Advances in Parallel Computing Technologies and Applications
D.J. Hemanth et al. (Eds.)

doi:10.3233/APC210116

Intrusion Detection System Using Convolutional Neural Network on UNSW NB15 Dataset

Mahalakshmi G [a,1], Uma E [b], Aroosiya M [c] and Vinitha M [c]
[a] *Teaching Fellow, Department of IST, Anna University, Tamil Nadu*
[b] *Assistant Professor, Department of IST, Anna University, Tamil Nadu*
[c] *Student, Department of IST, Anna University, Tamil Nadu*

Abstract. Networks have an important role in our modern life. In the network, Cyber security plays a crucial role in Internet security. An Intrusion Detection System (IDS) acts as a cyber security system which monitors and detects any security threats for software and hardware running on the network. There we have many existing IDS but still we face challenges in improving accuracy in detecting security vulnerabilities, not enough methods to reduce the level of alertness and detecting intrusion attacks. Many researchers have tried to solve the above problems by focusing on developing IDSs by machine learning methods. Machine learning methods can detect datas from past experience and differentiate normal and abnormal data. In our work, the Convolutional Neural Network(CNN) deep learning method was developed in solving the problem of identifying intrusion in a network. Using the UNSW NB15 public dataset we trained the CNN algorithm. The Dataset contains binary types of '0' and '1' in general for normal and attack datas. The experimental results showed that the proposed model achieves maximum accuracy in detection and we also performed evaluation metrics to analyze the performance of the CNN algorithm.

Keywords. intrusion detection, anomaly detection, deep learning, convolution neural network, UNSW NB15.

1. Introduction

Networks play an important role in our current life, using the network we transfer datas easily while transferring datas we can face many security threats to avoid any vulnerabilities we use cyber security. Cyber security is a technique which prevents anonymous attacks in networks like anti-virus software, firewalls etc. But they are not strong enough to detect a new type of attack. To improve the network security Intrusion Detection System(IDS) is introduced. IDS used to detect, monitor and analyze any vulnerabilities for both software and hardware running during a network. The following Figure 1 shows the overview of intrusion detection: here the firewall acts as Intrusion Detector; it stands between networks and filters traffic that might be unhealthy. Network Security can be

[1] Mahalakshmi G, Department of IST, Anna University, Tamil Nadu.
E-mail: mlakshmig27@gmail.com.

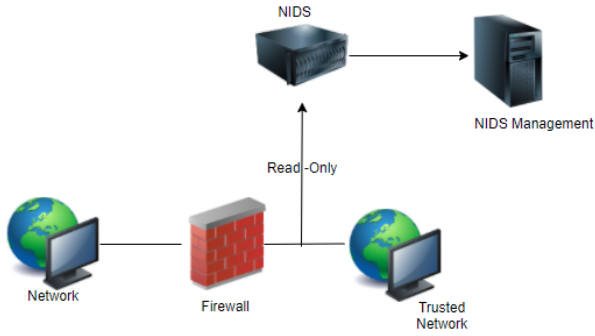

Figure 1. Overview of Intrusion Detection System.

monitored by administrators in network and security officers to provide a protected environment for user accounts, their online resources, personal details and passwords. IDS can be divided into two types by their approach:

1. Misuse Detection: It always uses signatures from previous data to detect intrusion; it may not be effective for new types of attacks.
2. Anomaly Detection: It uses an unusual pattern to detect attacks.

Here we use Anomaly detection for IDS. Attackers may act in two ways to try out their attacks in networks; 1) They create unavailability of network service for users. 2) Violating their personal information on the network. There are many types of attacks in networks. Denial of Service (DoS) is one among the frequent cases of attacks on network resources. It makes the network unavailable for users and creates traffic to crash their system. There are different sorts of DoS attacks, and each type has its own behaviour by intruding network resources of users for their own purpose which is to render the network unavailable for its users. Remote to Local (R2L) is one type of network attack, attackers send some type of files to gain unauthorized access to enter victim machines. User to Root (U2R) attacks are similar to r2l attacks; it enters the root machine illegally to crash the local machines. Probing is another type of attack in which intruders scan victim devices to find any weak spot in their machine to gain illegal access for their future attacks .There are many examples that represent probing over a network, like nmap, portsweep, ipsweep.

Table 1. Different Classes of Attacks

Attacks	**Description**
Denial of Service (DoS)	An attacker tries to shut down the victim's machine when they need to and create a traffic jam to crash the machine.
User to Root (U2R)	An attacker enters the local machine to crash the root privilege.
*R*emote to Local (R2L)	An Attacker tries to get unauthorized access to gain information into the victim machine.
Probe	An attacker attacks in which intruders scan victim devices to find any weak spot in their machine to gain illegal access for their future attacks mple port-scan and ping-sweep.

By using IDS we can prevent this type of attack. IDS uses classification techniques to form some sort of decision about every packet the network undergoes whether it's a

traditional packet or an attack packet. Software to detect network intrusion and protect the network from unauthorized users. Intrusion learning models are used to create a predictive model to prevent attacks and it distinguishes the connection as "good" for normal and "bad" for intrusion.

2. Related Work

This chapter gives a survey of literature work done by other researchers. I've learned some existing techniques from their research work, few of them are discussed below.

Coelho et al, [1] used homogeneity of data cluster and label to form a semi-supervised data for feature selection. This method enhances the performance of the feature selection process. Mutual Information is employed during a Forward-Backward search process so as to gauge the relevance of every feature to info distribution and therefore the existent labels, during a context of few labeled and many unlabeled instances.

Gharaee and Hossein [2] proposed a genetic algorithm and SVM with a new feature selection technique to improve the IDS. The new feature selection method based on a genetic algorithm with innovative fitness function to increase the true positive rate and reduce the false positive simultaneously reduces the time taken for execution. They performed their work on KDD CUP 99 and UNSWNB 15 dataset.

Gul and Adali [3] proposed a feature selection process for Intrusion Detection. Feature selection is an important process before classification is performed. When selecting the important feature it will reduce the execution time and increase the accuracy of the model.

Zhang and Wang [4] proposed an effective wrapper based feature selection to increase the accuracy of the algorithm. The wrapper method feature selection is based on Bayesian Network classifier.

Moustafa et al, [5] compared the signature based network intrusion detection that Anomaly based detection is more efficient. Anomaly does not follow patterns like signature based detection. The Authors evaluate their classification algorithm with two benchmark datasets of Network Intrusion Detection System (NIDS) NSL-KDD and KDD99 and find out that the datasets may be lacking in accuracy because of poor recent attack types, so the author used UNSW NB15 dataset. The author shows that evaluation of UNSW NB15 is done in three aspects to find its complexity. Also the system designed by [6] offered higher accuracy based on optimiztion in real time.

Intruders use more enhanced techniques to break the security so enhancement in IDS is needed . Primartha and Tama [7] used three different (UNSW NB15, GPRS, and NSL-KDD) datasets to perform classification process using Random forest, Naive Bayes, and Neural Network to get high accuracy and low warning rate and K-cross validation is done.

Selvakumar et al, [8] proposed a novel intelligent intrusion detection for multi-class classification data. They have used the KDD CUP dataset. The dataset is preprocessed and FR algorithm is applied to get best features for classification. they get 99.7% accuracy for intrusion detection. Compared with existing models they achieved a high accuracy rate.

Belouch et al, [9] proposed a two-stage classifier supported RepTree algorithm and protocol subset for network intrusion detection systems. To gauge the performance of

their approach, they used UNSW-NB15 and NSL KDD dataset. The feature technique is used to reduce the get best features here they get 20 best features out of 40. They have achieved detection accuracy of 88.95% and 89.85% on the UNSW-NB15 and NSL-KDD dataset.

Dhanabal and Shantharajah [10] used an NSL-KDD dataset and applied a different classification algorithm to detect the effectiveness of the classification algorithm in anomaly detection.

Tama et al [11] proposed hybrid feature selection and two-level classifier ensembles algorithm to improve the IDS. They have used NSL-KDD and UNSW NB15 dataset to perform their algorithm. In hybrid feature selection there are three methods(genetic algorithm, particle swarm optimization, ant colony algorithm) used to reduce the size of features in the datasets.

Selvakumar et al [12] proposed the FRNN approach to improve accuracy by reducing false positives in Wireless Sensor Network (WSN). They have used a traced dataset and applied Allen's interval algebra for preprocessing and selected important features using the Fuzzy algorithm. They have achieved 99.87% accuracy compared with existing models.

Vanthana et al, [13] proposed an optimal packet concept to increase the effectiveness in the intrusion detection. They introduce an indexing technique to reduce complexity and increase the accuracy in network intrusion detection. They use traced file datas.

Dahiya and Srivastava [14] proposed a framework during which a feature reduction algorithm is employed for reducing the smaller features than applied the supervised data processing techniques on UNSW-NB15 network dataset for fast, efficient and accurate detection of intrusion within the Netflow records using Spark.

Osama Faker [15] combined big data and IDS to create an efficient IDS for a large number of datas. Here, CICIDS2017 and UNSW NB15 datasets are used to perform the classification. homogenetic metrics are used to select the best feature for classification and there are three algorithms used for classification techniques are Deep Feed-Forward Neural Network (DNN), Random Forest and Gradient Boosting Tree. They get a high accuracy rate and 5-fold cross validation is done on Machine learning models.

3. Proposed Algorithm

In existing machine learning based IDS, always depending on the previous data may not be effective for newly generated attacks. The proposed deep learning model is dynamic and it can also be used for unusual patterns.

3.1. Convolution Neural Network (CNN)

In this proposed work Convolution Neural Network (CNN) used as a learning model for classification in IDS. Convolution Neural Networks (CNN) is designed to mimic the human visual system (HVS). It is made up of several neurons with learning weights and biases. CNN accepts a large number of inputs and takes the weighted sum of those inputs and sends them to the activation function to give output. A CNN is stacked with alternate convolution, activation pooling and fully connected layer. Figure 2 represent the CNN and their Layers like convolutional and pooling layer

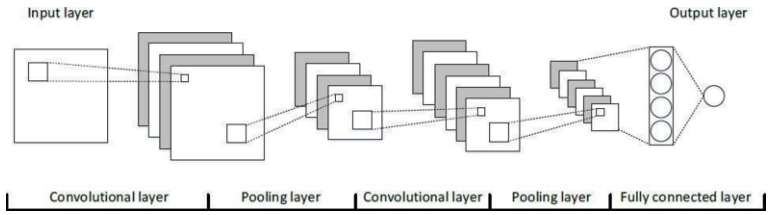

Figure 2. CNN Layout.

3.2. System Architecture

Figure 3 represents the system architecture. when the unsw nb15 dataset is given as an input and the given input splitted into trained and test data. CNN algorithm applied to the splitted data and we get a trained model. That trained model and test data are compared and we performed evaluation metrics for that data.

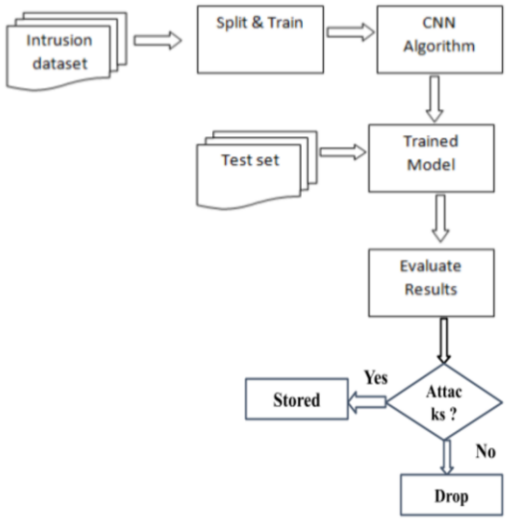

Figure 3. Architecture Diagram.

4. Implementation

Here the dataset is taken from kdd.ics.uci.edu.The downloaded dataset contains train and test data and the outputs are classified into different classes with binary value “0” and”1” for normal and attacked data. The train dataset is considered as the train set and the test dataset is considered as the test set. CNN is applied for the classification process in this work and evaluates the algorithm with performance metrics.

4.1. Dataset Detail

The algorithm is trained using the UNSW NB15 public dataset. The dataset contains 1,75,341 recorded datas as training data and 82,332 recorded datas as testing data. It has 49 features and it is categorized into six groups like flow, time, content, etc. It recorded 9 different types of recent and common attacks like Dos, fuzzers,backdoors, worms, etc. the output categorized into binary values as "0" and "1" for attack and normal data.

4.2. Functional Requirement

4.2.1. Data Collection

The data collection is a process of collecting relevant data for their work. The job of this process is to collect and analyze the data whether the data needed for their work. Here we collected the unsw nb15 network intrusion dataset from uci.edu.

4.2.2. Data Visualization

The purpose of visualization is for easy understanding. visualization can be shown in graphs, diagrams, slides, etc. Here we represent the graph. It shows the intrusion accuracy of the learning model which we have used.

4.2.3. Data Preprocessing

The purpose of preprocessing is to convert the raw data into the form that fits for machine learning models. For data preprocessing the data will be normalized and categorical features converted into numerical form by data encoding method.

4.2.4. Dataset Splitting

After preprocessing the dataset will be splitted into training and testing data. Training data is modeled into a form of algorithm fit. Testing data is used to evaluate the training data.

4.2.5. Model Training

After preprocessing and data will be splitted into training and testing data. A deep learning model is applied to the training data and we will get a trained model.

4.2.6. Model Evaluation

We will compare the trained model with testing data and determine the accuracy of the deep learning model. Evaluation metrics will be evaluated to find the performance of the algorithm. Here we used MSE, MAE, R-Square, RMSE for evaluation metrics.

5. Result

The proposed work is done with a deep learning model to improve the Intrusion Detection System. Here, CNN deep learning model is used to find the accuracy of IDS from UNSW NB15 dataset. The proposed work is done in python 3.7 with libraries of

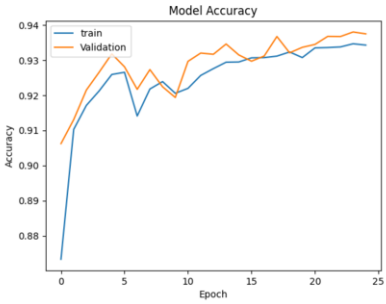

Figure 4. CNN Model Accuracy

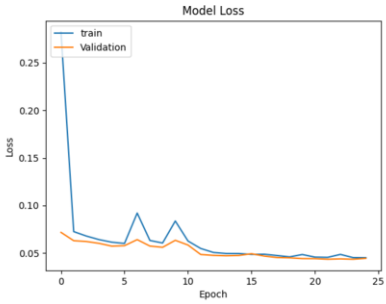

Figure 5. CNN Model Loss

keras, tensorflow, matplotlib and other mandatory files. Here we get 93.5% of algorithm accuracy using CNN and Evaluation metrics calculated to find the performance of the algorithm. The result shows that intrusion detection is efficient using CNN algorithm. The Figure 4 shows the model accuracy of training and validation datas by using CNN algorithm with 25 epoch. The Figure 5 shows the model loss accuracy of training and validation set using CNN algorithm with 25 epoch. The Figure 6 shows the evaluation metrics of CNN algorithm to analyze the performance of the model.

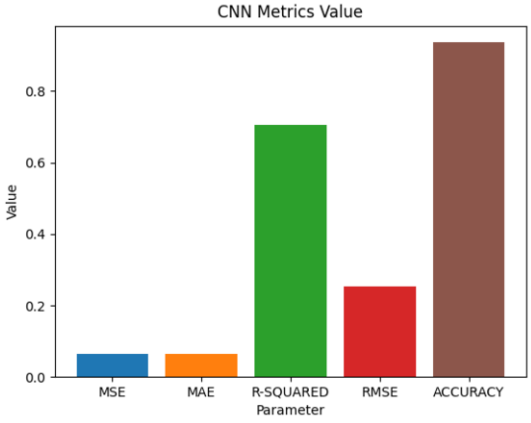

Figure 6. CNN Evaluation Metrics.

Evaluation metrics are used to survey the classification of the statistical learning model. Evaluating the learning models or algorithms is consequential for any project. There are many distinct types of evaluation metrics available to test a model. Mean Square Error (MSE) is a mean of Squared Error it is the difference between actual and predictive value. Mean Absolute Error (MAE) measures the difference between two variables and absolute error of each prediction error. R-Squared measures the goodness of fit of a regression model. Root Mean Square Error (RMSE) measures the square root of MSE value. Accuracy is the total number of predicted values by Total number of original value

6. Conclusion

The proposed work is to improve intrusion detection efficiency though we have many existing IDS mostly developed in the Machine learning algorithm that fails to provide strong IDS to prevent from newly formed attacks because it mostly depends on previous data. Here, CNN deep learning model is used for developing the IDS. By using UNSW NB15 network intrusion public Dataset we perform the classification technique by applying CNN algorithm and we get 93.5% accuracy. The accuracy shows that CNN is efficient in Intrusion detection and evaluation metrics also performed to analyse the performance of the model.

References

[1] Coelho F, de Pádua Braga A, Verleysen M. Cluster homogeneity as a semi-supervised principle for feature selection using mutual information. In European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning 2012 (pp. 507-512).

[2] Gharaee H, Hosseinvand H. A new feature selection IDS based on genetic algorithm and SVM. In2016 8th International Symposium on Telecommunications (IST) 2016 Sep 27 (pp. 139-144). IEEE.

[3] Gül A, Adalı E. A feature selection algorithm for IDS. In2017 International Conference on Computer Science and Engineering (UBMK) 2017 Oct 5 (pp. 816-820). IEEE.

[4] Zhang F, Wang D. An effective feature selection approach for network intrusion detection. In2013 IEEE eighth international conference on networking, architecture and storage 2013 Jul 17 (pp. 307-311). IEEE.

[5] Moustafa N, Slay J. The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. Information Security Journal: A Global Perspective. 2016 Apr 4;25:18-31.

[6] Yang R. UAV landmark detection on fast region-based CNN. Arabian Journal of Geosciences. 2021 Jun;14(12):1-9.

[7] Primartha R, Tama BA. Anomaly detection using random forest: A performance revisited. In2017 International conference on data and software engineering (ICoDSE) 2017 Nov 1 (pp. 1-6). IEEE.

[8] Selvakumar K, Sairamesh L, Kannan A. Wise intrusion detection system using fuzzy rough set-based feature extraction and classification algorithms. International Journal of Operational Research. 2019;35(1):87-107.

[9] Belouch M, El Hadaj S, Idhammad M. A two-stage classifier approach using reptree algorithm for network intrusion detection. International Journal of Advanced Computer Science and Applications. 2017 Jul;8(6):389-94.

[10] Dhanabal L, Shantharajah SP. A study on NSL-KDD dataset for intrusion detection system based on classification algorithms. International journal of advanced research in computer and communication engineering. 2015 Jun 6;4(6):446-52.

[11] Tama BA, Comuzzi M, Rhee KH. TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. IEEE Access. 2019 Jul 11;7:94497-507.

[12] Selvakumar K, Karuppiah M, SaiRamesh L, Islam SH, Hassan MM, Fortino G, Choo KK. Intelligent temporal classification and fuzzy rough set-based feature selection algorithm for intrusion detection system in WSNs. Information Sciences. 2019 Sep 1;497:77-90.

[13] Vanthana G, Muthurajkumar S, Sairamesh L, Rakesh R, Kannan A. Optimal packet classification techniques for performance enhancement and intrusion detection. Advances in Natural and Applied Sciences. 2015 Jun 1;9(6 SE):311-6.

[14] Dahiya P, Srivastava DK. Network intrusion detection in big dataset using spark. Procedia computer science. 2018 Jan 1;132:253-62.

[15] Faker OM. Intrusion detection using big data and deep learning techniques (Master's thesis).

[16] Dhanalakshmi B, SaiRamesh L, Selvakumar K. Intelligent energy-aware and secured QoS routing protocol with dynamic mobility estimation for wireless sensor networks. Wireless Networks. 2021 Feb;27(2):1503-14.