Advances in Computer Engineering and Science (ACES) (2026)
Volume -1, Issues – 2

DDoS Attack Detection Using Long Short-Term Memory Based on Hybrid Grey Wolf Optimization and Tabu Search

S RAVISHANKAR
Department of Computer Science
Sona College of Arts and Science, Salem, Tamil Nadu, India
E-Mail: ravirohan83@gmail.com

P KANMANI
Department of Computer Science
Thiruvalluvar Government Arts College, Rasipuram, Tamil Nadu, India
E-Mail: pushpakanmani@gmail.com

Abstract: A Distributed Denial of Service (DDoS) attack in a Hadoop environment can have serious consequences, as it can disrupt the availability and performance of critical services and applications that rely on the Hadoop infrastructure for data storage and processing. An intriguing technique that makes use of deep learning capabilities to examine network traffic patterns and spot unusual behavior suggestive of an attack is the use of Long Short-Term Memory (LSTM) networks for DDoS attack detection in a Hadoop context. However, the LSTM algorithm has several shortcomings such as low accuracy, and convergence rate due to its improper selection of hyperparameters. Hence, the optimized LSTM method is proposed based on hybrid grey wolf optimization (GWO) and tabu search (TS) called GWO-TS. The hybrid GWO-TS method is used to optimize the hyperparameters of LSTM which is applied to detect the DDoS attack in the Hadoop environment. The experimental results show that the developed optimized LSTM produced high detection accuracy with fast convergence when compared to literature algorithms.

Keywords: DDoS attack detection; Hadoop environment; long-short term memory; grey wolf optimization; tabu search;

I. INTRODUCTION

The Hadoop infrastructure, tools, and services are being improved by several open-source communities, IT companies, and academic institutions. Open-source platforms, which aid in the advancement of big data, facilitate the sharing of big data advances. Users are having trouble since a Hadoop framework was built using multiple versions of components from different sources. Within the Hadoop architecture, there is a danger of version incompatibility because each Hadoop element has its curve of maturity. Security concerns are further increased by the integration of several modules from various suppliers using various technologies on a single platform [1]. For businesses that keep their critical data in the Hadoop environment, HDFS security is essential. HDFS is susceptible to a variety of attacks, including the DDoS attack, which works by crashing data or saturating the target with traffic. The HDFS Name Node is susceptible to DDoS attacks. The Name Node will work with the Job Tracker to perform data processing tasks. The read-write function of HDFS can be stopped by a DDoS attack on the Name Node, which will then have an impact on the data processing work [2, 3]. Hence, DDoS attack detection is a major consideration in the HDFS environment which can help to mitigate the traffic and enhance the performance of HDFS during data processing work. Recently, deep learning approaches have been getting more attention to detect abnormal incoming data on HDFS environments.

Recently, the recurrent neural network (RNN) is a kind of DL approach that is applied to solve a variety of problems with high accuracy including DDoS attack detection [4, 5]. The LSTM is a kind of RNN algorithm which used to handle sequential data by maintaining an internal state that captures information about previous elements in the sequence [6]. LSTM can process sequential data online, meaning they can make predictions at each time step as new data arrives. Hyperparameters are parameters that are set before training a model and determine how the model learns during training. For LSTMs, there are several hyperparameters that you need to consider when designing and training your model. These hyperparameters influence the behavior and performance of the RNN. However, the hyperparameters of RNN approaches disturb the performance, and selecting optimal parameters are challenging task [7].

Swarm Intelligence (SI) strategies are influenced by how social creatures like ants, bees, and birds behave in unison. [8]. These approaches can be applied to hyperparameter optimization for RNNs to efficiently search for optimal hyperparameters. Recently, GWO approach is a recently developed approach that is applied to all real-world applications. Particularly for well-behaved optimization problems, GWO usually converges fast to a close to optimal solution. Its method of updating positions by the delta, omega, alpha, and beta wolves enables efficient exploitation of places within the search space that show promise. However, the GWO difficult to maintain population diversity when updating GWO wolf locations based on alpha, beta, delta, and omega wolves to maintain population variety. Hence, GWO has a low convergence rate and population diversity. Since the GWO tends to converge to local optima, it performs poorly in many complex settings even if GWO has proven to be an effective search strategy for solving optimization problems.

Hence, the present research work focused on enhancing the convergence rate and population diversity of GWO by using the TS algorithm. The TS algorithm is used to find the optimal position of the grey wolf to manage the exploitation and exploration ability of GWO. The hybrid approach called GWO-TS is used to optimize the hyperparameters of LSTM for enhancing the detection accuracy and convergence rate. The optimized LSTM was used to detect the DDoS attack and compared with well-known detection methods. Attack detection methods examine incoming packets to find attacks when there is a deviation from the expected flow. The aim is to enhance the performance of LSTM by obtaining optimal hyperparameters for achieving a high DDoS detection rate. The contributions are as follows,

- The optimized LSTM approach is used to detect DDoS attacks based on a new hybrid approach
- The new hybrid approach uses the merits of both GWO and TS named GWO-TS for finding the optimal hyperparameter of LSTM.

- The new optimized LSTM method was compared with some variants of LSTM and the conventional approach for analyzing its performance.
- Two DDoS attack datasets are considered to analyze the performance of detection algorithms and evaluated by using four kinds of performance measures.

II. Related works

In DDoS attack detection research, the related works section aims to give a thorough overview of the body of knowledge about current approaches, methodology, and literature in the subject. V. K. Kamboj et al. (2018) [9] created a novel hybrid technique known as hGWO-PS, which is based on pattern search and the Grey Wolf Optimizer (GWO). The suggested hybrid GWO-PS method's exploitation phase is more effective experimentally than the conventional, previously published meta-heuristics search strategy. However, as the number of fitness evaluations has increased, the algorithm's processing time has marginally increased as well. D. Vasan et al. (2020) [10] offer a potent cross-architecture threat-hunting method (MTHAEL) for Internet of Things malware. By leveraging a layered ensemble of heterogeneous feature selection and state-of-the-art techniques to learn varying degrees of semantic features, the MTHAEL demonstrated enhanced malware detection above current methods. On various IoT architectures, MTHAEL is an efficient optimization of convolutional neural networks (CNNs) and RNNs with consistently low processing overheads and good classification accuracies. An extensive experimental evaluation is conducted using an IoT cross-architecture dataset and demonstrated that the MTHAEL achieves 99.98% classification accuracy.

W. Glenn et al. (2016) [11] examined how attacks affect how long it takes for a Map Reduce task to finish when a compromised node in a Hadoop cluster occurs. Three attack strategies were examined: (1) slowing the delivery of packets sent to the master node, attacking the master node, and hindering all arriving data from the master node aside from special messages that convey the slave node's condition. We applied these attacks on several cluster configurations with three, six, and nine slave nodes as well as a single mother node in the testbed to assess their effects. This information demonstrates how various assaults can impact MapReduce performance by lengthening the computation time of MapReduce tasks. S. Ahmad et al. (2018) [12] provided a study of DDoS attacks on Hadoop using a variety of models and techniques to examine Hadoop's behavior and effects. When a system malfunctions or crashes, Hadoop services become unavailable and cause resource disruptions. S. Paul et al. (2022) [13] offers an effective testbed for creating data sets, detecting attacks within the cluster, and creating internal attacks. Additionally, our approach locates the attacked nodes. The author developed a new insider attack known as the BUSY YARN attack. Comparable insider attacks, when the target node or nodes are unpredictable, can be identified using the framework.

K. B. Virupakshar et al. (2020) [14] created a novel technique for detecting DDoS attacks that is intended to identify instances of connection and bandwidth flooding. These types of assaults are directed at network attacks that are specifically designed to detect flooding of connections and bandwidth. Additionally, a system with a DDoS detection system that uses raw socket programming and an OpenStack-integrated firewall is needed. DDoS attacks are eventually discovered, and the private cloud administrator is alerted. model that has been trained. When DDoS attacks are eventually discovered, the private cloud administrator is alerted. D. Lightbody et al. (2024) [15] present Dragon_Pi, a side-channel power consumption-based intrusion detection dataset created for the Internet of Things devices. Dragon_Pi is a set of power consumption traces—both under assault and normal—from two different testbeds, one with a Raspberry Pi and the other with a DragonBoard 410c. This dataset is used to train Dragon_Slice, an unsupervised convolutional autoencoder (CAE) that is only trained on Dragon_Pi's held-out normal slices to detect anomalies. This study uses two iterations of the Dragon_Slice network.

G. Sagar et al. (2019) proposed three steps of a new malware detection model. The information gain (IG) and term frequency-inverse document frequency (TF-IDF) features are extracted during the feature extraction phase. More significantly, the Holoentropy evaluation is applied to the IG feature. Principal Component Analysis (PCA) is used for feature selection after the feature extraction step. Lastly, an optimized activation function is employed using a Deep Belief Network (DBN) to complete the classification process. This research aims to provide a new hybrid algorithm that combines the idea of ALO with the Glowworm Swarm Algorithm (GSO) to solve this optimization problem. The suggested Lion Updated GSO (LU-GSO) performs better than other traditional models based on several evaluation metrics, demonstrating improvements over others [16]. M. M. Rathore et al. (2016) [17] created an IDS architecture with four layers: the capturing layer, the processing or Hadoop layer, the filtration and load balancing layer, and the decision-making layer. Additionally, a feature selection strategy based on the study of DARPA datasets and the parameters FSR and BER is proposed to pick nine parameters for classification. In addition, the suggested system is assessed using five main machine-learning methodologies.

III. Problem Definition

A DDoS attack on Hadoop clusters, which are dispersed systems made to handle and store massive amounts of data across numerous workstations, is the goal of the attack. The HDFS, YARN, and other associated services like HBase or Hive are just a few examples of the components of the Hadoop ecosystem that the attacker may target with vulnerabilities. A botnet, made up of infected PCs or Internet of Things devices, could be used by the attacker to overwhelm the Hadoop cluster with a large amount of malicious traffic or requests. The attack aims to flood the Hadoop cluster with requests or packets that are faulty to overload its CPU, memory, and network bandwidth. A denial-of-service scenario can arise from the DDoS attack, preventing authorized users from using the Hadoop cluster's resources. Important business processes that depend on the Hadoop infrastructure for analytics and data processing may be affected by this. Data integrity may be in danger from a DDoS attack on Hadoop clusters in addition to availability issues. Data loss or illegal access could result from the deluge of requests or traffic, which has the potential to alter or distort the stored data. Such attacks can be identified and stopped by using intrusion detection and prevention systems (IDPS) and keeping Hadoop ecosystem components updated with security patches.

Another bot from the previous quarter, known by the moniker DemonBot, attracted notice for using a flaw in the way YARN remote commands were executed to take control of Hadoop clusters. Though theoretically simple, this bot's target selection makes it dangerous: Because Hadoop clusters are built to handle Big Data, they have a significant computing power advantage. Additionally, because they are cloud-integrated, they

can greatly increase DDoS attacks. Up to one million viruses can occur per day on the 70 active servers that Radware is now keeping an eye on. DemonBot may easily be re-aimed at a greater number of targets because it is compatible with most IoT devices in addition to Hadoop clusters ("https://securelist.com/ddos-attacks-in-q4-2018/89565/").

IV. Research Methods

A. LSTM

Hochreiter and Schmidhuber's novel RNN architecture, LSTM, can recognize long-term dependencies. As can be seen, depending on the gate's state at any given instant in time t, several gates that control the cell can either keep the value or rearrange it. The forget (f_t), input (i_t), and output gate(o_t) are applied as three gates to the cell. Additionally, there is a candidate value entrance modulation gate. The gates are determined as follows,

$$i_t = \sigma(W_{x,i}x_t + W_{i,h}h_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{f,i}x_t + W_{f,h}h_{t-1} + b_f) \quad (2)$$

$$o_t = \sigma(W_{o,i}x_t + W_{o,h}h_{t-1} + b_o) \quad (3)$$

$$c'_t = tanh(W_{c',i}x_t + W_{c',h}h_{t-1} + b_{c'}) \quad (4)$$

Here, W - synaptic weight, x_t- input samples, b - the bias vectors, the vector c'_t contains fresh candidates. The prior output of the LSTM is represented by h_{t-1}. $\sigma(.)$and $tanh()$ are the sigmoid and tangent hyperbolic activation functions. The first stage in the LSTM algorithm is determining how much of the prior memory rate will be deducted from the state of the cell. This decision is made by the forget gate. How much of the new data is saved in the next step is decided by the input gate. The condition of the cell is then ascertained using the following phrase:

$$c_t = f_t \odot c_{t-1} + i_t \odot c'_t \quad (5)$$

Where, $\odot$ is the elementwise product. h_t – is the hidden output defined as follows,

$$h_t = o_t \odot tanh(c_t) \quad (6)$$

The traditional RNN is impacted by the vanishing gradient problem. It becomes difficult to train the network using backpropagation of errors. Further, LSTM is used to prevent the vanishing gradient problem.

B. Grey wolf optimization (GWO)

The GWO is to imitate the predatory behavior and leadership hierarchy of wolves. It then makes use of the grey wolf abilities, such as hunting, encirclement, and search and grab, to accomplish the optimization. The N wolf with D search space that the position of the i^{th} wolf can be defined as: $X_i = (X_{i1}, X_{i2}, X_{i3}, \ldots, X_{id})$. The alpha (α) wolf is determined by considering the solution that fits the best to mathematically denote the social hierarchy of wolves. As a result, beta (β) and delta (δ) wolves are the names of the second and third-best solutions, respectively. It is presumed that the remaining candidate solutions are omega (ω) wolves. Grey wolf encirclement behavior can be analytically modeled as follows:

$$D = \left|C \times X_p(t) - X(t)\right| \quad (7)$$

$$X(t+1) = X_P(t) - A \times D \quad (8)$$

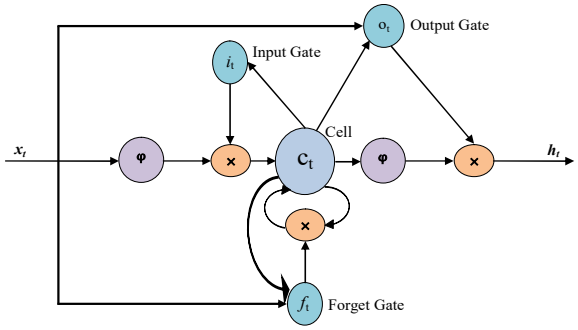

Fig. 1 : Architecture of LSTM

Where $X_p(t)$ - represents the position of the prey at t^{th} iterations. The $X(t)$ is the grey wolf position. The C is a control coefficient determined as follows,

$$C = 2r_1 \quad (9)$$

Where, r_1 is the random variable between [0,1]. The convergence factor A is calculated as follows

$$A = 2ar_2 - a \quad (10)$$

$$a = 2\left(1 - \frac{t}{T\,max}\right) \quad (11)$$

Where, a is the control coefficient that linearly decreases from 2 to 0. ($a_{max} = 2\ and\ a_{min} = 0$) . The outside wolves follow the leader wolf to surround the prey when the grey wolves catch prey. Then, to catch the prey, the alpha wolf leads the beta and delta wolves. Since the gray wolves are the closest to the prey, their positions can be used to determine the prey's location. The particular mathematical model is this one:

$$D_\alpha = |C_1 \times X_\alpha(t) - X(t)| \quad (12)$$

$$D_\beta = \left|C_2 \times X_\beta(t) - X(t)\right| \quad (13)$$

$$D_\delta = |C_3 \times X_\delta(t) - X(t)| \quad (14)$$

$$X_1 = X_\alpha - A_1 \times D_\alpha \quad (15)$$

$$X_2 = X_\beta - A_2 \times D_\beta \quad (16)$$

$$X_3 = X_\delta - A_3 \times D_\delta \quad (17)$$

$$X(t+1) = \frac{X_1 + X_2 + X_3}{3} \quad (18)$$

The position of the wolves as they approach the prey is estimated after determining the distance between $X(t)$ and the wolves.

C. Tabu search

Glover (1986) created tabu search (TS), which has been used in a variety of challenging optimization situations. The iterative process known as TS was created to address optimization issues. TS begins with a randomly selected solution and assesses the solution's fitness function. Next, every potential neighbor of the provided solution is created and assessed. A neighbor is a solution that can be obtained by a straightforward transformation from the current solution. Select the best neighbor if it isn't in the tabu list as the new current solution. Solutions that have already been

investigated are tracked by the tabu list, which also prevents TS from going over them again. TS will therefore increase if the best neighbor solution outperforms the present strategy. Local minima can be overcome in this fashion. Reversing these decisions or actions is therefore forbidden and categorized as tabu. If introducing some ambition criteria that permit overriding the current tabu status still results in a greater fit relative to the fitness of the current optimum, then that move may be justified. The method finishes when there are no more neighbors (all are tabu) or when no improvements are identified after a set number of iterations. If not, the algorithm carries out the TS steps.

D. Hybrid GWO-TS

Although GWO has high convergent qualities and is effective at solving optimization problems, within a few generations the population variety will drop off dramatically, and the algorithm may lead to an early convergence to a local optimum. TS is a stochastic technique that can potentially converge asymptotically to a global optimum solution, even though it will take a long time to reach the near-global minimum. By using TS as a local improvement strategy within GWO, the algorithm can maintain population variation while avoiding the creation of spurious local optima. The TS was applied to pick the new current grey.

E. Proposed IGWO-LSTM

An inventive method that combines deep learning with metaheuristic optimization to improve the performance of LSTM models for DDoS attack detection in a Hadoop environment is the combination of LSTM and hybrid GWO-TS. Determine the number of layers, hidden units, activation functions, and other details while designing the LSTM model architecture. The foundational model for DDoS attack detection will be this architecture. To optimize the LSTM model's hyperparameters, apply GWO. Grey wolf social behavior serves as the inspiration for the population-based optimization method known as GWO. It can be used to find the ideal combination of to optimize the LSTM model's performance. Start the gray wolf population using random solutions that correspond to various LSTM model hyperparameter sets.

By using the corresponding hyperparameters to train the LSTM model on the training data and validating it on a different validation set, you can assess the fitness of each grey wolf solution. Reposition each grey wolf according to the pack's social rank and fitness rating. This entails modifying the hyperparameters to achieve better results and raise the LSTM model's performance. Iteratively update phases and repeat the fitness assessment until a termination criterion—such as accomplishing a desired performance level or a maximum number of iterations—is satisfied. Choose the GWO-identified optimal solution, which corresponds to the collection of hyperparameters that produce the best LSTM model performance on the validation set.

V. EXPERIMENTAL RESULTS AND ANALYSIS

Examine the experimental data to determine the advantages and disadvantages of the LSTM+IGWO strategy. Determine the situations in which the model detects DDoS attacks well and the situations in which it might falter or generate false positives. Two types of datasets, namely NSL-KDD and CIC-IDS2017, are taken into consideration. Examine how well the LSTM+GWO strategy performs in Hadoop systems when compared to other cutting-edge DDoS attack detection techniques or baseline methodologies. This comparison aids in ascertaining whether the

Table 1 : Parameters setting

Parameter(s)	Values
Hidden layers	2
Epochs	100
Batch size	60
Activation function (hidden layer)	Tanh and Signomoid
Dropout	0.35
Optimizer	IGWO
Learning rate	0.6

LSTM+GWO technique offers appreciable gains in efficiency or accuracy of detection. Talk about the ramifications of the experimental findings, such as possible causes affecting the LSTM+GWO approach's performance, scaling issues, and practical difficulties in real-world implementation.

A. Datasets collections

Numerous intrusion detection assessment datasets include both normal and abnormal network traffic data. Our research focuses on detecting DoS/DDoS attacks; hence, we chose two datasets: NSL-KDD and CIC-IDS2017.

1) *NSL-KDD* ("https://www.unb.ca/cic/datasets/nsl.html"): A well-liked benchmark dataset for assessing IDSs, especially in the context of network, is the NSL-KDD dataset. By eliminating redundant records, offering a more evenly distributed range of attack types, and upgrading the dataset to reflect more recent network traffic patterns, the NSL-KDD dataset overcomes these problems. Entirely divided into training and testing sets, the NSL-KDD dataset comprises over 22,000 data examples. Two thousand instances make up the testing set and about twelve thousand instances make up the training set.
2) *CIC-IDS2017* ("https://www.unb.ca/cic/datasets/ids-2017.html") was collected over five days in 2017 during various cyberattacks and non-attack days which includes both consistent and DDoS attack data [18]. 225,742 samples, 85 network flow features, and label attributes are all included in the dataset. Due to the uneven nature of this dataset, we altered the training dataset for this study such that the proportion of attack and normal data was balanced, resulting in a reduction of the number of instances to 80 characteristics. The datasets are divided into two categories: harmful and

TABLE 2 : RESULTS ANALYSIS OF DDOS ATTACK DETECTION METHODS FOR THE NSL-KDD DATASET

Methods	Accuracy	Recall	Precision	F-Score
LSTM+GWO-TS	90.99	90.34	90.67	89.99
LSTM+GWO	89.64	89.49	89.45	88.75
LSTM+PSO	89.57	89.22	88.24	87.79
LSTM+GA	88.97	88.97	87.75	87.45
LSTM	87.59	87.49	86.79	86.57
ERNN	86.96	86.56	86.65	85.57

TABLE 3: RESULTS ANALYSIS OF DDOS ATTACK DETECTION METHODS FOR THE NSL-KDD DATASET

Methods	Accuracy	Recall	Precision	F-Score
LSTM+GWO-TS	90.59	90.49	91.87	91.79
LSTM+GWO	90.66	89.97	89.79	90.55
LSTM+PSO	89.97	89.89	88.85	89.78
LSTM+GA	89.54	88.67	87.85	88.79
LSTM	88.99	88.49	87.14	86.59
ERNN	87.98	87.79	86.98	85.89

legitimate networks ('including email, SSH, FTP, HTTP, and HTTPS protocols').

B. Preprocessing

To eliminate the impact of the original feature value scales, the numerical features have been standardized. For every feature, we have applied the Min-Max Normalization, which rescales the feature range to fall inside the interval [0, 1]. The formula for Min-Max Normalization is represented by the following equation [19, 20]:

$$x\_new = \frac{x - min(x)}{max(x) - min(x)} \quad (19)$$

The dataset was then divided in half: 20% was reserved for the testing dataset, which was used to evaluate the models with additional, as-yet-unviewed data, while the remaining 80% was used to train and analyze the selected machine learning approaches.

C. Parameter settings

In LSTM networks, hyperparameters are crucial in defining the architecture, training behavior, and, in the end, the model's performance on a particular task. Table 1 shows the parameters setting of the LSTM network.

D. Performance measures

Performance metrics are essential for detecting DDoS attacks because they offer valuable information about how the network and its resources behave. The prediction algorithm's performance was evaluated in this study using four performance measures.

- A prediction algorithm's total performance is evaluated based on its accuracy, which may be thought of in the following way:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (20)$$

- Recall: The capacity to predict positive cases—like the real positive rate, which can be interpreted in the following ways—is known as

$$Recall = \frac{TP}{TP+F} \times 100\% \quad (21)$$

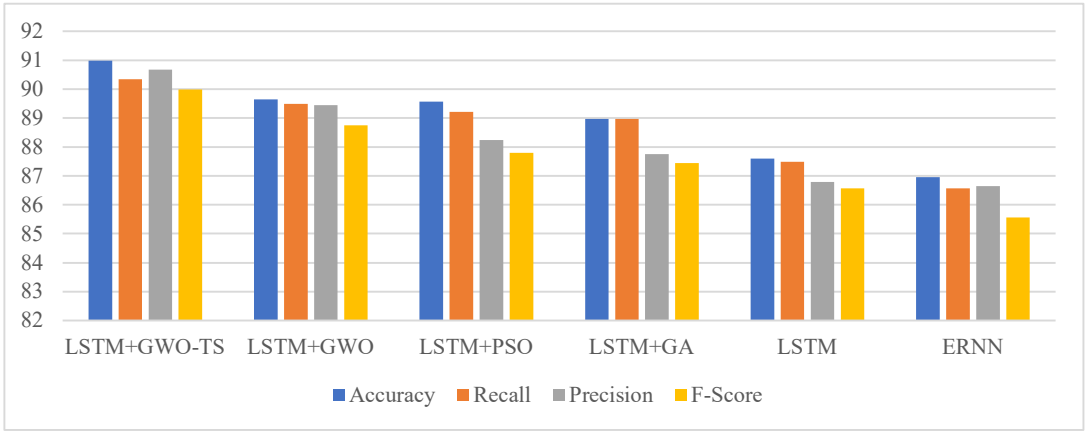

Fig. 2 : Results analysis of DDoS attack detection methods for NSL-KDD dataset

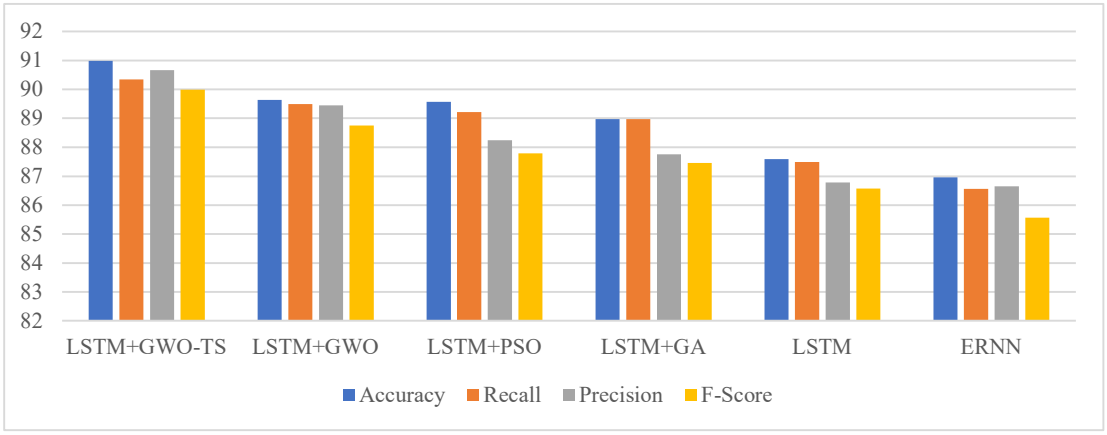

Fig. 3 : Results analysis of DDoS attack detection methods for CIC-IDS2017 dataset

- Precision: Precision, also referred to as a positive predictive value, is the number of appropriate examples among the recovered samples and is defined as follows,

$$Pr\,e\,cision = \frac{TP}{TP+F} \times 100\% \quad (22)$$

- F-score: It is calculated by taking the harmonic mean of the precision and sensitivity, assigning roughly equal weight to each. This makes it possible to compare prototypes, describe the performance of a prototype, and combine recall and precision into a single score.

$$F - score = 2 * \frac{Pr\,ecisio\ \times Re\,ca}{Pr\,ecision+Re\,ca} \times 100\% \quad (23)$$

Where, True Positives (TP): The quantity of DDoS attack cases that were correctly predicted. False Positives (FP): The number of occurrences that are mistakenly categorized as DDoS attacks. True Negatives (TN): The quantity of exactly predicted normal cases. False Negatives (FN): The number of samples that are incorrectly categorized as normal while they are DDoS attacks.

E. Results analysis

Several factors must be considered when analyzing the impacts of a DDoS attack on a Hadoop cluster. These include the attack's nature, how it affects the Hadoop architecture, and possible defenses. Understanding the efficacy of detection systems, optimizing detection algorithms, and enhancing overall cybersecurity posture all depend on the role that outcomes analysis plays in DDoS attack detection. The present research work proposed an optimized DDoS attack detection approach based on optimized LSTM. Tables 2 and 3 show performance results analysis for detection methods for NSL-KDD and CIC-IDS2017 datasets. Also, a graphical representation of the compared methods is shown in Figures 1 and 2. According to the experimental results, the developed method such as LSTM+GWO-TS produced high detection accuracy with a fast convergence rate.

VI. Conclusion

DDoS attacks are a tactic used by attackers to block Hadoop accessibility. An improved DDoS attack detection method for the Hadoop environment is presented in the current work. The recommended method of using the hybrid GWO-TS algorithm is utilized to assign more appropriate hyperparameters of LSTM. Four performance analyzers are used to evaluate the proposed GWO-TS-based LSTM scheme's performance over the two DDoS attack datasets. The experimental results showed that the GWO-TS-based LSTM performed better in terms of generalization than other variations and conventional detection systems.

REFERENCES

[1] M. Liu, Z. Xue, X. Xu, C. Zhong, and J. Chen, "Host-based intrusion detection system with system calls: Review and future trends," *ACM Computing Surveys (CSUR),* vol. 51, no. 5, pp. 1-36, 2018.

[2] G. S. Bhathal and A. Singh, "Big Data: Hadoop framework vulnerabilities, security issues and attacks," *Array,* vol. 1, p. 100002, 2019.

[3] K. Vijayakumari, J. Gnanadurai, S. Usharani, and K. Velusamy, "Deep Learning Models for Intelligent IoT Ecosystems," in *Intelligent Mobile and IoT Ecosystems*: Chapman and Hall/CRC, 2026, pp. 131-149.

[4] M. Malini and N. Chandrakala, "DDoS attack detection in the cloud environment using an optimised long short-term memory with an improved firefly algorithm," *International Journal of Communication Networks and Distributed Systems,* vol. 32, no. 1, pp. 58-87, 2026.

[5] X. Liang and T. Znati, "A long short-term memory enabled framework for DDoS detection," in *2019 IEEE global communications conference (GLOBECOM)*, 2019: IEEE, pp. 1-6.

[6] B. Fathimamary, M. Nasreen, and K. Velusamy, "An Efficient DDoS Attack Detection Using Optimized Long Short-Term Optimization Based on Improved Brainstorm Optimization," *Indian Journal of Science and Technology,* vol. 19, no. 5, pp. 298-312, 2026.

[7] M. Sinthuja and K. Suthendran, "DDoS attack detection using enhanced long-short term memory with hybrid machine learning algorithms," in *2022 3rd International Conference on Smart Electronics and Communication (ICOSEC)*, 2022: IEEE, pp. 1213-1218.

[8] S. Wang *et al.*, "Prediction and optimization model of sustainable concrete properties using machine learning, deep learning and swarm intelligence: A review," *Journal of Building Engineering,* p. 108065, 2023.

[9] V. K. Kamboj, A. Bhadoria, and N. Gupta, "A novel hybrid GWO-PS algorithm for standard benchmark optimization problems," *INAE Letters,* vol. 3, no. 4, pp. 217-241, 2018.

[10] D. Vasan, M. Alazab, S. Venkatraman, J. Akram, and Z. Qin, "MTHAEL: Cross-architecture IoT malware detection based on neural network advanced ensemble learning," *IEEE Transactions on Computers,* vol. 69, no. 11, pp. 1654-1667, 2020.

[11] W. Glenn and W. Yu, "Cyber attacks on MapReduce computation time in a Hadoop cluster," *Big Data Concepts, Theories, and Applications,* pp. 257-279, 2016.

[12] S. Ahmad, A. Yasin, and Q. Shafi, "DDoS attacks analysis in bigdata (hadoop) environment," in *2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 2018: IEEE, pp. 495-501.

[13] S. Paul, S. Saha, and R. T. Goswami, "Detection of Unknown Insider Attack on Components of Big Data System: A Smart System Application for Big Data Cluster."

[14] K. B. Virupakshar, M. Asundi, K. Channal, P. Shettar, S. Patil, and D. Narayan, "Distributed denial of service (DDoS) attacks detection system for OpenStack-based private cloud," *Procedia Computer Science,* vol. 167, pp. 2297-2307, 2020.

[15] D. Lightbody, D.-M. Ngo, A. Temko, C. C. Murphy, and E. Popovici, "Dragon_Pi: IoT Side-Channel Power Data Intrusion Detection Dataset and Unsupervised Convolutional Autoencoder for Intrusion Detection," *Future Internet,* vol. 16, no. 3, p. 88, 2024.

[16] G. Sagar, "Malware detection using optimized activation-based deep belief network: An application on Internet of Things," *Journal of Information &*

Knowledge Management, vol. 18, no. 04, p. 1950042, 2019.

[17] M. M. Rathore, A. Ahmad, and A. Paul, "Real time intrusion detection system for ultra-high-speed big data environments," *The Journal of Supercomputing,* vol. 72, pp. 3489-3510, 2016.

[18] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *ICISSp,* vol. 1, pp. 108-116, 2018.

[19] K. Velusamy and R. Amalraj, "Cascade correlation neural network with deterministic weight modification for predicting stock market price," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1110, no. 1: IOP Publishing, p. 012005.

[20] K. Velusamy and R. Amalraj, "Performance of the cascade correlation neural network for predicting the stock price," in *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, 2017: IEEE, pp. 1-6.