Dynamic Feature Optimization with Moth-Flame Optimization Algorithm for Intrusion Detection using Deep Learning

[1]*Bhawana Choudhary, M. Tech Scholar, Computer Science & Engineering Department, NRI Institute of Research and Technology, Bhopal*
[2]*DR P.K Sharma, Associate Professor & HOD, Computer Science & Engineering Department, NRI Institute of Research and Technology, Bhopal*

Abstract-The issue of network security has drawn increasing attention as the Internet has grown rapidly. An important area of study in network security is the detection of anomalous behaviour in networks. Intrusion Detection Systems (IDSs) are used to analyse network data and identify unusual network behaviour. IDSs can be classified into two main categories: signature-based and anomaly-based detection systems. Signature-based detection systems, such as Snort intrusion detection systems, create libraries of signatures for known malicious behaviours and compare network data against these signatures to detect intrusions. This paper proposes a feature optimization-based intrusion detection approach. For feature optimization, the Moth-Flame Optimization (MFO) algorithm is employed. The feature optimization algorithm reduces complex features and improves the detection process. For classification, a Convolutional Neural Network (CNN) is used, which enhances the system's detection capacity. The proposed algorithm was tested using MATLAB2018R software with the KDDCUP2003 dataset. It was compared with existing algorithms such as CNN and CNN-GUR. The performance analysis suggests that the proposed algorithm is more efficient than the existing algorithms.

Keywords: - IDS, MFO, Swarm Intelligence, Deep Learning, KDDCUP

INTRODUCTION

The increasing rate of internet traffic, comprising various types of data, includes both normal and abnormal traffic. Abnormal traffic often carries numerous threats and cyber-attack files, which can compromise public systems and networks. To prevent these threats, several methods are employed, including firewalls, intrusion detection systems (IDS), and antivirus software [1,2]. The concept of an IDS was first proposed by James P. Anderson in 1980 for the security of public and private networks. IDS can be categorized into three forms: Host-Based Intrusion Detection Systems (HIDS)[3], these systems monitor and analyze the internals of a computing system rather than network traffic. They can detect malicious activities by examining logs, file integrity, and system calls. Network-Based Intrusion Detection Systems (NIDS)[4]these systems monitor and analyze network traffic to identify suspicious activities. They operate at the network level and can detect threats that originate from outside the system. Hybrid Intrusion Detection Systems [5], these systems combine the features of both HIDS and NIDS, providing a more comprehensive security solution by monitoring both network traffic and system activities. By employing these IDS types, along with other security measures like firewalls and antivirus software, organizations can better protect their networks from cyber threats and ensure the integrity and confidentiality of their data [6,7]. Accurate intrusion detection is a very challenging task due to the complex features of internet traffic. Recently, several authors have employed feature optimization-based intrusion detection systems to address this issue. Feature optimization reduces the number of complex features, improving the feature mapping's search space during the classification process [8]. The classification process utilizes various machine learning and deep learning algorithms. Deep learning algorithms, in particular, increase the capacity for detection and enhance the performance of intrusion detection systems. This paper proposes a feature optimization-based deep learning algorithm for the detection of intrusions [9,10]. The proposed method employs the Moth Flame Optimization (MFO) algorithm and Convolutional Neural Network (CNN) algorithm for the classification and detection of intrusions. To reduce the dimensionality of the data and eliminate irrelevant or redundant features. This process simplifies the feature space, making it easier for the detection algorithm to process the data and identify

potential intrusions accurately. The MFO algorithm mimics the navigation strategy of moths in nature, which follow a logarithmic spiral path towards a light source. In the context of intrusion detection, MFO helps in identifying the most relevant features that contribute to accurate intrusion detection. CNNs are a class of deep learning algorithms particularly effective in recognizing patterns and structures in data. They are widely used in image processing and are now being adapted for analyzing network traffic data to detect anomalies and intrusions. By optimizing the feature set, the CNN can focus on the most relevant data, improving the accuracy of intrusion detection. The combination of MFO and CNN leads to more efficient processing and better performance in real-time intrusion detection scenarios. The proposed method can handle large volumes of internet traffic data, making it suitable for deployment in modern, high-traffic network environments. The integration of Moth Flame Optimization (MFO) and Convolutional Neural Networks (CNN) offers a robust solution for intrusion detection. By optimizing the feature set and employing advanced deep learning techniques, this approach aims to enhance the accuracy and efficiency of detecting intrusions in complex internet traffic. The rest of the article is organized as follows: Section II covers related work, Section III presents the proposed methodology for intrusion detection, Section IV provides an experimental analysis of the proposed algorithm, and Section V concludes the paper and discusses future directions.

II. RELATED WORK

The complex features of internet traffic data slow down the process of intrusion detection. The process of feature optimization reduces the features of internet traffic and improves the detection of intrusions. Recently, several authors have proposed swarm intelligence and deep learning-based algorithms for intrusion detection. In [1], security and privacy recommendations for e-health systems are discussed, focusing on integrating common medical sensors and constraints on network development. It addresses challenges in developing communication networks for healthcare 4.0 and IoMT, emphasizing data and algorithmic biases leading to healthcare disparities. Additionally, it examines limitations of 5G communications in healthcare 4.0. In [2], multivariate analysis for anomaly detection in Healthcare IoT is presented, highlighting security vulnerabilities and the potential exploitation of internet connections by hackers to steal patient information. In [3], cloud computing is characterized as a framework for resource availability, introducing a Modified Metaheuristics with Weighted Majority Voting Ensemble Deep Learning model. It discusses physical and functional diversity challenges in IoT Intrusion Detection Systems (IDS) mechanisms and issues with unrealistic use of attributes for IDS self-protection. In [4], an intrusion detection model using PSO, BA, and RF is proposed, addressing challenges in ML-based models for IIoT network security. It notes low processing ability due to energy constraints in IoT devices and challenges in data analytics due to heterogeneous data, which affect intrusion detection system performance. In [5], the utilization of neural networks with one hidden layer for classification is explored, addressing issues such as over-fitting during training and NP-hard problems in adapting neural networks for practical applications. In [6], an ensemble-based voting classifier for intrusion detection using the Ton-IoT dataset is introduced, highlighting shortcomings of existing intrusion detection systems and evaluating the proposed approach through statistical analysis. In [7], the development of hybridized AOA metaheuristics for multi-layer ANN training is focused on, emphasizing limitations in optimizing hyperparameters due to scope and focusing on MLP with a single hidden layer. In [8], kernel techniques for feature selection to enhance classifier performance are discussed, noting cloud computing limitations such as latency, connectivity, and mobility, and constraints in computational resources for edge devices. In [9], ML classifier accuracy is enhanced using the marine predator's optimization algorithm, optimizing learning parameters for efficient ECG feature classification. It discusses challenges specific to SVM with noisy data, RF with interpretability, and the importance of parameter tuning for ML efficiency. In [10], IoT-sensed COVID data from remote areas is gathered for analysis, reviewing access control models and highlighting challenges with encryption methods' time and cost consumption. In [11], machine learning methods for detecting middle box attacks in IoT networks are evaluated, addressing research gaps using the Bot-IoT dataset and inefficiencies in SSI detection methods due to network data volume and cyber-attack diversity. In [12], Bluetooth security in smart healthcare systems is focused on, presenting the Blue Tack dataset for Bluetooth

attacks in IoMT networks without specifying limitations. In [13], blockchain and FL-based IDS are integrated into IoT security, proposing a framework to enhance security in healthcare systems despite increased execution time complexity due to blockchain algorithms. In [14], machine learning for detecting attacks in water system infrastructure is proposed, outlining data pre-processing workflows and noting limitations such as the absence of hyperparameters in ML modeling and unexplored deep learning algorithms. In [15], a feature selection model based on distribution density functions is introduced, enhancing prediction of behavioral analysis through optimal feature selection despite challenges with unstructured data and existing behavioral intrusion models. In [16], an ML-based IDS for IoT to detect RPL network attacks is proposed, utilizing genetic recursive feature selection and a fuzzy k-nearest neighbor classifier to address vulnerabilities in RPL protocol and router resources. In [17], the BHS-ALOHDL technique with ALO-FSS, HDL classification, and FPA hyperparameter tuning is introduced to enhance security and privacy of medical data in IoT healthcare systems without specific limitations. In [18], security challenges in Cloud-IoT systems for healthcare applications are addressed, developing a Machine Learning-based IDS for Internet of Medical Things and highlighting gaps in real-world implementation and comparison with other IDS. In [19], MCAD for healthcare systems using machine learning in SDNs is proposed, achieving high F1-scores and throughput but noting gaps in addressing insider threats and protecting data. In [20], automated disease classification for accurate predictions in healthcare is focused on, utilizing federated learning for secure data handling despite challenges with cloud-based medical records and matrix aggregation. In [21], edge-based computing for patient data collection is designed, employing federated learning to retrain local ML models and highlighting challenges in privacy and data reduction in edge and fog computing models. In [22], a fog architecture strategy for unforeseen catastrophic events is presented, using machine learning to detect vulnerabilities in IoT systems for cyber defense while noting limited discussions on privacy concerns and ethical implications of IoT technologies in healthcare devices.

III. PROPOSED METHODOLOGY

This section describes the proposed algorithm for intrusion detection systems, which encapsulates the Moth Flame Optimization (MFO) and Convolutional Neural Network (CNN) algorithms. The MFO algorithm is employed for the feature optimization process. The primary goal of feature optimization is to reduce the number of unwanted or irrelevant features in internet traffic data. By optimizing the feature set, the algorithm can focus on the most significant features, improving the efficiency and accuracy of intrusion detection. Convolutional Neural Networks (CNNs), a class of deep learning algorithms, are used for the classification task. In this context, the CNN algorithm performs binary classification, categorizing internet traffic as either normal or abnormal. The optimized feature set obtained from the MFO algorithm is used as input for the CNN, which then processes this data to detect intrusions. The proposed algorithm combines the strengths of Moth Flame Optimization and Convolutional Neural Networks to create an efficient and accurate intrusion detection system. The MFO algorithm enhances the feature set by removing irrelevant data, and the CNN algorithm leverages this optimized feature set to perform precise binary classification of internet traffic, effectively identifying potential intrusions. The proposed model shown in figure1. The process of feature selection for the classification and detection by MFO algorithm. The description of MFO[16,17,18] algorithm describe here.

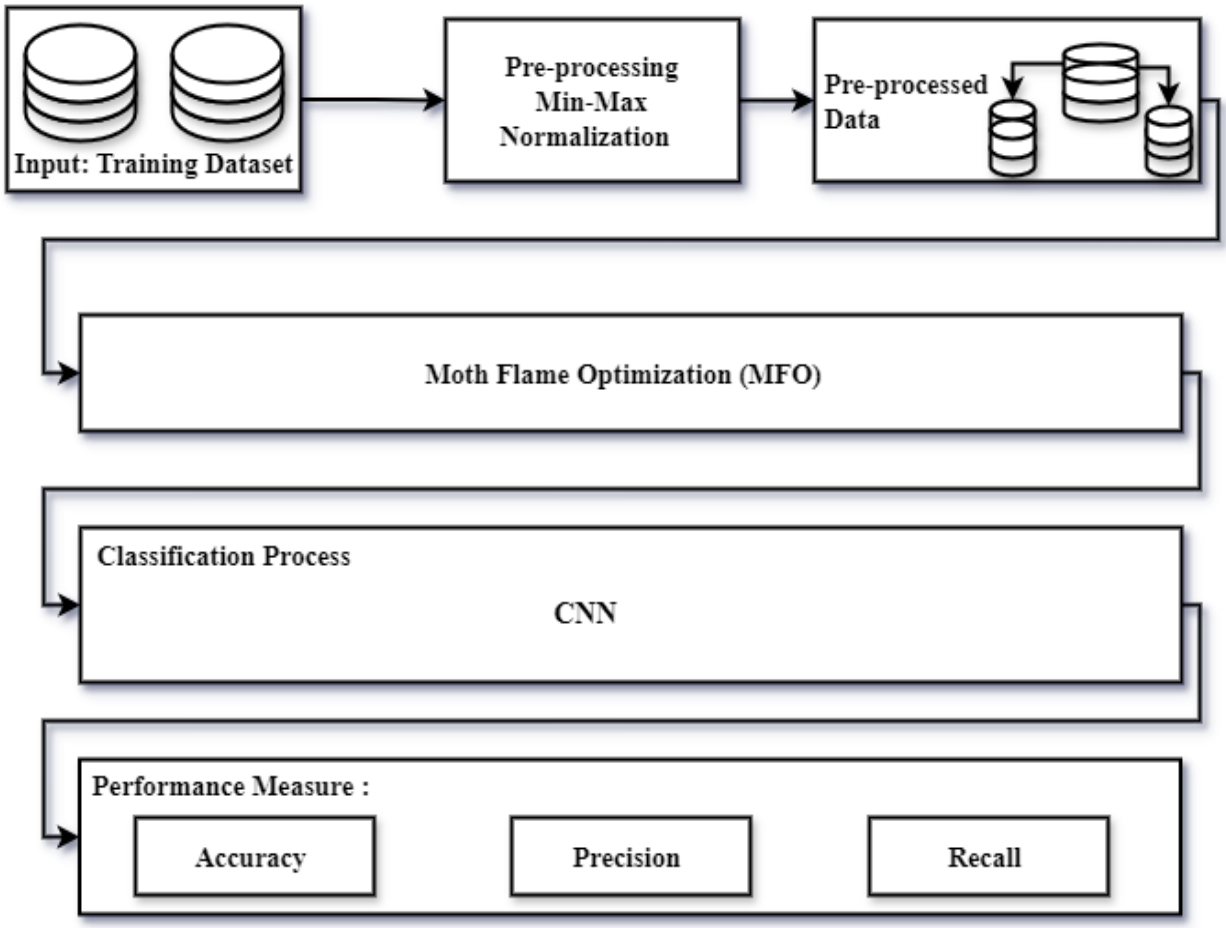

Figure 1 proposed model of intrusion detection systems

Moth-flame optimization algorithm is dynamic population based meta-heuristic function. The processing of algorithm describes here

The set of moths is defined as M, in which Mi is the i-th moth and Mij is the corresponding position of the i-th moth. Now OM define as fitness constraints

$$M = \begin{pmatrix} M \\ . \\ . \\ Mi \end{pmatrix} = \begin{pmatrix} M1 & & M1j \\ & . & \\ & . & \\ Mi1 \dots\dots\dots & & Mij \end{pmatrix}$$

The set of flame is defined as F, in which Fi is the i-th flame and Fij is the corresponding position of the i-th flame. Now OF is define as fitness constraints

$$F = \begin{pmatrix} F \\ . \\ . \\ Fi \end{pmatrix} = \begin{pmatrix} F1 & & F1j \\ & . & \\ & . & \\ Fi1 \dots\dots\dots & & Fij \end{pmatrix}$$

The algorithm describes the global optimal solution as

$$MFO = (I, P, T) \dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

$$I: \varphi \to \{M, OM\} \dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

$$P: M \to M \dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

$$T: M \to \{true, false\} \dots\dots\dots\dots\dots\dots\dots\dots\dots (4)$$

The processing of algorithm as

M=I()

While T(M) is equal to false

M=P(M);

End

Update the position of flames as

Mi=S(Mi,Fi)……..(5)

$$S(Mi, Fj) = D_{ie^{bt}} \cos(2\Pi t) + Fj \dots\dots\dots\dots\dots\dots\dots\dots\dots (6)$$

$$Di = |Fj - Mi| \dots\dots\dots\dots\dots\dots\dots\dots\dots (7)$$

The flame is updated as

$$flame\ no = round\left(N - L\frac{N-1}{T}\right) \dots\dots\dots\dots\dots\dots\dots\dots\dots (8)$$

Where N is number of initial flames, T is total number of iterations, and L is current number of the iterations.

Processing of MFO algorithm

1. Define value of M according to formula (2) and estimate OM as M
2. The position of M and OM is constant and F and OF can be found by matching sequence of M and OM
3. By formula (8) estimates the numbers of moth and the end moths' flames removed
4. The distance between moths is calculated by formula (7)
5. Update the value of moths according to formula (6)
6. By M estimate OM
7. Decide the end condition is met, otherwise go to step 2

The process of feature selection describes here
The feature set of intruders mapped as $(X_i \in R^D, y_i \in R)$,i=1…….m
Here Xi is feature set the range between 0 to 41, R is relation belongs to data.

$select\ X(f)$ if $* = X$

$[x^1, \ldots\ldots x^k] \leftarrow [rand(1,k) \times (p-w)] + 1$

$f \leftarrow n$ the optimal features of set

For $i \leftarrow 1\ to\ 41\ do$

$f \cdot \in X^D \leftarrow S *= X$

$41^i_* \in R^D \leftarrow moth\ baise$

$G \in R^D \leftarrow set\ of\ features$

End for

Input sample of features as $x^1_*, \ldots\ldots x^m_*$

$F_{MFO} \in R^{D \times x} \leftarrow \emptyset([x^1_*, \ldots\ldots x^m_*])$ $w = new\ feature\ set$

$W \in R^D \leftarrow x^{-1}$

$F \in R^{d\square,} \leftarrow W^T\ \emptyset(G)$

For $optimal \leftarrow 1\ to\ AX_c\ do$

IV. EXPERIMENTAL ANALYSIS

To analyzed the performance of proposed algorithm for intrusion detection system using MATLAB software. MATLAB is well-known data and algorithm analysis tools and supported various function of machine learning. The system configuration of machine for the simulation process windows 10 operating system, 16GB RAM and I7 processor. For the analysis of algorithm employed two reputed dataset of intrusion detection systems are KDDCUP2203. The empirical evaluation of results measure in form of precision, accuracy and recall[19,20,21].

$$Accuracy = \frac{Total\ No.of\ Correctly\ Classified\ Instances}{Total\ No.of\ Instances} \times 100$$

$$pricision = \frac{TP}{TP + FN} \times 100$$

$$Recall = \frac{TN}{TN + FP} \times 100$$

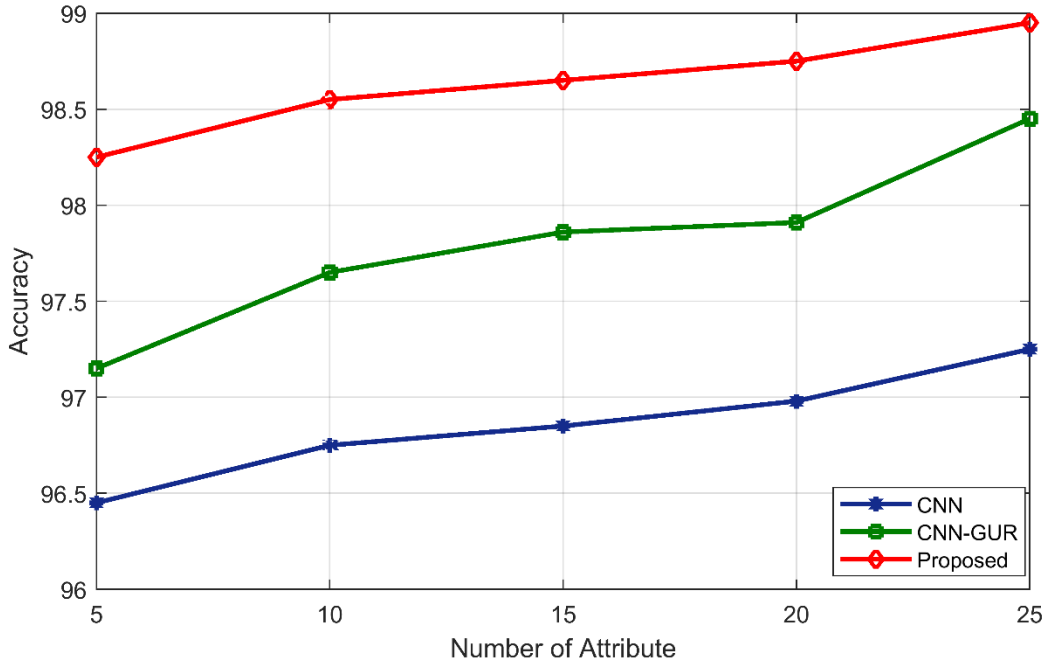

Fig 2: Performance analysis of accuracy of KDDCUP2003 dataset

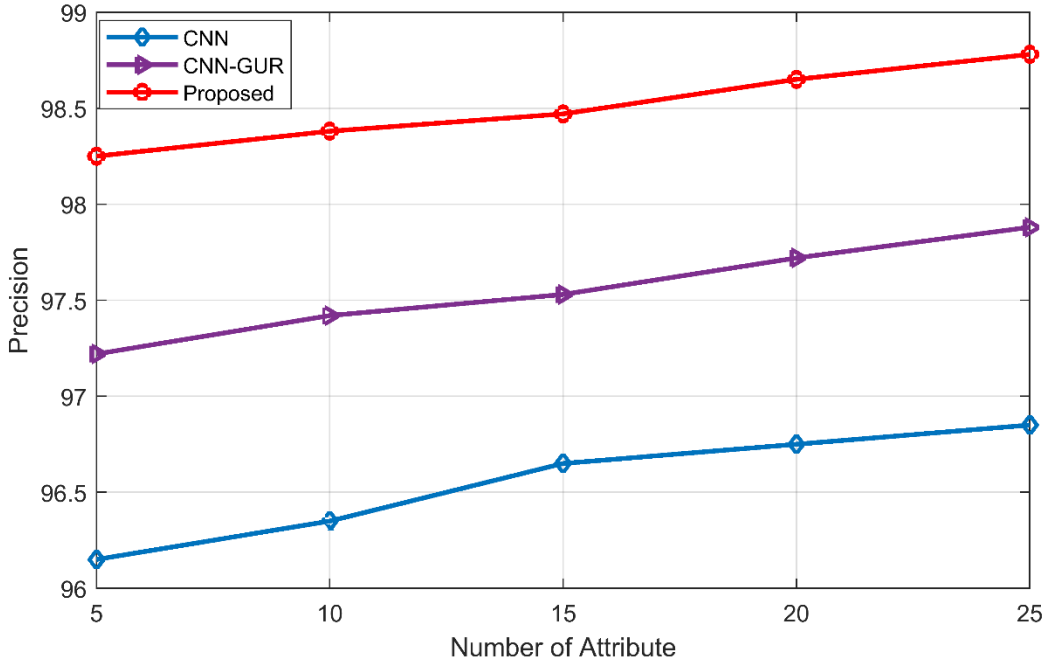

Fig 3: Performance analysis of precision with KDDCUP2003 dataset

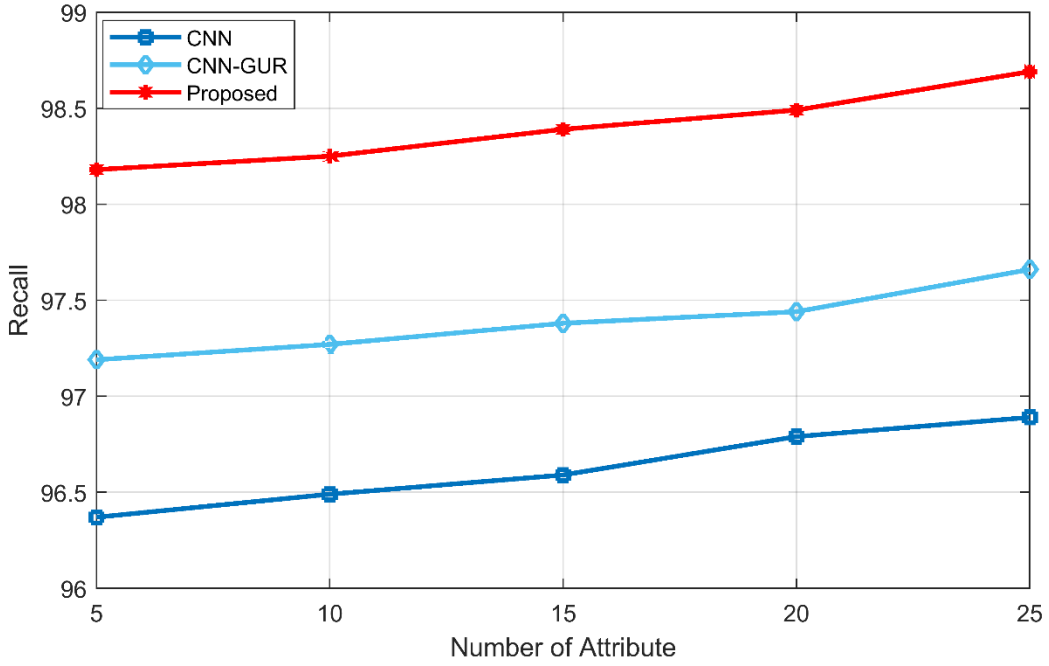

Fig 4: Performance analysis of recall with KDDCUP2003 dataset

V. CONCLUSION & FUTURE SCOPE

The proposed algorithm for feature selection enhances the capacity of the intrusion detection system (IDS). The detection rate varies depending on the range of features selected, which is controlled by the Moth Flame Optimization (MFO) algorithm. The MFO algorithm selects the optimal feature set for training and testing the ensemble classifier, reducing the variance of data features. The MFO algorithm is applied to different patterns of feature sets, such as 5, 10, 15, 20, 25, and 30 features. The variable feature sets impact the detection accuracy and recall, validating the effectiveness of the MFO algorithm compared to other feature selection

methods like Particle Swarm Optimization (PSO) and Genetic Algorithm (GA). The PSO and GA algorithms are applied to the same sets of features as the MFO algorithm. However, some feature selections using GA and PSO face problems with variance, which can decrease classification performance. The results are shown in a comparative analysis. For validation, the KDDCUP2023 dataset is used to evaluate the performance of the CNN and the MFO feature selector.

REFERENCE

[1] Osama, Manar, Abdelhamied A. Ateya, Mohammed S. Sayed, Mohamed Hammad, Paweł Pławiak, Ahmed A. Abd El-Latif, and Rania A. Elsayed. "Internet of medical things and healthcare 4.0: Trends, requirements, challenges, and research directions." Sensors 23, no. 17 (2023): 7435.

[2] Huang, Hsiao-Ching, I-Hsien Liu, Meng-Huan Lee, and Jung-Shian Li. "Anomaly Detection on Network Traffic for the Healthcare Internet of Things." Engineering Proceedings 55, no. 1 (2023): 3.

[3] Ragab, Mahmoud, Sultanah M. Alshammari, and Abdullah S. Al-Ghamdi. "Modified Metaheuristics with Weighted Majority Voting Ensemble Deep Learning Model for Intrusion Detection System." Computer Systems Science & Engineering 47, no. 2 (2023).

[4] Gaber, Tarek, Joseph B. Awotunde, Sakinat O. Folorunso, Sunday A. Ajagbe, and Esraa Eldesouky. "Industrial internet of things intrusion detection method using machine learning and optimization techniques." Wireless Communications and Mobile Computing 2023, no. 1 (2023): 3939895.

[5] Gajevic, Masa, Nemanja Milutinovic, Jelena Krstovic, Luka Jovanovic, Marina Marjanovic, and Catalin Stoean. "Artificial neural network tuning by improved sine cosine algorithm for healthcare 4.0." In Proceedings of the 1st international conference on innovation in information technology and business (ICIITB 2022), vol. 104, p. 289. Springer Nature, 2023.

[6] Khafaga, Doaa Sami, Faten Khalid Karim, Abdelaziz A. Abdelhamid, El-Sayed M. El-kenawy, Hend K. Alkahtani, Nima Khodadadi, Mohammed Hadwan, and Abdelhameed Ibrahim. "Voting Classifier and Metaheuristic Optimization for Network Intrusion Detection." Computers, Materials & Continua 74, no. 2 (2023).

[7] Stankovic, Marko, Jelena Gavrilovic, Dijana Jovanovic, Miodrag Zivkovic, Milos Antonijevic, Nebojsa Bacanin, and Milos Stankovic. "Tuning multi-layer perceptron by hybridized arithmetic optimization algorithm for healthcare 4.0." Procedia Computer Science 215 (2022): 51-60.

[8] Taouali, Okba, Sawcen Bacha, Khaoula Ben Abdellafou, Ahamed Aljuhani, Kamel Zidi, Rehab Alanazi, and Mohamed Faouzi Harkat. "Intelligent Intrusion Detection System for the Internet of Medical Things Based on Data-Driven Techniques." Computer Systems Science & Engineering 47, no. 2 (2023).

[9] Hassaballah, Mahmoud, Yaser M. Wazery, Ibrahim E. Ibrahim, and Aly Farag. "Ecg heartbeat classification using machine learning and metaheuristic optimization for smart healthcare systems." Bioengineering 10, no. 4 (2023): 429.

[10] Almalawi, Abdulmohsen, Asif Irshad Khan, Fawaz Alsolami, Yoosef B. Abushark, and Ahmed S. Alfakeeh. "Managing security of healthcare data for a modern healthcare system." Sensors 23, no. 7 (2023): 3612.

[11] Al Abdulwahid, Abdulwahid. "Detection of Middlebox-Based Attacks in Healthcare Internet of Things Using Multiple Machine Learning Models." Computational Intelligence and Neuroscience 2022, no. 1 (2022): 2037954.

[12] Zubair, Mohammed, Ali Ghubaish, Devrim Unal, Abdulla Al-Ali, Thomas Reimann, Guillaume Alinier, Mohammad Hammoudeh, and Junaid Qadir. "Secure Bluetooth communication in smart healthcare systems: a novel community dataset and intrusion detection system." Sensors 22, no. 21 (2022): 8280.

[13] Ashraf, Eman, Nihal FF Areed, Hanaa Salem, Ehab H. Abdelhay, and Ahmed Farouk. "Fidchain: Federated intrusion detection system for blockchain-enabled iot healthcare applications." In Healthcare, vol. 10, no. 6, p. 1110. MDPI, 2022.

[14] Mboweni, Ignitious V., Daniel T. Ramotsoela, and Adnan M. Abu-Mahfouz. "Hydraulic data preprocessing for machine learning-based intrusion detection in cyber-physical systems." Mathematics 11, no. 8 (2023): 1846.

[15] Yusof, Mohammad Hafiz Mohd, Abdullah Mohd Zin, and Nurhizam Safie Mohd Satar. "Behavioral Intrusion Prediction Model on Bayesian Network over Healthcare Infrastructure." Computers, Materials & Continua 72, no. 2 (2022).

[16] Raghavendra, T., M. Anand, M. Selvi, K. Thangaramya, SVN Santhosh Kumar, and A. Kannan. "An intelligent RPL attack detection using machine learning-based intrusion detection system for Internet of Things." Procedia Computer Science 215 (2022): 61-70.

[17] Alamro, Hayam, Radwa Marzouk, Nuha Alruwais, Noha Negm, Sumayh S. Aljameel, Majdi Khalid, Manar Ahmed Hamza, and Mohamed Ibrahim Alsaid. "Modelling of Blockchain Assisted Intrusion Detection on IoT Healthcare System using Ant Lion Optimizer with Hybrid Deep Learning." IEEE Access (2023).

[18] Khaldi, Miloud, Nadir Mahammed, Mohammed Abdrrahim Lahmar, and Fadela Djelloul Daouadji. "An Intrusion Detection System for Healthcare Applications using Machine Learning." In IAM, pp. 94-101. 2023.

[19] Halman, Laila M., and Mohammed JF Alenazi. "MCAD: a machine learning based cyberattacks detector in software-defined networking (SDN) for healthcare systems." IEEE Access 11 (2023): 37052-37067.

[20] Abbas, Sagheer, Ghassan F. Issa, Areej Fatima, Tahir Abbas, Taher M. Ghazal, Munir Ahmad, Chan Yeob Yeun, and Muhammad Adnan Khan. "Fused weighted federated deep extreme machine learning based on intelligent lung cancer disease prediction model for healthcare 5.0." International Journal of Intelligent Systems 2023, no. 1 (2023): 2599161.

[21] Alnaim, Abdulrahman K., and Ahmed M. Alwakeel. "Machine-learning-based IoT–edge computing healthcare solutions." Electronics 12, no. 4 (2023): 1027.

[22] Manimaran, M., Murali Dhar, Roger Norabuena-Figueroa, R. Mahaveerakannan, S. Saraswathi, and K. Selvakumarasamy. "Implementing Machine Learning-based Autonomic Cyber Defense for IoT-enabled Healthcare Devices." Journal of Artificial Intelligence and Technology 3, no. 4 (2023): 162-172.