Enhanced Intrusion Detection System using Feature Selection Method and Ensemble Learning Algorithms

Manal Abdullah
Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Saudi Arabia
maaabdullah@kau.edu.sa

Arwa Alshannaq
Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Saudi Arabia
aalshannaq@stu.kau.edu.sa

Asmaa Balamash
Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Saudi Arabia
Abalamash0003@stu.kau.edu.sa

Soad Almabdy
Faculty of Computing and Information Technology
King Abdul-Aziz University
Jeddah, Saudi Arabia
salmabdy@kau.edu.sa

Abstract— **The main goal of Intrusion Detection Systems (IDSs) is to detect intrusions. This kind of detection system represents a significant tool in traditional computer based systems for ensuring cyber security. IDS model can be faster and reach more accurate detection rates, by selecting the most related features from the input dataset. Feature selection is an important stage of any IDs to select the optimal subset of features that enhance the process of the training model to become faster and reduce the complexity while preserving or enhancing the performance of the system. In this paper, we proposed a method that based on dividing the input dataset into different subsets according to each attack. Then we performed a feature selection technique using information gain filter for each subset. Then the optimal features set is generated by combining the list of features sets that obtained for each attack. Experimental results that conducted on NSL-KDD dataset shows that the proposed method for feature selection with fewer features, make an improvement to the system accuracy while decreasing the complexity. Moreover, a comparative study is performed to the efficiency of technique for feature selection using different classification methods. To enhance the overall performance, another stage is conducted using Random Forest and PART on voting learning algorithm. The results indicate that the best accuracy is achieved when using the product probability rule.**

Keywords-Intrusion Detection Systems, NSL-KDD, Feature Selection, Supervised Learning, Classification.

I. Introduction

Wireless sensor networks (WSNs) comprise of tiny sensor nodes or devices that have radio, processor, memory; battery as well as sensor hardware. The widespread deployment of these sensor nodes makes it possible for environmental monitoring. These small devices are resource inhibited in terms of the speed of the processor, the range of the radio, memory as well as power. This nature of resource inhibition makes designers design systems that are application specific. While the Wireless Sensor Networks are not protected, and the transmitted medium is wireless, this raises the vulnerability to attacks. WSNs are being gradually embraced also in applications which are very sensitive for instance in detection of forest fires [1], power transmission as well as distribution [2], localization [3], applications of the military [4], Critical-infrastructures (CIs) [5] and Underwater Wireless Sensor Networks (Underwater WSNs) [6].

Lack of proper security measures can lead to launching of different types of attacks in environments that are hostile. These kinds of attacks can interrupt the WSNs from working normally and can defeat the deployment's purpose. Consequently, security is a significant networks feature. The shortage of means makes the creators use primitives of security which are traditional such as encryption and one-way functions cautiously. Detection of intrusion is seen as the defense's second line which matches the security primitives. For practicality in implementing WSNs, intrusions detection ideas need to be lightweight, scalable as well as distributed. This paper proposes such approaches in the detection of anomaly intrusion in WSNs. In this kind of context, it is very important to make sure that there is the protection of the sensor network from threats emanating from cyber−security. Regrettably, the achievement of this objective is a bit of a challenge due to features number of WSNs, highest important one being: inadequate computational resources, inhibiting the execution of robust mechanisms that are cryptographic; and their distribution in environments that are wild and unattended, where it is possible for the enemy to access the sensor nodes physically, for instance, reading cryptographic keys straight from the memory. The fast technology development over the Internet makes the security of a computer serious issue. Currently, Intelligence which is artificial, data mining as well as machine learning algorithms are exposed to a broad investigation in ID with stress on enhancing the detection accuracy as well as create a model that is immune for IDS. In addition to detection abilities, IDSs also offers extra mechanisms, for instance, diagnosis as well as prevention. Wireless sensor networks' IDSs architectures are presently being examined and various solutions have been recommended in the research.

This paper concentrates on building IDs for WSN. To construct an Intrusion Detection System model quicker with more correct rates of detection, choice of features that are vital from the input dataset is extremely important. Learning process's feature selection while designing the model indicates a decrease in computational rate and improves precision. The main objective of this paper is determining the greatest suitable features to use in the identification of attack in a dataset of NSL KDD as well as WEKA [7] tool is used for analysis. Different performance metrics are used to assess the performance of each classifier such as: precision, recall, F-measure, false positive rate, overall accuracy (ACC) and ROC curve. NSL KDD dataset [8] is a common dataset for revealing of the anomaly, particularly for identifying the intrusion. This dataset comprises of forty-one features that resemble different types of the network traffic. The network traffic is divided into dual classes, one being the normal class while the other is referred to as the anomaly class. The anomaly class usually depicts intrusions or attacks that originate from the network at the time of taking records for the network traffic. In relation to these attacks, the NSL KDD dataset is additionally categorized into four main attack classifications such as the DoS, in addition to probing. Further classifications comprise of users to root (U2R), as well as remote to local (R2L). The DoS attack renders the unavailability of crucial services to genuine users through the bombardment of the attack packets that are found on the computing and also on network resources. Instances of DoS attacks contain backland, and smurf. Moreover, teardrop, plus neptune attacks are also examples of such attacks. Due to the high levels of the risks that are found in other types of the DoS attacks that relate to computer expenses, the paper primarily dealt on the DoS attacks, as stated in the 2014 document [9]. A DoS attack is viewed as a major concern for authentic operators retrieving services through the Internet. DoS attacks render the unattainability of services to users through limiting network and also the system resources. While a lot of investigation has been performed by dint of network security professionals to defeat the DoS attack concerns, DoS attacks are still on the rise and have a more significant detrimental influence as time passes.

The organization of the paper as following. Section 2 presents an intrusion detection overview, reviews related work. Section 3 describes IDS proposed model, and Sect. 4 is analysis the experimental results obtained. Finally, Section 5 states the conclusions.

II. Literature Review

Intrusion detection system uses machine learning algorithms or classifiers to learn system normal or abnormal behavior and build models that help to classify new traffic. Developing an optimal machine learning based detection systems directs research to examine the performance of a single machine learning algorithm or multiple algorithms to all four major attack categories rather than to a single attack category. Some of the algorithms and methods used by the researchers in this filed will be mentioned. Also, we will try to focus on the researches that used NSL-KDD for analyzing their experimental results.

Hota and Shrivas, 2014 [10], proposed a model that used different feature selection techniques to remove the irrelevant features in the dataset and developed a classifier that is more robust and effective. The methods that were used combined with classifier are Info Gain, Correlation, Relief and Symmetrical Uncertainty. Their experimental work was divided into two parts: The first one is building multiclass classifier based on various decision tree techniques such as ID3, CART, REP Tree, REP Tree and C4.5. The second one is applying feature selection technique on the best model obtained which was here C4.5. Their experimental analysis was conducted using WEKA tool. The results showed that C4.5 with Info Gain had better results and achieved highest accuracy of 99.68% with only 17 features. However, in case of using11 features, Symmetrical Uncertainty achieved 99.64% accuracy.

Deshmukh, 2014 [11], developed IDS using Naive Bayes classifier with different pre-processing methods. Authors used NSL-KDD dataset and WEKA for their experimental analysis. They compared their results with other classification algorithms such as NB TREE and AD Tree. The results showed that with respect to the TP rate of all algorithms, the execution time of Naïve Bayes is less.

Noureldien Yousif, 2016 [12], examined the performance of seven supervised machine learning algorithms in detecting the DoS attacks using NSL-KDD dataset. The experiments were conducted by using for training step the Train+20 percent file and for testing using Test-21 file. they used 10-fold cross validation in test and evaluate the methods to confirm that techniques will achieve on undetected data. Their results showed that Random Committee was the best algorithm for detecting smurf attack with accuracy of 98.6161%. At the average rate, the PART algorithm was the best for detecting the Dos Attacks, however, Input Mapped algorithm was the worst.

Jabbar and Samreen, 2016 [13], have presented a novel approach for ID using alternating decision trees (ADT) to classify the various types of attacks while it is usually used for binary classification problems. The results showed that their proposed model produced higher detection rate and reduces the false alarm rate in classification of IDS attacks.

Paulauskas and Auskalnis, 2017 [14], analyses the initial data pre-processing influence on attack detection accuracy by using of ensemble, that are depend on the idea of combining multiple weaker learners to create a stronger learner, model of four different classifiers: J48, C5.0, Naïve Bayes and PART. Min-Max normalization as well as Z-Score standardization was applied in pre-processing stage. They compared their proposed model with and without pre-processing techniques using more than one classifier. Their results showed that their proposed classifier ensemble model produces more accurate results. After they presented their results, they were warned not to use only the NSL-KDDTrain+ dataset for both training and testing because even without pre-processing methods, it leads to get 99% of accuracy. Therefore, NSL-KDDTest+ dataset must be used for model assessment. In this case the performance of the real model can be tested to detect a new type of attack.

Wang, 2017 [15], suggested an SVM based intrusion detection technique that considers pre-processing data utilizing converting the usual attributes by the logarithms of the marginal density ratios that exploits the classification information that is included in each feature. This resulting in data that has high quality and concise which in turn achieved a better detection performance in addition to reducing the training time required for the SVM detection model.

Yin, et al., 2017 [16], have explored how to model an IDS based on deep learning approach using recurrent neural networks (RNN-IDS) because of its potential of extracting better representations for the data and create better models. They pre-processed the dataset using Numericalization technique because the input value of RNN-IDS should be a numeric matrix. The results showed that RNN-IDS has great accuracy rate and detection rate with a low false positive rate compared with traditional classification methods.

Feature selection as a vital part of any IDS can assist make the procedure of training the model less multifaceted and faster while preserving or even enhancing the total performance of the system. Shahbaz et al. [17] suggested an efficient algorithm for feature selection by considered the correlation between the behavior class label and a subset of attribute to resolve the problem of dimensionality lessening and to defining good features. The outcomes revealed that the proposed model has considerably minimal training time while preserving accuracy with precision. Additionally, several feature selection methods are tested with varying classifiers regarding the detection rate. The comparison outcomes reveal that J48 classifier accomplishes well with the proposed feature selection method.

Similarly, the study in [18] proposed a new intelligent IDS that works on reduced number of features. First, authors perform feature ranking on the basis of information-gain and correlation. Feature reduction is then done by combining ranks obtained from both information gain and correlation using a novel approach to identify useful and useless features. These reduced features are then fed to a feed forward neural network for training and testing on KDD99 dataset. The method uses pre-processing to eliminate redundant and irrelevant data from the dataset in order to improve resource utilization and reduce time complexity. The performance of the feature reduced system is actually better than system without feature reduction. According to the feature optimization selection problems of the rare attack categories detection the researchers in [19] used the cascaded SVM classifiers to classify the non-rare attack categories and using BN classifiers to classify rare attack categories, combining with cascaded GFR feature selection method (CGFR) The experimental results showed that the CGFR feature selection is effective and accurate in IDS.

Redundant as well as irrelevant characteristics in data have resulted in a constant problem in network traffic classification. To combat this concern, Ambusaidi et al. [20] offered a supervised filter-based feature selection algorithm that methodically picks the ideal feature for categorization. The Flexible Mutual Information Feature Selection (FMIFS) that has been proposed to lessen the redundancy among features. FMIFS is then combined with the Least Square Support Vector Machine based IDS(LSSVM) technique to develop an IDS. The role of the model is appraised by means of three intrusion identification datasets, that is to say, KDD Cup 99, NSL-KDD plus Kyoto 2006+ datasets. The appraisal outcomes revealed that characteristic selection algorithm gives other essential characteristics for LSSVM-IDS to accomplish enhanced accurateness and lessen computational expenses in contrary to the state-of-the-art techniques.

Ikram and Cherukuri,2017 [21], proposed an ID model using Chi-Square attribute selection and multi-class support vector machine (SVM). The main idea behind this model is to construct a multi class SVM which has not been adopted for IDS so far to decrease the training and testing time and increase the individual classification accuracy of the network attacks.

In [22], Khammassi and Krichen have applied a wrapper methods based on a genetic algorithm as a search strategy and logistic regression as a learning algorithm for network IDSs to choice the best subset of features. The proposed approach is based on three stages: a pre-processing phase, a feature selection phase, and a classification stage the experiment will be conducted on the KDD99 dataset and the UNSW-NB15 dataset. The results showed that accuracy of classification equal to 99.90 %, 0.105 % FAR and 99.81% DR with a subset of only 18 features for the KDD99 dataset. Furthermore, the selected subset provides a good DR for DoS category with 99.98%. The obtained results for the UNSW-NB15 provided the lowest FAR with 6.39% and a good classification accuracy compared to the other mentioned approaches with a subset composed of 20 features.

From this inspiration, we are trying to find out which of classification algorithms that we select will give better results after selecting the features that have a strong correlation in the training dataset. In this work, researchers will try to conduct some experiments to differentiate and discover the normal and abnormal behavior.

III. Proposed IDS Methodology

The main goal of the research, is to build a framework of intrusion detection with minimum number of features in the dataset. The previous researches showed that only a subset of these features is related to ID. So, the aim is to reduce the data set dimensionality to build a better classifier in a reasonable time. The proposed approach consists of four main phases: The first phase is to select the related features for each attack using feature selection method. Then combining the different features to obtain the optimal set of features for all attacks. The final set of features is fed to the classification stage. Finally, the model is tested using a test dataset. The framework of the proposed methodology is shown in Fig. 1.

A. Selecting the Related Features for Each Attack

While the network intrusion system deals with a large amount of raw data, the feature selection is becoming a basic step in building such system. Feature selection is related to a number of methods and techniques that are used to eliminate the irrelevant and redundant features. The dimensionality of the data set has a big effect in the model complexity that leads to low classification accuracy, and high computational cost and time. The aim of these methods also is to select the optimal features which will enhance the model's performance. There are two general categories of methods for feature selection, filter methods and wrapper methods [23]. In the Filter algorithms an

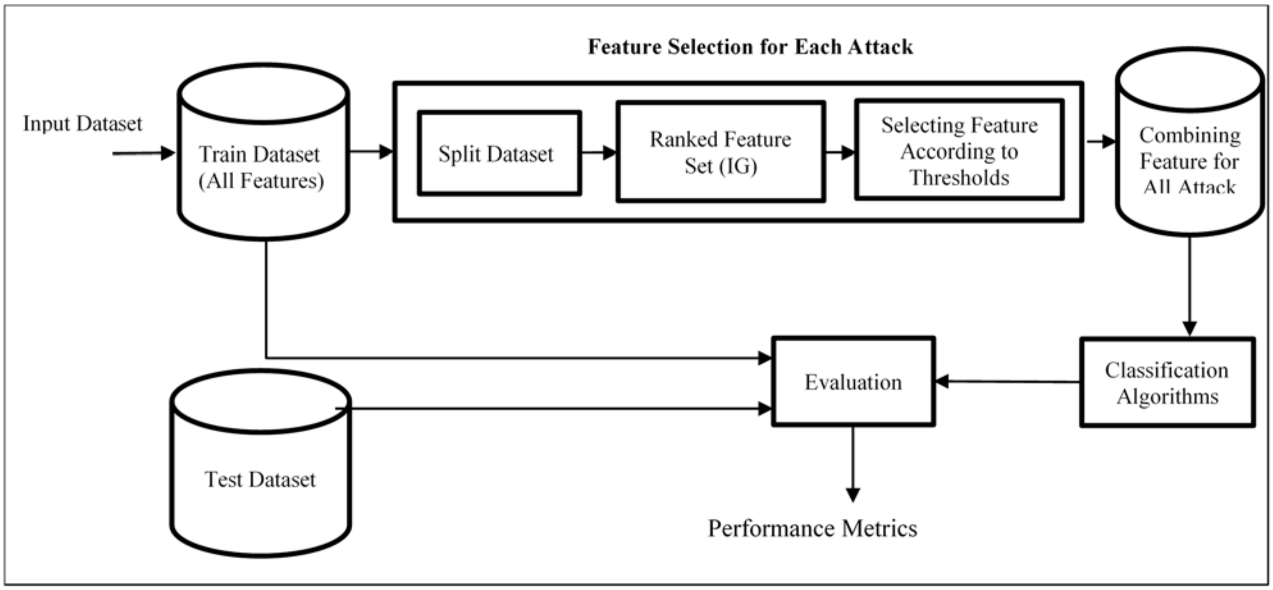
Feature Selection for Each Attack
Input Dataset
Train Dataset (All Features)
Split Dataset
Ranked Feature Set (IG)
Selecting Feature According to Thresholds
Combining Feature for All Attack
Classification Algorithms
Evaluation
Test Dataset
Performance Metrics

Figure 1. *Framework of The Proposed Model of IDS*

independent measure is utilized (such as, information, distance, or consistency) which are used to estimate the relation of a set of features, while wrapper algorithms use of one of learning algorithms to make the evaluation of the feature's value. In this study, Information Gain (IG) will be used to select the subset of related features. IG is often cost less and faster than the wrapper methods.

Information gain is computed for each individual attribute in the training dataset related to one class. If the ranked value is high that means a feature is highly distinctive this class. Otherwise if the value is less than the predetermined threshold, it will be removed from the feature space. To obtain a better threshold value, the distribution of the IG values is examined and tested with different threshold values on the training dataset.

The IG of a feature t, overall classes is known by equation (1).

$$IG(t) = -\sum_{i=1}^{m} p(c_i)\log p(c_i) + p(t)\sum_{i=1}^{m} p(c_i\backslash t)\log p(c_i\backslash t) + p(\bar{t})\sum_{i=1}^{m} p(c_i\backslash \bar{t})\log p(c_i\backslash \bar{t}) \quad (1)$$

Where:

- c_i represents (i) category.
- P(c_i): probability that a random instance document belongs to class c_i.
- P(t) and P($\bar{t}$) probability of the occurrence of the feature w in a randomly selected document.
- P(c_i|t): probability that a randomly selected document belongs to class c_i if document has the feature w.
- m is the number of classes.

The selection features stage for each attack is divided into three main steps as follows:

Step1: The training dataset is divided into 22 datasets. Each dataset file contains the records of one attack records merged with the normal records. If the whole dataset is used without splitting, then the selection features method will be biased to the most frequent attacks. So, this step is essential to obtain more accurate results.

Step2: Each file then is used as an input to IG method to select the most relevant features of that attack. For example, the spy attack has the related features ranked as shown in Table 1.

Step3: A ranked feature list is generated, and according to some thresholds, a number of features are eliminated. From the list in Table I, it can be noticed that the most relevant features for spy attack are features 38 and 39, if we take the threshold equal to 0.003. So, we can take the best two features and eliminate the others.

TABLE I. SPY RANKED RELATED FEATURES

Ranked Value	Feature Number	Feature Name
0.004029	**38**	**dst_host_serror_rate**
0.0036057	**39**	**dst_host_srv_serror_rate**
0.0018171	3	Service
0.0012618	18	num_shells
0.0011184	15	su_attempted
0.0008256	19	num_access_files
0.0001008	2	protocol_type

B. *Combining the Different Set of Features for All Attacks*

In this step, a combined list of features for all attacks is generated from the obtained subsets. For some attacks the highest rank of the first three features are selected. But for another set of attacks, like land attack, one feature has been taken, since it's rank is equal to 1, while the ranks for other features were very low. That means this feature can fully discriminate this attack.

C. *Classification of the Training Dataset*

The final combined subset is used as an input to the classification stage. The results of three different classifiers have been considered to make the comparative study. These classifiers are J48, Random-Forest (RF) and Partial Decision List (PART). After conducting the experiments, the best two classifiers results are chosen. The next step, is to use the vote ensemble method to enhance the performance of the model.

- **J48 classifier:** C4.5 (J48) is an algorithm developed by Ross Quinlan that used to generate a decision tree. This algorithm becomes a popular in classification and Data Mining. The gain ratio method is used in this algorithm as a criterion for splitting the data set. Some normalization techniques are applied to the information gain using a "split information" value.
- **Random Forest:** is related to a machine learning method which makes a combination between decision tree and ensemble methods. The input of the forest that represent the features are picked randomly to build the composed trees. The generation process of the forest constructs a collection of trees with controlled variance. majority voting or weighted voting can be used to decide the resulting prediction.
- **Partial Decision List (PART):** PART is an algorithm of decision–list based on partial decision tree, joining the advantages of both classifier C4.5 and PIPPER. A pruned decision tree is created for all existing instances, for the leaf node building a rule corresponding with the largest coverage, after that discarding the tree and continuing.
- **Ensemble classifier:** An ensemble classifier consists of the combination of multiple weak machine learning algorithms (known as weak learners) to improve the classification performance. The combination of weak

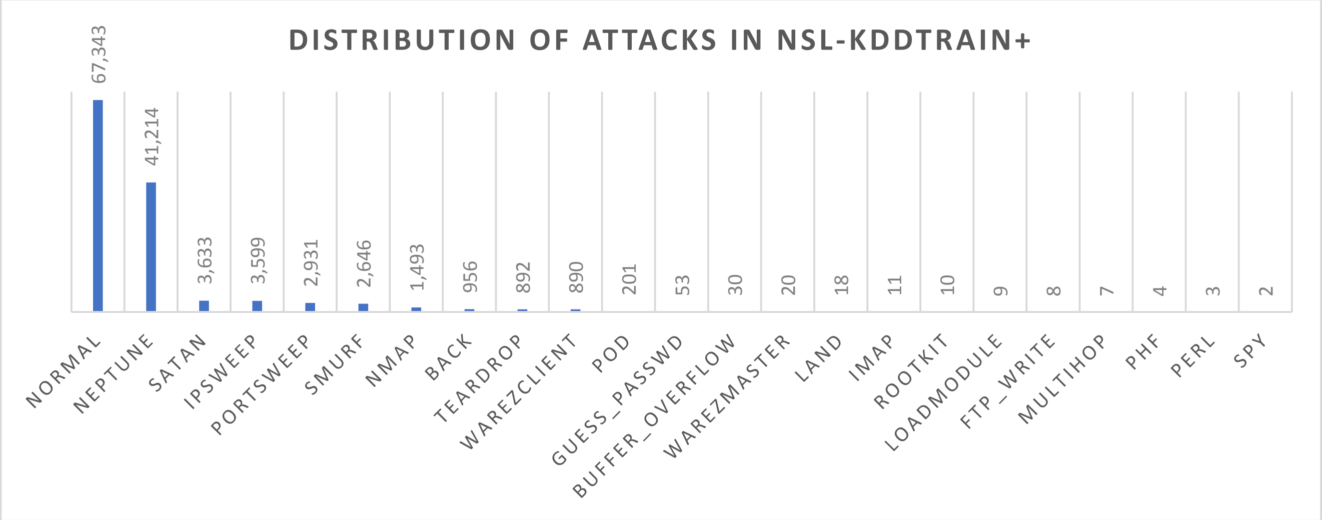

Figure 2. *Distribution of Attacks in NSL-KDDTrain+*

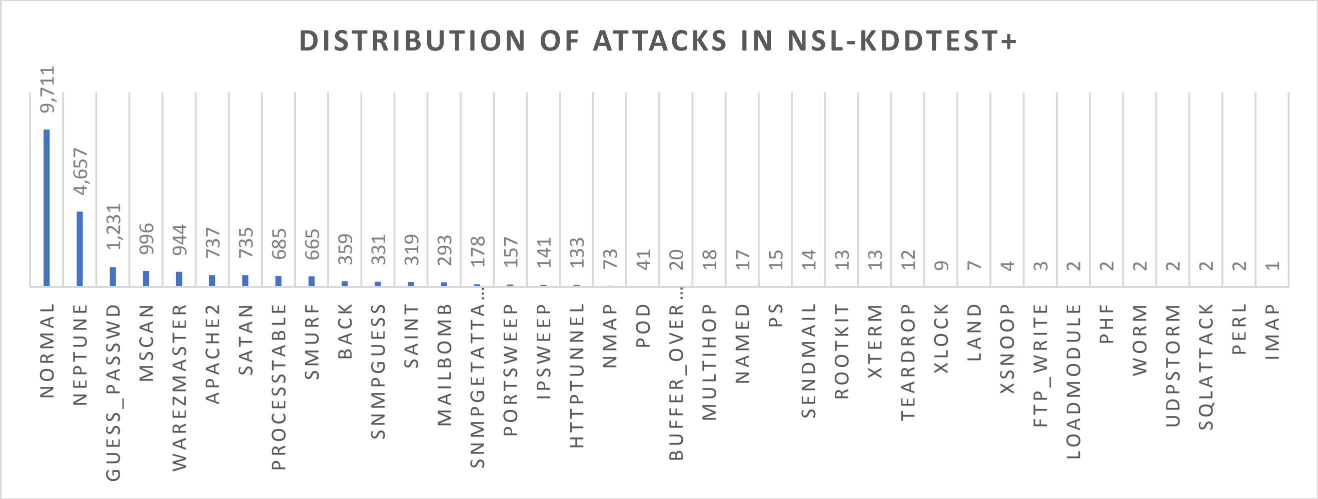

Figure 3. *Distribution of Attacks in NSL-KDDTest+*

learners can be based on different strategies such as majority vote, boosting, or bagging.

D. Testing the Model

In this stage, a test dataset KDD-Test is used to evaluate the model which has been generated by the vote ensemble method. The test dataset file is different from the training dataset and has an extra number of attacks. After that the performance evaluation of the model is conducting using some measures such as accuracy, and area under the ROC.

IV. Results and Analysis

In this section, experiments results analysis is discussed. All experiments were conducted using platform of Windows with configuration of Intel® core™ i7 CPU 2.70 GHZ, 8 GB RAM. WEKA tool was used to evaluate the method and perform feature selection. In order to select the optimal training parameters, a 10-fold cross validation (CV) is performed on the training dataset.

A. Dataset Description

All experiments are carried out on NSL-KDD datasets [8]. NSL-KDD is a refined version of the KDD'99 dataset. It overcomes some inherent problems in the original KDD dataset. Redundant records in the training set have been removed so that the classifiers produce unbiased results. There is no duplicate data in the improved testing set. Therefore, the biased influence on the performance of the learners has been significantly reduced. Each connection in this dataset contains 41 features. Researchers in this work carry out the experiments using the KDDTrain and KDDTest data. The different attacks are listed in Table II. The Distribution of Attacks in NSL-KDDTrain+ and NSL-KDDTest+ files are shown in Fig 2 and Fig 3.

TABLE II. Attacks in NSL_KDD Training Dataset

Attack Type	Attack Name
DOS	Neptune, Smurf, Pod, Teardrop, Land, Back
Probe	Port-sweep, IP-sweep, Nmap, Satan
R2L	Guess-password, Ftp-write, Imap, Phf, Multihop, spy, warezclient, Warezmaster
U2R	Buffer-overflow, Load-module, Perl, Rootkit

B. Evaluation Metrics

The performance evaluation of the proposed model, used different performance metrics such as: precision (equation 2), recall (equation 3), F-measure (equation 4), true negative rate, false positive rate and overall accuracy (ACC) (equation 5) that known as correctly classified instances (CC). In addition, presented Received Operating Characteristics (ROC) of the system. The ROC curve is computed by drawing the relation between true positive rate and false positive rate in y-axis and x-axis, respectively.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F_{measuer} = \frac{2 \times Precision \times Recall}{Recall + Precision} \tag{4}$$

$$Accuracy = \frac{Number\ of\ Correct\ Classified\ Connections}{Number\ of\ Connections} \times 100\% \tag{5}$$

Where:

- TP: related to the true positive.
- FP: related to the false positive.
- FN: related to the false negative.

C. Results Analysis

After making many experiments on the combined list. The optimal number of combined features is equal to 28 features. These features as well as its number in the DS are listed in Table III.

TABLE III. The Final Selected Features

Feature Number	Feature Name
1	duration
2	protocol_type
3	services
4	flag
5	src_bytes
6	dst_bytes
7	land
8	wrong_fragment
9	urgent
10	hot
11	num_failed_logins
13	Num_compromised
14	Root_shell
17	num_file_creations
18	num_shells
19	num_access_files
26	srv_serror_rate
29	same_srv_rate
30	diff_srv_rate
31	srv_diff_host_rate
32	dst_host_count
33	dst_host_srv_count
34	dst_host_same_srv_rate
36	dst_host_same_src_port_ra
37	dst_host_srv_diff_host_rat
38	dst_host_serror_rate
39	dst_host_srv_serror_rate
41	dst_host_srv_rerror_rate

In Table IV, comparing the accuracy and different evaluation metrics with two sets of attributes against using the all dataset with 41 attributes according to PART classifier with two test option cross validation and NSL-KDD Test +. As observed, for the accuracy is shown. The performance of proposed technique compared in terms of using cross validation test and testing dataset. The result shows that high accuracy with (99.7984%) is obtained when using set of 19 feature with cross validation test, while using 28 features, the accuracy is (86.66%) when using NSL-KDD Test + dataset.

On the other hand, the results of the comparison between the performance of three classification algorithms with the proposed method, and both CV and testing are presented in Table V.

As a comparison, we used various popular classifiers algorithms. These classifiers are J48, Random-Forest (RF) and Partial Decision List (PART). The highest testing accuracy with (86.66%) is achieved by PART algorithm, whereas the highest obtained accuracy from CV with (99.78%) by using RF. Fig. 4 shows a comparison of classification algorithms in term of accuracy with test option cross validation and NSL-KDD Test +. According to these results, the best two classifiers (PART and RF) have been chosen to manipulate the voting ensemble algorithm. Table VI demonstrates the performance of using voting learning algorithm for Random Forest and PART to improve the obtained accuracy for the system of intrusion detection. It was noticed that, when Random Forest and PART classifiers are used under different combination methods, the accuracy of the model is enhanced. Table VI shows also that the accuracy in CV is the same while using the three rules. But when the supplied test dataset is being used, a different behavior is noticed for the three rules. The best accuracy is achieved when using the product probability rule. Finally, the area under the ROC curves as shown in Fig. 5 is calculated for each attack classes in the dataset based on cross validation and NSL-KDD Test. The results also show that, the ROC values for DoS and probe attacks are almost the same in the two test options, but the values fluctuate with R2L and U2R attacks.

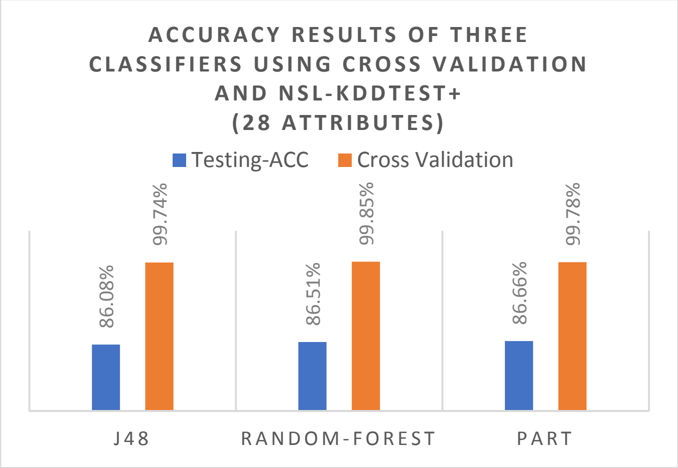

Figure 4. *Accuracy Results of Three Classifiers*

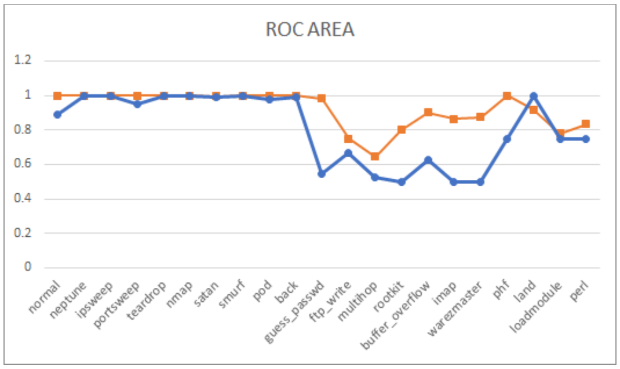

Figure 5. *Final ROC Area for each Class for CV and NSL-KDDTest+*

V. Conclusion and Future Work

IDS is used to secure the computer based systems against a lot of cyber-attacks. Feature selection at the beginning stage of machine learning approach has proven to enhance the detection performance. In the research, we have proposed feature selection approach using information gain methods that was calculated for each attack in the NSL-KDD dataset to identify the optimal feature set for each presented attack and select these features according to some thresholds. Then combining the feature list for all attacks. The experiment result shows that the highest accuracy obtained when using Random Forest and PART classifiers under combination methods namely the product probability rule.

As a future work, it is suggested to use the adaptive boost learning algorithm in the feature selection stage instead of using IG. This will increase the efficiency of the detection system.

TABLE IV. Results with Different Number of Features Using PART

Feature set	Test Option	Correctly Classified	Incorrectly Classified	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
19	Cross Validation	125719	254	99.7984 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16231	2563	86.3627 %	0.864	0.124	0.794	0.864	0.814	0.856
28	Cross Validation	125701	272	99.7841 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16287	2507	86.6606 %	0.867	0.108	0.850	0.867	0.823	0.880
41	Cross Validation	125714	259	99.7944 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16283	2511	86.6394 %	0.866	0.124	0.881	0.866	0.818	0.857

TABLE V. CROSS-VALIDATION AND TEST RESULTS OF THREE CLASSIFIERS

Classifier Name	Test Option	Correctly Classified	Incorrectly Classified	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
J48	Cross Validation	125644	329	99.7388 %	0.997	0.002	0.997	0.997	0.997	0.999
	NSL-KDD Test +	16178	2616	86.0807 %	0.861	0.119	0.774	0.861	0.814	0.840
Random-Forest	Cross Validation	125785	188	99.8508 %	0.999	0.001	0.998	0.999	0.998	1.000
	NSL-KDD Test +	16259	2535	86.5117%	0.865	0.112	0.831	0.865	0.819	0.943
PART	Cross Validation	125701	272	99.7841 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16287	2507	86.6606 %	0.867	0.108	0.850	0.867	0.823	0.880

TABLE VI. CROSS-VALIDATION AND TEST RESULTS USING VOTE METHOD WITH (RF+PART)

Combination Rule	Test Option	Correctly Classified	Incorrectly Classified	Accuracy	TP	FP	Precision	Recall	F-Measure	ROC Area
Majority Voting	Cross Validation	125743	230	99.8174 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16292	2502	86.6872 %	0.867	0.108	0.850	0.867	0.823	0.847
Product probability	Cross Validation	125737	225	99.8127 %	0.998	0.001	0.998	0.998	0.998	0.999
	NSL-KDD Test +	16294	2496	86.6979 %	0.867	0.108	0.851	0.867	0.823	0.884
Average probability	Cross Validation	125743	230	99.8174 %	0.998	0.001	0.998	0.998	0.998	1.000
	NSL-KDD Test +	16292	2502	86.6872 %	0.867	0.108	0.850	0.867	0.823	0.947

REFERENCES

[1] P. D´ıaz-Ram´ırez, A., Tafoya, L.A., Atempa, J.A., Mej´ıa-Alvarez, “Wireless sensor networks and fusion information methods for forest fire detection,” *Procedia Technol. 3*, pp. 69–79, 2012.

[2] A. Isaac, S., Hancke, G., Madhoo, H., Khatri, “A survey of wireless sensor network applications from a power utility’s distribution perspective,” *AFRICON 2001*, pp. 1–5, 2011.

[3] B. . Mao, G., Fidan, B., Anderson, “Wireless sensor network localization techniques. Computer Networks,” vol. 10, no. 51, pp. 2529–2553, 2007.

[4] V. Durisic, M., Tafa, Z., Dimic, G., Milutinovic, “A survey of military applications of wireless sensor networks,” in *2012 Mediterranean Conference on Embedded Com- puting, MECO*, 2012, pp. 196–199.

[5] L. Afzaal, M., Di Sarno, C., Coppolino, L., D’Antonio, S., Romano, “A resilient architecture for forensic storage of events in critical infrastructures.,” in *2012 IEEE 14th International Symposium on High-Assurance Systems Engineering, HASE*, 2012, pp. 48–55.

[6] D. Wahid, A., Kim, “Connectivity-based routing protocol for underwater wireless sensor networks,” in *2012 International Conference on ICT Convergence, ICTC*, 2012, pp. 589–590.

[7] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[8] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in *IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA 2009.*, 2009, pp. 1–6.

[9] P. Institute, “2014 Global report on the cost of cyber crime,” 2014.

[10] H. S. Hota and A. K. Shrivas, “Decision Tree Techniques Applied on NSL-KDD Data and Its Comparison with Various Feature Selection Techniques,” in *Advanced Computing, Networking and Informatics-Volume 1: Advanced Computing and Informatics Proceedings of the Second International Conference on Advanced Computing, Networking and Informatics (ICACNI-2014)*, 2014, pp. 205–211.

[11] D. H. Deshmukh, T. Ghorpade, and P. Padiya, “Intrusion detection system by improved preprocessing methods and Na #x00EF;ve Bayes classifier using NSL-KDD 99 Dataset,” in *2014 International Conference on Electronics and Communication Systems (ICECS)*, 2014, pp. 1–7.

[12] I. M. Y. Noureldien A. Noureldien, “Accuracy of Machine Learning Algorithms in Detecting DoS Attacks Types,” *Sci. Technol.*, vol. 6, no. 4, pp. 89–92, 2016.

[13] M. A. Jabbar and S. Samreen, “Intelligent network intrusion detection using alternating decision trees,” in *2016 International Conference on Circuits, Controls, Communications and Computing (I4C)*, 2016, pp. 1–6.

[14] N. Paulauskas and J. Auskalnis, “Analysis of data pre-processing influence on intrusion detection using NSL-KDD dataset,” in *2017 Open Conference of Electrical, Electronic and Information Sciences (eStream)*, 2017, pp. 1–5.

[15] H. Wang, J. Gu, and S. Wang, “An effective intrusion detection framework based on SVM with feature augmentation,” *Knowledge-Based Syst.*, vol. 136, no. Supplement C, pp. 130–139, 2017.

[16] C. Yin, Y. Zhu, J. Fei, and X. He, “A Deep Learning Approach for Intrusion Detection Using Recurrent Neural Networks,” *IEEE Access*, vol. 5, pp. 21954–21961, 2017.

[17] M. B. Shahbaz, Xianbin Wang, A. Behnad, and J. Samarabandu, “On efficiency enhancement of the correlation-based feature selection for intrusion detection systems,” *2016 IEEE 7th Annu. Inf. Technol. Electron. Mob. Commun. Conf.*, pp. 1–7, 2016.

[18] Akashdeep, I. Manzoor, and N. Kumar, “A feature reduced intrusion detection system using ANN classifier,” *Expert Syst. Appl.*, vol. 88, pp. 249–257, 2017.

[19] Y. Sun and F. Liu, “A & ascaded) eature 6 election $ pproach in 1 etwork , ntrusion â€™ etection,” pp. 119–124, 2015.

[20] M. Ambusaidi, X. He, P. Nanda, and Z. Tan, “Building an intrusion detection system using a filter-based feature selection algorithm,” *IEEE Trans. Comput.*, vol. PP, no. 99, p. 1, 2016.

[21] I. Sumaiya Thaseen and C. Aswani Kumar, “Intrusion detection model using fusion of chi-square feature selection and multi class SVM,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 4, pp. 462–472, 2017.

[22] C. Khammassi and S. Krichen, “A GA-LR wrapper approach for feature selection in network intrusion detection,” *Comput. Secur.*, vol. 70, pp. 255–277, 2017.

[23] F. Amiri, M. R. Yousefi, C. Lucas, A. Shakery, and N. Yazdani, “Mutual information-based feature selection for intrusion detection systems,” *J. Netw. Comput. Appl.*, vol. 34, no. 4, pp. 1184–1199, 2011.

AUTHORS PROFILE

Authors Profile …