.097. The observation at round 10 it is reduced to 0.021. When the number of rounds is increased to 50, the average loss value is 0.007. When the number of rounds reaches 100, the average loss observed is the least with 0.0044. These observations are recorded when D2 is used for experiments.

Table 6. Performance of CNN with FL

Model & Dataset	Accuracy (%)
CNN with D1	95.3856
CNN with D2	94.0512

As presented in Table 6, the performance of the CNN model with the two data distributions is provided in terms of accuracy achieved in hand-written digit recognition.

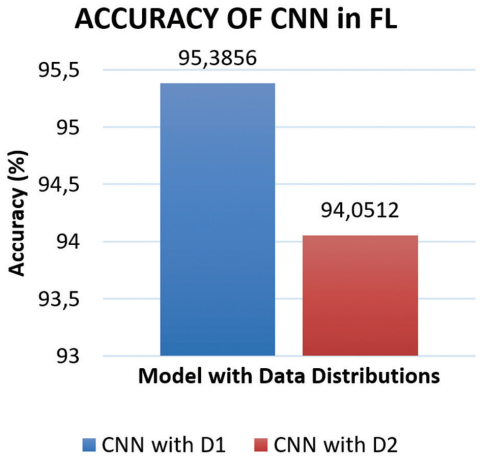


Fig. 6. Accuracy exhibited by CNN with FL when two data distributions are used

As presented in Fig. 6, the accuracy of CNN model in FL with two data distributions is compared. CNN model with D1 achieved better performance with 95.38% accuracy. With D2, the CNN model in FL could achieve 94.05% accuracy.

4.3. RESULTS OF MLP MODEL

Experimental results of FL with MLP model are presented in this section. It provides average loss dynamics and accuracy of the MLP model for two dataset distributions namely D1 and D2.

Table 7. Average loss exhibited by MLP for D1

# Rounds	Average Loss
Round 1	0.607
Round 10	0.059
Round 20	0.032
Round 30	0.026
Round 40	0.027
Round 50	0.017
Round 60	0.018
Round 70	0.013
Round 80	0.01
Round 90	0.013
Round 100	0.008

As presented in Table 7 the average loss exhibited by MLP in FL against different numbers of rounds is provided for D1.

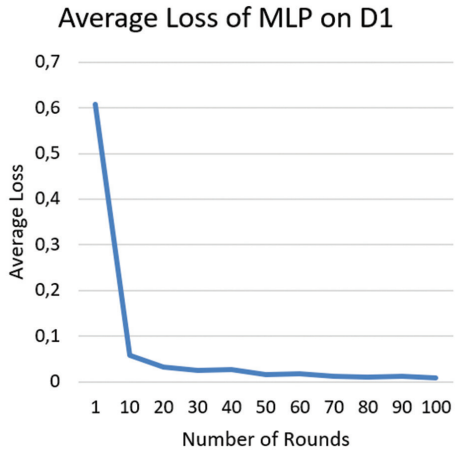


Fig. 7. Average loss of MLP in FL against number of rounds when D1 is used

As presented in Fig. 7, the average loss value exhibited by MLP in FL is gradually decreased as the number of rounds is increased. At round 1 the average loss is exhibited as 0.607. The observation at round 10 it is reduced to 0.059. When the number of rounds is increased to 50, the average loss value is 0.017. When the number of rounds reaches 100, the average loss observed is the least with 0.008. These observations are recorded when D1 is used for experiments. Less average loss indicates better performance.

Table 8. Average loss exhibited by MLP for D2

# Rounds	Average Loss
Round 1	0.125
Round 10	0.022
Round 20	0.013
Round 30	0.011
Round 40	0.014
Round 50	0.005
Round 60	0.008
Round 70	0.004
Round 80	0.012
Round 90	0.006
Round 100	0.008

As presented in Table 8 the average loss exhibited by MLP in FL against different numbers of rounds is provided for D2.

As presented in Fig. 8, the average loss value exhibited by MLP in FL is gradually decreased as the number of rounds is increased. At round 1 the average loss is exhibited as 0.125. The observation at round 10 it is reduced to 0.022. When the number of rounds is increased to 50, the average loss value is 0.005. When the number of rounds reaches 100, the average loss observed is the least with 0.008. These observations are recorded when D2 is used for experiments.

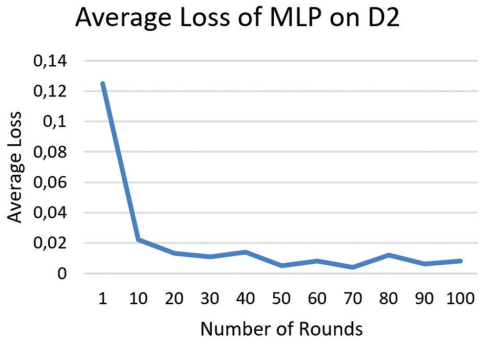


Fig. 8. Average loss of MLP in FL against number of rounds when D2 is used

Table 9. Performance of MLP with FL

Model & Dataset	Accuracy (%)
MLP with D1	93.3216
MLP with D2	90.4032

As presented in Table 9, the performance of the MLP model with the two data distributions is provided in terms of accuracy achieved in hand-written digit recognition.

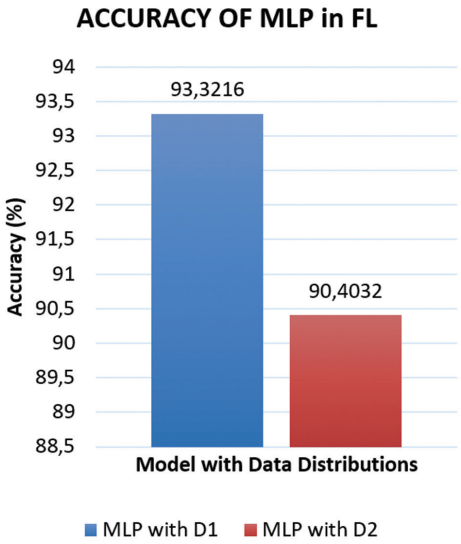


Fig. 9. Accuracy exhibited by MLP with FL when two data distributions are used

As presented in Fig. 9, the accuracy of the CNN model in FL with two data distributions is compared. MLP model with D1 achieved better performance with 93.32% accuracy. With D2, the MLP model in FL could achieve 90.40% accuracy.

4.4. PERFORMANCE COMPARISON

The performance of MLP and CNN models in FL is evaluated in terms of accuracy. The observations are made in this section with two data distributions.

As presented in Table 10, a performance comparison between MLP and CNN in FL is made in terms of accuracy in handwritten digit recognition.

Table 10. Performance comparison between MLP and CNN in FL

Model & Dataset	Accuracy (%)
MLP with D2	90.4032
MLP with D1	93.3216
CNN with D2	94.0512
CNN with D1	95.3856

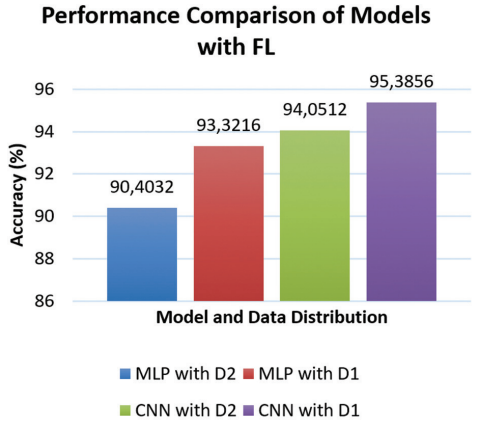


Fig. 10. Accuracy exhibited by MLP and CNN with FL when two data distributions are used

As presented in Fig. 10, the two models such as MLP and CNN are used in FL with two data distributions. The accuracy of MLP with D2 is 90.40%, MLP with D1 93.32%, CNN with D2 94.05% and CNN with D1 95.38%. Highest accuracy achieved by the CNN model with D2 is 95.38%.

Table 11. Performance comparison with state-of-the-art

FL Model	Accuracy (%)
Chen et al. [42]	93.2145
Ng et al. [43]	93.4231
FA-PLP (Proposed)	95.3856

Our results are compared with state-of-the-art methods such as Chen et al. [42] and Ng et al. [43] as presented in Table 11.

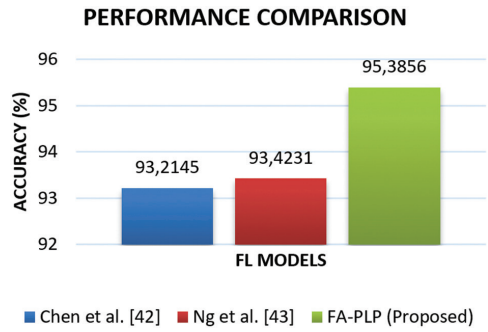


Fig. 11. Performance comparison of FL models

The performance of the proposed model named FA-PLP is compared against existing models. The results revealed that FP-PLP outperforms other models in terms of accuracy with 95.3866%.

5. CONCLUSION AND FUTURE WORK

In this paper, we not only implement an FL framework but propose a methodology for preventing privacy leakage while realizing machine learning-based automatic hand-written digit recognition. Our framework supports the FL of deep networks where models trained locally are averaged. Two models Convolutional Neural Network (CNN) and Multilayer Perceptron (MLP) are implemented with FL. We proposed an algorithm, Federated Averaging with Privacy Leakage Prevention (FA-PLP), for model averaging to be done by the server. Our algorithm exploits differential privacy (DP) for realizing model averaging while getting rid of chances of privacy leakage. We evaluated our framework with two distributions of the MNIST dataset. Our empirical results revealed that FA-PLP outperforms existing FL techniques in terms of communication cost, accuracy and privacy leakage prevention. Our framework with the CNN model could achieve the highest accuracy of 95.38%. In future, we intend to improve our framework further by considering the security concerns of FL as well. We also elaborate on different privacy attack scenarios and system behaviours in our future research.

6. REFERENCES

[1] Z. Chunyi, F. Anmin, Y. Shui, Y. Wei, W. Huaqun, Z. Yuqing, "Privacy-Preserving Federated Learning in Fog Computing", IEEE Internet of Things Journal, Vol. 7, No. 11, 2020, pp. 10782-10793.

[2] Y. Li, X. Tao, X. Zhang, J. Liu, J. Xu, "Privacy-Preserved Federated Learning for Autonomous Driving", IEEE Transactions on Intelligent Transportation Systems, Vol. 23, No. 7, 2022, pp. 8423-8434.

[3] L. Yi, Y.J. Q. James, K. Jiawen, N. Dusit, Z. Shuyu, "Privacy-preserving Traffic Flow Prediction: A Federated Learning Approach", IEEE Internet of Things Journal, Vol. 7, No. 8, 2020, pp. 7751-7763.

[4] Z. Yang, Z. Jun, J. Linshan, T. Rui, N. Dusit, L. Zengxiang, L. Lingjuan, L. Yingbo, "Privacy-Preserving Blockchain-Based Federated Learning for IoT Devices", IEEE Internet of Things Journal, Vol. 8, No. 3, 2020, pp. 1817-1829.

[5] F. Anmin, Z. Xianglong, X. Naixue, G. Yansong, W. Huaqun,Z. Jing, "VFL: A Verifiable Federated Learning with Privacy-Preserving for Big Data in Industrial IoT", IEEE Transactions on Industrial Informatics, Vol. 18, No. 5, 2022, pp. 3316-3326.

[6] F. Chen, G. Yuanbo, W. Na, J. Ankang, "Highly efficient federated learning with strong privacy pres-

ervation in cloud computing", Computers & Security, Vol. 96, 2020.

[7] B. Zhao, K. Fan, K. Yang, Z. Wang, H. Li, Y. Yang, "Anonymous and Privacy-Preserving Federated Learning With Industrial Big Data", IEEE Transactions on Industrial Informatics, Vol. 17, No. 9, 2021, pp. 6314-6323.

[8] C. Yu, L. Fang, L. Tong, X. Tao, L. Zheli, L. Jin, "A Training-integrity Privacy-preserving Federated Learning Scheme with Trusted Execution Environment", Information Sciences, Vol. 522, 2020, pp. 69-79.

[9] A. Elgabli, J. Park, C. B. Issaid, M. Bennis, "Harnessing Wireless Channels for Scalable and Privacy-Preserving Federated Learning", IEEE Transactions on Communications, Vol. 69, No. 8, 2021, pp. 5194-5208.

[10] H. Yang, J. Zhao, Z. Xiong, K. Y. Lam, S. Sun, L. Xiao, "Privacy-Preserving Federated Learning for UAV-Enabled Networks: Learning-Based Joint Scheduling and Resource Management", IEEE Journal on Selected Areas in Communications, Vol. 39, No. 10, 2021, pp. 3144-3159.

[11] L. Yunlong, H. Xiaohong, D. Yueyue, M. Sabita, Z. Yan, "Blockchain and Federated Learning for Privacy-preserved Data Sharing in Industrial IoT", IEEE Transactions on Industrial Informatics, Vol. 16, No. 6, 2019, pp. 4177-4186.

[12] L. Xiaoxiao, G. Yufeng, D. Nicha, S. H. Lawrence, V. Pamela, D. S. James, "Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results", Medical Image Analysis, Vol. 65, 2020.

[13] L. Xiaofeng, L. Yuying, L. Pietro, H. Pan, "Privacy-Preserving Asynchronous Federated Learning Mechanism for Edge Network Computing", IEEE Access, Vol. 8, 2020, pp. 48970-48981.

[14] T. U. Islam, R. Ghasemi, N. Mohammed, "Privacy-Preserving Federated Learning Model for Healthcare Data", Proceedings of the IEEE 12th Annual Computing and Communication Workshop and Conference, Las Vegas, NV, USA, 26-29 January 2022.

[15] M. A. P. Chamikara, P. Bertok, I. Khalil, D. Liu, S. Camtepe, "Privacy-preserving distributed machine learning with federated learning", Computer Communications, Vol. 171, 2021, pp. 112-125.

[16] L. Zhang, J. Xu, P. Vijayakumar, P. K. Sharma, U. Ghosh, "Homomorphic Encryption-based Privacy-preserving Federated Learning in IoT-enabled Healthcare System", IEEE Transactions on Network Science and Engineering, Vol. 5, No. 2, 2022, pp. 2864-2880.

[17] C. Jiang, C. Xu, Y. Zhang, "PFLM: Privacy-preserving federated learning with membership proof", Information Sciences, Vol. 576, 2021, pp. 288-311.

[18] Q. Yuanhang, H. M. Shamim, N. Jiangtian, L. Xuandi, "Privacy-preserving blockchain-based federated learning for traffic flow prediction", Future Generation Computer Systems, Vol. 117, 2021, pp. 328-337.

[19] L. Yin, J. Feng, H. Xun, Z. Sun, X. Cheng, "A Privacy-Preserving Federated Learning for Multiparty Data Sharing in Social IoTs", IEEE Transactions on Network Science and Engineering, Vol. 8, No. 3, 2021, pp. 2706-2718.

[20] L. Yunlong, H. Xiaohong, D. Yueyue, M. Sabita, Z. Yan, "Federated Learning for Data Privacy Preservation in Vehicular Cyber-Physical Systems", IEEE Network, Vol. 34, No. 3, 2020, pp. 50-56.

[21] X. Ma, Y. Zhou, L. Wang, M. Miao, "Privacy-preserving Byzantine-robust federated learning", Computer Standards & Interfaces, Vol. 80, 2022.

[22] L. Xiaoyuan, L. Hongwei, X. Guowen, L. Rongxing, H. Miao, "Adaptive privacy-preserving federated learning", Peer-to-Peer Networking and Applications, Vol. 13, 2020, pp. 2356-2366.

[23] L. Shixiang, Z. Gao, Q. Xu, C. Jiang, A. Zhang, X. Wang, "Class-Imbalance Privacy-Preserving Federated Learning for Decentralized Fault Diagnosis With Biometric Authentication", IEEE Transactions on Industrial Informatics, Vol. 18, No. 12, 2022, pp. 9101-9111.

[24] K. Wei, J. Li, M. Ding, M. Chuan, H. Su, B. Zhang, H. V. Poor, "User-Level Privacy-Preserving Federated Learning: Analysis and Performance Optimization", IEEE Transactions on Mobile Computing, Vol. 21, No. 9, 2021, pp. 3388-3401.

[25] M. Ali, F. Naeem, M. Tariq, G. Kaddoum, "Federated Learning for Privacy Preservation in Smart Healthcare Systems: A Comprehensive Survey", IEEE

Journal of Biomedical and Health Informatics, Vol. 27, No. 2, 2022, pp. 1-10.

[26] Q. Kong, F. Yin, R. Lu, B. Li, X. Wang, S. Cui, P. Zhang, "Privacy-Preserving Aggregation for Federated Learning-Based Navigation in Vehicular Fog", IEEE Transactions on Industrial Informatics, Vol. 17, No. 12, 2021, pp. 8453-8463.

[27] G. Han, T. Zhang, Y. Zhang, G. Xu, J. Sun, J. Cao, "Verifiable and privacy-preserving federated learning without fully trusted centres", Journal of Ambient Intelligence and Humanized Computing, Vol. 13, 2021, pp. 1431-1441.

[28] C. Fang Y. Guo, Y. Hu, B. Ma, L. Feng, A. Yin, "Privacy-preserving and communication-efficient federated learning in the Internet of Things", Computers & Security, Vol. 103, 2021.

[29] L. Tian, S. A. Kumar, T. Ameet, S. Virginia, "Federated Learning: Challenges, Methods, and Future Directions", IEEE Signal Processing Magazine, Vol. 37, No. 3, 2020, pp. 50-60.

[30] K. Cheng, T. Fan, Y. Jin, Y. Liu, T. Chen, D. Papadopoulos, Q. Yang, "SecureBoost: A Lossless Federated Learning Framework", IEEE Intelligent Systems, Vol. 36, 2021, pp. 87-98.

[31] Z. Huafei, M. Goh, R. Siow, N. Wee-Keong, "Privacy-Preserving Weighted Federated Learning Within the Secret Sharing Framework", IEEE Access, Vol. 8, 2020, pp. 198275-198284.

[32] F. Wang, H. Zhu, R. Lu, Y. Zheng, H. Li, "A privacy-preserving and non-interactive federated learning scheme for regression training with gradient descent", Information Sciences, Vol. 552, 2021, pp. 183-200.

[33] L. Li, F. Yuxi, T. Mike, L. Kuo-Yi, "A review of applications in federated learning", Computers & Industrial Engineering, Vol. 149, 2020.

[34] A. Lakhan, M. A. Mohammed, J. Nedoma, A. Lakhan, R. Martinek, P. Tiwari, A. Vidyarthi, A. Alkhayyat, W. Wang, "Federated-Learning Based Privacy Preservation and Fraud-Enabled Blockchain IoMT System for Healthcare", IEEE Journal of Biomedical and Health Informatics, Vol. 27, No. 2, 2023, pp. 664-672.

[35] X. Jie, G. S. Benjamin, S. Chang, W. Peter, B. Jiang, W. Fei, "Federated Learning for Healthcare Informatics", Journal of Healthcare Informatics Research, Vol. 5, 2020, pp. 1-19.

[36] G. Liu, C. Wang, X. Ma, Y. Yang, "Keep Your Data Locally: Federated-Learning-Based Data Privacy Preservation in Edge Computing", IEEE Network, Vol. 35, No. 2, 2021, pp. 60-66.

[37] L. Zengpeng, S. Vishal, P. M. Saraju, "Preserving Data Privacy via Federated Learning: Challenges and Solutions", IEEE Consumer Electronics Magazine, Vol. 9, No. 3, 2020, pp. 8-16.

[38] Y. Dong, X. Chen, L. Shen, D. Wang, "EaSTFLy: Efficient and secure ternary federated learning", Computers & Security, Vol. 94, 2020.

[39] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. S. Quek, H. V. Poor, "Federated Learning with Differential Privacy: Algorithms and Performance Analysis", IEEE Transactions on Information Forensics and Security, Vol. 15, 2020, pp. 3454-3469.

[40] M. Aledhari, R. Razzak, R. M. Parizi, F. Saeed, "Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications", IEEE Access, Vol. 8, 2020, pp. 140699-140725.

[41] The MNIST Database. Retrieved from http://yann.lecun.com/exdb/mnist/ (accessed: 2024)

[42] Y. Chen, X. Sun, Y. Jin, "Communication-efficient federated deep learning With layerwise asynchronous model update and temporally weighted aggregation," IEEE Transactions on Neural Networks and Learning Systems, Vol. 31, No. 10, 2020, pp. 4229-4238.

[43] J. S. Ng et al. "Joint auction-coalition formation framework for communication-efficient federated learning in UAV-enabled Internet of Vehicles," IEEE Transactions on Intelligent Transportation Systems, Vol. 22, No. 4, pp. 2326-2344.

A Mighty Image Retrieval Descriptor Based on Machine Learning and Gaussian Derivative Filter

Original Scientific Paper

El Aroussi El Mehdi
Chouaib Doukkali University,
ELITES Laboratory, Departement of Computer Science and Mathematics Higher School of Technology
El Jadida, Morocco
Elaroussi.e@ucd.ac.ma

Barakat Latifa
Chouaib Doukkali University,
Management of Sustainable Agriculture Laboratory, Higher School of Technology
El Jadida, Morocco
barakatlati@gmail.com

Silkan Hassan
Chouaib Doukkali University,
LaROSERI Laboratory, Department of computer sciences, Faculty of Sciences,
El Jadida, Morocco
silkan_h@yahoo.fr

Abstract *– The development of new image descriptor has always been an important topic to improve the efficiency of content-based image classification and retrieval. Improvements and developments in machine learning and deep learning algorithms as well as artificial intelligence algorithms are widely used by researchers to obtain effective CBIR descriptors. In our article, we will present a robust image descriptor, extended by machine learning and deep learning algorithms. The descriptor is provided through a Gaussian derivative filter scaffold named GDF-HOG with an enhanced convolutional neural network (CNN) AlexNet, to reduce the dimensions we used the principal component analysis algorithm. The experimental results were carried out on Oliva and Torralba, Caltech-101, Wang and Coil100 datasets. Experiments show that the accuracy of the proposed method is 98.23% for Coil-100%, 95.92% for Corel-1000, value 87.17 and 94.6% for Oliva and Torralba. In comparison our results with other descriptor image classifiers show that they achieved accuracy increases of 0.12% on average and up to 3.23%. These experimental results affirm the advantage of the proposed descriptor over existing systems based in terms of average accuracy. the proposed descriptor improves the precision, and also reduces the complexity of the calculation.*

Keywords: *Federated Learning, Machine Learning, Deep Learning, Privacy, Collaborative Machine Learning*

Received: Received: August 12, 2023; Received in revised form: October 19, 2023; Accepted: October 20, 2023

1. INTRODUCTION

Nowadays, the amount of digital images in the form of personalized and corporate collections has increased enormously thanks to the widespread and easy use of the Internet and the enormous use of audiovisual data in digital format for communications. Hence, there is a increasing demand for powerful image indexing and retrieval in an automatic way. Nevertheless, with such use and availability of images, the solutions based on textual images (grace of keywords) become impassable and inappropriate for indexing and retrieving images. To overcome this problem content-based image retrieval (CBIR) has become a great research interest among research communities [1,2,3]. Content-based image indexing and search discriptors generate considerable image representations by considering the visual features of images, i.e. salient points ,texture and shape [4-14] , It brings similar images using distance as a semantic result. The audacious increase in image descriptors has been an active field of research and will help to increase the performance vast actions in computer vision. various systems such as cale-invariant feature transform (SIFT) [15], speeded up robust features (SURF) [16] ,co-occurrence matrix (GLCM) [45] , local binary patterns (LBP) [17] , GIST algorithms were used for CBIR systems [18]. Such hand-crafted feature generation algorithms are still used in Machine Learning [19–22]. Each of these

descriptors has abnormalities, such as a large capacity of the feature vector , is that it cannot describe the characteristics of textures efficiently and distinctively and is mathematically weak and sensitive to noise. In this article we propose an efficient image descriptor to overcome these problems. This descriptor is created by combining Gaussian derivative filters named GDF-HOG more AlexNet CNN Enhanced. Also , for dimensional reduction . Principal Component Analysis (PCA) was applied. Detailed experimental analysis is performed using precision and recall on four datasets : Coil100, Corel1000, OT and FP datasets. The remainder of this article is put away as pursuant. Related works in the field of CBIR are presented in section 2. For section 3 we will present and discuss the proposed system. In section 4, the experimental results are reported, these results are compared to the experimental results existing systems in section 5. Finally, the conclusion and future perspectives are presented in section 6.

2. RELATED WORK

Simulation The deepening of automated CBIR systems has been an attractive field of study owing to its wide range of application in critical fields such as space imagery, bioinformatics, medical imaging, online surveillance and security. interior, etc.

There are many CBIR approaches described in the literature [23-28]. have been published until today to converge on the problem of image indexing and image retrieval descriptors in a more efficient and faster way. in general, the early work of CBIR applied to a single collection of features among the various features. Usually, it is difficult to obtain acceptable recovery results by applying only one characteristic. This is the reason why several scientists have employed a conjunction of systems to state new ones in order to speed up the performance and to consecrate it for astonishing cases [29]. Some of them are represented as follows:

In the work of [30] proposed a method in which uses Gaussian derivative filters named GDF-HOG a novel extension in which the local texture patterns are subjected to further treatment and then computed in Gaussian derivative filters way. In [31] showed a virtual solution for recovering semantically similar images from large image databases with respect to any solicitation image. to reduce discrepancy between low-level and high-level attributes. Genetic algorithms and support vector machines are used In [32], three image attribute s have been suggested for sovereign automatic overlay of images. To distill the color feature the color co-occurrence matrix (CCM) was used, while the difference between the scanning pattern pixels (DBPSP) is used for the texture features.

In the work of [33], a method based on the mixture of features extracted from two networks used for face discovery has been proposed. In works, the mixture of convolutional and system neural networks is used to realize a new image system. In [34], propose a system due to which the features of the image are compiled using SURF systems with HOG, the features of two descriptors are plenipotentiaries to convolutional spaces and feature vectors of measures 1×2016 and 1×1024 are created. The performances of these algorithms have exposed that they are efficient systems for the exploration of categories, the main problem found is the unequal dimension. In this paper, we propose a new algorithm to solve this unequal dimension problem.

3. PROPOSED METHOD

Our proposed method is graphically illustrated as Fig. 1. Overall, the proposed framework consists of four parts, in the first step the images are studied and scaled to $227 \times 227 \times 3$ using MATLAB's bicubic intercalation. Then all images are sent to an in-depth feature extractor (an enriched AlexNet CNN) and crafting systems such as HOG. First, the enriched CNN AlexNet hosted the images and explored its models, and finally suggested a feature vector of dimension 1×64 [33]. On the other hand, we use GDF-HOG an extension in which the local texture patterns are subjected to a complementary emolument then calculated in the manner of Gaussian derivative filters [30], thereafter, we use the PCA algorithm at the end to reduce the dimensions of the characteristics given by GDF -HOG Descriptor and also we use in order to match the dimension by GDF-HOG descriptor with deep feature vector . To finish, the deep aspect vector and the GDF-HOG-PCA descriptor essential vector are combined to have an efficient image system.

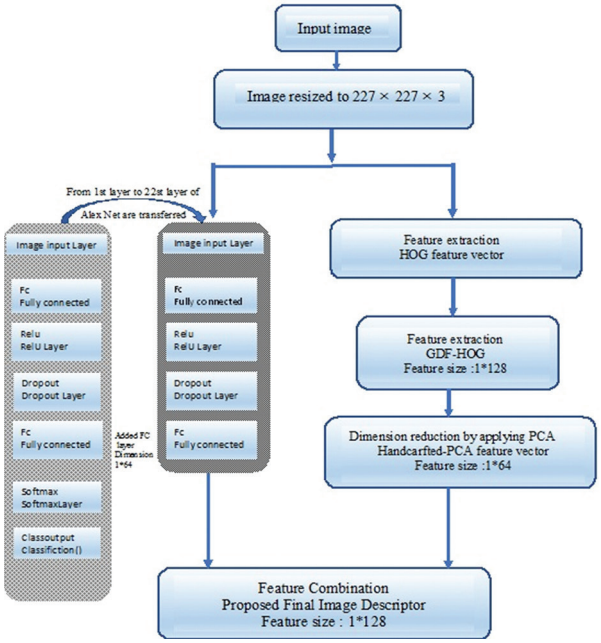


Fig. 1. Proposed Final Image Descriptor

The choice AlexNet CNN, HOG, GDF and PCA in this research due to some reasons. These reasons are described as follow:

The PCA algorithm is a means of vector downscaling that is commonly practiced to reduce the size of large data sets, by metamorphosing large sets of elements into sets with fewer elements without changing most of the information in the data set. data set. big set. we have exploited the PCA algorithm for different arguments, for example the reduction of the computation volume and the learning times, the simplification of the models, etc. [34].

The CNN Alexnet includes 25 layers, in order to create an Alexnet -improved, three last layers (23rd, 24th and 25th layers) of Alexnet CNN have been eliminated but other layers (22 layers - from the 1st to the 22nd) are transferred. Then, at the end of the Alexnet CNN diapers transferred, a fully connected layer (FC) from dimension1*64 was added [54].

The AlexNet CNN not only reduces the number of parameters and the proportion of full connection layer parameters, but also improves the automatic detection of interesting and super-scale features to the exclusion of human control [33].

The basic concepts of the Histogram of Oriented Gradients (HOG) are the regional characteristics and mode of the objects, which experience the marked by the assignment of local intensity gradients or edge administrations [35]. The orientations of the gradients are robust to all lighting variations, since the training histogram gives rotation invariance. the window-based HOG algorithm concerted locally to a stale point of interest. The advantage of this algorithm is that it ends up with local cells that are invariant to the geometric and photometric change, to the derogation of the orientation of the object [36,37].

The GDF algorithm yields amplification in which local texture patterns are kneed to additional processing after computed in the form of Gaussian derivative filters. He practiced the algorithm of Gaussian derivative filters to draw and catalog texture images, even if the dimensions of the image modify because the absolute state of the form does not modify. The first and second Gaussian derivative filters can be rotated at any angle by linear combination of two basis filters [38]. Gradient calculation which is calculated via Gaussian function and two-dimensional convolution gives more overwhelming texture and intensity factors than conventional gradient. Thus, Gaussian derivative filters are usually a suitable exemplar for extracting fundamental properties from texture pretexts.

4. EXPERIMENTAL RESULTS AND SIMULATION

In the experimental part the proposed system was implemented on the Anaconda software for an environment of Python, a computer system with 8 GB 1600 MHZ DDR RAM , Intl HD Graphics 5000 15366Mo graphics ,processor IntelCore i5,1.40 GHz central processor. accomplishment of exfiltration does not depend exclusively on a skilful description of the characteristics, but also on effective measures of similarity. In our experiments, we have used the measures of similarity mainly the overriding ones, including square chord distance for classification [39], extended Canberra distance [40], Euclidean distance has also been used [41,42] . In this paper, a detailed experimental analysis is performed using average mean precision (mAP) and recall criteria to quantify the proposed descriptor for archiving and CBIR [43] on four datasets: Coil100, Corel -1000, FP (Catech101) and OT. For our experiments, reducing image to size 227*227is grown by resizing the MATLAB load employing bicubic interpolation.

In this study, we estimate the achievements of the various basic steps and their absorption of functionalities, as appropriate in paragraph 4.2, on four most used datasets: COIL-100 [44], Corel-1000 (Wang) [48], FP (Catlech-101) [46] and OT [47]. The amounts of images on board these datasets are 7200, 10000, 380 and 2688 respectively. Each dataset contains color images that represent various features. These datasets are depicted in detail below:

COIL-100 is a database [44] of 100 uses of color images. Objects were placed on a motorized turntable against a dark background and images were taken at internal exposures of 5 degrees. This dataset was used in a real-time 100-use recognition system in which a sensor in the system could identify the object and display its angular pose. There are 7,200 images of 100 objects. Each object was rotated 360 degrees to vary the pose of the object against a stationary color camera. Images of the objects were taken at 5 degree exposure intervals. This corresponds to 72 exposures per object. These images were then normalized in size. Objects have a wide variety of complex geometric characteristics and reflectance.

Corel-1000 (Wang) is an image dataset containing 1000 of the Corel photo gallery [48] with ground truth. the images are collected in ten groups just like (Africa, beach, monuments, buses, dinosaurs, elephants, flowers, horses, mountains and food), there are 100 images of size 256x384 or 384x256 for each group. The images of the same group are admired as similar images. The images are subdivided into ten groups so that it is almost certain that a user will want to find the other images in a group if the query comes from one of these ten groups.

The FP(Caltech-101) [46] dataset is a widely used dataset for object identification missions, it includes almost 9,000 images of 101 classes of objects (e.g., "helicopter", " elephant" and "chair", etc.) and a background class that dominate images that are not part of the 101 object classes. For each category of objects, there are around 40 to 800 images, while most classes have around 50 images. images are 300×200 pixel dimensions.. The categories were chosen to reflect a variety of real-world objects, and the images themselves were carefully selected and annotated to provide a challenging benchmark for object recognition algorithms.

The Oliva & Torralba (OT) dataset globally includes 2,688 color images [47]. The dataset has eight classes, namely coast, forest, mountain, countryside, highway, inner city, high-rise building and street. These images are of JPG types with a dimension of 265 × 265 pixels.

4.1. EFFECT OF DISTANCE MEASUREMENTS ON THE SIMULATION AND EVALUATION OF THE PROPOSED SYSTEM FOR IMAGE RETRIEVAL

In this paragraph, we demonstrate that the performance of the proposed descriptor on the four datasets with six different measures of similarity has been evaluated and compared to the best existing similar methods. for image classification [43]. In these experiments, we randomly selected 10 images of any class as search images. Mean values of precision and recall are shown for N = 10. The value N = 10 is taken because later in Table 5 we will compare our results with other methods.

4.1.1. Proposed system performance on the COIL-100 dataset for the CBIR

The performance of the proposed descriptor on the COIL-100 dataset and for the six distance offerings for CBIR has been proven in Table 1. It is observed from Table 1, that the best mAP and the best average recall for recovery. Ten vertices are collected for the square chord distance which is 98.75%, followed by the extended Canberra distance measurement which gives a value of 98.26% for accuracy. The accuracy value achieved using Euclidean distance is 97.32%. The distance with a value of 92.53%. For all relative images, the best mAP and best average recall are achieved for the square chord distance which is 92.15%, followed by the extended Canberra distance measurement which gives a value of 91.71%, for the Euclidean distance measure which gives a value of 90.84. %.

Table1. Performance of proposed approach on Coil100 dataset on Mean Average Precision

Distance metrics	Proposed method	
	10-top (mAP)	All relative (mAP)
Square Chord	98.75	93.34
Euclidean Distance	97.32	91.84
Extended Canberra	98.26	92.71
L1	93.53	88.64
L2	92.53	87.39
X^2	96.8	90.18

4.1.2. Proposed system performance on the Wang dataset for the CBIR

The performance of the proposed descriptor on the Wang dataset and for the different distance measures are represented in Table 2. According to this table, the average mAP and recall using the square chord distance for the search of the top ten are 96.39% and for all similar images they are 92.15% for extended Canberra distance the mAP and mean Recall which is 96.16% for 10-top retrieval and 91.93% for all relative images. we observe for the euclidean distance that the mAP and mean Recall which is 95.9% for 10-top retrieval and 91.69% for all relative images. For the other distances the average mAP and Recall for the recovery of the top 10 relative images for the distance of X^2, L1 and L2 is 91.8%, 91.5% and 90.1% respectively, and for the recovery of all the relative images is 87% for X^2 , 85.5% for L2 and 86% for L1.

Table 2. Performance of proposed approach on Wang dataset on Mean Average Precision

Distance metrics	Proposed method	
	10-top (mAP)	All relative (mAP)
Square Chord	96.39	92.15
Euclidean Distance	95.9	91.69
Extended Canberra	96.16	91.93
L1	91.53	86.21
L2	90.1	85.48
X^2	91.80	87.14

4.1.3. Proposed system performance on the Caltech-101 dataset for the CBIR

For OT(Caltech-101) dataset the visual results of proposed system for the various similarity measures are depicted in Table 3. According to Table 3, it can be concluded for top ten image retrieval that, the best mean average precision (mAP) is provides for the square-chord distance which is 95.18%, followed by the euclidean distance measure which yields a value of 94.23% for precision. The precision value provides by using extended Canberra is 94.29%. followed by the L1 distance for which this values is 89.05% which is superior than the performance provides by L2 distance measure with a value the 88.53% ,which is slightly lower than the performance provides by X^2 which is 89.36%. When we look for the overall results, the square-chord distance measure provides best results which is 89.15%. The second best result for the euclidean distance measure which is 86.24%, The mean average precision (mAP) values for the extended Canberra distance , L1, L2 and for X^2 distance are 88.56% , 84.46% , 82.7% and 84.52%, respectively.

Table 3. Performance of proposed approach on FP (Caltech101) dataset on Mean Average Precision

Distance metrics	Proposed method	
	10-top (mAP)	All relative (mAP)
Square Chord	95.18	89.15
Euclidean Distance	94.23	86.24
Extended Canberra	94.29	88.56
L1	89.05	84.46
L2	88.53	82.7
X^2	89.36	84.52

4.1.4. Proposed system performance on the Oliva and Torralba (OT) dataset for the CBIR

The average mean precision (mAP) using the different methods on the Oliva and Torralba (OT) datasets are marked in Table 4. respectively. The qualita-

tive propensity of these performances is similarly the same as that found for the COIL-100, Corel-1000 and FP (Caltech-101) datasets.

The proposed system produces an excellent collection balance for the square-chord distance for 10-top retrieval which is 90.3% and for all relative images are 85.92%, followed by the extended Canberra distance measure which yields a value of 88.37% for 10-top retrieval and 84.02% for all relative images , in third place the euclidean distance measure which yields a value of 87.56% for 10-top retrieval and which yields 87.56% for all relative images. The L1 distance measure and L2 distance measure provides mean average precision results which is 82.68% and 81.19% for 10-top retrieval which is slightly lower than the results acquired by X^2 distance measure which is 86.14% for 10-top retrieva. For all relative images the L1 distance measure, L2 distance measure and X^2 distance measure provides mean average precision results which is 79.48%, 78.47% and 80.26% respectively.

Table 4. Performance of proposed approach on Oliva and Torralba (OT) dataset on Mean Average Precision

Distance metrics	Proposed method	
	10-top (mAP)	All relative (mAP)
Square Chord	90.3	85.92
Euclidean Distance	87.56	83.62
Extended Canberra	88.37	84.02
L1	82.68	79.48
L2	81.19	78.47
X^2	86.14	80.26

4.2. THE PROPOSED APPROACH'S MAP-MEAN RECALL CURVES FOR SQUARE-CHORD DISTANCE

We will illustrate the proposed descriptor map averaged recall figures for square chord distance and for ten vertex retrieval across the four databases for CBIR Coil-100, Corel-1000, FP, and OT. In figure 2, the power of the proposed descriptor is noted on the Coil-100 dataset. In proportion to this figure, the average mAP and recall are 96.02%. The proposed descriptor power on the Corel-1000 dataset is expressed in Fig. 3. According to this figure, the average mAP and recall for the recovery of ten peaks is 93.91%. For the FP dataset, the power of the proposed descriptor is demonstrated in Fig. 4. Based on this figure, the mean mAP and recall are shown as 86.86%. In Fig. 5, the proposed descriptor power is shown on the OT dataset. According to this graph, the average mAP and recall is 96.02%. Considering the result of the four graphs. We can conclude that the proposed methodology perfectly recovers many relevant images with a very high rate on different image datasets. The proposed descriptor graph fruits for the CBIR on the four image datasets are shown in Figs. 6, 7, 8. For each dataset, an image is randomly named and content-based image trapping is performed. At the end, the ten images most similar to the requested image are collected and displayed.

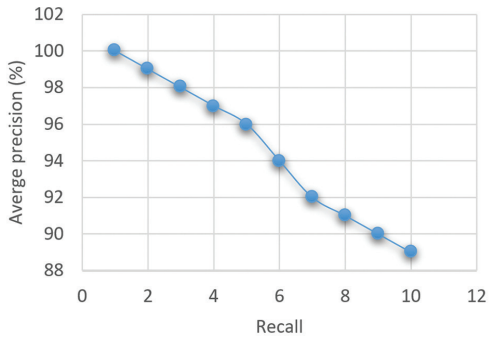


Fig. 2. Average Descriptor Accuracy Rate Curve on the Coil-100 Dataset

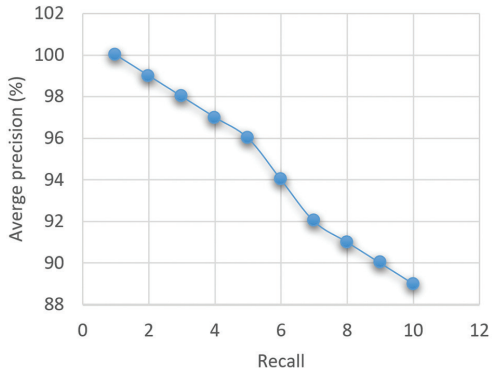


Fig. 3. Average Descriptor Accuracy Rate Curve on the Wang Dataset

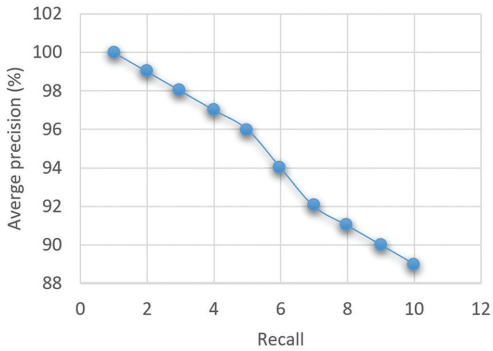


Fig. 4. Average Descriptor Accuracy Rate Curve on the Coil-100 Dataset method for FP (Caltech-101) dataset

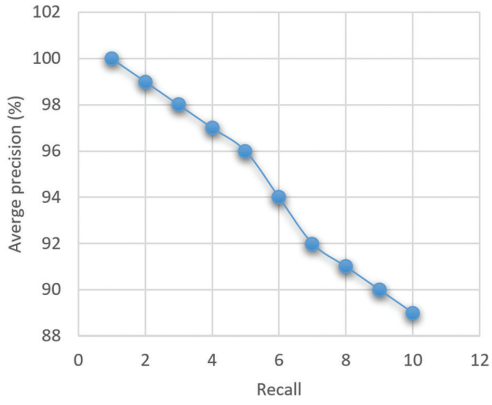


Fig. 5. Average Descriptor Accuracy Rate Curve on the Oliva and Torralba(OT) Dataset

According to Figs. 6, 7, 8, it can be said that in more cases, the images most similar to the query image were retrieved and placed in the first position, which is the goal of an effective CBIR system.

Fig. 6. Examples of the proposed descriptor visual results on the Wang dataset

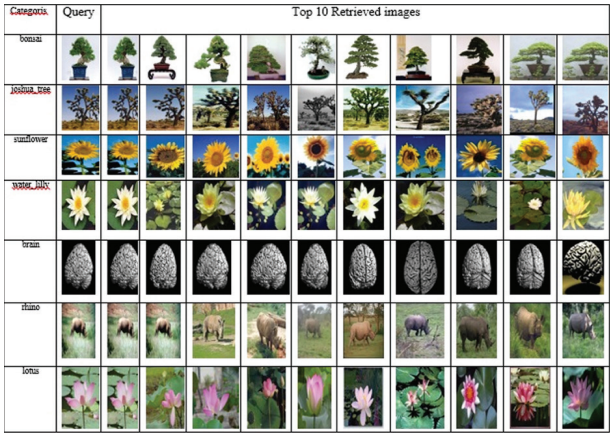


Fig. 7. Examples of the proposed descriptor visual results on the FP dataset

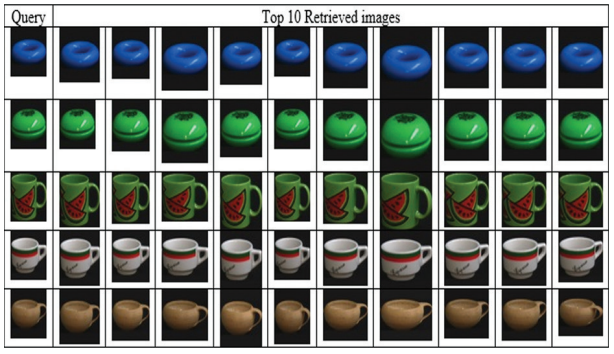


Fig. 8. Examples of the proposed descriptor visual results on the.Coil-100 dataset

4.3. COMPARISON RESULTS OF VARIOUS APPROACHES

In this part, the descriptor efficiencies proposed using the Coil100, Wang, OT and FP databases were estimated and evaluated with other descriptors for CBIR. In this regard, the technique is compared to [33, 49, 50, 51, 52, 53]. All the experiments below were carried out under the same conditions. the reason for our choice to compare with these discriptors is that these systems declare their balance sheets on the same databases and that they also use the Euclidean distance measure. The average mean precision values (mAP) are presented in Table 5. The performance of the proposed descriptor in terms of ranking compared to the other available descriptors is proven. From this table, it is concluded that the proposed descriptor has a higher accuracy compared to other existing descriptors.

For the Coil-100 dataset, our proposed method give higher recovery performance for which the average accuracy is 98.23%. The results obtained from the average precision for approach [53] and for approach [51] are respectively 81% and 95%.

For the Corel-1000 dataset, the results show that the proposed system is more efficient in terms of average precision than the other systems, with the average precision obtained being 95.92%. The mAP values obtained for the other approaches are 91.87% for the AlexNet CNN[49] approach, 80.61%, 95.80%, 66.5% and 73.27% respectively for the HOG + SURF approaches [50], [53], [51] and [03]. On the FP dataset (Caltech-101) also reflect a trend similar to that obtained for the Wang and Corel-1000 datasets. The average accuracies obtained by the different approaches in the FP dataset (Caltech-101) are respectively 87.17%, 81.80%, 86.86% and 76.39% for the proposed approach, [33], [49] and [52]. For the OT (Oliva and Torralba) dataset, again, the proposed method gives the best result with a value of 94.6%.

According to the results, one can easily monitor that the proposed method has the highest mAP-average rate. Therefore, it can be concluded that the proposed method is an operational method for image classification and retrieval.

Table 5. Comparison the proposed method with other standard retrieval systems in datasets for CBIR

	Coil-100	Corel-1000	Caltech-101	Oliva and Torralba
Proposed method	98.23	95.92	87.17	94.6
AlexNet CNN [49]	…….	91.87	81.80	92.30
HOG + SURF [50]	…….	80.61	…….	…….
AlexNet+ HOG [33]	…….	95.80	86.86	93.91
Co-occurance matrix[52]	…….	…….	76.39	78.83
H.Color + 2D.F.C.G [53]	81	66.5	……	……
2-D histogram + S.M+ GLCM [51]	95	73.27	……	……

5. CONCLUSION

In this study, we investigated the performance of a high-performance image overlay descriptor. The proposed descriptor was created using a combination of Gaussian derivative filters named GDF-HOG with

an improved AlexNet convolutional neural network (CNN), principal component analysis (PCA) algorithm was used for dimension reduction. In the present analysis, it is observed that the proposed descriptor gives analog image retrieval results to current descriptors similar to the proposed one. we even analyzed different distances from the similarity measurements, which gives very high results for our descriptor and we also monitor that the square chord distance measurement gives excellent results on the Coil100 we obtained an average score of 98.23%, on Wang we obtained an average score of 95.92%, on Caltech-101 we obtained an average score of 87.17% and on Oliva and Torralba (OT) we obtained an average score of 94 .6%. which exceeds between 0.12% and 3.3% other descriptors . The design and explanation of computer-aided diagnosis (CAD) systems has currently become a priority that researchers have focused on. In these systems, data descriptors play an essential function. For our future research, we will study the extension of the proposed descriptor on the CAD system, and we can also use other new powerful convolutional neural networks.

6. REFERENCES

[1] V. N. Gudivada, V. V. Raghavan, "Content based image retrieval systems", Computer, Vol. 28, 1995, pp. 18-22.

[2] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-based image retrieval at the end of the early years", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, 2000, pp. 1349-1380.

[3] R. Datta, J. Li, J. Z. Wang, "Content-based image retrieval: approaches and trends of the new age", Proceedings of the 7th ACM SIGMM International Workshop on Multimedia Information Retrieval, New York, NY, USA, 1-2 August 2005, pp. 253-262.

[4] Z. Lei, L. Fuzong, Z. Bo, "A CBIR method based on color-spatial feature", Proceedings of the IEEE Region 10 Conference, Cheju Island, South Korea, 15-17 September 1999, pp. 166-169.

[5] J. R. Smith, S. F. Chang, "Tools and Techniques for Color Image Retrieval", SPIE Conference Proceedings, Vol. 2670, 1996, pp. 2-7.

[6] K. N. Plataniotis, A. N. Venetsanopoulos, "Color Image Processing and Applications", Springer, 2000.

[7] N. Chitaliya, A. Trivedi, "Comparative analysis using fast discrete Curvelet transform via wrapping and discrete Contourlet transform for feature extraction and recognition", Proceedings of the International Conference on Intelligent Systems and Signal Processing, Gujarat, India, 1-2 March 2013, pp. 154-159.

[8] A. Barley, C. Town, "Combinations of Feature Descriptors for Texture Image Classification", Journal of Data Analysis and Information Processing, Vol. 2, 2014, pp. 67-76.

[9] I. J. Sumana, M. M. Islam, D. Zhang, G. Lu, "Content based image retrieval using curvelet transform", Proceedings of the IEEE 10th Workshop on Multimedia Signal Processing, Cairns, Australia, 8-10 October 2008, pp. 11-16.

[10] J. Zhang, T. Tan, "Brief review of invariant texture analysis methods", Pattern Recognition, Vol. 35, 2002, pp. 735-747.

[11] M. Yang, K. Kpalma, J. Ronsin, "A survey of shape feature extraction techniques", Pattern Recognition Techniques, Technology and Applications, InTech Open, 2008, pp. 43-90.

[12] D. Zhang, G. Lu, "Shape-based image retrieval using generic Fourier descriptor", Signal Processing: Image Communication, Vol. 17, 2002, pp. 825-848.

[13] E. Vimina, K. P. Jacob, "Content Based Image Retrieval Using Low Level Features of Automatically Extracted Regions of Interest", Journal of Image and Graphics, Vol. 1, 2013, pp. 7-11.

[14] K. Velmurugan, L. D. S. S. Baboo, "Content-based image retrieval using SURF and colour moments", Global Journal of Computer Science and Technology, Vol. 11, 2011, pp. 1-4.

[15] J. Liu, S. Zhang, W. Liu, C. Deng, Y. Zheng, D. N. Metaxas, "Scalable mammogram retrieval using composite anchor graph hashing with iterative quantization", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 27, No. 11, 2016, pp. 2450-2460.

[16] Y.-H. Lee, Y. Kim, "Efficient image retrieval using advanced SURF and DCD on mobile platform", Multimedia Tools and Applications, Vol. 74, 2015, pp. 2289-2299.

[17] P. Mohanaiah, P. Sathyanarayana, L. Gurukumar, "Image texture feature extraction using GLCM approach", International Journal of Scientific and Research Publications, Vol. 3, No. 5, 2013, pp. 1-5.

[18] Z. Camlica, H. R. Tizhoosh, F. Khalvati, "Medical image classification via SVM using LBP features from saliency-based folded data", Proceedings of the IEEE 14th International Conference on Machine Learning and Applications, Miami, FL, USA, 9-11 December 2015. pp. 128-132.

[19] A. S. Vijendran, S. V. Kumar, "A new content based image retrieval system by HOG of wavelet sub bands", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 8, No. 4, 2015, pp. 297-306.

[20] P. Dhar, "A new flower classification system using LBP and SURF features", International Journal of Image, Graphics and Signal Processing, Vol. 11, No. 5, 2019, p. 13.

[21] C. Gonzalez-Arias, C. C. Viáfara, J. J. Coronado, F. Martinez, "Automatic classification of severe and mild wear in worn surface images using histograms of oriented gradients as descriptor", Wear, Vol. 426-427, Part B, 2019, pp. 1702-1711.

[22] A. Shinde, A. Rahulkar, C. Patil, "Content based medical image retrieval based on new efficient local neighborhood wavelet feature descriptor", Biomedical Engineering Letters, Vol. 9, No. 3, 2019, pp. 387-394.

[23] S. Fekri-Ershad, "Developing a Gender Classification Approach in Human Face Images Using Modified Local Binary Patterns and Tani-moto Based Nearest Neighbor Algorithm", arXiv:2001.10966, 2020.

[24] R. M. Kumar, K. Sreekumar, "A survey on image feature descriptors", International Journal of Computer Science and Information Technologies, Vol. 5, 2014, pp. 7668-7673.

[25] A. S. Nair, R. Jacob, "A Survey on Feature Descriptors for Texture Image Classification", International Research Journal of Engineering and Technology, Vol. 4, No. 2, 2017.

[26] Y. Liu, D. Zhang, G. Lu, W.-Y. Ma, "A survey of content-based image retrieval with high-level semantics", Pattern Recognition, Vol. 40, No. 1, 2007, pp. 262-282.

[27] M. S. Lew, N. Sebe, C. Djeraba, R. Jain, "Content-based multimedia information retrieval: State of the art and challenges", ACM Transactions on Multimedia Computing, Communications, and Applications, Vol. 2, No. 1, 2006, pp. 1-19.

[28] El M. El Aroussi, S. Hassan, "Image Retrieval System Based on Color and Texture Features", Proceedings of ESAI: Embedded Systems and Artificial Intelligence, Fez, Morocco, 2-3 May 2019, pp. 475-487.

[29] S. Antani, R. Kasturi, R. Jain, "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video", Pattern Recognition, Vol. 35, No. 4, 2002, pp. 945-965.

[30] A. Shakarami, H. Tarrah, A. Mahdavi-Hormat, "A CAD system for diagnosing Alzheimer's disease using 2D slices and an improved AlexNet-SVM method", Optik, Vol. 212, 2020, p. 164237.

[31] K. Hanbay, N. Alpaslan, M. F. Talu, D. Hanbay, A. Karci, A. F. Kocamaz, "Continuous rotation invariant features for gradient-based texture classification", Computer Vision and Image Understanding, Vol. 132, 2015, pp. 87-101.

[32] D. Nister, H. Stewenius, "Scalable recognition with a vocabulary tree", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, NY, USA, 17-22 June 2006, pp. 2161-2168.

[33] H. Jegou, M. Douze, C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search", Proceedings of the 10th European Conference on Computer Vision, Marseille, France, 12-18 October 2008, pp. 304-317.

[34] A. Shakarami, H. Tarrah, "An efficient image descriptor for image classification and CBIR", Optik, Vol. 214, 2020, p. 164833.

[35] S. Arefnezhad, S. Samiee, A. Eichberger, A. Nahvi, "Driver drowsiness detection based on steering wheel data applying adaptive neuro-fuzzy feature selection", Sensors, Vol. 19, No. 4, 2019, p. 943.

[36] Y. Liu, Y. Ge, F. Wang, Q. Liu, Y. Lei, D. Zhang, G. Lu, "A rotation invariant HOG descriptor for Tire pattern image classification", Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Brighton, UK, 12-17 May 2019, pp. 2412-2416.

[37] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection", Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20-25 June 2005, pp. 886-893.

[38] W. T. Freeman, E. H. Adelson, "The design and use of steerable filters", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 13, No. 9, 1991, pp. 891-906.

[39] V. S. Katti, S. Sushitha, S. Dhareshwar, K. Sowmya, "Implementation of Dalal and Triggs Algorithm to Detect and Track Human and Non-Human Classifications by Using Histogram-Oriented Gradient Approach", Proceedings of the Third International Conference on ICT for Sustainable Development, Goa, India, 30-31 August 2018, pp. 759-770.

[40] R. Arandjelovic, A. Zisserman, "Three things everyone should know to improve object retrieval", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16-21 June 2012, pp. 2911-2918.

[41] Y. Rubner, J. Puzicha, C. Tomasi, J. M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture", Computer Vision and Image Understanding, Vol. 84, No. 1, 2001, pp. 25-43.

[42] S. Antani, R. Kasturi, R. Jian, "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video", Pattern Recognition, Vol. 35, No. 4, 2002, pp. 945-965.

[43] M. Kokare, B. N. Chatterji, P. K. Biswas, "Comparison of similarity metrics for texture image retrieval", Proceedings of the IEEE Conference on Convergent Technologies for the Asia-Pacific Region, Bangalore, India, 15-17 October 2003, pp. 571-575.

[44] J. Han, K.-K. Ma, "Rotation-invariant and scale-invariant Gabor features for texture image retrieval", Image and Vision Computing, Vol. 25, No. 9, 2007, pp. 1474-1481.

[45] S. A. Nene, S. K. Nayar, H. Murase, "Columbia Object Image Library (COIL-100)", Technical Report CUCS-006-96, February 1996.

[46] J. Z. Wang, J. Li, G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 9, 2001, pp. 947-963.

[47] G. Griffin, A. Holub, P. Perona, "Caltech-101 Object Category Dataset", CaltechDATA, Technical Report 7694, California Institute of Technology, Pasadena, CA, USA, 2007.

[48] A. Oliva, A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope", International Journal of Computer Vision, Vol. 42, No. 3, 2001, pp. 145-175.

[49] A. Shah, R. Naseem, S. Iqbal, M. A. Shah, "Improving CBIR accuracy using convolutional neural network for feature extraction", Proceedings of the 13th International Conference on In Emerging Technologies, Islamabad, Pakistan, 27-28 December 2017, pp. 1-5.

[50] Z. Mehmood, F. Abbas, T. Mahmood, M. A. Javid, A. Rehmen, T. Nawaz, "Content-based image retrieval based on visual words fusion versus features fusion of local and global features", Arabian Journal for Science and Engineering, Vol. 12, 2018, pp. 7265-7284.

[51] El M. El Aroussi, El H. Noureddine, S. Hassan, "Content-based image retrieval approach using color and texture applied to two databases (Coil-100 and Wang)", Proceedings of the Ninth International Conference on Soft Computing and Pattern Recognition, Marrakech, Morocco, 11-13 December 2018, pp. 49-59.

[52] J. F. Serrano-Talamantes, C. Aviles-Cruz, J. Villegas-Cortez, J. H. Sossa-Azuela, "Self organizing natural scene image retrieval", Expert Systems with Applications, Vol. 40, No. 7, 2013, pp. 2398-2409.

[53] El M. El Aroussi, El H. Noureddine, S. Hassan, R. Mohammed, "New Index and Search Descriptor Combined Image of Text and Color Applied to Two Databases (Coil-100 and Corel-DB)", International Journal of Applied Mathematics & Statistics, Vol. 57, No. 1, 2018, pp. 113-127.

Performance Measurement of Small Cell Power Management Mechanism in 5G Cellular Networks using Firefly Algorithm

Original Scientific Paper

J. Premalatha
Department of Electronics and Communication Engineering,
Sathyabama Institute of Science and Technology,
Chennai 600 119.
premalathajeyaraman@gmail.com

A. Sahaya Anselin Nisha
Department of Electronics and Communication Engineering,
Sathyabama Institute of Science and Technology,
Chennai 600 119.
anselinnisha.ece@sathyabama.ac.in

Sanjaikanth E Vadakkethil Somanathan Pillai
School of Electrical Engineering and Computer Science,
University of North Dakota, Grand Forks, North Dakota, USA.
s.evadakkethil@und.edu

A. Bhuvanesh
Department of Electrical and Electronics Engineering,
PSN College of Engineering and Technology,
Tirunelveli 627 152.
bhuvanesh.ananthan@gmail.com

Abstract *– In cellular networks, with the increase in demand, designing a base station (BS) with less energy consumption remains a challenge for researchers. Also, in a heterogeneous network that is dense in nature, the distribution of numerous small BS has become a challenging issue in terms of expanding the cost of energy. In this paper, we investigate an optimized nature-based cluster sleep technique for reducing the power consumption in the BS as well as the interference in the network. The small BS are grouped along with the interference, which is assumed to be the cluster, which is quite large, where the fire fly (FF) algorithm is applied to frame the sleep technique for the small BS. These FF algorithms, which are based on fire fly attractiveness behavior, improve connectivity among the base stations in an energy-efficient way. The outcomes reveal that the projected sleep technique with the FF algorithm reduces the power consumed by the BS and also gives satisfactory performance for mobile users. The results were compared with the other techniques, such as BS conventional sleep mode and BS sleep mode with LEACH. The proposed method outperformed the other techniques.*

Keywords: *Firefly algorithm, Base station, sleeps technique, power consumption, and heterogeneous network*

Received: November 10, 2023; Received in revised form: February 21, 2024; Accepted: February 21, 2024

1. INTRODUCTION

In cellular networks, reducing energy consumption is a challenging topic of interest and is beneficial for both telecommunication operators and the global environment [1]. Also, in recent years, there has been a tremendous increase in the usage of mobile data, which is predominantly determined by smart phones, which offer user-friendly internet access and a variety of multimedia applications. On the whole, information and communication technology (ICT) is accountable for about 2% of CO2 emissions globally, and it will reach 4% in 2021 [1]. The conventional BSs have not been able to offer quality of service (QOS) to mobile users. According to the 2012 census, there were nearly 5.8 million conventional BSs worldwide, and it was expected to be more than 10 million in 2020 [2]. As of now, the global number of small BS (SBS) has now exceeded the conventional numbers. Thus, the increase in energy demand over the past few years has given way to green communication in cellular networks. And it is a well-known fact that the cellular network BS is the one that consumes two-thirds of the energy consumed by the whole radio access network. Consequently, reducing the energy utilized by the BS has become the main topic of research.

Energy-efficient BS can be achieved from many perspectives, like using energy-efficient power amplifiers, making use of renewable resources, and also deploying relays and small BSs. Cell zooming can also be used to reduce the energy consumption of BS. In practice, cell zooming reduces the number of active BS when there is low traffic. At the point when few BSs are switched off, the remaining active BSs tend to zoom out for an uninterrupted quality of service (QOS). It is necessary

to control the transmission power of the cellular network, as 50–60% of total energy is consumed by the processing circuit and cooling system when the BS is in a working state [3]. According to the data set presented in [4], the data traffic during the day is much larger than the data traffic at night. And also, it slightly varies from normal work days to the end of the week. As discussed, earlier SBS can be maintained and deployed easily as compared to conventional ones, which also require low transmission power. Of the advantages stated above, these SBS form a heterogeneous network (HNET) along with the macro-BS (MBS). Basically, the main idea behind SBS is to reduce the heavy load encountered by MBS for a better QOS. But on the other side, due to the large number of SBS, the newly formed HNET experienced severe interference in terms of both cross-tier interference and co-tier interference [5].

The source of intrusion is the variance in power among the nodes due to the deployment of cells, which are not planned beforehand as they can be switched on and off at any time or moved anywhere. These interferences may greatly reduce the HNET's functioning. Further, more severe interference leads to radio link failure in mobile equipment (ME), and due to unreliable control channels, the user might not continue to use the existing service or be unable to request a new service. To avoid all these problems, inter-cell interference coordination can be used for its proper operation. FF is a bio-roused method that has been utilized for settling nonlinear optimization issues. It depends on perceptions from the social bug settlements, where every person (for example, a firefly sparkling through bioluminescence) seems to work for its own advantage, but then the gathering in general performs to be profoundly coordinated [6]. FF algorithm firefly's brightness relies upon the fitness work. The objective of the FF is to achieve effective self-coordination among BSs. Fitness esteem chooses the brightness of the BS; henceforth, the fireflies with lesser fitness esteem move towards more prominent fitness esteem. BSs are considered haphazardly conveyed fireflies [7]. In [8], the constraint of firefly measurement is repealed by utilizing the hybrid approach of particle swarm optimization (PSO) and firefly measurement, which incorporates the usefulness of PSO in firefly measurement and, additionally, works on the conduct of fireflies engaged in the search for a better solution.

In [9], the author sets up a cluster for BS using the FF algorithm that minimizes the cost function. The objective of the FF algorithm is to observe the particle position of those outcomes for the best assessment of guaranteed fitness function. At the point when groups are framed with the FF algorithm, all bunch hubs initiate the transmission of information to their individual group heads. Group heads gather information from hubs and move to the base station for less energy utilization [10]. The author in [11] has suggested a method for energy-efficient clustering where they start by producing the irregular population of n fireflies. Every particle computes its light force (fitness). Without fail, all fireflies are arranged by request, diminishing as indicated by their fitness and view as the best one. Later, with a pairwise correlation of the light power, the firefly with less light pushes toward a more splendid one. This development relies on the distance between two fireflies. During the process, the best-up-to this point arrangement is refreshed until terminal measures are fulfilled. Firefly measurement is by all accounts an ideal improvement apparatus peculiarity because of the impact of the allure work, which is exceptional to the FF conduct [12]. Firefly doesn't retain or recollect any set of experiences of better circumstances, and they may wind up missing their circumstances [13]. Energy consumption of the node is estimated on the premise of transmission. The energy examination additionally demonstrates that the energy consumed-through correlation among LEACH, direct transmission, and the fast firefly algorithm performs better-through-less energy [14].

The Python implementation of the framework is used to assess its performance using real-world network construction datasets from a 5G operator. Through thorough simulations, the benefits given by network slicing are studied in terms of the attained data rates for V2X, blocking likelihood, and handover ratio through various mixtures of traffic types. The findings showed that when network traffic load in a region of interest and end users' quality of service are taken into consideration, appropriate resource splitting is crucial for slicing across V2X and other varieties of services.

This paper deals with the sleep mode technique used to reduce BS power consumption by the FF algorithm used to formulate the sleep technique. Interference is reduced by the FF algorithm, and the proposed FF algorithm optimizes the power consumed by the base station. The remainder of the paper is structured as follows: In Section II, the proposed model and general problem are discussed. Section III discusses the FF algorithm to formulate the sleep technique, and interference reduction is discussed. Section IV demonstrates the performance and simulation results of the proposed algorithm. Section V deals with the conclusion of the paper.

2. 5G SMALL CELL NETWORK MODEL

Let us consider a downlink model from the above fig. 1 of a HNET. In this model, each tier is considered a cellular network, with macro and small BS having their own prescribed radius. The mobile equipment that connects the concerned BS is named macro mobile equipment (MME) and small mobile equipment (SME). In Fig. 1, it is assumed that the mobile equipment is distributed uniformly, and each mobile equipment is associated with SBS and MBS. The whole frequency band is shared by both MBS and SBS, which are in dissimilar clusters. The bandwidth of the system is given by Bs, and the frequency reuse factor is one for the system.

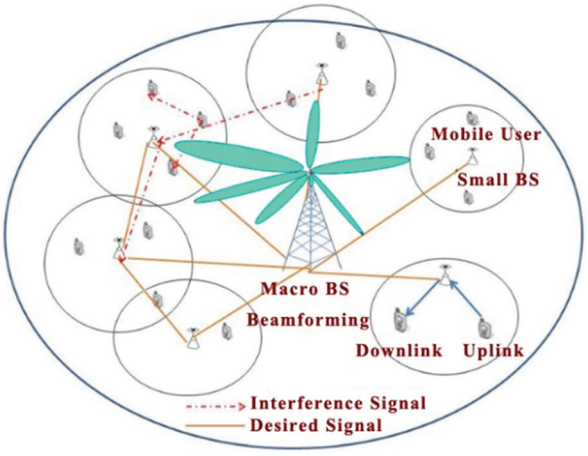


Fig. 1. Small cell and macro cell network model in 5G systems

The transmission power of *SBS b*, and *MBS m* is given by P_b and P_m respectively. The path loss between mobile equipment and BS for HNET outdoor network is given by the equation,

$$P_L = C + 10\ log_{10}(R^{\eta}) + S_r \quad (1)$$

where, P_L is path loss between mobile equipment and *BS*, *R* is distance amid *BS* and mobile equipment, η is the path loss component and S_r deals with random shadowing and it is zero mean random variable. After measuring the path loss channel gain is measured by

$C_g = 10^{-P_L/10}$(2). Where P_L is the value of path loss. Next step is to determine the signal to interference and noise ratio (*SINR*). In general, *SINR* is measured for typical mobile user w.r.t *BS y* is given by

$$SINR\ (y) = \frac{Wy/Plf(|y|)}{Np+Tp-y/Plf(|y|)} \quad (2)$$

where, N_p is the noise power which is a constant, T_p is total received power from the whole network and is given by $T_p = \sum_{y\epsilon\varphi} W_y / P_{lf}(y)$ where φ is related to poisson point process where W_y is assumed as $\{W_y\}_y \epsilon \varphi$ and is given by a collection of random variables which are identically and independently distributed.

The path loss function P_{lf} is given by

$P_{lf}(|y|) = (C|y|)^{\alpha}$ with the constants $C>0$ and $\alpha>2$ (3)

In this proposed method *SINR* is given for *ME e* that connects to *SBS d* as

$$\text{SINR (y)} = \frac{P_{Signal}}{P_{Interference} + P_{Noise}}$$
$$SINR_{d,e} = \frac{P_{d,e} C_{g\ d,e}}{\sum_{m=1}^{N} P_{m,e}\ C_{g\ m,e} + \sum_{b=1,i\neq d}^{M} q_{b,e}\ P_{b,e}\ C_{g\ b,e} + \mathrm{Np}^2} \quad (4)$$

Here $P_{d,e}$ is transmission power of *SBS d* with *ME e* , $C_{gd,e}$ is the channel gain between *ME e* and *SBS d*, $P_{m,e}$ is transmission power of *MBS m* when is related with *Me e*. similarly $C_{g\ m,e}$ is the channel gain of *MBS m* and *ME e*. $P_{b,e}$ is the transmission power of *SBS b* when is related with *ME e*. similarly $C_{g\ m,e}$ is the channel gain amid *SBS b* and *ME e*. *N* denotes the number of *MBS* and *N* denotes the number of *SBS* (*N*=6). *Np* is the noise power of the network and $q_{b,e}$ =1 implies the connection between *ME e* and *SBS b*.

Similarly, the *SINR* for *MBS* m *w.r t ME t* is given by,

$$SINR_{m,t} = \frac{P_{m,t} C_{g\ m,t}}{\sum_{j=1,j\neq m}^{N} P_{j,t} C_{g\ j,t} + \sum_{b=1}^{N} q_{b,t} P_{b,t} C_{g\ b,t} + \mathrm{Np}^2} \quad (5)$$

3. POWER MANAGEMENT MODEL FOR 5G SYSTEMS

Basically, the power consumed by a base station depends on two types of power consumption. One is the dynamic power, and the other is the static power. Static power consumption at the base station is active even if there is no connection from users. On the other hand, a dynamic base station is a function of load or traffic [14]. The power consumption of *SBS* is given by

$$P_{b,e} = \begin{cases} \alpha P_{SBS} + P_{am} & SBS\ in\ active\ mode \\ P_{sm} & BS\ in\ sleeping\ mode \end{cases} \quad (6)$$

where, $P_{b,e}$ is the transmission power of *SBS* when it is related with *ME e*, α is a constant which is allied with usage of data traffic. P_{SBS} is the *SBS* transmission power. P_{am} denotes the power consumed by *SBS* in active mode which is static in nature. This P_{am} is independent of the transmission power. P_{am} is the power consumption of *SBS* in sleeping mode. The small cell base station components are shown in Fig. 2.

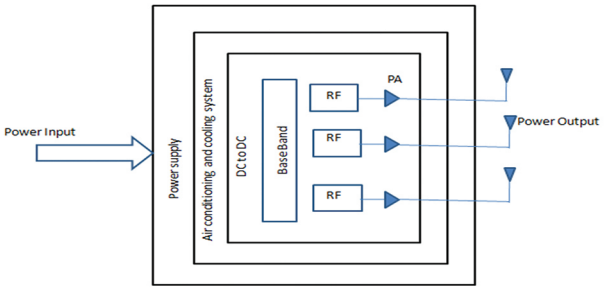


Fig. 2. Small cell base station components

4. FF ALGORITHM FOR SLEEP TECHNIQUE FOR OPTIMIZATION

The *FF* algorithm is a metaheuristic type of swarm intelligence technique where the behavior of *FF* is followed. *FF* is a non-linear algorithm that has multiple agents and is based on swarm intelligence algorithms. The *FF* [15] algorithm is one that is derived from nature, as it is enthused by the behavior of fireflies. Fireflies are insects or beetles that have wings that produce light and blink at it. This light does not have any ultraviolet or infrared frequencies; rather, it is produced chemically from the lower abdomen, which is called bioluminescence. The *FF* algorithm, which was first introduced by Yang [16], is based on bioluminescent communication and was assumed with the following formulations:

Fireflies will be attracted by every other firefly in spite of the sex since it is unisexual in nature.

Brightness and attractiveness are proportional to each other; a brighter firefly will attract a less bright firefly. However, when the distance between two fireflies is increased, their attractiveness decreases.

On the other hand, it will move around randomly if the level of brightness is the same.

Thus, when we relate the brightness of fireflies to their objective function, their attractiveness makes them competent to divide themselves into smaller groups, and each subgroup swarms around the neighborhood model.

Here, the brightness B_r of a firefly at a point is defined as

$$B_r(p) \alpha f_n(p) \tag{7}$$

Where p is dimensional point in dimensional space and $f_n(p)$ is the fitness function which is defined. $B_r(p)$ is directly proportional to the value of $f_n(p)$.

As discussed, earlier attractiveness A_r depends on the distance amid two fireflies and the brightness is indirectly proportional. The attractiveness decreases between the fireflies as the distance increases. Thus, attractiveness equation is defined by

$$A_r(p) = A_{ro} e^{-\gamma d^2} \tag{8}$$

where, A_{ro} is the attractiveness at d=0 and γ is constant value. The movement of firefly a toward more attractive firefly b is given by the equation

$$p_a^{i+1} = p_a^i + A_r e^{-\gamma d_{ab}^2}\left(p_b^i - p_a^i\right) \tag{9}$$

where, d_{ab} is the distance amid fireflies a and b, i is the iteration number.

The brightest firefly moves randomly and given by the equation

$$p_a^{i+1} = p_a^i + \alpha\varepsilon_i \tag{10}$$

were, $\alpha\varepsilon i$ randomization parameter.

Firefly measurement is productive and simple to execute. It is likewise reasonable for parallel execution. Nonetheless, investigations show that it is delayed in convergence and effectively gets caught in the neighborhood ideal for multimodal issues.

Likewise, the updates exclusively rely on current execution, and no memory of past best solutions or exhibitions is kept. That might prompt losing better solutions. Besides, since the boundaries are fixed, the search conduct stays very similar for any condition in all emphases. Subsequently, changing the standard firefly measurement to support its exhibition has been one of the examination issues. The network power management optimization flow in small cell 5G systems is shown in Fig. 3.

An *FF* algorithm is implemented in the proposed methodology, where power is consumed efficiently in cellular *BS*. Initially, the *BS* in the network is grouped, and every *BS* in the group shares information related to residual energy, its distance from other *BS* in the group, and the number of retransmissions. This information is used to choose the active *BS*. After every round, this information is modernized on every *BS*, and regrouping and macro-*BS* selection are carried out. In the firefly-based method, the value of residual energy plays an important role, as this value is shared between the other *BS* in the network. The distance between any two *BS* in the group is measured. Based on the values of residual energy and later, an active *BS* is found in the network, from macro *BS* to Femto *BS*. The *BS* with low power is enticed toward the high-power *BS*, and the attractiveness factor is measured. Any two Femto *BS* having the same power can be selected randomly. For 5G beam-forming heterogeneous networks, an outage probability analysis is proposed, which consists of a cellular multi-layer network. For the proposed beamforming heterogeneous network, the connotation possibility, the number of users in a layer, and the serving *BS*s probability density function are derived [17, 18].

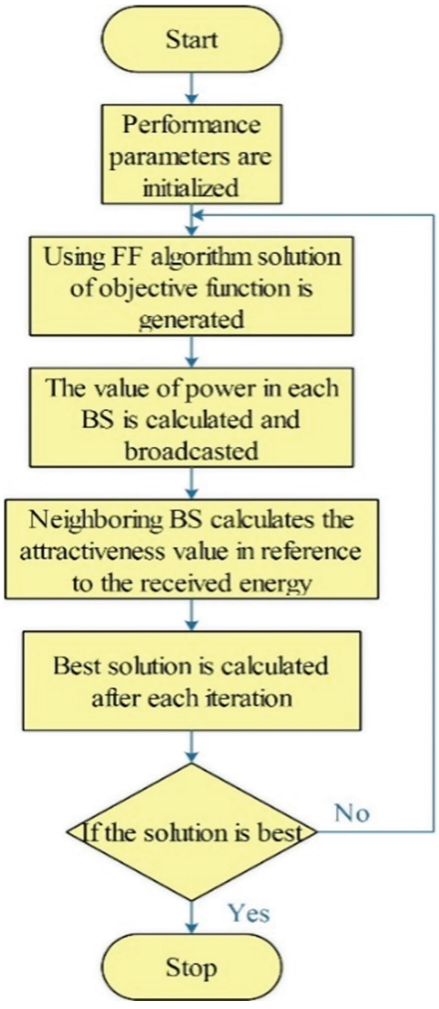


Fig. 3. Network power management optimization flow in small cell 5G systems

Likewise, the updates exclusively rely on current execution, and no memory of past best solutions or exhibitions is kept. That might prompt losing better solutions. Besides, since the boundaries are fixed, the search conduct stays very similar for any condition in all

emphases. Subsequently, changing the standard firefly measurement to support its exhibition has been one of the examination issues. The network power management optimization flow in small cell 5G systems is shown in Fig. 3.

An *FF* algorithm is implemented in the proposed methodology, where power is consumed efficiently in cellular *BS*. Initially, the *BS* in the network is grouped, and every *BS* in the group shares information related to residual energy, its distance from other *BS* in the group, and the number of retransmissions. This information is used to choose the active *BS*. After every round, this information is modernized on every *BS*, and regrouping and macro-*BS* selection are carried out. In the firefly-based method, the value of residual energy plays an important role, as this value is shared between the other *BS* in the network. The distance between any two *BS* in the group is measured. Based on the values of residual energy and later, an active *BS* is found in the network, from macro *BS* to Femto *BS*. The *BS* with low power is enticed toward the high-power *BS*, and the attractiveness factor is measured. Any two Femto *BS* having the same power can be selected randomly. For 5G beam-forming heterogeneous networks, an outage probability analysis is proposed, which consists of a cellular multi-layer network. For the proposed beamforming heterogeneous network, the connotation possibility, the number of users in a layer, and the serving *BS*s probability density function are derived [17, 18].

5. RESULTS AND DISCUSSION

Fig. 4 displays the distribution of macro- and small-cell base station nodes in the 5G network. Macro Base Stations (*BS*s) are used as a baseline and provide uniform coverage.Micro and pico/femto (often also referred to as small) cells are equipped with lower power *BS*s which are deployed in hotspots to increase capacity, or in dead spots unreachable by macro *BS*s in order to increase coverage.

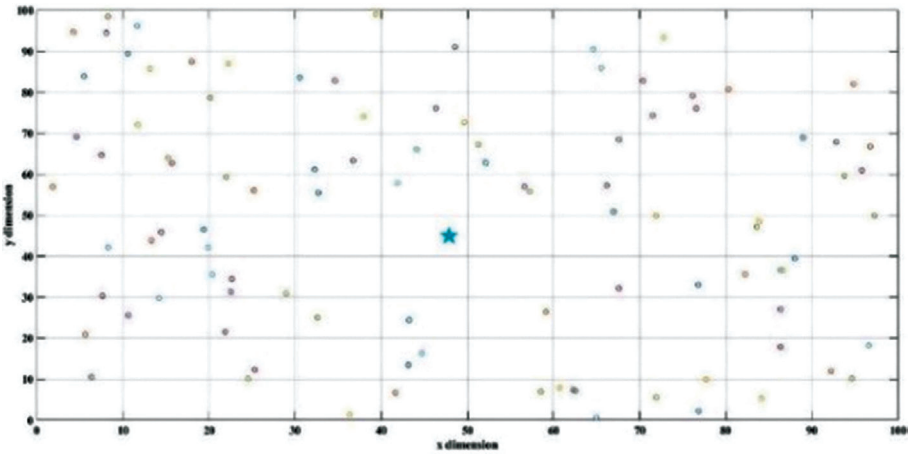


Fig. 4. Distribution of Macro and Small Cell Base Station Nodes in 5G Network

Fig. 5 illustrates the number of active base stations with 25% initial base station energy. The results imply that the proposed *FA* outperformed the existing methods. Fig. 6 shows the residual energy in a small cell 5G network with 25% initial base station energy. The total energy consumed by the base station during its operational time can be estimated by multiplying the energy consumption rate with operational time.

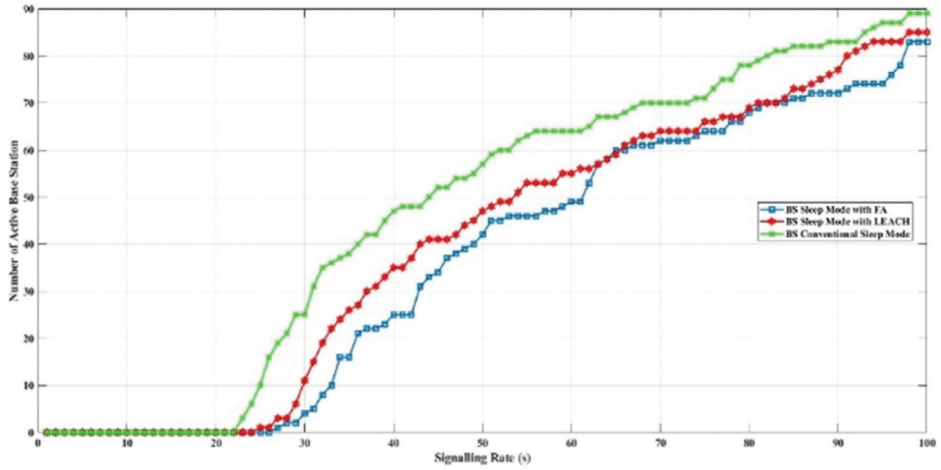


Fig. 5. Number of active base station with 25% initial base station energy

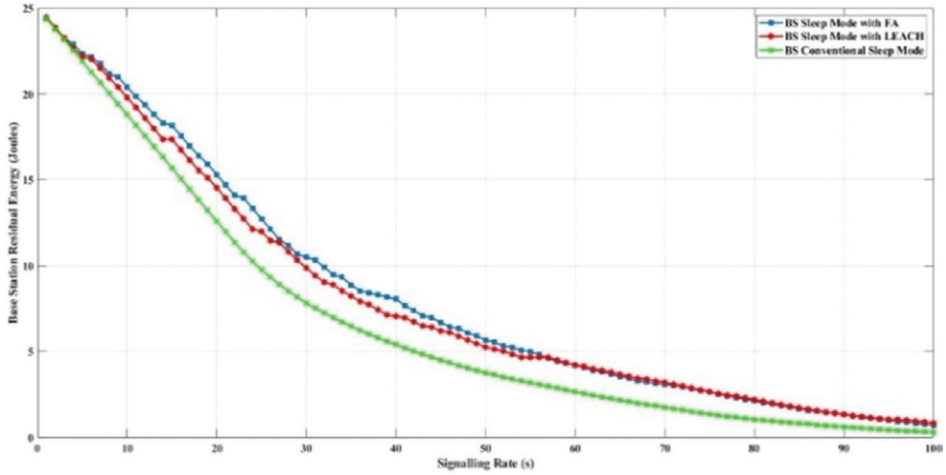


Fig. 6. Residual energy in small cell 5G network with 25% initial base station energy

The number of inactive 5G base stations with 25% initial energy is shown in Fig. 7. If a base station is considered inactive when its energy level is below 25% of the total capacity, then the number of inactive base stations can be calculated by multiplying total number of base stations with probability that a base station is inactive due to having less than 25% energy. Fig. 8 shows the throughput of a small-cell 5G network with 25% initial energy. It is denoted that the proposed *FA* has more number of throughput than other methods.

The distribution of small base stations is a dynamic process that considers the evolving needs of users, traffic patterns, and the characteristics of the deployment area. The goal is to create a flexible and adaptive network that efficiently meets the demands of 5G services.

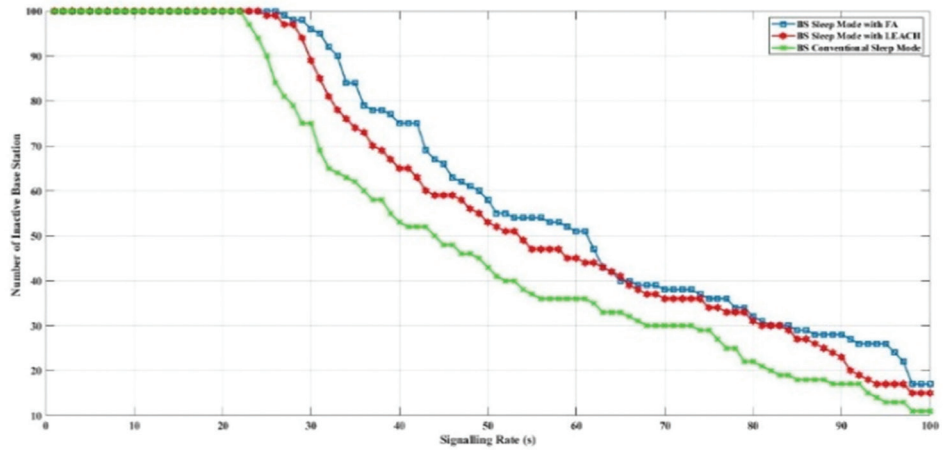


Fig. 7. Number of inactive 5G base station with 25% initial energy

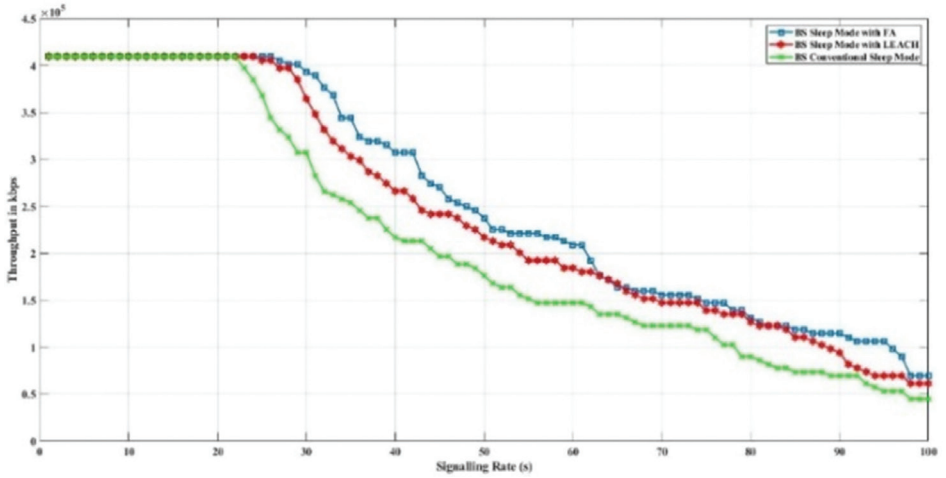


Fig. 8. Throughput of small cell 5G network with 25% initial energy

Fig. 9 displays the small-cell 5G network with 50% initial base station energy.

The residual energy in a small cell 5G network with 50% initial base station energy is displayed in Fig. 10. The number of inactive 5G base stations with 50% initial energy is shown in Fig. 11. Fig. 12 shows the throughput of a small-cell 5G network with 50% initial energy. The small-cell 5G network with 75% initial base station energy is illustrated in Fig. 13. Fig. 14 displays the residual energy in a small-cell 5G network with 75% initial base station energy. Fig. 15 shows the number of inactive 5G base stations with 75% initial energy. Fig. 16 shows the throughput of a small cell 5G network with 75% initial energy.

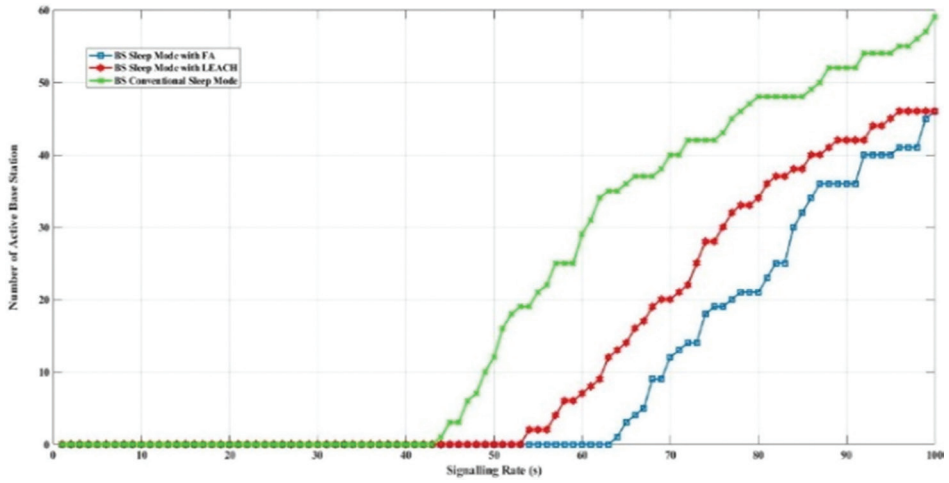


Fig. 9. Small cell 5G network with 50% initial base station energy

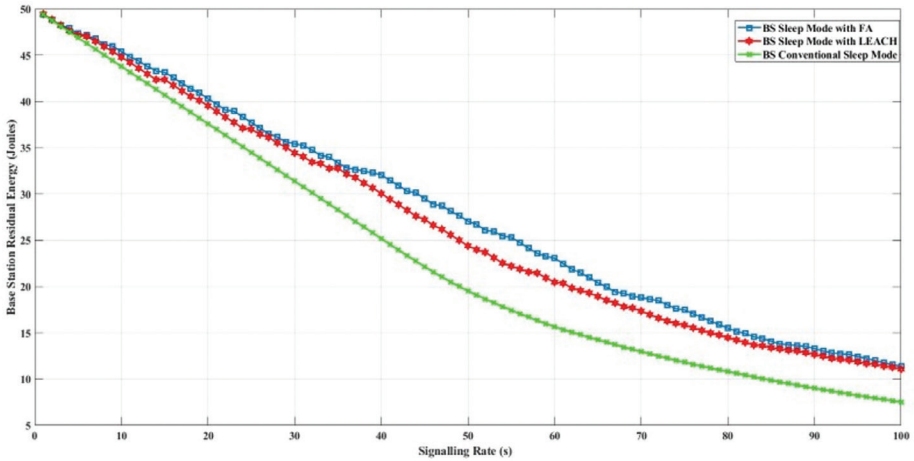


Fig. 10. Residual energy in small cell 5G network with 50% initial base station energy

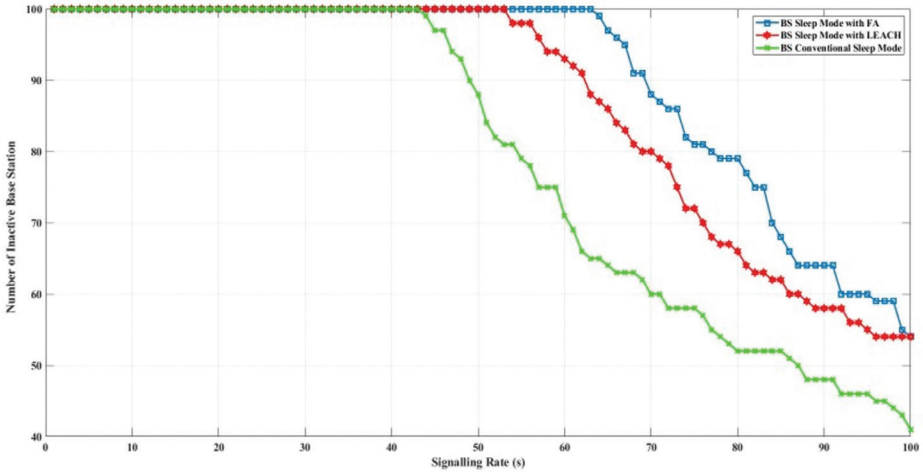


Fig. 11. Number of inactive 5G base station with 50% initial energy

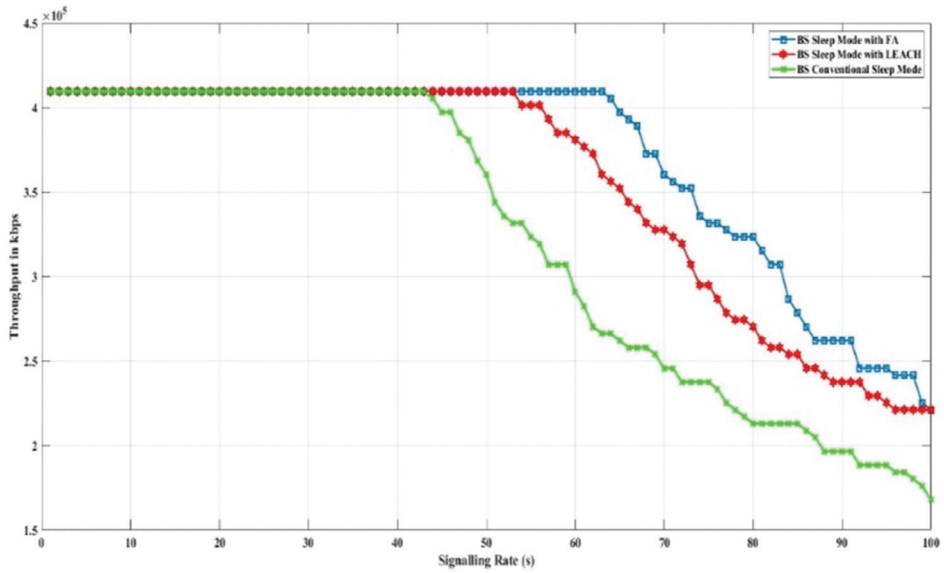


Fig. 12. Throughput of small cell 5G network with 50% initial energy

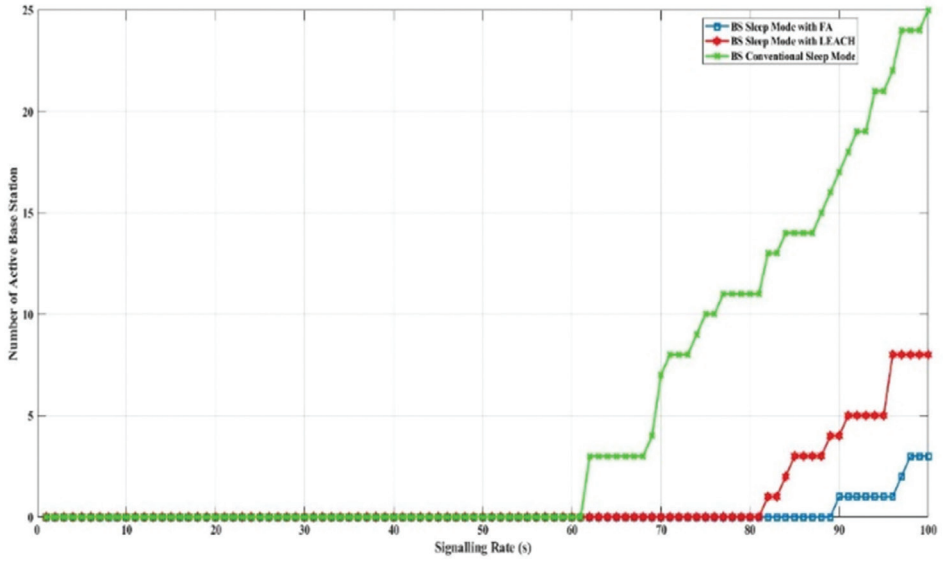


Fig. 13. Small cell 5G network with 75% initial base station energy

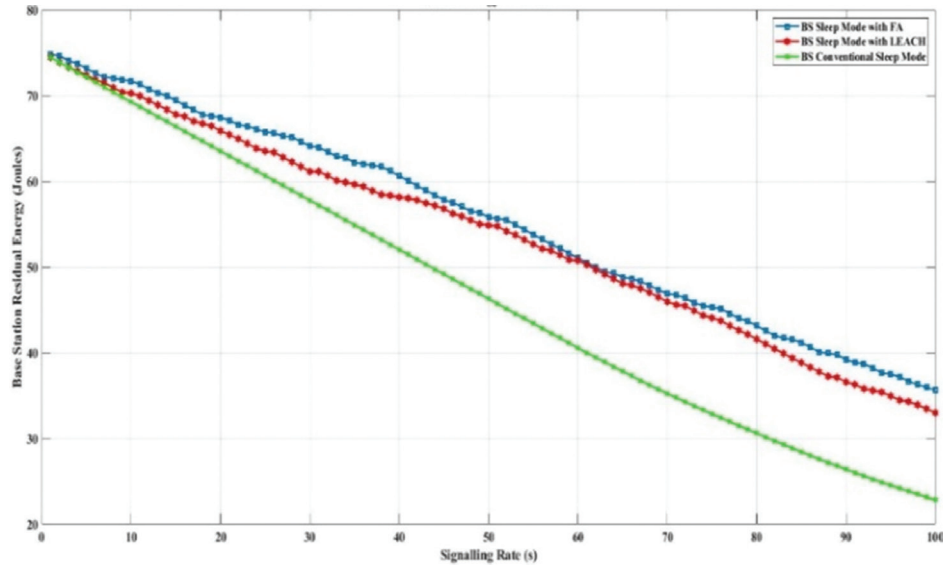


Fig. 14. Residual energy in small cell 5G network with 75% initial base station energy

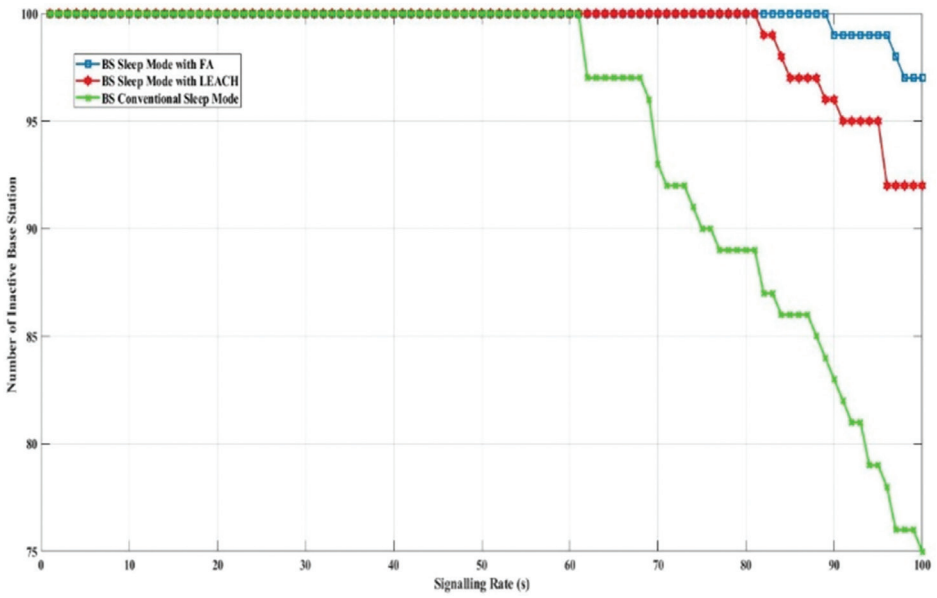


Fig. 15. Number of inactive 5G base station with 75% initial energy

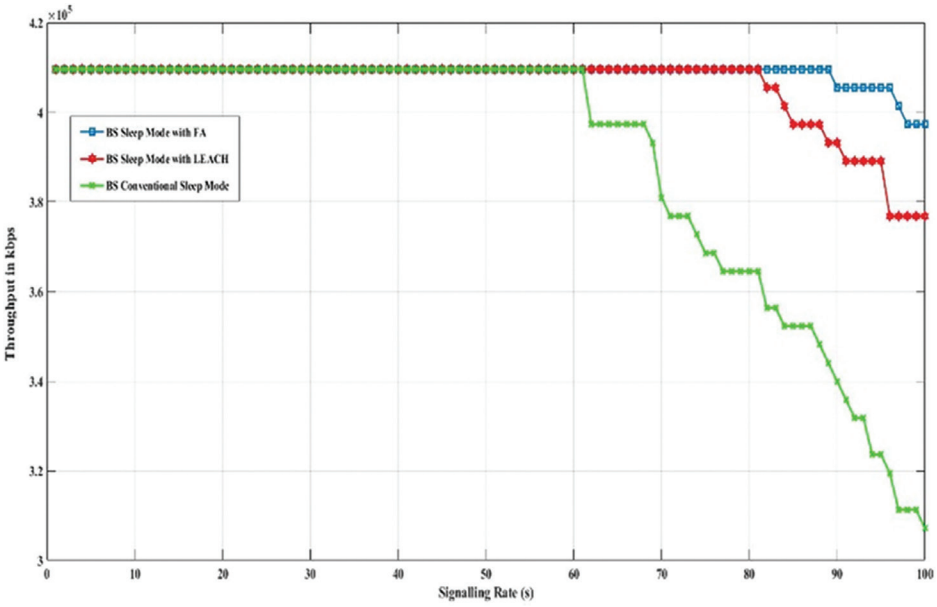


Fig. 16. Throughput of small cell 5G network with 75% initial energy

Fig. 17 displays the small-cell 5G network with inverse Gaussian traffic arrival. Actually, in an inverse Gaussian traffic arrival model, the inter-arrival times of packets follow an inverse Gaussian distribution. This type of traffic model can be used to represent bursty traffic patterns commonly observed in communication networks. Fig. 18 shows the residual energy in a small-cell 5G network with inverse Gaussian traffic arrival. This process typically involves simulation or analytical modeling, where you simulate the behavior of the network over time and observe the energy dynamics. Depending on the complexity of the model and the simulation environment, this calculation may require advanced tools such as network simulators or custom software implementations. Fig. 19 displays the number of inactive 5G base stations with an inverse Gaussian traffic arrival. This process typically involves simulation or analytical modeling, where you simulate the behavior of the network over time and observe the energy dynamics. Depending on the complexity of the model and the simulation environment, this calculation may require advanced tools such as network simulators or custom software implementations. The throughput of a small-cell 5G network with inverse Gaussian traffic arrival is shown in Fig. 20. The calculation of the throughput of a small-cell 5G network with inverse Gaussian traffic arrival involves modeling the traffic arrival pattern, resource allocation, and network conditions.

From this research, it is experienced that managing residual energy in small cell 5G networks with inverse Gaussian traffic arrival requires adaptive energy management strategies, accurate energy prediction

models, and careful optimization to balance energy efficiency with network performance requirements. The attained results address these challenges, operators can maximize the operational lifetime and sustainability of small cell deployments while ensuring high-quality service delivery to users.

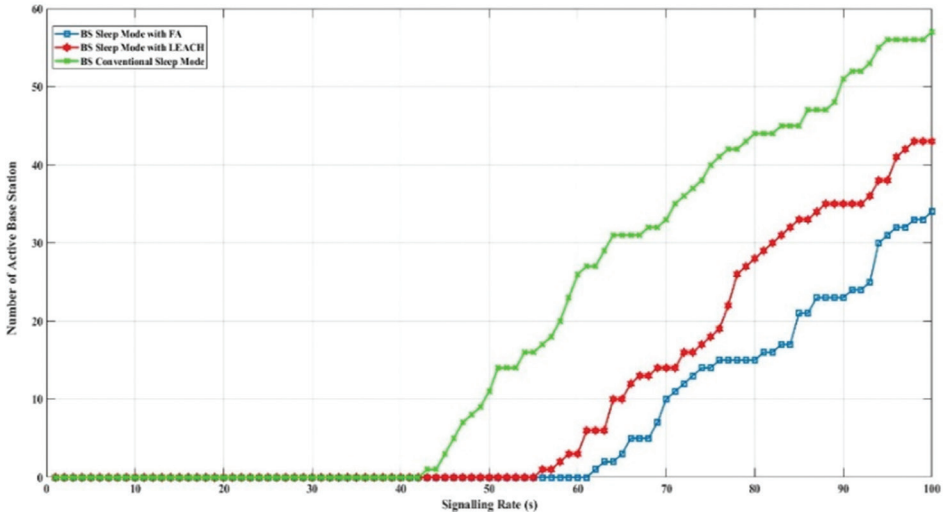


Fig. 17. Small cell 5G network with inverse Gaussian traffic arrival

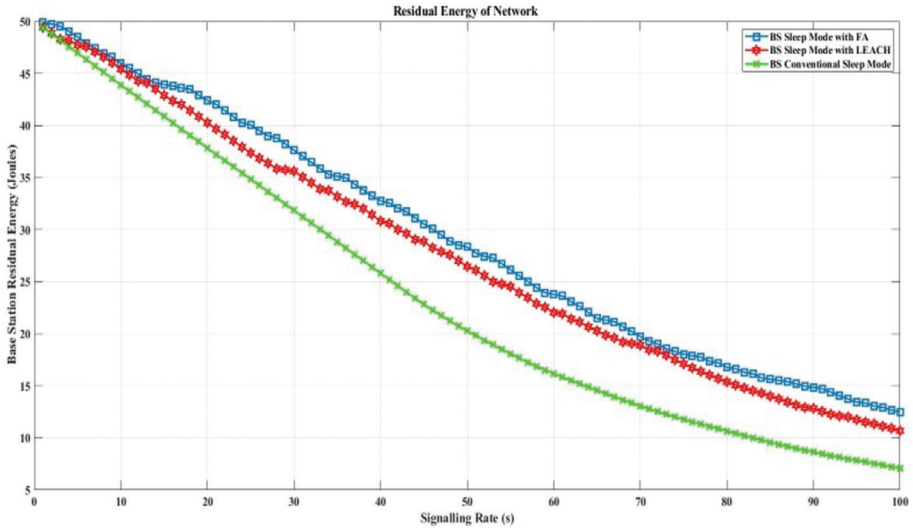


Fig. 18. Residual energy in small cell 5G network with inverse Gaussian traffic arrival

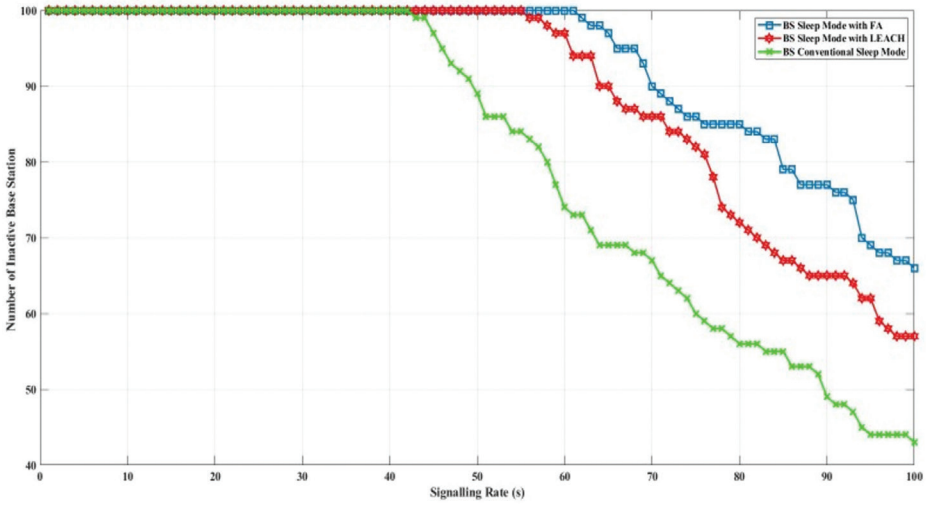


Fig. 19. Number of inactive 5G base station with inverse Gaussian traffic arrival

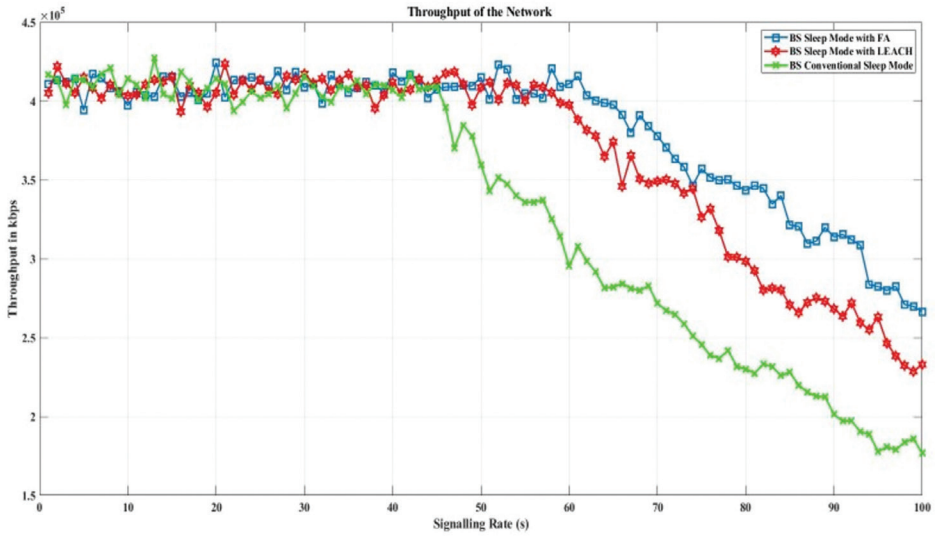


Fig. 20. Throughput of small cell 5G network with inverse Gaussian traffic arrival

In 5G cellular networks, lowering base station power consumption is an important objective for a number of reasons, including the opportunity to deploy more energy-efficient networks and environmental sustainability, as well as operating cost reductions. An optimization technique inspired by nature, the Firefly Algorithm (FF), can be used to improve a number of variables, including wireless network power usage. Energy efficiency, cost savings, environmental impact, extended network lifespan, capacity and performance optimization, adaptability to dynamic environments, regulatory compliance, trade-offs, and challenges are some possible effects and advantages of using the Firefly Algorithm for power optimization in 5G base stations. In conclusion, using the Firefly Algorithm to lower 5G base station power usage can have a variety of advantageous effects, including sustainable environmental and economic gains. In summary, leveraging the firefly algorithm for performance measurement and optimization of small cell power management in 5G networks can lead to significant improvements in network efficiency, quality of service, energy efficiency, capacity optimization, and overall network performance.

6. CONCLUSION

In order to improve interior user coverage and cell capacity in the 5G network, low-power small-cell base stations are deployed in residential and commercial buildings. However, power consumption from macro- and small-cells has grown more than the former, and this is a possible issue that the proposed 5G network aims to address. In order to save energy in 5G networks, we presented firefly optimization-based power management in this study. Comparing the suggested firefly optimization to traditional power management strategies, simulation results demonstrate a notable improvement in energy conservation with increased throughput and decreased latency. In a 5G network, cutting power and limiting interference has several advantages, including lower operating costs, environmental sustainability, better network performance, increased spectrum efficiency, and an improved user experience. These elements support a 5G network's overall performance and competitiveness.

7. REFERENCES

[1] Y.-H. Choi, "Energy Efficient Operation of Cellular Network Using On/Off Base Stations", International Journal of Distributed Sensor Networks, Vol. 11, No. 8, 2015, pp. 1-7.

[2] J. Malmodin, Å. Moberg, D. Lundén, G. Finnveden, N. Lövehagen, "Greenhouse gas emissions and operational electricity use in the ICT and entertainment & media sectors", Journal of Industrial Ecology, Vol. 14, No. 5, 2010, pp. 770-790.

[3] J. Lorincz, T. Garma, G. Petrovic, "Measurements and modelling of base station power consumption under real traffic loads", Sensors, Vol. 12, No. 4, 2012, pp. 4281-4310.

[4] D. Willkomm, S. Machiraju, J. Bolot, A. Wolisz, "Primary users in cellular networks: A large-scale measurement study", Proceedings of the 3rd IEEE Symposium on New Frontiers in Dynamic Spectrum Access Networks, Chicago, IL, USA, 14-17 October 2008, pp. 1-11.

[5] D. Lopez-Perez, I. Guvenc, G. De la Roche, M. Kountouris, T. Q. Quek, J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks", IEEE Wireless communications, Vol. 18, No. 3, 2011, pp. 22-30.

[6] N. Jabeur, "A firefly-inspired micro and macro clustering approach for wireless sensor networks", Procedia Computer Science, Vol. 98, 2016, pp. 132-139.

[7] K.M. Mamatha, M. Kiran, "Firefly Algorithm for Self Organization of Mobile Wireless Sensor Network", Journal of Communications, Vol. 15, No. 3, 2020, pp. 270-275.

[8] B. Pitchaimanickam, G. Murugaboopathi, "A hybrid firefly algorithm with particle swarm optimization for energy efficient optimal cluster head selection in wireless sensor networks", Neural Computing and Applications, Vol. 32, 2020, pp. 7709-7723.

[9] N. V. S. N. Sarma, M. Gopi, "Implementation of energy efficient clustering using firefly algorithm in wireless sensor networks", International Proceedings of Computer Science and Information Technology, Vol. 59, 2014, p. 1.

[10] R. Kaur, A. Mittal, R. Aggarwal, "Fire Fly Optimization Algorithm based Clustering by Preventing Residual Nodes in Mobile Wireless Sensor Networks", Indian Journal of Science and Technology, Vol. 9, 2016, p. 33.

[11] B. Mostafa, C. Saad, H. Abderrahmane, "Firefly algorithm solution to improving threshold distributed energy efficient clustering algorithm for heterogeneous wireless sensor networks", IAES International Journal of Artificial Intelligence, Vol. 6, No. 3, 2017, p. 91.

[12] R. Tao, J. Zhang, X. Chu, "An energy saving small cell sleeping mechanism with cell expansion in heterogeneous networks", Proceedings of the IEEE 83rd Vehicular Technology Conference, Nanjing, China, 15-18 May 2016, pp. 1-5.

[13] P. N. Sarma, M. Gopi, "Energy efficient clustering using jumper firefly algorithm in wireless sensor networks", arXiv:1405.1818, 2014.

[14] C. Desset et al. "Flexible power modeling of LTE base stations", Proceedings of the IEEE Wireless Communications and Networking Conference, Paris, France, 1-4 April 2012, pp. 2858-2862.

[15] M. Yan, C. A. Chan, A. F. Gygax, J. Yan, L. Campbell, A. Nirmalathas, C. Leckie, "Modeling the total energy consumption of mobile network services and applications", Energies, Vol. 12, No. 1, 2019, p. 184.

[16] N. F. Johari, A. M. Zain, N. H. Mustaffa, A. Udin, "Firefly Algorithm for Optimization Problem", Applied Mechanics and Materials, Vol. 421, 2013, pp. 512-517.

[17] S. Bagal, V. Hayagreev, S. Nazare, T. Raikar, P. Hegde, "Energy Efficient Beamforming for 5G", Proceedings of the International Conference on Recent Trends on Electronics, Information, Communication & Technology, Bangalore, India, 27-28 August 2021, pp. 928-933.

[18] Y. Xie, B. Li, X. Zuo, M. Yang, Z. Yan, Q. Xue, "Outage analysis for 5G beamforming heterogeneous networks", Proceedings of the IEEE International Conference on Signal Processing, Communications and Computing, Hong Kong, 5-8 August 2016, pp. 1-6.

Adaptive Speech Coding Method Based on Singular Value Decomposition and Grey Wolf Optimization for Arabic Language

Original Scientific Paper

Hassan Kassim Albahadily
University of Mustansiriyah, College of Science, Department of Computer Science
Baghdad, Iraq
hassan@uomustansiriyah.edu.iq

Alaa A. Jabbar Altaay
University of Mustansiriyah, College of Science, Department of Computer Science
Baghdad, Iraq
alaaaltaay@uomustansiriyah.edu.iq

Jamal N. Hasoon
University of Mustansiriyah, College of Science, Department of Computer Science
Baghdad, Iraq
Jamal.hasoon@uomustansiriyah.edu.iq

Abstract *– Speech coding plays a crucial role in maintaining speech quality while optimizing network resources and expediting transmission, as well as facilitating the storage of speech data. In this paper, an adaptive method for speech coding using singular value decomposition (SVD), grey wolf optimization (GWO), and run-length encoding (RLE) was proposed. The proposed method streamlines the speech matrix through preprocessing, converting it into short intervals. Subsequently, each interval undergoes decomposition using SVD, followed by optimization of compression quality using GWO. Finally, RLE is employed as the last step to increase space-saving. The developed method was conducted on two datasets: Quran and LibriSpeech using PSNR, PSEQ, and MOS tests. The results demonstrate promising outcomes, achieving space-saving up to 89.80, 84.04, 74.76, 67.24, and 59.52, respectively, for different values of quality (10, 20, 30, 40, and 50). GWO was used to optimize the quality factor which varies in each block, further increasing the space-saving up to 85.77. The average value of PSNR was equal to 21.3, MOS = 4.71, and PSEQ was equal to 3.95. Lastly, the RLE method effectively reduced the size of speech matrices, yielding a highly satisfactory space saving of up to 90.77, while maintaining excellent speech quality.*

Keywords: *Adaptive speech compression, Singular value decomposition (SVD), Grey wolf optimization (GWO)*

Received: November 12, 2023; Received in revised form: January 12, 2024; Accepted: March 16, 2024

1. INTRODUCTION

Speech is a form of audio data with specific requirements that must be met for the compressed data to be understandable [1]. Speech compression plays a vital role in modern digital life, as it minimizes storage requirements and facilitates transmission over networks. In both scenarios, the main objective is to reduce costs and save time. Speech compression is particularly important in applications such as teleconferencing, where transmitting large amounts of data is not cost-effective. Therefore, any method capable of reducing transferred data is considered cost-effective.

Adaptive speech coding methods have a crucial role in speech compression. These methods have enabled efficient speech transmission across various communication systems, including mobile networks, voice-over IP (VoIP), and streaming services [2].

Various methods are used for speech coding, including transformations along with optimization techniques such as discrete cosine transform (DCT), discrete wavelet transform (DWT), and singular value decomposition (SVD) [3].

Transformation is commonly used in compression, occurring after the preprocessing stage, to convert the data distribution from the time domain into the frequency domain to identify the potential for quantizing new data distribution in a way that effectively reduces its size [4].

Certain compression methods use swarm intelligence in the quantization process to achieve the desired quality of compressed speech [5].

SVD is a powerful mathematical tool commonly used for data compression. In the realm of speech compression, SVD is useful for reducing the dimensionality of speech, thereby aiding in the reduction of storage and transmission signals associated with speech data [6].

The singular vectors are a set of orthonormal vectors that span the same space as the original matrix, while the singular values are scalars that represent the relative importance of each singular vector [7]. Widely used in data science and engineering, SVD is a powerful tool for analyzing and manipulating matrices. SVD finds applications in various fields, including image and signal processing, data compression, machine learning, and many other fields [8].

Optimization is required to enhance the compression process. Grey wolf optimization (GWO) is a metaheuristic optimization algorithm introduced as a novel technique for resolving complex issues [9]. It has proven to be an effective optimization method [10], performing well in both unimodal and multimodal problem scenarios. GWO has successfully tackled optimization problems such as feature selection, image compression, and power system optimizations [11]. One notable advantage of GWO is simplicity and ease of implementation, with only a few parameters requiring tuning [12].

The structure of this paper comprises an introduction, a literature review, sections on SVD, GWO, RLE, the proposed method, and a final segment dedicated to results and conclusions.

The contribution of this work lies in highlighting the significance of utilizing the GWO optimization algorithm in conjunction with the SVD method to enhance speech compression. This involves selecting an optimal quality key to maximize compression efficiency. The proposed method is specifically applied to a distinct type of speech (the Quranic intonation of the Arabic language).

2. LITERATURE REVIEW

Recently, several studies have investigated adaptive speech coding methods, and a selection of these efforts is outlined below.

Hosny et al. [13] introduced two voice compression methods based on wavelet transforms, zero wavelet transform and average zero wavelet transform. These methods decompose the speech signal into multiple-resolution components, eliminating low-energy components to improve compression. This approach achieved a space-saving of 16.56 with a PSNR = 27.

Vig and Chauhan [14] proposed a hybrid wavelet method for voice reduction, breaking down the speech signal into multi-resolution components and eliminating low-energy signals using Walsh and DCT. By adjusting the threshold, this method achieved a space-saving of 72.10 with a PSNR of 49.71.

Alsalam et al. [15] employed contourlet and wavelet transforms for voice compression. The one-dimensional wavelet-transformed voice is converted into a two-dimensional array for contourlet transformation, followed by applying the contourlet transform on the high wavelet coefficients. This method achieved a space-saving of 53 with SNR = 33.

Vatsa and Sahu [16] proposed a speech compression method using discrete wavelet transform (DWT) and DCT with RLE and Huffman encoding to remove redundancies. The developed method was evaluated through compression factor, PSNR, MOS, and normalized root mean square error, achieving a space-saving of up to 29.11 with a PSNR of 16.39.

Bousselmi [17] developed an adaptive speech compression method based on the discrete wave atoms transform. This approach involves truncating signals based on the SNR and then using RLE and Huffman coding. The researchers found that the wave atom transform outperforms other wave transforms, achieving a notable space-saving of up to 10.78 with a PSNR of 36.74.

Abduljaleel [18] proposed a method for compressing and encrypting speech signals based on Sudoku, fuzzy C-means, and the Threefish cipher. The initial step involves removing low frequencies, followed by the fuzzy C-means method. This method successfully achieved a space-saving of 50.20 with a PSNR of 41.40.

Elaydi [19] introduced a lossy compression scheme using the DWT, resulting in a space-saving of 3.33 with a PSNR of 44.85.

The mentioned studies developed new techniques of speech compression using SVD, DCT, and DWT with some modifications or enhancements. All developed methods were tested on different datasets using objective tests like SNR, PSNR, and Compression factor and they achieved a good ratio of compression.

3. SINGULAR VALUE DECOMPOSITION

SVD is a mathematical tool providing substantial theoretical and technical insights into linear transformations with algebraic features [20]. It is decomposing a matrix into three matrixes returning the original matrix if they are combined.

The SVD decomposition of a matrix A can be represented by the following equation.

$$A = U\,S\,V^T \tag{1}$$

Where:

$U = m \times m$ matrix of orthonormal eigenvectors of AAT

$S = n \times n$ matrix of diagonal elements.

VT = the transpose of an $n \times n$ matrix containing the orthonormal eigenvectors of ATA.

Fig. 1 illustrates the SVD for matrix A [20].

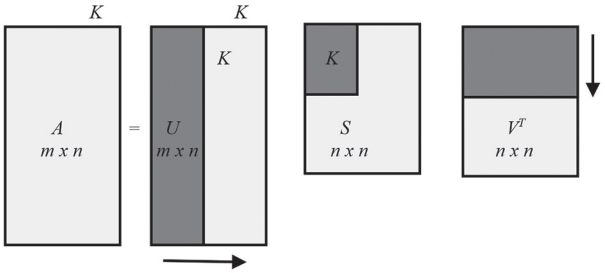


Fig. 1. SVD of matrix A

K represents the number of columns in the first matrix U, the number of elements from the diagonal of the second matrix S, and the number of rows in the third matrix V. The K parameter controls the quality of compressed speech and how many columns will be used in the decomposition process.

4. GREY WOLF OPTIMIZATION (GWO)

GWO is one of the swarm intelligent algorithms (metaheuristic algorithm) invented by Mirjalili *et al*. [21], which is modelling the hunting behavior of grey wolves [22]. The algorithm is designed to mimic the hierarchical structure and hunting strategy observed in grey wolves in the wild [23]. The GWO algorithm consists of four main components: social hierarchy, tracking/hunting, surrounding, and attacking prey [24, 25].

Grey wolves are categorized into four types based on their social hierarchy: alpha (α), beta (β), delta (δ), and omega (ω). The leadership hierarchy of wolves is illustrated in Fig. 2.

Fig. 2. The wolf leadership hierarchy

The primary decision-maker and leader is the alpha wolf, with beta and delta wolves supporting the alpha wolf in this role. The three leadership wolves (α, β, and δ), possessing the highest fitness levels, take charge of the hunting and optimization process, while the omega (ω) wolves follow their lead.

The following equations can be utilized to quantitatively represent the surrounding prey process:

$$X(t+1)=X(t)-A.D \tag{2}$$

where X_p represents the location of the prey, A defines the coefficient vector, and D is defined as:

$$parentD=|C.X_p\,(t)-X(t)| \tag{3}$$

where C stands for the coefficient vector, X defines the location of the grey wolf, and t represents the current iteration. The coefficient vectors A and C are determined by the following equations:

$$A=2a.r_1-a \tag{4}$$

where elements of a are linearly reduced from 2 to 0 over the sequence of iterations, and $r1$ and $r2$ define the random vectors in the range [0, 1].

Hunting: In terms of hunting, the first three prominent solutions (α, β, and δ) attained are stored and induce other search agents (including ω) to adjust their positions, considering the position of the best search agent. The positions of the grey wolves can be updated according to the following equations.

$$X(t+1)=(X\alpha+X\beta+X\delta\,)/3 \tag{5}$$

5. RUN-LENGTH ENCODING (RLE)

RLE is a lossless approach that provides reasonable space-saving for specific data types by replacing consecutive data values in a file with a count number (run) and its value. The implementation of RLE sometimes depends on the data type being compressed. The operation of this method is illustrated in Fig. 3.

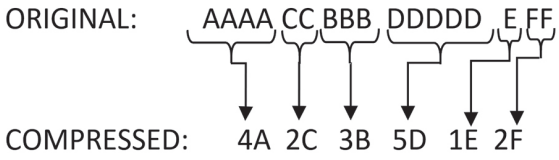


Fig. 3. The work of RLE algorithm

6. THE PROPOSED METHOD

The proposed adaptive method is introduced for speech compression, starting with the necessary preprocessing steps. This involves framing, removing silent intervals, and then reshaping a specific number of samples into a two-dimensional matrix for the decomposition process.

The framework of the proposed method is shown in Fig. 4.

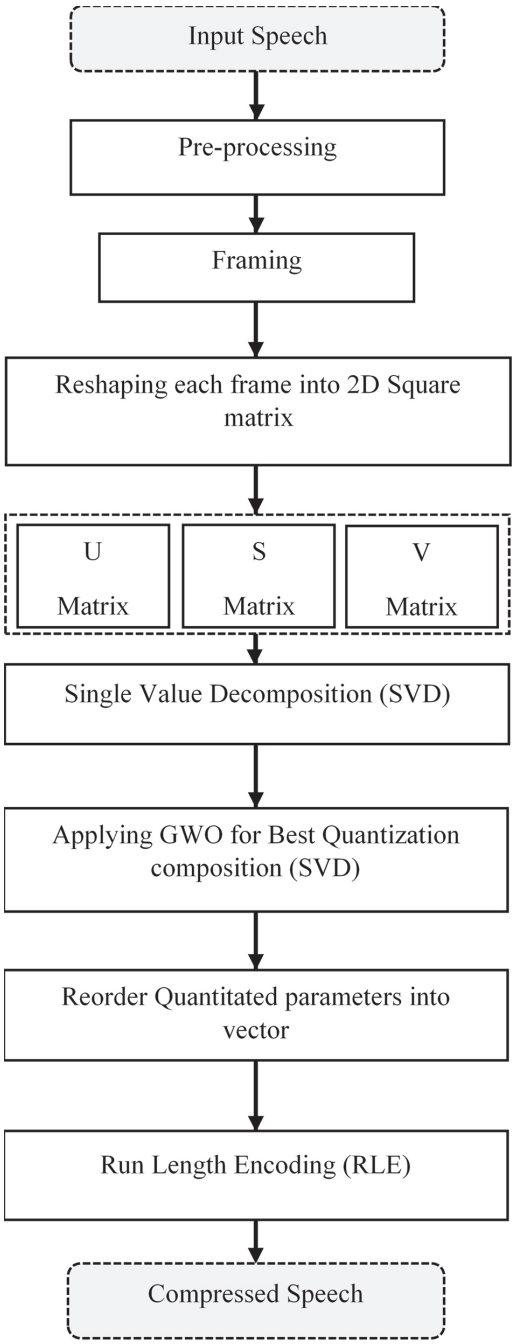


Fig. 4. The framework of the proposed method

6.1. PREPROCESSING OF SPEECH

The preprocessing of input speech involves two main steps: filtering the speech with a noise-removing filter called the Wiener filter and removing silence intervals.

The Wiener filter is designed to reduce the impact of additive noise in a signal while preserving the desired signal components. In the context of speech processing, the Wiener filter estimates the power spectral density of both the noise and the signal. Subsequently, it applies a frequency-dependent gain to the noisy signal, aiming to minimize the mean square error between the estimated clean signal and the noisy signal. This process enhances the SNR of the speech, making downstream speech processing tasks more effective.

Silence intervals, which are non-speech segments that contain no useful information, are then removed to reduce the computational data dimensionality, focusing solely on the speech data. This method involves establishing a threshold based on the amplitude or energy of the speech signal. Segments that fall below this threshold are identified as silence and removed. This step is beneficial in decreasing the amount of unnecessary data that needs processing, particularly in applications where only the speech content is relevant.

6.2. FRAMING OF SPEECH

Framing is the process of dividing a continuous stream of data representing speech signals into smaller segments called frames. This process is defined by specifying the size of the square block, such as 256×256, which is equivalent to 65,536 samples. After removing silent intervals, the remaining samples are partitioned into frames, each equal to this specified value. If the size is less than the selected value, the final frame is padded.

Before entering the decomposition process, the frames are reshaped into two-dimensional matrices. These matrices can have their dimensions (number of rows and columns) adjusted according to specific requirements by modifying the arguments passed to the reshape method.

6.3. SINGULAR VALUE DECOMPOSITION OF BLOCK RESHAPING

The SVD process is used to decompose a two-dimensional matrix into a one-dimensional matrix, allowing for efficient data manipulation. Each block will select k rows from the three matrices and ignore the remaining rows. The process of SVD decomposition is shown in Fig. 5.

6.4. APPLYING GWO FOR BEST QUANTIZATION COMPOSITION

The selection of k values for all blocks is enhanced through the application of GWO. GWO randomly selects numbers and then optimizes them to strike a balance between the output size and the quality of speech compression. This process involves using GWO to fine-tune the "k" values for data blocks, ensuring the desired balance between reducing data size and preserving information integrity is achieved.

6.5. REORDERING QUANTITATED PARAMETERS

In this step, the chosen parameters are rearranged into a single vector, which represents the compressed speech. The process is applied to each block, from start to end, and all other values are quantized to a specific decimal digit. This step is important for the following stage of lossless compression.

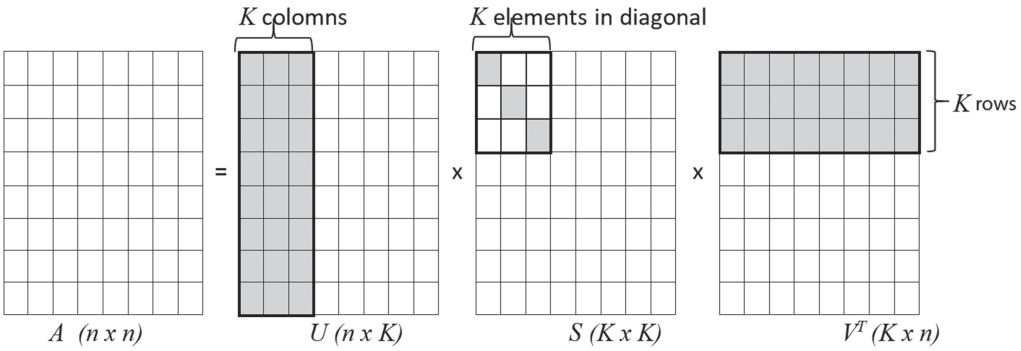


Fig. 5. Reshaping the matrix using SVD

6.6. RUN-LENGTH ENCODING OF QUANTIZED VECTOR

RLE is a highly efficient compression algorithm, especially effective when there is a long sequence of identical values in the data. It excels at reducing redundancy and significantly decreases the overall size of the data. This technique is applied to the truncated parameters, resulting in compressed data represented in pairs indicating the run and its length.

7. RESULTS

To assess the efficiency of the proposed method and validate the results, the method has been implemented on two datasets. The first dataset comprises twenty speech files of Quranic intonation S1-S5 with varying durations (3, 6, 9, and 12 seconds), frequency of 8000 KHz, and mono channel. The second dataset is LibriSpeech L1-L5 with durations (3, 6, 9, and 12 seconds), frequency of 8000 KHz, mono channel, yielding 300 seconds in total.

The experiments were conducted on the datasets using MATLAB 2020b on a computer with 2.4 GHz CPU frequency and 16 GB of memory under the Windows 10 operating system.

The space-saving factor was used as the reduction in size relative to the uncompressed size as shown in the following equation.

$$\text{Space Saving} = 1 - \frac{compresed\ Size}{uncompressed\ Size} \quad (6)$$

The obtained results, after applying SVD on the datasets, are presented in Table 1 and Table 2.

Table 1. The space-saving ratio for Quran intonation

	Duration (Second)			
***K* values**	**3**	**6**	**9**	**12**
10	89.91	88.29	89.77	91.61
20	84.16	84.10	83.69	84.23
30	75.34	76.10	74.65	72.94
40	66.29	67.35	68.16	67.16
50	59.08	59.23	60.20	59.54

The values in Table 1 represent the space-saving factor for test speech files of duration 3, 6, 9, and 12 seconds with five different values of the quality parameter (K) =10, 20, 30, 40, and 50. The results indicate that the average space-saving value is 89.90 when the quality factor K is set to 10. Subsequently, depending on the chosen compression quality K, the ratio decreases to 84.04, 74.76, 67.24, and 59.52 when K is set to 20, 30, 40, and 50, respectively.

Similar results were calculated for the dataset LibriSpeech as shown in Table 2.

Table 2. The space-saving ratio for LibriSpeech

	Duration (Second)			
***K* values**	**3**	**6**	**9**	**12**
10	91.57	90.29	90.15	90.22
20	83.29	82.22	82.06	84.21
30	75.04	75.55	75.52	75.43
40	68.33	66.82	67.39	68.04
50	60.31	60.82	59.11	59.87

The space saving is increased when K is set to a low value, implying good quality and the ratio decreases when K is set to a high value, indicating low speech quality. The relationship between the average space-saving and the quality factor K of values 10–50 is illustrated in Fig. 6.

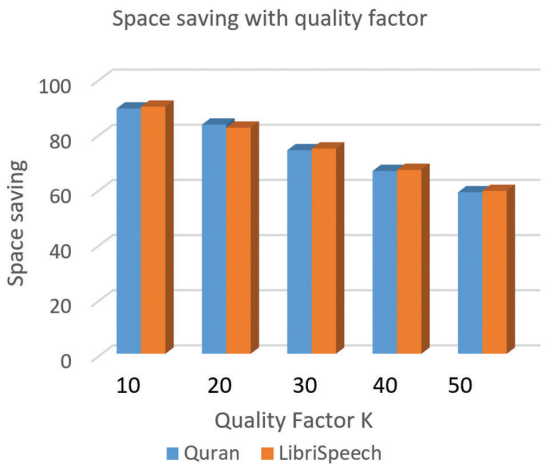


Fig. 6. The average space-saving of speech

To assess the quality of compressed speech and validate the results of the proposed method, several tests will be conducted on the recovered speech files. One commonly used test is PSNR, which measures the ratio between the maximum possible value of a signal and the power of distorting noise that impacts its quality.

The PSNR values are shown in Table 3 and Table 4.

Table 3. PSNR (dB) for Quran intonation

	Duration (Second)			
K values	3	6	9	12
10	6.93	4.76	4.70	4.42
20	9.14	10.96	10.38	8.26
30	17.11	16.17	14.34	17.54
40	17.31	16.40	15.04	21.05
50	28.86	24.77	23.09	27.17

Table 3 represents the PSNR values of test speech files of duration 3, 6, 9, and 12 seconds with five different values of quality parameter (K) =10, 20, 30, 40, and 50. The results indicate that the average PSNR value is 5.20 when the quality factor K is set to 10. Subsequently, depending on the chosen compression quality K, PSNR increases to 9.68, 16.29, 17.45, and 25.97 when K is set to 20, 30, 40, and 50, respectively.

Similar results were found for the dataset LibriSpeech as shown in Table 4.

Table 4. PSNR (dB) for LibriSpeech

	Duration (Second)			
K values	3	6	9	12
10	6.31	6.26	4.47	5.01
20	8.71	8.28	7.51	10.70
30	12.37	16.21	10.73	14.31
40	17.19	14.27	20.57	16.88
50	29.75	23.01	23.78	26.75

PSNR is increased when K is set to a high value, showing high noise and low speech quality. The relationship between the average PSNR and the quality factor K of values 10–50 is illustrated in Fig. 7.

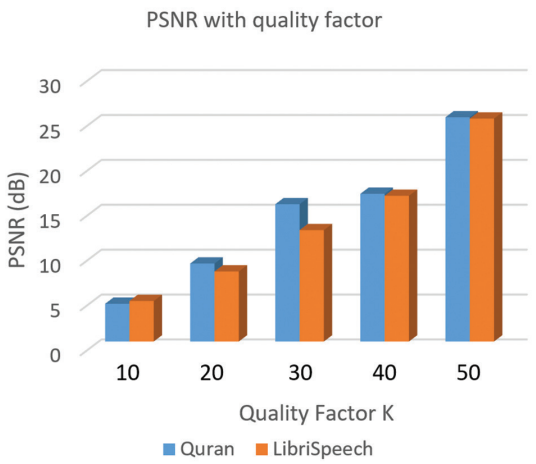


Fig. 7. The mean PSNR of the compressed speech

The speech frames contain varying data, and it is important to select a varying value of quality factor K according to the contents of the frame. The GWO is employed to make the selection of K varying, depending on the frame significance. The space-saving increases when using GWO optimization, albeit with a trade-off of relatively subdued speech quality. The results after applying GWO optimization are presented in Table 5 and Table 6.

Table 5. Space-saving with GWO for Quran intonation

	Duration (Second)			
Speech Files	3	6	9	12
S1	86.63	84.29	86.13	86.04
S2	85.21	84.91	85.64	85.88
S3	86.94	86.16	85.68	85.47
S4	84.97	85.28	86.12	85.79
S5	86.33	85.35	85.94	86.58
Average	86.02	85.20	85.90	85.95

Table 5 represents the space-saving factor after using GWO for test speech files of duration 3, 6, 9, and 12 seconds. The average values are listed in the last row. The results indicate that the average space-saving value is almost the same for the test files with an average = 85.77.

Similar results were calculated for the dataset LibriSpeech as shown in Table 6.

Table 6. Space-saving with GWO for LibriSpeech

	Duration (Second)			
Speech Files	3	6	9	12
L1	82.50	79.15	85.87	78.04
L2	84.51	80.15	82.11	80.53
L3	80.93	84.99	77.36	82.38
L4	80.84	84.87	80.17	77.33
L5	85.36	77.35	85.60	85.97
Average	82.83	81.30	82.22	80.85

The space-saving is increased for all files because the value of quality factor K is selected as the best value for each frame. The relationship between the average space-saving after using GWO is illustrated in Fig. 8.

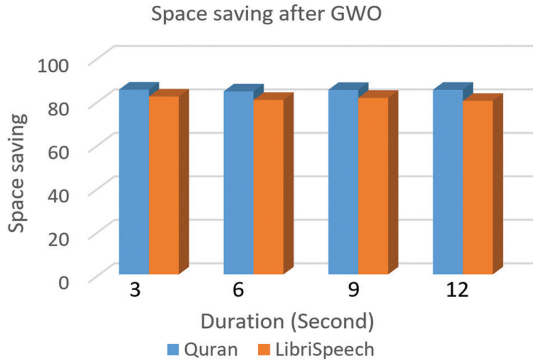


Fig. 8. Space-saving of speech after GWO

The subjective test will be applied to check the quality of speech. The first test is Mean Opinion Score (MOS) which is used to evaluate results and measure the quality of compressed speech. MOS consists of five values (5 = excellent, 4 = very good, 3 = fair, 2 = poor, 1 = bad) to express the quality of speech as perceived by listeners [26]. The MOS test involves presenting both the original and compressed speech to ten native speakers individuals (five males and five females), who then assign values from 1 to 5 based on the quality they perceive. Forty speech files were used in the test, each file contains one sentence and several words depending on the duration of the file. The results of the MOS test are shown in Table 7 and Table 8.

Table 7. MOS test for Quran intonation

	Duration (Second)			
Speech Files	**3**	**6**	**9**	**12**
S1	3.76	4.36	3.54	3.93
S2	3.66	4.44	4.04	4.57
S3	4.02	4.20	4.55	4.18
S4	3.93	4.78	4.15	4.24
S5	4.64	3.70	3.58	4.95

Table 7 represents the average values of MOS for test speech files of duration 3, 6, 9, and 12 seconds. The results show that the average MOS values are 4, 4.30, 3.97, and 4.37 which indicate very good quality with an average value =4.71.

Similar results were calculated for the dataset LibriSpeech as shown in Table 8.

Table 8. MOS test for LibriSpeech

	Duration (Second)			
Speech Files	**3**	**6**	**9**	**12**
L1	4.84	4.60	4.70	3.76
L2	4.16	4.21	4.44	4.69
L3	3.74	3.85	4.06	3.96
L4	4.23	4.59	4.10	4.15
L5	4.48	4.11	4.55	4.43

The results obtained from Table 8 indicate that the quality was acceptable, and all compressed speech files were nearly identical to the original files.

Another subjective test is the Perceptual Evaluation of Speech Quality PSEQ, which analyzes speech signals and considers a good result if the score is above 3.5. The results of PESQ are shown in Table 9.

Table 9. PESQ test for Quran and LibriSpeech

Quran intonation		**LibriSpeech**	
Speech File	**PESQ**	**Speech File**	**PESQ**
S1	3.983	**L1**	3.822
S2	3.992	**L2**	3.914
S3	3.906	**L3**	3.899
S4	3.921	**L4**	3.904
S5	3.932	**L5**	3.903

Table 9 represents the average values of PESQ for test speech files of duration 3, 6, 9, and 12 seconds for both datasets Quran (S1-S5) and LibriSpeech (L1-L5). The results show that the average PESQ value is 3.95 for Quran files and 3.92 for the LibriSpeech dataset, which indicates very good quality for the compressed speech.

The final step involves applying the lossless compression method RLE, which relies on identifying repeated similar values present in the speech. The results from the RLE method increased the space-saving by 5–7%, resulting in a final space-saving of 90–92% for the collected dataset and 85-90% for the LibriSpeech dataset.

The suggested method has demonstrated promising results when compared with other referenced efforts. The comparison between the suggested method and other methods is presented in Table 10.

Table 10. Performance comparison with related work

Reference	**Compression Method**	**Space Saving**	**MOS**	**PSNR**
[12]	ZWT,AZWT	85.44	3.8	27.0
[13]	DCT	72.10	-	49.71
[22]	DWT, DCT	70.89	-	16.39
[23]	DWAT	89.22	-	36.74
[24]	C-Means	50.20	-	41.40
[25]	DWT	66.7	-	44.85
Suggested Method	SVD	85.36	4.16	21.3

According to the results from Table 10, and comparing the suggested method with the related studies, the suggested method provides good results and can be used in different cases and applications.

To validate the results and ensure the stability of the suggested method, the experiment was tested on the datasets by the authors as a trial presentation, then repeated two times with a total time equal to 150 minutes, in addition to the subjective test which takes 30 minutes which confirming the reliability of the tests to validate the judgments of the obtained results and the suggested method.

8. CONCLUSION

An adaptive method that combines SVD, GWO, and RLE has been proposed for compressing speech signals. The method has shown promising results, achieving up to a 92% reduction in the size of speech files compared to their original size. The quality of the compressed speech was evaluated using PSNR, which yielded a value of 21.3. The validation test was supported by the subjective tests MOS which was 4.71 and PESQ which yielded 3.95, indicating excellent speech quality.

For future work, it is essential to explore the application of the proposed method in speech storage and over VoIP protocols for transferring audio files through Internet-of-Things applications after encrypting the files. Additionally, it is recommended to utilize optimi-

zation algorithms based on AI and DNA, with a focus on saving the most frequently repeated words and generating their compressed equivalents.

ACKNOWLEDGMENT

The authors would like to thank the University of Mustansiriyah, Baghdad, Iraq, for the support and provision of open resources that facilitated the completion of this work.

9. REFERENCES

[1] R. Cox, C. Kamm, L. Rabiner, J. Schroeter J. Wilpon, "Speech and language processing for next-millennium communications services", Proceedings of the IEEE, Vol. 88, No. 8, 2000, pp. 1314-1337.

[2] H. Xie, Z. Qin, G. Li, B. Juang, "Deep Learning Enabled Semantic Communication Systems", IEEE Transactions on Signal Processing, Vol. 69, 2021, pp. 2663-2675.

[3] J. James, V. Jyothi, "A Comparative Study of Speech Compression using Different Transform Techniques", International Journal of Computer Applications, Vol. 97, No. 2, 2014, pp. 16-20.

[4] V. Maider, J. Amigo, "Pre-processing of hyperspectral images Essential steps before image analysis", Chemometrics and Intelligent Laboratory Systems, Vol. 117, 2012, pp. 138-148.

[5] M. Tuba, N. Bacanin, "JPEG quantization tables selection by the firefly algorithm", Proceedings of the International Conference on Multimedia Computing and Systems, Marrakech, Morocco, 14-16 April 2014, pp. 153-158.

[6] M. Wall, A. Rechtsteiner, L. Rocha, "Singular Value Decomposition and Principal Component Analysis", A Practical Approach to Microarray Data Analysis, Springer, 2003, pp. 91-109.

[7] S. Berkant, L. Eldén, "Handwritten digit classification using higher order singular value decomposition", Pattern Recognition, Vol. 40, No. 3, 2007, pp. 993-1003.

[8] A. Cichocki, "Tensor Decompositions for Signal Processing Applications: From two-way to multiway component analysis", IEEE Signal Processing Magazine, Vol. 32, No. 2, 2015, pp. 145-163.

[9] M. Hemis, B. Boudraa, T. Merazi-Meksen, "Intelligent audio watermarking algorithm using Multi-objective Particle Swarm Optimization", Proceedings of the 4th International Conference on Electrical Engineering, Boumerdes, Algeria, 13-15 December 2015, pp. 1-5.

[10] A. Kaur, S. Sharma, A. Mishra, "An Efficient Opposition Based Grey Wolf Optimizer for Weight Adaptation in Cooperative Spectrum Sensing", Wireless Personal Communications, Vol. 118, 2021, pp. 2345-2364.

[11] L. Abualigah, A. Khader, E. Hanandeh, "A new feature selection method to improve the document clustering using particle swarm optimization algorithm", Journal of Computational Science, Vol. 25, 2018, pp. 456-466.

[12] A. Bilal, G. Sun, S. Mazhar, A. Imran, "Improved Grey Wolf Optimization-Based Feature Selection and Classification Using CNN for Diabetic Retinopathy Detection", Evolutionary Computing and Mobile Sustainable Networks, Vol. 116, Springer, 2021.

[13] N. Hosny, S. El-Ramly M. El-Said, "Novel techniques for speech compression using wavelet transform", Proceedings of the Eleventh International Conference on Microelectronics, Kuwait, 22-24 November 1999, pp. 225-229.

[14] R. Vig, S. Chauhan, "Speech Compression using Multi-Resolution Hybrid Wavelet using DCT and Walsh Transforms", Procedia Computer Science, Vol. 132, 2018, pp. 1404-1411.

[15] A. Alsalam, S. Razoqi, E. Ahmed, "Effects of Using Static Methods with Contourlet Transformation on Speech Compression", Iraqi Journal of Science, Vol. 62, No. 8, 2021, pp. 2784-2795.

[16] S. Vatsa, O. Sahu, "Speech Compression Using Discrete Wavelet Transform and Discrete Cosine Transform", International Journal of Engineering Research and Technology, Vol. 1, No. 5, 2012.

[17] S. Bousselmi, N. Aloui, A. Cherif, "Adaptive Speech Compression Based on Discrete Wave Atoms Transform", International Journal of Electrical and Computer Engineering, Vol. 6, No. 5, 2016, pp. 2150-2157.

[18] I. Abduljaleel, A. Khaleel, "Developed a Speech Signal Compression and Encryption Based on Sudoku, Fuzzy C-means and ThreefishF Cipher",

International Journal of Electrical and Computer Engineering, Vol. 11, No. 6, 2021, pp. 5049-5059.

[19] H. Elaydi, M. Jaber, M. Tanboura, “Speech Compression Using Wavelets“, Proceedings of the International Arab Conference on Information Technology, Ajman, UAE, 6-8 December 2003, pp. 1-8.

[20] N. Albatayneh, K. Ghauth, F. Chua, “A Semantic Content-Based Forum Recommender System Architecture Based on Content-Based Filtering and Latent Semantic Analysis”, Advances in Soft Computing, Vol. 287, 2014, pp. 369-378.

[21] K. Srikanth, "Meta-heuristic Framework: Quantum Inspired Binary Grey Wolf Optimizer for Unit Commitment Problem", Computers and Electrical Engineering, Vol. 70, 2018, pp. 243-260.

[22] J. Neeraj, “Applications of Grey Wolf Optimization in Control System”, Acadimic Publishing, 2018.

[23] S. Mirjalili, A. Lewis, “Grey Wolf Optimizer”, Advances in Engineering Software, Vol. 69, 2014, pp. 46-61.

[24] L. Wong, "Grey Wolf Optimizer for Solving Economic Dispatch Problems", Proceedings of the IEEE International Conference on Power and Energy, Kuching, Malaysia, 1-3 December 2014.

[25] B. Yang, "Grouped grey wolf optimizer for maximum power point tracking of doubly-fed induction generator-based wind turbine", Energy Conversion and Management, Vol. 133, 2017, pp. 427-443.

[26] P. Loizou, "Speech quality assessment", Multimedia Analysis, Processing and Communications, Springer, 2011, pp. 623-654.

Speed Control of Switched Reluctance Motor using Adaptive Fuzzy Backstepping Sliding Mode Control

Case Study

Nha Phi Hoang
Ha Noi University of Industry
Ha Noi capital, Viet Nam
nhaph@haui.edu.vn
Hung Pham Van
Ha Noi University of Industry
Ha Noi capital, Viet Nam
phamvanhung@haui.edu.vn

Abstract *– The Switched Reluctance Motors (SRMs), with many outstanding advantages, are gradually being widely applied in industries, households, and recommended in many works. However, most studies in the field of SRM control only concentrate on the mathematical model of the motor itself, neglecting the nonlinearity introduced by the inverter, which is responsible for switching between phases to drive the motor. This paper proposes an adaptive backstepping sliding mode control algorithm based on the SRM nonlinear model that combines both the motor and the inverter. Firstly, a backstepping sliding mode controller is used to track the desired value and ensure the stability of the system according to the Lyapunov criterion. Secondly, a fuzzy logic system is added to adjust the controller parameters to account for uncertainty and external disturbance, as well as to minimize the chattering phenomenon. Finally, a few simulation scenarios are performed to assess the effectiveness of the proposed controller. The simulation results clearly demonstrate that the proposed controller surpasses the previously published H infinity controller in terms of speed control quality for the combined nonlinear model of SRM. The proposed controller exhibits zero steady-state error, zero overshoot, and a short settling time of approximately 0.5 seconds. Moreover, the system's output quickly stabilizes when affected by disturbance noise.*

Keywords: *switched reluctance motors, adaptive control, backstepping sliding mode control, fuzzy logic system*

Received: August 24, 2023; Received in revised form: March 16, 2024; Accepted: March 19, 2024

1. INTRODUCTION

Switched reluctance motor have been proposed since 1946, and there are two types: rotary switched reluctance motors (SRMs) and linear switched reluctance motor (LSRM). Both types consist of a stator and rotor, with windings only on the stator poles, and they do not use permanent magnets. These motors operate through an inverter, which switches between phases.

SRMs offer several advantages due to their operational principles and structure [1]. These advantages include a high starting torque [2], a simple structure, low manufacturing costs, and high stability [3]. As a result of these benefits, SRMs are gradually finding wider applications, particularly in the field of electric vehicles for tourism [4]. However, SRMs also come with certain drawbacks, including significant pulsating torque [5], challenging control requirements, and high nonlinear characteristics [6]. The strong nonlinearity of SRMs can be attributed to their inherent structure, combined with the phase-switched converter, which introduces resonance nonlinearity [7]. The inherent structure of SRMs, in conjunction with the phase-switched converter, contributes to the pronounced nonlinearity of these motors. Consequently, when controlling an SRM, it is crucial to consider the impact of nonlinearities in the motor's kinetics, arising from the simultaneous excitation of stator phases [8]. Addressing this issue poses a key challenge that needs to be resolved [9]. Several studies have proposed mathematical models for SRMs and control strategies for switched reluctance motor drive systems based on these models [10-24]. While [10, 17, 24] address the basic learning model of the SRM, [11] presents the nonlinear model of the switched reluctance motor with consideration of the influence of mutual inductance between phases. The issue of flux is also mentioned in [12, 13], where the authors use third-order Fourier series analysis to approximate the flux curve and compare it with the flux characteristic in the Matlab library. Then, direct torque control is applied to improve the control quality

of hybrid electric vehicle systems using SRM. However, due to the difficulty in determining the flux characteristics caused by unknown parameters, the process of determining the SRM model encounters difficulties. Therefore, the authors in [14, 18] linearized this characteristic, while [15] used a neural network to identify the model. Related to the modeling of the switched reluctance motor, the document [16] considers the friction coefficient in the model and proposes the ovel Direct Instantaneous Torque Control (DITC) to reduce torque ripple. In the field of SRM control, [19] proposes the use of a PID controller combined with a genetic algorithm applied to the linear SRM model to achieve good control performance. However, it should be noted that the SRM model used in this study is an ideal case taken from the Matlab library. Another approach investigated in the literature is the application of predictive control combined with torque control to optimize flux, as studied in [22]. This method aims to enhance the performance of SRM control by considering both torque and flux optimization.

Additionally, there are other research directions exploring sensorless SRM control using nonlinear state observers, as mentioned in [20] and [23]. These methods aim to achieve control without the need for external sensors, relying on observer-based techniques to estimate the motor's states. Furthermore, the use of sliding mode observers for SRM control is discussed in [21]. Sliding mode control is a robust control technique that can handle system uncertainties and disturbances effectively. However, the majority of these studies have primarily concentrated on the mathematical model of the motor itself, overlooking the nonlinearity introduced by the inverter. In this paper, we aim to develop a control algorithm for the nonlinear model of SRM that takes into account both the motor and the inverter.

The research group led by Rigatos was the first to publish a comprehensive mathematical model that incorporates both the motor and the switch (inverter) [25]. However, their approach treated the SRM as a combined linear model for control algorithm design. Specifically, they used the H infinity nonlinear feedback controller for the combined linear model of SRM and proved its stability using Lyapunov theory, but the consideration of nonlinearity was incomplete. However, there are still some limitations in the control quality of the SRM system, specifically regarding large overshoot (around 20%) and long settling time (around 7 seconds). In particular, when changing the setpoint value, the overshoot can reach up to 50%. Building on this research [25], we retained the combined model of SRMs and applied an adaptive Backstepping sliding mode control algorithm to enhance the performance of the SRM drive system.

Several published works, including references [26-29], have employed the Backstepping nonlinear algorithm for speed control of SRMs. The Backstepping algorithm for SRMs was first introduced in [27], where a Backstepping controller combined with a state observer was proposed to stabilize the speed control. Subsequently, [26] further proposed a Backstepping control scheme combined with a state observer to achieve speed stabilization. In [28], the Backstepping algorithm was presented for SRM control considering the saliency effect. Additionally, the Backstepping technique was used in [29] to reduce circuit current ripple and improve control performance. However, it has been observed that the Backstepping control algorithm has limited adaptability to slow load noise. In the most recent research [30], a Backstepping combined sliding mode controller is chosen to address this limitation. However, the sliding control coefficient of the controller is quite difficult to select and maintain fixed during the control process, leading to limitations in the ability to respond quickly and reduce vibration for the SRM when the sliding controller changes state at the working point. To overcome this difficulty, this paper proposes the use of a fuzzy controller to flexibly adjust the parameters of the sliding control signal. By incorporating fuzzy controller into Backstepping combined sliding mode controller, the sliding surfaces can be adjusted automatically to limit the chattering phenomenon that can be caused by using the function sgn(S) and a model that includes a switch. Consequently, this paper proposes the use of a backstepping adaptive control algorithm based on fuzzy logic to address the aforementioned issues and improve control quality, even in the presence of significant noise.

Following the introductory section, the paper proceeds to present the combined nonlinear model of the SRM in Section 2, and subsequently introduces the adaptive backstepping sliding mode control algorithm based on the fuzzy logic system in Section 3. Finally, in Section 4, the simulation results obtained with the proposed controller are presented. This section evaluates various performance metrics such as settling time, overshoot, and steady-state error under different load and speed setpoint.

2. THE COMBINED NONLINEAR MODEL OF SWITCHED RELUCTANCE MOTORS

In this specialized research paper on SRM control, the mathematical model of SRMs is derived from the fundamental equations of electrical machines. The dynamics of reluctance motors include equations of voltage, equation of torque and equation of mechanics, which are represented as in (1)

$$\begin{cases} u_j = R.i_j + \dfrac{d\psi_j}{dt} \\ T_j(\theta, i_j) = \dfrac{\partial \mathrm{W}'_j}{\partial \theta} \\ J\dfrac{d^2\theta}{dt^2} = T_e - T_l \end{cases} \quad (1)$$

where j = 1, 2, 3, 4. (consider with 4-phase switching reluctance motor). In (1), uj is the voltage of phase j, R is the resistance of phase j, i_j is the current of phase j, θ is the rotor angular, T_j is the torque of phase j, the load torque T_l, the moment of inertia J and ψ_j is flux of phase j in, determined by (2)

$$\psi_j = \int_0^{2\pi} (u_j - R.i_j)dt \tag{2}$$

The magnetic field counterpart, W'_j, is determined by (3)

$$\partial \mathrm{W}'_j(\theta, i_j) = \int_0^{i_j} \psi_j(\theta, i_j)di_j \tag{3}$$

Which is a nonlinear function of current if the magnetic circuit is linear, the total torque T_e produced is equal to the sum of moments in the phases, as expressed in (4).

$$T_e(\theta, i_1, i_2, i_3, i_4) = \sum_{j=1}^{4} T_j(\theta, i_j) \tag{4}$$

To effectively control the switched reluctance motor, it is crucial to accurately determine the magnetic flux characteristic, denoted as $\psi_j(\theta,i_j)$. For ease of research and development of control algorithms, it is common to approximate the magnetic flux characteristic as a continuous function, as demonstrated in reference [31], which can be expressed as follows:

$$\psi_j(\theta, i_j) = \psi_s(1 - e^{-i_j f_j(\theta)}) \tag{5}$$

where j = 1, 2, 3, 4, ψ_s represents the saturation flux. The equation is also used in [32] and [33] for online parameter identification of the model and performance optimization in angle control. If we ignore the higher - order components in the Fourier series, we get the function $f_j(\theta)$ in (6)

$$f_j(\theta) = a + b\sin[N_r\theta - (j-1)\frac{2\pi}{n}] \tag{6}$$

where N_r is the number of rotor poles, n represents the number of phases, while a and b are coefficients obtained through the transformation of the Fourier series [34].

The moment of phase j is expressed as follows

$$T_j(\theta, i_j) = \frac{\psi_s}{f_j^2(\theta)}\frac{df_j(\theta)}{d\theta}\{1 - [1 + i_j f_j(\theta)]e^{-i_j f_j(\theta)}\} \tag{7}$$

To represent the SRM system in a mathematical model with state variables such as position, velocity, and current, we can derive the state space equations from equations (1) and (4). The state space equation of the switched reluctance motor includes the following equations, where ω represents the rotor velocity

$$\begin{cases} \dfrac{d\theta}{dt} = \omega \\ \dfrac{d\omega}{dt} = \dfrac{1}{J}\left\{\sum_{j=1}^{4} T_j(\theta, i_j) - T_l(\theta, \omega)\right\} \\ \dfrac{di_j}{dt} = -\left(\dfrac{\partial \psi_j}{\partial i_j}\right)^{-1}\left(Ri_j + \dfrac{\partial \psi_j}{\partial \theta}\omega\right) + \left(\dfrac{\partial \psi_j}{\partial i_j}\right)^{-1} u_j \end{cases} \tag{8}$$

The state model of the switched reluctance motor drive system is presented below based on [25]. Considering an 8/6 switched reluctance motor with 4 phases the state vector is defined as follows $x = [\theta, \omega, i_1, i_2, i_3, i_4]^T = [x_1, x_2, x_3, x_4, x_5, x_6]^T$. The equation of the motor's state as follows.

$$\dot{x}_1 = x_2 \tag{9}$$

$$\dot{x}_2 = \frac{1}{J}\left[T_1(\theta, x_3) + T_2(\theta, x_4) + T_3(\theta, x_5) + T_4(\theta, x_6) - T_l(x_1, x_2)\right]$$
$$= \frac{1}{J}\begin{bmatrix} \dfrac{\psi_s}{f_1^2(x_1)}\dfrac{\partial f_1(x_1)}{\partial x_1}\left\{1 - [1 + x_3 f_1(x_1)]e^{-x_3 f_1(x_1)}\right\} + \\ \dfrac{\psi_s}{f_2^2(x_1)}\dfrac{\partial f_2(x_1)}{\partial x_1}\left\{1 - [1 + x_4 f_2(x_1)]e^{-x_4 f_2(x_1)}\right\} + \\ \dfrac{\psi_s}{f_3^2(x_1)}\dfrac{\partial f_3(x_1)}{\partial x_1}\left\{1 - [1 + x_5 f_3(x_1)]e^{-x_5 f_3(x_1)}\right\} + \\ \dfrac{\psi_s}{f_4^2(x_1)}\dfrac{\partial f_4(x_1)}{\partial x_1}\left\{1 - [1 + x_6 f_4(x_1)]e^{-x_6 f_4(x_1)}\right\} \\ -Bx_2 - mgl\sin(x_1) \end{bmatrix} \tag{10}$$

and $\dot{x}_3$, $\dot{x}_4$, $\dot{x}_5$, $\dot{x}_6$ are

$$\dot{x}_{j+2} = \left[\psi_s e^{-x_{j+2} f_j(x_1)} f_j(x_1)\right]^{-1} u_j + \left[-\psi_s e^{-x_{j+2} f_j(x_1)} f_j(x_1)\right]^{-1} \left[\left(\psi_s e^{-x_{j+2} f_j(x_1)}\right)\left(x_{j+2}\frac{\partial f_j(x_1)}{\partial x_1}\right)x_2 + Rx_{j+2}\right] \tag{11}$$

$$\frac{\partial f_j}{\partial x_1} = bN_r\cos\left(N_r x_1 - (j-1)\frac{2\pi}{4}\right) \qquad j = 1,2,3,4 \tag{12}$$

It is mentioned that in the state-space description provided above, the term Bx_2 represents the damping effect that opposes the rotational motion of the machine, while $mglsin(x_1)$ corresponds to the mechanical load torque, for instance in the case that the SRM lifts a rod of length l with a mass m attached to its end [25].

From (10) we put

$$g_j(\boldsymbol{x}) = \frac{1}{J}\left[\frac{\psi_s}{f_j^2(x_1)}\frac{\partial f_j(x_1)}{\partial x_1}\left\{1 - e^{-x_{j+2} f_j(x_1)}\right\}\right]$$
$$h_j(\boldsymbol{x}) = \frac{1}{J}\left[\frac{\psi_s}{f_j^2(x_1)}\frac{\partial f_j(x_1)}{\partial x_1}\left\{-f_j(x_1)e^{-x_{j+2} f_j(x_1)}\right\}\right] \tag{13}$$

where j = 1, 2, 3, 4

Equation (10) can be rewritten as

$$\dot{x}_2 = \sum_{j=1}^{4}\left[g_j(\boldsymbol{x}) + h_j(\boldsymbol{x})x_{j+2}\right] - \frac{B}{J}x_2 - \frac{mgl}{J}\sin(x_1) \tag{14}$$

Differentiating equation (14) with respect to time, we get

$$\ddot{x}_2 = \sum_{j=1}^{4}\left[\dot{g}_j(\boldsymbol{x}) + \dot{h}_j(\boldsymbol{x})x_{j+2} + h_j(\boldsymbol{x})\dot{x}_{j+2}\right] - \frac{B}{J}\dot{x}_2 - \frac{mgl}{J}\cos(x_1)\dot{x}_1 \tag{15}$$

From equation (11), we set:

$$p_j(\boldsymbol{x}) = \left[-\psi_s e^{-x_{j+2} f_j(x_1)} f_j(x_1)\right]^{-1}\begin{bmatrix} Rx_{j+2} + \left(\psi_s e^{-x_{j+2} f_j(x_1)}\right) \\ \left(x_{j+2}\dfrac{\partial f_j(x_1)}{\partial x_1}\right)x_2 \end{bmatrix}$$
$$q_j(\boldsymbol{x}) = \left[\psi_s e^{-x_{j+2} f_j(x_1)} f_j(x_1)\right]^{-1} \tag{16}$$

We can rewrite equations (11) as follows

$$\dot{x}_{j+2} = p_j(\boldsymbol{x}) + q_j(\boldsymbol{x})u_j \qquad j = 1,2,3,4 \tag{17}$$

Replace (17) and (16) into (15), we have

$$\ddot{x}_2 = \sum_{j=1}^{4}\left[\dot{g}_j(\boldsymbol{x}) + \dot{h}_j(\boldsymbol{x})x_{j+2} + h_j(\boldsymbol{x})p_j(\boldsymbol{x}) + h_j(\boldsymbol{x})q_j(\boldsymbol{x})u_j\right] - \frac{B}{J}\dot{x}_2 - \frac{mgl}{J}\cos(x_1)\dot{x}_1 \tag{18}$$

The switched reluctance motor operates on the principle of supplying voltage to each phase. For an SRM with a pole configuration of 8/6 and a phase number of 4, we can derive the following expression for each phase *j*, where *j* can take values 1, 2, 3, or 4.

$$u_j = k_j.u \tag{19}$$

Where k_j represents the phase transition key, it is a variable that can only have the values of 0 or 1. Equation (18) can be reformulatedas follows

$$\ddot{x}_2 = \sum_{j=1}^{4}\left[\dot{g}_j(\boldsymbol{x}) + \dot{h}_j(\boldsymbol{x})x_{j+2} + h_j(\boldsymbol{x})p_j(\boldsymbol{x})\right] + \sum_{j=1}^{4}\left[h_j(\boldsymbol{x})q_j(\boldsymbol{x})k_j\right]u - \frac{B}{J}\dot{x}_2 - \frac{mgl}{J}\cos(x_1)\dot{x}_1 \tag{20}$$

Set

$$F(\boldsymbol{x}) = \sum_{j=1}^{4}\left[\dot{g}_j(\boldsymbol{x}) + \dot{h}_j(\boldsymbol{x})x_{j+2} + h_j(\boldsymbol{x})p_j(\boldsymbol{x})\right]$$
$$G(\boldsymbol{x}) = \sum_{j=1}^{4}\left[h_j(\boldsymbol{x})q_j(\boldsymbol{x})k_j\right] \tag{21}$$

We can express equation (20) in a different form as follows

$$\ddot{x}_2 = F(\boldsymbol{x}) + G(\boldsymbol{x})u - \frac{B}{J}\dot{x}_2 - \frac{mgl}{J}\cos(x_1)\dot{x}_1 \tag{22}$$

Set

$$f(\boldsymbol{x}) = \mathrm{F}(\boldsymbol{x}) - \frac{B}{J}\dot{x}_2 - \frac{mgl}{J}\cos(x_1)\dot{x}_1$$
$$g(\boldsymbol{x}) = G(\boldsymbol{x}) \tag{23}$$

We have

$$\ddot{x}_2 = f(\boldsymbol{x}) + g(\boldsymbol{x})u \tag{24}$$

For backstepping to be applied, one must rewrite (23) using a strict feedback form as follows

$$\begin{cases}\dot{z}_1 = z_2 \\ \dot{z}_2 = f(\boldsymbol{x}) + g(\boldsymbol{x})u\end{cases} \tag{25}$$

with *f*(*x*), *g*(*x*) are defined in (23).

Due to the challenges associated with synthesizing a stable speed controller for the switched reluctance motor (SRM), a design method that combines the backstepping technique with fuzzy adaptive sliding control is considered suitable. This is because the nonlinear state model of the SRM, represented by equation (25) in the form of 2nd order tight backpropagation, is highly susceptible to external noise. By combining these techniques, it is possible to effectively address these challenges and improve the overall performance of the SRM speed controller.

3. BACKSTEPPING SLIDING ADAPTIVE CONTROLLER BASED ON FUZZY LOGIC SYSTEM

3.1 SYNTHESIS OF BACKSTEPPING SLIDING MODE CONTROLLER FOR SRM

By utilizing the backstepping and sliding technique, the controller is designed for the nonlinear model (equation 25) as follows

Step 1: Put

$$e_1 = z_1 - z_{1d} \tag{26}$$

where Z_{1d} represents the setpoint of speed. Differentiating equation (26) with respect to time, we get

$$\dot{e}_1 = \dot{z}_1 - \dot{z}_{1d} = \dot{z}_1 - \dot{z}_{1d} \tag{27}$$

Put

$$e_2 = z_2 - \alpha \tag{28}$$

where α is the virtual control signal then we have

$$\dot{e}_1 = \dot{z}_1 - \dot{z}_{1d} = e_2 + \alpha - \dot{z}_{1d} \tag{29}$$

To obtain $e_1 \to 0$, we consider the Lyapunov function candidate of e_1 as follows

$$V_1 = \frac{1}{2}e_1^2 \tag{30}$$

Differentiating equation (30) with respect to time, we have

$$\dot{V}_1 = e_1\dot{e}_1 = e_1\left(e_2 + \alpha - \dot{z}_{1d}\right) \tag{31}$$

To have $\dot{V}_1 = -c_1 e_1^2 + e_1 e_2$ with $c_1 > 0$, the virtual control signal is

$$\alpha = -c_1 e_1 + \dot{z}_{1d} \tag{32}$$

Step 2: The sliding surface is defined as follows:

$$S = \mu e_1 + e_2; \mu > 0 \tag{33}$$

To ensure a stable closed system and tracking error of zero, we determine the sliding control signal *u*(*t*) by defining a Lyapunov function for the closed system, as shown in equation (34).

$$V = V_1 + \frac{1}{2}S^2 \tag{34}$$

Differentiating equation (34) with respect to time, we have

$$\begin{aligned}\dot{V} &= -c_1 e_1^2 + e_1 e_2 + S\dot{S} = -c_1 e_1^2 + e_1 e_2 + S\left(\mu\dot{e}_1 + \dot{e}_2\right) \\ &= -c_1 e_1^2 + e_1 e_2 + S\left(\mu\dot{e}_1 + f(\boldsymbol{x}) + g(\boldsymbol{x})u - \dot{\alpha}\right) \\ &= -c_1 e_1^2 - c_2 e_2^2 - KS\,\mathrm{sgn}(S);\ K \geq 0\end{aligned} \tag{35}$$

if

$$e_1 e_2 + c_2 e_2^2 + S\left(K\,\mathrm{sgn}(S) + \mu\dot{e}_1 + f(\boldsymbol{x}) + g(\boldsymbol{x})u - \dot{\alpha}\right) = 0 \tag{36}$$

Therefore, we select the control signal as depicted in equation (37)

$$u = -\frac{e_2\left(e_1 + c_2 e_2\right)}{Sg(\boldsymbol{x})} - \frac{K\,\mathrm{sgn}(S) + \mu\dot{e}_1 + f(\boldsymbol{x}) - \dot{\alpha}}{g(\boldsymbol{x})} \quad (37)$$

Theorem: The proposed controller (37) guarantees asymptotical stability of SRM system with the nonlinear state model (25).

Proof: Choose a Lyapunov function for a closed system of the following form:

$$V = V_1 + \frac{1}{2}S^2 \quad (38)$$

Derivative V with respect to time, we have

$$\begin{aligned}\dot{V} &= \dot{V}_1 + S\dot{S} = -c_1 e_1^2 + e_1 e_2 + S\left(\mu\dot{e}_1 + \dot{e}_2\right) \\ &= -c_1 e_1^2 + e_1 e_2 + S\left(\mu\dot{e}_1 + f\left(\boldsymbol{x}\right) + g\left(\boldsymbol{x}\right)u - \dot{\alpha}\right)\end{aligned} \quad (39)$$

Replace u in (37) into (39), we have:

$$\dot{V} = -c_1 e_1^2 - c_2 e_2^2 - KS\,\mathrm{sgn}(S) \le 0 \quad (40)$$

Therefore, the SRM system is asymptotically stable.

3.2. SYNTHESIS OF BACKSTEPPING ADAPTIVE CONTROLLER FOR SRM BASED ON FUZZY LOGIC SYSTEM

One of the disadvantages of sliding control is the phenomenon known as "chattering," which refers to the shaking of the system when it is close to the working point due to high-frequency sign changes in the control signal function. To overcome this drawback, the paper proposes the addition of a fuzzy logic system to adjust the gain of the sliding control component. The fuzzy logic system used in this paper is based on the Sugeno model [35, 36].

The input to the fuzzy logic system consists of the speed state and angular acceleration of the reluctance motor, while the output is the gain factor.

Each input language variable contains three triangular fuzzy sets with names corresponding to the digitized fuzzy sets [-1 0 1] with values [-10 0 10] belongs to the real number line R, respectively (as shown in Fig. 1). The notation [-1 0 1] represents the linguistic terms [small - zero - big]. One the other hand, the output variables are represented by constants with a digitized name of [-2 -1 0 1 2], which are interpreted as [very small - small - zero - big - very big]. The corresponding real values on the R scale are [2 1 0.01 1 2], as shown in Fig. 2. Furthermore, Table 1 provides the fuzzy tuning rules for the parameter K.

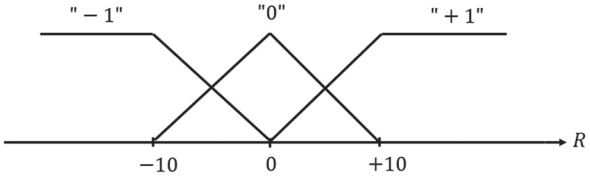


Fig. 1. Fuzzy set of input language variables e_1, $e^{\cdot}_1$

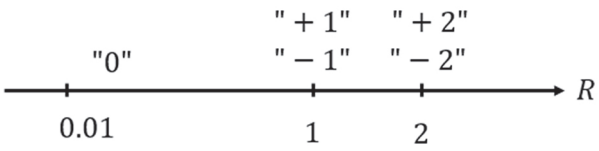


Fig. 2. Constant value for output variable

Table 1. Fuzzy tuning rules for the parameter K

K		e_1		
		-1	0	1
$e^{\cdot}_1$	0	-1	-2	
	0	1	0	-1
	-1	2	1	0

3.3. CONTROLLER STRUCTURE DIAGRAM

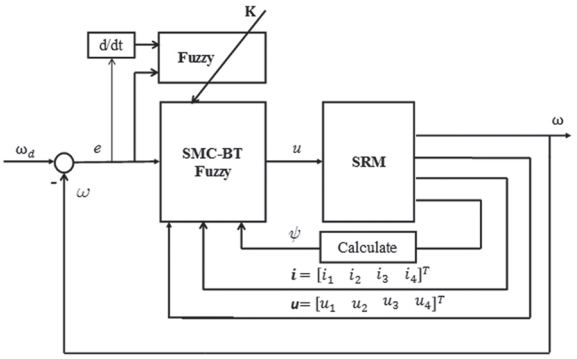


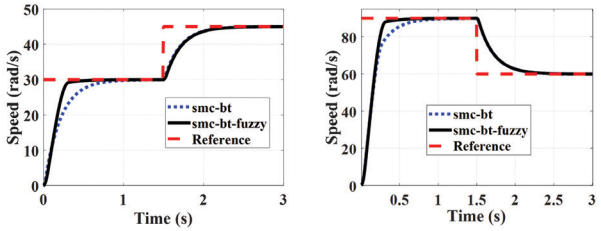
Fig. 3. The proposed control structure diagram for the SRM

Fig. 3 illustrates the structure of the adaptive backstepping sliding mode control algorithm based on a fuzzy logic system for the SRM, as described in sections 3.1 and 3.2. The backstepping sliding mode controller's gain K is adjusted by a fuzzy logic controller to address the phenomenon of "chattering" caused by high-frequency sign changes in the control signal function. It is assumed that the state variables of the SRM are directly observable for the implementation of the control.

4. THE SIMULATION RESULTS

This section compares the performance of the system using the backstepping adaptive controller based on fuzzy logic (smc-bt-fuzzy) with the system using the backstepping sliding mode controller (smc-bt) under different scenarios. Additionally, the results obtained in [25] are also compared. The parameters of the SRM are obtained from [23]. Figs. 4a and 4b illustrate the response of the system when the setpoint of speed changes from 30 rad/s to 45 rad/s and from 90 rad/s to 60 rad/s, respectively. The results indicate that both controllers, smc-bt-fuzzy and smc-bt, effectively track the setpoint of the SRM speed with zero overshoot and zero steady-state error. They also exhibit similar control quality in the case of a second setpoint where the speed reference value changes minimally. The specific

control qualities are presented in Table 2, which show that the proposed controller can achieve a controlled variable that reaches 90% and 95% of the reference speed within approximately 0.1 s and 0.18 s, respectively. This performance is faster compared to the smc-bt controller.



(a) (b)

Fig. 4. System speed response to variable setpoint of speed

Table 2. Control quality of systems when changing the setpoint of speed

Setpoint changes	In case of the fist setpoint		In case of the second setpoint	
Controller	smc-bt-fuzzy	smc-bt	smc-bt-fuzzy	smc-bt
Over shoot	0%	0%	0%	0%
Settling time	0.18s	0.38s	0.57s	0.57s
Steady-state error	0	0	0	0
Rise time	0.1s	0.5s	0.3s	0.3s

The faster rise time achieved by the smc-bt-fuzzy controller suggests improved dynamic response and quicker attainment of the desired speed compared to the smc-bt controller. Notably, these results are superior to those obtained in [25], which utilized a nonlinear H-infinity controller that produced an overshoot of approximately 20% and a settling time of approximately 7 s.

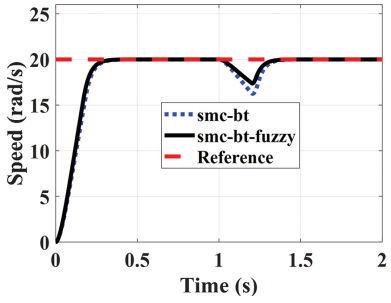
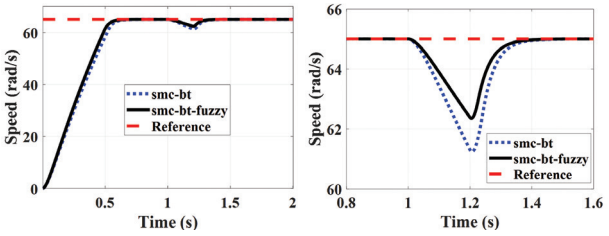


Fig. 5. Speed response at 20 rad/s with increasing load

Next, the system's response is analyzed in the presence of load disturbances. Specifically, Figs. 5 and 6 illustrate the scenario where the load increases while the system operates at speeds of 20 and 65 rad/s, respectively. Conversely, Figs. 7 and 8 represent the case where the system operates at speeds of 22 and 82 rad/s, respectively, while undergoing a sudden decrease in load.



(a) (b)

Fig. 6. Speed response at 65 rad/s with increasing load

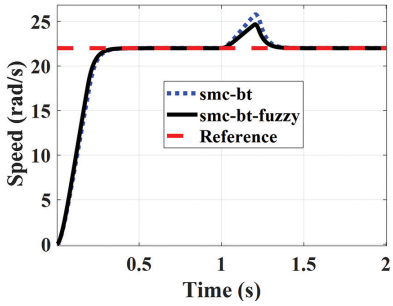
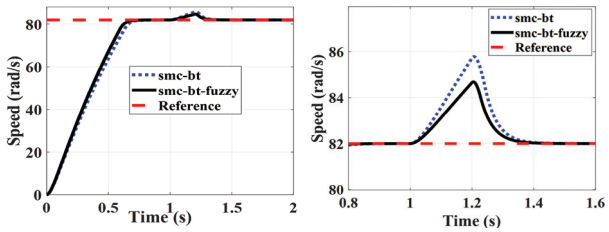


Fig. 7. Speed response at 22 rad/s under load reduction



(a) (b)

Fig. 8. Speed response at 82 rad/s under load reduction

Fig. 6b is an enlarged image of Fig. 6a, providing a clearer assessment of the control quality of the two systems during the load increase process. Similarly, Fig. 8b is an enlarged image of Fig. 8a, aiding in the evaluation of the control quality during the load reduction process.

Table 3. Control quality of two systems under load change

Load changes	In case of increased load		In case of reduced load	
Controller	smc-bt-fuzzy	smc-bt	smc-bt-fuzzy	smc-bt
Over shoot	4%	6%	2.7%	5%
Settling time	0.35s	0.37s	0.3s	0.35s
Steady-state error	0	0	0	0

The results in Fig. 6b, Fig. 8b and Table 3 indicate that the backstepping adaptive controller based on fuzzy logic helps the system return to the desired positions with zero steady-state error. Additionally, it exhibits comparable setting times but achieves a reduction in overshoot by 30% to 50% compared to the backstepping sliding mode controller.

5. CONCLUSION

In this paper, we present the application of the adaptive backstepping sliding mode control algorithm based on fuzzy logic for controlling the speed of an SRM drive system. This approach considers the inherent nonlinearity introduced by the inverter to improve the overall performance of the SRM drive system. The simulation results, which consider changes in the setpoint of speed and variable load, demonstrate the excellent performance of the proposed controller in comparison to the backstepping sliding mode controller and the nonlinear H-infinity controller mentioned in [25]. These results indicate considerable potential for the future development and application of novel algorithms in SRM systems. Further research to enhance the overall performance through torque control and experimental testing on real hardware will be conducted by the authors in the future.

Appendix A. SRM and simulatuon parameters

Number of rotor poles 6	J=6.8x103 kg/m2
Number of stator poles 8	a=1.5x103 H
Number of phases 4	b=1.364x103 H
Power 5.5 HP	B=0.2
Peak current 9A	l=2 m
Stator winding resistance 0.72 Ω	c1=2
Aligned phase inductance 130 mH	c2=0.1
Unaligned phase inductance 12 mH	T=0.025

6. REFERENCES

[1] D. F. Valencia, R. Tarvirdilu-Asl, C. Garcia, J. Rodriguez, A. Emadi, "Vision, Challenges, and Future trends of model predictive control in switched reluctance motor drives", IEEE Acces, Vol. 9, 2021, pp. 69926-69937.

[2] P. H. Nha, D. Q. Thuy, "Improving the characteristics of switched reluctance motor", Automatic Control and System Engineering Journal, Vol. 16, No. 2, 2016, pp. 59-66.

[3] Y. Lan et al. "Switched reluctance motors and drive systems for electric vehicle power trains: state of the art analysis and future trends", Energies, Vol. 14, No. 8, 2021, pp. 1-29.

[4] P. Azer, B. Bilgin, A. Emad, "Mutually coupled switched reluctance motor: Fundamentals, control, modeling, state of the Art review and future trends", IEEE Access, Vol. 7, 2019, pp. 100099-100112.

[5] J. Song, S. Song, B. Qiu, "Application of an adaptive PI controller for a switched reluctance motor drive", Proceedings of the 2nd IEEE Annual Conference on Power Electronics, Auckland, New Zealand, 5-8 December 2016, pp. 1-5.

[6] B. Fabianski, "Optimal control of switched reluctance motor drive with use of simplified nonlinear reference model", Proceedings of the 7th IEEE International Conference on Mechatronics - Mechatronika, Prague, Czech Republic, 7-9 December 2016.

[7] H. N. Huang, K. W. Hu, Y. W. Wu, T. L. Jang, C. M. Liaw, "A current control scheme with back EMF cancellation and tracking error adapted communication shift for switched reluctance motor drive", IEEE Transactions on Industrial Electronics, Vol. 69, No. 2, 2016, pp. 7381-7392.

[8] D. Ronanki, S. S. Williamson, "Comparative Analysis of DITC and DTFC of Switched Reluctance Motor for EV Applications", Proceedings of the IEEE ICIT International Conference on Industrial Technology, Ontario, Canada, March 2017, pp. 509-514.

[9] M. Bychkov, A. Fedorenko, A. Krasovsky, E. Gorbanova, "Torque control of switched reluctance drive in generating mode", Proceedings of the 25th IEEE International Workshop on Electric Drives: Optimization in Control of Electric Drives, Moscow, Russia, 31 January - 2 February 2018.

[10] A. Berdai et al. "Similarity and Comparison of the Electrodynamics Characteristics of Switched Reluctance Motors SRM with Those of Series DC Motors", Engineering, Vol. 7, No. 1, 2015, pp. 36-45.

[11] A. Nirgude, M. Murali, N. Chaithanya, S. Kulkarni, V. B. Bhole, S. R. Patel, "Nonlinear mathematical modeling and simulation of switched reluctance motor", Proceedings of the IEEE International Conference on Power Electronics, Drives and Energy Systems, Trivandrum, India, 14-17 December 2016, pp. 1-6.

[12] J. A. Makwana, P. Agarwal, S. P. Srivastava, "Modeling and Simulation of Switched Reluctance Motor", Lecture Notes in Electrical Engineering, Vol. 442, 2018, Springer, pp. 545-558.

[13] X. Sun, K. Diao, Z. Yang, G. Lei, Y. Guo, J. Zhu, "Direct Torque Control Based on a Fast Modeling Method for a Segmented-Rotor Switched Reluctance Motor in HEV Application", IEEE Journal of Emerging and Selected Topics in Power Electronics, Vol. 9, No. 1, 2019, pp. 232-241.

[14] H. Abdelmaksoud, M. Zaky, "Design of an adaptive flux observer for sensorless switched reluctance motors using lyapunov theory", Advances in Electrical and Computer Engineering, Vol. 20, No. 2, 2020, pp. 123-130.

[15] O. Ustun, "A nonlinear full model of switched reluctance motor with artificial neural network", Energy Conversion and Management, Vol. 50, 2018, pp. 2413-2421.

[16] S. Wang, Z. Hu, X. Cui, "Research on novel direct instantaneous torque control strategy for switched reluctance motor", IEEE Access, Vol. 8, 2020, pp. 66910-66916.

[17] G. Fang et al. "Advance control of switched reluctance motors: a review on current regulation, torque control and vibration suppression", IEEE Open Journal of the Industrial Electronics Society, Vol. 2, 2021, pp. 280-301.

[18] X. Shao, F. Naghdy, H. Du, Y. Qin, "Coupling effect between road excitation and an in-wheel switched reluctance motor on vehicle ride comfort and active suspension control", Journal of Sound and Vibration, Vol. 443, 2019, pp. 683-702.

[19] H. E. A. Ibrahim, M. S. S. Ahmed, K. M. Awad, "Speed control of switched reluctance motor using genetic algorithm and ant colony based on optimizing PID controller", Proceedings of the International Conference on Applied Mathematics, Computational Science and Systems Engineering, Athens, Greece, 28-30 December 2018.

[20] Y. Nakazwa, S. Matsunaga, "Position sensorless control of switched reluctance motor using state observer", Proceedings of the 22nd International conference on electrical machines and systems, Harbin, China, 11-14 August 2019, pp. 20-23.

[21] J. Sun et al. "Sliding mode observer based position estimation for sensorless control of the planar switched reluctance motor", IEEE Access, Vol. 7, 2019, pp. 61034-61045.

[22] C. Shang, A. Xu, L. Huang, J. Chen, "Flux Linkage Optimization for direct torque control of switched reluctance motor based on model predictive control", IEĘ Transaction on Electrical and Electronic Engineering, Vol. 14, No. 7, 2019, pp. 1105-1113.

[23] R. Ortega, A. Sarr, A. Bobtsov, I. Bahri, D. Diallo, "Adaptive state observer for sensorless control switched reluctance motors", International Journal of Robust and Nonlinear control, Vol. 29, No. 4, 2019, pp. 990-1006.

[24] M. Divandari, B. Rezaie, A. R. Noei, "Speed control of switched reluctance motor via fuzzy fast terminal sliding mode control", Computers and Electrical Engineering, Vol. 80, 2019, pp. 1-16.

[25] G. Rigatos, P. Siano, S. Ademi, "Nonlinear H-infinity control for switched reluctance machines", Nonlinear Engineering, Vol. 9, 2019, pp. 14-27.

[26] P. H. Nha, P. H. Phi, D. Q. Thuy, P. X. Dat, L. X. Hai, "Backstepping control using nonlinear state observer for Switched reluctance motor", Vietnam Journal of Science and Technology, Vol. 60, No. 3, 2022, pp. 554-568.

[27] M. T. Alrifai, J. H. Chow, D. A. Torrey, "Backstepping nonlinear speed controller for switched-reluctance motors", IEE Proceedings - Electric Power Applications, Vol. 150, No. 2, 2003, pp. 193-200.

[28] C. H. Lin, "Adaptive nonlinear backstepping control using mended recurrent Romanovski polynomials neural network and mended particle swarm optimization for switched reluctance motor drive system", Transactions of the Institute of Measurement and Control, Vol. 41, No. 14, 2019, pp. 4114-4128.

[29] C. H. Lin, J. C. Ting, "Novel nonlinear backstepping control of Synchronous reluctance motor drive system for postion tracking of periodic reference input torque ripple consideration", International Journal of Control, Automation and Systems, Vol. 17, No. 1, 2019, pp. 1-17.

[30] N. P. Hoang, H. P. Van, H. L. Xuan, "Backstepping Sliding Mode Controller of Switched Reluctance Motors with Combined Nonlinear Model", SSRG International Journal of Electrical and Electronic Engineering, Vol. 10, No. 8, 2023, pp. 235-242.

[31] M. Ilic-Spong, R. Marino, S. M. Peresada, D. G. Taylor, "Feedback Linearizing Control of Switched Reluctance Motors", IEEE Transactions on Automatic Control, Vol. 32, No. 5, 1987, pp. 371-379.

[32] S. Mir, I. Husain, M. E. Elbuluk, "Switched reluctance motor modeling with on-line parameter

identification", IEEE Transactions on Industry Applications, Vol. 34, No. 4, 1998, pp. 776-783.

[33] C. Mademlis, I. Kioskeridis, "Performance optimization in switched reluctance motor drives with online commutation angle control", IEEE Transactions on Energy Conversion, Vol. 18, No. 3, 2003, pp. 448-457.

[34] H. Hannoun, M. Hilairet, C. Marchand, "High performance current control of a switched reluctance machine based on a gain-scheduling PI controller", Control Engineering Practice, Vol. 19, 2011, pp. 1377-1386.

[35] P. P. Angelov, D. P. Filev, "An approach to online identification of Takagi-Sugeno fuzzy models", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 34, No. 1, 2004, pp. 484-498.

[36] M. Bernal, T. M. Guerra, "Generalized Nonquadratic Stability of Continuous-Time Takagi–Sugeno Models", IEEE Transactions on Fuzzy Systems, Vol. 18, No. 4, 2010, pp. 815-822.

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia.

About this Journal

The International Journal of Electrical and Computer Engineering Systems publishes original research in the form of full papers, case studies, reviews and surveys. It covers theory and application of electrical and computer engineering, synergy of computer systems and computational methods with electrical and electronic systems, as well as interdisciplinary research.

Topics of interest include, but are not limited to:

- Power systems
- Renewable electricity production
- Power electronics
- Electrical drives
- Industrial electronics
- Communication systems
- Advanced modulation techniques
- RFID devices and systems
- Signal and data processing
- Image processing
- Multimedia systems
- Microelectronics
- Instrumentation and measurement
- Control systems
- Robotics
- Modeling and simulation
- Modern computer architectures
- Computer networks
- Embedded systems
- High-performance computing
- Parallel and distributed computer systems
- Human-computer systems
- Intelligent systems
- Multi-agent and holonic systems
- Real-time systems
- Software engineering
- Internet and web applications and systems
- Applications of computer systems in engineering and related disciplines
- Mathematical models of engineering systems
- Engineering management
- Engineering education

Paper Submission

Authors are invited to submit original, unpublished research papers that are not being considered by another journal or any other publisher. Manuscripts must be submitted in doc, docx, rtf or pdf format, and limited to 30 one-column double-spaced pages. All figures and tables must be cited and placed in the body of the paper. Provide contact information of all authors and designate the corresponding author who should submit the manuscript to https://ijeces.ferit.hr. The corresponding author is responsible for ensuring that the article's publication has been approved by all coauthors and by the institutions of the authors if required. All enquiries concerning the publication of accepted papers should be sent to ijeces@ferit.hr.

The following information should be included in the submission:

- paper title;
- full name of each author;
- full institutional mailing addresses;
- e-mail addresses of each author;
- abstract (should be self-contained and not exceed 150 words). Introduction should have no subheadings;
- manuscript should contain one to five alphabetically ordered keywords;
- all abbreviations used in the manuscript should be explained by first appearance;
- all acknowledgments should be included at the end of the paper:
- authors are responsible for ensuring that the information in each reference is complete and accurate. All references must be numbered consecutively and citations of references in text should be identified using numbers in square brackets. All references should be cited within the text;
- each figure should be integrated in the text and cited in a consecutive order. Upon acceptance of the paper, each figure should be of high quality in one of the following formats: EPS, WMF, BMP and TIFF;
- corrected proofs must be returned to the publisher within 7 days of receipt.

Peer Review

All manuscripts are subject to peer review and must meet academic standards. Submissions will be first considered by an editor-in-chief and if not rejected right away, then they will be reviewed by anonymous reviewers. The submitting author will be asked to provide the names of 5 proposed reviewers including their e-mail addresses. The proposed reviewers should be in the research field of the manuscript. They should not be affiliated to the same institution of the manuscript author(s) and should not have had any collaboration with any of the authors during the last 3 years.

Author Benefits

The corresponding author will be provided with a .pdf file of the article or alternatively one hardcopy of the journal free of charge.

Units of Measurement

Units of measurement should be presented simply and concisely using System International (SI) units.

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003

Published: semiannually

Copyright

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.

Postal Address

Faculty of Electrical Engineering,
Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b
31000 Osijek, Croatia

IJECES Copyright Transfer Form

(Please, read this carefully)

This form is intended for all accepted material submitted to the IJECES journal and must accompany any such material before publication.

TITLE OF ARTICLE (hereinafter referred to as "the Work"):

COMPLETE LIST OF AUTHORS:

The undersigned hereby assigns to the IJECES all rights under copyright that may exist in and to the above Work, and any revised or expanded works submitted to the IJECES by the undersigned based on the Work. The undersigned hereby warrants that the Work is original and that he/she is the author of the complete Work and all incorporated parts of the Work. Otherwise he/she warrants that necessary permissions have been obtained for those parts of works originating from other authors or publishers.

Authors retain all proprietary rights in any process or procedure described in the Work. Authors may reproduce or authorize others to reproduce the Work or derivative works for the author's personal use or for company use, provided that the source and the IJECES copyright notice are indicated, the copies are not used in any way that implies IJECES endorsement of a product or service of any author, and the copies themselves are not offered for sale. In the case of a Work performed under a special government contract or grant, the IJECES recognizes that the government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official government purposes only, if the contract/grant so requires. For all uses not covered previously, authors must ask for permission from the IJECES to reproduce or authorize the reproduction of the Work or material extracted from the Work. Although authors are permitted to re-use all or portions of the Work in other works, this excludes granting third-party requests for reprinting, republishing, or other types of re-use. The IJECES must handle all such third-party requests. The IJECES distributes its publication by various means and media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various collections, databases and other publications. The IJECES publisher requires that the consent of the first-named author be sought as a condition to granting reprint or republication rights to others or for permitting use of a Work for promotion or marketing purposes. If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IJECES publisher assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

Authors of IJECES journal articles and other material must ensure that their Work meets originality, authorship, author responsibilities and author misconduct requirements. It is the responsibility of the authors, not the IJECES publisher, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it.

- The undersigned represents that he/she has the authority to make and execute this assignment.
- For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
- The undersigned agrees to indemnify and hold harmless the IJECES publisher from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.

_____ _____

Author/Authorized Agent **Date**

CONTACT
International Journal of Electrical and Computer Engineering Systems (IJECES)
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Josip Juraj Strossmayer University of Osijek
Kneza Trpimira 2b
31000 Osijek, Croatia
Phone: +38531224600,
Fax: +38531224605,
e-mail: ijeces@ferit.hr