iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.51858

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Analysis of Efficient Intrusion Detection System using Ensemble Learning

Dr. R. Neelaveni[1], Abhinav[2], Sahas[3]
Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, India

Abstract: ***Our increasingly connected world continues to face an ever-growing amount of network-based attacks. Intrusion detection systems (IDS) are essential security technology for detecting these attacks. Although numerous machine learning-based IDS have been proposed for detecting malicious network traffic, most have difficulty properly detecting and classifying the more uncommon attack types. The research in Cyber Security has raised the need to address the cybercrimes that have caused the requisition of intellectual properties such as the breakdown of computer systems and impairment of important data compromising the confidentiality authenticity and integrity of the user. Considering these scenarios, securing the computer systems and the user using an Intrusion Detection System (IDS) is essential. The performance of IDS was studied by developing an IDS dataset consisting of network traffic features to learn the attack patterns. Intrusion detection is a classification problem wherein various Ensemble Learning (ML) and Data Mining (DM) techniques are applied to classify the network data into normal and attack traffic. Moreover, the types of network attacks changed over the years, so updating the datasets used for evaluating IDS is necessary.***

Keywords: ***Data Processing, Data Mining, Network Intrusion detection systems, Intrusion detection system, Ensemble learning***

I. INTRODUCTION

The increasing volume and sophistication of network-based attacks motivate the development of effective techniques and tools to prevent service disruption unauthorized access and the disclosure of sensitive information. An Intrusion Detection System (IDS) is an important defence tool against sophisticated and increasing network attacks but these systems especially Machine Learning (ML) based systems require large reliable and valid network traffic datasets to be effective. Although the majority of recently available datasets cover a range of network attack types and traffic patterns and include information about the attacking infrastructure modern networks are increasingly diversified such that existing datasets are often not enough to develop effective classification mechanisms. These datasets often lack traffic diversity and volume or fail to cover the full scope of known attack types. To cope with these new changes we require a more dynamic dataset that will improve the ability of an IDS to detect intrusions. Using deep learning techniques such as Generative Adversarial Networks (GANs) we can fabricate additional data using existing datasets to increase the classification accuracy of an IDS, especially for rare attack categories. Two methods of IDS are Signature-based Intrusion Detection Systems (SNIDS) and Anomaly-based Intrusion Detection Systems (ANIDS).

The SNIDS approach is effective for known threats as it looks for specific patterns (or signatures) such as byte sequences in network traffic or known malicious instructions sequences used by malware. Conversely, the ANIDS approach uses ML algorithms to analyze and monitor the network traffic in order to detect any suspicious activity thus being an effective method for catching unknown attacks.

The emergence of deep learning and its integration with Reinforcement Learning (RL) has created a class of Deep Reinforcement Learning (DRL) methods that are able to detect the most recent and sophisticated types of network attacks. DRL combines artificial neural networks with a framework of RL that helps software agents (or learning entities) learn how to reach their goals. DRL combines function approximation and target optimization mapping states and actions to the rewards they lead to. This results in a policy that our learning agents can follow to make the best decisions given the current state. To detect network attacks DRL is used to train an agent such that given a state represented as a collection of feature values will take the best action (which in our case acts as a classification of attack type) in order to recognize an attack. Each network is different in that its behavior and patterns evolve gradually. Naturally, vulnerabilities also evolve. The performance of IDS classification accuracy suffers as existing datasets gradually become out of date invalid and unreliable. Moreover, reliable data cannot often be shared due to privacy concerns. Existing publicly available datasets do not include all of the existing network attack types let alone the unknown vulnerabilities and attacks.

II. LITERATURE SURVEY

Hongyu Yang and Fengyan Wang [1] The diversification of wireless network traffic attack characteristics has led to problems what traditional intrusion detection technology with a high false positive rate, low detection efficiency, and poor generalization ability. In order to enhance security and improve the detection ability of malicious intrusion behaviour in a wireless network, this paper proposes a wireless network intrusion detection method based on an improved convolutional neural network (ICNN). First, the network traffic data is characterized and pre-processed, and then modelled the network intrusion traffic data by ICNN. The low-level intrusion traffic data is abstractly represented as advanced features by CNN, which extracted autonomously the sample features, and optimised network parameters by stochastic gradient descent algorithm to converge the model. Finally, we conducted a sample test to detect the intrusion behaviour of the network. The simulation results show that the method proposed in our paper has higher detection accuracy and true positive rate together with a lower false positive rate. The test results on the test set KDD Test + in our paper show that compared with the traditional models, the detection accuracy is 8.82% and 0.51% higher than that of LeNet-5 and DBN, respectively, and the recall rate is 4.24% and 1.16% higher than that of LeNet-5 and RNN, respectively, while the false positive rate is lower than the other three types of models. It also has a big advantage compared to the IDABCNN and NIDMBCNN methods. Aiming at the problem of wireless network intrusion detection technology based on the deep learning method that has low detection efficiency and is prone to face over-thing and gen- realization issues in the model training process. This paper proposes a wireless network intrusion detection based on an improved convolutional neural network. The classification training and test experiments are carried out in IBWNIDM using the pre-processed training set and test set data. The experimental results show that the accuracy and true positive rate of intrusion detection of IBWNIDM are higher and the false positive rate is lower. Overall, IBWNIDM leverages the advantage of the deep learning model for feature extraction of sample data.

Wenming Wang, Haiping Huang, Qi Li, Fan He and Chao Sha [2] Intrusion detection as one of the most important approaches to guarantee wireless sensing network security has been studied adequately in previous work. However, with the development of electronic anti-reconnaissance technology, the intruder may obtain the location information of detection nodes and perform path planning to avoid being detected. Such an intruder is defined as an empowered intruder who will bring new challenges to traditional intrusion detection methods. Moreover, some subareas may have coverage holes due to the random initial deployment of detection nodes, and the desired effect of detection cannot be achieved. To address these issues, we propose a vehicle collaboration sensing network model, where mobile sensing vehicles and static sensor nodes cooperate to provide intrusion detection against empowered intruders. Our proposal (named IDEI) consists of a target pursuit algorithm of mobile sensing vehicles and a sleep-scheduling strategy of static nodes. Mobile sensing vehicles will track the empowered intruder and fill up the coverage breaches, while static nodes follow a sleep-scheduling mechanism and will be awakened by detection nodes nearby when the intruder is detected. Simulation experiments are conducted to compare our proposal with existing methods such as KMsn and MTTA in terms of intrusion detection performance, energy consumption and the moving distance of sensor nodes. The parameter sensitivity of IDEI is also studied with extensive simulations. The theoretical analysis and simulation results indicate that our proposal can achieve better efficiency and availability. In this paper, we first put forward the model of an empowered intruder. Compared with naive intruders, the empowered intruder can locate detection nodes nearby and escape from them to reduce the probability of being detected. Aiming at the challenge brought by the empowered intruder, a dis- tributed intrusion detection scheme IDEI based on a vehicle collaboration sensing network is proposed. Mobile sensing vehicles are utilized to track the empowered intruder to achieve high-quality monitoring and a sleep-scheduling mechanism is designed for static sensors to reduce energy consumption. In addition, there is a mobile sensing vehicle that acts as an edge computing node in each monitoring area to satisfy the requirements of low latency and high-quality service.

Jaime Zuniga-Mejia, Rafaela Villalpando-Hernandez, Cesar Vargas-Rosales and Andreas Spanias [3] Reconfigurable wireless networks, such as ad hoc or wireless sensor networks, do not rely on fixed

infrastructure. Nodes must cooperate in the multi-hop routing process. This dynamic and open nature makes reconfigurable networks vulnerable to routing attacks that could degrade significantly network performance. Intrusion detection systems consist of a set of techniques designed to identify hostile behaviour. In this paper, there are several approaches for intrusion detection in reconfigurable network routing such as collaborative, statistical, or machine learning-based techniques. In this paper, we introduce a new approach to intrusion detection for reconfigurable network routing based on linear systems theory. Using this approach, we can discriminate routing attacks by considering the system's z-plane poles. The z-plane can be thought of as a two-dimensional feature space that arises naturally. It is independent of the number of network attack detection metrics and does not require extra dimensionality reduction. Two different host-based intrusion detection techniques, inspired by this new linear systems perspective, are presented, and analysed through a case study.

The case study considers the effects of attack severity and node mobility on the attack detection performance. High attack detection accuracy was obtained without increasing packet overhead for both techniques by analysing locally available information. In this work, we proposed two different IDS for routing in RWN based on the same perspective of considering a network node as a linear system. This new perspective allows us to gain some intuitive understanding of the problem. Additionally, by using the system poles on the z-plane as the feature space for attack detection, we can represent all the relevant information in two dimensions. This two-dimensional feature space is guaranteed to be independent of the number of input and output signals considered relevant network metrics for a given attack detection. Good detection accuracy was obtained for both attack detection techniques. For more elaborate scenarios than the simple case presented in Section IV, we need to consider additional inputs. The root locus approach is more robust to mobility and has a lower computational cost when compared to the black box method and hence more feasible for low-power devices.

Liqun Yang, Jianqiang Li, Liang Yin, Zhonghao Sun, Yufei Zhao and Zhoujun Li [4] With the development of wireless network techniques, the number of cyber-attack increases significantly, which has seriously threatened the security of Wireless Local Area Networks (WLAN). Traditional intrusion detection technology is a prevalent area of study for numerous years, but it may not have a good detection performance in a real-time way. Therefore, it is urgent to design a detection mechanism to detect the attacks timely. In this paper, we exploit a CDBN (Conditional Deep Belief Network)-based intrusion detection mechanism to recognize the attack features and perform the wireless network intrusion detection in real-time. To avoid the impact of the imbalanced dataset and the data redundancy on the detection accuracy, a window-based instance selection algorithm Sam Select is adopted to undersample the majority class data samples and a Stacked Contractive Auto-Encoder (SCAE) algorithm is proposed to reduce the dimension of the data samples. By doing so, our proposed mechanism can effectively detect the potential attack and achieve high accuracy. The experiment results show that CDBN can be effectively combined with Sam Select and SCAE, and the proposed mechanism has a high detection speed and accuracy, with an average detection time of 1.14 ms and a detection accuracy of 0.974. In this paper, we propose an improved Deep Belief Network-based scheme for detecting wireless network intrusion. Our proposed scheme employs Conditional Deep Belief Network (CDBN) to efficiently learn the temporal behaviour features between the experimental data. We adopt a window-based under-sampling algorithm Sam Select to balance the numbers of the normal samples and that of the attack samples in the AWID training dataset. We use Stacked Contractive Auto-encoder (SCAE) algorithm to eliminate the redundancy of experimental data. In the simulations, we illustrate our work by four cases, and the first two cases show that SCAE is feasible to reduce the dimensionality with an average reconstruction error of 0.058. The detection accuracy increases as the number of the hidden layers of CDBN increases, and the highest detection accuracy is 0.974 when the number of hidden layers is 6.

Zhendong Wang, Yong Zeng, Yaodi Liu and Dahai Li [5] Deep learning has become a research hotspot in the field of network intrusion detection. In order to further improve the detection accuracy and performance, we proposed an intrusion detection model based on an improved deep belief network (DBN). Traditional neural network training methods, like Back Propagation (BP), start to train a model with preset parameters such as the randomly initialized weights and thresholds, which may bring some issues, e.g., attracting the model to the local optimal solutions, or requiring a long training period. We use the Kernel-based Extreme Learning Machine (KELM) with the supervised learning ability to replace the BP algorithm in DBN in a bid to ameliorate the situation. Considering the problem of poor classification performance usually caused by randomly initializing kernel parameters with KELM, an enhanced grey wolf optimizer (EGWO) is designed to optimize the parameters of KELM. In order to improve the searchability and optimization ability of the traditional grey wolf optimizer algorithm, a novel optimization strategy combining inner and outer hunting is introduced. Experiments on KDDCup99, NSL-KDD, UNSW-NB15 and CICIDS2017 datasets show that the proposed DBN-EGWO- KELM algorithm has greater advantages in terms of its accuracy, precision, true positive rate, false positive rate and other evaluation indices compared with BP, RBF, SVM, KELM, LIBSVM, CNN, DBN- KELM and other intrusion detection models, and can effectively meet the requirements of intrusion detection of complex networks. We propose an intrusion detection method based on an improved deep belief network. A novel kernel extreme learning machine classification model is designed using enhanced grey wolf optimizer optimization, which extracts Z. Wang et al.: DBN Integrating Improved KELM for Network ID Z. Wang et al.: DBN Integrating Improved KELM for Network ID data features by employing the dimensionality reduction ability of DBN for complex high-dimensional network intrusion data features. The combination with the enhanced grey wolf optimizer is viable to optimize the kernel extreme learning machine classification model for the purpose of improving the performance of KLEM.

Manuel Lopez-Martin, Antonio Sanchez-Esguevillas, Juan Ignacio Arribas and Belen Carro [6] Network intrusion detection focuses on classifying network traffic as either normal or attack carriers. The classification is based on information extracted from the network flow packets. This is a complex classification problem with unbalanced datasets and noisy data. This work extends the classic radial basis function (RBF) neural network by including it as a policy network in an offline reinforcement learning algorithm. With this approach, all parameters of the radial basis functions (along with the network weights) are learned end-to-end by gradient descent without external optimization. We further explore how additional dense hidden layers and the number of radial base kernels influence the results. This novel approach is applied to five prominent intrusion detection datasets (NSL-KDD, UNSW-NB15, AWID, CICIDS2017 and CICDDOS2019) achieving better performance metrics than alternative state-of-the-art models. Each dataset provides different restrictions and challenges allowing a better validation of results. Analysis of the results shows that the proposed architectures are excellent candidates for designing classifiers with the constraints imposed by network intrusion detection. We discuss the importance of dataset imbalance and how the proposed methods may be critically important for unbalanced datasets. Network intrusion detection is an increasingly important problem in modern data networking, and it is an active research eld in which many types of machine learning and deep learning models have been applied. We propose novel extensions to the RBFNN model. These extensions are based on an end-to-end training scheme using gradient descent for all the parameters of the network: the network weights, and the centres and dispersion parameters of the radial basis functions. This end-to-end training scheme allows us to pro- pose several alternative loss functions, different from the cross-entropy generally used for classification. It also allows a complete RBFNN network to be included as the policy network of an offline reinforcement learning model where the loss function is replaced by a reward function that is not necessarily differentiable. This approach also offers the opportunity to include additional dense hidden layers after the initial RBF layer.

Kaiyuan Jiang, Wenya Wang, Aili Wang and Haibin Wu [7] Intrusion detection system (IDS) plays an important role in network security by discovering and preventing malicious activities. Due to the complex and time-varying network environment, the network intrusion samples are submerged into a large number of normal samples, which leads to insufficient samples for model training and detection results with a high false detection rate. According to the problem of data imbalance, we propose a network intrusion detection algorithm that combined hybrid sampling with the deep hierarchical network. Firstly, we use the one-side selection (OSS) to reduce the noise samples in the majority category, and then increase the minority samples by Synthetic Minority Over-sampling Technique (SMOTE). In this way, a balanced dataset can be established to make the model fully learn the features of minority samples and greatly reduce the model training time. Secondly, we use a convolution neural network (CNN) to extract spatial features and Bi-directional long short-term memory (BiLSTM) to extract temporal features, which forms a deep hierarchical network model. The proposed network intrusion detection algorithm was verified by experiments on the NSL-KDD and UNSW-NB15 datasets, and the classification accuracy can achieve 83.58% and 77.16%, respectively. In this paper, a novel method for an intrusion detection system based on the combination of hybrid sampling and the deep hierarchical network has been proposed and discussed. Firstly, we combine OSS and SMOTE to construct a balanced dataset for model training. It can reduce the training time of the model and solves the common problems to some extent of inadequate training from unbalanced samples. In addition, a network data preprocessing method is established for com- the proposed deep hierarchical network model. Then, classify the input data through the hierarchical network constructed by CNN and BiLSTM. The model extracts features automatically through repeated multi-level learning by taking advantage of the outstanding features of deep learning. Two intrusion datasets (NSL-KDD and UNSW-NB15) have been employed to evaluate the performance of the proposed approach. Based on the statistical significance tests, it could be concluded that the proposed approach outperforms other classes, such as Random Forest, LeNet, AlexNet, CNN and BiLSTM. The proposed method yields a superior K. Jiang et al.

Tao Duan, Youhui Tian, Hanrui Zhang, Yaozong Liu, Qianmu Li, Jian Jiang and Zongsheng Shi [8] Intrusion detection technology, as an active and effective dynamic network defence technology, has rapidly become a hot research topic in the field of network security since it was proposed. However, current intrusion detection still faces some problems and challenges that affect its detection performance. Especially with the rapid development of the current network, the volume and dimension of network data are increasing day by day, and the network is full of a large number of unlabeled data, which brings great pressure on the data processing methods of IDS. In view of the tremendous pressure of intrusion detection brought by the current complex and high-dimensional network environment, this paper provides a feasible solution. Firstly, this paper briefly outlines the necessity of feature learning, the shortcomings of traditional feature learning methods and the new breakthroughs brought by deep belief networks in feature learning, and focuses on the principle and working mechanism of deep belief networks and Principal Component Analysis (PCA).

Then, it constructs the intrusion detection model based on PCA-BP and DBN respectively. And through the experimental evaluation of the two detection models, a comparative experiment between deep belief network and principal component analysis is constructed. The experimental results show that deep belief network has unique advantages and good performance in feature learning. Therefore, deep belief networks can be applied in the field of intrusion detection to extract effective features from the current high-dimensional and redundant network data, thereby improving the detection performance of IDS and its adaptability to the current complex and high-dimensional network environment. Due to the high dimensionality and redundancy of current network data, feature learning has become a necessary process for current intrusion detection data processing models. However, the traditional feature learning methods have T. Duan et al.: Intelligent Processing of Intrusion Detection Data certain shortcomings. And the advent of deep learning has brought new directions to feature learning. In order to verify the unique advantages of deep learning-related technologies in feature learning, this paper constructs a PCA-BP-based intrusion detection data processing model and a DBN-based intrusion detection data processing model.

Ying Zhang, Peisong Li and Xinheng Wang and [9] With the advent of the Internet of Things (IoT), the security of the network layer in the IoT is getting more and more attention. Traditional intrusion detection technologies cannot be well adapted to the complex Internet environment of IoT. For the deep learning algorithm of intrusion detection, a neural network structure may have fifteen detection accuracy for one kind of attack, but it may not have a good detection effect when facing other attacks. Therefore, it is urgent to design a self-adaptive model to change the network structure for different attack types. This paper presents an intrusion detection model based on an improved genetic algorithm (GA) and deep belief network (DBN). Facing different types of attacks, through multiple iterations of the GA, the optimal number of hidden layers and a number of neurons in each layer are generated adaptively so that the intrusion detection model based on the DBN achieves a high detection rate with a compact structure. Finally, the NSL-KDD dataset was used to simulate and evaluate the model and algorithms. The experimental results show that the improved intrusion detection model combined with DBN can effectively improve the recognition rate of intrusion attacks and reduce the complexity of the neural network structure. Through GA, the optimal individuals can be generated by iterations. DBN can effectively process highly complex and high-dimensional data, and the classification results are very good. In this paper, the improved genetic algorithm is combined with the deep belief networks, GA performs multiple iterations to produce an optimal network structure, and DBN then uses the optimal network structure as an intrusion detection model to classify the attacks. In this way, facing different types of attacks, the problem of how to select an appropriate neural network structure when using deep learning methods for intrusion detection is solved, thus it improves the classic- cation accuracy and generalization of the model, and reduces the complexity of network structure. This method has many advantages: on the one hand, the specific network structure generated for specific attack types is higher in classification accuracy than other network structures, which can reach more than 99% of the detection rate.

III. EXISTING SYSTEM

Attacks in wireless sensor networks (WSNs) aim to prevent or eradicate the network's ability to perform its anticipated functions. Intrusion detection is a defence used in wireless sensor networks that can detect unknown attacks. Due to the incredible development in computer-related applications and massive Internet usage, it is indispensable to provide host and network security. The development of hacking technology tries to compromise computer security through intrusion. An intrusion detection system (IDS) was employed with the help of machine learning (ML) Algorithms to detect intrusions in the network. Classic ML algorithms like support vector machine (SVM) K-nearest neighbour (KNN) and filter-based feature selection often led to poor accuracy and misclassification of intrusions. This article proposes a novel framework for IDS that can be enabled by Boruta feature selection with a grid search random forest (BFSGSRF) algorithm to overcome these issues. The performance of BFS-GSRF is compared with ML algorithms like linear discriminant analysis (LDA) and classification and regression tree (CART) etc. The proposed work was implemented and tested on network security laboratory knowledge on discovery dataset (NSL-KDD). The experimental results show that the proposed model BFS-GSRF yields higher accuracy (i.e. %) in detecting attacks and it is superior to LDA CART and other existing algorithms. For classification, random forest with grid search (RFGS) is used. RF is one of the widely used algorithms for classification problems. This algorithm will create several classification trees for predicting the target class. Based on the majority of the vote the final prediction was made. Parameter optimization is used to improve the accuracy of the RF algorithm. The grid search method is used in RF to obtain the classification model with higher accuracy for tuning the parameter. The randomly based search method is more efficient than the grid-based search method for hyperparameter optimization. Two discrete integer parameters such as tree and entry are used to tune the parameter. The main objective of the optimization is to minimize the out-of-bag (OOB) error.

After multiple runs, the optimal parameters value is chosen based on the pair that produces the lowest OOB error. CART classification is a supervised nonlinear algorithm used for classification and regression. This algorithm constructs the binary decision tree by splitting the attributes which are considered a node. The whole tree from root to leaf contains a learning sample. For classification, the target variable in CART should be categorical whereas the target variable for the regression tree should be continuous. Here the target variable is categorical for performing classification on intrusion detection. The metric Gini index is used to perform the classification task.

A. Existing System Issues

1) Correctly classified records cannot outperform.
2) No statistically relevant performing variations among the various algorithms.
3) Poor application performance.
4) High complexity, inaccuracy, and inadequacy.
5) Cannot meet current network business demands.
6) The computation burden may limit its further application for real scenarios.

IV. PROPOSED ARCHITECTURE

An intrusion detection dataset can be developed by collecting information from varied sources such as network traffic flows that contain information about the host user behaviour and system configurations. This information is required to study the attack patterns and abnormal activity of various network attacks. The network activity is collected through a router or network switch. After collecting the incoming and outgoing network traffic network flow analysis is performed to study the network traffic. Flow analysis can be described as the process of analyzing the network packet information such as source IP address destination IP address source port number destination port number type of network services to name a few. The network host delivers the system configurations and user information that cannot be extracted from the network flow analysis. For instance, information obtained through failed login attempts by observing the intrusion activity. The evaluation of an IDS model can be performed by implementing Machine Learning (ML) and Data Mining (DM) techniques to classify the network traffic into the benign and malicious traffic flow. The ML and DM techniques implemented on the IDS datasets contain labelled data and network traffic features. These help the classifier to learn different attack patterns to detect a particular attack. The features of the dataset help the classifier to learn the normal traffic patterns as well as attack patterns through which the classifier is able to classify the input data. The dataset used for training the classifier is built by monitoring the network traffic for a particular interval of time. The dataset consists of normal network traffic and anomalous network traffic which helps the classifier to identify the patterns of the data with a sufficient amount of examples. The data collected is divided into a training set and a test set for training and testing the classifier respectively. Thus various ML and DM techniques are used for developing an IDS. A single performance metric is not sufficient to measure the efficiency of the algorithm. It is necessary to consider the confusion matrix and find the number of false positives and false negatives to derive other performance metrics such as Detection Rate (DR) False Positive Rate (FPR) precision and recall. The accuracy of a particular attack type is also a critical aspect as the classifier may give better accuracy for one attack type but may fail for classifying the other.

A. Architecture Diagram

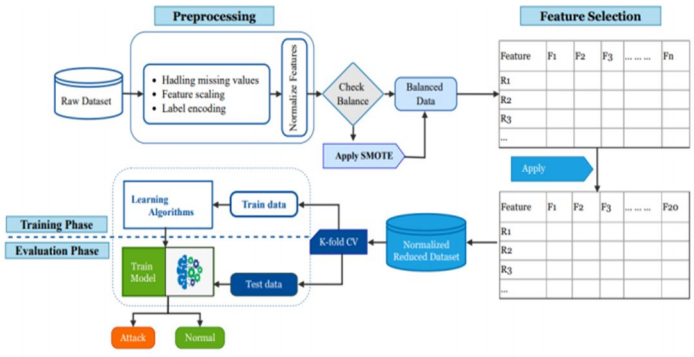


FIGURE 1

B. Algorithm

Ensemble Learning Algorithm

C. Advantages Of Proposed Algorithm

1) Handling multi-dimensional and multi-variety data
2) Efficient Handling of Data
3) Continuous Improvement

D. Modules

1) Different types of feature pre-processing are required for different data types and different machine learning models. Some pre-processing methods are common for all data types. Feature scaling is a method used to normalize the range of independent variables or features of data. It is commonly referred to as normalization. Feature scaling impacts non-tree-based models more than tree-based models. Thus, if you want to achieve good results using a non-tree-based model, you should consider normalizing your numerical features. GAN is a class of algorithmic machine learning frameworks having two neural networks that connect and can analyze, capture, and copy the variations within a dataset. Further, both neural networks work against one another in GAN machine learning, hence called adversarial networks. Synthetic Minority Oversampling Technique (SMOTE) is a statistical technique for increasing the number of cases in your dataset in a balanced way. The component works by generating new instances from existing minority cases that you supply as input.

2) Feature selection is a method of selecting a subset of the underlying features in order to minimize the feature space to the smallest possible size based on some criteria. Feature extraction is a technique for creating a new set of features that can be utilized alone or in combination. Moreover, it can locate and choose the far more beneficial properties inside data. It's an important stage in the learning workflow since it assists in minimizing the fitting problems, reducing adaptation efficiency on the testing data, reducing training duration, and reducing model interpretability. There are three main kinds of feature selection methods: filter-based, wrapper-based, and embedded feature selection. Build-in feature selection is available in the embedded feature selection method, which helps to build a model without applying any additional feature selection method. To choose features, the filter-based feature technique employs assessment criteria, including information analysis as well as distance assessment. The wrapper-based feature selection approach builds a subset of features in a particular way before evaluating feature selection using the findings of classifiers. Using the embedded feature selection approach, certain properties can be dynamically removed within classifier construction, allowing feature selection and classification to be done simultaneously. The variable selection, also known as attribute selection or feature selection is the process of choosing the most important features of a given dataset. In a network intrusion detection dataset, there might be several features that do not contribute to the detection of intrusion. So in order to reduce overfitting, improve the accuracy of the model, and reduce the training time, we can carry out feature selection before training the model. In this paper, we use an analyzer function to evaluate the performance of algorithms with different subsets of the dataset to find the best one that results in better accuracy.

3) Hyperparameter tuning involves finding the best set of parameters to give to our algorithm to achieve the best accuracy measures. Generally, there are two techniques that are used for this purpose, namely grid search and random search. In the grid search method, every possible list of values with every combination is evaluated, and in the random search method, random combinations of parameters are tested to find the best possible values for the model. In this work, we utilized the random search method to find the best hyperparameters in single decision tree experiments. In a Random Forest algorithm, parameters such as the number of trees and the depth of the tree can be examined.

E. Evaluation Metrics

We used the accuracy and F1-score (which combines precision and recall) metrics to evaluate the performance of our DRL model and other ML algorithms. While the accuracy score only measures the percentage of correctly classified samples, this selection of performance metrics allows us to also evaluate the percentage of samples that were incorrectly classified. This is especially important for NIDS as the accuracy performance metric is not enough to evaluate imbalanced datasets such as network traffic data which generally include significantly more normal traffic. These performance metrics are derived from the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values. Fig. 1 presents this confusion matrix used by our evaluation method.

1) *Accuracy:* Accuracy measures the number of correct predictions out of the total predictions made by the model. In this case, accuracy measures the model's ability to correctly identify normal and attack traffic records.
2) *Precision:* Precision measures the number of correct positive predictions out of the total number of positive predictions. In this case, precision measures the model's degree of correctness in predicting attack records over the total number of attacks predicted.
3) *Recall:* Recall measures the number of correct positive predictions out of the total number of positive instances in the dataset. In this case, recall measures the model's ability to correctly identify attack traffic records. From this definition, recall is also referred to as the true positive rate, detection rate, or sensitivity.
4) *F1-score:* F1-score is the harmonic mean of the precision and recall values, essentially a combined measure of the two performance metrics. F1-score quantifies how discriminative the model is and acts as a good indicator of performance since a decrease in either precision or recall results in a significant decrease in the F1-score. In addition, for multiclass classification, we present both the unweighted and weighted F1 scores. The weighted F1- score accounts for label imbalance by considering the number of instances of each label when calculating the average F1 score.

V. COMPARISON

The performance of both algorithms are stated below

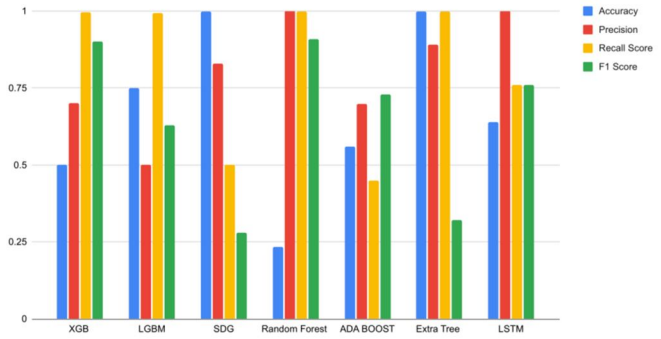


Figure 2 (existing system)

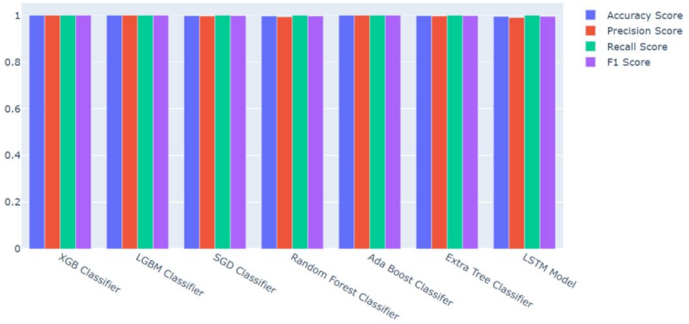


Figure 3 (proposed system)

VI. CONCLUSION AND FUTURE WORK

A. *Conclusion*

The study reviews the datasets developed in the field of Intrusion Detection Systems (IDS). These datasets have been used for performance evaluation of the EL and DM-based IDS. The study revealed that there is a need to update the underlying dataset to identify the recent attacks in the field of IDS with improved performance. This is because the attackers execute attacks by using varied processes and technologies. Moreover, the pattern of executing different attacks simulates the need to have datasets with realistic network scenarios. To fulfil, the requirement of building an intrusion detection dataset with realistic network traffic and updated network attacks CSE-CIC-IDS on AWS datasets have been introduced. This paper reviews the characteristics of these datasets and also discusses a few shortcomings.

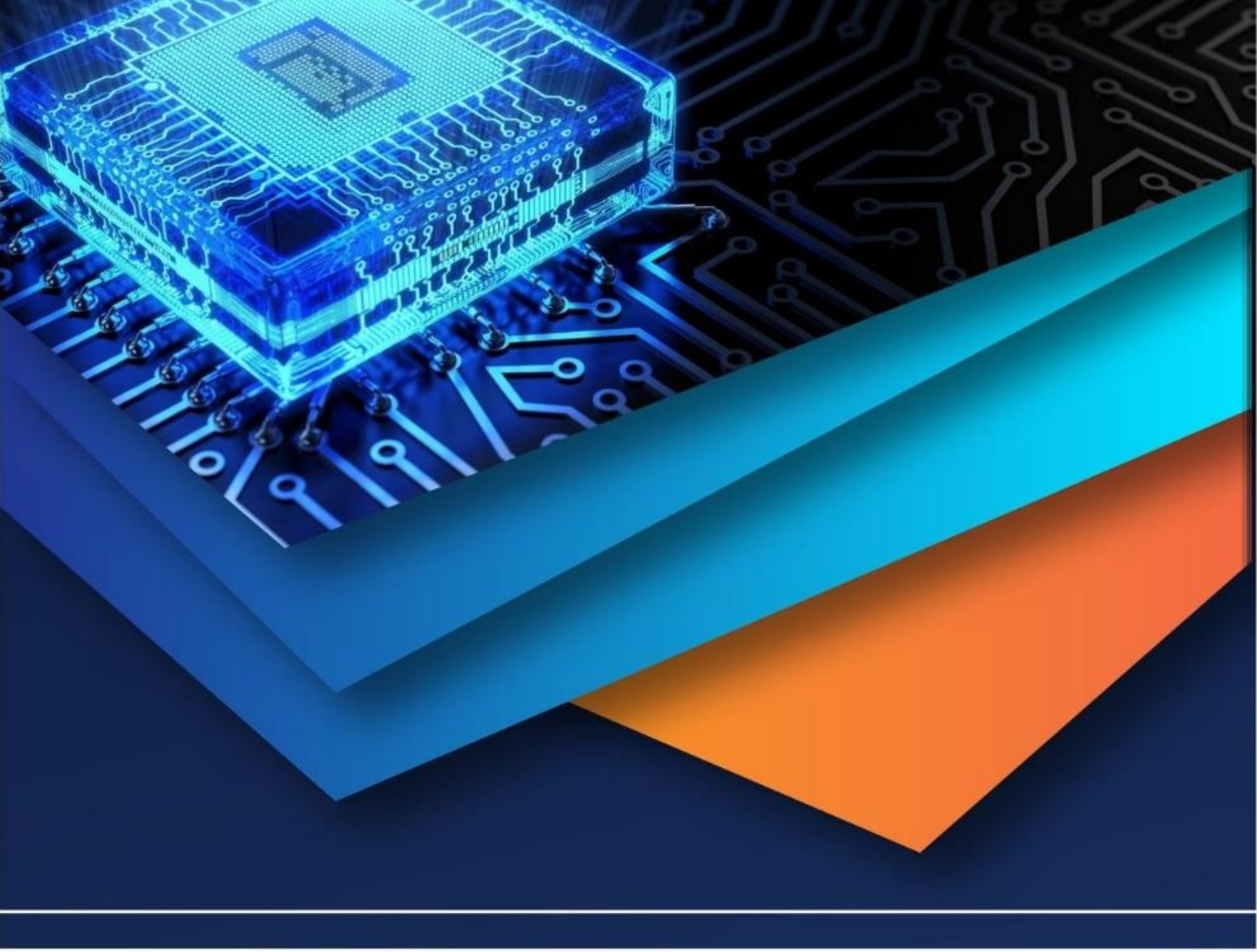
B. Future Work

In the future, we focus on studying the performance of these datasets with various ML and DM techniques along with incorporating feature engineering and data sampling to address the shortcomings of these datasets.

REFERENCES

[1] K. Zheng et al., Algorithms to speedup pattern matching for network May, doi:, ., /j.comcom
[2] Y. Xu and H. Zhao, Intrusion detection alarm altering technology based on ant colony clustering algorithm, in Proc., th Int. Conf. Intel. Syst.
[3] L. Dhanabal and S. P. Shantharajah, A study on NSL-KDD dataset for intrusion detection system based on classification algorithms, Int. J. Adv.
[4] T. Wang, G. Zhang, A. Liu, M. Z. A. Bhuiyan, and Q. Jin, A secure IoT service architecture with an efficient balance dynamics based on cloud and Jun
[5] T. Clouqueur, V. Phipatanasuphorn, and P. Ramanathan, Sensor deployment strategy for detection of targets traversing a region, Mobile. Netw.
[6] W. Heinzelman, A. Chandrakasan, and H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in Proc. , rd Annu. Hawaii Int. Conf. Syst. Sci., Aug., p.,
[7] F. L. Fessant, A. Papadimitriou, A. C. Viana, C. Sengul, and E. Palomar, A sinkhole resilient protocol for wireless sensor networks: Performance Jan. ,
[8] L. Nishani and M. Biba, Machine learning for intrusion detection in, T. Clausen, Optimized link state routing protocol, IETF Internet-Draft, Fremont, CA, USA, Tech. Rep., Jul.,
[9] G. Indirani and K. Selvakumar, A swarm-based efficient distributed intrusion detection system for mobile ad hoc networks (MANET), Int.
[10] L. Ljung, System Identification: Theory for the User. Upper Saddle River, NJ, USA: Prentice Hall,.
[11] H. H. Pajouh, G. Dastghaibyfard, and S. Hashemi, Two-tier network anomaly detection model: A machine learning approach, J. Intell. Inf
[12] S. Park, S. Seo, and J. Kim, Network intrusion detection using stacked Oct. , . , H. Saxena and V. Richariya, Intrusion detection in KDD, a dataset using SVM-PSO and feature reduction with information gain, Int. J. Comput.
[13] Kim, M. Park, and D. H. Lee, AI-IDS: Application of deep learning to Apr.,
[14] L. Thurner, A. Scheidler, F. Schfer, J.-H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, PandapowerAn open-source python tool for convenient modelling, analysis, and optimization of electric power Nov,.
[15] Y. Su, J. Li, A. Plaza, A. Marinoni, P. Gamba, and S. Chakraborty, DAEN: Deep autoencoder networks for hyperspectral unmixing, Jul.,
[16] L. Zhao, Z. Wang, X. Wang, and Q. Liu, Driver drowsiness detection using facial dynamic fusion information and a DBN, IET Intel. Transp.
[17] X. Sun and W. Xu, Fast implementation of DeLongs algorithm for comparing the areas under correlated receiver operating characteristic Jul. , .
[18] S. F. Jilani, Q. H. Abbasi, and A. Alomainy, Inkjet-printed millimetre- , G., ., .
[19] M. Kumar and A. K. Singh, Distributed intrusion detection system using blockchain and cloud computing infrastructure, in Proc. , th Int. , ., /ICOEI, ., .
[20] W. Wang, X. Du, D. Shan, R. Qin, and N. Wang, Cloud intrusion detection method based on stacked contractive auto-encoder and support vector, /TCC., .,.
[21] Z. Yan and Y. Xu, A multi-agent deep reinforcement learning method for cooperative load frequency control of a multi-area power system, , ., /TPWRS., ., .
[22] C. Li, J. Wang, H. Wang, M. Zhao, W. Li, and X. Deng, Visual-texual emotion analysis with deep coupled video and danmu neural networks, , ., /TMM., ., .
[23] K. Zhu, Z. Chen, Y. Peng, and L. Zhang, Mobile edge assisted literal multi-dimensional anomaly detection of in-vehicle network using LSTM, , ., /TVT., ., .
[24] B. Riyaz and S. Ganapathy, A deep learning approach for effective intro- , P. Mishra, V. Varadharajan, U. Tupakula, and E. S. Pilli, A detailed investigation and analysis of using machine learning techniques for intrusion, st Quart., , doi: , ., /cost., ., .
[25] K. Nugroho, E. Noersasongko, Purwanto, Muljono, and H. A. Santoso, Javanese gender speech recognition using deep learning and singu- lar value decomposition, in Proc. Int. Seminar Appl. Technol. Inf. TIC., .,
[26] L. Ertz, M. Steinbach, and V. Kumar, Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data, in Proc. SIAM Int.
[27] H.-T. Li, C.-Y. Chou, Y.-T. Chen, S.-H. Wang, and A.-Y. Wu, Robust and lightweight ensemble extreme learning machine engine based on , ., /TCSI., ., ., A. Aldweesh, A. Derhab, and A. Z. Emam, Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and doi: , ., /j.knosys., ., .
[28] S. Gamage and J. Samarabandu, Deep learning methods in net- work intrusion detection: A survey and an objective comparison, , ., /j.jnca., .,
[29] M. Lopez-Martin, B. Carro, and A. Sanchez-Esguevillas, Application of deep reinforcement learning to intrusion detection for supervised prob- , ., /j.eswa., .,
[30] S. Levine, A. Kumar, G. Tucker, and J. Fu, Ofine reinforcement learning: Tutorial, review, and perspectives on open problems, , arXiv:, ., . , A. AbuGhazleh, M. Almiani, B. Magableh, and A. Razaque, Intelligent intrusion detection using radial basis function neural network, in Proc. , ., /SDS., ., .
[31] Z. Yang, X. Wei, L. Bi, D. Shi, and H. Li, An intrusion detection system based on RBF neural network, in Proc. , th Int. Conf. Com- , ., /CSCWD., ., .
[32] L. Lv, W. Wang, Z. Zhang, and X. Liu, A novel intrusion detec- tion system based on an optimal hybrid kernel extreme learning , ., /j.knosys., ., .
[33] T. Poggio and F. Girosi, Networks for approximation and learning, Proc.
[34] C.-G. Li, M. Wang, Z.-J. Huang, and Z.-F. Zhang, An actor-critic reinforcement learning algorithm based on adaptive RBF network, in Proc. , ., /ICMLC., ., .
[35] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, Kitsune: An ensem- ble of autoencoders for online network intrusion detection, , arXiv:, .,
[36] T. Thi Nguyen and V. Janapa Reddi, Deep reinforcement learning for cyber security, , arXiv:,.,
[37] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, Toward generating a new intrusion detection dataset and intrusion trafc characterization, , ., /, .
[38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, and S. Ghemawat, Ten- sorFlow: Large-scale machine learning on heterogeneous distributed sys- tems, , arXiv:, ., .
[39] H. van Hasselt, A. Guez, and D. Silver, Deep reinforcement learning with double Q-learning, , arXiv:, ., .
[40] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, Playing atari with deep reinforcement learning, , arXiv:, ., . , R. Boutaba, M. A. Salahuddin, N. Limam, S. Ayoubi, N. Shahriar, F. Estrada-Solano, and O. M. Caicedo, A comprehensive survey on machine learning for networking: Evolution, applications and research, ., /s, -, -, -
[41] M. Azizjon, A. Jumabek, and W. Kim, D CNN-based network intrusion detection with normalization on imbalanced data, in Proc. Int. , ., /ICAIIC, ., ., .
[42] J. E. B. Maia, V. R. S. Laboreiro, F. E. Chaves, F. J. A. Maia, T. G. N. Silva, and T. N. Ferreira, a Performance comparison between edited kNN and MQ-RBFN for regression and classification tasks, in Proc. , st Brazilian , .

[43] avuolu, A new hybrid approach for intrusion detection using Jul. , . 44. S. Thaseen, C. A. Kumar, and A. Ahmad, Integrated intrusion detection model using chi-square feature selection and an ensemble of classifiers, , I. S. Thaseen and C. A. Kumar, Intrusion detection model using fusion of chi-square feature selection and multi-class SVM, J. King Saud Univ.
[44] Z. Liu, Z. Lai, and W. Ou Structured optimal graph based sparse fea- May, Art. no. , doi:, ., /j.sigpro., ,.
[45] J. Kim, J. Kim, H. Le T. Thu, and H. Kim, Long short-term memory recurrent neural network classier for intrusion detection, in Proc. Int.
[46] M. Wang and J. Li, Network intrusion detection system based on con-,.
[47] R. Zazo, P. S. Nidadavolu, N. Chen, J. Gonzalez-Rodriguez, and N. Dehak, Age estimation in short speech utterances based on LSTM,
[48] S. Wan, Y. Xia, L. Qi, Y.-H. Yang, and M. Atiquzzaman, Auto-mated colourization of a grayscale image with seed points propagation, TMM., ,.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY