IJIE

http://penerbit.uthm.edu.my/ojs/index.php/ijie
ISSN : 2229-838X e-ISSN : 2600-7916

The International Journal of Integrated Engineering

A Mechatronics System based on Feature Selection and AI for IoT Intrusion Detection Applications

Montassar Aidi Sharif[1*]

[1]Electronic and Control Engineering Department, Technical Engineering College/Kirkuk, Northern Technical University, IRAQ

Corresponding Author*

DOI: https://doi.org/10.30880/ijie.2022.14.06.026
Received 30 April 2022; Accepted 13 June 2022; Available online 10 November 2022

Abstract: In today's rapid development of the Internet, people's daily lives have become easier, but on the other hand, people's privacy is also faced with potential threats if necessary, security measures are not taken. To detect or stop cyberattacks in this area, network intrusion detection systems (IDS) can be equipped with machine learning algorithms to improve accuracy and speed. Recent research on intrusion and anomaly detection has shown that machine learning (ML) algorithms are widely used to detect malicious web traffic, using neural networks to learn models to visualize the sequence of connections between computers on a network. By analyzing and selecting the correct features, dense attacks can be detected more accurately, ultimately reducing misclassification rates and improving accuracy. In this study, we propose a Teacher-Student Feature Selection (TSFS) method that first uses the Isomap method to extract and select features in low dimensions and the best display, and then performs classification. The artificial neural MLP-Net for classification is used to minimize diagnostic errors. Although the teacher-student scheme is not new, to our knowledge, this is the first time this scheme has been used to select features in an intruder alert system. The proposed method can be used to select monitored and unmonitored features. The method is evaluated on different datasets and compared with the state-of-the-art feature selection methods available. The results show that the method performs better in classification, clustering, and error detection. Furthermore, experimental evaluations show that the method is less sensitive to parameter selection.

Keywords: Detections system, Internet of Things (IoT), MLP neural network

1. Introduction

Data mining uses various methods and algorithms to extract patterns from stored data. Researchers analyze these algorithms in different fields and describe future research paths. Data mining is a new model that has been successfully applied in various fields due to the importance of knowledge in decision making. The size of the data is increasing in different applications, for example, the image captured by a typical camera has nearly one million features. Because direct processing of high-dimensional data is very expensive and often has redundant or correlated features. Therefore, the intrinsic dimension of the data is usually much smaller than the original feature space. Therefore, dimensions have several advantages:

1) Reduced data storage.
2) simplify machine learning models that lead to increased generalizability.
3) Reduce the computational complexity of training and testing machine learning models.

Broadly speaking, downsizing falls into two categories: feature selection and feature extraction. Feature extraction methods attempt to find linear or nonlinear mappings that can map basic dimensional data to low-dimensional subspaces.

*Corresponding author: msharif@ntu.edu.iq

These methods all use attributes to get the best performance. The resulting space is not necessarily important or difficult to interpret. On the other hand, the feature selection method selects a subset of high-quality main features based on certain criteria, retains the main features, and does not change the dimensionality reduction results of the main feature evaluation. In this way, they can obtain meaningful information that experts in the field can interpret. Another benefit of feature selection is that you only need to collect data after selecting the following attributes. The attributes used for collection or calculation can be useful because sometimes measuring all features is expensive or almost impossible [1]. The structures of the Internet of Things (IoT) are networks of homogeneously associated devices which transfer information among themselves to offer smarter services to users. Today, the Internet of Things has seen more innovative work in different fields of application. Smart home appliances and foundations, smart security and surveillance, smart street traffic, and the management and crisis reaction systems that are related to health are only a few instances of IoT. The Internet of Things depends on wireless technologies and is used in many places where there are devices joined to the net, and smart services like a smart house, smart cities, smart health, etc. are provided. Therefore, the collection of raw data will be done directly from the user, and this will increase the threat of privacy violations. Some IoT devices interact with people's daily lives. If bad people (hackers) use these devices (like smart locks, smart lights, and built-in cameras, etc.), they can cause major privacy problems [2].

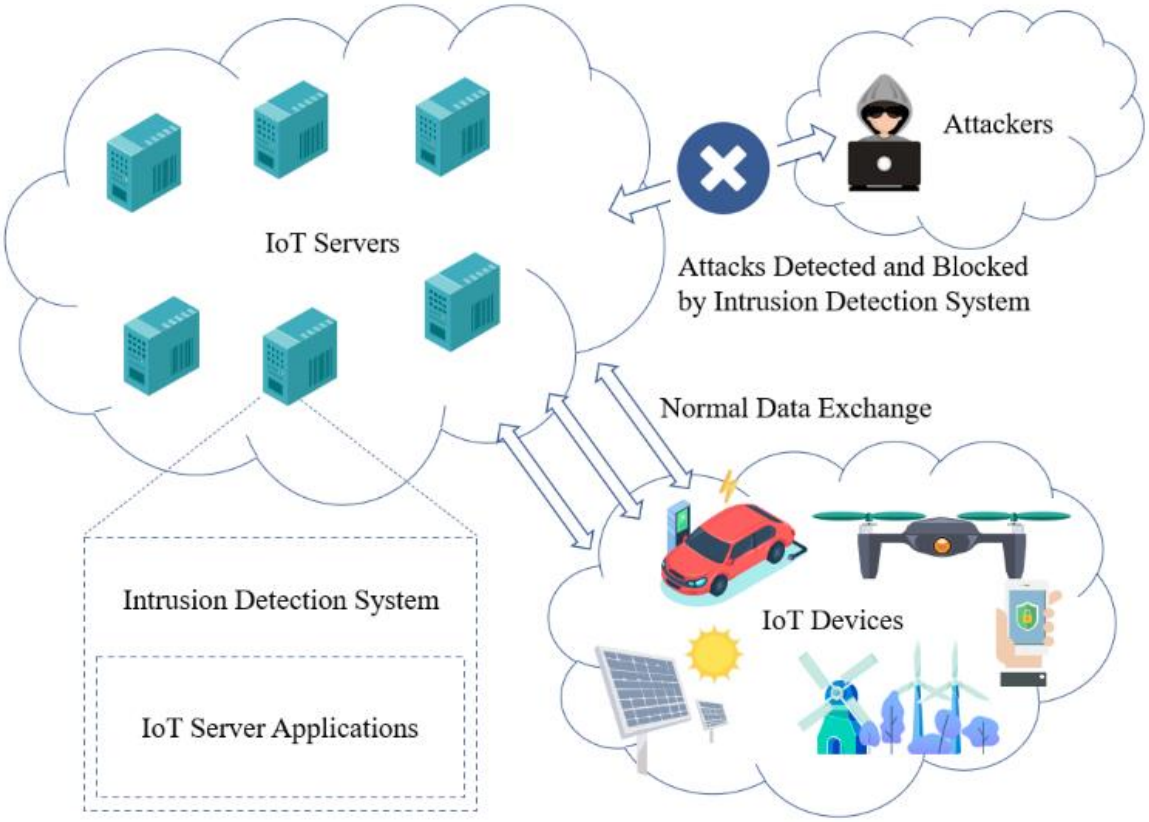

Fig. 1 - The proposed electronic system

Artificial neural networks are algorithms that can easily mimic and solve complex problems in animal brain processes. Perceptron neural networks consist of artificial neurons or nodes. The processing unit is the information associated with the hierarchy of these synaptic weights (joints). Neurons can filter and transfer information in a monitored manner, creating predictive models that classify data stored in memory.

The proposed system (Fig. 1) in this paper tries to cover three challenges and gaps in the methods of detecting the penetration of network attacks. Therefore, the importance and necessity of this research will be threefold.

- Reducing Computational Complexity
- Increase generalization power
- Reduce dimensions and select features in order to maintain the proximity of data to each other

The intrusion detection system has various stages such as classification, preprocessing, data collection, and feature selection (FS), which use different methods in each phase. Feature selection focuses on selecting groups of variables from the input (input) that can effectively build up the input data, reduce the influence of noise-independent variables, and provide excellent predictive results. In general, the dependent variable acts as chimes because it does not provide any additional information about the class. This means that the content of any piece of information can be obtained from

some unique features, including the maximum information about the rank. Therefore, by removing the dependent variable, one could minimize the amount of data and improve classification performances.

Classification is a branch of supervised learning because the sample dataset is used to learn the group structure, just like a teacher supervises his students for a specific purpose. While the problem of clustering is to identify groups The problem of classification is to learn the structure of a sample dataset that has been divided into distinct groups called categories or classes. Learning these categories is usually done through models. The model is used to estimate group IDs (or class labels) from one or more previously unseen data samples with unknown labels. So the input to a classification problem is a dataset of samples that has been divided into different classes. This data is called instructions, and the group identifiers for these classes are called class labels. In the majority of cases, class labels have a distinct semantic significance in the context of a certain application, such as a group of clients interested in a specific product or a group of data items with the required quality. The training model is the model that has been trained. Test datasets are a collection of unseen data pieces that need to be categorized. The learner refers to the algorithm which provides the training pattern for prediction.

The majority of classification algorithms contain two steps:

1) **Training stage**: an instructive demonstrate is made from preparing tests. Instinctively, this step can be caught on as a scientific demonstrate summarizing the tag bunches within the preparing information set.
2) **Test stage**: is utilized to find the class labels (or group IDs) of test samples.

In this paper, after examining the basic concepts and background issues in the introduction, in the second part, a summary of the DDoS (Distributed Denial of Service) attacks is discussed. After that, the related articles are surveyed and reviewed. Then, In the third part of the paper, the proposed method is evaluated by MATLAB simulator. While the method is used and the various simulation parameters are briefly expressed and the proposed design in the fourth section is evaluated from different aspects and we present the experimental evaluations of the proposed system. Finally, in the fifth section, the conclusion and summary of the article and suggestions for future research are stated.

2. Related Work

DDoS represents an essential web attacks, and the goal of the enemy is to destroy the targeted service for legitimate users [3]. The high frequency of the method is the important reason in the number of ways such attacks have been generated and performed. In fact, you can create DDoS attacks at any level of the OSI communication model. Therefore, the attacker found several ways to successfully deny service to the target victim.

In fact, there are many types of DDoS protection in traditional systems, even advanced intrusion detection and defense systems. Similarly, DDoS protection by AI and machine learning is also attracting the attention of researchers, following the achievement of machine learning algorithms and the success of their implementation to solve important problems in various fields. The demand for machine learning in this field makes DDoS attacks more complex and deceptive than ever. It is easy to hide less information that identifies the new attack as the actual state of the packet. For example, a technique called fast flux can be used to spoof or change IP addresses on the same stream at high speed [4].

In general, the purpose of a DDoS attack is to use multiple, distributed sources to reduce the service of the target. A common instant of this kind of attack is a flood attack that overwhelms the network traffic could be sent to the victim. The main reason behind DDoS attack is based on the fact that large amounts of resources are being used in multiple locations targeting victims. Botnets are often used to launch DDoS attacks because they can typically use many dangerous hosts (also known as zombies). Prevention of DDoS attacks is a main focus of the researchers [5]. Because DDoS attacks are humorous, we are going to concentrate on developing specialized solutions that can professionally prevent and detect the attacks of DDoS. The research revealed the following effective DDoS attack mechanisms [5]:

1) DDoS mechanisms must not interfere with the legitimate activities of users.
2) DDoS mechanisms must be able to prevent attacks from the network and outside the network.
3) Mechanisms must meet the performance and scalability required for modern data centers.
4) DDoS protection mechanisms should be able to be introduced as cheaply as additional hardware. Also, scaling doesn't seem to cause big changes such as increased overhead.
5) Mechanisms should be strong, adaptable and flexible.
6) Low (false) detection rate and high detection rate are also desirable functions of efficient DDoS protection mechanism.

The aforementioned requirements should only be used as a starting point for developing a DDoS protection system.

Based on the network context in which it is deployed and the level of protection necessary, the intensity of each requirement might vary. Nevertheless, there is no foolproof method for preventing DDoS attacks.

The increasing complexity of DDoS attacks makes developing a faultless DDoS attack protection solution difficult. In order to detect the invasion of the phenomenon, Aldwiri et al. In paper [6] used an artificial bee colony (ABC) to search and compare.

Nancy and others in paper [7] suggested a framework for collection and identification of functions. And in the choice of functions, Nancy et al. developed a new paradigm called the DRFSA (Dynamic Return Algorithm Selection Function).

Both packaging and filtering approaches take advantage of this model. A very smart decision tree for identification has grown by extending the standard selection tree technique with time guides and fuzzy guides.

Eduardo de la Hose et al. in paper [8] have proposed a classification method for detecting network anomalies. By choosing feature selection, they used the Fishers differentiation and Principal Components Analysis (PCA). Network transaction can be then classified as an abnormal or normal contingency self-organizing maps and noise cancellation.

Sunil Nilkanth Pawar et al. in paper [9] they used variable length chromosomes (vlc) for developing the Intrusion Detection System (IDS) in the Genetic algorithm -based network is proposed. Used to produce fewer chromosome bases with corresponding specifications. To define the fitness of each law, an efficient fitness function is used. Each chromosome would have one or more rules in it at least.

In the study [10], they use some classification tools in the form of Discovery to evaluate the big data tools they are about to discover. The function of the detection method is the classification whether the input of the network traffic is normal according to each feature that defines each network mode. The author concludes that the random classifier gives the excellent results in terms of execution time and accuracy. The benefit of this work is to solve the problem of low accuracy and prediction time in IDS. The comprehensive performance has been rated according to the diagnosis construction time, accuracy, and prediction time.

3. The Proposed System

The student-teacher method is used and explained here. An attempt is made to first reduce the specificity of the input data by using one of the dimensional reduction methods, which includes extracting and selecting the feature in the teacher-student section. The classification section will then attempt to learn a space similar to the learned space using a neural network, which is a multi-layered perceptron (MLP).

The three steps of the proposed method are as follows:

Teacher stage: The instructor is often a state-of-the-art approach that attempts to get the great show of facts in small dimensions. Typically, the instructor for one-of-a-kind programs has to be one of a kind in line with the wishes of the hassle and may be supervised or unsupervised. For example, a characteristic extraction approach which include Isomap may be used as an instructor approach. Extraction is executed through this set of rules withinside the instructor segment to keep the most variance of the characteristic values.

Student stage: To select a feature, it is necessary to have a method that can select important features in the input layer with appropriate performance. The PCA method is an algorithm for reducing the size (feature space) of a data set. At the end of this step, the attributes are ranked based on the attribute score and then the attributes with the highest score are selected.

Classification stage: In the second one stage, after the low-dimensional codes are normalized and the scholar community is skilled primarily based totally on those low-dimensional codes. Next, an MLP multilayer neural community technique is used for classification.

- **Extracting and selecting features:**

Usually, real-global records are excessive-dimensional in nature and really hard to recognize and examine. There are some of recognized dimensional strategies that try and clear up this hassle and permit customers to higher examine or visualize complicated records sets. The fundamental factor analysis (PCA) approach has been broadly utilized in records mining to have a look at the shape of records. In PCA, new orthogonal variables (latent variables or fundamental components) are acquired with the aid of using maximizing the variance of the records. Principal Component Analysis (PCA) is a set of rules for lowering the scale of a records set from n top wherein p approach does now no longer lessen the precise properties, however reduces the scale with the aid of using extracting the properties. Isomap approach is primarily based totally at the concept of looking at the hassle of making a deformation from excessive to low dimensions as a graph hassle. The Isomap set of rules extends the classical strategies of fundamental factor analysis (PCA) and multidimensional scaling (MDS) to a category of nonlinear manifolds. The entire isometric characteristic mapping set of rules has 3 steps:

1) Estimating neighborhood graph.
2) Calculate the shortest route diagram according to the neighborhood diagram.
3) Construction of the lowest embedded dimensions.

In classification algorithms, the number one records set is split into units of schooling records and the experimental records set. Multilayer perceptrons are a category of synthetic Feedforward neural networks.

In type algorithms, the number one records set is split into schooling records units and an experimental records set. The version is constructed the use of the schooling records set and the experimental records set is used to validate and calculate the accuracy of the version. there are not unusual place neural community architectures: CNNs (convolutional neural networks) and RNNs (recurrent neural networks). CNNs are used to discover visible styles at once from variable

pixel images. MLPs are designed to discover time collection styles fashioned with the aid of using audio / speech symbols or shapes. Both CNN and MLP are unique varieties of multilayer neural networks. They are skilled with the aid of using the republishing algorithm.

Multilayer perceptron can be defined as a category of synthetic Feedforward neural-network. An MLP includes at the least 3 node-layer: the first layer (input layer), a hidden layer, and the last layer which can be named as an output layer. With the exception of enter nodes, every node is a neuron that makes use of a nonlinear activation function. MLP makes use of a supervised studying method referred to as payback for training. Its more than one layers and nonlinear activation distinguish MLP from a linear perceptron. It can genuinely distinguish records that isn't linearly separable.

Explanation of the teacher-student method:

Knowledge distillation (KD) is a special type of knowledge transfer developed in recent years. The basic idea of KD consists of two network structures called teachers and students. It is used to train the student network to make predictions of the same teacher. The teacher is a strong role model, and the student is a simple role model. The most common goals considered in the TS model include designing simple models, designing models with limited data access, requiring short-term training, and designing models with higher accuracy.

The proposed method, consists of 3 different steps. Each of these steps is described in Figure (1).

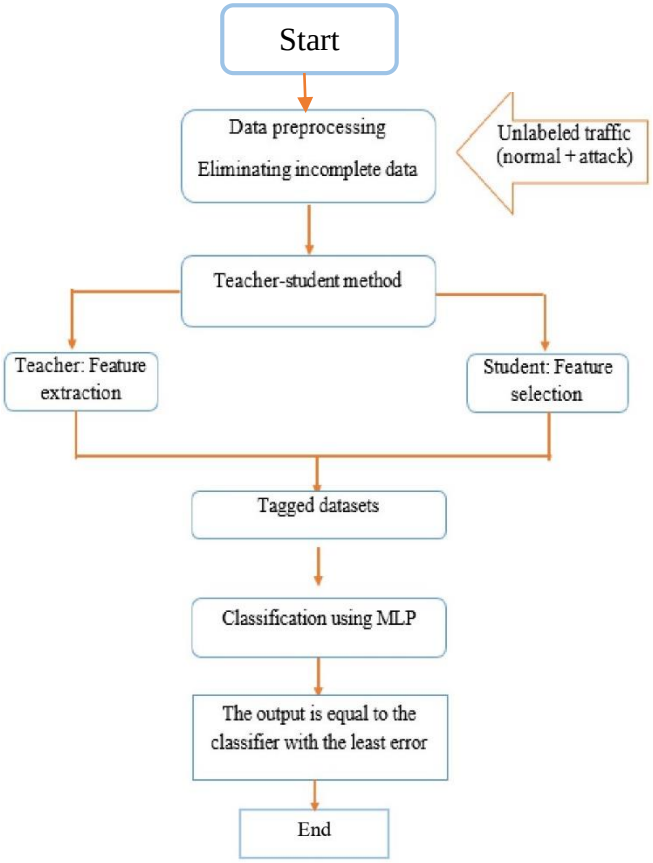

Fig. 2 - Flowchart of the method

The steps of the proposed method are described below:

- step 1 - data preprocessing: In this step, the data is refined. In fact, this step is data preprocessing. This two-step involves deleting incomplete data as well as normalizing the data. Incomplete data in the data set is always inevitable. This can be due to human error, system failure, data transfer problems or other issues. But incomplete data can have bad consequences for the system. This problem becomes more acute when the system work based on machine learning. Because in such systems, the accuracy of the information is assumed and any defect in the received data can lead to incorrect output of the system. In the proposed method, the data received from the input is checked and the existence of any incomplete data is detected, and this data is destroyed. But the second part also includes data normalization. The data used in any numerical algorithm can be either binary or continuous. Therefore, in preprocessing, the main problem is the conversion of the data to binary form or to normalized

continuous form. To convert numbers and to normalize attributes, first a statistical analysis is performed on each attribute based on the data in the data set and a maximum value is set for the values of each feature. Then, based on the equation (1), the numbers are converted to normalized form.

$$\text{normal data}_i = \frac{(\text{data}_i - \text{mindata})}{(\text{maxdata} - \min \text{data})} \quad (1)$$

After processing the data and converting the numbers to the normal form, the data is ready to use.

- Step 2- Teacher-Student Method: Extract and feature selection: In this step, the preprocessed data is entered and the information of different sessions is extracted and selected using their features. To do this, we used the Isomap algorithm. For this purpose, the received data is in the form of a matrix set of features and this data is reduced in size by entering the algorithm. The main challenge is to downsize of the data structure protection and make the transformation with minimal data loss. The main purpose of Isomap is to discover a suboptimal space and maintain geodetic distance between data points. The Isomap algorithm is as follows:

    1) The local diagram K created the nearest neighborhood (KNN) for all data points.
    2) Estimated the geodetic distance between all data points.
    3) Finally, MDS is utilized to the resulting geodetic matrix to detect d-dimensional embedding by combined execution.

- Step 3- Classification with MLP neural network: In this step, the normalized and reduced data are analyzed using the MLP neural network. These classifications try to be trained using the sessions of each cluster of information. Once the additional data properties are removed, they will be provided to the multilayer perceptron neural network. Neural networks belong to the category of machine learning algorithms that try to predict the process of converting output from input data using regression by receiving input and output data. In this case, with the new input data, the output can be predicted. In multilayer perceptron neural networks, the number of layers is limited and is determined by the user. In each layer there are a number of neurons that process the data received from the previous layer and output to the next layer. It should be noted that in most applications an input layer is used whose number of neurons is determined by the user. In addition to the input layer, there is also an output layer (the last layer of the network) which has the number its neurons equally to the number of outputs. A multilayer perceptron neural network in the first layer used to estimate regression uses sigmoid functions and uses the Identical function for the neurons entities in the output layer.

    - **Data segmentation:**

    A multilayer perceptron neural network receives input data and divides them into three parts. These parts are:

    1) Training data: This data is used to train neurons. In fact, in this part of the data, the input and output are specified and will be provided to the neurons so that with the help of this data, they can adjust their sigmoid functions.
    2) Validation data: This data, unlike training data, is unknown. This means that the output is hidden and only the input is given to the neurons. This data is used to measure the function of neurons during the training process. In fact, it is the data that determines when learning process is ended. When the error rate is less than a certain amount or the number of times the error rate exceeds a certain number, during the training process, the training process will be terminated.
    3) Test data: This data, like the validation data, is unknown. But the difference with the validation data is that it is provided to the neurons after the training process, and it is with this data that the prediction error of the multilayer perceptron neural network is determined.

    - **Neural network training:**

    After removing additional features from the data and applying them as input to the neural network, as well as determining how to divide the data and the number of neurons in the neural network, it is time to train the neural network. In this step, a neural network tries to adjust the sigmoid function of each neuron in such a way that it has the best estimate of the input-to-output data. Neural network training continues until one of the termination conditions is reached, which can be one of the following:

    1) Specific number of repetitions
    2) Specific training time
    3) Considered performance
    4) Specific validation number
    5) Specific gradient amount

In case of any of these cases, the training process will be stopped, and the trained neural network will be provided.

4. Evaluate The Proposed Method

We used this method (already has been described in section 3) to examine the efficiency by using the MATLAB programming language. The experiments were performed in an environment with the conditions satisfies the proposed method.

To evaluate methods that are categorical in nature, four criteria are often used: False Positive, True Positive, False Negative, and True Negative which is defined in the following of each of these criteria.

True Positive: In this case, the predicted category for the data sample pair is the same and the actual category of the data sample pair is the same. This condition is said to be correctly predicted positive in the correction.

False positive: In this case, the predicted category for the data sample pair is the same but the actual category of the data sample pair is different. This condition is said to be incorrectly predicted positive in the correction.

True Negative: In this case, the predicted category for the data sample pair is not the same and the actual category of the data sample pair is not the same. This condition is said to be correctly predicted negative in the correction.

False Negative: In this case, the predicted category for the data sample pair is not the same, but the actual category of the data sample pair is the same. This condition is said to be incorrectly predicted negative in the correction.

Now, according to these four parameters, the following four evaluation criteria are introduced:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (2)$$

$$\text{Recall} = \frac{\mathbf{TP}}{\mathbf{TP+FN}} \quad (3)$$

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad (4)$$

$$\mathbf{F1} = \frac{\mathbf{2 \times Precision \times Reacall}}{\mathbf{Precision + Reacall}} \quad (5)$$

The KDD Cup dataset and was one of the benchmark databases offered on the famous UCI web, which is intended for analysis, is the data set used for the studies. This dataset, which several current and recent publications have used in the field of attack detection, including [38-41], contains 41 features and 4969 data samples. Features of each session include session duration, protocol type, service, active flags, number of bytes sent / received, session location, and more. In addition to 41 attributes, one attribute represents the session class. In fact, the meeting class indicates that the meeting is healthy or aggressive. Therefore, each session will be one of two healthy sets or attacks. In this data set, the types of attacks mentioned include the table 1.

Table 1 - Types of attacks in the KDD Cup dataset

Imap R2l	Back Dos	Buffer_Overflow U2r	Ftp_Write R2l
Warezclient R2l	Guess_Passwd R2l	Warezmaster R2l	Rootkit U2r
Teardrop Dos	Spy R2l	Smurf Dos	Satan Probe
Portsweep Probe	Pod Dos	Phf R2l	Perl U2r
Nmap Probe	Neptune Dos	Multihop R2l	Ipsweep Probe
Loadmodule U2r	Land Dos		

The evaluation of the proposed method as well as the comparable method is done based on two different scenarios. In these evaluations, the first scenario works on the size of the training set. In this scenario, first 30% of the data set is randomly selected and entered into the evaluated systems as a training set. It is clear that in this case the size of the test set is 70%. In the next experiment, the size of the training set changes to 40% and of course the size of the test set to

60%. This process continues until the size of the training set reaches 80%. It should be noted that in this scenario, the size of the number of features is equal to 10% of the total features.

4.1 Variable Size of Training Set

The results of all four evaluation criteria for the first scenario are shown in Fig (3) to (6).

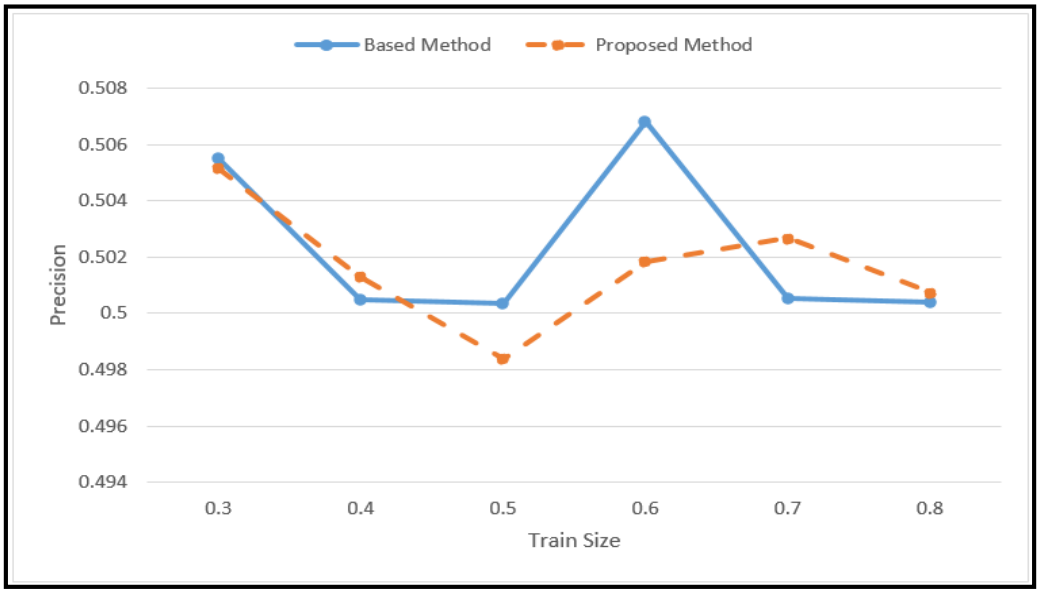

Fig. 3 - Precision comparison in the first scenario for the methods being compared

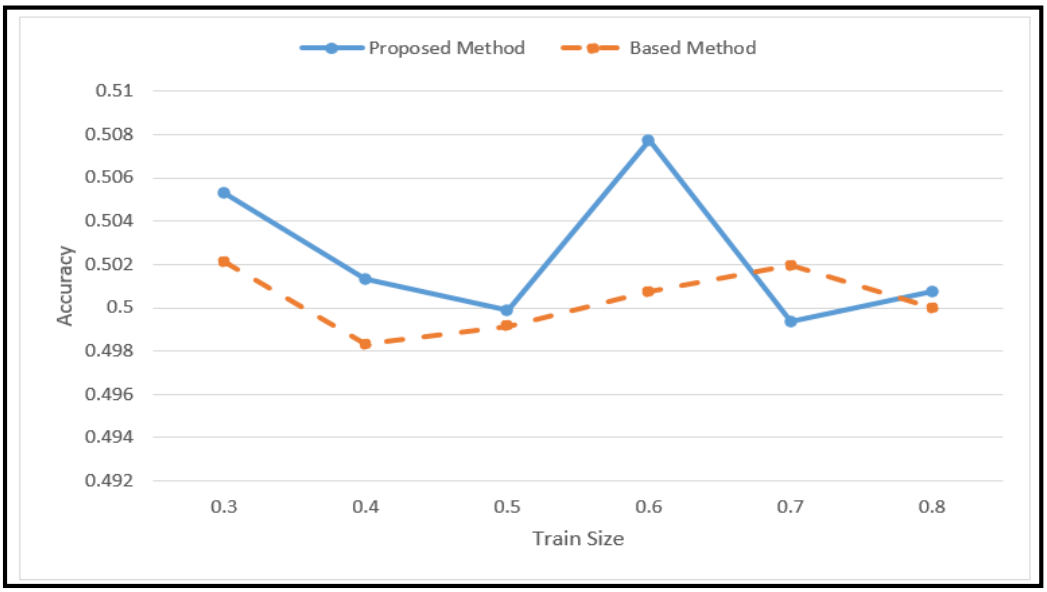

Fig. 4 - Accuracy comparison in the first scenario for the methods being compared

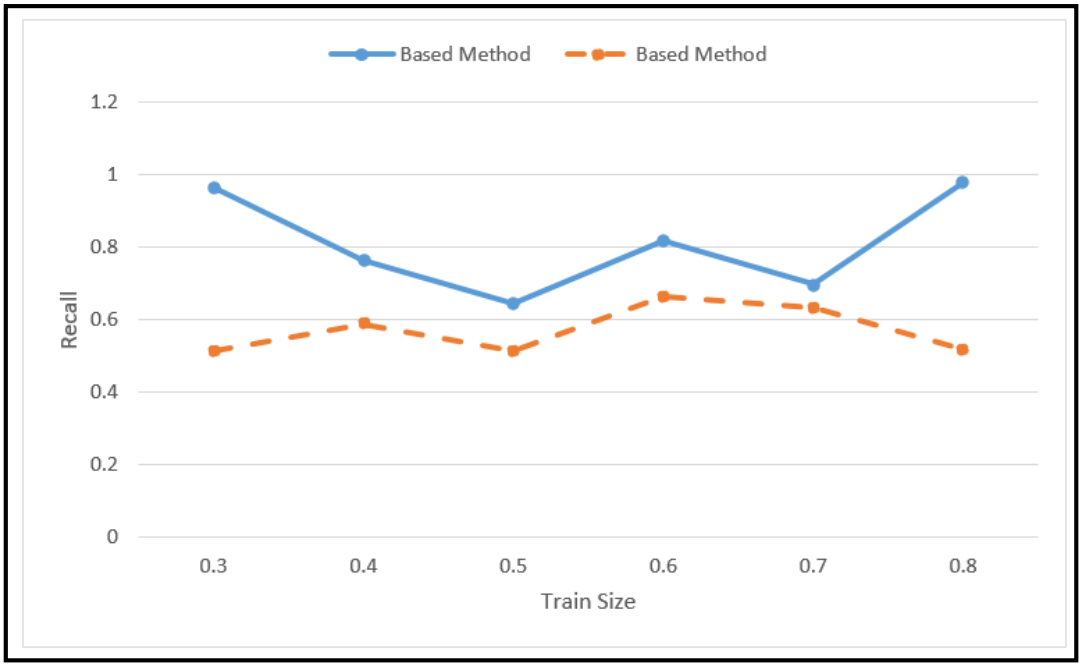

Fig. 5 - Recall comparison in the first scenario for the methods being compared

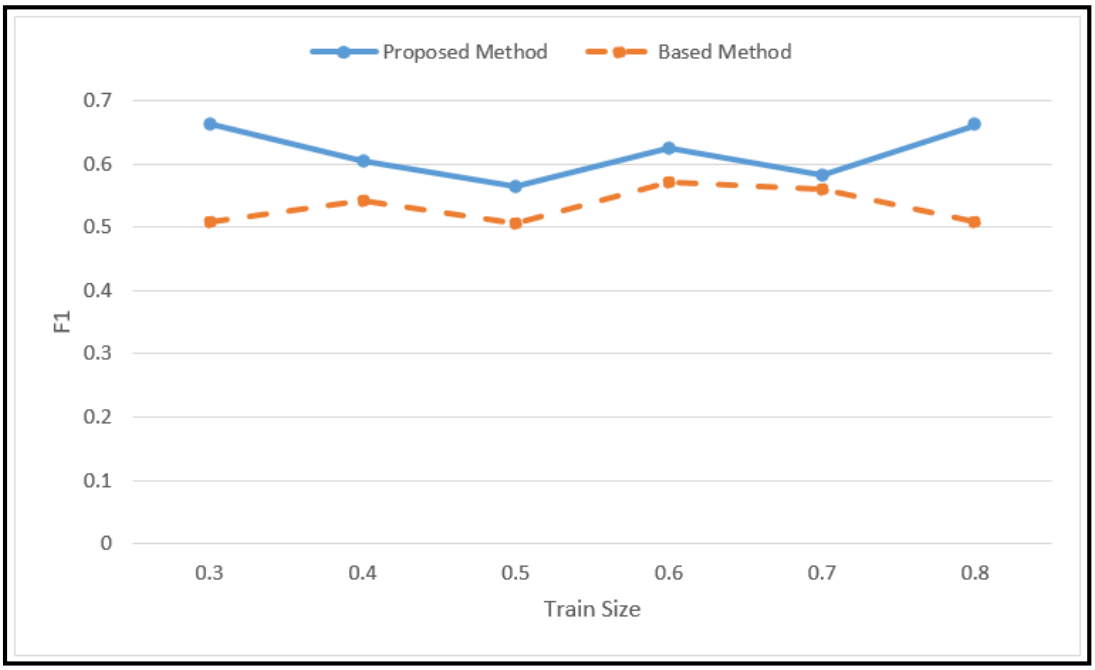

Fig. 6 -. Comparison of Measure in the first scenario for the methods being compared

In the results of evaluations performed in the above figures, the proposed approach has been able to show better results than the basic method. With the exception of some limited cases in the Precision and Recall criteria, the proposed method has shown better results than the compared method. This indicates that the proposed method has been able to train its classifier well for different sizes of training data sets and achieve better results than the comparison method. In fact, the size of the training set did not have much effect on the quality of the proposed system. The purpose of this scenario is to investigate the scalability of the methods compared to the size of the training set and the proposed approach has shown its scalability well.

4.2 Variable Size Number of Features

In this scenario, unlike the first scenario, the size of the training set is fixed, but it is the number of selected features that has changed. In fact, in this scenario, first the size of the training set is 70% of the data and, of course, 30% is intended for testing. But for the number of features, first 10%, then 20%, and so on until 70% of the number of features remain in the system. The results of this scenario can be seen in Figures (7) to (10).

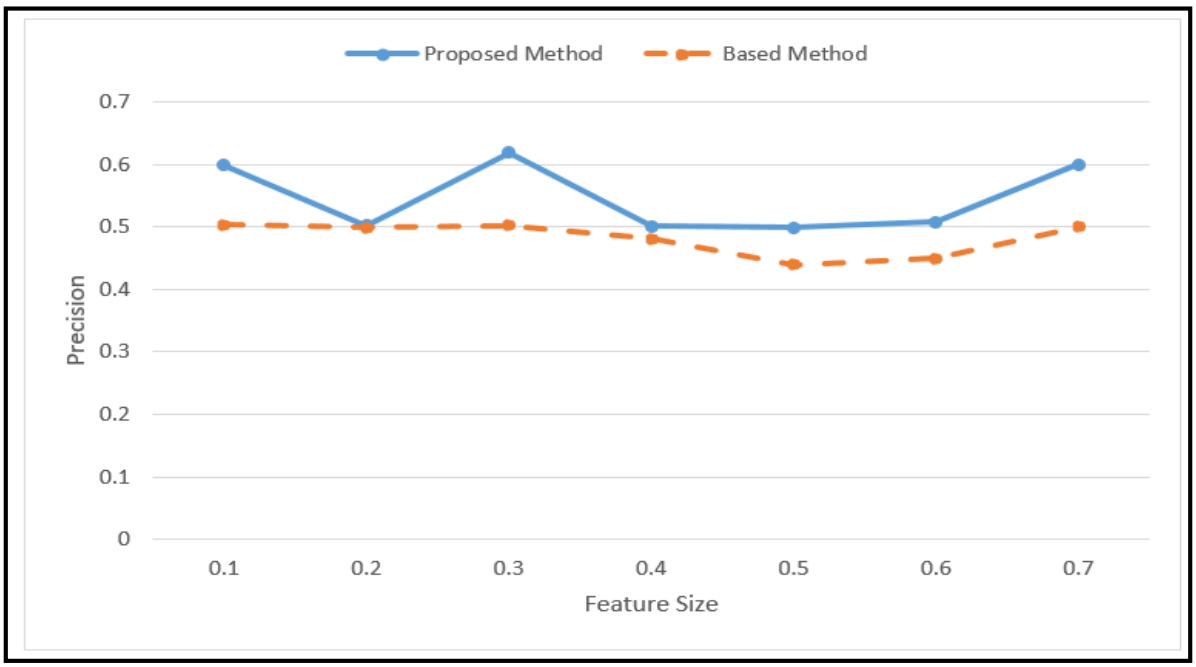

Fig. 7 - Comparison of Precision for a number of different features

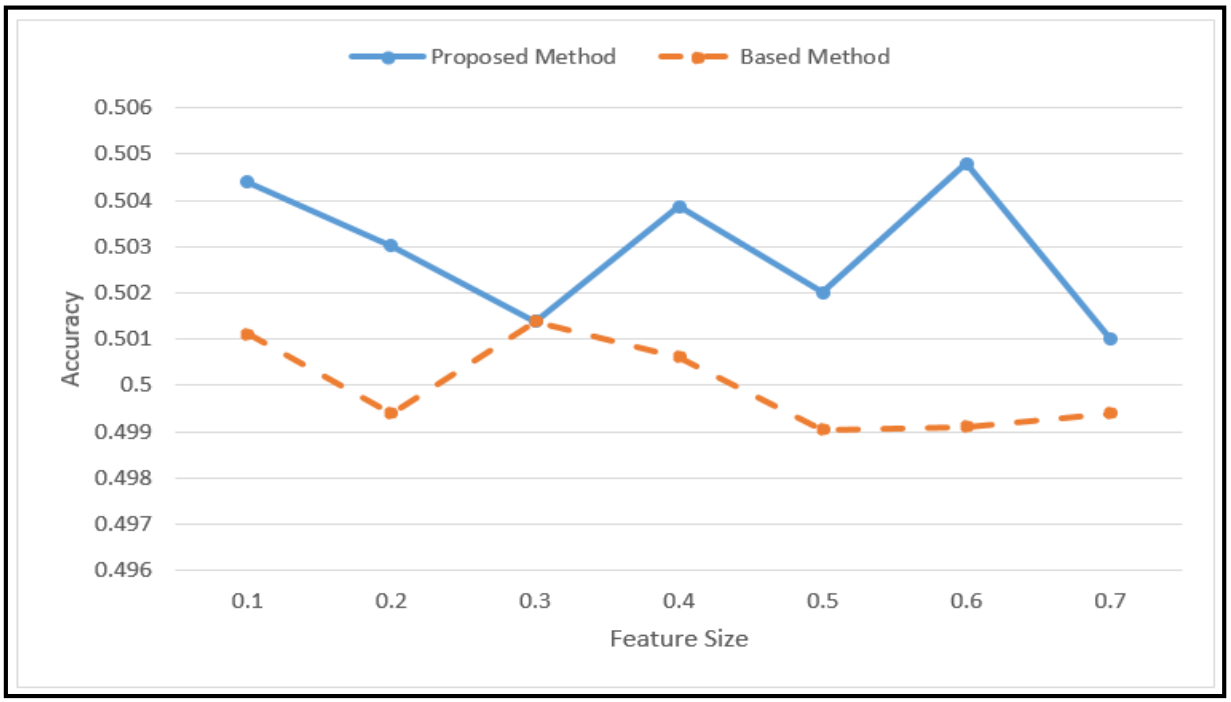

Fig. 8 - Accuracy Comparison Jump for a number of various features

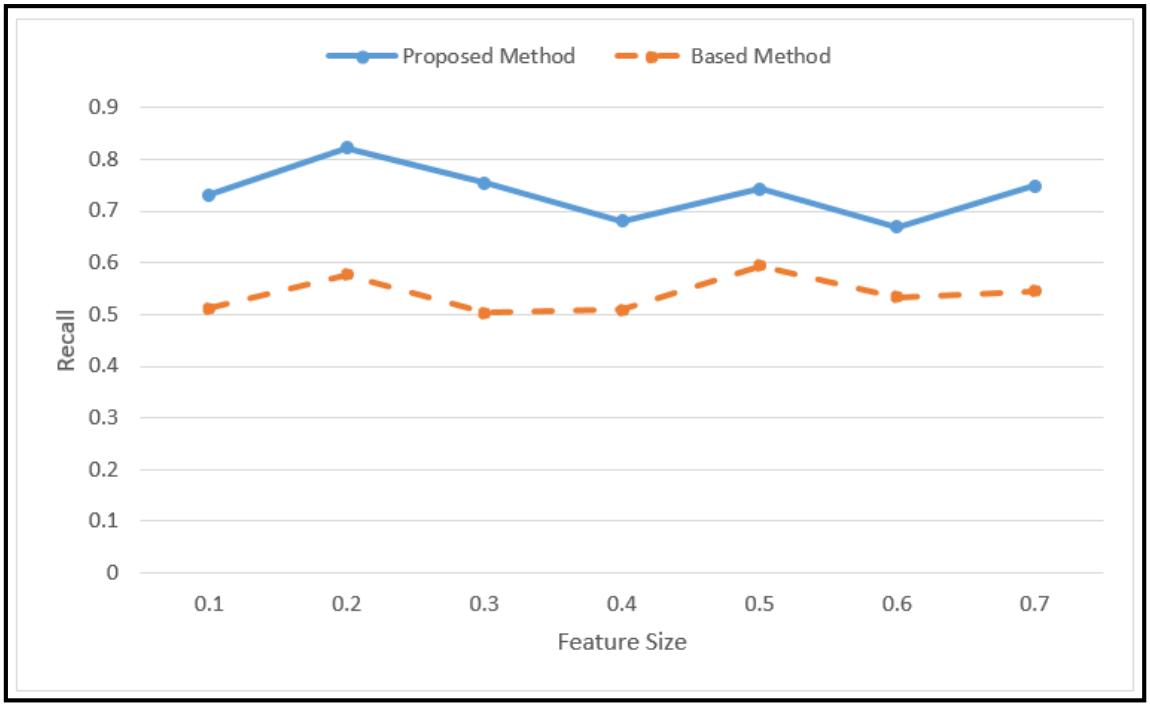

Fig. 9 - Comparison of Recall for various features

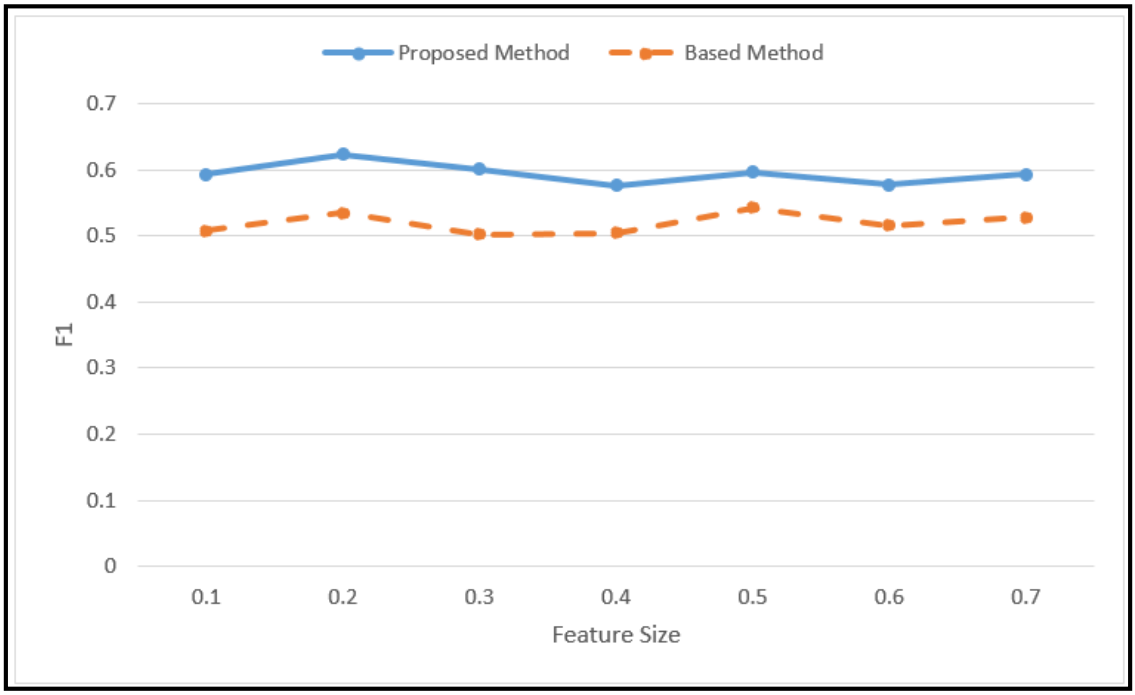

Fig. 10 - Comparison of F1 for a number of different features

In the results of this scenario, as in the previous scenario, the proposed method will provide better performance than the compared approach. The purpose of this scenario is to examine the impact of the number of features on system performance. As shown in these results, in all the studied periods, the proposed method has been able to perform more favorably in all evaluation criteria. This shows that the methods used to select the features as well as the classification used have been able to communicate better with each other and thus result in the success of the system.

5. Conclusion

The latest developments in various fields have led to the collection of a large amount of data, and the data is usually stored in various formats, such as sound recordings, files, images, audio and video. The collected data is used in the decision-making process, but such a large amount of data makes data management/analysis complicated and difficult. Using data to make better decisions requires extracting knowledge from large repositories in the right way. Data mining techniques can be used to extract valuable and meaningful insights from large amounts of data. An analytical tool that provides important information and insights, and can improve the decision-making process.

In this paper, an algorithm is used to extract and select a feature from the so-called teacher-student approach and to classify the MLP neural network approach. In the proposed approach, after entering the data into the system and pre-processing them, in the teacher section, we extract the features and also reduce the features. Then, the student section goes to the classification stage by dividing the data set, which has been reduced in complexity by reducing its features. In this step, the training data is first used to create a neural network model based on multilayer perceptron neural networks. This trained neural network is then used to categorize the experimental data.

In order to evaluate the proposed approach, along with one of the latest articles in the field under discussion, this paper has been implemented by MATLAB programming language. In the experimental results performed on a benchmark data set and on the conventional classification evaluation criteria, namely Precision, Recall, Accuracy and F-Measure in the form of two different scenarios, it was found that the proposed approach could achieve better results compared to the basic method.

6. Future works

Feature reduction is still an unresolved issue in the minds of many researchers. Therefore, interested researchers can seek to improve the performance of these algorithms by providing new solutions. One of these methods is feature selection. Feature selection is a search issue. Therefore, we suggest to interested researchers that by combining new evolutionary algorithms such as fireflies, gray wolves, or bats, they seek to provide approaches to select useful features from the data set. Because these algorithms have good performance for searching complex spaces.

Acknowledgment

The authors fully acknowledged Northern Technical University for supporting this work.

References

[1] J. Mirzaei, A., Pourahmadi, V., Soltani, M., & Sheikhzadeh, H. (2020). Deep feature selection using a teacher-student network. Neurocomputing, 383, 396-408

[2] Alaba, F. A., Othman, M., Hashem, I. A. T., & Alotaibi, F. (2017). Internet of Things security: A survey. Journal of Network and Computer Applications, 88, 10-28.

[3] Aamir, M., Zaidi, M.A., 2013. A survey on DDoS attack and defense strategies: from traditional schemes to current techniques. Interdisciplinary Inf. Sci. 19 (2), 173– 200.

[4] Kim, J., Sim, A., Tierney, B., Suh, S., Kim, I., 2018. ‘‘Multivariate network traffic analysis using clustered patterns,”. Computing, 1–23.

[5] Bawany, N. Z., Shamsi, J. A., & Salah, K. (2017). DDoS attack detection and mitigation using SDN: methods, practices, and solutions. Arabian Journal for Science and Engineering, 42(2), 425-441.

[6] M. Aldwairi, Y. Khamayseh, M. Al-Masri, Application of artificial bee colony for intrusion detection systems, Sec. Commun. Netw. 8 (16) (2015) 2730–2740; (b) I.

[7] P. Nancy, S. Muthurajkumar, S. Ganapathy, S. V. N. Santhosh Kumar, M. Selvi, and K. Arputharaj, ‘‘Intrusion detection using dynamic feature selection and fuzzy temporal decision tree classification for wireless sensor networks,’’ IET Commun., vol. 14, no. 5, pp. 888–895, Mar. 2020.

[8] Eduardo De la Hoz, Emiro De La Hoz, Andreıs´ Ortiz, Julio Ortega, Beatriz Prieto, PCA Filtering and Probabilistic SOM for Network Intrusion Detection, vol. 164, Special Issue: Advances in Computational Intelligence in Elsevier—Neurocomputing, 2015, pp. 71–81.

[9] Sunil Nilkanth Pawar, Rajankumar Sadashivrao Bichkar, Genetic algorithm with variable length chromosomes for network intrusion detection, Int. J. Autom. Comput. 12 (3) (2015) 337–342.

[10] Belouch, M., El Hadaj, S., & Idhammad, M. (2018). Performance evaluation of intrusion detection based on machine learning using Apache Spark. Procedia Computer Science, 127, 1-6.

Abbreviation

IDS Intrusion detection system

ML Machine learning

TSFS Teacher-Student Feature Selection

IoT Internet of things

AI Artificial Intelligence