Attention-based RNN architecture for detecting multi-step cyber-attack using PSO metaheuristic

Conference Paper · April 2023

DOI: 10.1109/ECCE57851.2023.10101590

4 authors:

Pritom Biswas
Khulna University of Engineering and Technology
3 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Kowshik Sankar Roy
Khulna University of Engineering and Technology
3 PUBLICATIONS **6** CITATIONS
SEE PROFILE

Md. Ebtidaul Karim
Khulna University of Engineering and Technology
11 PUBLICATIONS **54** CITATIONS
SEE PROFILE

Shah Muhammad Azmat Ullah
Khulna University of Engineering and Technology
6 PUBLICATIONS **73** CITATIONS
SEE PROFILE

Attention-based RNN architecture for detecting multi-step cyber-attack using PSO metaheuristic

Pritom Biswas Udas[1], Kowshik Sankar Roy[1], Md. Ebtidaul Karim[1], Shah Muhammad Azmat Ullah[1]
[1]Dept of Electronics and Communication Engineering,
Khulna University of Engineering & Technology, Bangladesh
Email: udas1609037@stud.kuet.ac.bd, roy1609001@stud.kuet.ac.bd, ebtikarim@ece.kuet.ac.bd, azmat@ece.kuet.ac.bd

***Abstract*— In recent years, the intensive usage of electronic devices called for a greater threat to preventing a massive volume of information generated by billions of users every second. Therefore, ensuring the stability of these data is deemed to be the cornerstone of the field of cyber security. However, the reliability of any cyber security system has often been compromised with the introduction of various malware and intrusive features within the system. To deal with such abnormal characteristics, an Intrusion Detection System (IDS) has played a vital role over the years. Countless work has continuously been performed to make the IDS more effective and reliable than ever. In this paper, an attention-based Recurrent Neural Network (RNN) model has been proposed for detecting various multi-step cyber-attacks in the network. Our classification model comprises a Long Short-Term Memory (LSTM) unit with an Attention layer. A metaheuristic approach, Particle Swarm Optimization (PSO), has been utilized to exploit the most effective and suitable features with a 72.73% reduction rate from the dataset along with reduced computational complexity and time consumption of around three times less as well as improved detection rate by greater than 1%. This proposed method's performance is evaluated against several evaluation metrics and further analyzed against several traditional classifiers. When compared to the corresponding values of different models on the same dataset, experimental results show significantly improved results in different aspects using the proposed approach.**

Keywords—Cyber Security, Deep Learning, Attention Layer, Feature Selection, Network Intrusion Detection, Particle Swarm Optimization

I. INTRODUCTION

Cyber-attacks have been identified as one of the top-rated risks and have become the norm in both the public and private sectors. The rate at which cyberattacks are increasing, making it difficult to keep track of the data and predict what kinds of future cyber threats may be in place. In addition to that, the ever-changing form of different types of cyberattacks continuously increasing the vulnerabilities of every major system. However, cyberattacks are not completely arbitrary rather carry certain specific tokens. A closer inspection of the numbers reveals patterns indicating early signs of strike. A cyber-attack is any malicious effort to enter a computer, computing system, or computer network. Attacks on computer systems are designed to prevent their normal operation, cause disruption, destroy them, or gain control over them, as well as to modify, block, erase, manipulate, or steal the data they contain. It can be of various types and classes. For instance, SQL Injection attacks, Password attacks, Phishing attacks, or other events can talk about where single action, possibly repeated, is required to threaten the system, which can be called single-step attacks. Advanced attackers, on the other hand, use a step-by-step process to try to break into a system. There are two reasons to pursue this course of action. First, advanced attackers often target large or medium-sized organizations with complicated network topology and several protection layers. Considering that the most valuable information assets are located in the most inaccessible regions of the network, it would be very hard to successfully execute an infiltration using a single-step attack. Second, if the attack is fragmented into several parts, it is more stealthy and harder for the victim to identify, particularly if some of the steps do not constitute a risk to the system on their own. These forms of attacks are generally outlined as multi-stage or multi-step attacks [1]. As a countermeasure to such attacks (both single-stage and multi-stage) in the system and further protect the system even before the occurrence of such events, the Intrusion Detection System has played an extensive role over the years.

In recent years, different Machine Learning (ML) techniques have been intensively used for classifying numerous single and multi-step cyber-attacks. ML approaches have the capability of detecting unknown attacks within the system by reaching certain conclusions from seen data. Owing to the success of ML methods, deep learning methods have been lately introduced in this field to enhance the performance of any IDS model. It has the benefit of processing more data with a variety of parameters and layers in a more efficient manner modeled after the human brain's neural activity. Although certain deep learning schemes have been continuously utilized for developing intelligent and effective IDS models, Recurrent Neural Networks (RNNs) have lately gained substantial importance due to their expanding computer resources among them. RNNs have played an important role in recent years in different fields, including the internet of things (IoT), smart grids, time series forecasting, along with others [2].

An attention-based recurrent neural network (RNN) model for intrusion detection systems has been suggested in this article. The overall contribution of this paper, which will be outlined in the following paragraphs, has been broken down into four distinct parts.

- A novel deep learning model incorporating LSTM and attention mechanism has been established to detect various multi-steps cyber-attacks in the field of IDS.
- Attention mechanism into the LSTM model has been introduced to emphasize the most important features of the input matrix. Which provides

979-8-3503-4536-0/23/$31.00 ©2023 IEEE

learning of important data from the large sequence of data package vectors.

- In order to reduce complexity and boost speed, Particle Swarm Optimization (PSO) has been utilized to reduce high-dimensional data into the optimal feature subset.
- For each class detection strategy, a set of performance parameters has been reviewed and analyzed. To test the robustness and superiority of the proposed model, the observational results have been evaluated against several individual classifiers' performance.

II. Related Works

The literature works that have been carried out in the IDS field over the years can be evaluated into three distinctive categories. The first type deals with the feature-matching approach, where the main idea is to look for the specific pattern of different characters or tokens among the given set of input data. This approach is also known as the pattern-matching method. The second type involves using the conventional ML approaches to find out the intrusive behaviors among the given data. In contrast, the last of these categories is concerned with detecting such intrusion in a more effective manner through incorporating deep learning approaches to build the IDS model.

For the sake of building a filter-based IDS, the authors in [3] incorporate a number of ML and deep learning approaches to find out the superior detection model. Their experimental results showed that the feed-forward deep neural network IDS performs better than the other models in detecting intrusions. The authors in [4] introduced an RNN-based classification model in their work for using the time series features in a proper way. The suggested RNN model has demonstrated numerous improvements over traditional ML algorithms, with varied numbers of hidden nodes and learning rates. Following the success of RNN, LSTM, which is the updated form of RNN, has been introduced in IDS models. In [5], the authors have proposed a PCA-based LSTM detection model for detecting binary and multiclass intrusion within the NSL-KDD dataset. Here, their primary goal was to find out the minimal features within the dataset while maintaining the functionality of the classification model. The authors in [6] consider an AdaBoost-based IDS model on the CICIDS dataset, which is a modern updated version of the IDS dataset that includes the recent footprint of various network attacks. A Deep Belief Network (DBN) incorporated with the backpropagation method was introduced by the authors in [7], where the performance of the proposed model was evaluated over the well-known NSL-KDD dataset. For the sake of handling the imbalanced network traffics, the authors in [8] utilize a CNN-based classification scheme. Prior to using CNN for balancing the network traffic, they applied the Synthetic Minority Oversampling Technique combined with Edited Nearest Neighbors (SMOTE-ENN) algorithm in their work. Feature extraction was accomplished using Principal Component Analysis (PCA) in research [9]. The most discriminatory features needed for the detection model were extracted using a feature selection approach that combined the Genetic Algorithm (GA) and PSO techniques. Both Artificial Neural Network (ANN) and Support Vector Machine (SVM) classification techniques were utilized to create intrusion detection with the KDDCUP dataset. The authors of [10] utilized three deep learning models: Deep Belief Networks, Long Short-Term Memory Recurrent Neural Networks (LSTM-RNN), and Deep Neural Networks (DNN) to investigate performance differences for intrusion detection using the NSL-KDD and CICIDS2017 datasets, both of which are rather popular. Additionally, employing double PSO, one for optimal feature selection and the other for hyperparameter optimization. The BAT model for intrusion detection was created by the authors of [11] with the help of the standard-setting NSL-KDD dataset. In this work, the authors have taken into account Bidirectional-LSTM and the attention layer to create a new kind of detection model. Additionally, a technique known as BAT-MC, which uses multiple convolution layers, is implemented. The hybrid categorization model has been studied both with and without the attention layer being analyzed. The authors in [12] have proposed a novel classification model named SPIDER, adopting a hybrid approach comprising several CNN, LSTM, GRU, Bi-LSTM, and Bi-GRU in a stacked configuration. In their work, PCA has been taken into consideration for reducing the dimensionality and additional complexity of the detection model. In addition, NSL-KDD and UNSW-NB15 datasets have been used to evaluate the proposed approach's overall performance for binary and multiclass detection schemes.

Apart from the aforementioned literature that has been discussed, there exists a multitude of other works that have been created in the field of IDS, combining various deep learning algorithms. However, the scope of building an effective detection model by incorporating all the recent extension of the field have yet to be well recognized in the scope of IDS. For the sake of conducting our observation through exploring the most recent addition in the field of RNN, we have decided to use the LSTM model with the aid of attention mechanism to build our classification model. Where, we have used the PSO metaheuristic method for finding out the best effective features suitable for the work.

III. Proposed Work

Our proposed approach is divided into two stages: data pre-processing and classification model for multi-step cyber-attack detection as shown in Fig. 1. The pre-processing stage is primarily concerned with the transformation of the features within the dataset which consists of two fundamental steps. Firstly, PSO, a wrapper-based feature selection algorithm, has been performed to determine the optimal amount of feature subsets that are necessary and appropriate for the classifier to classify the connection or activity as normal or as an attack and to reduce operational and time complexity. Second, the process of standardizing data has been done for selected features. Here, the primary objective is to convert all the selected features from the dataset to an equal scale, so it becomes easier for the model to analyze. After finding the optimal subset of features using PSO and standardizing those features, a neural network for classification purposes has been established. Here, we have proposed an attention-based recurrent neural network comprising of LSTM for multi-step cyber-attack detection. Upon completing the pre-processing stage all these transformed features are eventually passed to the detection model for classification purpose.

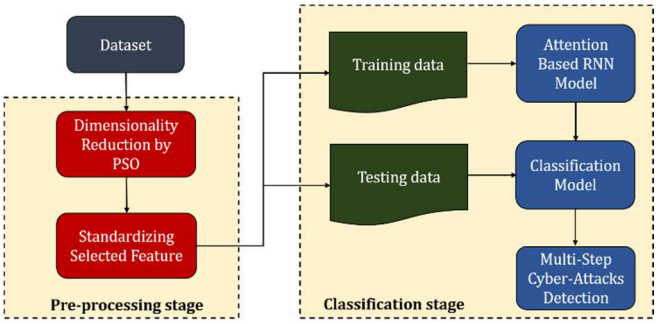

Fig. 1. Flowchart of the proposed approach

A. Data Pre-Processing

Regarding the feature selection process using the PSO algorithm, Table I displays the chosen subset of features and the rate at which features were reduced for the dataset. The Feature Reduction Rate (FRR) indicates the proportion of the total number of features that were removed [13]. Eq. (1) is used to determine the value of FRR.

$$FRR = 1 - \frac{Number\ of\ Selected\ Features}{Number\ of\ All\ Features} \quad (1)$$

TABLE I. THE FEATURE SELECTION OF THE DATASET

No. of features	Selected features	No. of selected features	Selected features (%)	FRR (%)
66	18	4,8,13,14,16,17,23,24,25, 26,34,35,37,40,51,53,57,58	27.27	72.73

Particle Swarm Optimization: The PSO was developed by Eberhat and Kennedy and is a population-based optimization method. PSO mimics social behaviors seen in nature, such as flocking birds and schooling fish. Candidate solutions are represented as particles in PSO's search space and form a population called a swarm. PSO initializes particles randomly. Particles travel across the search space, adjusting their positions depending on their own and their neighbors' experiences [14]. During the movement, the current position of particle *i* and its velocity are expressed in (2) and (3):

$$y_i = \{y_{i1}, y_{i2}, \ldots, y_{iD}\} \quad (2)$$

$$v_i = \{v_{i1}, v_{i2}, \ldots, v_{iD}\} \quad (3)$$

The best prior location of a particle is recorded as the personal best, or pbest, while the best position acquired by the swarm so far is the global best, or gbest. PSO iteratively updates particle positions and velocities in accordance with the following equations (4) and (5), searching for the best possible solution.

$$y_{id}^{t+1} = y_{id}^{t} + v_{id}^{t+1} \quad (4)$$

$$v_{id}^{t+1} = w * v_{id}^{t} + c_1 * r_{1i} * (p_{id} - y_{id}^{t}) + c_2 * r_{2i} * (p_{gd} - y_{id}^{t}) \quad (5)$$

where t represents the *t*th cycle of the evolutionary process. The *d*th dimension of the search space is denoted by d ∈ D. *w* is the inertia weight. c_1 and c_2 are constants of acceleration. r_{1i} and r_{2i} are random values distributed evenly in the interval [0, 1]. In the *d*th dimension, p_{id} and p_{gd} represent the constituents of pbest and gbest.

Data Standardization: This process has been adopted following the operation of feature selection. In order to make it easier for most algorithms to process a dataset, it is best to normalize the features within it to the same scale. Furthermore, it raises the possibility of bettering the model's performance and accuracy by decreasing data bias. The formula for standardization is shown in equation (6) where x is the value of the number feature.

$$Xstandardization = \frac{x - x\_mean}{std(x)} \quad (6)$$

B. Our Proposed Classification Model

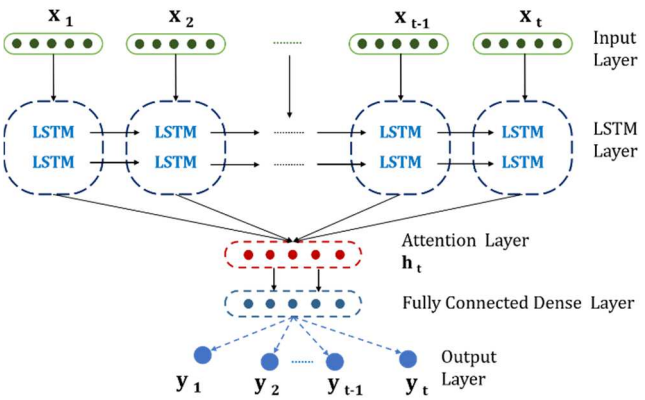

Fig. 2. Architecture of the cyber-attack detection model

As shown in Fig. 2 from bottom to top, our proposed model consists of four components, including the input layer, LSTM layer, attention layer and output layer. Upon selecting the most salient features for the detection model utilizing PSO, these are then channeled into the classification model. At the input layer, each traffic byte is altered into decimal data format. Following the numerical representation of the traffic byte, standardization operations are carried out. The LSTM layer uses the LSTM model to extract features from each packet's traffic bytes. Because LSTM is suited to the structure of network traffic, it can learn the sequential properties contained inside communication bytes. The attention layer analyzes the most important packet vectors to provide most useful features that are more prominent for the detection of malicious traffic. The selected features that most efficiently determines network traffic behavior are obtained at the output layer when the features produced by the attention mechanism are transferred into a fully connected layer. For the purpose of avoiding complexities in the work, grid search technique has been used to optimize the necessary hyperparameters in the model. The values of the proposed model's hyperparameters are presented in Table II.

TABLE II. FUNDAMENTAL HYPERPARAMETER VALUES OF THE LEARNING MODEL

HYPER-PARAMETERS	VALUES
LSTM	Activation = tanh, Neurons=64
Dense (1) Dense (2)	Activation = ReLu, Neurons = 32 Activation = SoftMax, Neurons= 6
Batch Size Learning Rate Optimizer Cost function Iterations	64 0.001 Adam Categorical cross entropy 100

LSTM Layer: Long Short-term Memory is a special kind of RNN that can cope up with long-term data dependencies. It is the improved variant of RNN that have been developed to overcome the vanishing gradient problems of conventional RNNs [15]. An LSTM module has a cell state as its principal component along with three gates which allows it to selectively learn, unlearn, or retain information. LSTM's cell state allows only a few linear interactions, preventing information from being altered. Each unit includes an input, output, and forget gate to add or erase cell state information. The forget gate uses a sigmoid function to forget prior cell state information. The input gate regulates the current cell state via point-wise sigmoid and tanh multiplication. The output gate determines what to send to the next hidden state.

Attention Layer: An attention layer is primarily responsible for focusing on certain parts of the input where the model needs to selectively process parts of the input that is more relevant to the task. The mechanism works by assigning a weight to each input element based on its relevance to the output. These weights are learned during training and are used to compute a weighted sum of the input elements. The resulting weighted sum is then passed through the rest of the neural network [16].

Dense Layer: A dense layer, also known as a fully connected layer, is a type of artificial neural network layer where each neuron in the layer is connected to every neuron in the previous layer. A dense layer is characterized by a set of weights and biases that are learned during training. These weights and biases determine the output of each neuron in the layer based on its input. The output of a dense layer can be computed as a linear combination of the inputs, followed by the application of an activation function.

IV. EVALUATION

A. Dataset Description

The dataset is collected from IEEE-DataPort named Multi-Step Cyber-Attack Dataset for intrusion detection (MSCAD). [17]. There are 66 feature columns and one 'Label' column in the dataset. The first scenario in MSCAD is a password cracking attack, while the second is a volume-based Distributed Denial of Service (DDoS) attack. Both scenarios are multi-step in nature. In the first scenario, an attacker uses brute force to crack passwords on any victim network host. This attack has three basic steps. First, the port scan was executed simultaneously. Second, to take offline copies of the web application pages, the website crawler HTTrack Website Copier was used. In the next phase, the attacker aimed to launch a DDoS attack on any host within the victim network. The volume-based DDoS attack was carried out in three stages. First, run the port scan attack (Full, SYN, FIN, and UDP Scan). Then, execute an HTTP Slowloris APP-based DDoS attack. Lastly, launch volume-based DDoS using the Radware tool. [18]. Table III provides a breakdown of the dataset's class labels. In an 80:20 split, we've separated the dataset into train data and test data.

TABLE III. NUMBER OF RECORDS OF CLASS LABELS

Data	Brute Force	HTTP DDoS	ICMP Flood	Normal	Port Scan	Web Crawling
Train Data	70642	523	35	22900	8916	23
Test Data	17860	118	10	5602	2165	5
Total	88502	641	45	28502	11081	28

B. Evalutaion Metrics

All of the experimental outcomes of this work have been included in well-established evaluation metrics. A brief explanation of each of these metrics has been provided below.

Accuracy: Accuracy is the ratio of samples that have been properly identified to the total number of samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

Precision: Precision measures a model's ability to correctly classify positive values.

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Recall: Recall is measurement of correct positive predictions divided by total positives.

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

F1 Score: The F1-score is a harmonic mean of the recall and accuracy scores.

$$F1 = \frac{2*Recall*Precision}{Recall+Precision} \quad (10)$$

Training Time: This is the duration of the models' training phase.

Testing Time: This is the duration consumed by the models' testing phase.

C. Experimental Results Analysis

With increasing dimensionality, the number of data points required for a classifier to perform well increases exponentially. Since the excess data points provide the classifier better ability of analyzing data and hence find the desired pattern from it. Numerous research has shown that the number of training sample dimensions increases the prediction power of any classifier. However, after a given period, performance drops. Which is known as the "curse of dimensionality" problem. In order to minimize the model's complexity and enhance its performance, PSO has been utilized to reduce the number of data dimensions. Where the

primary motive is to prevent computational inefficiency while maintaining classification performance.

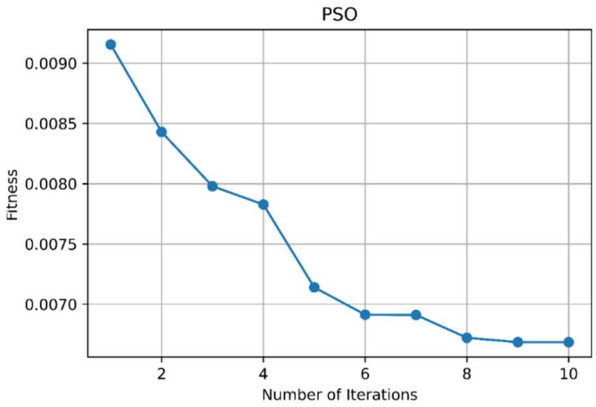

Fig. 3. Iteration vs Fitness curve

Graphical representation in Fig. 3 depicts the fitness value of global best or gbest to the iteration of particle fitness evaluation within the train and test dataset. For selecting the best subset of features from the dataset, 18 features have been selected upon completing ten iterations with a 72.23% feature reduction rate.

The most important factor in determining a model's efficiency in detecting cyber-attacks depends on how well it performs on evaluation metrics. Evidently, higher values for Accuracy, Precision, Recall, and F1-Score demonstrates the prominent effectiveness of the classifier.

The corresponding confusion matrix resulting from classification with the optimum feature selected by the PSO algorithm is shown in Fig. 4.

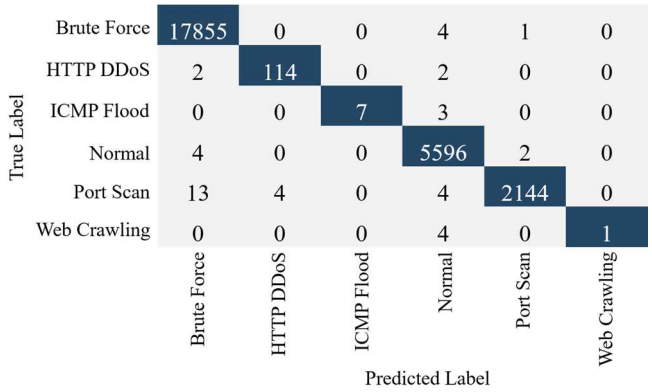

Fig. 4: Confusion Matrix of multiclass classification with reduced features

Table IV presents the values of the evaluation metrics for the experiment. Except for the times spent in training and testing phase, which are represented in seconds, all other numbers are shown in percentages. As per our prediction, all the metrics for the dataset demonstrate superiority of the deep learning model with reduced feature subsets by outperforming the same model with actual features. The PSO-based technique creates a smaller dataset with an ideal feature subset, which streamlines the model's structure and uses less time during training and testing. The results show that the four metrics (Accuracy, Precision, Recall and F1Score) were enhanced by roughly 1% in the deep learning model with PSO's optimal feature subsets compared to the equivalent values of the same model without PSO. Additionally, training and testing time is about three times and two times lower for reduced features than for entire features.

TABLE IV. EXPERIMENTAL RESULTS FOR DIFFERENT CLASSIFICATION

Metric (%)	Full Features (66)	Reduced Features (18)
Accuracy	98.73	99.83
Precision	98.71	99.88
Recall	98.73	99.83
F1 Score	98.72	99.85
Training time (sec)	12879	4032
Testing time (sec)	855	314

From the confusion matrix, we have calculated the detection rate of each class which includes normal and five attack categories. In Table V the detection rate, also called recall or true positive rate (TPR) of each class has been shown as percentage. Due to having comparatively fewer amount of data in the training set, a relatively degrading performance has been observed for ICMP Flood and Web Crawling compared to other class labels.

TABLE V. DETECTION RATE OF EACH CLASS

Class Name	Detection Rate (%)
Brute Force	99.97
HTTP DDoS	99.61
ICMP Flood	70.00
Normal	99.89
Port Scan	99.03
Web Crawling	20.01

Fig. 5 represents the results of an analysis of how well the suggested deep learning scheme performs in comparison to the performance of the conventional ML and Deep Learning classifiers. Here the performance has been evaluated against five well-known ML methods: Support vector Machine (SVM), K-Nearest Neighbor (KNN), Random Forrest (RF), Decision Tree (DT), and Naïve Bayes (NB). Besides these conventional classifiers, the performance of DNN (Deep Neural Networks) along with RNN (Recurrent Neural Networks) classifiers have been evaluated where DNN has been created with sequential dense layers and RNN with SimpleRNN layers. Observational data demonstrate that the proposed deep learning model outperforms practically all assessment criteria with higher result values.

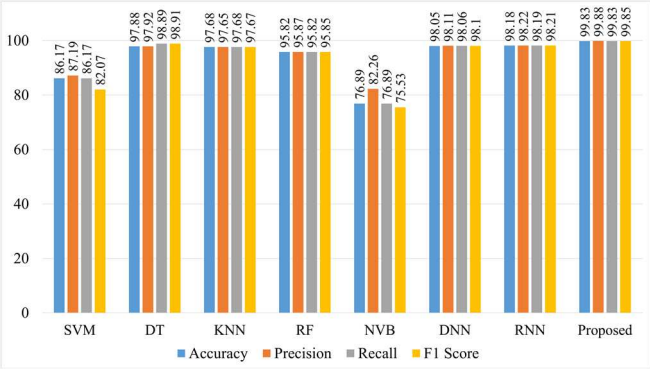

Fig. 5. Comparative analysis between different classification schemes

V. Conclusion

In recent years, numerous forms of cyberattacks including different multi-stage attacks have been growing at an exponential rate alongside the proliferation of both applications and networks. Therefore, it is becoming more vital to find a suitable Intrusion Detection System (IDS) solution for safeguarding networks and devices. In this study, we propose a deep learning-based classification model in order to detect multi-step cyber-attacks within an IDS dataset. Multi-Step Cyber-Attack Dataset (MSCAD), which is a new benchmark, has been utilized. Here, the primary architecture of the detection model is comprised of a sequence of LSTM, attention, and two back-to-back dense layers. PSO has been used to draw out the optimal subset of features from the dataset for ensuring both functionality and reliability of the proposed approach and reducing operational complexity and time consumption. The experimental findings demonstrate that our approach outperformed models without a pre-processing phase in terms of network intrusion detection. Hence, we think that the approach that was suggested is an effective system for solving the problem of cyber-attack detection.

References

[1] J. Navarro, A. Deruyver, and P. Parrend, “A systematic survey on multi-step attack detection,” *Comput. Secur.*, vol. 76, pp. 214–249, 2018.

[2] M. E. Karim, M. M. S. Maswood, S. Das and A. G. Alharbi, "BHyPreC: A Novel Bi-LSTM Based Hybrid Recurrent Neural Network Model to Predict the CPU Workload of Cloud Virtual Machine," in IEEE Access, vol. 9, pp. 131476-131495, 2021.

[3] S. M. Kasongo and Y. Sun, “A deep learning method with filter based feature engineering for wireless intrusion detection system,” IEEE *Access*, vol. 7, pp. 38597–38607, 2019.

[4] C. Yin, Y. Zhu, J. Fei, and X. He, “A deep learning approach for intrusion detection using recurrent neural networks,” *IEEE Access*, vol. 5, pp. 21954–21961, 2017.

[5] F. Laghrissi, S. Douzi, K. Douzi, and B. Hssina, “Intrusion detection systems using long short-term memory (LSTM),” *J. Big Data*, vol. 8, no. 1, 2021.

[6] A. Yulianto, P. Sukarno, and N. A. Suwastika, “Improving AdaBoost-based intrusion detection system (IDS) performance on CIC IDS 2017 dataset,” *J. Phys. Conf. Ser.*, vol. 1192, p. 012018, 2019.

[7] H. Yang, G. Qin, and L. Ye, “Combined wireless network intrusion detection model based on deep learning,” *IEEE Access*, vol. 7, pp. 82624–82632, 2019.

[8] X. Zhang, J. Ran, and J. Mi, “An intrusion detection system based on convolutional neural network for imbalanced network traffic,” in *2019 IEEE* 7th *International Conference on Computer Science and Network Technology (ICCSNT)*, 2019.

[9] I. Ahmad, “Feature selection using particle swarm optimization in intrusion detection,” *Int. J. Distrib. Sens. Netw.*, vol. 2015, pp. 1–8, 2015.

[10] W. Elmasry, A. Akbulut, and A. H. Zaim, “Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic,” *Comput. netw.*, vol. 168, no. 107042, p. 107042, 2020.

[11] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, “BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset,” *IEEE* Access, vol. 8, pp. 29575–29585, 2020.

[12] P. B. Udas, M. E. Karim, and K. S. Roy, “SPIDER: A shallow PCA based network intrusion detection system with enhanced recurrent neural networks,” Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 10, pp. 10246–10272, Nov. 2022.

[13] K. S. Roy, Md. E. Karim, and P. B. Udas, “Exploiting Deep Learning Based Classification Model for Detecting Fraudulent Schemes over Ethereum Blockchain,” 2022 4th International Conference on Sustainable Technologies for Industry 4.0 (STI), Dhaka, Bangladesh, 2022, pp. 1-6.

[14] B. Xue, M. Zhang, and W. N. Browne, “Particle swarm optimisation for feature selection in classification: Novel initialisation and updating mechanisms,” *Appl. Soft Comput.*, vol. 18, pp. 261–276, 2014.

[15] M. E. Karim and S. Ahmed, "A Deep Learning-Based Approach for Stock Price Prediction Using Bidirectional Gated Recurrent Unit and Bidirectional Long Short Term Memory Model," 2021 2nd Global Conference for Advancement in Technology (GCAT), 2021.

[16] A. Vaswani *et al.*, “Attention is all you need,” 2017.

[17] M. A. Almseidin, J. A.-S. Al-Sawwa, and M. A. Alkasassbeh, “MSCAD.” IEEE DataPort, 18-Jun-2022.

[18] M. Almseidin, J. Al-Sawwa, and M. Alkasassbeh, “Generating a benchmark cyber multi-step attacks dataset for intrusion detection,” *J. Intell. Fuzzy Syst.*, vol. 43, no. 3, pp. 3679–3694, 2022.