Journal of Optimization of Soft Computing (JOSC)

Vol. 1, Issue 2, pp: (39-45), December-2023

Journal homepage: https://sanad.iau.ir/en/Journal/josc

Research paper

An Intrusion Detection System for Network Cyber Security Using Hybrid Feature Selection Algorithms

Zahraa Oday Kamil[1], Golnaz Aghaee Ghazvini[2,*]

1Department of Computer engineering, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

2Department of Computer engineering, Dolatabad Branch, Islamic Azad University, Isfahan, Iran.

Article Info

Article History:
Received: 2024/2/29
Revised:
Accepted: 2024/4/1

Keywords:
intrusion detection; Minimum Redundancy Maximum Relevance; Gray Wolf Optimization; Support Vector Machines.

**Corresponding Author's Email Address: golenaz.aghaee@gmail.com*

Abstract

One of the most important challenges of the expansion of the Internet and virtual space is cyber-attacks. These attacks are becoming new every day and it is becoming more difficult to deal with them. As a result, methods should be used to detect them, which can detect all types of cyber-attacks in the shortest possible time and with proper accuracy. Nowadays, machine learning methods are usually used to detect cyber-attacks. But since the data related to cyber-attacks have many characteristics and are kind of bulky data, as a result, the accuracy of conventional machine learning methods to detect them is usually low. In this research, we have used a hybrid feature selection method to select optimal features from the database related to cyber-attacks, which increases the accuracy of attack detection by classification models. In the proposed feature selection method, first the features that have the least redundancy with each other and at the same time are most related to the category variables (labels) are selected by the MRMR algorithm. Then, using a wrapper feature selection method based on the gray wolf optimization (GWO) algorithm to select a subset of the features selected from the previous step, which maximizes the accuracy of the SVM classifier model, is used this subset has optimal features by which the SVM model is trained. As a result, the accuracy of detecting cyber-attacks by the SVM model increases. According to the simulation results, the average accuracy of the proposed method for detecting cyber-attacks is 99.84%, which has improved compared to the intrusion detection methods of the reference article.

1. Introduction

The military, economy, social sector, and so on now all rely heavily on computer networks as essential tools. It guarantees the communication, coordination, and cooperation between these many industries. Naturally, the number of users rises in tandem with the remarkable improvements of networks. Numerous benefits resulted from the internet's wide accessibility and massive rise in computer network usage. However, the rise in computer hacking has become a serious problem. Various security solutions, including firewalls, antivirus software, internet security tools, and network intrusion detection systems (IDS) are designed to protect computer servers and clients globally against anomalous invasions. On social networks, users who are recognized and/or anonymous may not always have the best of intentions. They are able to take advantage of system and network weaknesses. Additionally, they have access to private or sensitive data that they can read, edit, or destroy. Therefore, in order to prevent prospective assaults on networks, the act of network security has become increasingly important [1].

IDS has been divided into two groups: anomaly behavior detection and signature-based anomaly detection. These two kinds differ from one another in their patterns. The signature-based intrusions periodically scan the network and attempt to match certain specified patterns there. On the other hand, anomalous network intrusion-based systems present typical traffic patterns and search for packets that are similar enough to be identified as intrusions [2].

As stated in [3], anomaly detection, sometimes referred to as undesired persuaded defects, assaults, or faults, is the essential component of intrusion detection.

By dividing packets into two categories—attacks and normal—IDS aims to stop intrusions from occurring in subsequent network transactions. Computational speed and comparison accuracy are two of the most crucial factors in the IDS. Given the abundance of characteristics in every transaction, an appropriate technique is needed to identify both the incursions and the complete feature set by deriving an effective subset of features.

Finding a subset of features (M) from an original set (N) where M<N is the process of feature extraction. Cropping features with more relevant information and those that include data about other features is the aim of feature selection. Moreover, the process time is being decreased with minimal accuracy loss by removing some of the unnecessary or redundant elements. In order to accomplish this goal, a few data mining techniques have been presented and used thus far for feature extraction [4].

Two techniques for feature selection in IDS are wrapper-based and filter-based methods. The wrapper-based strategy leverages machine learning techniques to obtain dependable intrusion detection features. Whereas the filter approach, which minimizes and effectively uses the features that are redundant or irrelevant, never makes use of machine learning [4].

The following describes the current study's structure: An overview of the literature on the application of wrapper and filter-based methods for creating effective NIDS is given in Section 2. A synopsis of the suggested model is provided in Section 3. The findings and discussion are presented in Section 4. Regarding Section 5, it offers the findings.

2. Related works

The Intrusion Detection System has been the subject of numerous prior studies that have been published in the literature. The intrusion detection system's abstract model was first put forth by Denning D.E. [5] in 1987. Intrusion detection is the first security defensive method for the computer system used in this paper. The concept is not dependent on any particular application environment, operating system, or type of intrusion or vulnerability in the system. It is a framework that can serve as a great illustration of how to create application systems for intrusion detection. Though other unidentified reasons that are not anomalous behaviors can also trigger the audit criteria in the proposed model. Furthermore, it remains to be demonstrated whether the approach can identify the greatest intrusion before significant harm is caused. In addition to focusing primarily on anomaly detection using data mining, Wu et al. [6] apply association rules in a forward implementation based on Trie trees. Aumreesh et al.'s review [7] highlights the several kinds of intrusion detection systems, including host-based, network-based, misuse-based, anomaly-based, and hybrid-based systems. It primarily concentrates on agent-based, behavior-based, and anomaly-based technologies in real network traffic. The advantages and disadvantages of the abuse detection strategy and the anomaly detection approach are compared by S. Northcutt et al. [8]. The disadvantage of the anomaly detection approach, as the author notes, is that the Intrusion Detection System may trigger a false positive alarm when it detects a new behavior for the first time. A review of using Machine Learning (ML) technologies in the Intrusion Detection System is provided by L. Haripriya and M.A. Jabbar [9].
Additionally, the rates of false negatives, false positives, and anomaly detection are comparatively considerably higher than the rates of misuse detection. Additionally, they go over how to apply machine learning to a system and provide a thorough comparison of different methods for an intrusion detection system that use machine learning. According to this paper, it can be challenging to train machine learning models when there is a shortage or unavailability of traffic data. Basant Subba et al. [10] provide an effective Artificial Neural Network (ANN) model for an intrusion detection system. A constraint of their methodology is that the suggested model necessitates an extensive training duration. Nevertheless, the inability to add more agents to the prior one won't affect the neural network's overall detection performance. The most popular

feature selection algorithm, Filter and wrapper, is described by Pan-Shi Tang et al. [11] in their work. The Genetic Algorithm-based selection method is also applied to a combination of two algorithms, and the results show that GA is significantly more efficient in picking features than the Filter and Wrapper algorithms. S. Aksoy et al. [12] and B. Kavitha et al. [13] provide a crucial technique for employing the Genetic Algorithm to choose the necessary subset of features. They think feature selection can eliminate unnecessary elements and significantly impact the development of an effective classification system in subsequent stages. Ketan Sanjay Desale and Roshani Ade [14] present a novel approach to feature selection utilizing a genetic algorithm and the mathematical intersection principle. Additionally, a variety of feature selection methods are examined, including IG, CAE, and CFS. Their results with J48 and Naive Bayes (NB), the other two commonly used classifiers, are compared. These articles provide a nice illustration of how to use a genetic algorithm to choose features.

3. Proposed method

Our goal in this work is to provide a high-precision system in order to increase cyber security in the network by using feature selection hybrid algorithms. As a result, in this work, we will use two algorithms - MRMR and Gray Wolf Optimization (GWO) in order to select an optimal subset of features. In this research, first, by using the MRMR algorithm, which uses the concepts of maximum similarity and minimum redundancy, the features with the least redundancy are selected. Then, with the help of the gray wolf optimization algorithm, an optimal subset of the features selected by the MRMR algorithm will be selected. The gray wolf optimization algorithm, which is based on the instinctive behavior of gray wolves, is a form of meta-heuristic algorithm with a hierarchical structure that is inspired by the hunting activity of gray wolves. This population-based approach is simple in its operation and can be easily extended to situations with different dimensions. Also, after choosing the optimal features, these features are classified by the support vector machine algorithm and all types of cyber threats are detected. With the help of mapping the feature space to the dimension with higher resolution, the support vector machine has a large capacity to recognize patterns and thus recognize attack patterns. This enables the SVM algorithm to accurately detect network intrusions. The diagram of the proposed method is shown in Figure (1).

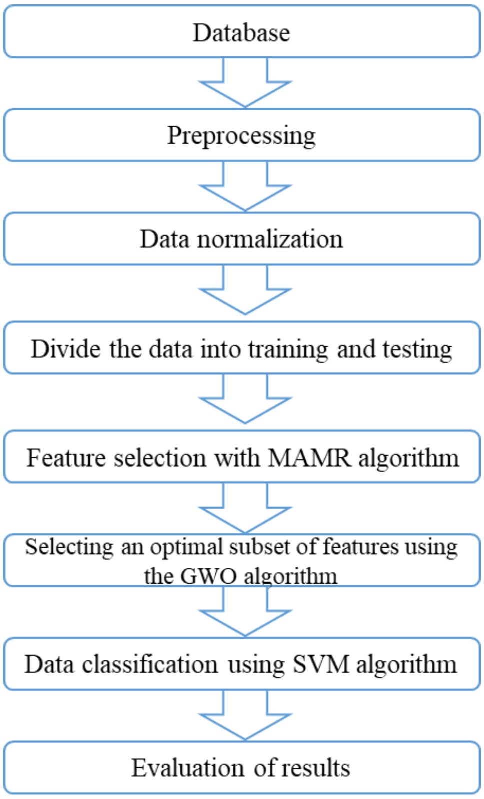
Database
Preprocessing
Data normalization
Divide the data into training and testing
Feature selection with MAMR algorithm
Selecting an optimal subset of features using the GWO algorithm
Data classification using SVM algorithm
Evaluation of results

Figure 1: Diagram of the proposed method

4. Steps of the proposed method

According to the diagram of the proposed method, the basic steps in the proposed method for detecting different types of cyber-attacks include pre-processing and dividing the data, selecting the combined feature and classifying the selected features. These steps are explained below.

4.1. Data preprocessing

First, the KDD database used in this research is uploaded to the software, and then the first step is to apply the necessary pre-processing on the data. Some of the necessary pre-processing are outlier data removal, missing data removal and data normalization. In this research, the Max-Min technique was used to normalize the data. Data normalization ensures that all features are in a certain range (for example -1 and 1) and as a result, the effect of all features in the classification is the same. The normalization of the Max-Min method is defined by (equation 1).

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}}(H - L) + L. \quad i = 1.2 \dots . N \qquad (1)$$

In equation (1), X_i is the actual value of a feature for the network input and $X_n$is the normalized expression of X_i. X_{min} and

X_{max}represent the minimum and maximum X values. The lower and upper limits of normalization are also indicated by L and H and have values of -1 and +1, respectively.

4.2. Dividing the data into two groups of training and testing

Title must be in 15 pt Regular font. Author name Since in the proposed method, a supervised machine learning model is used for data classification, as a result, training and test data are needed to train and evaluate the performance of the proposed model. In this research, we have divided the database samples into two groups of training and testing with a ratio of 66 to 34. The stages of feature selection and SVM model training are first completed by the training samples and then the proposed model is evaluated on the test samples.

4.3. feature selection by MRMR-GWO method

As stated before, the basis of the proposed method in this research is a combined feature selection technique, which improves the performance of the machine learning model in detecting cyber-attacks. Feature selection methods are classified into two general categories, the first category is filter methods and the second category is wrapper methods. The methods that perform feature selection without considering the classification model and only by considering the features and labels are filter methods and in contrast to the methods that select the features by considering They carefully choose the classification model; they are Wrapper methods. In this research, we have used the MRMR algorithm as a filter method and the GWO algorithm as a wrapper method. The diagram of feature selection method in this research is shown in figure (2).

In the proposed feature selection method, the features that are most related to the category variable (label) and at the same time have the least redundancy between the features are selected. In fact, in the MRMR algorithm, the relationship between features with data labels and the redundancy between features are calculated based on mutual information and according to relations (2) and (3).

$$W_I = \frac{1}{|s|^2}\sum_{i,j\in s} I(i,j) \quad (2)$$

$$V_I = \frac{1}{|s|}\sum_{i\in s} I(h,i) \quad (3)$$

In the above relations, S is a specific set of attributes and h is a class variable. Also, W_I is the redundancy between features and V_I is the relation of S with class variable h. Mutual information between two variables $I(x,y)$ is also calculated by the following equation:

$$I(x,y) = \sum_{i,j} \log\frac{p(x_i,y_j)}{p(x_i)p(y_j)} \quad (4)$$

As a result, first all the features are ranked by the MRMR algorithm and a number of those that are most related to the category variable and have the least redundancy among themselves are selected.

In the following, the relationship between the features selected by the MRMR algorithm and the classifier model, which is the SVM algorithm in this research, is examined by a Wrapper method, and the best features that lead to the highest accuracy in the SVM algorithm are selected. In this research, GWO meta-heuristic algorithm has been used to investigate the relationship between features and the classifier model.

The gray wolf algorithm works to select the final feature by first defining a feature space with the dimensions of the features selected from the MRMR algorithm and generating feature vectors with specific dimensions where each wolf is a representative of a feature vector. The wolves move in the feature space and their information is updated according to the objective function, and the quality of each feature vector is evaluated. In this technique, the objective function in the gray wolf algorithm is the accuracy of the classification model i.e. SVM. As a result, the extraction of the features that lead to the maximum of the objective function or the accuracy of the SVM algorithm is selected as the optimal feature vector.

Briefly, in the gray wolf algorithm, it starts from a random population of wolves with random feature vectors, and then these wolves move randomly in the feature space and update the improved feature vectors. In this way, the alpha wolf keeps the best feature vector and the other wolves replace the alpha wolf if the feature vector improves. This algorithm is used for feature selection in various problems due to its fast convergence capability and no need for many parameters.

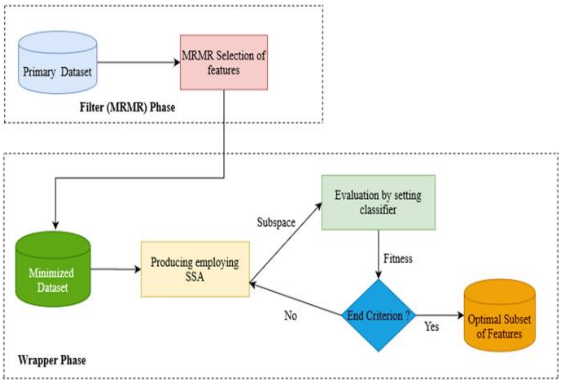

Figure 2: Diagram of feature extraction method.

4.4. Classification using SVM algorithm

In this research, the classifier model, which the wrapper feature selection technique is also performed according to, is the SVM algorithm. SVM is an effective method to build a classifier. Its purpose is to create a decision boundary between two classes to allow prediction of labels from one or more vectors. This decision boundary is called a hyperplane and is oriented so that it is farthest from the nearest data points of each class. These closest points are called support vectors. Assuming we have a labeled estimator dataset:

$$(x_1, y_1 \ldots \ldots (x_n \cdot y_n) \cdot x_i \in R^d \text{ and } y_i \in (-1.+1) \tag{5}$$

where x_1 represents the feature vector and y_i is the class label (negative or positive) of estimator combination i. Therefore, the desired hyperplane is defined as follows.

$$wx^T + b = 0 \tag{6}$$

where w is the weight vector, x is the input feature vector, and b is the orientation. w and b satisfy all the following inequalities for all components of the estimator set:

$$wx_i^T + b \geq +1 \text{ if } y_i = 1 wx_i^T + b \leq -1 \text{ if } y_i = -1 \tag{7}$$

The purpose of SVM model estimation is to find w and b so that it separates the data super plane and maximizes the boundary $1/||w||2$. Hence, vectors x_i with $(wx^T + b) = 1$ y_i are called support vectors.

4.5. Evaluation of the proposed model

After selecting the best features and training the network using training data, the last step is dedicated to testing the proposed network. This step is done with the help of test data that do not have output labels. In this step, the test data is entered into the classifier model and the accuracy of the trained network is evaluated to detect different types of cyber-attacks on this data.

5. Simulation and Results

In this research, the two-class NSL-KDD dataset has been used. This database is the updated version of the KDDCup99 database that was presented in 2009 in order to categorize different cyber-attacks. NSL-KDD bank consists of data related to normal mode and attack mode. This data bank is registered in two different groups, consisting of KDDTrain+, KDDTest+, and KDDTrain set is used for network training, and KDDTest set is used to evaluate the proposed technique. In order to evaluate the results of the proposed method, the following criteria are used.

$$Accuracy(acc) = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Precision = TP(TP + FP) \tag{9}$$

$$Recall = TP/(TP + FN) \tag{10}$$

$$F1\ score = \frac{2 * (Recall * Precision)}{(Recall + Precision)} \tag{11}$$

In the above relations, TP represents the number of correct identification of attacks, TN is the number of correct identifications of normal traffic, FP is the number of false identification of attacks, and FN is the number of false identifications of normal traffic.

5.1. Simulation settings

As stated in the proposed method chapter, we have used a two-stage hybrid feature selection technique. The NSL-KDD database has 41 features. In the simulation, we first ranked the features by the MRMR algorithm and selected the top 20 features. Then, in the next step, 8 optimal features from the 20 initially selected features have been selected by the gray wolf algorithm and entered into the SVM algorithm for classification. Also, in the simulation, we have considered 10,000 samples for training the SVM algorithm and 5,000 samples for testing the proposed model. Finally, Table (1) presents the parameters of the gray wolf algorithm for feature selection using the wrapper method.

Table 1: Basic parameters of the gray wolf algorithm

the amount of	parameter
50	The number of repetitions
20	population size
[20-1]	The allowed interval for changing variables
accuracy	The objective function

5.2. Simulation results

In this section, the results for the test data are presented. As stated earlier, the labels of this part of the data have not been seen before by the proposed algorithm, and the proposed method should predict the label of each sample from the optimal features selected by the MRMR-GWO method. Cyber-attack or normal traffic. Figure (3) shows the confusion matrix on the test data. As can be seen, there are 2145 samples in the first row of the confusion matrix, which are samples related to cyber-attacks, and in the second row, there are 2855 samples, which are samples related to normal traffic. In both groups, only 2 samples were misdiagnosed. The overall accuracy of the proposed method for detecting examples of cyber-attacks and normal traffic is 99.9%.

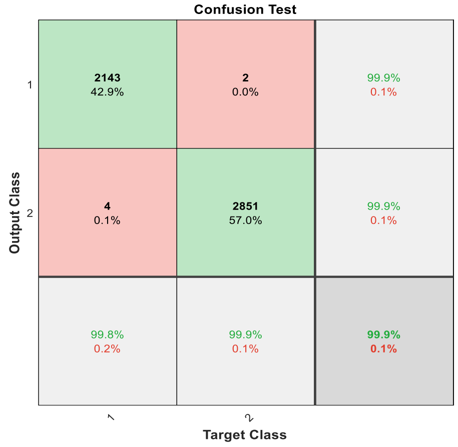

Figure 3: Confusion matrix for test data

Figure (4) also shows the diagram related to the numerical values of the evaluation criteria. The exact values of the criteria of precision, accuracy, recall and F-score are equal to 99.88% according to figure (4). These results are for one simulation run.

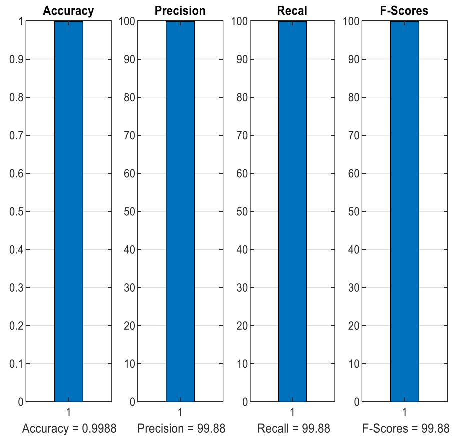

Figure 4: Numerical values of evaluation criteria on test data

Also, the ROC curve related to the test data can be seen in Figure (5). In this graph, the true positive rate (TPR) is plotted against the false positive rate (FPR) at different classification thresholds. In the ROC diagram, the closer the line of a class is to the intentional axis and the farther it is from the minor diameter, the more accurate it is in recognizing the samples of that class. In figure (5), it is clear that the accuracy of detecting samples of both classes is equal.

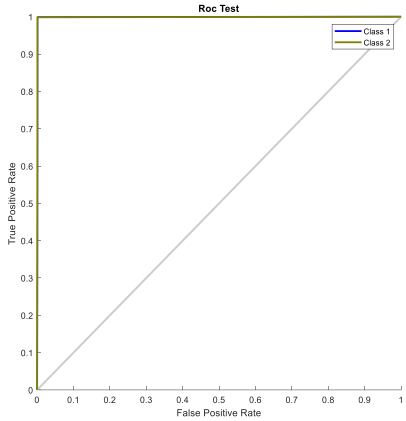

Figure 5: ROC curve on test data

5.2.1. Comparison of results

Finally, in this part, we have compared the results of the proposed method for detecting cyber-attacks with other methods. The simulation results are presented in terms of intrusion detection accuracy, and the important point is that the results of the proposed method are the result of the average of 40 simulation program executions. As it is known, the proposed method has an average accuracy of 99.84%, which has improved compared to other methods in the reference article.

Table 2: Comparison of the proposed method with other methods presented in the reference article [100]

Method	accuracy
Bayesian network	85.76
J48	96.43
SMO	95.99
Reference Article Method (NB-GA) [۱۰۰]	99.73
suggested method (MRMR-Bagging)	99.84

Based on the simulation results, the average accuracy of the proposed model for detecting cyber-attacks is 99.84%, which is higher than the compared methods. This high accuracy in the proposed method is due to the two-stage feature selection method, which makes the classifier model to be trained with the best features and the relationship between the features with the classification variable in the best way for The SVM model is specified.

6. Conclusion

In this research, a method based on machine learning to detect cyber-attacks is presented. A combined feature selection technique is used in the proposed method. This technique has the ability to select the best features, which leads to increase the accuracy of the classifier model, among all the features of the database. In this proposed method, we first rank the features of the NSL-KDD database by the MRMR algorithm and select 20 of the features that are most related to tags and at the same time have the least redundancy with other features. have, we choose. Then, among the 20 selected features, we select a subset of 8 features that maximize the accuracy of the SVM algorithm as optimal features. Then the training samples with these features are entered into the SVM algorithm and the learning process is completed.

Reference

[1] Folorunso, O., O.O. Akande, A.O. Ogunde and O.R. Vincent 2010. ID-SOMGA: A self organising migrating genetic algorithm-based solution for intrusion detection. Comput. Inform. Sci., 3: 80-92.

[2] Trair, D., W. Ma, D. Sharma and T. Nguyen, 2007. Fuzzy vector quantization for network intrusion detection. Proceedings of the IEEE International Conference on Granular Computing, Nov. 2-4, IEEE Xplore Press, Fremont, CA., pp: 566-566. DOI: 10.1109/GrC.2007.124

[3] Lazarevic, A., L. Ertoz, V. Kumar, A. Ozgur and J. Srivastava, 2003. A comparative study of anomaly detection schemes in network intrusion detection. Proceedings of the 3rd SIAM International Conference on Data Mining, (CDM’ 03), SIAM.

[4] 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 2nd SKLOIS conference on Information Security and Cryptology, Nov. 29-Dec. 1, Springer Berlin Heidelberg, Beijing, China, pp: 153-167.

[5] D. E. Denning, “An intrusion detection model,” IEEE Transactions on Software Engineering, vol. 13, no. 2, pp. 222-232, 1987.

[6] W. Gongxing and H. Yimin, “Design of a new intrusion detection system based on database,” in Proc. 2009 International Conference on Signal Processing Systems, 2009, pp. 814-817.

[7] A. K. Saxena, S. Sinha, and P. Shukla, “General study of intrusion detection system and survey of agent-based intrusion detection system,” in Proc. 2017 International Conference on Computing, Communication and Automation (ICCCA), 2017, pp. 421-471.

[8] S. Northcutt and J. Novak, “Network intrusion detection,” IEEE Network, vol. 8, no. 3, pp. 26-41, 2003.

[9] L. Haripriya and M. A. Jabbar, “Role of machine learning in intrusion detection system: Review,” in Proc. 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 925-929.

[10] M. B. Subba, S. Biswas, and S. Karmakar, “A neural network-based system for intrusion detection and attack classification,” in Proc. 2016 Twenty Second National Conference on Communication (NCC), 2016, pp. 1-6.

[11] P. S. Tang, X. L. Tang, and Z. Y. Tao, “Research on feature selection algorithm based on mutual information and genetic algorithm,” in Proc. 2014 11th International Computer Conference on Wavelet Active Media Technology and Information Processing, 2014.

[12] S. Aksoy, “Feature reduction and selection,” Department of Computer Engineering, Bilkent University, 2008.

[13] B. Kavitha, S. Karthikeyan, and B. Chitra, “Efficient intrusion detection with reduced dimension using data mining classification methods and their performance comparison,” in Proc. International Conference on Business Administration and Information Processing, 2010, pp. 96-101.

[14] K. S. Desale and R. Ade, “Genetic algorithm-based feature selection approach for effective intrusion detection system,” in Proc. 2015 International Conference on Computer Communication and Informatics (ICCCI), 2015, pp. 1-6.