International Journal of Information Security and Privacy
Volume 16 • Issue 1

Anomaly-Based Intrusion Detection Using Machine Learning:
An Ensemble Approach

R. Lalduhsaka, Mizoram University, India*
Nilutpol Bora, Delhi Technological University, India
Ajoy Kumar Khan, Mizoram University, India

ABSTRACT

Intrusion detection systems were developed to detect any suspicious traffic in the network. Conventional intrusion detection comes with its sets of limitations. The authors aimed to improve anomaly-based intrusion detection using an ensemble approach of machine learning. In this article, CICIDS2017 and CICIDS 2018 datasets have been used for implementing the proposed method. Random forest regressor is used for feature selection. Three machine learning algorithms (i.e., naïve bayes, QDA, and ID3) are selected and combined (ensembled) for their low computational cost. The ensemble algorithm results are compared with the standalone algorithms. With the ensembled method, classification accuracy of 98.3% and 95.1%, with FAR of 2% and 6.9% were achieved on CICIDS 2017 and CICIDS 2018 datasets respectively. Naïve bayes, QDA, and ID3 have classification accuracies of 82%, 84.7%, and 95.8% respectively on CICIDS 2017; 68.3%, 68.4%, and 94.4% respectively on CICIDS 2018; false alarm rates of 54.9%, 55.5%, and 20.6% respectively on CICIDS 2017; and 3.6%, 3.7%, and 7.1% respectively on CICIDS 2018.

KEYWORDS

CICIDS Datasets, Computational Cost, False Alarm Rate, Feature Selection, Random Forest Regressor

INTRODUCTION

Intrusion detection (ID) systems are renowned solutions for detecting malicious activities in a network. These ID systems have become an essential component of defense to network security infrastructure. The importance of these ID systems grows with the exponential growth of network attacks in modern networking systems. In 1931, John Anderson published the first significant paper on ID, Computer Security surveillance, and threat monitoring emphasizing the importance of such systems in security (Xu, Shen, Du & Zhang, 2018). An ID system usually monitors all internal and external packets of a network to detect whether a packet has a sign of violations (Modi et al., 2013). An ID system must be able to determine various kinds of attacks and send alarms when detecting them on the network.

ID systems are generally categorized into two types, based on the methodologies used; Signature-based method (Maleh et al., 2015) and Anomaly-based method (Zhang & Chen, 2017). Signature-based ID systems uses a pattern matching technique to detect an intrusion in the network. This is done by utilizing a database of known attack signatures that are compared with the network traffic; triggering

DOI: 10.4018/IJISP.311466 *Corresponding Author

an alert when a match is found. It has an extremely low false alarm rate (Maleh et al., 2021) and is efficient in detecting known attacks. However, since this class of ID systems are solely based on previous information available, they are ineffective against new attacks that are not available in the signature database. In contrast to signature-based ID systems, there is an Anomaly-based ID system that is capable of triggering unknown attacks (Fayssal, Hariri, & Al-Nashif, 2007). Anomaly-based or behavior-based ID systems analyze the network traffic on the basis of the behavior of the network. It defines the normal behavior of a network, and if any abnormal behavior is detected or the network deviates from the normal behavior of the network, an alert is raised. However, these systems are not very accurate when it comes to ID since the profiling of a network is a complex process and often leads to a high false alarm rate (Haq et al., 2015). Most approaches which are currently used in ID systems cannot properly deal with the complex and dynamic nature of malicious threats. Therefore, various methods of Machine Learning techniques are being sought after, to achieve a better detection rate (DR), false alarm rate, and computation costs (Zamani & Movahedi, 2013). The traditional ID techniques have few limitations in protecting a system; most notably, when systems are facing a high volume of malicious attacks (DoS/DDoS); systems can obtain high values of False Positives and False Negatives (Khan & Kim, 2021). Recently, numerous researchers have used Machine Learning techniques for ID systems to improve ID rates. Several studies have been done to enhance and apply this method to the ID systems (Wagh, Pachghare, & Kolhe, 2013). The Machine Learning models were found to have many issues that slow down the training process; these issues included the size of the dataset and the optimal parameters for the most suitable model. These kinds of problems prompted the researchers to look for the most effective methodology. However simple Machine Learning approaches are limited (Mahesh, 2020), while intrusion methods are expanding and growing complex. The authors make use of multiple Machine Learning classification algorithms to make a better prediction model on the network traffic for intrusions.

Different studies have been conducted focusing on how ID can make use of Machine Learning to detect zero-day attacks (Patidar & Khandelwal, 2018; Smys, 2019), which are the attacks not yet recognized by the ID systems. These attacks are unrecognizable in signature-based ID systems; however, anomaly-based ID systems are known to flag these attacks as well. But such anomaly-based systems were initially difficult to trigger an anomaly effectively (Buczak & Guven, 2015). This is where Machine Learning can help in improving the anomaly-based ID systems, by letting the system figure out what kind of traffic is classified as benign and what triggers an attack alert.

The main contribution of this work is the development of an efficient model for anomaly-based ID systems. And also tackle the limitations of existing ID systems such as false alarm rate, computation cost, and DR. In order to achieve this, the author uses random forest regressor feature selection method to select important features where 7 - 8 features are selected from around 80 features. Ensemble machine learning technique is used for final evaluation. Traditional Machine Learning Algorithms are implemented first and three algorithms with the lowest computation cost viz. Naïve Bayes, ID3 and QDA are selected for blending. The proposed method is implemented on CICIDS 2017 and CICIDS 2018 datasets. The ensemble method gives better results in both the dataset than the traditional machine learning algorithms individually.

This rest of the paper is organized as follows: Related Works highlights the recent researches on ID systems, the classification algorithms and the dataset used along with their performance. The Proposed Methodology section covers the steps of the proposed method, algorithms used and datasets analysis. Results and Discussion section gives the performance metrics of the proposed methods and comparison with its standalone algorithms as well as the recent researches presented in the related works. Lastly, the Conclusion section highlight the works presented in this paper and future scope of the work.

RELATED WORKS

ID systems had been in the field of research for more than three decades, and the research base on machine learning and deep learning is growing largely in the past two decades. Anomaly-based ID system is mainly developed for network ID systems and is also mainly using Machine Learning and Deep Learning techniques. There are different datasets for ID systems and the NSL-KDD dataset is considered the benchmark dataset for ID in the last decade (Javaid, Niyaz, Sun & Alam, 2016).

Latifur Khan, Mamoun, and Bhavani (2007) used Clustering Trees based on SVM to reduce the training time of SVM on the DARPA98 dataset. The accuracy rate for normal and Probe was observed as 98% and 88% respectively, but for DoS, U2R, and R2L, the accuracies were very low at 39%, 23%, and 15% respectively. The Clustering Trees-based SVM however, improve the pure SVM average accuracy from 57.6% to 69.8% as well as training time was reduced from 17.34 hours to 13.18 hours.

A study focusing on the importance of feature reduction for training an ID model was published in 2012 (S. Mukherjee and N. Sharma, 2012). Feature-Vitality Based Reduction Method (FVBRM) was used to reduce the NSL-KDD datasets from 41 features to 24. Naïve Bayes Classifiers' mean accuracy was improved from 95.11% to 97.78% when feature reduction was used in contrast to the selection of all attributes of the dataset. However, the U2R attack detection accuracy was low at 64%. The study implies that the selection of reduced features gives a better overall performance in designing an ID system in terms of efficiency and effectiveness.

Yassin, Udzir, Muda, Sulaiman (2013) proposed an integrated machine learning algorithm using the K-Means classifier and Naïve Bayes on KDD CUP '99 dataset. The proposed method is to address the high false alarm rate and to surpass the detection accuracies of the existing ID system. This model significantly improves the accuracy, DR up to 99% and 98.8%, respectively, while also decreasing the false alarm rate to 0.5%.

Kostas (2018) used seven machine learning methods; Naive-Bayes (NB), Random Forest (RF), K-Nearest Neighbors (KNN), Multilayer perceptron, Adaboost, ID3, and Quadratic Discriminant Analysis (QDA) independent to each other in order to compare their performance. For Feature reduction, Random Forest Regressor was used on CICIDS 2017 dataset and achieved an F-score as Naive Bayes: 0.86, QDA: 0.86, Random Forest: 0.94, ID3: 0.95, AdaBoost: 0.94, MLP: 0.83, and KNN: 0.97. The study also shows that the computation time of AdaBoost, MLP, and KNN are extremely long in comparison to NB, QDA, and ID3.

Gautam and Amit (2018) used Information Gain to select important features on the KDDCUP99 dataset, and an ensemble of Naïve Bayes, Adaptive Boost, and PART (Partial Decision Tree) for classification. The ensemble method improved the classification results of the imbalanced data and thus surpasses all the lone classifiers (Naïve Bayes, Adaptive Boost and PART).

Tama, Comuzzi, & Rhee (2019) proposed a two-level ensemble with hybrid feature selection using three evolutionary algorithms viz. Particle Swarm Optimization, Genetic Algorithm, and Ant Colony Algorithm. A Two-level ensemble using rotation forest and Bagging on NSL-KDD and UNSW-NB15 dataset achieving classification accuracy of 85.8% and 91.27% respectively which surpass the state of art individual and meta-classifier.

Yang, Sheng & Wang (2020) proposed a parallel Quadratic ensemble method on CICIDS 2017. The authors use Gradient Boosting Decision Tree to deal with the spatial data and Gated Recurrent Unit to deal with the temporal data. Combining both the algorithm to make a quadratic ensemble method. The proposed model achieves classification accuracies of 99.9% for different kinds of attacks on the CICIDS 2017 dataset.

Beulah and Punithavathani (2020) proposed a clustering-based outlier detection (CBOD) method for classifying attacks and benign in the NSL-KDD dataset. They used a hybrid of four built-in feature selection methods of WEKA for feature selection and selected 6 most relevant features. They achieve an accuracy of 97.96% and a low false alarm rate of 1.88%.

Mohammadpour, Ling, Liew and Alihossein (2020) proposed a Mean Convolution Layer (CNN-MCL) which is based on CNN architecture. It was developed for learning the anomalies' features

and identifying their abnormality. The proposed method is implemented on CICIDS 2017 dataset and achieved classification accuracy of 99.46%.

Varzaneh and Kuchaki (2021) proposed a new fuzzy rule-based classification system based on Genetic Algorithm. The proposed method is implemented on KDDCUP99 dataset. They achieve DR of 95.33% and False Alarm rate of 0.18%.

Zhao, P., Fan, Z., Cao, Z., & Li, X. (2022) proposed an Intrusion Detection model based on Temporal Convolutional Networks (TCN). Attention mechanism is applied to the proposed model. KDD CUP99 dataset and UNSW-NB15 dataset were used and achieved accuracy of 92.8% and 72.92% respectively.

PROPOSED METHODOLOGY

The proposed method is the combination of three Machine Learning classification algorithms with the purpose of improving the existing solutions. The proposed model uses Random Forest Regressor feature selection technique and then the ensemble technique of machine learning. The ensemble technique uses predictions from different models and learns how to best combine these models using some other machine learning model. In particular, for this study, the authors used the blending ensemble technique to combine these different machine learning models to make the ID system classify between attack and benign network traffic data.

In this paper, CICIDS 2017 and CICIDS 2018 datasets have been used for training and evaluating the proposed model. The dataset consisting of around 3 million and 6 million traffic data records respectively have been analyzed for discrepancies, cleaned, and preprocessed. This preprocessing is done to remove any inconsistent or incomplete records.

The architecture of the proposed ID System is as shown in Figure 1. The traffic data is initially preprocessed and cleaned, then the entire records labels are converted into 1 (attack) and 0 (benign) labels for classification. Random forest regressor is used to calculate the feature importance and an appropriate number of features were selected based on the weight of the features. This data was used for implementing machine learning algorithsm, the authors used the blending ensemble model; Naïve Bayes, QDA, ID3 for level-1 i.e., base models (Level-1) and random forest classifier as Meta-Model (level-2) for final predictions. The algorithms were selected from various traditional machine learning algorithms based on their respective performance and time taken to train the models.

Datasets

ID models especially anomaly-based requires an abundant of data for training the predictive model, consisting of both normal traffic and abnormal traffic (attacks). For this purpose, different datasets have been produced to simulate a real-world scenario of a network, there are multiple available datasets for ID. The authors have decided to use CICIDS 2017 for training and evaluation for it is one of the most recent and most realistic datasets for ID which is available publicly (Ankit Thakkar & Rikita Lohiya, 2020). It also contains newer types of attacks that are not available in the older datasets. CICIDS 2018 is also evaluated with the proposed model in order to validate the proposed model. Details of the dataset is shown in Table 1.

CICIDS 2018 dataset is much bigger than CICIDS 2017 as the traffic flow capture duration is 10 days and 5 days respectively. The number of attack infrastructure and victim infrastructure of the network where the traffic is captured is also much larger in CICIDS 2018. The CICIDS datasets were selected in this study due to their wide variety of attacks and protocol range in comparison to the older and benchmark dataset. The data collected had been obtained from a real-world using variety of computer systems and servers, and labelled which makes it an appropriate choice for machine learning purposes, especially in ID. Firstly, the CICIDS 2017 pre-processed data (CSV files) is selected as the dataset to be used in the implementation due to its size.

Figure 1. Proposed model

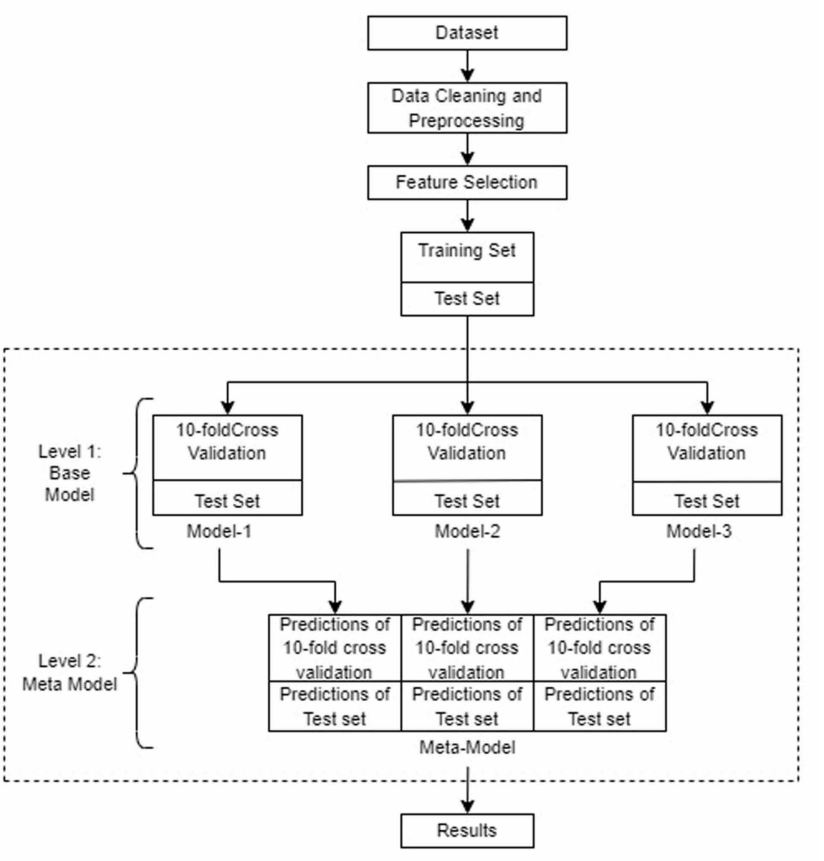


Table 1. Details of CICIDS 2017 and CICIDS 2018

Dataset	CICIDS 2017	CICIDS 2018
Capture duration	5 days	10 Days
Number of Classes	15	18
Number of features	80	83
Number of records/instances	28,30,743	162,33,022
Number of attacks	5,57,646 (19%)	27,59,364 (17%)
Number of Benign	22,73,097 (81%)	134,73,658 ((83%)
Data size (csv)	885Mb	7Gb
Types of attacks	DoS DDoS Brute Force Botnet Infiltration Web Attack	DoS DDoS Brute Force Botnet Infiltration Web Attack

Data Preparation

The CICIDS 2017 dataset consists of 8 CSV files, containing 5 days network traffic stream as briefed in Table 1. These CSV files were combined, forming a single CSV consisting of 28,30,743 records. On experimentation, there were few incomplete records seen in the dataset, which were removed from the file. It was observed that "Fwd Header Length" attribute of the dataset existed twice in the set, which was corrected by removing one copy of the attribute. "Flow Bytes/s" and "Flow Packets/s" contained infinity and NaN values (Missing values) which were converted into integer -1 and 0 respectively. It was also seen that multiple attributes contained string values, including the Label column, which was handled by converting all the benign stream records into integer 0 and the remaining attack records into integer 1, to be able to make these records fit for training.

In order to evaluate the performance of the model, the CICIDS 2017 dataset is fed to the model to compute the performance of the learning, done by the model. The dataset is hence split into two partitions: a training set and a test set. The machine learning algorithms try to learn from the training set and verified on the test set. However, the CICIDS 2017 dataset does not have these pre-defined splits, hence the data is divided into two parts manually using python libraries. Sklearn uses the train_test_split (Pedegrosa et al., 2011) function to split the dataset randomly such that 80% of the data is used for training the model i.e., training set and the remaining 20% of data as the test set. The results obtained over test data is the performance of that model, which is used to calculate the performance metrics.

The same method of preparation of dataset which includes data cleaning, converting character data into machine-readable data, and conversion of labelled data into 0 (benign) and 1 (attack) is performed on CICIDS 2018. This dataset consists of 10 CSV files of 10 days network traffic stream, out of which 7 CSV files were combined forming 1 CSV file; 3 CSV files were excluded due to its size and no presence of attack in the file. The combined CSV file contains 6.1 million records which are around 38% of the total records in the CICIDS 2018 dataset. It also contains a greater number of sub-types of attacks. The same ratio of train_test_split is used as in CICIDS 2017.

Feature Selection

CICIDS 2017 dataset consisted of 80 features with 28,30,743 records; CICIDS 2018 dataset consisted of 83 features with 162,33,022 records out of which 61,87,054 (38%) records were used for implementation. Using the dataset with all the features to train the machine learning model is computationally costly and also results in extended training times. Feature Selection is hence applied to the datasets in order to determine which features are important to define a traffic record as an attack or benign.

Random Forest Regressor fits plenty of decision trees on different sub-samples of the dataset. Random Forest Regressor is applied to the dataset to calculate the importance (weight) of each feature in the dataset. Random Forest Regressor uses averaging to control overfitting and also to improve the accuracy. The dataset contains only 1 (attack) and 0 (benign) labels; hence the feature importance will be focusing only on the importance of classifying between attack or benign data. The features list along with their weight importance is tabulated in Table 2.

Implementation Using Machine Learning

The Anomaly-based ID System devised uses the ensemble technique of machine learning, which is a technique that uses predictions from different models and learns how to best combine these models using some other machine learning model. This results in improving upon the effectiveness that can be achieved by any of these models independently. In particular, for this study, the authors used the blending ensemble technique to combine these different machine learning models. The combination of the algorithms makes the proposed ID model to classify between attack and benign network traffic data.

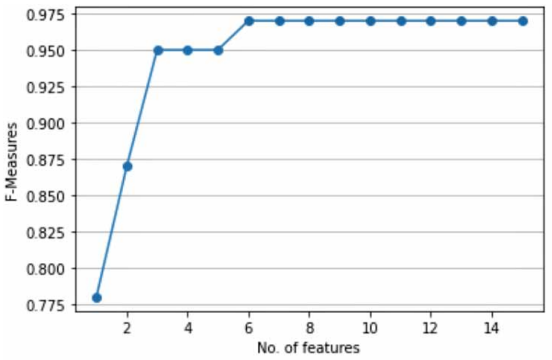
Table 2. Weight of features (up to 12 features)

CICIDS 2017		CICIDS 2018	
Features	**Weight**	**Features**	**Weight**
Bwd Packet Length Std	0.24662	Dst Port	0.275073
Flow bytes/s	0.17877	Flow IAT Max	0.066496
Total Length of Fwd Packets	0.10241	Flow Byts/s	0.048871
Fwd Packet Length Std	0.06388	Flow IAT Mean	0.012347
Flow IAT Std	0.00989	Bwd Pkt Len Min	0.013189
Flow IAT Min	0.00694	Fwd Pkt Len Max	0.006626
Flow IAT Total	0.00512	Bwd Pkt Len Mean	0.005703
Flow Duration	0.00415	Bwd Pkt Len Std	0.003946
Bwd Packet Length Max	0.00400	Fwd Pkt Len Mean	0.003443
Flow IAT Max	0.00357	Tot Bwd Pkts	0.002907
Flow IAT Mean	0.00326	Bwd Pkt Len Max	0.000898
Total Length of Bwd Packets	0.00130	TotLen Bwd Pkts	0.000376

The machine learning algorithms used in the level-1 are selected on the basis of experimentation performed on various Machine Learning classification algorithms; Model-1 is Naïve Bayes, Model-2 is QDA and Model-3 is ID3. These 3 algorithms are selected based on the results (accuracy & training time) achieved when applying them on the given dataset: Naïve Bayes (F-Measure: 0.86, Time: 1.825s), Quadratic Discriminant Analysis (QDA) (F-Measure: 0.86, Time: 2.37s) and Iterative Dichotomiser 3 (ID3) (F-Measure: 0.95, Time: 9.51s) (Kostas, 2018; and Haq et al., 2015).

These three base models are trained through 10-fold cross-validation, and examined over their respective test sets. These results are stacked to create new training and test set for the level-2 meta-model as shown in Figure 1. The meta-model is trained using the Random forest classifier algorithm, and prediction of this meta-model determines the final performance metrics of the system. The set of attributes to be used is selected on an experimental basis, where the authors chose the appropriate number of features to be used in the implementation using the feature importance computed as shown in Table 2 and the change of F-1 scores according to the number of features which is given in Figure 2.

Figure 2. Change in F-measure with number of features used (CICIDS 2017)



It is evident in the graph that there is little to no change in the F-scores after the selection of the 6 most important features from the feature importance shown in Table 2. Hence, the authors used the top 7 and top 8 features which makes up roughly 96% of feature importance to get optimal results for the machine learning implementation of the proposed ID model. This method of feature selection removes around 70 features in these CICIDS datasets.

RESULTS AND DISCUSSION

The effectiveness of the proposed ID model is seen through experiments applying on the CICIDS 2017 and CICIDS 2018 datasets via normal and attack classification. Classification accuracy, precision, recall, F-measure, False alarm rate ROC curve are used in this paper for evaluation. The computation time of the models are also shown in Table 4 and Table 5. The implementation of the proposed method was carried out on a PC with Intel core i5-10th Gen, 8Gb RAM and GPU of MX-350 running windows 10.

Performance Metrics

The performance of the model devised in this study is evaluated according to the following measures -accuracy, precision, recall, f-measure, false alarm rate and ROC curves. The given criteria range from 0 to 1, where 0 describes least and 1 describes the highest performance. The predictions obtained by the model are evaluated using a confusion matrix. The distribution of the confusion matrix is shown in Table 3.

Table 3. Illustration of Confusion Matrix

Actually Negative	**True Negative (TN)**	**False Positive (FP)**
Actual Positive	False Negative (FN)	True Positive (TP)
	Predicted Negative	**Predicted Positive**

These elements of the confusion matrix help in defining the performance measures- accuracy, precision, recall, f-measure, and false alarm rate as follows:

Accuracy: It is a metric that measures how correctly the classifier works, measuring the percentage of the data which are correctly classified. In another word, it is the ratio of successfully categorized data to the total data.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}$$

Precision: It is a measure that gives the percentage of the actual true positive identified by the classifier from all the instances which the classifier identified as positive. It is the ratio of true positive to the sum of true positive and false positives.

Recall: It is also called True Positive Rate (TPR). It gives the percentage of actual true positive from the actual positive in the data i.e True positives plus False negatives.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F -measure: The F-measure mixes the properties of the previous two measures as the harmonic mean of precision and recall. (Bhuyan, Bhattacharyya & Kalita, 2013)

$$\mathrm{F-measure} = \frac{2}{\frac{1}{\mathrm{Precision}} + \frac{1}{\mathrm{Recall}}}$$

False Positive Rate (FPR): Also called False Alarm Rate (FAR) is the rate at which there is wrong or false alarm. It is the ratio of the False positive by the sum of False positive and the True Negative.

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

ROC curve: Receiver Operating Characteristic (ROC) curve is a graph that shows the performance of the classification algorithm at various threshold. It is a curve that plots between the TPR and the FPR.

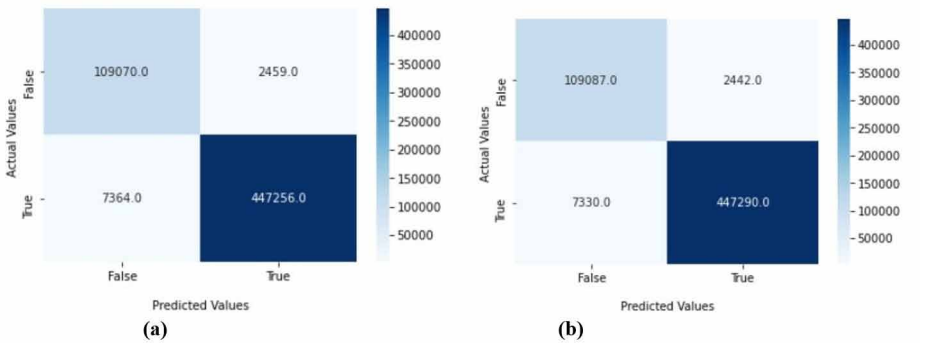
When the precision and recall both reach 1, the F-measure is the maximum, i.e., the classifier has no false alarms and detects all of the attacks. Thus, a classifier is expected to obtain F-measure as high as possible. The FAR is the rate for the false alarm and does not reflects the false negatives.

Evaluation of Machine Learning Model

The analysis done is carried out in accordance with the performance metrics discussed in the above section. CICIDS 2017 as well as CICIDS 2018 dataset (but not simultaneously) after data cleaning and preprocessing is split into a training set and test set; the training set is to train the model and the test set is to test the performance of the model.

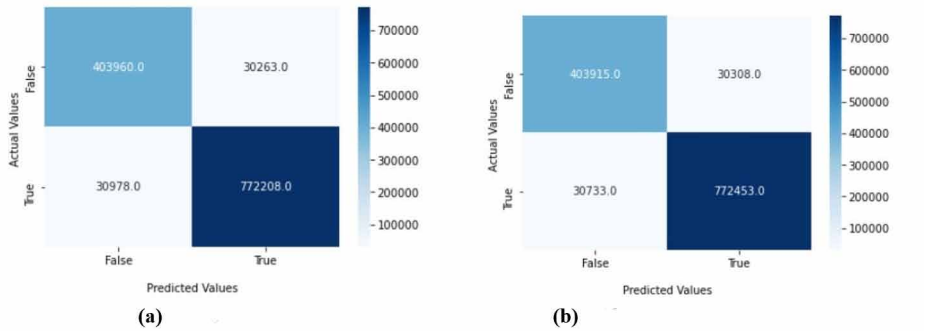
The confusion matrix for the ensembled model given in Figure 3 and Figure 4 illustrates the True Positives (TP), False-negatives (FN), False-positives (FP), and True negatives (TN) predicted over the CICIDS 2017 and CICIDS 2018 dataset.

Figure 3. Confusion Matrices of the proposed ensemble method on CICIDS 2017



It can be observed that the CICIDS 2018 has higher FP as well as FN, which affect the overall performance. The confusion matrix obtained is used to calculate the performance metrics. The performances of the proposed model are tabulated in Table 4 and Table 5 along with the individual algorithms.

Figure 4. Confusion Matrices of the proposed ensembled method on CICIDS 2018



It is observed from Table 4 and Table 5 that the ensemble technique significantly improves the classification results upon the selected individual machine learning algorithms. It is also evident that addition of one feature i.e. from 7 features to 8 features, does not have a huge impact on the classification results. There is little to no change in the outcome and may not have high positive impact on increasing the number of features. In Table 4, the proposed model has lower computation time than the ID3 while having better results. The overall performance of the ensemble method surpasses the individual algorithms, especially in the FAR. However, in Table 5, the FAR for NB and QDA are lower than FAR of the ensembled method even though their overall performances are not good. The proposed model gives an accuracy of 98.3% and 95.1% with low false alarm rate of 2% and 6.9% for CICIDS 2017 and CICIDS 2018 respectively. The training time of the ensembled algorithm is higher than that of the individual algorithms, but this is very less and negligible in comparison to neural networks.

Table 4. Performance of the proposed ensemble method (CICIDS 2017)

No of features	Algorithms	Accuracy	Precision	Recall	F1-Score	False Alarm Rate	Training Time (Sec)
With 7 Features	**Ensemble Method**	**0.983**	**0.966**	**0.981**	**0.973**	**0.022**	**10.15**
	Naïve Bayes	0.820	0.712	0.681	0.693	0.549	1.50
	QDA	0.847	0.779	0.695	0.721	0.555	2.04
	ID3	0.958	0.971	0.896	0.928	0.206	13.13
With 8 features	**Ensemble Method**	**0.983**	**0.966**	**0.981**	**0.973**	**0.021**	**11.21**
	Naïve Bayes	0.817	0.707	0.681	0.692	0.544	1.57
	QDA	0.842	0.759	0.692	0.716	0.556	2.18
	ID3	0.959	0.968	0.901	0.930	0.193	15.86

The graphical presentation of the performance evaluation of the proposed method and other individual algorithms can be visualized in Figure 5 and Figure 6. The proposed ensembled method has the best result in every aspect in CICIDS 2017 dataset. However, in CICIDS 2018, though the proposed method has the best result in overall classification accuracy, it has higher FAR than that of Naïve Bayes and QDA. This is because the Naïve Bayes and QDA algorithm has very less false positive and very high false negative.

Table 5. Performance of the proposed ensemble method (CICIDS 2018)

No of features	Algorithms	Accuracy	Precision	Recall	F1-Score	False Alarm Rate	Training Time (Sec)
With 7 Features	**Ensemble Method**	**0.951**	**0.946**	**0.946**	**0.946**	**0.069**	**37.2**
	Naïve Bayes	0.683	0.745	0.748	0.683	0.036	3.55
	QDA	0.684	0.746	0.748	0.684	0.037	4.33
	ID3	0.944	0.940	0.937	0.938	0.088	28.4
With 8 features	**Ensemble Method**	**0.951**	**0.946**	**0.946**	**0.946**	**0.069**	**37.87**
	Naïve Bayes	0.687	0.747	0.750	0.687	0.037	3.54
	QDA	0.682	0.745	0.747	0.682	0.037	5.16
	ID3	0.948	0.942	0.943	0.943	0.071	31.08

Figure 5. Performance evaluation of different algorithms on CICIDS 2017

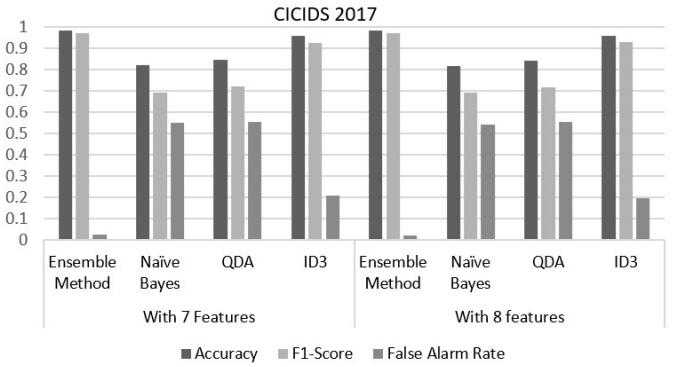
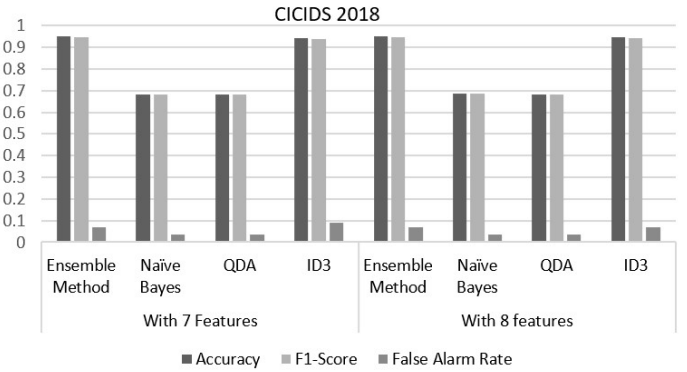
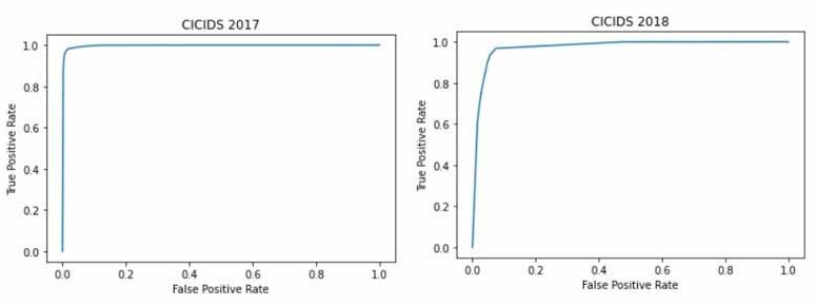


Figure 6. Performance evaluation of different algorithms on CICIDS 2018



The ROC curve plots the TPR and the FPR at different classification threshold. The more the curve bends towards the top left corner, the better is the curve. The ROC curve of the ensemble method on CICIDS 2017 and CICIDS 2018 is depicted in Figure 7.

Figure 7. ROC curve of ensembled method in CICIDS 2017 (a) and CICIDS 2018 (b)



CICIDS 2018 consisted of a new and wider range of protocols and attacks. The proposed model is trained and designed for CICIDS 2017 dataset and CICIDS 2018 dataset is used for validating the model as CICIDS 2017 and CICIDS 2018 dataset have the same source of network traffic. The ability to ensemble multiple machine learning algorithms produces more efficient models even with algorithms with weak performance metrices. The comparison of the proposed work with other related works is shown in Table 6. Some related works with Deep Learning models is not added in this comparative analysis as the computation costs are very high in comparison to traditional Machine Learning Algorithms, and one of the objectives of this paper is to develop a model with efficient computation cost. Table 6 contain related researches of ID model with Machine Learning algorithms on different datasets and their respective outcome.

Table 6. Comparative analysis of the proposed work with other related works.

Authors and Year	Research Aspects/Model	Datasets used	Results
Latifur Khan, Mamoun, and Bhavani (2007)	Clustering Trees based on SVM	DARPA 98	Average Acc=69.8%
S. Mukherjee and N. Sharma (2012)	Feature-Vitality Based Reduction Method (FVBRM) And Naïve Bayes	NSL-KDD	Acc=97.78%
Yassin, Udzir, Muda, Sulaiman (2013)	K-Means and NB	KDDCUP 99	K Means DR=99% NB DR=98.8%
Gautam and Amit (2018)	Ensemble of Naïve Bayes, Adaptive Boost, and PART	KDDCUP 99	Acc=99.97%
Kostas (2018)	Different traditional Machine Learning Algorithms	CICIDS 2017	KNN Acc=97% (highest)
Tama, Comuzzi, & Rhee (2019)	Ensemble of rotation forest and Bagging	NSL-KDD and UNSW-NB15	NSL-KDD Acc=85.8% UNSW-NB15 Acc=91.27%
Beulah and Punithavathani (2020	Clustering-based outlier detection (CBOD)	NSL-KDD	Acc=97.96% FAR=1.88%
Our Paper	*Ensemble of NB, QDA and ID3using Random Forest*	*CICIDS 2017 and CICIDS 2018*	*CICIDS 2017 Acc=98.3% CICIDS 2018 Acc=95.1%*

CONCLUSION

In this study, it was aimed to detect anomalies on a network by creating an ID model using machine learning techniques, addressing the high false alarm rate, computation time and improving the

classification accuracy. For this, an anomaly-based network ID system model was developed using ensemble techniques of Machine Learning. For the implementation, the authors used three different base models, Naïve Bayes, Quadratic Discriminant Analysis (QDA) and Iterative Dichotomiser 3 (ID3) which were combined using another machine learning algorithm; Random Forest Classifier. The training and testing of the model have been carried out using the CICIDS 2017 dataset. Feature selection has been done using Random Forest Regressor; this is to reduce the computation cost and also removal of features with less importance. The same proposed method was implemented on CICIDS 2018 dataset for validation and the results proves to be positive.

The proposed ensemble model outperforms the individual machine learning algorithms in classification accuracy and False Alarm Rate, with little to no increase in computational cost. Individually, Naïve Bayes and QDA algorithm does not have high classification results on the CICIDS datasets, but when these algorithms are blended with ID3, the outcome of the combined algorithms surpass all the individual classification results. Feature selection is done using the weight importance of the features evaluated using Random Forest Regressor. Out of around 80 features, 7 features with highest weights are selected for implementation as their weight consist of around 95% of the total weight. This reduction in features reduces the computational cost to around one tenth of the computational cost with full feature.

The future work may include feature subset selection, which will search the best subset of features, this will result in higher overall classification accuracy, which will also have a significant decrease in computational cost. The detection of particular attack by training the model for only one type of attack at a time will give more detailed information for that particular attack type.

REFERENCES

Beulah, J. R., & Punithavathani, D. S. (2020). An efficient mixed attribute outlier detection method for identifying network intrusions. *International Journal of Information Security and Privacy*, *14*(3), 115–133. doi:10.4018/IJISP.2020070107

Bhuyan, M. H., Bhattacharyya, D. K., & Kalita, J. K. (2013). Network anomaly detection: Methods, systems and tools. *IEEE Communications Surveys and Tutorials*, *16*(1), 303–336. doi:10.1109/SURV.2013.052213.00046

Buczak, A. L., & Guven, E. (2015). A survey of data mining and machine learning methods for cyber security intrusion detection. *IEEE Communications Surveys and Tutorials*, *18*(2), 1153–1176. doi:10.1109/COMST.2015.2494502

CIC-IDS2017 Dataset. (2017). Retrieved from https://www.unb.ca/cic/datasets/ids-2017.html

CIC-IDS2018 Dataset. (2018). Retrieved from https://www.unb.ca/cic/datasets/ids-2018.html

Fayssal, S., Hariri, S., & Al-Nashif, Y. (2007, August). Anomaly-based behavior analysis of wireless network security. In *2007 Fourth Annual International Conference on Mobile and Ubiquitous Systems: Networking & Services (MobiQuitous)* (pp. 1-8). IEEE. doi:10.1109/MOBIQ.2007.4451054

Gautam, R. K. S., & Doegar, E. A. (2018, January). An ensemble approach for intrusion detection system using machine learning algorithms. In *2018 8th International conference on cloud computing, data science & engineering (confluence)* (pp. 14-15). IEEE. doi:10.1109/CONFLUENCE.2018.8442693

Haq, N. F., Onik, A. R., Hridoy, M. A. K., Rafni, M., Shah, F. M., & Farid, D. M. (2015). Application of machine learning approaches in intrusion detection system: A survey. *IJARAI-International Journal of Advanced Research in Artificial Intelligence*, *4*(3), 9–18. doi:10.14569/IJARAI.2015.040302

Javaid, A., Niyaz, Q., Sun, W., & Alam, M. (2016). A deep learning approach for network intrusion detection system. *Eai Endorsed Transactions on Security and Safety*, *3*(9), e2. doi:10.4108/eai.3-12-2015.2262516

Khan, L., Awad, M., & Thuraisingham, B. (2007). A new intrusion detection system using support vector machines and hierarchical clustering. *The VLDB Journal*, *16*(4), 507–521. doi:10.1007/s00778-006-0002-5

Kostas, K. (2018). Anomaly detection in networks using machine learning. *Research Proposal*, *23*, 343.

Khan, M. A., & Kim, Y. (2021). Deep learning-based hybrid intelligent intrusion detection system. *Computers. Materials & Continua*, *68*(1), 671–687. doi:10.32604/cmc.2021.015647

Maleh, Y., Sahid, A., & Belaissaoui, M. (2021). Optimized Machine Learning Techniques for IoT 6LoWPAN Cyber Attacks Detection. In *Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020). SoCPaR 2020. Advances in Intelligent Systems and Computing* (vol. 1383). Springer. doi:10.1007/978-3-030-73689-7_64

Maleh, Y., Ezzati, A., Qasmaoui, Y., & Mbida, M. (2015). A global hybrid intrusion detection system for wireless sensor networks. *Procedia Computer Science*, *52*, 1047–1052. doi:10.1016/j.procs.2015.05.108

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research*, *9*, 381-386.

Modi, C., Patel, D., Borisaniya, B., Patel, H., Patel, A., & Rajarajan, M. (2013). A survey of intrusion detection techniques in cloud. *Journal of Network and Computer Applications*, *36*(1), 42–57. doi:10.1016/j.jnca.2012.05.003

Mohammadpour, L., Ling, T. C., Liew, C. S., & Aryanfar, A. (2020). A Mean Convolutional Layer for Intrusion Detection System. *Security and Communication Networks*, *2020*(2020), 8891185. doi:10.1155/2020/8891185

Mukherjee, S., & Sharma, N. (2012). Intrusion detection using naive Bayes classifier with feature reduction. *Procedia Technology*, *4*, 119–128. doi:10.1016/j.protcy.2012.05.017

Patidar, P., & Khandelwal, H. (2019). Zero-day attack detection using machine learning techniques. *International Journal of Research and Analytical Reviews*, *6*(1), 1364–1367.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research*, *12*, 2825-2830.

Smys, S. (2019). DDOS attack detection in telecommunication network using machine learning. *Journal of Ubiquitous Computing and Communication Technologies*, *1*(01), 33–44. doi:10.36548/jucct.2019.1.004

Sy, B. K. (2005, July). Signature-based approach for intrusion detection. In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 526-536). Springer. doi:10.1007/11510888_52

Tama, B. A., Comuzzi, M., & Rhee, K. H. (2019). TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. *IEEE Access: Practical Innovations, Open Solutions*, *7*, 94497–94507. doi:10.1109/ACCESS.2019.2928048

Thakkar, A., & Lohiya, R. (2020). A review of the advancement in intrusion detection datasets. *Procedia Computer Science*, *167*, 636–645. doi:10.1016/j.procs.2020.03.330

Varzaneh, Z. A., & Rafsanjani, K. (2021, June 29). Intrusion Detection System Using a New Fuzzy Rule-based Classification System Based on Genetic Algorithm. *Intelligent Decision Technologies*, *15*(2), 231–237. doi:10.3233/IDT-200036

Wagh, S. K., Pachghare, V. K., & Kolhe, S. R. (2013). Survey on intrusion detection system using machine learning techniques. *International Journal of Computers and Applications*, *78*(16). Advance online publication. doi:10.5120/13608-1412

Xu, C., Shen, J., Du, X., & Zhang, F. (2018). An intrusion detection system using a deep neural network with gated recurrent units. *IEEE Access: Practical Innovations, Open Solutions*, *6*, 48697–48707. doi:10.1109/ACCESS.2018.2867564

Yang, J., Sheng, Y., & Wang, J. (2020). A GBDT-Paralleled Quadratic Ensemble Learning for Intrusion Detection System. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 175467–175482. doi:10.1109/ACCESS.2020.3026044

Yassin, W., Udzir, N. I., Muda, Z., & Sulaiman, M. N. (2013). Anomaly-based intrusion detection through k-means clustering and naives bayes classification. *7th International Conference on IT in Asia (CITA).* doi:10.1109/CITA.2011.5999520

Zamani, M., & Movahedi, M. (2013). *Machine learning techniques for intrusion detection*. doi:/arXiv.1312.217710.48550

Zhang, X., & Chen, J. (2017, May). Deep learning based intelligent intrusion detection. In *IEEE 9th international conference on communication software and networks (ICCSN)* (pp. 1133-1137). IEEE. doi:10.1109/ICCSN.2017.8230287

Zhao, P., Fan, Z., Cao, Z., & Li, X. (2022). Intrusion Detection Model Using Temporal Convolutional Network Blend Into Attention Mechanism. *International Journal of Information Security and Privacy*, *16*(1), 1–20. doi:10.4018/IJISP.290832

R. Lalduhsaka is currently pursuing PhD in the Department of Computer Engineering in Mizoram University, India. He received his M.Tech degree from NIT Silchar, India in 2019. His current research interests include Machine Learning and Network Security.

Nilutpol Bora is currently pursuing Master Degree in Information System from Delhi Technological University, India. He received his B.Tech Degree in Computer Engineering Department from Mizoram University, India in 2021. His current research interests include Network Security, IoT and Machine Learning.

Ajoy Kumar khan received PhD degree from Assam University, Silchar in 2015. He joined as Assistant Professor in the Department of Computer Science and Engineering, Assam University (Central University under Govt of India) in 2009 and presently he is serving as Associate Professor in the Department of Computer Engineering in Mizoram university (Central University under Govt of India). He has completed 3 Govt. of India sponsored research project in the area of VLSI and Network Security. His current research interests include Network Security and Machine Learning.