JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 36, 293-308 (2020)
DOI: 10.6688/JISE.202003_36(2).0009

Computational Intelligence Algorithms to Handle Dimensionality Reduction for Enhancing Intrusion Detection System

HUSAM IBRAHIEM ALSAADI[1,2], RAFAH M. ALMUTTAIRI[3],
OGUZ BAYAT[1] AND OSMAN NURI UCANI[1]
[1]*Faculty of Engineering*
Altinbas University
Istanbul, 34676 Turkey
[2]*Faculty of Basic Education*
University of Mustansiriy
Baghdad, 10052 Iraq
[3]*College of Information Technology*
University of Babylon
Babylon, 51002 Iraq
E-mail: husam.alsaadi@ogr.altinbas.edu.tr; rafahmohammed@gmail.com;
oguz.bayat@altinbas.edu.tr; osman.ucan@altinbas.edu.tr

In this paper, propose to use computational intelligence models to improve intrusion detection system, the computational intelligence algorithms are used as preprocessing steps for selecting most significant features from network data. Two computational intelligence algorithms, namely Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) are implemented to generate subset of relevant features. The computational intelligence approaches have been applied to optimize the classification of algorithms. The most significant features obtained from computational intelligence is fed into the classification algorithm. Novelty of this presents research of use computational intelligence algorithms namely Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) for handling dimensionality reduction. The dimensionality reduction is obstructed time processing of classification algorithms. Three classification algorithms namely K-Nearest Neighbors (KNN), Support Vector Machine (SVM) and Naïve Bayes (NB) are implemented for intrusion detection system. Benchmark datasets, namely, KDD cup and NSL-KDD datasets are used to demonstrate and validate the performance of the proposed model for intrusion detection. From the empirical results, it is observed that the classification algorithm has improved the intrusion detection system with using computational intelligence algorithms. A comparative result analysis between the proposed model and different existing models is presented. It is concluded that the proposed model has outperformed of conventional models.

Keywords: computational intelligence algorithm, classification algorithms, intrusion detection system, support vector machine, *K*-nearest neighbors

1. INTRODUCTION

Data Security has seen huge advancement over the most recent couple of decades. The technologies have been developed so that the accessibility of electronic information handling frameworks/ information systems got inside the range of independent venture and home clients. These information systems got interconnected through an overall net-

Received September 23, 2019; revised September 28 & October 11, 2019; accepted October 17, 2019.
Communicated by Osamah Ibrahim Khalaf.

work by and large known as web. Furthermore, Enormous development of, broad utilization of web has changed the information prepared and directed organizations over the web in the most recent decade [1]. To ensure the computers and the cost of harms which is caused by such unapproved get to, a compelling and proficient intrusion detection system should be utilized to be protected the security of information systems. Presently a-days intrusion detection systems have developed to be an essential part of system security foundation [2]. The idea behind using data mining approaches is to allow system to detect the unknown attacks and also their variations. The aim target of this presents research work is to use the computational intelligence models to improve the classification algorithms for detecting intrusion. The paper is organized as follows in Section 1 presented an introduction.

2. RELATED WORK

Intrusion is fundamentally an effective sequence of linked occurrences that deliberately cause damage or assault, attempting to make the scheme unusable, accessing unlawful data, or manipulating such data, both successful and ineffective. The methods of data mining were proposed to detect and classify the intrusions; it can be obtained individually or as tools. Intrusions detection system presents possibility for numerous research problems which is discussed in this section. Panda *et al.* [3] and Yoshimasa *et al.* [4] introduced generic framework of Network Intrusion Detection System (NIDS) using Naïve Bayes algorithm. It is noted that the proposed model performs good according to false positive rate, cost, and computational time. P. Amudha *et al.* [5] used Random forest tree algorithm for classification IDS. Authors noted that the proposed system gives better detection with respective to evaluation metrics for two types of attacks. Different classification algorithms are proposed to detect intrusion [6-9]. Bukhtoyarov *et al.* [10] implemented probability based on neural networks structures for detection NIDS. To development neural network ensembles, the GPEN implements genetic programming process to discover a most favorable function for coming together the over classifiers into an ensemble. This KDD Cup 1999 data set is used, the results of this research is compared with results of those papers [11]. From prediction results, it investigated that the accuracy of the model for detection Probe attacks with research paper is available in [12]. Cordeiro *et al.* [12] proposed PSO algorithm for enhancing the classification algorithms. Norouzian *et al.* [13] proposed Multi-Layer Perception Neural Network algorithm for classification intrusion in to attack and normal. Dong *et al.* [14] used SVM algorithms to classify the IDS for enhancing the network security. From experimental analysis noted that this algorithm is significantly privileged which is compared with existing classification algorithm with accuracy and speed of building model. Feature selection is extremely significant step in data analysis; particularly feature selection algorithm is dealing with high dimensional space of data. The main ideal behind of using feature selection method is to simplify of data set for reducing its dimensionality space of data and find out relevant essential features with given good classification accuracy. ACO is applied in numbers of research papers used to improve intrusion detection system for selecting significant features in [15-17]. The ACO algorithm is inspired that is based on social behavior of ant colonies [18]. In literatures numbers of experimental results are shown ACOs superiority over existing method [19]. In [20], two features selection

methods are proposed namely PCA and PSO to select most significant features from IDS data for enhancing IDS. The SVM is processed for prediction purpose. They used standard KDD cup dataset foresting algorithm. The PSO approach is important for selecting significant features. It is considered for enhancing the classification algorithm intrusion detection system. From the literature, it is observed that the PSO method an effective and efficient algorithm for selecting features [21, 22].

3. METHODOLOGY

Automatic classification of IDS using machine learning algorithms is the main target of the presents research work. The main objective of presents research to analyze dimensionality reduction and it impacts the classification algorithms. In order to examine the proposed model two standard intrusion detection data sets are conducted. These data sets have normal and attack packets in order to help improving the detection of intrusion. To analyze the data, the preprocessing stage is required for putting the data in proper format. Two computational intelligence algorithms namely ACO and PSO are considered to select significant features from dataset; due to the network data has many different formats and dimension. The complexity of classification algorithm is greatly reduced if the numbers of features in data set is reduced. Handling dimensionality reductions improve the classification results and time processing. These algorithms use to enhance time of building the model. These features are fed to the classification algorithms for detecting the intrusion. Three classification algorithms have considered for classification network data as normal or attack packets. The evaluation metrics are used to evaluate the analysis results. Finally, a comparative analysis results between the proposed systems and different existing models is presented. The detailed description for proposed model is next.

3.1 Data Sets

Two standard network traffic data sets were used, KDD cup and NSL-KDD. The description of these data sets is presented in the next subsections. The data sets have three anomalies features namely protocol types, services and flag. We have developed this algorithm for convert these anomalies features into numerical features. This algorithm has helped to improve classification convert by these features.

(A) KDD cup'99 data set

The KDD (Data Mining and Knowledge Discovery) cup data set has employed in the 3rd international knowledge discovery and data mining tools completion for developing intrusion detection and discovery robust data mining algorithm for distinguishing between normal and attack packets. In 1998, the DARPA intrusion detection evaluation program was developed simulation form collection data from Local-Area Network (LAN) by Lincoln Lab. KDD cup data set was created by using five million connection records; features were extracted from network connection. The connection has been extracted values from IP addresses of sequence of TCP packets at starting and ending at some well defined times for the presents research intrusion detection dataset has collected from standard KDD cup. This contains three major intrusion namely Denial of Service (DOS), Probe, User to Root (U2R) and Remote to Local (U2R). KDD cup is represented by 41 features [23].

(B) NSL-KDD data set

The NSL-KDD is advanced versions for KDD cup data for analyzing and detecting intrusion in network. The NSL-KDD data set is proposed by McHugh [8]. The 4,898,431 entries are available in NSL-KDD data set; these entries are used for training and testing. Furthermore, each record contains 41 features and these features labeled as either normal or attacks. The NSL-KDD dataset contains three major intrusions namely (DOS), Probe, and (U2R) & (U2R) [23].

3.2 Preprocessing

Preprocessing is a main stage in data analysis; it is employed to manage real world datasets into an intelligible format. Undoubtedly, the most of real world datasets have been imperfect, noisy and very difficult to determine the behavior of this data. Preprocessing play vital role for analysis patterns from network data for achieving accurate results. Therefore, the preprocessing steps are essential part in IDS to improve the data mining algorithms for classification of intrusions from network datasets. In the present research the computational intelligence algorithms are proposed to select the significant features from dataset. Description of information gain method is presented in the next subsections.

(A) Ant colony optimization (ACO)

ACO algorithm is one of the most important probabilities techniques that are used to solve computation problem. It is applied to find the best for solving the problem based on the rules of real ants. The ACO algorithm was developed by Dorigo in 1992 in his Ph.D. thesis [24], the algorithm was focusing to find out the best path in graph by using the behavior of ants for seeking the best path between their colony and source of food. The Ants traversal through graph where the less numbers of nodes found, the graph nodes are entirely connected to permit features to be coming features.

We have applied ACO algorithm to select the significant features from different dataset that have been used. 8 most significant features have selected from KDD cup is shown in Table 1. Table 2 shows the subset features selection from NSL-KDD data set. The ACO is employed to discover the space of subset from among of all features. These significant features are fed into the classification algorithms to build robust IDS system. It is observed that the time processing to select the features is more suitable.

(B) Particles swarm optimization (PSO) method

The PSO is a population based on computation intelligence proposed by Eberhat and Kennedy [25]. PSO is an effective and esteemed global search system [26]. PSO algorithm is called reasonable algorithm due to the accompanying reasons: simple encoding of feature, global search, being reasonable computationally, less parameters and less demanding execution for addressing and selection of significant feature issues [27].

We have applied PSO algorithm for enhancing the classification of IDS. The important 8 features have been selected by using PSO algorithm. Table 3 shows the significant features obtained by PSO method for KDD CUP. Most significant features selection using PSO algorithm for NSL-KDD data set is demonstrated in Table 4. The time processing of selecting the significant features is very less.

Table 1. Most significant features of (KDD CUP) using ACO method.

Attacks	Features numbers	Features Name
DOS	34	dst_host_same_srv_rate
	4	Flag
	27	rerror_rate
	37	dst_host_srv_diff_host_rate
	41	dst_host_srv_rerror_rate
	19	num_access_files' real
	3	Service
	21	is_host_login
Probe	5	src_bytes
	26	same_srv_rate
	25	srv_rerror_rate
	31	dst_host_same_srv_rate
	9	logged_in
	18	is_host_login
	3	dst_bytes
	14	num_file_creations
U2R and R2L	40	dst_host_rerror_rate
	39	dst_host_srv_serror_rate
	10	Hot
	8	wrong_fragment
	14	root_shell
	3	Flag
	16	num_root
	40	Service

Table 2. Most significant features of (NSL-KDD) using ACO method.

Attacks	Features numbers	Features Name
DOS	5	wrong_fragment
	37	dst_host_rerror_rate
	34	dst_host_srv_diff_host_rate
	10	num_compromised
	31	dst_host_same_srv_rate
	41	Flag
	30	dst_host_srv_count
	40	Service
Probe	39	protocol_type
	9	logged_in
	33	dst_host_same_src_port_rate
	31	dst_host_same_srv_rate
	12	su_attempted
	40	Service
	14	num_file_creations
	20	Count
U2R and R2L	31	dst_host_same_srv_rate
	41	Flag
	8	num_failed_logins
	32	dst_host_diff_srv_rate
	7	Hot
	33	dst_host_same_src_port_rate
	4	Land
	40	Service

Table 3. Most significant features of (KDD CUP) using PSO method.

Attacks	Features numbers	Features Name
DOS	7	Land
	38	dst_host_serror_rate
	4	Flag
	25	serror_rate
	11	num_failed_logins
	12	logged_in
	15	su_attempted
	34	dst_host_same_srv_rate
Probe	19	num_access_files' real
	13	num_compromised
	11	num_failed_logins
	12	logged_in
	14	root_shell
	32	dst_host_count
	17	num_file_creations
	18	num_shells
U2R and R2L	18	num_shells
	17	serror_rate
	34	dst_host_same_srv_rate
	8	wrong_fragment
	27	rerror_rate
	5	src_bytes
	4	Flag
	25	num_file_creations

Table 4. Most significant features of (NSL-KDD) using PSO method

Attacks	Features numbers	Features Name
DOS	4	Lan
	34	dst_host_srv_diff_host_rate
	22	serror_rate
	19	is_guest_login
	11	root_shell
	5	wrong_fragment
	18	num_failed_logins
	17	Hot
Probe	16	num_access_files
	12	su_attempted
	11	root_shell
	36	dst_host_srv_serror_rate
	14	num_file_creations
	38	dst_host_srv_rerror_rate
	8	num_failed_logins
	24	rerror_rate
U2R and R2L	33	dst_host_same_src_port_rate
	8	num_failed_logins
	16	num_access_files
	15	num_shells
	14	num_file_creations
	18	is_host_login
	11	root_shell
	7	Hot

3.3 Classification Algorithms

In this section, we discuss details description of classification algorithms that are used for automatic IDS. Three classifiers namely Support Vector Machine (SVM), K-Nearest Neighbor (K-NN) and Naïve Bayes (NB) are used to detect intrusion.

(A) Support Vector Machine (SVM)

SVM was proposed by Vapnik [28] in 1963. It is one significant supervised machine learning algorithm used for large dataset and gives more accurate results. The SVM has been designed for dichotomist classification problem such as binary classification with two classes or with multi classes. The SVM is used to find out the optimal dichotomist hyper plane that can help to maximize the margin which can make the largest separation of two or more classes. In order to classify two classes, two parallel hyper planes are constructed; the SVM tries to find out separating hyper lance and maximize the distance between these two hyper planes. Hence the hyper plane has the largest distance this is called good separation [29]. The SVM obtained lower error when the margin is large.

(B) K-Nearest Neighbor (K-NN)

K-nearest neighbor (KNN) classification algorithm is one of the powerful data mining algorithms. It is theoretically mature with low complexity. The fundamental idea of KNN algorithm is that if training data, if most of K values NN belong to same class with training dataset. The Nearest Neighbor can be the single or multidimensional feature vector space that is employed to find out the training data closet and the closest criteria is Euclidean distance of the sample space. Euclidean distance method is considered in the K-NN algorithm for discovering the closet point between the features.

(C) Naives Bayes algorithm

The Naïve Bayes algorithm is Bayesian probability machine learning algorithm. The main idea of the algorithm is used as conditional probability for classification network data as normal and attacks. The NB algorithm is very robust independence assumption [30, 31].

4. PERFORMANCE MEASURES

The performance measures have been carried out to test the results of proposed model. The Accuracy, False Positive, Precision, True Positive and Time are used. The equations performance measures are as follow:

$$\text{False Positive Rate (FPR)} = \frac{FP}{TN + FP} \%100 \tag{1}$$

$$\text{True Positive Rate (TPR)} = \frac{TP}{TN + FN} \%100 \tag{2}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \%100 \tag{3}$$

$$\text{Precision} = \frac{TP}{TP + FP} \%100 \tag{4}$$

True negative (TN): Correctly classified of valid records as normal record.
True positive (TP): Correctly classified of attack records as attacks.
False positive (FP): The percentage of incorrect records normal data as attacks.
False negative (FN): The percentage of incorrect records attacks as normal record.

5. EXPERMENTAL ANALYSES

The experiment was conducted and evaluated by using KDD cup and NSL-KDD datasets. The MATLAB R2016a with −64 windows 7 Ultimate with the core i7 processor and 8 GB RAM is used to compute all programs. In this experiment, 31 major attacks are considered in KDD CUP data and NSL-KDD datasets. Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) algorithms are applied to select most significant features. Theses algorithms are used to analyze dimensionality reduction for improving the classification results. Throughout the optimization of classification algorithms, the accuracy of detection rate for each attack and normal class in to two datasets are enhanceed. The main object of the proposed model is employed to integrate the intelligence models with classification algorithms for increasing the performance of classification.

By gradually removing the less important features; consequently, the integrated model has ability to decide the important features. In this, work three experiments have conducted for each dataset, we have been decided to work with individual attack in dataset. The dataset has three major attacks namely DOS, Probe, U2R & R2L attacks. In KDD dataset, the Dos attack contains 17722 record packets and it is divided into 12405 for training and 5317 for testing, the Probe attack contains 12111 record, this data is divided the training 8477 and testing 3634, the U2R and R2L together contains 9189 record, it is divided into 6432 for training and for testing 2757.

Similarly, NSL-KDD data has four major attacks, the DOS attack contains 29175, it is divided into 20422 and 8753 with respective to training and testing respectively. The Probe attack contains 20292 record, it is divided into 14204 training and 6088 testing, also U2R and R2L together contains 34378, the data is divided into 24064 For training and 10314 for testing.

Table 5 indicates that the results obtained by using KNN classifier with respective to the computational intelligence algorithms for features selection. The KDD CUP data set has 41 features, the processing of all features is very complex due to time cost. The 8 subset features are obtained from ACO and PSO computational intelligence algorithms. The data set has been divided into 70% training data and 30% testing data. From the empirical results, it is observed that the KNN with PSO and ACO algorithm is performed 100% out according to the accuracy metrics using DOS attacks and normal datasets. Furthermore, the KNN classifier with PSO, algorithm using Probe attacks with normal network data obtained best result with 99.26% in terms of accuracy metrics. The KNN with ACO algorithms give good accuracy 100% using U2R and R2L attacks and normal dataset. Table 6 shows the results of KNN classifier with PSO and ACO using NSL-KDD data set. As the results show, KNN with PSO and ACO methods is the best in the DOS attack; it obtained 100% according to accuracy. The KNN classifier with feature selection using ACO algorithm gives best accuracy 100% in Prop attack. However, the accuracy of KNN with PSO algorithm is best in the U2R and R2L attack; the accuracy is

99.96%. Figs. 1-3 illustrate the performance of KNN algorithm with computational intelligence methods using KDD CUP dataset. The performance of KNN with computational intelligence methods using NSL-KDD cup dataset is shown in Figs. 4-6.

Table 5. Results of KNN with CIMs for DOS, Probe and U2R & R2L attack in KDD CUP.

Attack	Data	Train data	Testing data	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Second
DOS	17722	12405	5317	KNN+PSO	100	99.96	99.96	100	0.89
				KNN+ACO	100	100	100	100	0.96
Probe	12111	8477	3634	KNN+PSO	99.26	97.49	96.83	98.42	0.043
				KNN+ACO	99.10	99.75	99.84	99.92	0.052
U2R & R2L	9189	6432	2757	KNN+PSO	99.24	98.33	93.58	99.03	0.046
				KNN+ACO	100	99.97	98.58	99.97	0.048

Table 6. Results of KNN with CIMs for DOS, Probe and U2R & R2L attack in NSL-KDD.

Attack	Data	Train data	Testing data	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Second
DOS	29175	20422	8753	KNN+PSO	100	98.07	98.07	98.69	0.085
				KNN+ACO	100	97.34	97.22	97.67	0.092
Probe	20292	14204	6088	KNN+PSO	97.28	98.62	97.56	97.23	0.076
				KNN+ACO	100	99.97	99.95	99.29	0.082
U2R & R2L	34378	24064	10314	KNN+PSO	99.96	99.82	99.96	98.02	0114
				KNN+ACO	99.52	99.65	99.85	99.79	0.122

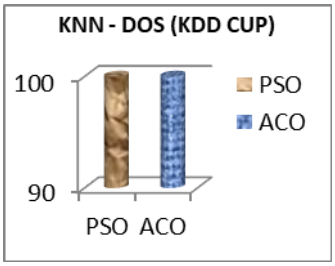

Fig. 1. Performance KNN-CIMs for DOS attack in KDD CUP.

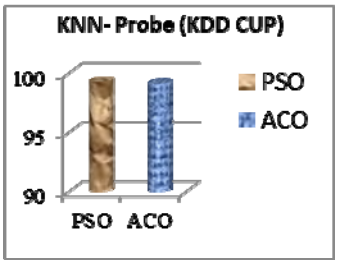

Fig. 2. Performance KNN-CIMs for Probe attack in KDD CUP.

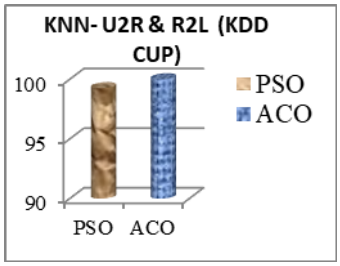

Fig. 3. Performance KNN-CIMs for U2R & R2L attack in KDD CUP.

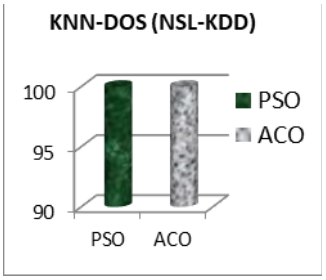

Fig. 4. Performance KNN with CIMs for DOS attack in NSL-KDD.

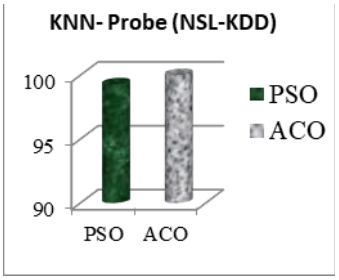

Fig. 5. Performance KNN with CIMs for Probe attack in NSL-KDD.

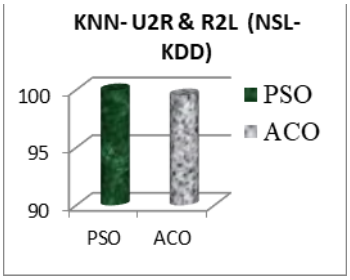

Fig. 6. Performance KNN with CIMs for U2R and R2L attack NSL-KDD.

In order to enhance the proposed model, we are decided to work with another classifier. The PSO and ACO feature selection method are proposed for improving the Naïve Bayes classifier. The computational intelligence methods assist to increase the accuracy of NB classifier and reduce the time of building the model. After obtaining the goodness features using PSO and ACO the features are fed into NB classifier. The results obtained by using NB classifier with PSO and ACO Methods using DOS, Probe and U2R and R2L with respective to KDD CUP is shown in Table 7. The NB algorithm with ACO algorithm using DOS attacks obtained superior accuracy is 99.92%. Whereas the results of NB algorithm with ACO algorithm is best using Probe attacks and normal dataset, it has obtained 99.39% with respective to accuracy. The NB with PSO algorithm is demonstrated the best by using U2R and R2L data with respect to accuracy 91.22%.

Furthermore, Table 8 shows the performance results of NB with intelligent models using NSL-KDD data set. We have applied this proposed model with all three types of attacks. From the experimental results, the NB classifier with ACO method using DOS attack and normal is a good and the accuracy result is 87.10%. In the Probe attack data, it is observed that the NB with PSO method has good accuracy 87.70%. The performance of NB with PSO method is the best by using U2R and R2L; it is accuracy with is 96.94%.

Table 7. Results of NB with CIMs for attacks in KDD CUP data set.

Attack	Data	Train data	Testing data	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Second
DOS	17722	12405	5317	KNN+PSO	98.92	99.98	99.98	100	0.060
				KNN+ACO	99.92	99.96	99.96	100	0.040
Probe	12111	8477	3634	KNN+PSO	98.92	98.23	99.02	100	0.023
				KNN+ACO	99.39	99.63	97.39	98.68	0.025
U2R & R2L	9189	6432	2757	KNN+PSO	91.22	90.44	95.05	99.79	0.018
				KNN+ACO	90.71	90.88	95.65	99.79	0.020

Table 8. Results of NB with CIMs for attacks in NSL-KDD data set.

Attack	Data	Train data	Testing data	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	T/Sec
DOS	29175	20422	8753	NB+PSO	80.22	73.13	97.92	98.96	0.075
				NB+ACO	87.10	69.02	88.23	95.45	0.078
Probe	20292	14204	6088	NB+PSO	87.70	86.73	81.98	96.57	0.055
				NB+ACO	87.68	89.04	76.98	94.31	0.062
U2R & R2L	34378	24064	10314	NB+PSO	96.94	90.92	90.92	99.73	0.095
				NB+ACO	73.49	93.39	71.69	98.67	0.082

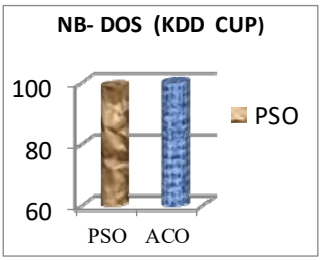

Fig. 7. Performance NB with CIMs for DOS attack in KDD CUP.

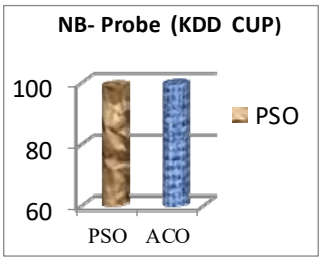

Fig. 8. Performance NB CIMs for Probe attack in KDD CUP.

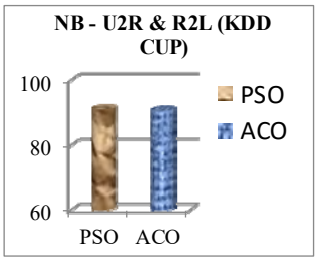

Fig. 9. Performance NB with CIMs for U2R and R2L in KDD CUP.

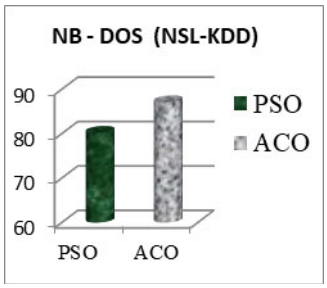

Fig. 10. Performance NB with CIMs for DOS attack in NSL-KDD.

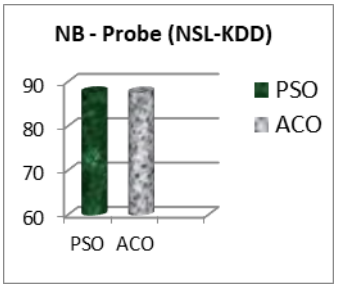

Fig. 11. Performance NB with CIMs for Probe attack in NSL-KDD.

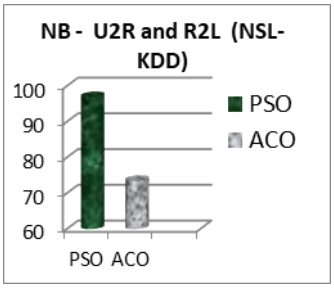

Fig. 12. Performance NB with CIMs for U2R & R2L in NSL-KDD.

Figs. 7-9 demonstrate the performance of NB method with computational algorithms using KDD CUP dataset. The performance of NB algorithm with computational algorithms using NSL-KDD dataset is displayed in Figs. 10-12.

Table 9 shows performance of SVM classifier CIMs for enhancing intruding detection system using KDD CUP. It is observed that the proposed model has outperformed better than the all existing algorithms. From the results obtained the SVM classifier with using subset feature of ACO. It is observed that the accuracy of SVM + ACO algorithms is best 89.98% in DOS attack. Whereas the Probe attacks demonstrated that SVM+ACO algorithm is good, the result is 98.10 % it according to the accuracy. The SVM+PSO algorithms using U2R and R2L are obtained good results, the results value is 89.95 with respective to accuracy metrics. Figs. 13-15 illustrate the performance of SVM with CIMs using KDD CUP dataset.

Table 9. Results of SVM with CIMs for attacks in KDD CUP data set.

Attack	Data	Train data	Testing data	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Second
DOS	17722	12405	5317	SVM+PSO	76.34	100	100	100	0.039
				SVM+ACO	89.98	100	100	100	0.039
Probe	12111	8477	3634	SVM+PSO	97.39	99.96	92.22	95.67	0.022
				SVM+ACO	98.10	99.45	96.04	97.92	0.025
U2R & R2L	9189	6432	2757	SVM+PSO	89.95	92.20	58.80	95.73	0.042
				SVM+ACO	89.92	100	54.98	88.41	0.042

Table 10. Results of SVM with CIMs for attacks in NSL-KDD data set.

Attack	Data	Train data	Testing data	Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	Second
DOS	29175	2042	8753	SVM+PSO	80.45	74.16	97.35	98.54	0.089
				SVM+ACO	89.98	87.80	93.98	95.49	0.076
Probe	20292	14204	6088	SVM+PSO	81.16	77.60	88.33	94.34	0.049
				SVM+ACO	81.01	77.60	80.33	94.34	0.051
U2R & R2L	34378	24064	10314	SVM+PSO	75.12	65.23	55.02	72.15	0.056
				SVM+ACO	78.37	59.32	59.32	60.08	0.062

Similarity, Table 10 shows the performance of SVM with CIMs using NSL-KDD dataset. From the empirical results, it is observed that the results of SVM+ACO algorithms in DOS attack is 89.98% with respect to accuracy metrics, this result is the best. The result of SVM+PSO with using Probe attack data is 81.16% compared with another feature selection methods. The SVM+ACO method obtained good performance using U2R and R2L attack data. The results value of integrated SVM+ACO algorithm method is 78.37 %. Figs. 16-18 display the graphical representations of SVM and computational intelligence algorithms using NSL-KDD.

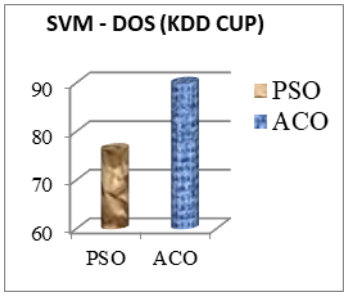

Fig. 13. Performance SVM with CIMs for DOS attack in KDD CUP.

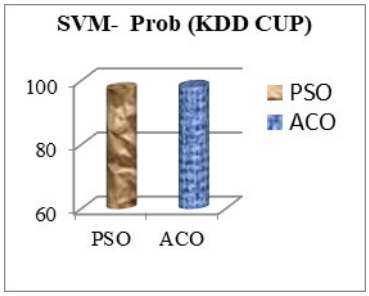

Fig. 14. Performance SVM with CIMs for Probe attack in KDD CUP.

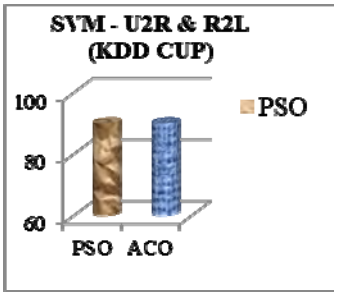

Fig. 15. Performance SVM-CIMs for U2R&R2L attack in KDD CUP.

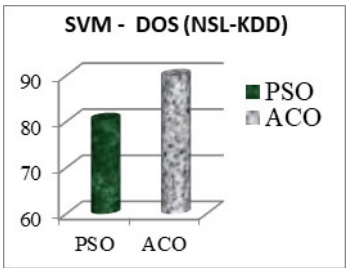

Fig. 16. Performance SVM with CIMs for DOS attack in NSL-KDD.

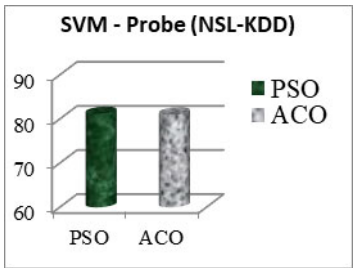

Fig. 17. Performance SVM with CIMs for Probe attack in NSL-KDD.

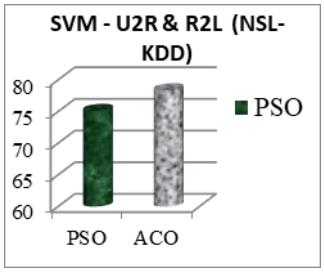

Fig. 18. Performance SVM-CIMs for U2R&R2L attack in NSL-KDD.

6. PERFORMANCE AND COMPARISON PROPOSED MODEL

The Comparison to evaluate and test the proposed model to detect intrusion according to classification accuracy Table 11 summarizes the results of proposed model against the existing models. From the results, it is observed that the results of the proposed model are better than all conventional algorithms. Fig. 19 illustrates the performance of the proposed model in comparison with different existing algorithms with respect to accuracy metrics. It proved that the proposed model is more accurate and takes less time to build model.

Table 11. Comparison for accuracy rate with other models that existing depend on the evaluation datasets.

Authors, years	Models	No. of Select features	Data set	Accuracy %
Rajinder, 2016 [36]	Best first search +NB Genetic search +NB Greedy stepwise +NB	6 15 6	NSL-KDD	82.01 89.85 82.01
Yinhui Li *et al.*, 2012 [33]	GFR+ SVM	19	K-DD CUP	98.67
Shenfield *et al.*, 2018 [35]	ANN			98.0
Mohammed A. *et al.*, 2015 [32]	wrapper and filter +LSSVM	6	K-DD CUP	99.90
Mukkamala *et al.*, 2005 [34]	SVM with PBR		K-DD CUP	99.59
Proposed system, 2019	Proposed model. We have the best results.	8 for PSO method 8 for ACO method	K-DD CUP NSL-KDD	100

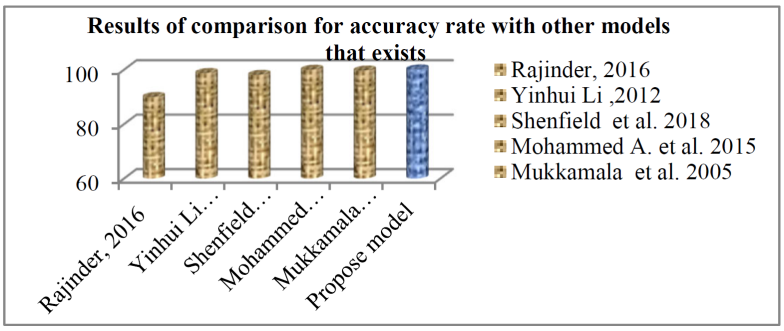

Fig. 19. Display the performance of proposed model against existing models.

7. CONCLUSIONS

One of biggest challenges of IDS is dimensionality reduction. For addressing the dimensionality reduction, we have applied different computational intelligence algorithms. These algorithms have robust search ability in the problem space and can efficiently find optimal subset features. The useless and irrelevant, redundant features are reduced. The PSO and ACO computational intelligence algorithms are employed for feature reduction and subset feature selection. 8 subset features have been selected using PSO and ACO algorithms. These optimal features are processed using classification algorithms. We have used three classification algorithms like KNN, NB and SVM, it is realized that the classification algorithms the obtained very good results and time of building model. Two datasets are considered to evaluate the proposed model; the dataset is divided into 7o% training and 30% testing dataset. A comparative results analysis between the proposed models with the existing models is presented. It is proved that the proposed model is extremely reliable to detect intrusion, in addition the proposed model base computational intelligence models outperformed method. It is demonstrated that the performance of classification is improved with using subset features. Furthermore, it is concluded that the proposed model is outperformed of existing models.

REFERENCES

1. A. N. Toosi and M. Kahani, "A new approach to intrusion detection based on an evolutionary soft computing model using neuron-fuzzy classifiers," *Computer Communications*, Vol. 30, 2007, pp. 2201-2212.
2. S. J. Horng, M. Y. Su, Y. H. Chen, T. W. Kao, R. J. Chen, J. L. Lai, and C. D. Perkasa, "A novel intrusion detection system based on hierarchical clustering and support vector machines," *Expert Systems with Applications*, Vol. 38, 2011, pp. 306-313.
3. M. Panda and M. R. Patra, "Network intrusion detection using naive bayes," *International Journal of Computer Science and Network Security*, Vol. 7, 2007, pp. 256-263.
4. Y. Tsuruokam and J. Tsujii, "Training a naive bayes classifier via the EM algorithm with a class distribution constraint," in *Proceedings of the 7th Conference on Natural Language Learning at HLTNAACL*, Vol. 4, 2003.
5. P. Amudha and H. A. Rauf, "Performance analysis of data mining approaches in intrusion detection," in *Proceedings of IEEE International Conference on Process Automation*, *Control and Computing*, 2011, pp. 1-6.
6. R. C. A. Naidu and P. S. Avadhani, "A comparison of data mining techniques for intrusion detection," in *Proceedings of IEEE International Conference on Advanced Communication Control and Computing Technologies*, 2012, pp. 41-44.
7. R. Chitrakar and H. Chuanhe, "Anomaly based intrusion detection using hybrid learning approach of combining *k*-medoids clustering and naïve bayes classification," in *Proceedings of IEEE 8th International Conference on Wireless Communications*, *Networking and Mobile Computing*, 2012.
8. R. Chitrakar and H. Chuanhe, "Anomaly detection using support vector machine classification with *k*-medoids clustering," in *Proceedings of IEEE International Conference on Third Asian Himalayas*, 2012.
9. E. Bahri, N. Harbi, and H. N. Huu, "Approach based ensemble methods for better and faster intrusion detection," in *Proceedings of the 4th International Conference on Computational Intelligence in Security for Information Systems*, 2011, pp. 17-24.
10. V. Bukhtoyarov and V. Zhukov, "Ensemble-distributed approach in classification problem solution for intrusion detection systems," in *Proceedings of International Conference on Intelligence Data Engineering and Automated Learning*, 2014, pp. 255-265.
11. A. J. Malik, W. Shahzad, and F. A. Khan, "Binary PSO and random forests algorithm for probe attack detection in a network," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2011, pp. 662-668.
12. Z. Cordeiro and G. L. Pappa, "A PSO algorithm for improving multi-view classification," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2011, pp. 925-932.
13. M. R. Norouzian and S. Merati, "Classifying attacks in a network intrusion detection system based on artificial neural networks," in *Proceedings of IEEE 13th International Conference on Advanced Communication Technology*, 2011, pp. 868-873.
14. J. Pu, L. Xiao, Y. Li, and X. Dong "A detection method of network intrusion based on SVM and ant colony algorithm," in *Proceedings of National Conference on Information Technology and Computer Science*, 2012, pp. 153-156.

15. S. M. Vieira, J. M. C. Sousa, and T. A. Runkler, "Fuzzy classification in ant feature selection," in *Proceedings of IEEE International Conference on Fuzzy Systems*, 2008, pp. 1763-1769.
16. M. H. Aghdam, N. Ghasem-Aghaee, and M. E. Basiri, "Application of ant colony optimization for feature selection in text categorization," in *Proceedings of IEEE Congress on Evolutionary Computation*, 2008, pp. 2872-2878.
17. E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, Oxford University Press, NY, 1999.
18. K. J. Lee, J. Joo, J. Yang, and V. Honavar, "Experimental comparison of feature subset selection using GA and ACO algorithm," *Advanced Data Mining and Application*, LNCS, 2006, pp. 465-472.
19. A. P. Engelbrecht, *Computational Intelligence: An Introduction*, 2nd ed., Wiley, NJ, 2007.
20. J. Kennedy, R. C. Eberhart, and Y. Shi, *Swarm Intelligence Evolutionary Computation Series*, Morgan Kaufmann Publishers, CA, 2001.
21. I. Ahmad and F. Amin, "Towards feature subset selection in intrusion detection," in *Proceedings of IEEE 7th Joint International Information Technology and Artificial Intelligence Conference*, 2014, pp. 68-73.
22. K. K. Vardhini and T. Sitamahalakshmi, "Enhanced intrusion detection system using data reduction: An ant colony optimization approach," *International Journal of Applied Engineering Research*, Vol. 12, 2017, pp. 1844-1847.
23. S. Vishwakarma, V. Sharma, and A. Tiwari, "An intrusion detection system using KNN-ACO algorithm," *International Journal of Computer Applications*, Vol. 171, 2017, pp. 11-18.
24. M. Dorigo and G. D. Caro, "Ant colony optimization: a new meta-heuristic," in *Proceedings of Congress on Evolutionary Computation*, 1999, Cat. No. 99TH8406.
25. J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of IEEE International Conference on Neural Networks*, 1995, pp. 1942-1948.
26. Y. Chung and N. Wahi, "A hybrid network intrusion detection system using simplified swarm optimization (SSO)," *Applied Soft Computing*, Vol. 12, 2012, pp. 3014-3022.
27. X. Wu and W. Banzhaf, "The use of computational intelligence in intrusion detection systems: A review," *Applied Soft Computing*, Vol. 10, 2010, pp. 1-35.
28. C. Cortes and V. Vapnik, "Support vector networks," *Machine Learning*, Vol. 20, 1995, pp. 273-279.
29. T. Shon, Y. Kim, C. Lee, and J. Moon, "A machine learning framework for network anomaly detection using SVM and GA," in *Proceedings of the 6th Annual IEEE SMC Information*, 2005.
30. R. Taruna and P. Trikha, "A framework: Intrusion detection in data mining," *International Journal of Research in Computer Engineering and Electronics*, Vol. 2, 2013.
31. M. Panda and M. Patra, "Network intrusion detection using naive bayes," *International Journal of Computer Science and Network Security*, Vol. 7, 2007.
32. M. A. Ambusaidi, X. He, Z. Tan, P. Nanda, L. F. Lu, and U. T. Naga, "A novel feature selection approach for intrusion detection data classification," in *Proceedings of the 3rd IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, 2014, pp. 82-86.

33. Y. Li, J. Xia, S. Zhang, J. Yan, X. Ai, and K. Dai, "An efficient intrusion detection system based on support vector machines and gradually feature removal method," *Journal of Expert Systems with Applications*, Vol. 39, 2012, pp. 424-430.
34. S. Mukkamala and A. H. Sung, "Significant feature selection using computational intelligent techniques for intrusion detection," *Advanced Methods for Knowledge Discovery from Complex Data*, Springer, 2005, pp. 285-306.
35. A. Shenfield, D. Day, and A. Ayesh, "Intelligent intrusion detection systems using artificial neural networks," *ICT Express*, Vol. 18, 2018, No. 30049-3.
36. R. Kaur, M. Sachdeva, and G. Kumar, "Study and comparison of feature selection approaches for intrusion detection," *International Journal of Computer Applications*, No. 2, 2016, pp. 1-7.

Husam Ibrahiem Alsaadi received the M.Sc. degree in Computer Science from Baghdad University, Iraq, 2004. He is currently pursuing the Ph.D. degree in the Department of Electronic and Computer Engineering, Altinbas University. His research interests in data mining, machine learning and data hidden.

Rafah M. Almuttairi currently works as a Director of Studies and Planning Department at University of Babylon, Iraq. She has completed her Ph.D. research in University of Hyderabad, India, 2012. She received her Master's degree in Computer Science from University of Baghdad, Iraq, 2003. She received her B.Sc. degree in Computer Science from University of Babylon, Iraq, 2001. In 2007 she has got a Diploma in Arabic-English translation from Osmaina University, India. She is a Professor. Her current research is interest in fuzzy decision making for grid resources enhancing IDS.

Oguz Bayat received the B.Sc. degree from Istanbul Technique University (ITU), Turkey in 2000. He has got M.Sc. Electrical Engineering from University of Hartford, USA, 2002. He received a doctorate degree in the North-eastern University, Electrical Engineering, USA, 2006. He has published more many papers in different fields. Now he is an Assistant Professor, Doctorate Supervisor and the Dean of the graduate institute, Altinbas University. His main research interests include signal processing, communication, and machine learning

Osman Nuri Ucani received the B.SC., M.SC. and Ph.D. from Technique Istanbul University (TIU), Turkey 1985, 1988 and 1995, in respectively. He has published more than 250 papers in different fields. Now he is a Professor, Doctorate Supervisor and the Dean of the Faculty of Engineering, Altinbas University. His main research interests include biomedical, image processing and machine learning.