Research Journal of Applied Sciences 11 (10): 910-920, 2016
ISSN: 1815-932X

An Evaluation Study on Performance Enhancement of Intrusion Detection Systems

[1]Adil M. Salman and [2]Safaa O. Al-mamory
[1]College of Info Technology, University of Babylon, Hillah, Iraq
[2]College of Business Informatics, University of Information Technology and Communications, Hillah, Iraq

Abstract: Intrusion Detection Systems (IDSs) rely on feature selection algorithms when selecting the most important features; this has an effect on both the accuracy and the time it takes to classify data. Several of these algorithms deal with a number of classes to classify the data. In this study we will evaluate several methods relating to feature selection which utilise adifferent number of classes of the classification in order to determine the optimal number of classes that deliver the best results basedon two criteria the overall accuracy and the time it takes to completethe classification.We utilised WEKA 3.8.0 software for datamining as well as to analyse two types of datasets which are KDD-CUP and NSL-KDD the datasets are each divided into three types based on (23, 5 and 2) classes. The reason behind choosing these numbers of classes is due to the fact that these datasets are available to the researchers on the internetat no cost.It was observed that through minimising the number of classes in classification algorithms, theresults arehighly accurate while training requires only ashort period of time; moreover, there are fewer selected features therefore the processing time is shorter.

Key words: Intrusion detection system, feature selection, network security, network flow, features

INTRODUCTION

Due to the fact that the internet is frequently usedwhile there is ahuge amount of data that is regularly transferred between the servers and clients this data could be lost and become a victim of various types of attacks these attacks are carried out by various intrusions and are steadily increasing. IDS is employed so as to prevent the intruders from accessing the computer systems or networks. IDSs classify the collected network dataset into various kinds such as normal and anomaly. Researchers have proposed numerous classification algorithms in order to assist in the design of an effective IDS. IDS have been classified into three types as follows (Patcha and Park, 2007; Bhuyan *et al*., 2014):

- Misuse-based detection system which utilises a rangeof signatures of known attacks and then identifies the patterns of the malicious traffic; however it cannot recognise the unknown attacks
- Anomaly-Based detection system works by collecting information about the network flow and constructing the flow profile before comparing activities against a (normal) baseline
- Hybrid detection system combines the techniques of the two approaches outlined above

Due to the significant size of the collected original datasets, feature selection methods are a common solution for this issue.These methods succeed in reducing the number of features by removing any redundant and irrelevant data; this technique results in increasing the speed at which the data is reprocessed (Chen *et al*., 2006).

Feature selection and classification algorithms are used alongside one another so as to provide the IDS withthe results, classifying the data to a number of classes depending on the type of data and the purpose that it is to be used for. This study will attempt to answer the question which relates to the number of classes needed in order to obtain aneffectiveoutcome?). Therefore, the aim of this study is to represent an overview of the most effectivenumber of classes that are utilisedin the classification algorithms whichprovidesa good result. We have conducted an experiment of (144) tests employingWEKA 3.8.0 (Waikato Environment for Knowledge Analysis) software for datamining by utlising twelve combinations of feature selection algorithms using the KDD-CUP and NSL-KDD datasets these datasets are

Corresponding Author: Adil M. Salman, College of Info Technology, University of Babylon, Hillah, Iraq

preprocessed to be classified into three types (23, 5 and 2) of classes before applying ten different classification algorithms on each test we can then apply (1440) tests. After conducting this study, it is evident that minimising the number of classes in classification algorithms provides more accurate results while training requires a shorter period of time as well as a fewer number of selected features which takes a less amount of time to process.

Literature review: Chen *et al.* (2006) surveyedseveral of the existing feature selection methods which are utilised in intrusion detection systems. These methods have been categorised into three broad categories: filter, wrapper and hybrid. The conclusion of this study is the identification of the trends and challenges associated withfeature selection research and the development of intrusion detection systems.

Garg and Khurana (2014a) utilised various ranking techniques so as to reduce the number of features as well as to determine the most important selected features by employing classification algorithms. The Boolean and operator were used to create an combination of the six reduced feture sets. The overall performance was analysed by using ten classification algorithms.The final result consisted of a combination of symmetric and Gain ratio while considering the top fifteen attributes which resulted in the most effective performance.

Garg and Khurana (2014b) presented the comparative performance of various classification algorithms using the NSL-KDD dataset. They employed WEKA software to evaluate these classifiers based on 41 attributes. Afterwards, they applied Garrett's ranking techniquein order to rank the different classifiers according to their performance. The rotation forest classification approach performed more effectively in comparison to the other methods.

Bjerkestrand *et al.* (2015) carried out an evaluation and comparison of various algorithms for feature selection and reduction by utilisingdatasets that were publicly available. They used three feature selection algorithms consisting of an attribute evaluator and a test method. The initial results suggestedthat the performance of the classifier was unaffected by reducing the number of attributes.

Oyebode *et al.* (2011) examined the accuracy of using data mining techniques in intrusion detection systems while investigating the accuracy of three classification techniques Naïve Bayes, Radial basis and rotation forest methodswere adopted to detect network access patterns using a KDD 1999 dataset. The models were tested and their classification accuracy was determined using detection accuracy as well as afalse positive rate so as to evaluate metrics.

Amrita and Ahmed (2012) proposed various feature selection methods which are outlined in this literature. There are three categories of approaches for selecting important features which are filter, wrapper and hybrid. The objective of this study is to present a survey of various feature selection methods for IDSutilising a KDD CUP'99 dataset based on these three categories and different evaluation criteria.

Araar and Bouslama (2014) presented a feature selectionusing a random forest technique while employing a KDD'99 dataset during which twenty important features wereselected. The performance and overall accuracy of the feature selection when utilising six representative classification techniques are compared; these techniques include decision trees, BayesNet, NaïveBayes, Rules, SVM and perceptron multi-layer network.The results revealed that J48 is the most effectiveclassifier model for IDS with a reduced number of features.

Singh and Kumar (2015) represented a review of the three kinds of feature selection techniques, includingfilter, weapper and hybrid methods. They reviewed a number of feature selection methods for IDS, utilising a KDD CUP'99 dataset with various evaluation criteria.

Kumar (2016) evaluated the performance of data mining classification algorithms, namely C4.5, J48, Nave Bayes, NB-Tree and Random Forest using a NSL KDD dataset while assessing the Correlation Feature Selection (CFS). The results demonstrate that NB-Tree and Random Forest outperforms the other two algorithms in terms of predictive accuracy and detection rate.

MATERIALS AND METHODS

Intrusion Detection Systems (IDS): An IDS is a software or hardware tool utilisedto detect unauthorised access to a computer system or network. It must be capable of detecting anyabnormal network traffic and computer usage. It collects data bymonitoring the network traffic. The collected network datais analysed to determine if there are any existing rule violations; when any type of violation is found, the IDS will raisean alert (Patcha and Park, 2007). These techniques were categorised into three types Misuse etection, anomaly detection and hybrid detection (Bhuyan *et al.*, 2014). The intrusion can be defined as a set of actions aimed to compromise the purpose of computer security such as confidentiality, integrity and availability (Heady *et al.*, 1990).

Datasets for intrusion detection: Datasets play asignificant role in the testing and validation of

any intrusion detection method. In order to evaluate the IDSs performance, it is important for the system to correctly identifythe attack as well as any normal data. There are several datasets which are available for the public to use in orderto test and evaluate intrusion detection methods. In this study, the following datasets will be utilised (Bhuyan *et al*., 2014; Gogoi *et al*., 2012; Tavallaee *et al*., 2009).

KDD-CUP'99 dataset: This data set was employedfor The third international knowledge discovery and data mining tools competition which was held in conjunction with KDD-99, The fifth international conference on knowledge discovery and data mining. Normal connections have a profile which is expected in a military network while attacks fall into one of four categories user to root; remote to local; Denial of service; and probe.This dataset was prepared by Stolfo *et al*. (2000), it is built on the data captured in the DARPA98 IDS evaluation programme. The KDD training dataset consists of approximately 4,900,000 single connection vectors each of which contains 41 features and is labeled as either normal or attack with a specific attack type. All of the detailsabout these extracted features are in Stolfo *et al*. (2000).

NSL-KDD dataset: This data set was created based on the original dataset KDD-CUP which consists ofa large number of redundant records;these can result in learning algorithms to be biased towards frequent records. In order to address this particular issue, one unique copy of each record is kept in a new dataset known as the NSL-KDD dataset.

Feature selection methods: These methods were proposed so as to solve the problem of dealing with the largecollection of raw data withinthe network flow. It was not a good idea to analyse the data that was initially collecteddue to its large size and the redundant and repeated data; therefore, it has to be optimised. An effective solution to this issue is adopting of one of thefeature selection methods. These methods were an important step in classification. Deciding upon the right set of features is both difficult and time consuming. Feature selection is also regarded as being a preprocessing step in machine learning to select a subset of relevant features for building robust learning models. It is important to determine an optimal set of features that accurately represents the characteristics of the traffic that is being evaluated. In terms of feature selection, numerousresearches have proposed identifying important intrusion features through filter, wrapper and hybrid approaches. The Filter method utilisesthe underlying characteristics of the training data in order to evaluate the relevance of the features or feature set by variousindependent measures such as (distance, correlation and consistency) measures. The wrapper method involves adopting a machine learning algorithm so as to evaluate the goodness of features or feature set. A hybrid method combines the wrapper and filter approaches (Chen *et al*., 2006; Garg and Kumar, 2014a, b; Lee and Stolfo, 1998; Sheen and Rajesh, 2008; Araujo *et al*., 2010).

RESULTS AND DISCUSSION

The experiment: We have conductedan experiment using WEKA 3.8.0 software for datamining by carrying out (144) tests of 12 algorithms for feature selection using two types of datasets, the KDD-CUP 99 and NSL-KDD datasets which are available online for free. Table 1 reveals the distribution of the records for (23, 5 and 2) classes, respectively.We will conduct a comparison between these datasets based on the number of classes employedto classify these datasetsin order to determine the ideal number of classesto be used by network's intrusion detection systems.

Feature selection and classification algorithms in WEKA: There are various algorithms designed for feature selection in WEKA software which allow us to select several of the most important features; these algorithms consist of two main types, attribute evaluators and search methods which are presented below:

Attribute evaluators: This is utilised for ranking all of the features according to various metric. There are several kinds of attribute evaluators available in WEKA we will employ two evaluators in our experiment which are (CfsSubsetEval, ConsistencySubsetEval) for more information about these evaluators, please refer to.

Search methods: These methods search the set of all of the possible features in order to determine the most effective set of features. Seven search methods will be used which are (BestFirst, Evolutionary Search, Genetic Search, Greedy Stepwise, PSO Search, Rank Search and Multiobjective Evolutionary Search) for more information, please refer to (Moraglio *et al*., 2008).

Classification algorithms: These algorithms are utilised in order to classify the datasets in a number of classes. We employed ten classifiers withinour experiment which are as follows. Trees category: (J48, Hoeffding Tree, random forest and random tree).

Table 1:Distribution of records for 23, 5 and 2 classes for two types of kdd 99 datasets

Specific 2 classes types	Specific 5 classes types	Specific 23 classes types	KDD-CUP'99			NSL-KDD'99		
			Train set	Test set	10%	Train set	Test set	20%
Normal	Normal	Normal	972781	60593	97278	67343	9711	13449
Anomaly	Dos	Back	2203	1098	2203	956	359	196
		Land	21	9	21	18	7	1
		Neptune	1072017	58001	107201	41214	4657	8282
		Pod	264	87	264	201	41	38
		Smurf	2807886	164091	280790	2646	665	529
		Teardrop	979	12	979	892	12	188
		Total dos instances	3883370	223298	391458	45927	5741	9234
	probe	Ipsweep	12481	306	1247	3599	141	710
		Nmap	2316	84	231	1493	73	301
		Portsweep	10413	354	1040	2931	157	587
		Satan	15892	1633	1589	3633	735	691
		Total probe instances	41102	2377	4107	11656	1106	2289
	r2l	Ftp_write	8	3	8	8	3	1
		Guess_passwd	53	4367	53	53	1231	10
		Imap	12	1	12	11	1	5
		Multihop	7	18	7	7	18	2
		Phf	4	2	4	4	2	2
		Spy	2	0	2	2	0	1
		Warezclient	1020	0	1020	890	0	181
		Warezmaster	20	1602	20	20	944	7
		Total r2l instances	1126	5993	1126	995	2199	209
	u2r	Buffer_overflow	30	22	30	30	20	6
		Loadmodule	9	2	9	9	2	1
		Perl	3	2	3	3	2	0
		Rootkit	10	13	10	10	13	4
		Total U2R instances	52	39	52	52	37	11
Total anomaly instances			3925650	231707	396743	58630	9083	11743
Total instances	4898431		292300	494021	125973	18794	25192	

- Rules category: (PART and OneR)
- Bayes category: (Naive bayes)
- Meta category: (Attribute selected classifier, random committee and random sub space)

For more information in regards to these classifiers, please refer to.

Implementation of the experiment: We have attempted various combinations of feature selection utilisingthe two datasets that were mentioned above; for the KDD-CUP 99 dataset we utilisedthe Test set as well as the 10% set, although we did not use the train set due to its large size and the fact that it takes a long time to provide only one result. For the NSL-KDD, we utilisedthe train set along with the 20% set. The computer that was used to implement the experiment contains8GB of RAM, Core i7/2.60GHz CPU and a SSD 256GB Hard disk.

Table 2 illustrates the result of the 48 tests showing the twelve combinations of the algorithms along with their selected features using the datasets that classified the features into 23 classes as was previously mentioned in Table 1.

Table 3 presents the results of the 48 tests that were conducted on the twelve combinations of the algorithms and their selected features by employing the datasets that classified the features into five classes as was previously mentioned in Table 1. Table 4 reveals the results of the 48 tests of the twelve combinations of the algorithms along with their selected features using the datasets that classified the features into two classes as was previously mentioned in.

After creating 144 files of the reduced data by utilisingthe feature selection algorithms which are presented above, we appliedten classification algorithms to each file using five cross-validations which are mentioned above at 6.1.3 which means that 1440 trials have been completed.

Table 5 reveals one example of these trialsfor the data generated by using the classification algorithm. The main performance metrics that we employed to build the model are Accuracy and Time as are presented below:

Accuracy: This can be defined as the percentage of the correct predictions. On the basis of the confusion matrix, it is calculated by using Equation:

$$\text{Accuracy} = TP+TN/n$$

where, n is a total number of instances.

Training time: This is the time taken by a classifier to build a model of the dataset. It ismeasured in seconds. For each result table, we calculatedthe average of both the

Table 2: Most important selected features using KDD datasets with 23 classes

Feature selection algorithms		Most important features selected			
		KDD-CUP 99		NSL-KDD	
Attribute evaluators	Search methods	Test set	10%	Train set	20%
Best first	Cfs subset eval	13 features selected* (2,3,4,5,6,7,8,14 21,24,3,0,33,36,)	11 features selected* (2,3,4,5,6,7,8,14,23,30,36)	19 features selected* (2,3,4,5,6,7,8,10, 12,23,25,29,30,35, 36,37,38,39,40)	17 features selected* (2,3,4,5,6,8,10,12,23,25, 29,30,35,36,37,38,40)
	Consistency subseteval	9 features selected* (1,3,5,6,23,24 34,35,40)	(1,3,5,6,12,33,35,36,37,38)	(1,3,5,6,23,32,33,35 36,37,38,39,40)	(1,3,4,6,23,24,33, 35,36,37,38)
Genetic sear ch	Cfs subset eval	13 features selected* (2,3,5,6,7,8,14,21,24 29,30,33,36)	17 features selected* (2,3,4,5,6,7,8,10,12,19, 23,29,30,31,33,36,38)	21 features selected* (2,3,4,5,6,7,8,10,11,12 15,23,,25,26,29,30,34 ,35,36,37,38)	21 features selected* (1,2,3,4,5,6,8,10,14,19 ,23,25,26,27,29 ,30,32, 35,36,37,38)
	Consistency subset eval	17 features selected* (3,5,6,7,10,15,23,24, 27,28,30,33,34,35,36 40,41)	22 features selected* (2,4,5,6,8,9,10,12,13 15,20,23,24,25,26,29, 32,33,35,37,39,41)	24 features selected* 3,4,5,6,7,9,10,11,12 13,15,17,20,23,24,26 27,31,32,33,35,36,37,41	19 features selected* (5,6,9,10,13,14,15,20,, 21 23,24,25,32,33,35, 37,38,40,41)
Rank search	Cfs subset eval	10 features selected* (1,3,5,6,23,24,33,34 35,40)	11 features selected* (1,3,5,6,12,23,33,35 36,37,38)	14 features selected* (1,3,5,6,12,23,32,33, 35 36,37,38,38,40)	11 features selected* (1,3,5,6,23,24,33,35, 36,37,38)
	Consistency subseteval	28 features selected* (1,2,3,4,5,6,7,8,10,11 12,13,18,21,22,24,25 27,28,29,30,33,34,35 36, 37,40,41)	28 features selected* (2,3,4,5,6,7,8,9,10,11 12,13,14,22 ,23,24,25 26,29,30,32,33,34,35 36,37,38,39)	21 features selected* (2,3,4,5,6,7,8,10,11,12 13,25,26 27,29,30,35,36 37,38,39)	22 features selected* (2,3,4,5,6,7,8,10,11 12,13,25,26 27,29 30,34,35,36
PSO search	Cfs subseteval	35 features selected* (1,2,3,4,5,6,7,8,10,11 12 13,14 16,18,21,22 24,25 26,27,28 29,30 31,32,33, 34,35,36,37 38,39,40,41)	26 features selected* (2,3,4,5,6,7,8,9,10,11 12,13,14,22,23,25,26 29 30,33,34,35,36,37 38,39)	32 features selected* (2,3,4,5,6,7,8,10,11,12 13,14,18,22,23,25,26,27 28,29,30,31,32,33,34,35 36,37,38,39,40,41)	34 features selected* (1,2,3,4,5,6,7,8,10,11,12 13,14,17 22,23,24,25,26 27,28,29,30,31,32,33,34 35,36,37,38,39,40,41)
	Consistency subseteval	13 Features selected* (2,3,4,5,6,7,8,14,21,23 29,34,36	19 Features selected* (2,3,4,5,6,7,8,9,10,12,23 24,25,29,30,32,34,35,36)	18 Features selected* (2,4,5,6,7,8,10,11,12,13 23,25,26,29,30,34,35,36 37,38	16 Features selected* (2,4,5,6,7,8,11,12,14 23, 26,29,30,35,36 37,38)
Evolutionary search	Cfs subset eval	15 features selected* (3,5,6,7,14,16,17,22 23,24,33,34,35,38,40)	21 features selected* (2,4,5,6,9,12,13,14,16,17 18,20,23,24,26,33,35 37,38 40,41)	20 features selected* (2,3,4,5,6,7,11,12,,13,14 18,22,23,32,33,35,37 ,38 39,40)	17 features selected* (2,3,5,6,7,13,14,20,2324 32,33,35,37,38 ,39,40)
	Consistency subset eval	18 features selected* (1,2,3,5,8,12,13,24,27 28,29 31,33,34,35,36 37,39)	17 features selected* (2,3,5,6,7,8,10,23,24,28,29 33,35,36,37,38,39)	21 features selected* (2,3,4,5,6,7,8,10,12,21,25 26,29,30,32,33,34,35 37 38,40)	24 features selected* (2,3,4,5,6,8,12,13,17,19 22,23,24,25,26,29,30,32 33,35,37,38,40,41)
Multi objective evolutionary search	Cfs subs eteva 1	18 features selected* (1,3,5,6,8,12,17,19,22 23,24,28,31,33,34,35 38,40)	18 features selected* (2,3,5,6,7,8,10,23,24 ,28 29,33,35,36,37,38,39)	23features selected* (1,3,5,6,7,10,12,13,19 ,22 24,26,27,29,30,31 32 33 35,36,37,38,39	18 features selected* (1,3,5,8,12,17,19,22,23 24,26,30,32,33,34,35,37 40)
Greedy stepwise	Consistency subseteval	15 features selected* (2,3,5,6,7,8,12,18,21 23,29,30,35,36,40)	12 features selected* (2,3,5,6,7,8,14,23,29,36 38,40)	22 features selected* (2,3,4,5,6,7,8,12,13,21,23 25,26,29,30,31,34,35,36,37	16 features selected* (2,3,4,5,8,10,12,25,29 30,31,33,35,36,37,39,39,41)

Table 3: Most important selected features using KDD datasets with 5 classes

Feature selection algorithms		Most important features selected			
		KDD-CUP 99		NSL-KDD	
Attribute evaluators	Search methods	Test set	10%	Train set	20%
Best first	Cfs subset eval	6 features selected* (1,5,6,12,23,32)	8 features selected* (3,6,12,14,23,31,32,37)	11 features selected* (3,4,5,6,12,14,26,29,30 37,38)	11 features selected* (3,4,5,6,12,14,26,29 30,37,38)
	Consistency subset eval	10 features selected* (1,3,5,6,23,24,30,34 35,40)	8 features selected* (3,5,6,12,23,33,35,40)	13 features selected* (1,3,5,6,12,23,25,32,33 35,37,39,40)	9 features selected* (1,3,5,6,23,32,33,35, 38)

Table 3: Continue

Feature selection algorithms		Most important features selected			
		KDD-CUP 99		NSL-KDD	
Attribute evaluators	Search methods	Test set	10%	Train set	20%
Genetic search	Cfs subset eval	11 features selected* (1,2,6,12,19,23,31,32, 34,37,39)	10 features selected* (2,5,6,11,12,22,23,30 31,37)	15 features selected* (2,3,4,5,6,12,14,23,26 29 30,35,37,38,39)	23 features selected* (3,4,5,6,7,8,9,12,15 22 ,25 26,29,30,31 32,33,34,35,36,38 39 41)
	Consistency subset eval	18 features selected* (1,3,5,6,7,8,12,13,15 23,24,25,29,33,34, 35,36)	16 features selected* (4,5,6,12,15,17,20,23, 27 ,29,31,32,35,37 38,41)	24 features selected* (2,3,5,6,7,8,11,12,15,17 20,21,22,23,25,28,29 32,33,34,35,37,39,40)	17 features selected* (2,5,6,9,10,12,21,23 24,25,26 29,32,33 35,38,41)
Rank search	Cfs subset eval	11 features selected* (1,3,5,6,23,24,30,33 34,35,40)	9 features selected* (3,5,6,12,13,23,33 35,40)	13 features selected* (1,3,5,6,12,23,25,32,33 35,37,39,40)	9 features selected* (1,3,5,6,23,32,33 35,38)
	Consistency subset eval	15 features selected* (1,2,3,5,6,9,11,12,14 16 18 22,31,32,37)	11 features selected* (3,5,6,9,11,12,14,22,31 32,37)	13 features selected* (3,4,5,6,12,14,25,26,29 30 ,37,38,39)	14 features selected* (3,4,5,6,9,11,12,25 26 29,30,37,38,39)
PSO search	Cfs subset eval	25 features selected* (1,2,3,5,6,9,10,11,12 14,16,17,18,19,21,22 23,24,30,31,32,33,35 36,37)	24 features selected* (1,2,3,5,6,9,10,11,12 14,15 ,16,17,18,19,22 23,24,31,32,35,36,37,38)	26 features selected* (2,3,4,5,6,8,9,10,11,12, 14,22,23,25,26,27 29,30 31,32,33,34,35,37,38,39)	26 features selected* (3,4,5,6,8,9,10,11,12 14,22,23,25,26,27, 28,29 30,31,32,33 34 35,37,38,39)
	Consistency subset eval	8 features selected* (1,5,6,12,23,29,31,32)	9 features selected* (1,3,5,6,12,23,31 32,37)	13 features selected* (3,4,5,6,12,14,26,29 30,35,37,38,39)	11 features selected* (4,5,6,12,14,26,29 30 35,37,39)
Evolutionary search	Cfs sub seteval	20 features selected* (3,4,5,6,13,14,15,16, 17,18,23,24,28,30,32 33,34 38,39,40)	18 features selected* (2,4,5,6,9,10,12,16,17 22 24,29,30,32,33,34 35,41)	19 features selected* (2,3,4,5,6,11,12,14,16 17,24,28,32,33,34,35 39,40,41	13 features selected* (2,5,6,13,15,20,23 24,32,33,35,38,41)
	Consistency subset eval	13 features selected* (1,2,3,5,6,11,12,13,15 21,23,32,37)	14 features selected* (1,2,3,5,6,11,12,15,22 24,27,31,32,37)	18 features selected* (3,4,5,6,8,9,12,15,17 25 26,29,30,33,34,37 38,39)	17 features selected* (3,4,5,6,8,10,12,25 26,29,30,31,33,34 37,38,39)
Multiobjective evolutionary search	Cfs subset eval	16 features selected* (1,3,5,12,18,20,21,22 23 24,26,30,33,34,35 36)	15 features selected* (1,3,5,6,12,18,21,22 24,26,30,34,35,36,37)	18 features selected* (2,3,5,6,9,10,12,13,20 22,23,31,32,33,36,37 39,40)	12 features selected* (3,5,6,12,17,23,31, 32,33,36,39,40)
Greedystepwise	Consistency subset eval	10 features selected* (1,3,5,6,12,23,26,31 32,37)	10 features selected* (2,3,6,9,12,15,23,31, 32,37)	16 features selected* (3,4,5,6,9,12,14,16,25 26,30,33,35,37, 38,39)	20 features selected* (3,4,5,6,9,10,12,16 25,26,29,30,31,32 34,35,37,38,39,40)

*return to [16] to see features name

Table 4: Most important selected features using kdd datasets with 2 classes

Feature selection algorithms		Most important features selected			
		KDD-CUP 99		NSL-KDD	
Attribute evaluators	Search methods	Test set	10%	Train set	20%
Best first	Cfs subset eval	6 features selected* (5,6,12,23,31,37)	5 features selected* (6,12,23,31,32)	6 features selected* (4,5,6,12,26,30)	8 features selected* (4,5,6,12,26,29,30,37)
	Consistency subset eval	9 features selected* (1,3,5,6,23,24,34,35 40)	7 features selected* (1,3,5,23,33,34,35)	10 features selected* (1,3,5,6,23,32,34,35 37,39)	10 features selected* (1,3,4,5,14,23,32,34, 35,37)
Genetic search	Cfs subset eval	9 features selected* (2,5,6,12,15,23,31,36 37)	12 features selected* (1,2,6,7,8,12,15,23 31,32,36,37)	15 features selected* (4,5,6,8,10,12,17,23 26,29,30,32,37,38,39)	15 features selected* (3,4,5,6,12,16,18,25, 26,29,30,31,36,37,38)
	Consistency subset eval	19 features selected* (1,2,3,5,6,11,21,23,24 29,30,33,34,35,37, 40,41)	20 features selected* (1,2,4,5,6,7,8,13,18, 21,24 ,26,28,29,32 33, 35,38,39,41)	22 features selected* (3,5,7,8,9,10,11,13 22,23,24,27,28,29,31 32,33,34 ,37,38 39,40)	19 features selected* (1,3,5,6,7,8,9,13,17,18 21,23,27,29,34,36,37 38,40)

Table 4: Continue

Feature selection algorithms		Most important features selected			
		KDD-CUP 99		NSL-KDD	
Attribute evaluators	Search methods	Test set	10%	Train set	20%
Rank search	Cfs subset eval	9 features selected* (1,3,5,6,23,24,34, 35,40)	8 features selected* (1,3,5,13,23,33,3 4,35)	11 features selected* (1,3,5,6,23,32,33,34 35,37,39)	10 features selected* (1,3,4,5,14,23,32,34 35,37)
	Consistency subset eval	6 features selected* (2,6,12,31,32,37)	6 features selected* (3,6,12,31,32,37)	12 features selected* (3,4,5,6,12,25,26,29 30,37 ,38,39)	12 features selected* (3,4,5,6,12,25,26,29 ,30,37,38,39)
PSO search	Cfs subset eval	23 features selected* (1,2,3,4,5,6,8,12,15,17 19,23,24,27,28,29,30 31.32,33,35,36,37)	23 features selected* (1,2,3,5,6,12,15,16, 17,18 ,19,23,24,26 31,32,33,34,35,36 37 38,39)	30 features selected* (1,2,3,4,5,6,8,10,12 13,15,16 ,19,23,25,26 27,28 ,29,30,31,32,33 34,35,36,37,38,39,41)	25 features selected* (3,4,5,6,8,12,15,16 23,25,26,27,28,29 30,31,32,33,34,35 36,37,38,39,41)
	Consistency subset eval	8 features selected* (2,6,12,23,29,31,32,37)	5 features selected* (3,6,12,31,32)	9 features selected* (4,5,6,12,26,29,30 37,39)	6 features selected* (4,5,6,12,29,39)
Evolutionary search	Cfs subset eval	20 features selected* (1,2,3,5,6,7,11,14,15 16,23,24,29,30,33,35 38 39,40,41)	16 features selected* (1,3,5,6,10,13,14,15 18,20,23,31,33,35 38,41)	19 features selected* (3,5,11,13,14,16,17 20, 21,23,31,32,33 34,35 37,38,39,40)	14 features selected* (1,3,4,5,6,14,24,25, 32,35,36,38,39,41)
	Consistency subset eval	11 features selected* (5,6,10,11,12,15,24, 31,32,36,37)	18 features selected* (1,2,3,4,5,6,10,11,12 16 22,23,25,31,32 33, 36,37)	9 features selected* (4,5,6,12,25,26,29 30, 37,39)	18 features selected* (3,4,5,6,8,17,19,23 25,26,29,30,33,34,37 38,39,41)
Multiobjective evolutionary search	Cfs subset eval	14 features selected* (1,3,5,6,11,16,18,20 23,24,34,35,36,40)	16 features selected* (1,2,3,5,10,16,17,24 27,28,31,33,34 35 ,37,39)	19 features selected* (1,2,3,5,12,13,20, 23,24,26 ,29,31,32 33,34,37,38,39,41)	13 features selected* (1,2,3,5,6,11,17,20 ,23,34,35,36,37)
Greedy stepwise	consistency subset eval	12 features selected* (2,3,5,6,8,12,22,23,31, 32,35,37)	8 features selected* (3,5,6,12,23,31 32,37)	19 features selected* (4,5,6,7,8,11,12,16 21 ,25,26,29,30,31 ,34,37,38,39,41)	9 features selected* (4,5,6,12,15,30,33, 37,39)

*return to [16] to see features name

Table 5: One example of the 144 files of resulted data by applying classification algorithms

Name of classifier	Accuracy	Time to build model
J48	99.9436	16.15
Hoeffding tree	99.466	68.92
Random forest	99.9576	149.42
Part	99.9367	17.91
Naive bayes	94.3329	1.25
Random tree	99.9141	2.45
Random committee	99.9501	25.29
Random subspace	99.9384	37.38
One R	98.858	1.23
Attribute selected classifier	99.947	10.43
The average	99.22444	33.043

accuracy and the training time so as to evaluate the feature selection methods as well as to determine the number of the classes.

Table 6 revealsa comparison between the feature selection algorithms based on the average of the accuracy of the ten classification algorithms for the two datasets using the (23, 5 and 2) classes. Table 7 shows a comparison between the feature selection algorithms based on the average training time of the ten classification algorithms for the two datasets using the classes (23, 5 and 2). We also calculated the average of the accuracy and the average of the training time for the data with 41features to evaluate the differences between the 41 features and the selected features. Table 8 reveals a comparison for the number of the most important selected features based on the used classes (23, 5 and 2) classes.

Analysis of experiment: In this study we will not evaluate the classification algorithms due to fact that this issue has been discussed in other studies therefore we will focus on the number of classes that are utilisedto classify the data and to determine if this number has any influence on the accuracy of the classification algorithm and the training time. As has been previously shown in the above tables, our experiment was carried out using two types of datasets which are KDD-CUP and NSL-KDD. We have preprocessed these datasets so that they are suitable for the experiment therefore, each dataset was classified into three types (23 classes, 5 classes and 2 classes). We then employed WEKA 3.8.0 Software for datamining which was applied 12 to the feature selection algorithms in order to receive the most important features for each type of dataset next we saved each result of the reduced data in a separate file which provided us with 144 files of the

Table 6: Comparison between feature selection algorithmsbased on the (23, 5and 2) classes with the average of the accuracy

	Average of the accuracy of 10 classification algorithms											
	KDD CUP 99						NSL-KDD 99					
	Test Set			10%			Train Set			20%		
Feature selection algorithms	23 Classes	5 Classes	2 Classes	23 Classes	5 Classes	2 Classes	23 classes	5 Classes	2 Classes	23 Classes	5 Classes	2 Classes
Best first+Cfs subset eval	99.56	98.23	98.79	99.29	98.91	98.80	94.25	97.20	97.03	94.73	95.68	97.14
Bestfirst+consistency subseteval	98.83	98.85	99.65	99.19	98.97	99.64	94.41	97.61	98.13	95.02	95.91	98.09
Genetic search+Cfs subset eval	99.33	98.39	99.50	99.18	98.98	99.41	93.82	97.16	98.07	93.60	96.93	97.83
Genetic search+consistency subseteval	99.07	97.91	99.55	99.19	98.61	99.60	95.22	97.58	98.50	90.77	95.74	98.17
Greedy stepwise+consistency subseteval	98.76	97.70	99.65	99.03	98.97	99.65	94.83	97.61	98.19	95.02	96.39	98.09
Ranksearch+Cfs subset eval	99.32	98.63	98.47	99.31	98.79	98.67	92.36	97.02	98.16	92.84	92.42	97.91
Ranksearch+consistency subseteval	99.16	99.06	99.57	99.17	99.20	99.61	93.56	97.33	98.44	93.51	96.83	98.03
Psosearch+Cfs subset eval	99.44	98.40	98.80	98.92	99.02	98.77	93.07	97.21	97.95	91.43	93.30	96.59
Psosearch+consistency subseteval	98.90	98.11	99.56	98.93	98.63	99.58	95.47	97.23	98.51	95.06	96.52	97.99
Evolutionary search+cfs	99.34	99.11	99.03	99.32	98.67	99.65	94.32	97.02	97.40	94.23	96.48	98.19
Evolutionary search+consistency subset eval	98.87	99.10	99.59	99.08	99.11	99.31	93.76	97.19	98.37	94.91	96.54	97.74
Multiobjectiveevolutionary search+Cfs subset eval	98.99	98.99	99.29	99.08	99.15	99.49	93.60	96.59	97.97	95.00	94.73	97.04
Average of the accuracy of selected features	99.13	98.54	99.29	99.14	98.92	99.35	94.06	97.23	98.06	93.84	95.62	97.73
Average of the accuracy of 41 features	99.22	98.25	99.57	98.99	98.92	99.61	93.22	97.14	98.40	93.51	96.93	97.92

Table 7: Comparison between feature selection algorithmsbased on the (23, 5and 2)classes with the average of the training time

	Average of training time of 10 classification algorithms											
	KDD CUP 99						NSL-KDD 99					
	Test set			10%			Train set			20%		
Feature selection algorithms	23 Classes	5 Classes	2 Classes	23 Classes	5 Classes	2 Classes	23 Classes	5 Classes	2 Classes	23 Classes	5 Classes	2 Classes
Best first+Cfs subset eval	10.87	11.49	6.42	15.16	16.78	9.54	20.97	9.82	8.66	1.95	0.94	0.91
Best first+consistency subseteval	11.10	9.69	6.90	17.66	12.17	9.46	15.07	13.64	11.03	1.67	0.91	0.83
Genetic search+cfs subseteval	13.56	11.49	7.38	21.79	15.48	14.98	21.31	13.44	18.52	2.12	1.45	1.02
Genetic search+consistency subseteval	16.85	13.96	9.20	31.76	25.75	22.29	23.31	21.21	17.25	2.31	1.43	1.35
Greedy stepwise+consistency subseteval	12.24	10.15	6.92	18.25	13.70	10.25	16.47	14.40	9.88	2.61	1.03	0.81
Rank search+cfs subseteval	20.89	11.96	5.81	32.88	14.44	9.04	21.65	12.41	9.22	2.15	1.05	0.90
Rank search+consistency subset eval	26.40	15.89	10.32	31.82	23.97	20.12	32.03	22.31	22.01	3.29	1.70	1.69
Psosearch+Cfs subset eval	11.97	11.11	7.02	29.84	17.33	8.78	24.10	10.32	12.73	1.85	1.16	0.78
PSO search+consistency subset eval	13.68	14.14	10.03	28.90	23.80	16.11	20.10	15.39	17.84	1.86	1.33	1.13
Evolutionary search+cfs subset eval	17.52	10.82	10.10	24.72	16.24	15.96	20.66	17.37	14.86	2.25	1.30	1.30
Evolutionary search+consistency subs eteval	16.49	11.43	8.40	26.36	17.99	16.88	24.42	17.65	19.89	1.92	1.07	0.97
Multiobjective evolutionary subs eteval	12.66	10.21	6.71	17.83	17.94	9.69	29.26	15.37	20.73	1.59	1.39	0.92
Average of the training time of selected features	15.35	11.86	7.93	24.75	17.96	13.59	22.44	15.28	15.22	2.13	1.23	1.05
Average of the training time of 41 feathers	33.04	24.18	16.06	51.16	34.29	29.08	42.23	34.11	31.08	3.86	2.65	2.38

Table 8: comparison between feature selection algorithmsfor the number of selected features based on the (23, 5and 2) classes

	No. of the most important selected features											
	KDD CUP'99						NSL-KDD'99					
	Test set			10%			Train set			20%		
Feature selection algorithms	23 Classes	5 Classes	2 Classes	23 Classes	5 Classes	2 Classes	23 Classes	5 Classes	2 Classes	23 Classes	5 Classes	2 Classes
Best first+Cfs subset eval	13	6	6	11	8	5	19	11	6	17	11	8
Best first+consistency subset eval	9	10	9	10	8	7	13	13	10	11	9	10
Genetic search+Cfs subset eval	13	11	9	17	10	12	21	15	15	21	23	15
Genetic search+consistency subseteval	17	18	17	22	16	20	24	24	22	19	17	19
Greedy stepwise+consistency subseteval	10	11	9	11	9	8	14	13	11	11	9	10
Ranksearch+Cfs subset eval	28	15	6	28	11	6	21	13	12	22	14	12
Ranksearch+consistency subset eval	35	25	23	26	24	23	32	26	30	34	26	25
PSO search+Cfs subset eval	13	8	8	19	9	5	18	13	9	16	11	6
PSO search+consistency subseteval	15	20	20	21	18	16	20	19	19	17	13	14
Evolutionary search+cfs subset eval	18	13	11	17	14	18	21	18	9	24	17	18
Evolutionary search+consistency subseteval	18	16	14	18	15	16	23	18	19	18	12	13
Multobjective evolution search Cfs subset eval	15	10	12	12	10	8	22	16	19	16	20	9
Average of numbers of selected features	17	13.6	12	17.7	12.7	12	20.7	16.6	15.1	18.8	15.2	13.3

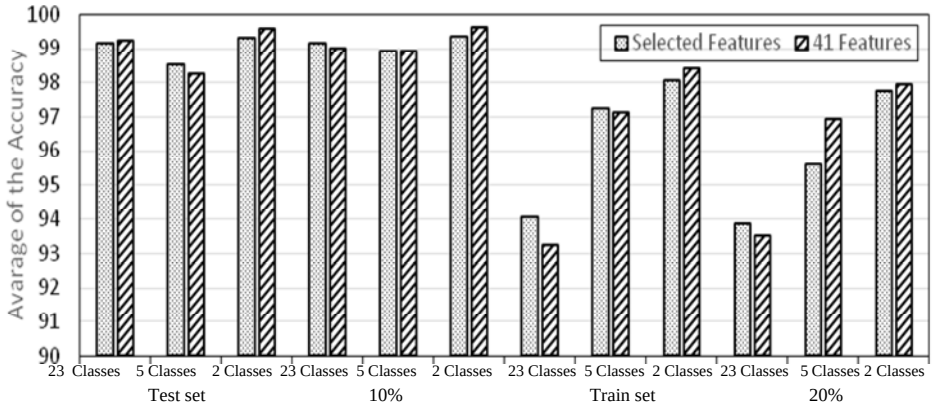

Fig. 1: Comparison between the orignal dataset and the reduced databased on the average of accuracy

reduced data. We then applied ten different classification algorithms on each reduced data file which meant thatthe number of applied algorithms was 1440. To summarise the results we have calculated the average of the accuracy of the classification algorithms as well as the average of the training time which is (the time to build the model). Table 5 illustrates an example of the collected data with the average of the accuracy and the average of the training time.

Figure 1 which is based on Table 6, presents the average of the accuracy of the classification algorithms that were applied on the reduced data files with selected important features and the original datasets we calculated the average of the results for both datasets it was evident that each dataset that usestwo classes is more accuratein comparison to the dataset that has 23 and 5 classes while there isa little difference between the averages of accuracy ofall of the datasets with selected features and the average of accuracy ofall of datasets that contains 41 features this means that the selected features provide thesame results as the original datasets.

Figure 2 which is based on Table 7 shows the average training time for the classification algorithms that were applied to the reduced data files with selected important features and the original datasets; we calculated the average of the results for both datasets while it

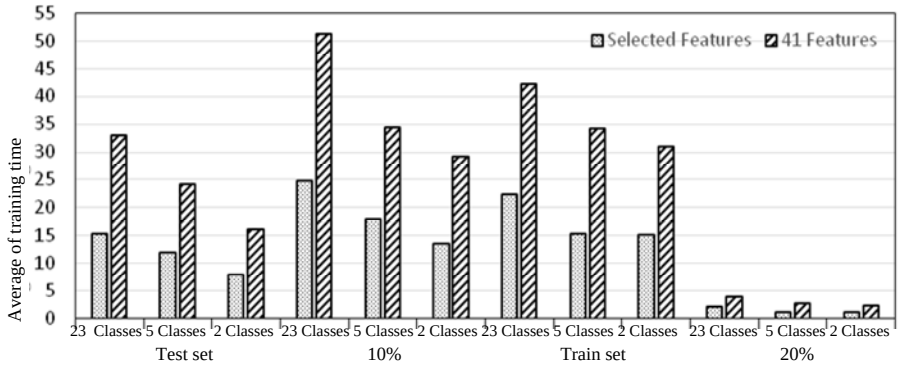

Fig. 2: Comparison between the orignal dataset and the reduced databased on the average of accuracy

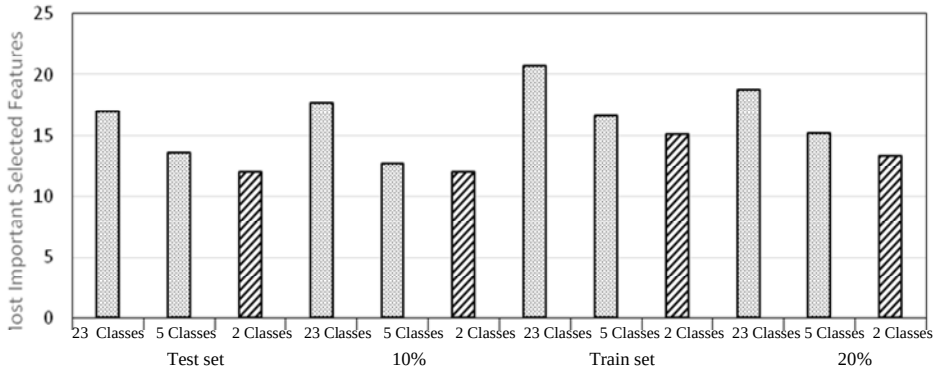

Fig. 3: Comparison between the number of classes and number of selected features

became evident that each dataset which utilises two classes has a lower average of training time in comparison to the dataset with 23 and 5 classes while there are various differences between the average training time for all of the datasets with selected features and the average training time for all of the datasets containing 41 features.

Figure 3 which is based on Table 8 reveals the comparison between the feature selection algorithms based on the number of classes and the number of the selected features the results clearly reveal that the datasets containingtwo classes have less important selected features in comparison to the datasets containingfive classes whilethe datasets with five classes have less important selected features than the datasets with 23 classes; this suggests that selecting fewer classes to classify the datasets provides results with less selected features; therefore there is less calculation time and higher accuracy as can be seen from observing other tables.

CONCLUSION

From this study, we concluded that whenever there are fewer numbers of classes utilisedin the classification algorithms, the results will be more accurate there is less training time and a lower number of selected features; this will therefore lead to less computation and preprocessing time.

RECOMENDATIONS

The future work will aim to evaluate another datasetby employing the same method while utilising a greater number of classification algorithms to ensure our final results.

REFERENCES

Amrita and P. Ahmed, 2012. A study of feature selection methods in intrusion detection system: A survey. Int. J. Comput. Sci. Eng. Inf. Technol. Res., 2: 1-25.

Araar, A. and R. Bouslama, 2014. A comparative study of classification models for detection in ip networks intrusions. J. Theoretical Appl. Inf. Technol., 64: 107-114.

Araujo, N., D.R. Oliveira, A.A. Shinoda and B. Bhargava, 2010. Identifying important characteristics in the KDD99 intrusion detection dataset by feature selection using a hybrid approach. Proceedings of the IEEE 17th International Conference on Telecommunications (ICT), April 4-7, 2010, IEEE, Cuiaba, Brazil, ISBN:978-1-4244-5246-0, pp: 552-558.

Bhuyan, M.H., D.K. Bhattacharyya and J.K. Kalita, 2014. Network anomaly detection: methods, systems and tools. IEEE. Commun. Surv. Tutorials, 16: 303-336.

Bjerkestrand, T., D. Tsaptsinos and E. Pfluegel, 2015. An evaluation of feature selection and reduction algorithms for network IDS data. Proceedings of the International Conference on Cyber Situational Awareness Data Analytics and Assessment (CyberSA), June 8-9, 2015, IEEE, London, UK., ISBN:978-0-9932-3380-7, pp: 1-2.

Chen, Y., Y. Li, X.Q. Cheng and L. Guo, 2006. Survey and taxonomy of feature selection algorithms in intrusion detection system. Proceedings of the International Conference on Information Security and Cryptology, November 29- December 1, 2006, Springer, Berlin, Germany, ISBN:978-3-540-49608-3, pp: 153-167.

Garg, T. and S.S. Khurana, 2014a. Comparison of classification techniques for intrusion detection dataset using WEKA. Proceedings of the IEEE Conference on Recent Advances and Innovations in Engineering (ICRAIE), May 9-11, 2014, IEEE, Bathinda, India, ISBN:978-1-4799-4040-0, pp: 1-5.

Garg, T. and Y. Kumar, 2014b. Combinational feature selection approach for network intrusion detection system. Proceedings of the 2014 International Conference on Parallel Distributed and Grid Computing (PDGC), December 11-13, 2014, IEEE, Kapurthala, India, ISBN:978-1-4799-7682-9, pp: 82-87.

Gogoi, P., M.H. Bhuyan, D.K. Bhattacharyya and J.K. Kalita, 2012. Packet and flow based network intrusion dataset. Proceedings of the International Conference on Contemporary Computing, August 6-8, 2012, Springer, Berlin, Germany, ISBN:978-3-642-32128-3, pp: 322-334.

Heady, R., G.F. Luger, A. Maccabe and M. Servilla, 1990. The architecture of a network level intrusion detection system. Department of Computer Science, College of Engineering, University of New Mexico, New Mexico. https://www.researchgate.net/profile/Mark_Servilla/publication/242637613_The_architecture_of_a_network_level_intrusion_detection_system/links/5564805a08ae06101ab

Kumar, M.S., 2016. A survey on improving classification performance using data pre processing and machine learning methods on NSL-KDD data. Int. J. Eng. Comput. Sci., 5: 16156-16161.

Lee, W. and S. Stolfo, 1998. Data mining approaches for intrusion detection. Proceedings of the 7th USENIX Security Symposium, January 26-29, 1998, USENIX Association, Berkeley, CA., USA., pp: 79-94.

Moraglio, A., C.D. Chio, J. Togelius and R. Poli, 2008. Geometric particle swarm optimization. J. Artif. Evol. Appl., 2008: 1-14.

Oyebode, E.O., S.G. Fashoto, O.A. Ojesanmi, O.E. Makinde and O. State, 2011. Intrusion detection system for computer network security 1. Aust. J. Basic Appl. Sci., 5: 1317-1320.

Patcha, A. and J.M. Park, 2007. An overview of anomaly detection techniques: Existing solutions and latest technological trends. Comput. Networks, 51: 3448-3470.

Sheen, S. and R. Rajesh, 2008. Network intrusion detection using feature selection and decision tree classifier. Proceedings of the IEEE Region 10 Conference TENCON, November 19-21, 2008, Hyderabad, pp: 1-4.

Singh, H. and D. Kumar, 2015. A study on performance analysis of various feature selection techniques in intrusion detection system. Int. J., 3: 50-54.

Stolfo, J., W. Fan, W. Lee, A. Prodromidis and P.K. Chan, 2000. Cost-based modeling and evaluation for data mining with application to fraud and intrusion detection: Results from the JAM Project by Salvatore. Computer Science Department, Columbia University, New York, USA. http://ids.cs.columbia.edu/sites/default/files/ada511232.pdf

Tavallaee, M., E. Bagheri, W. Lu and A.A. Ghorbani, 2009. A detailed analysis of the KDD CUP 99 data set. Proceedings of the 2nd IEEE Symposium on Computational Intelligence for Security and Defence Applications, July 8-10 2009, Ottawa, Canada -.