Journal of Advanced Zoology

ISSN: 0253-7214
*Volume **44** Issue **S-6 Year 2023** Page **1010:1016***

NIDS: An Efficient Network Intrusion Detection Model for Security of Big Data Using Different Machine Learning classifiers

A. P. Bhuvaneswari[1] Dr.R.Praveen Sam[2] Dr.C.Shoba Bindu[3]

[1,3]Dept. of Computer Science and Engineering, JNTUA University, Ananthapuramu, A.P, India.
[2]Dept. of Computer Science and Engineering, G.Pulla Reddy Engineering College, Kurnool, A.P, India.

**Corresponding author's E-mail: A. P. Bhuvaneswari*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 30 Nov 2023

Abstract

Security of the big data is one of the important challenges which needs to be addressed by designing an efficient network intrusion model for detecting the unauthenticated intruders in the network. The model should be able to detect the validity of the packet. The detection of intrusions in network was already represented by multiple researchers using different algorithms which still needs instant addressing. Proposing a machine learning classifier algorithm for intrusion detection. The KDD intrusion dataset is used in training the machine for identifying the different intrusions of the network traffic. The machine must be trained efficiently using the different classification algorithms and the security for the data needs to be attained by identifying the invalid network packets. The experimental results demonstrate that the random forest ensemble machine learning classifier is having highest accuracy of 0.2 % when compared with the existing research results in the identification of different intrusions towards the network packets.

Keywords: *Intrusion detection, Network, Machine learning, Classification algorithms, Accuracy*

1. Introduction

In the present days, the security of the data plays an important role which needs to be protected with at most care especially in the case of health information. So, we need an efficient security model which can identify the network packets validity in accessing the data. By using machine learning techniques, we are improving the quality of detection of different intrusions to prevent the loss of crucial data. The machines are trained properly and tested for the detection of invalid packets while accessing the data through the networks.

Network Security

The computer networks allow a secure end-to-end communication using network infrastructure between the applications. With the huge usage of network many types of attacks are occurring now a days which is challenging the availability, confidentiality, Integrity, Authenticity of the available information. To protect the information the network traffic must be observed and the detection of invalid packets coming from the intruders need to be removed to safeguard the data. For that many security algorithms have been designed for resolving authentication problems. Machines are trained effectively with different classifiers in identifying the invalid packet and tested the accuracy in identifying the intruder packets. Compared with the traditional methods the machine learning algorithms can identifying the intruder packets with high accuracy [1-2]. IDS using machine learning will act as a shield in protecting the network from different attackers and hackers.

Machine Learning

The extraction of knowledge from the available data by machines is known as machine learning. The machines can gain the knowledge from the data, identify the existing patterns, and can make decisions with minimal human intervention. Machines need to be trained and tested for the accuracy of the results. Big data is defined to have more volume, velocity, and value. Big data is to be processed for attaining a good quality for training the machine to achieve good accurate results. For accessing the available information more attacks will happen with the network infrastructure which needs to be

protected. Machine learning algorithms are used for quick and automatic and accurate results from large scale of data. In machine learning, the dataset will be divided into the dependent target and independent feature variables. The machines are trained with the possible number of attacks to the networks with the available information of past or historical data and tested for the accurate results. Machine learning algorithms which can learn from the available data and whose performance can be easily measured are called as supervised machine learning algorithms. On the other side where data inputs are known and no one to training on the results can be understood by applying unsupervised algorithms for knowing the hidden patterns and classifying the data based on the knowledge gained and which are harder to understand and evaluate. Reinforcement algorithms that improve upon itself and learns from the new situations by using trial-and-error methods.

If the given data is high-dimensional it must be reduced to low-dimensional to receive good results in the prediction of the attacks for safeguarding the network infrastructure in protecting the available information. The available dataset must be pre-processed for removing the unwanted and the duplicate and irrelevant data to attain the data having a good quality to train the machine. Later the pre-processed data is tested with different supervised classification algorithms for calculating the accuracy of the prediction of the attacks from the intruders. The paper is divided into five sections addressing the recent works in section 2 and proposed methodology in section 3 and all implemented results are alleviated in section 4 and conclusion is given at section 5.

Related Works

There are several algorithms that have been developed to provide network security from the unauthorized to modify and steal the sensitive data like the health data. Now a day's data can be accessed from multiple sources like mobile devices, sensors etc., which provide threats to the data. Organizations safeguard their networks by using a network security tools such as firewalls or IDS. which can cause major damage if not detected early. Some of the recent related works are reviewed in this section. Ahmed S. Alzahrani et.al.[3] Proposes an effective model for identifying the intruders in networks by employing the machine learning strategies by examining the important and discriminative features by applying the structural sparse logistic regression and SVM methods with radial basis function. Experimental results are attained by using NSL-KDD dataset.

Dr. D. Bhavana et.al.[4] The paper discusses about the three datasets like KDDCUP 99, NSLKDD and UNSW-NB15. The experimental work is done in the Google's Tensor Flow Platform with Gradient Boosted Decision Trees achieving highest accuracy when compared with different machine learning classifiers. Nine common cyber attacks are discussed in this paper. Gautam M. Borkar et.al.[5] presented a Novel method of clustering which reduced the time factor a lot and scalability is also improved. A two-stage clustering technique applying SVM classifier algorithm, and the acknowledgement of the attack is sent to the sensor nodes for secure packet transmission in python platform. Nivedita S Naganhall et.al.[6] presented paper Distributed denial of service DDOS, Servers unable to deal with huge requests and enters offline for a long period and to detect and deal with such requests the traffic is to be monitored which involves much data, so the machine learning algorithms are applied in detection of bad intruders and removed from the network. Using Spyder tool the paper distinguishes the normal and malicious traffic.

C4.5 and NB are used and C4.5 is more efficient in detection of DDOS attack. Dataset is pre-processed and the feature selection algorithms are applied to reduce the dimension and detection time for attaining good accuracy. Eirini Anthi et al.[7] presents three layer IDS where the first layer classifies the type of the normal IOT devices and detecting the wireless attacks for the devices and classifies the types of attacks happened to the devices. Tested with the 8 devices and 12 network attacks are identified from four attacks which can distinguish between the malicious and benign and in future expected to deal with real time traffic so that can understand better about even more different attackers and how to protect the valuable information.

2. Materials And Methods

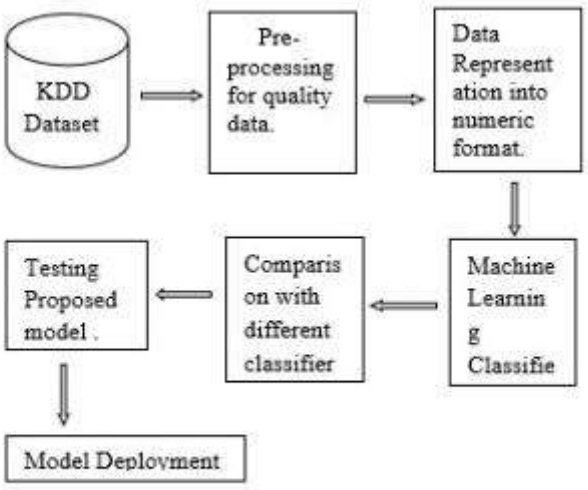


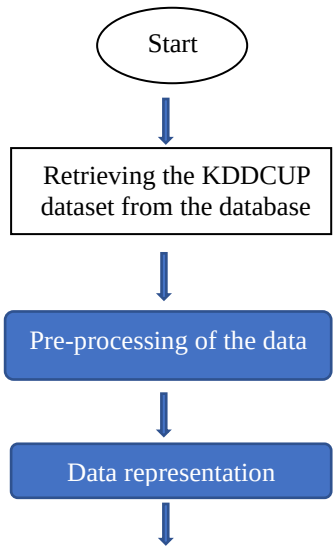
Figure 1: Proposed NIDS model for network security.

For detecting the network intruders various machine learning models are proposed which can detect the network traffic for unknown users' access with a good accuracy. To bring even more accuracy NIDS model is proposed as data is more important.

The proposed work deals with the prediction of attacks in protection of underlying network infrastructure from unauthorized access, modification & destruction for secure platform. Network traffic must be monitored for suspicious activity and detected and removed. The KDDCUP 99 dataset is retrieved from the UCI Repository having 42 features. The quality of the data should be good to retrieve accurate results. Pre-processing of the data should be done to remove the invalid data, unwanted outliers, and the missing values in the data so that the machine can be trained properly, and the accurate prediction results can be obtained.

After Pre-processing the data should be in numeric format for the analysis of machine in understanding the different attacks from the KDDCUP datasets. As some of the features are represented as categorical and object datatypes the whole data of the dataset is converted into numeric format so that the machine can make the predictions after training. So, the machine is trained properly and tested for accurate results.

Intrusion detection system is important in monitoring the network traffic to generate alerts when any attacks appear. Network intrusion detection system efficiently monitors the network traffic for the detection of unwanted packets. The machines must train properly in identifying the invalid packets and the results should be accurate. The KDDCUP data is further applied with the supervised machine learning classifiers by dividing the data into 70 and 30% ratio for training and test data. The different machine learning classifiers are tested for good accuracy and other checked with other performance measures and the data model is evaluated. The workflow of NIDS is depicted in fig 1.



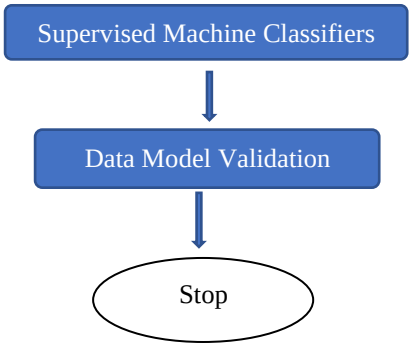


Figure 2: Flow chart of the proposed system

Pre-processing of dataset

The dataset should be pre-processed to remove the irrelevant data like missing data, outlier data and unwanted data. There are different stages of dealing with the data like cleaning the data by removing the unwanted and reducing the dimensions of the data and transforming the data into one scale. Later the dataset has to be split into training and test dataset and the machine is trained.The dataset consists of categorical features which need to be encoded into numerical values.

NIDS types

A Network Intrusion Detection System identifies the different attacks by monitoring the incoming packets of network. An agent which is present on the client machine which can detect the bad traffic even though it is encrypted is a host -based detection system of intrusion. A perimeter intrusion detection system detects the attempts made other than regular channel to the main server it triggers an alarm by acting like a fencing to the server. A virtual machine-based intrusion detection system is deployed remotely via a virtual machine with a good internet connection.

Representation of Data

Machines are good in learning from the data represented in one format usually the numeric format. As we are receiving data from different sectors different representations of the same data will cause more duplication, more storage space, much of processing time and inaccurate results. So, data pre-processing plays a major role in attaining the quality data. Regardless of the types of data features we collected the representation of will have an enormous effect of the performance of the machine learning models. Representation of data in a right way will have a greater influence on the performance of a supervised model than the exact parameters chosen. All the categorical features of the dataset are scaled into one format for the better understandability by the machine.

Supervised Machine Learning

Supervised machine learning is used in the prediction of a new and unseen data based on the knowledge gained by the machine using training data. Supervised machine learning classifiers are used in basically classifying the different network intruders based on the available data. Later the accuracy of the test data is checked for efficient results. There are different classifiers like KNN, LR, Decision tree, SVM, MLP, RF, GB etc., which can be used to check for the accurate results. We are using Random Forest machine learning classifiers here which is gives us with accurate results compared with other classifiers.

Data set description

For our research KDDCUP Dataset is considered as the best dataset obtained from UCI repository. It contains 494020 records of 42 features out of this one attribute is class variable. The objective of the dataset is to detect the different intruders for the network traffic based on the available data records. A predictive model for distinguishing between the bad intrusions and the normal connections.

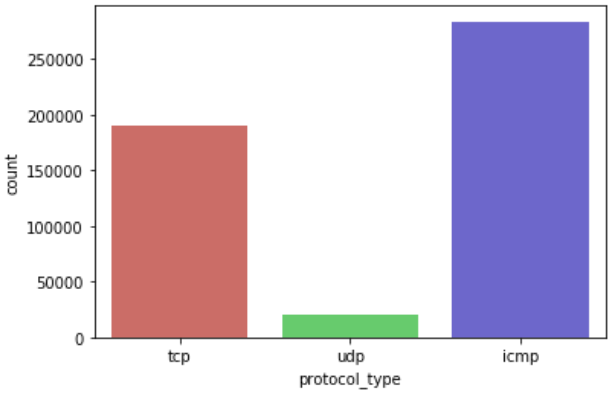


Figure 3: Different protocol types

3. Results and Discussion

The proposed algorithm was simulated in the Jupyter notebook. The results that were obtained using the NIDS was compared with the other machine learning algorithms. NIDS model after applying the classifiers the accuracy of the predictions is evaluated by the confusion matrix by predicting the four outcomes**.** In relation to surgical interventions, both Yuan ([15]) and Sulejmanajic [33] highlight the importance of a careful and minimally invasive surgical technique to reduce the risk of postoperative complications in patients with CKD. This precise and meticulous surgical approach is essential to preserve tissue integrity and minimize risks associated with procedures, emphasizing the importance of specialized and personalized care for this vulnerable population.

Table 1: Confusion_matrix

	Predicted Negative	**Predicted positive**
Actual Negative	TN-True Negative	FP-False Positive
Actual Positive	FN-False Negative	TP-True Positive

Accuracy=TP+TN

TP+TN+FP+FN

Recall= TP

TP+TN

Precision= TP

TP+FP

F1-score=Precision. Recall

Precision + Recall

TPR= True Positives / All Positives

FPR= False Positives / All negatives

Table 2: Accuracy results for KDDCUP using NIDS

Parameter	**Random Forest Model**
Accuracy	0.9999
Precision	1.000
Recall	1.000
F1_score	1.000
Confusion Matrix	[148196 0 3 7]

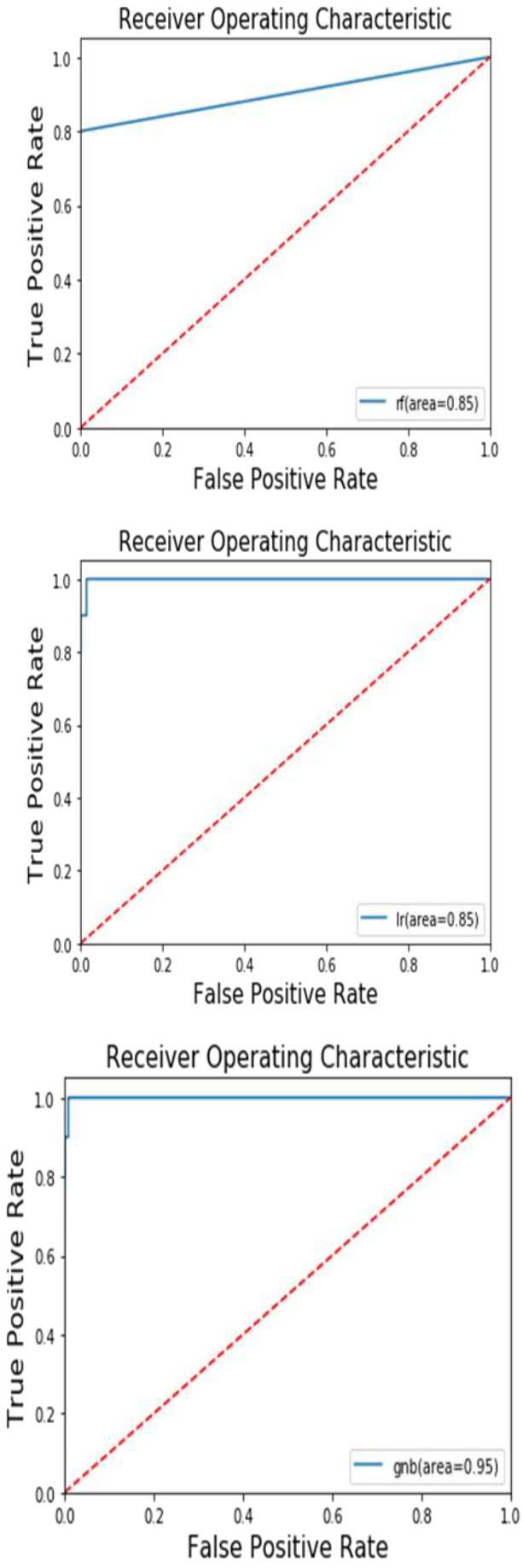


Figure 4: ROC Curve for KDDCUP dataset for different classifiers

Table 3: Comparison of results with other algorithms

Author	Proposed Algorithm	Accuracy
NIDS model	Random Forest	99.997%
Almseidin	Random Forest	93.77%
Nivedita S Naganhalli	C4.5 decision tree	99.0%
Eirini Anthi	Three layer IDS	98.0%
Sarika Choudhary	DNN	Above 90%
Ch. Aishwarya	Random Forest	99.99%

4. Conclusion

This paper has dealt with the development of a accurate prediction model for the detection of the different attackers of the network traffic for the safe guard of the valuable information. The NIDS model was trained with the KDDCUP dataset and then the results are tested for the detection of different bad attacks and the normal traffic. The pre-processed data is classified for good accuracy using the supervised machine learning algorithms. Among the different classification algorithms, the random forest algorithm attains very good accuracy in detecting the attackers and protecting the data. Later in future we should take care of how this valuable information can be secured using remaining other different types of attacks. The random forest algorithm yielded promising results of obtaining an accuracy of 0.999 when compared with the existing algorithms. This justifies the performance of the NIDS algorithm.

References:

[1] Francesco Restuccia, Salvatore DOro, and Tommaso Melodia. Securing the internet of things in the age of machine learning and software-defined networking. IEEE Internet of Things Journal, 5(6):4829–4842, 2018.

[2] Mohammad Reza Mahmoudi, Shahab S. Band. "Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries." Alexandria Engineering Journal , Vol.60 , pp.457-464, 2020.

[3] Mohammad Almseidin, Maen Alzubi, Szilveszter Kovacs, Mouhammd Alkasassbeh. " Evaluation of Machine Learning Algorithms for Intrusion Detection System." European Union Journal ,2020.

[4] Dr. D. Bhavana, Dr. K. Kishore Kumar, Vishnu Chilakala, Hemanth Gupta Chithirala, Tejesh Reddy Meka. "A Comparison Of Various Machine Learning Algorithms In Designing An Intrusion Detection System." International Journal of Scientific & Technology Research Volume 8, Issue 12, December 2019.

[5] G.M. Borkar ,Leena H.Patil, Dilip Dalgade, Anukull Hatke. "A novel clustering approach and adaptive SVM classifier for intrusion detection in WSN: A data mining concept." Sustainable Computing: Informatics and Systems 23 (2019) 120–135

[6] Nivedita S Naganhalli, Dr Sujata Terdal. "Network Intrusion Detection Using Supervised Machine Learning Technique." International Journal of Scientific & Technology Research Volume 8, Issue 09, September 2019.

[7] Eirini Anthi, Lowri Williams, and Pete Burnap. Pulse: An adaptive intrusion detection for the internet of things, 2018.

[8] Rohan Doshi, Noah Apthorpe, and Nick Feamster. Machine learning ddos detection for consumer internet of things devices. arXiv preprint arXiv:1804.04159, 2018.

[9] Amar Amouri, Vishwa T Alaparthy, and Salvatore D Morgera. Cross layer-based intrusion detection based on network behavior for iot. In Wireless and Microwave Technology Conference (WAMICON), 2018 IEEE 19th, pages 1–4. IEEE, 2018.

[10] Yair Meidan, Michael Bohadana, Yael Mathov, Yisroel Mirsky, Asaf Shabtai, Dominik Breitenbacher, and Yuval Elovici. N-baiotnetworkbased detection of iot botnet attacks using deep autoencoders. IEEE Pervasive Computing, 17(3):12–22, 2018.

[11] Olivier Brun, Yonghua Yin, and Erol Gelenbe. Deep learning with dense random neural network for detecting attacks against iot-connected home environments. Procedia computer science, 134:458–463, 2018.

[12] Christopher D McDermott, Farzan Majdani, and Andrei V Petrovski. Botnet detection in the internet of things using deep learning approaches. In 2018 International Joint Conference on Neural Networks (IJCNN), pages 1–8. IEEE, 2018.