Polytechnic Journal. 2022. 12(2): 1-7 100-113
ISSN: 2313-5727
http://journals.epu.edu.iq/index.php/polytechnic

POLYTECHNIC JOURNAL

RESEARCH ARTICLE

Anomaly-based Network Intrusion Detection System using Deep Intelligent Technique

Sardar KH. Hassan[1], **Muhammadamin A. Daneshwar**[2]

[1] *Department of Computer Science, Soran University, Erbil, Kurdistan Region, Iraq*

[2] *Department of Computer Science, Soran University, Erbil, Kurdistan Region, Iraq*

Corresponding author:
Sardar KH. Hassan,
Department of Computer Science,
Soran University, Erbil, Kurdistan Region, Iraq.

E-mail: skh210h@cs.soran.edu.iq

Received: 18 Jun 2022
Accepted: 7 Sep 2022
Published: 1 February 2023

DOI
10.25156/ptj.v12n2y2022.pp100-113

ABSTRACT

Background and objectives: Computer systems and network infrastructures are still exposed to many security risks and cyber-attack vulnerabilities despite advancements of information security. Traditional signature-based intrusion detection systems and security solutions by matching rule-based mechanism and prior knowledge are insufficient of fully protecting computer networks against novel attacks. For this purpose, Anomaly-based Network Intrusion Detection System (A-NIDS) as cyber security tool is considered for identifying and detecting anomalous behavior in the flow-based network traffic alongside with firewalls and other security measures. The main objective of the research is to improve the detection rate and reduce false-positive rates of the classifier using anomaly-based technique.

Methods: an intelligent technique using deep learning algorithm and mutual information feature selection (MIFS) method to select optimal features on the benchmark datasets. Proposed method accurately capable of classifying normal and anomalous states of the data packets in a comprehensive way by combination of Long-Short term memory (LSTM) algorithm and MIFS method.

Results: The model achieved encouraging results in terms of accuracy 99.79%, 0.002 false-positive rate with minimum time compared to other models recorded only 81.75s on CSE-CIC-IDS2018 dataset. At the end of the study, comparative studies are conducted to verify the effectiveness of proposed method on three realistic and latest intrusion detection datasets, named CSE_CIC-IDS2018, CIC-IDS2017, and NF-CSE-CIC-IDS2018 dataset.

Conclusions: Proposed model in a combination of LSTM NN and Feature selection method (MIFS) increased detection rate and reduced false-positive alarms, also the model able to detect low frequent attacks while other existing models are suffering from.

Key Words: Intrusion Detection System, Anomaly detection, Intelligent Technique, Cyber-Attack, Deep Learning, Machine Learning

1. INTRODUCTION

The development of global interconnected networks plays a critical role in expanding digital community and simplify different daily routine such as online transactions, personal communications, digital affairs, etc. With the emerging technology protecting and confidentiality of the data is become challenging tasks need to be considered rigorously. However, despite technological advancements, information security and privacy of the data are faced many security risks and vulnerabilities as well. Network intruders continuously endeavor to breach into computer systems and network infrastructures via bypassing firewalls and other security solutions to

unauthorized access. Cyber-attacks imposed many security risks and left devastated consequences behind. On other hand, implementing security principles like data encryption, digital certificates, authentication, and firewall are not provide sufficient guarantee. Therefore, Intrusion Detection System (IDS) as a dominant security tool introduced by James P. Anderson in 1980s to strengthen the security level by identifying attacks on the systems. Distributions of the IDS systems such as Host-Based IDS (HIDS), Network-Based IDS (NIDS), Protocol-Based IDS (PIDS) and Hybrid-Based IDS (HIDS) are constructed to implemented in many fields. NIDS systems are designed to recognize security threats in the network infrastructures. However, due to diversity of network attacks IDS systems are suffering from many challenging issues.(Ambusaidi *et al.*, 2016)

Traditional rule-based techniques highly depending on prior knowledge, matching audit record, and known signatures are unable to identify novel attacks (Peddabachigari *et al.*, 2007). Updating rule-set and optimizing thresholds are required security experts and time-consuming. In addition, cyber criminals always change their ways of attacking and designing novel attacks. Thus, intelligent techniques by applying strong hypothetical, mathematical calculations and effective feature engineering techniques come up to tackle the common problems and limitations as possible.

Intelligent NIDS systems are capable of classify normal and abnormal activities in the network. Deviations from out of normal range are alerted as an attack(Kanimozhi and Jacob, 2020). Most of the existing NIDS systems are signature-based which are highly effective against known threats, but fails if any unknown threats are occur. In this study, we are investigating an effective technique to detecting anomaly-based network intrusions detection technique using deep learning neural networks. LSTM and mutual information feature selection (MIFS) technique are utilized to recognize anomalous activity in the network. Main objective of the study is to improve detection rate and reducing false-positive alarms.

1.1 MOTIVATION

In spite of development of security trends and data protection technique, still organizations and personal peers are threatened by intruders over internet. This development is trade-off which give better opportunity to hackers to stealing sensitive data and breaches into systems any time. Are existing signature-based techniques being sufficient and provide excellent performance of novel attacks or not? What happens when a novel or zero-day cyber-attack is launched? Currently there is a large movement in the intelligent techniques to recognizing cyber-attack and improve current methods in terms of accuracy and false positive rates. A well-designed model for anomaly-based network intrusion detection system is a crucial improvement over the current advancements that responses of detect zero-day attacks and prevent malicious activity inside computer systems. However, several problems including low detection rate and high false-positive with the current case of anomaly detection have been identified. Therefore, recently intelligent techniques using statistics and strong mathematical theory provide high performance and the results are satisfactory.

1.2 CONTRIBUTIONS

The majority of researches do their experiments on a portion of the dataset or train their models using only separate csv files from the dataset that only contain certain types of attack and benign traffic. Additionally, including the entire dataset into our model during training enables it to learn how to separate particular attack types from others. When we evaluate the results and contrast them with existing researches, this approach is thought to be the most effective at detecting intrusions. On the other hand, some attack classes have the trait of having much lower sample percentages than the others. We have not applied over-sampling technique to low frequency attacks like SQL Injection, Web filtering, and infiltration to be more nature and represent realistic attack. Instead, we left the data in its original forms, and the detection is encouraging even for low frequency attacks. In addition, as mentioned in the literature section that most related approaches used the deprecated KDDCup99 and NSLKDD datasets, which did not represent recent attacks. We made advantage of the most recent benchmark dataset that captured from real network flow using CICFlowMeter tool to be more realistic and representative. Also, we combined six separate sub-datasets from different dates for better consistency of the model. Another significant problem that most previous researches have is the use of binary classification to assess the success of the model. Our model is accurate in recognizing different attack classes with low proportions. In order to provide more accurate findings for our research, we incorporated additional measures including accuracy, precision, FPR, and ROC-AUC.

1.3 Objectives of the research

The main objective of the research is to improve the detection rate and reduce false-positive rates of the classifier using anomaly-based technique. Also, introducing efficient feature selection technique such as Mutual Information Feature Selection (MIFS) to select optimal features for time and computational complexity reductions. Some of the existing models are suffering from recognizing rare attacks due to low proportions in the real world. On the other hand, novel attacks are not recognized by traditional signature-based intrusion detection systems such as Snort and Suricata. This research is a kind of effective way to find anomalies and malicious behavior amongst the data samples.

The rest of the paper is organized as follows: The related works is provided in Section 2. Section 3 describes the proposed methodology. Section 4 describes and discusses experimental setups and results. Section 5 Finally, the conclusion and summary about approaches and works mentioned in the study is given.

2. Related works

The advancements in information technology and cyber community are appreciated, however exposes many security risks. Intrusion detection systems (IDSs) as security tools are hot topic for many researchers and security consultants recently. Many approaches have been proposed to design an effective model to accurately detecting malicious activity in the network. Most recent studies conducted in the field are outlined in this regard:

Rosay, et al presented an intelligent IDS technique using Multi-layer Perceptron (MLP) on two recent datasets. The researchers preferred flow-based data packets instead of packet-based, but could lead to some limitations due to huge vast number packets in gigabit network throughput. The proposed model provided high performance in terms of accuracy 95.4% and false positive rate 0.67% compared to other machine learning algorithms such as Decision tree (DT) and Random Forest (RF). In addition, the authors indicated that the model performance on CIC-IDS2017 is worse than CSE-CIC-IDS2018 dataset due to highly imbalanced limitation in the dataset and lack of data instances for infiltration and SQL injection.(Rosay *et al.*, 2021).

Moualla and his team presented an approach to improve the performance of machine learning-based NIDS to detect novel attacks on the Australian UNSW-NB15 benchmark dataset. The study primarily concentrated on designing robust model to provide higher detection and lower false alarms. initially, SMOTE is used to oversampling classes imbalances whose percentages are less than 2% including (analysis, backdoors, shellcode, worms). Second, extremely randomized tree classifier and extreme learning machine (ELM) are used to extract most important features. The results are satisfactory and the performances are improved up to 98% overall, while reducing the FPR for the lowest rate to 0.015.(Moualla, Khorzom and Jafar, 2021).

In another study, Zhou, et al built an efficient IDS based on ensemble classifier and feature selection technique. The authors declared high dimensionality of instances, deficient performance of individual classifiers, and lack of adaptable models for novel attacks are the primary reasons to build an efficient model. In addition, the technique of voting (use power of several classifiers to build an appropriate decision) is responsible for attack identification based on probability distribution strategy. proposed CFS-BA outperformed other related deep learning models (DEMISe and HELAD), some existing ML algorithms, and even state-of-art models. However, the proposed framework provides less performance on unseen datasets including KDDTest+ and KDDTest-21 prepared for ML algorithms.(Zhou *et al.*, 2020).

(Yin *et al.*, 2017) presented a deep learning approach for NIDS based on recurrent neural networks (RNN-IDS) using full feature set. Researchers extended total 41 features into 122 dimensional features for better understanding the nature of the attacks. Experimental results illustrated excellent accuracy and low false-positive rates. However, evaluation measures dramatically degraded on KDDTest+ and KDDTest-21 which are completely unseen subset of KDD dataset.

De la Hoz and his colleagues proposed an ensembled classification model by combining principal component analysis (PCA) and Probabilistic self-organized map (PSOM) to detect anomalous behavior in the network on the NSLKDD benchmark dataset. The results demonstrated the effectiveness of ensemble PSOM+PCA+FDR method to select 8 features out of 41. The model achieved best performance among existing discriminative feature engineering methods in terms of sensitivity, accuracy, and specificity.(De la Hoz *et al.*, 2015)

Ambusaidi, et al concentrated on an ensemble approach of mutual information (MI) algorithm and Least Square support vector machine (LS-SVM)

which effectively discriminate normal and abnormal data packets. the proposed approach outperformed many state-of-art models. The results and comparisons are promising in terms of low computational cost, less processing time, and high classification rates. Empirically the proposed model achieved up to 99.77% detection rate overall on KDDCUP99, NSL-KDD, and Kyoto 2006+ datasets. In addition, the model obtained superiority over state-of-art methods in U2R and R2L attack category by achieving 22.11% and 88.3, respectively.(Ambusaidi *et al.*, 2016).

(Vinayakumar *et al.*, 2019) developed a Network and Host-based IDS scalable framework called Scale-Hybrid-IDS-AlertNet to identify unknown and novel cyber-attacks. For better recognition of low frequent attacks like U2R and R2L more neurons are allocated per each layer, while high frequent attack type like DOS allocated less. Empirical results show satisfactory performances on KDDCUP99, CICIDS2017, and WSN-DS datasets, however the results were degraded for Kyoto dataset and LKDD and UNSW-NB15 recorded the worst performance for both binary and multi-class classification.

In another recent study, (Al-Daweri, Abdullah and Zainol Ariffin, 2021) proposed a homogenous ensemble method for adaptively detect novel attacks on UKM-IDS20 dataset. The ensemble method comprises of four classifiers to detect four types of attack class in the dataset. The authors stated self-adaptive IDS method is crucial in protecting network infrastructure from any novel attacks. The results on proposed HOE-DANN adaptive method achieved high detection rate 96%, accuracy 94% in testing set, while suffering from high false positive 7.57%.

(Oliveira *et al.*, 2021) proposed an approach using Long-Short Term Memory (LSTM) to detect anomalous behavior in the CIDDS-001 dataset. The proposed LSTM outflanks Multi-layer Perceptron (MLP) in a comparative study. The authors stated, LSTM is reliable and has strong learning capabilities to get relationships of sequential data packets. At last, the model performance compared with random Forest (RF) and (MLP) and achieved higher results in term of f1-score and precision for multi-flow, while RF gets better score in single-flow.

3. Methodology

This section contains all implementation and experimental setups for anomaly-based network intrusion detection system includes the dataset description, data preprocessing, proposed method, and comparative studies. The model trained by Lenovo ThinkPad laptop with hardware capabilities of 8 Gigabyte of RAM, Core i7 2.60GHz CPU, and 64-bit windows 10 home operation system. All experimental results conducted on google Collab python 3 environment.

3.1 Data Collection

Data collection for intrusion detection system is counted as an initial and essential step. NIDS systems are trained with latest labelled dataset that captured via TCPdump on the real network flow. The captured data connections from TCPdump are saved as Pcap file then via CICFlowMeter4.0 network analysis tool are converted to dataframe with 81 attributes. Proposed dataset includes most realistic and latest attack connections.

3.2 Data Preprocessing

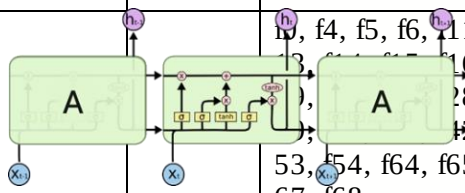
The collected data from real traffic flow contains raw data including basic features requires further preprocessing and preparation before fetched into ML models as described below:

3.2.1 Data Conversion

Thus, every symbolic and string characters in the dataset should be converted into numerical values. IDS datasets are including the symbolic features such as protocol types (e.g., TCP, ICMP and UDP) and network services such as (e.g., FTP, Telnet, HTTP, SSH, and so on) need to be converted from categorical into numerical forms. In this case, protocol types are considered as nominal data. CICflowmeter is a software used to establish csv file from dumped TCP raw captured data during network connections. Thus, Protocol types are automatically converting into integer numerical values using label encoding inside CICflowmeter. Label encoding is useful for non-ordered case sensitive categorical data by alphabetically assigning each category to an integer number.

3.2.2 Data Normalization

Existing dataset contains different attributes in different scales which are highly affect the model performance. In this case, normalization as preprocessing phase provides flexibility and preserves smooth relationships between different data scales in ML and DL models. Therefore, min-max normalization method is applied to lies features in a more reliable and suitable range of values between

Techniques	# Features	Selected features
MIFS	15	f69, f66, f1, f35, f0, f19, f17, f2, f16, f37, f38, f67, f40, f24, f47
		 f0, f4, f5, f6, f11, f12, f 13, f14, f15, f16, f18, f 19, f20, f28, f37,f 41, f42, f48, f 53, f54, f64, f65, f66, f 67, f68
CFS	39	f4, f6, f10, f11, f12, f1 4, f19, f20, f21, f22, f2 3, f24, f25, f29, f30, f3 5, f36, f37, f39,f40, f4 1, f42, f43, f45, f50, f5 1, f53, f54, f55, f62,f6 3, f64, f65, f67, f72, f7 3, f74, f76, f77
FFS	20	f0, f4, f10, f28, f30, f3 1, f33, f34, f38, f40, f4 2, f51, f57, f58, f59, f62, f64, f66, f67, f68

interval range of [0, 1]. The Min-Max scaler of the normalization function is illustrated as in eq. (1):

$$\text{Xnormalized} = \frac{\text{X} - \text{Xmin}}{Xmax - Xmin} \qquad (1)$$

3.2.3 Feature selection

Feature engineering as an essential preprocessing task provide some advantages including computational cost reduction, less time, and enhancing performance of the model. This study selected the most influential feature selection technique among different types of feature selection techniques such as Mutual Information Feature Selection (MIFS), Importance Feature Selection (IFS), Correlation-based Features Selection (CFS), and Forward Feature Selection (FFS) after conducting a comparative performance analysis for CSE-CIC-IDS2018 and CIC-IDS2017 datasets. Experimental results show MIFS outperform other techniques by electing only 15 optimal attributes out of 81. Thus, we employed MIFS entropy-based feature selection techniques via selecting maximum score of the features. Table 1 shows the results of selecting optimal features according to different techniques.

Table 1: Feature ranking results on the CSE-CIC-IDS2018

In this section, we discussing about proposed method using deep neural network algorithm called LSTM which is special kind of Recurrent Neural Network (RNN) and MIFS feature selection technique to identify network intrusions with high detection and low false-positive rates.

Figure 1: Simple architecture of LSTM network

3.3 Proposed model

In this research, we proposed LSTM as a classifier to distinguish between normal and abnormal data packets. As we know network flow is sequential and regulated as time series manner, therefore, we selected LSTM due to its reliability, fast training ability and strong memory capabilities to recognize previous state and relationships of sequential patterns in data packet. LSTM is a special kind of RNN, introduced by Hochreiter and Schmidhuber in 1997, designed to avoid the long-term dependency issues. Unlike RNN, LSTM can memorize and learn data for long periods. RNN architecture consists of hidden layers have a simple structure composed of a single tanh layer, while the LSTM architecture is constituted of 4 hidden layers and more complex. Cell state is the key principle of LSTM. Sigmoid function is used to add or remove information from the cell or protect current state of the gate. Inner structure of LSTM is consisting of many gates can control the incoming states of the input data described as below (Debasish Kalita, 2022):

- **Forget gate layer**: input and previous hidden layer data will be monitored to decide which information is going to remove from the cell state via sigmoid function, in case of one it keeps it while 0 means delete it. Calculation formula is as in eq. (2): $f_t=\sigma(W_f.[h_t-1,x_t]+b_f)$ (2)
- **Input or update gate layer**: refers to storing data in the cell state. Initially information of input layer will be updated via sigmoid function, then new vectors added via a Tanh layer to the cell state. Thus, LSTM decide about to forgot the information or updating as a new vector according to eq. (3) and (4):

$$i_t=\sigma(W_i.[h_t-1, x_t] + b_i) \qquad (3)$$
$$\hat{C}t=\tanh(W_c.[h_t-1,x_t] + b_C) \qquad (4)$$

- **Output Layer**: by executing sigmoid function decides which part of the cell LSTM is going to output and the result is crossed through a Tanh

Class	Label Encoder
Benign	0
DoS attacks-GoldenEye	4
DoS attacks-Hulk	5
Bot	1
FTP-BruteForce	8
SSH-Bruteforce	11
DoS attacks-Slowloris	7
DoS attacks-SlowHTTPTest	6
Infilteration	9
Brute Force-Web	2
Brute Force-XSS	3
SQL Injection	10

layer to output only the information we decide to pass to the next neuron. It is calculated as eq. (5):

$$O_t=\sigma(W_o[h_t-1,x_t]+b_o) \text{ and}$$
$$h_t=O_t*\tanh(C_t) \quad (5)$$

The proposed model consists of 4 dense layers including input and output layer. The designed model gives enough layer and neurons to effectively classify normal and abnormal status of the data packets also excellent detection on different attacks category such Dos, Portscan, Infiltration, WebAttack and so on. Hidden layers are also constituted of 128 and 64 neurons for better performance and reducing computational time perspectives.

Figure 2 shows general framework of the proposed NIDS model in the study, illustrating preprocessing phases, input layer, hidden layers and finally output layer for multi-class classification. The number of input layers and output neurons are fully customizable according to multi-class, binary classification or feature selection methods.

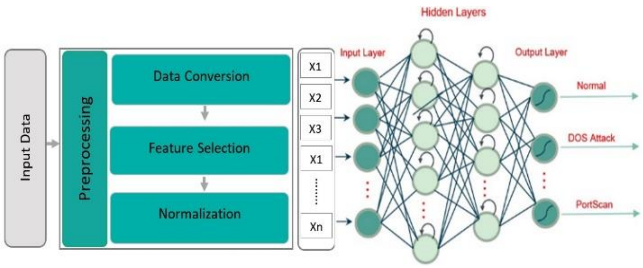


Figure 2: Illustrate the framework of the proposed NIDS model

3.4 Description of the Benchmark Datasets

Lack of representative dataset is one of the common issues for many researchers. Training intelligent models on different dataset containing multiple attacks dramatically increase the learning process and provide generalization capability for the NIDS models. Several public benchmark datasets mentioned in the study like, KDDCUP99 generated by DARPA in 1998. It contains different attacks such as Neptune-DoS, pod-DoS, SmurfDoS, and buffer-overflow, but suffering from many duplicate data instances. NSLKDD which is a revised version of kDDCUP99 dataset suggested to solve some of the inherent problems of the KDD'99 dataset and is one of the most used datasets due to balancing and class distribution of the dataset. The main issue in the dataset is low frequency for U2R and R2L category class.(Govindarajan and Chandrasekaran, 2011).

Table 2: CSE-CICIDS2018 class labels

The CIC-IDS2017 dataset was published by Canadian Institute for Cybersecurity in 2017 contains 2,830,743 records of 78 features, which covers necessary criteria with updated attack categories such as DDoS, Brute Force, XSS, SQL Injection, Infiltration, Port Scan, and Botnet. It is an up-to-dataset but contains many missing records and highly imbalanced class distribution. [5] Also, Australian Centre for Cyber Security created another dataset based on raw network packets named UNSW-NB15 in 2015. It contains nine different attacks, includes DoS, worms, Backdoors, and Fuzzers; comprised of 2,540,044 realistic modern normal and abnormal network activities. However, this dataset has two key issues, class imbalance and class overlap.(Zoghi and Serpen, 2021).

The UKM-IDS20 dataset is one of the latest realistic datasets for intrusion detection systems developed by Kebangsaan University of Malaysia in 2020. The dataset was collected and generated using real-world network traffic flow that contains 46 features and covers four types of attacks, namely ARP poisoning, DoS, Scans, and Exploits. However, the dataset only contains 4 types of attacks and labeling class records are done manually which are counted as the common imitation of the dataset.(Al-Daweri, Abdullah and Zainol Ariffin, 2021).

In addition, another newly published dataset represents more universal NIDS datasets containing flows from multiple network setups and different attack settings. The attack categories have been modified to combine all parent categories such as Dos category (DoS Attacks-Hulk, DoS attacks-SlowHTTPTest, DoS Attacks-GoldenEye and DoS attacks-Slowloris, DDos category (DDOS attack-LOIC-UDP, DDOS attack-HOIC and DDoS attacks-LOIC-HTTP), and finally, SQL Injection attacks located under injection attacks category.

The NF-CSE-CICIDS2018 dataset has a total of 11,994,893 records, out of which 9,208,048 (76.77%) are benign flows and 2,786,845 (23.23%) are attacks. Real network traffic flow via CICFlowMeter tool is directly converted from TCPdump to the csv file. Then the csv file with 10 features used as a comprehensive dataset for machine learning and data mining analysis.

CSE-CIC-IDS2018 is a popular dataset created in a collaborative project between the Communications Security Establishment (CSE) & the Canadian Institute for Cybersecurity (CIC) using the Amazon Web Services platform in 2018. The dataset enhanced in consideration with the standards of CSECIC IDS2017 contains about 16,000,000 instances. The captured network packets are separated on 8 different csv file including different attack classes. The main issue is highly imbalanced class distribution of the attacks in the dataset, which need to be considered before applying any model.(Ravikumar, 2021).

In this study, we used CSE-CICIDS2018 dataset as primary data source for anomaly-based NIDS detection. For the sake of use and better analysis to recognize all attack classes, we established a new dataset by combining all 8 documents and decrease the normal packets proportions to be easily processed by implementation software. The combined dataset refers different attack categories such as (Dos, DDos, Bot, BruteForce, infiltration, and SQL injection) as shown in table2. At the end of the research a comparative study applied for the performance and detection of different attack on CSE-CIC-IDS2018, CIC-IDS2017 and most up-to-date dataset for intrusion detection systems named NF-CSE-CIC-IDS2018 which contained only 10 attributes for analysis.

Figure 3 illustrate the percentage distribution of benign and attack classes frequency after assigning percentage ratio for each class in order to show class imbalance in the dataset. Despite low frequency for some attack classes such as SQL Injection, brute Force-XSS, Brute Force-Web, and Infilteration the model could detect them accurately. For binary classification all class labels are assigned as 1 for attack and 0 indicated for benign data samples. CSE-CIC-IDS2018 dataset contains high proportions of benign samples comparing to other classes.

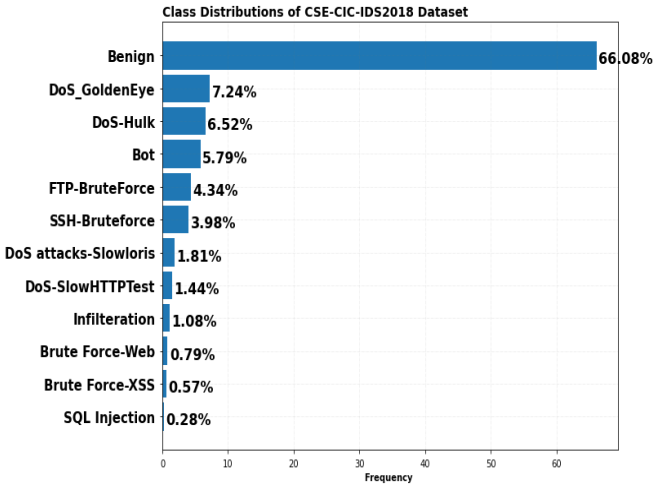


Figure 3: class distribution of CSE-CIC-IDS2018

4. Experimental Setups and Results

This section demonstrates all the experimental setups and data analysis for presenting the performance of the proposed model in classification cyber-attack based on most realistic datasets including (CIC-IDS2017, CSE-CIC-IDS2018, and NF-CSE-CIC-IDS2018). In this section, we conducted binary and multi-class classification for the aforementioned datasets. LSTM as a deep learning NN and several feature selection techniques are used to assess the proposed model in terms of accuracy, detection rate, and time-consuming. As a preprocessing technique, feature selection techniques such as MIFS, IFS, CFS, and FFS are used to extract optimal features. MIFS technique achieved optimal subset of attributes with less time than other techniques also, provide better understanding of the different attack types. Other feature selection techniques are good but suffering from lack of identifying some attacks and exceeded in time. Deep learning algorithms require huge data instances to provide better understanding of all latent patterns and finding strong relationships between features which leads to show higher performance. Proposed model with original features achieved high performance but took more time which is kind of critical aspect during real-time predictions.

In addition, determining number of layers and neuron per each layer is highly efficient for the classifier in terms of better understanding of the instances and computational time complexity Adam and rmsprop with 0.001 learning rate are used as optimization algorithm which achieved optimal solution during training the dataset, the performance of Adam is little higher than RMSProp. The number

of hidden layers, neurons per each layer, learning rate, and total epochs highly affect the efficiency of the model.

4.1 Performance Evaluation

Several evaluation metrics have been conducted to assess the effectiveness and performance analysis of the proposed classifier. In this case, the accuracy, detection rate, misclassification metrics (false-positive rate and confusion matrix), and roc curves are as the principal components of supervised evaluation metric are accomplished to estimate the overall performance during testing and validation phases. K-folded cross validation which is one of the important metrics has been conducted to estimate the overall efficiency of the model. Evaluation metrics are defined as follows in eq. (6) to (8):

TP: Attack samples that are truly identified as attack
FP: Normal samples are classified as attack
TN: Normal samples are truly identified as normal
FN: Attack samples are identified as normal

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (6)$$

$$Detection\_Rate = \frac{TP}{TP+FN} \quad (7)$$

$$False\_Positive = \frac{FP}{FP+TN} \quad (8)$$

Confusion Matrix calculate the total classified and misclassified samples in each class.

In addition, Receiver Operating Characteristics (ROC) curve is kind of visualization concept as binary classifier system to illustrate the true positive and false positive ratio on X-axis and Y-axis at each points respectively. One common measure with the ROC curve is the area under the curve (AUC) with values between [0, 1]. Higher AUC (more than 0.5) measures how well-trained classifiers are by allocating higher probability for correct predictions and lower probability for incorrect ones. A badly trained classifier has a diagonal line ROC curve with AUC close to 0.5

4.2 Results and Discussions

This section discusses all experimental and evaluation results for proposed method using the aforementioned datasets. Three experiments were conducted to evaluate the effectiveness of the proposed method. According to the table 3 and table 4 evaluation metrics such as accuracy, FPR, and time are presented to estimate the proposed method based on different feature selection techniques for binary and multi-class classification.

Table 3: Performance of the methods on CSE-CIC-IDS2018 (Multi-class)

	CSE-CIC-IDS2018			CIC-IDS2017		
Method	**Accuracy**	**FPR**	**Time**	**Accuracy**	**FPR**	**Time**
LSTM+ Original features	99.8	0.001	101.53	98.5	0.014	143.54
LSTM + MIFS	99.7	0.002	64.82	97.9	0.020	63.92
LSTM + IFS	99.5	0.004	82.93	98.2	0.017	65.44
LSTM + CFS	99.3	0.006	80.06	96.6	0.033	83.89
LSTM + FFS	99.4	0.005	72.18	94.9	0.050	82.99

Table 4: Performance of the classifiers on CSE-CIC-IDS2018 (Binary classification)

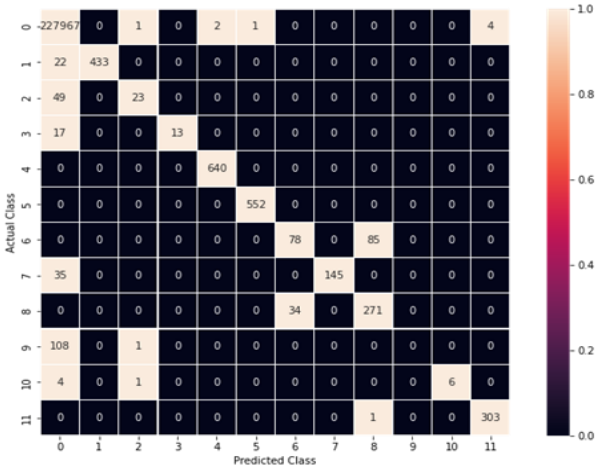
	CSE-CIC-IDS2018			CIC-IDS2017		
Method	**Accuracy**	**FPR**	**Time**	**Accuracy**	**FPR**	**Time**
LSTM + original features	99.88	0.001	112.72	98.82	0.011	79.92
LSTM + MIFS	99.88	0.001	66.37	98.28	0.017	62.66
LSTM + IFS	99.51	0.004	64.51	98.67	0.013	83.52
LSTM + CFS	99.42	0.005	76.80	96.62	0.033	84.03
LSTM + FFS	99.52	0.004	66.98	98.24	0.017	63.90

Table 3 shows the performance analysis of the proposed method deploying with several feature selection techniques for multi-class classification. Performance analysis of the classifier estimated

according to accuracy, false-positive rate and time-consuming. Proposed LSTM + MIFS with accuracy 99.7%, 0.002 FPR, and 64.82 seconds on CSE-CIC-IDS2018 dataset outperform others in terms of time and accuracy, except LSTM with original features which is very close. Proposed method with original attributes achieved better results on both datasets compared to other methods but requires more time which is critical. Proposed model with MIFS technique provides less time on both CSE-CIC-IDS2018 and CIC-IDS2017 dataset.

4.3 Recognition of Attacks

As demonstrated in figure (4, 5) proposed method accurately classify different types of attacks according to the multi-class confusion matrix recoding accuracy of 99.8% and 97.2% on both CSE-CIC-IDS2018 and CIC-IDS2017 datasets respectively. Classifying different attack categories is more difficult for the classifier due to low frequent



attack and lack of representative data instances of the attacks. It clearly obvious the classifier distinguished almost all types of attacks well. In addition, for binary classification both conducted confusion matrix illustrated how the classifier is more efficient and accurate for the binary classification as demonstrated at figure 6.

Figure 4: Multi-Class confusion matrix of proposed model on CSE-CIC-IDS2018

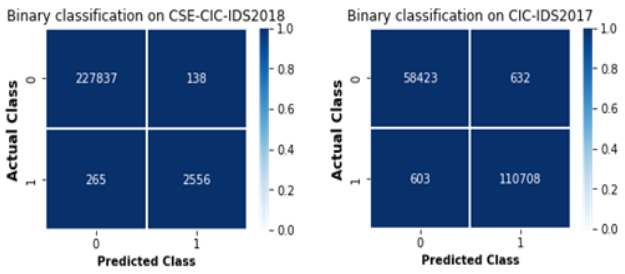


Figure 6: Binary confusion matrix of proposed model on CSE-CIC-IDS2018 and CIC-IDS2017

Table 4 Summarize the performance analysis of the proposed method with applying different feature selection techniques for binary classification. Proposed LSTM + MIFS with accuracy 99.8%, 0.001 FPR, and 66.37seconds on CSE-CIC-IDS2018 dataset outperform others in terms of Accuracy, FPR rate, and time-consuming. However, achieved excellent result with less time on CIC-IDS2017 dataset. Deep learning requires more data instances with more attributes for better understanding and learning trends but its time consuming and counted as limitation as clearly shows on CIC-IDS2017 dataset which outperform other methods by recording more time. Experimental setups conducted on 1,160,435 and 851828 samples by applying 10 epochs for every method on CSE-CIC-IDS2018 and CIC-IDS2017 respectively.

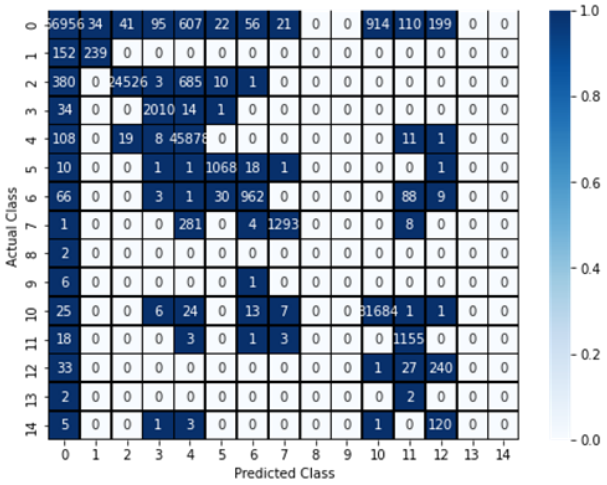


Figure 5: Multi-Class confusion matrix of proposed model on CIC-IDS2017

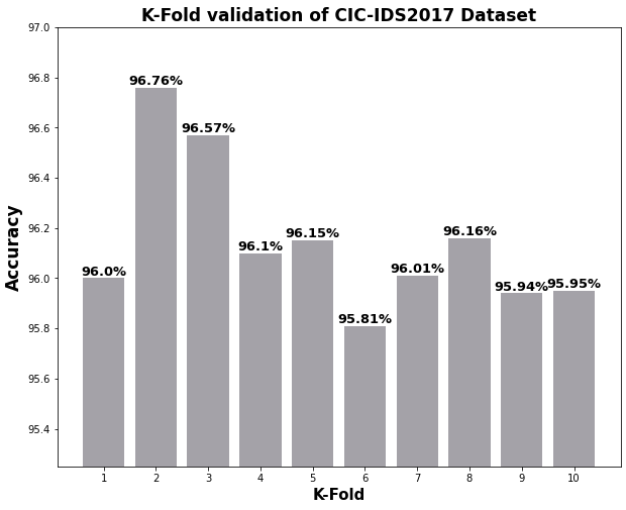


Figure 7: Accuracy vs 10-Fold using MIFS for CSE-CIC-IDS2017

Figure 8: Accuracy vs 10-Fold using MIFS for CIC-IDS2018

Table5 demonstrate the performance analysis of different types of feature selection methods on CSE-CIC-IDS2018 dataset for multi-class classification. MIFS based-on entropy by selecting subset of only 15 attributes
outperform other techniques by achieving 0.998 score. Selecting optimal 15 attributes highly reduce the time complexity compared to other methods.

Table 5: Scoring feature selection techniques on CSE-CIC-IDS2018

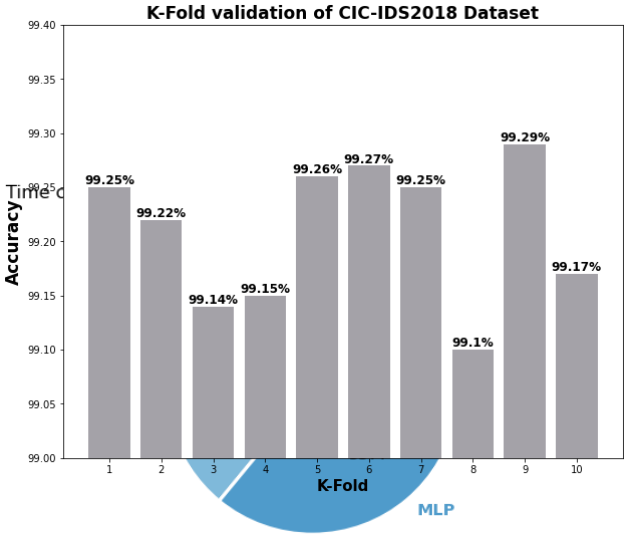
F.S techniques	MFS	IFS	CFS	FFS
No. of attributes	15	29	39	20
Score	0.998	0.995	0.993	0.995

4.4 K-Fold validation performance

In another experimental study, the proposed method has shown encouraging performance on both datasets using stratified K-Fold which is an enhanced version of K-Fold cross-validation mainly used to preserve percentage of samples for each class. It maintains the same class ratio throughout the K-folds as the ratio in the original dataset. While SMOTE is an effective technique to solve imbalance problems in the dataset, however we have not applied during training and testing sets due to preserve original nature ratio of some attacks such infiltration and SQL injection. The experiments conducted on both datasets applied by 10-Fold cross-validation, estimation and evaluation metric in terms of accuracy calculated at each validation. The results shown in Figure (7, 8) are indicated for excellent performance of proposed method achieved higher than 95% and 99%in each validation test for both datasets respectively.

4.5 Time Complexity

Computational time complexity is one of the critical aspects in many data mining and machine learning application especially real-time and time series detection. Detecting cyber-attack in real-time network flow is challenging for security consultants and network engineers. The results in Figures (9, 10) illustrated pie chart time computational complexity for proposed method and other base classifiers using original features and MIFS method. Proposed method finished the training process with lowest time only consuming 112 seconds for original features and 82 seconds for



only 15 features out of 78 using MIFS method. The two pie chart graphs demonstrate the superiority of MIFS method over original features regarding time-consuming. While, MLP recorded the worst results for both comparisons.

Figure 9: Illustrate time-consuming comparison (original features)

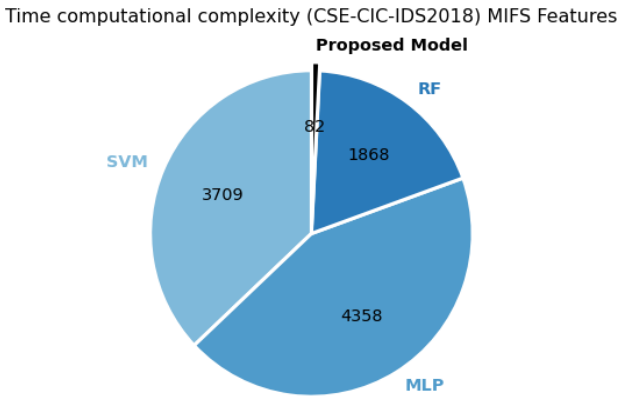


Figure 10: Illustrate time-consuming comparison (MIFS Features)

4.6 Comparative study

In order for further demonstration and effectiveness of the proposed method, this section discusses the comparative analysis of the proposed and the existing models for both binary and multi-class classification. We have investigated comparative analysis based on NIDS datasets using 10-fold cross-validation, ROC

curves, and performance comparison with other existing models. Figure 11 demonstrate the performance comparison of ROC curves on CSE-CIC-IDS2018 for proposed method and several base classifiers regarding normal and abnormal binary attack detection. Proposed method achieved maximum area under ROC curve by recording 0.98 score value. It clearly shows with the changing rate of False-positive, rate of
True-positive of the proposed method is gets much higher than others.

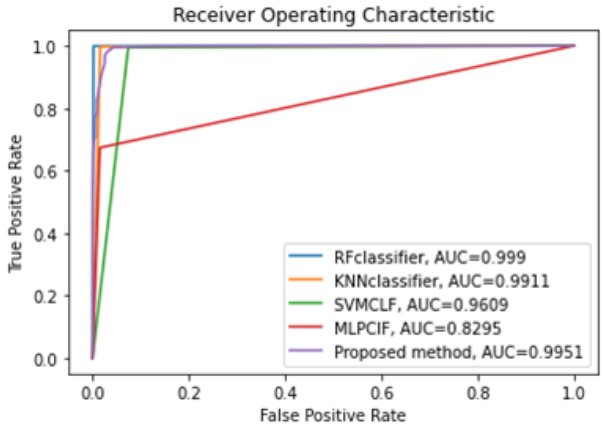


Figure 11: Performance comparison of ROC curves on (CSE-CIC-IDS2018)

In addition, in figure 12 another ROC curve comparison accomplished on CIC-IDS2017 dataset, proposed method shows excellent performance by achieving 0.995 score which is very close to RFclassifier in first order by scoring 0.999. However, MLP classifier achieved worst performance only by scoring 0.96 and 0.82 on both datasets respectively.

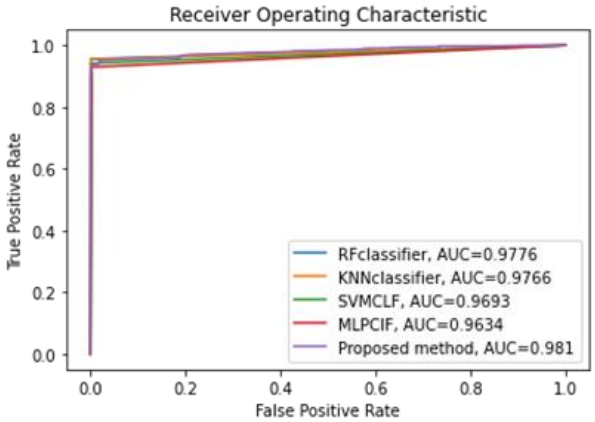


Figure 12: Performance comparison of ROC curves on (CIC-IDS2017)

In another comparison, in figure 13 a comparative analysis conducted for the proposed and the existing models with different K-folds. Proposed method recorded highest accuracy in each 10-fold cross-validation regarding accuracy compared with other. MLP recorded the worst result in all folds.

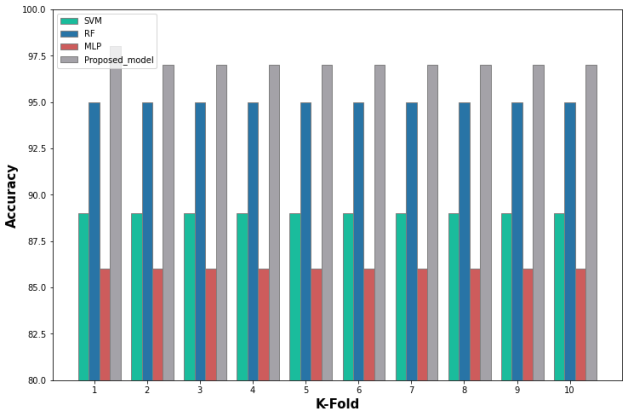


Figure 13: Comparative analysis for K-folds cross-validation

4.7 Additional comparisons

For further comparisons, the performance of the proposed method compared with several models based on evaluation metrics such as accuracy, precision, FPR, and time consuming for different datasets including different features selection techniques. All comparisons are applied on both CSE-CIC-IDS2018, CIC-IDS2017, and NF- CSE-CIC-IDS2018 dataset including 1,160,435, 851828, and 840,000 samples respectively. In table 6, proposed method in multi-class classification outperforms others in terms of accuracy achieving 99.89%, 0.0016 of FPR, and only 111.9 seconds using original features.

On the other hand, Table 7 shows how the MIFS method reduced time complexity into only 81.75 seconds, while the results are very close to the maximum values achieved by SVM. In both comparisons, MLP obtained the highest time during training data instances. In another comparison as shows in table (8, 9) demonstrate the performance comparisons for CIC-IDS2017 dataset using original

features and MIFS method respectively. Thus, proposed method conducted encouraging performance applying feature selection methods. While, Logistic regression (LR) provides lowest time but poor performance. Consequently, the results of Proposed method are better at all.

Table 6: Performance analysis of the models for the CSE-CIC-IDS2018 (original features)

Multi-class classification				
Model	Accuracy	Precision	FPR	Time
SVM	99.8	**99.9**	0.001	4738.7
MLP	99.7	99.9	0.002	5377.1
RF	99.8	99.8	0.001	1913.8
LR	99.6	99.5	0.003	157.3
Proposed Model	**99.8**	99.8	**0.001**	**111.9**

Table 7: Performance analysis of the models for the CSE-CIC-IDS2018 (MIFS method)

Multi-class classification				
Model	Accuracy	Precision	FPR	Time
SVM	**99.94**	**99.93**	**0.001**	4708.82
MLP	99.62	99.55	0.003	4358.25
RF	99.85	99.85	0.001	1867.96
LR	98.86	97.91	0.011	101.67
Proposed Model	99.79	99.78	0.002	**81.75**

Table 8: Performance analysis of the models for the CIC-IDS2017 (original features)

Multi-class classification				
Model	Accuracy	Precision	FPR	Time
SVM	98.65	98.77	0.013	9549.35
MLP	98.32	98.25	0.016	664.29
RF	**99.87**	**99.81**	**0.001**	2944.42
LR	88.98	89.0	0.114	**31.80**
Proposed Model	98.69	98.55	0.013	83.56

Table 9: Performance analysis of the models for the CIC-IDS2017 (MIFS method)

Multi-class classification				
Models	Accuracy	Precision	FPR	Time
SVM	97.42	97.46	0.025	6290.41
MLP	97.44	97.35	0.025	717.17
RF	**99.89**	**99.82**	**0.001**	563.83
LR	73.75	73.01	0.261	**24.19**
Proposed Model	97.28	97.06	0.027	**82.78**

Table 10: Performance analysis of the models for the CSE-CIC-IDS2018 (original features)

Binary classification				
Models	Accuracy	Precision	FPR	Time
SVM	99.64	99.62	0.003	4543.41
MLP	99.88	**99.88**	0.001	191.29
RF	99.81	99.81	0.001	98.01
LR	99.38	99.34	0.006	**19.53**
Proposed Model	**99.89**	99.87	**0.001**	72.06

Table 11: Performance analysis of the models for the CIC-IDS2017 (original features)

Binary classification				
Models	Accuracy	Precision	FPR	Time
SVM	92.49	92.52	0.075	8290.42
MLP	98.70	98.71	0.012	768.58
RF	**99.57**	**99.57**	**0.004**	85.02
LR	92.49	92.52	0.075	**33.10**
Proposed Model	98.79	98.92	0.012	90.96

Table 12: Performance analysis of the classifiers for the NF-CSE-CIC-IDS2018 (10 features)

Binary classification				
Models	**Accuracy**	**Precision**	**FPR**	**Time**
SVM	99.02	99.04	0.009	9994.05
MLP	99.07	99.12	0.009	568.21
RF	99.28	99.33	0.007	106.95
LR	97.36	97.40	0.026	**16.16**
Proposed Model	**99.39**	**99.50**	**0.006**	53.44

Table 10 and table 11 illustrate binary classification using original features having 78 features. The results of proposed method in table 10 are satisfactory by achieving 99.89% accuracy, the least FPR 0.001, and time on CSE-CIC-IDS2018 dataset. However, according to binary classification on CIC-IDS2017 dataset in table 11 RF outperform proposed method by recording little higher value.

Another comparison in table 12 accomplished to estimate the performance of the proposed method and other base classifiers on newly published dataset called NF-CSE-CIC-IDS2018 with only 10 features. Experimental results are conducted on 840,000 under data instances. Higher accuracy rate and lower FPR, and detection time recoded by proposed method were encouraging.

All the existing models and proposed model results are promising in terms of accuracy, false-positive rates. However, time complexity which is one of the critical aspects for attack detection is very high on compared models. Performance analysis of the proposed model on three latest and realistic datasets estimated providing excellent anomalous behavior detection on the network traffic flow.

4.8 NIDS LIMITATIONS AND ISSUES

Network intrusion detection system (NIDS) is one of the most common types of IDS system due to huge interconnection between personal network peers and enterprise devices. Hackers and network intruders always change their way of attacking to breech the systems. Recently, Artificial Intelligence (AI) has an essential impact on identifying and detecting network intrusions. IDS systems are integrating with intelligent techniques to protect network infrastructures from public intruders. However, these intelligent techniques are not perfect but remain many difficulties and shortcoming behind. Misclassification and high false-positive are the most common issues which most of the intelligent algorithms are suffering from. Hyperparameter tunning and time consuming is another limitation in deep learning techniques. Also, preparing a well representative balanced and realistic dataset which represents every attack is quite challenging for many researchers and technicians. In addition, the proposed approaches are provided deficient performance on real-time traffic analysis due to some issues including low frequent attack, novel attacks, huge amount of data packets inspection, attack detection based on stream data, and latent patterns in payload shellcodes.

5. CONCLUSIONS

In this study, we proposed an intelligent technique based on anomaly detection in NIDS system using deep learning algorithm. Proposed method with combination of LSTM and MIFS feature selection technique outperform other existing models in a supervised learning. Mutual information feature selection is selected optimal features among several method which improve the model performance regarding time reduction. LSTM networks by providing long and short memory dependencies able to memorize sequential and previous session of the data in the network. Experimental results based on different datasets show the superiority of deep learning over ML base classifiers in complex problems and sequential cyber-attack detection. In addition, traditional IDS systems based on experts and prior knowledge are rule-based, complex and time consumable to match sequential network packets with all rules in the database. Whilst, network intruders are always change their way of attacking based on revealing limitation of the systems and design robust intelligent techniques to be anonymous. Thus, Artificial Intelligence with powerful and accurate techniques are replaced eventually. For these purposes, ML and DL algorithms provide effective techniques to accurately detect diverse types of attacks in terms of accuracy, false positive rate, and misclassification rates. However, intelligent techniques are suffering from many limitations including misclassification and high false-positive rates and also poor performance in real-time detection. As a conclusion, more effective intelligent techniques and well-organized structures with expert knowledge need to be considered for better detection of any attacks in real-time applications.

5.1 Future work

As improvement for our model in the future work, we will use meta-heuristic optimization algorithm to train LSTM for better classification and detection rate particularly in time series data. Real-time data intrusion detection techniques are used unsupervised learning in time series manner. Also, we try to make a hybrid system of signature-based intrusion systems like snort and anomaly-based intelligent technique using machine or deep learning. In addition, we try to train our proposed model in real-time anomaly detection as LSTM algorithm is highly utilized in sequential data problems.

References

Al-Daweri, M. S., Abdullah, S. and Zainol Ariffin, K. A. (2021) 'An adaptive method and a new dataset, UKM-IDS20, for the network intrusion detection system', Computer Communications, 180(February), pp. 57–76. doi: 10.1016/j.comcom.2021.09.007.

Ambusaidi, M. A. et al. (2016) 'Building an intrusion detection system using a filter-based feature selection algorithm', IEEE Transactions on Computers, 65(10), pp. 2986–2998. doi: 10.1109/TC.2016.2519914.

Debasish Kalita (2022) An Overview on Long Short Term Memory (LSTM), March 11, 2022. Available at: https://www.analyticsvidhya.com/blog/2022/03/an-overview-on-long-short-term-memory-lstm/.

Govindarajan, M. and Chandrasekaran, R. (2011) 'Intrusion detection using neural based hybrid classification methods', Computer Networks, 55(8), pp. 1662–1671. doi: 10.1016/j.comnet.2010.12.008.

Kanimozhi, V. and Jacob, T. P. (2020) 'Artificial Intelligence outflanks all other machine learning classifiers in Network Intrusion Detection System on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing', ICT Express, (xxxx). doi: 10.1016/j.icte.2020.12.004.

De la Hoz, E. et al. (2015) 'PCA filtering and probabilistic SOM for network intrusion detection', Neurocomputing, 164, pp. 71–81. doi: 10.1016/j.neucom.2014.09.083.

Moualla, S., Khorzom, K. and Jafar, A. (2021) 'Improving the Performance of Machine Learning-Based Network Intrusion Detection Systems on the UNSW-NB15 Dataset', Computational Intelligence and Neuroscience, 2021. doi: 10.1155/2021/5557577.

Oliveira, N. et al. (2021) 'Intelligent cyber attack detection and classification for network-based intrusion detection systems', Applied Sciences (Switzerland), 11(4), pp. 1–21. doi: 10.3390/app11041674.

Peddabachigari, S. et al. (2007) 'Modeling intrusion detection system using hybrid intelligent systems', Journal of Network and Computer Applications, 30(1), pp. 114–132. doi: 10.1016/j.jnca.2005.06.003.

Ravikumar, D. (2021) 'Towards Enhancement of Machine Learning Techniques Using CSE-CIC-IDS2018 Cybersecurity Dataset'. Available at: https://search.proquest.com/openview/6223ef80659ae148bd418cd6fb01b2fd/1?pq-origsite=gscholar&cbl=18750&diss=y.

Rosay, A. et al. (2021) 'Multi-layer perceptron for network intrusion detection: From a study on two recent data sets to deployment on automotive processor', Annales des Telecommunications/Annals of Telecommunications. doi: 10.1007/s12243-021-00852-0.

Vinayakumar, R. et al. (2019) 'Deep Learning Approach for Intelligent Intrusion Detection System', IEEE Access, 7, pp. 41525–41550. doi: 10.1109/ACCESS.2019.2895334.

Yin, C. et al. (2017) 'A Deep Learning Approach for Intrusion

Detection Using Recurrent Neural Networks', IEEE Access, 5, pp. 21954–21961. doi: 10.1109/ACCESS.2017.2762418.

Zhou, Y. et al. (2020) 'Building an efficient intrusion detection system based on feature selection and ensemble classifier', Computer Networks, 174(April). doi: 10.1016/j.comnet.2020.107247.

Zoghi, Z. and Serpen, G. (2021) 'UNSW-NB15 Computer Security Dataset: Analysis through Visualization'. Available at: http://arxiv.org/abs/2101.05067.