IMPROVING INTRUSION DETECTION SYSTEM USING THE COMBINATION OF NEURAL NETWORK AND GENETIC ALGORITHM

Amin Dastanpour[1], Amirabbas Farizani[1], Raja Azlina Raja Mahmood[2]

[1]Computer Department, Kerman institute of higher education, Kerman, Iran
[2]Department of Communication Technology and Network, Faculty of Computer Science and Information Technology, University Putra Malaysia

ABSTRACT

One of the essential issues in network-based systems is a fault attack which is caused by intrusion. It is the responsibility of intrusion detection to provide capabilities such as adaptation, fault tolerance, high computational speed, and error resilience in the face of noisy information. Thus, the construction of an efficient intrusion detection model is highly appreciated to increase the detection rates as well as to decrease false detection. Currently, researchers are more focusing on abnormal behaviour of network as this system can easily recognize new attacks without updating the daily recognized databases. However, the capability of current developing machine learning algorithms suffers from inefficient use of intrusion detection particularly once it involved some huge datasets of irrelevant and redundant features. The main objective of this thesis is to achieve the higher detection rate with lower false detection for attack recognition in order to support efficient application of intrusion detection. To achieve this goal, a new machine learning model was designed and developed to provide intelligent recognition with new attack patterns. New proposed and improved algorithm GA-ANN is constructed to support the proof of concept. For evaluation, five datasets namely KDD CUP 99 from the online data repositories are used in the experiment. In the above scenarios, GA-ANN provides the highest detection rate for pattern recognition which was 98.98% based on 18 selected features. This means that the proposed IDS model is significant and increases the network security.

KEYWORDS

Intrusion detection system, Neural network, Genetic algorithm

1. INTRODUCTION

To prevent the attacks of these types of attackers, the administrator uses the intrusion detection system. According to the report published by Symantec (anti-virus company) for the month of November 2023, the number of internal attacks has increased, 5965 new malicious attacks have been discovered[1]. They are first trained with known inputs to "learn" attack patterns and then validated with unknown input attacks. In addition to the ability of machine learning algorithms to detect new attack patterns, it is noteworthy that these algorithms can be used on huge data sets with irrelevant and redundant features and consider only a few important features to optimize the detection process[2]. The above context, it can be concluded that in anomaly-based IDS, many features must be considered to detect specific attacks, and the system must deal with a huge amount of network traffic[3]. In anomaly-based IDS, there are many features that must be considered to detect specific attacks, and the system must deal with a huge amount of network traffic with highly unbalanced data distribution[4]. To check the capability of machine learning methods as an intrusion detection system, the ANN method has been used and validated in this

DOI: 10.5121/ijnsa.2024.16602

research[5]. To secure network systems, intrusion detection system (IDS) is commonly used as network security insurance. This study is to support IDS and improve the behaviours of secure networks from attack patterns[6] .

IDS have two types; network based (NIDS) and host based (HIDS) intrusion detection systems. IDS has two methods for detection, namely Signature-based IDS and Statistical anomaly-based IDS [7].

1.1. Network Intrusion Detection Systems

Network intrusion detection systems (NIDS) are placed at a strategic point or points within the network to monitor traffic from all devices on the network [8]. It performs to analyse a passing traffic on the entire subnet, works in a promiscuous mode, and matches the traffic that is passed on the subnets to the library of known attacks. Once the attack is identified or abnormal behaviour is discovered, the alert is sent to the administrator [9]. Example of using NIDS is being installed in the subnet, where firewalls are located in order to see whether someone is trying to break into the firewall [10].

1.2. Host Intrusion Detection Systems

Host Intrusion Detection Systems (HIDS) run in individual hosts or devices in the network [11]. A HIDS monitors only the inbound and outbound packets from the device and alerts the user or administrator if suspicious activity is detected. It takes a snapshot from existing system files and matches it to the previous snapshot [12]. If the critical system files were modified or deleted, the alert is sent to the administrator to investigate. An example of HIDS usage can be seen on mission critical machines, which are not expected to change their configurations [13].

1.3. Signature-Based IDS

A signature-based IDS monitor packets of the network and compare them with a database of signatures or attributes from known malicious threats. This is similar way to the most antivirus software detects malware.

The main signature-based IDS problem is that there will be a lag between a new threat being discovered and the signature for detecting the threat being applied to your IDS. Therefore, IDS would not be able to detect the new threat during that lag time [14].

1.4. Statistical Anomaly-Based IDS

An anomaly-based IDS monitors network traffic and compare it with an determined baseline[15]. The baseline identifies “what is “normal” for that network?”, “what sort of bandwidth is generally used?”, “what protocols are used?”, “what ports and devices generally connect to each other?”, and alert the administrator or user whenever traffic is detected as an anomalous or significantly different compared to the baseline. The main drawback of anomaly-based IDS is that it may raise a false positive alarm for a permissible use of bandwidth if the baselines are not intelligently configured [16].

1.5. Using Machine Learning for Statistical Anomaly-Based IDS

In anomaly-based IDS, there is too many attributes need to be considered when attacks are identified. The system must deal with huge amounts of network traffic with highly imbalanced data distribution. Thus, recognition of normal versus from abnormal behaviour is challenge[17]. With huge amounts of traffics, more data must be considered when forming patterns. This increases the probability of high false positive rate.

The goal of machine learning is to learn and/or discover as well as adapted to changing circumstances over time and improves its performance on certain tasks over time. In the intrusion detection system, the machine learning algorithms are firstly trained with known input to “learn” the attack patterns and then were validated with unknown input attacks. Besides the capability of machine learning algorithms for recognition of new attack patterns, it is remarkable to note that these algorithms could also be used in a case of huge datasets with irrelevant and redundant features and consider only few important features in order to optimize the detection process [18]. One of the important issues of machine learning is classification. The question raises form this statement is “How to classify a set of categories (sub-populations)?” For example, when the email received, it would be assigning into "spam" or "non-spam" classes or assigning a diagnosis to a given patient as described by observed characteristics of the patient (gender, blood pressure, presence or absence of certain symptoms). [19].

In the terminology of machine learning, classification is considered an example of supervised learning. Supervised learning are learning from a training set of correctly identified observations [20]. The most important classification algorithms are Artificial Neural Networks (ANN).

To improve the efficiency of machine learning model, optimization techniques are linked to the model. Optimization is the selection of a best element (with regard to some criteria) from some set of available alternatives [21]. In the simplest case, an optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from an allowing set and computing the value of the objective function. In an overall, optimization techniques are employed to find out the "best solution “for objective function while all constraints are satisfied [22]. Among various optimization algorithms Gravitational Search Algorithm (GSA) and Genetic Algorithm (GA) are the so efficient and powerful algorithms.

From the above background description, it can be concluded that in anomaly-based IDS, there are too many attributes need to be considered for identifying specific attacks and system must deal with huge amounts of network traffic. Thus, the recognition of normal versus abnormal behaviour is making challenge. With huge traffics, more data have to be considered when forming patterns, which increased risk of false positive rate. For this reason, many researchers use machine learning technique, and the output results show that techniques can give high accuracy results in detecting anomalies.

In signature-based systems, attack patterns or behaviours of intruders are modelled, and the system will alert once a match is detected. The limitation of these systems is that they can only detect known attacks and require the attack signatures to be updated frequently. To explain more, anomaly detection systems firstly have to create a baseline profile of the normal behaviour in system or network. Afterwards, if any activity deviates from the normal profile, it will be known as an intrusion. Anomaly detection systems are able to detect unknown attacks and thus more efficient than the signature based. In anomaly-based IDS, there are too many attributes need to be considered for identifying specific attacks and system has to deal with huge amounts of network traffic with highly imbalanced data distribution. Thus the recognition of normal versus abnormal behaviour is making challenge[17]. With huge traffics, more data have to be considered when forming patterns, which increased risk of false positive rate. Despite the forgoing challenges, many researcher use machine learning technique and the output results show that these techniques can give high accuracy results in detecting anomalies [23].

2. RESEARCH BACKGROUND

Research work 1 (Abraham & Grosan, 2006)[24]

In Distributed Intrusion Detection System (DIDS), the conventional intrusion detection system is embedded inside intelligent agents and deployed on a large network. Several data mining algorithms are applied to the audit data to compute models that accurately capture the actual behavior of intrusions as well as normal activities. Audit data analysis and mining combine association rules and classification algorithms to discover attacks in audit data. A distributed identifier (DIDS) consists of several IDSs in a large network (network), all of which are connected to each other or to a central server that facilitates advanced network monitoring. In a distributed environment, DIDS is implemented using common intelligent agents distributed across the network(s). In this paper, we presented a framework for distributed intrusion detection systems (DIDS) using several soft computing paradigms .

Research work 2 (Amiri, Yousefi, Lucas, Shakery, & Yazdani, 2011)[25]

This paper proposes a feature selection approach based on linear correlation to construct the NID model. The first layer selects a feature subset based on the analysis of Pearson correlation coefficients between features. While the second layer selects a new set of features from the subset of edited features of the first layer. This article formulates and validates a method to select the subset of optimal features based on the analysis of Pearson correlation coefficients. In this paper, a feature selection method based on linear correlation is proposed to construct the NID model. It consists of two layers, where the first layer selects a feature subset based on the analysis of Pearson correlation coefficients between features. By analyzing the Pearson correlation coefficients between the selected features and classes .

Research work 3 (Lei & Ghorbani, 2012)[26]

In this research, we propose two new clustering algorithms, improved competitive learning network (ICLN) and supervised improved competitive learning network (SICLN), for fraud detection and network intrusion detection. ICLN is an unsupervised clustering algorithm that applies new rules to the standard competitive learning neural network (SCLN). The network neurons in ICLN are trained to represent the data center with a new reward-punishment update rule. This new update rule overcomes the instability of SCLN. SICLN is a supervised version of ICLN. In SICLN, the new supervised update rule uses data labels to guide the training process to achieve a better clustering result. SICLN can be applied to both labeled and unlabeled data and is highly robust to missing labels or delays. In addition, SICLN is able to regenerate itself, so it is completely independent of the initial number of clusters. To evaluate the proposed algorithms, experimental comparisons have been made on research data and real-world data in fraud detection and network intrusion detection. The results show that both ICLN and SICLN have high performance, and SICLN outperforms traditional unsupervised clustering algorithms .

Research work 4 (Sree Kala & Christy, 2019)[27]

The neural network has the function of pattern recognition and can be used in the field of intrusion detection classification. At the same time, neural network has self-learning and adaptive capacity. As long as the system audit data and the network data packet are provided, the neural network can extract the normal user model or system attribute from it and distinguish the attack mode from the abnormal activity. Probabilistic neural network (PNN) is a type of artificial neural network first proposed by Dr. Donald F. Specht in 1989, which has the characteristics of simple structure, faster convergence, and wide application. An intrusion detection model is presented to perform network intrusion classification. Finally, the experiment was performed based on the KDD cup data set. In order to improve the network intrusion classification performance, probabilistic neural network is applied in the field of intrusion detection. By analyzing the theory of probabilistic neural network, a kind of network intrusion classification model based on probabilistic neural network has been presented to solve the two problems of classification and multi-class classification in intrusion detection. Finally, the experiment was performed on the KDD Cup 99 database.

The results show that the proposed model performs better. In this paper, probabilistic neural network is applied in the context of network intrusion classification. In this paper, probabilistic neural network is applied in the context of network intrusion classification. The classification test was performed on the KDD Cup 99 database, which showed that the proposed model has a higher intrusion detection rate .

Research work 5 (Kshirsagar et al., 2022)[28]

In the first step, fuzzy clustering technique is used to generate different training subsets. However, for ANN-based IDS, the detection accuracy, especially for low-frequency attacks, and the detection stability still need to be improved. In this paper, we propose a new approach, called FC-ANN, based on ANN and fuzzy clustering, to solve the problem and help IDS to achieve higher detection rate, lower false positive rate and stronger stability. The general method of FC-ANN is as follows: First, the fuzzy clustering technique is used to generate different training subsets. In this paper, we propose a new intrusion detection approach, called FC-ANN, based on ANN and fuzzy clustering. Experimental results using the KDD CUP 1999 database show the effectiveness of our new approach, especially for low-frequency attacks, namely R2L and U2R attacks, in terms of detection accuracy and detection stability .

However, there is no paper to review and understand the current state of using machine learning techniques to solve intrusion detection problems. IDSs can be used to detect various types of malicious network communications and use of computer systems, while a typical firewall cannot do this. In the literature, a large number of anomaly detection systems have been developed based on many different machine learning techniques. However, a review of these different machine learning techniques in the field of intrusion detection is lacking. We consider many machine learning techniques used in the field of intrusion detection for investigation, including single, hybrid, and ensemble classifiers. According to the comparative results of related work, the development of intrusion detection systems using machine learning techniques still needs research. Table 1 shows the research criteria .

Table 1 Research criteria

Papers	Paper NO.1	Paper NO.2	Paper NO.3	Paper NO.4	Paper NO.5	
Name of algorithm	(LCFS)	(FR2)	(FC-ANN)	(PNN)	(ICLN)	(SICLN)
High detection rate	✓	✓	✓	✓	✓	✓
Low false alarm			✓			
Minimum number of features	✓					

3. Proposed Intrusion Detection Model

The outline of this research is shown in Figure 1. This research examines the use of machine learning algorithms to build intrusion detection systems. First, benchmark data such as KDD CUP 99 is collected[29]. In this research, artificial neural networks (ANN) are used as intrusion

detection systems. Machine learning algorithms learn from the database in the training phase and make predictions or decisions based on example inputs. ANNs have been optimized and validated to increase the efficiency of systems and use fewer features for detection. Also, genetic algorithms (GA) are used as optimization techniques to optimize recognition patterns. The optimized models are called GA-ANN and their results are compared based on the detection rate and the number of detection features. Figure 1 shows the main model of IDS.
The parameter and setting of GA is shown in table 2.

Table 2 Parameters of Genetic Algorithm

Genetic algorithm Parameter	Genetic algorithm Amount
Number of iterations	49 in this case
Population size	20
Mutation rate	0.15
Fraction of population of kept	0.5
Total number of bits in a chromosome	41
Crossover	single point
Type of GA	Binary GA

In this research, stopping criteria is: If fi - fi-5 ≤ 10-6 then stop iterations. A fitness function is a particular type of objective function that is used to summaries, as a single figure of merit, how close a given design solution is to achieving the set aims. Each design solution, therefore, needs to be awarded a figure of merit to indicate how close it came to meeting the overall specification, and this is generated by applying the fitness function to the test, or simulation, results obtained from that solution. In this paper F is fitness function and equation (1) are fitness function in this study.

$$f = \frac{\alpha}{A} - \frac{\beta}{B} \qquad (1)$$

Where,α the number of correctly identified attacks, A is the total number of attacks in the training database, βis the number of normal connections incorrectly characterized as attacks, andB is the total number of normal connections in the training database.

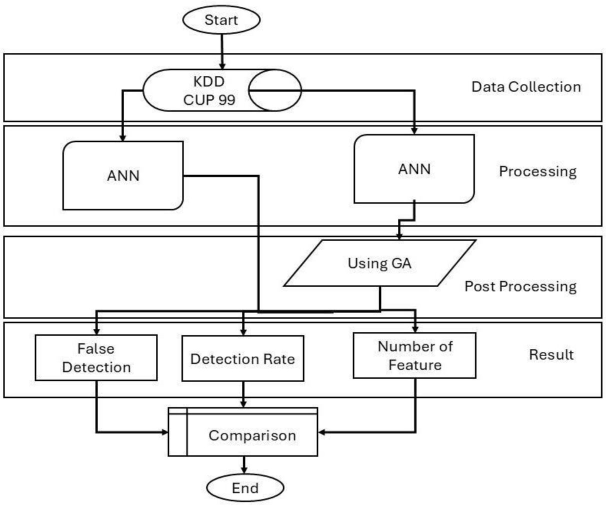

Figure 1 Main Model of IDS

3.1. Working Principle of GA

The working principle of GA is explained in [30], Genetic Algorithm begins with a set of suitable solutions for the problem. Each solution is represented by a chromosome-like data structure. Solutions from one population are selected and used to generate a new population. This is motivated by the possibility that the new population will be better than the old one. Solutions are selected according to their fitness to generate new population; more suitable they are more chances they have to reproduce. This is repeated until some condition (e.g. fixed number of generations reached or improvement of the best solution etc.) is satisfied.

The pseudo-code for GA is as shown below Pseudo code:

```
BEGIN
INITIALISE population with random candidate solutions.
EVALUATE each candidate;
REPEAT UNTIL (terminate condition) is satisfied DO
1.        SELECT parents;
2.        RECOMBINE pairs of parents;
3.        MUTATE the resulting offspring;
4.        SELECT individuals or the next generation;
END
```

3.2. Mutation

Mutation is a genetic operator used to maintain genetic diversity from one generation of a population of genetic algorithm chromosomes to the next. It is analogous to biological mutation. Mutation alters one or more gene values in a chromosome from its initial state. In mutation, the solution may change entirely from the previous solution. Hence GA can come to better solution by using mutation. Mutation occurs during evolution according to a user-definable mutation probability. This probability should be set low. If it is set too high, the search will turn into a primitive random search. The classic example of a mutation operator involves a probability that an arbitrary bit in a genetic sequence will be changed from its original state.
Bit string mutation

The mutation of bit strings ensue through bit flips at random positions.
Example:

1 0 1 0 0 1 0

↓

1 0 1 0 1 1 0

The probability of a mutation of a bit is ^{1}, where L is the length of the binary vector. Thus, a mutation rate $_{L}$ of 1 per mutation and individual selected for mutation is reached.

3.3. Selection

Selection is the stage of a genetic algorithm in which individual genomes are chosen from a population for later breeding (using the crossover operator).A generic selection procedure may be implemented as follows: The fitness function is evaluated for each individual, providing fitness values, which are then normalized. Normalization means dividing the fitness value of each individual by the sum of all fitness values, so that the sum of all resulting fitness values.The population is sorted by descending fitness values.

Accumulated normalized fitness values are computed (the accumulated fitness value of an individual is the sum of its own fitness value plus the fitness values of all the previous individuals). The accumulated fitness of the last individual should be 1 (otherwise something went wrong in the normalization step).

A random number R between 0 and 1 is chosen.
The selected individual is the first one whose accumulated normalized value is greater than R.

3.4. Crossover

In this paper used the Crossover technique. In genetic algorithms, crossover is a genetic operator used to vary the programming of a chromosome or chromosomes from one generation to the next. In this paper used fitness proportionate selection as a selection method of chromosomes for crossover part. Fitness proportionate selection is the individual is selected on the basis of fitness. The probability of an individual to be selected increases with the fitness of the individual greater or less than its competitor's fitness. Fitness proportionate selection, also known as roulette wheel selection, is a genetic operator used in genetic algorithms for selecting potentially useful solutions for recombination. In fitness proportionate selection, as in all selection methods, the fitness function assigns fitness to possible solutions or chromosomes. This fitness level is used to associate a probability of selection with each individual chromosome.

4. Genetic Algorithm and Artificial Neural Network (GA-ANN)

ANN can be used to identify and classify data. However, in order to classify and identify, a large database is required by ANN. To optimize this type of data and overcome the accuracy problem of artificial neural network, this paper proposes the use of Genetic Algorithm (GA). The purpose of this paper is to propose the use of genetic algorithm to improve the ANN mechanism. Genetic algorithm is one of the most popular and widely used algorithms for machine learning. Genetic algorithm is a heuristic and adaptive algorithm for work and search, which is based on evolutionary ideas of natural genetics. In GA, a solution is provided by each of these individuals to the problem. Since GA is a parallel algorithm and is able to find a solution to a problem with multiple subsets, it is considered suitable for IDS. Moreover, GA is able to suggest a solution in a solution with optimal value. Another feature of GA is that it is a suitable method for IDS, especially for detecting attacks based on human behavior. In machine learning, performance is usually evaluated according to the ability to reproduce known knowledge, while the key task is to discover previously unknown knowledge. Evaluated with respect to known knowledge, an uninformed (unsupervised) method easily outperforms supervised methods, and supervised methods cannot be used due to unavailability of training data. Figure 2 shows the flowcharts of GA-ANN.

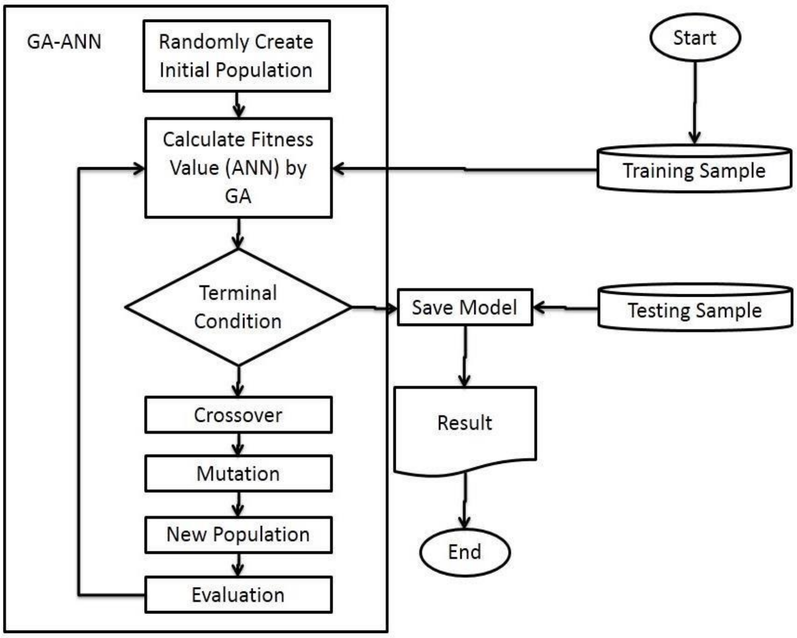

Figure 2 Flowcharts of GA-ANN

5. Evaluation of the Proposed IDS Model

In this section, the results of the improved algorithm and the new algorithm for intrusion detection systems (IDS) are reviewed. A comparison is made between machine learning algorithms in terms of high accuracy and low false rate, and the results of each algorithm using the IDS database are presented in terms of detection rate and false detections. Table 3 shows the results of ANN and GA-ANN.

Table 3 Results of ANN and GA-ANN.

Algorithm	**KDD CUP 99 Database**		
	Detection rate	False Alarm	Number of Features
ANN	97.8%	0.059%	41
GA - ANN	98.98%	0.057%	18

The GA-ANN algorithm introduced in this thesis has been able to achieve the research goal with the highest accuracy (98.98%) and the lowest false detection rate (0.057%) in the KDD database. This algorithm has significantly improved compared to other machine learning algorithms and also provides the highest recognition rate using only 18 features .

6. Evaluation of ANN

Figure 3 shows the result of ANN detection in KDD dataset in each features. The column show detection rate in percentage and rows show the number of features in KDD dataset. It can be clearly seen that the amount of KDD dataset features are 41. In the detection rate part, the minimum of detection rate is 81.1 % of detection. On the other hand, the highest detection rate for ANN in KDD dataset is 97.8%. ANN by KDD dataset in the feature number 40 can get the minimum result of detection but by the feature number 23 of KDD dataset can find best result as mention in top. Although ANN can find good result in features number 3 (96.4%), 12 (94.8%), 32 (91.78%) and 36 (92.48%), still not the high as feature number 23 (97.8%). It can clearly be seen that there has been a large increase in the number of detection rate between feature number 23 and 40 (increase 16.7 % of detection).

Figure of 4 shows the Process of ANN False detection in KDD dataset with each features The column show false detection rate in percentage and rows show the number of features in KDD dataset. It can be clearly seen that the amount of KDD dataset features are 41. In the false detection rate part, the minimum of false detection rate is 0.059% of detection. On the other hand, the highest false detection rate for ANN in KDD dataset is 0.34%. ANN by KDD dataset in the feature number 23 can get the minimum result of false detection but by the feature number 13 of KDD dataset can find highest result of false detection as mention in top. Although ANN can find good result of false detection in features number 3 (0.06%) and 12 (0.096%), still not the low as feature number 23 (0.059%). It can clearly be seen that there has been a large decrease in the number of false detection rate between feature number 23 and 13 (increase 0.281 % of false detection). To sum up, as these 2 figure shows, ANN can achieve 97.8% of detection as the best result for detection rate and in false detection can achieve 0.059 as best result for false detection in feature number 23 in KDD dataset.

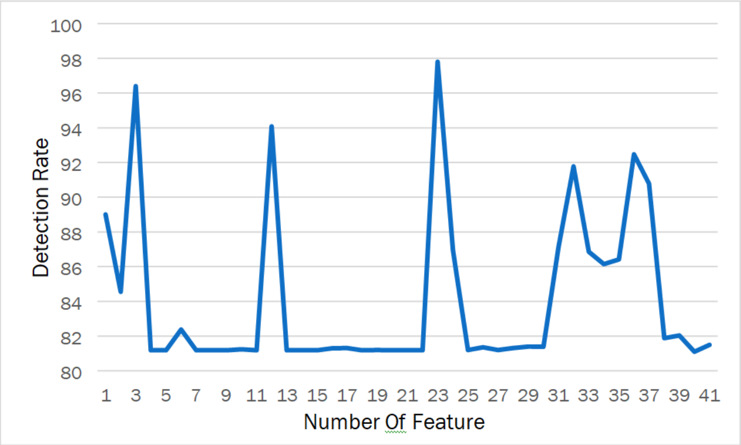

Figure 3 Result of ANN detection in KDD dataset

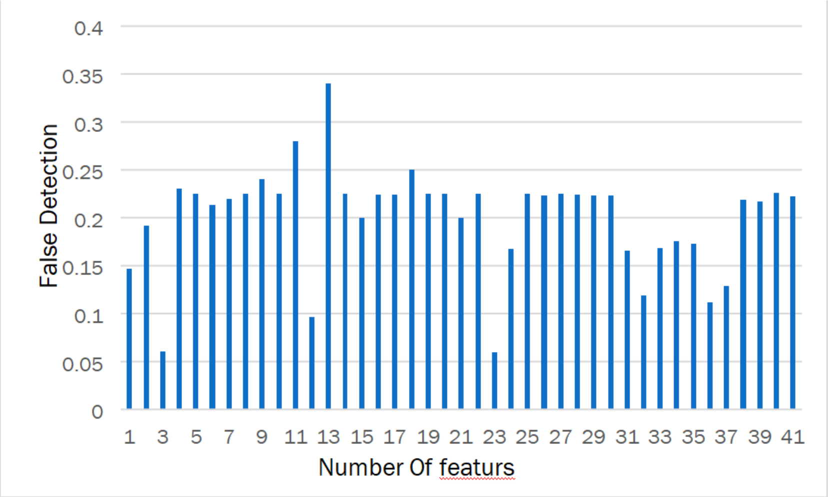

Figure 4 Process of ANN False detection in KDD dataset

7. Evaluation of GA-ANN

It shows the detection result of GA-ANN on KDD database in each feature. The column shows the recognition rate in percentage and the rows show the number of features of the KDD database. It can be clearly seen that the feature value of the KDD database is 41. On the other hand, the highest detection rate for GA-ANN in the KDD database is 99. Although GA-ANN can achieve good results in feature number 6 (98. It can be clearly seen that there is a large increase in the number of detection rates between feature number 1 and there have been 18 (increase 1. 4 tries to show the performance of GA in ANN result to improve the detection rate of ANN in KDD database. In IDS with ANN classifier algorithm, ANN results are not completely satisfactory because it cannot achieve high rate The detection rate is achieved and needs to be changed. 20 shows the detection rate of ANN with GA optimization. In this study, after installing GA to support ANN, the system can achieve a detection rate of 98.98% in feature number 18. It shows that the system can achieve a recognition rate of 98.98% with only 18 features, so the GA can also minimize the number of features in the machine learning test , GA-ANN can reach a recognition rate of 98 with only 18 features. Figure 3 shows the result of GA-ANN in KDD Database and in figure 6 GA-ANN comparison with ANN in detection rate of KDD database.

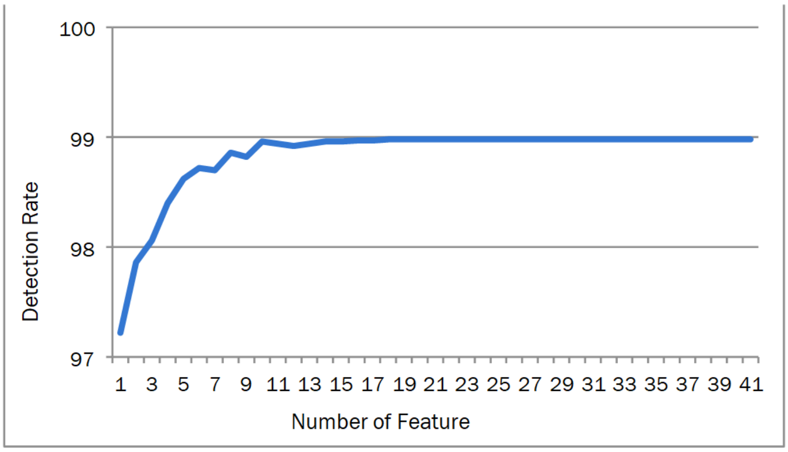

Figure 5 Result of GA-ANN in KDD

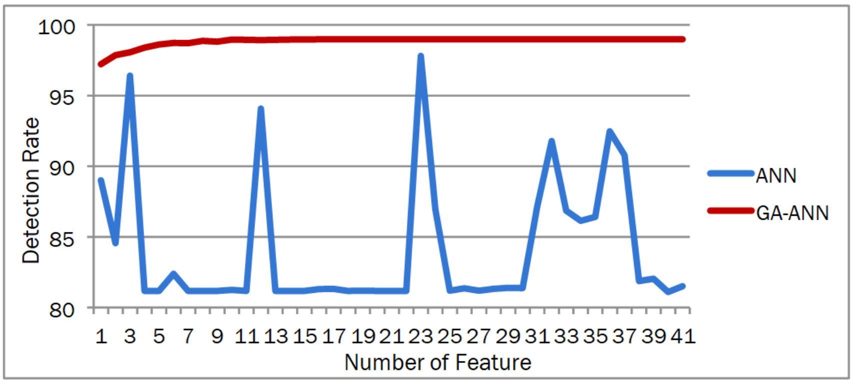

Figure 6 GA-ANN comparison with ANN in detection rate

8. Comparison of Algorithms

The main problem of the previous methods in the intrusion detection system is that their detection rate is not good enough and the false detection rate is also high. To improve IDS problems, GA-ANN models have been used in this research to improve the performance of intrusion detection systems by increasing the detection rate and reducing false detection. The results of other research and the results of GA-ANN used in this research are shown in Table 3, which compares GA-ANN and other detection algorithms of other researchers on the KDD database in each feature. The column shows the recognition rate in percentage and the rows show the number of features of the KDD database. Table 4 shows the process of GA-ANN with other researchers' algorithms in false detection rate on KDD database **.**

Table 4. The result of detection rate and false detection rate with the optimal number of features

Algorithm	**Detection Rate**	**False Alarm**	**Number of Feature**
SICLN	97%	1.05%	20
ICLN	97%	2.03%	24
PNN	97%	0.83%	32
FC-ANN	96%	0.41%	41
FR2	98%	1.03%	31
LCFS	90%	0.64%	21
ANN	97.8%	0.059%	41
GA - ANN	98.98%	0.057%	18

One of the most obvious problems in investigating intrusion detection systems is the number of detection features. Because most features require a lot of computer memory space. To optimize the number of these features, the optimization technique (GA) was used to improve the performance of the systems. To explain more, researchers use more than 20 features for diagnosis. But by optimizing the intrusion detection system, the optimal number of features was determined. Using fewer diagnostic features has advantages, including the need for less computer memory space as well as increased system diagnostic speed. Accordingly, machine learning methods such as ANN are first used as intrusion detection systems, and then optimization techniques such as GA are linked to the systems to find the optimal number of feature detection. Meanwhile, optimal models such as GA-ANN have been developed and tested in this research. The main goal of these models is to improve the recognition rate by optimal pattern recognition. By comparing the results of the GA-ANN model used in the current research, the results of the detection rate with the optimal number of features from a total of 18 features in Table 6. 8 is shown. In the research, it is evident that the developed and developed models provide a recognition rate of approximately 98.98% with fewer features than ever seen before .

9. Conclusions

Anomaly detection can detect unknown attacks based on audit, suffer from the disadvantages of high false alarm and is limited by training data, IDS is implemented using some machine learning techniques that protect the network and host from intruders before launching an attack. Protocols used by packages using machine learning techniques for intrusion detection can automatically build the model based on the training database, which includes data samples that can be described using a set of attributes (attributes) and associated labels. Various machine learning techniques are used and built as intrusion detection systems such as neural network (ANN) and genetic algorithm (GA). To achieve low false detection percentage in KDD data set to improve false alarm or error of intrusion detection system by comparing other mentioned algorithms could perform better in this problem of intrusion detection. This new machine learning algorithm (GA-ANN) has been applied and tested on small systems and simulations.

REFERENCES

[1] Thakkar, A., and Lohiya, R.: 'A review on machine learning and deep learning perspectives of IDS for IoT: recent updates, security issues, and challenges', Archives of Computational Methods in Engineering, 2021, 28, (4), pp. 3211-3243

[2] Sun, P., Liu, P., Li, Q., Liu, C., Lu, X., Hao, R., and Chen, J.: 'DL-IDS: Extracting Features Using CNN-LSTM Hybrid Network for Intrusion Detection System', Security and communication networks, 2020, 2020, (1), pp. 8890306

[3] Khraisat, A., and Alazab, A.: 'A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges', Cybersecurity, 2021, 4, pp. 1-27

[4] Kilincer, I.F., Ertam, F., and Sengur, A.: 'Machine learning methods for cyber security intrusion detection: Datasets and comparative study', Computer Networks, 2021, 188, pp. 107840

[5] Hosseini, S., and Zade, B.M.H.: 'New hybrid method for attack detection using combination of evolutionary algorithms, SVM, and ANN', Computer Networks, 2020, 173, pp. 107168

[6] Choraś, M., and Pawlicki, M.: 'Intrusion detection approach based on optimised artificial neural network', Neurocomputing, 2021, 452, pp. 705-715

[7] Sundaram, A.: 'An introduction to intrusion detection', Crossroads, 1996, 2, (4), pp. 3-7

[8] Wang, K., and Stolfo, S.J.: 'Anomalous payload-based network intrusion detection', in Editor (Ed.)^(Eds.): 'Book Anomalous payload-based network intrusion detection' (Springer, 2004, edn.), pp. 203-222

[9] Garcia-Teodoro, P., Diaz-Verdejo, J., Maciá-Fernández, G., and Vázquez, E.: 'Anomaly-based network intrusion detection: Techniques, systems and challenges', computers & security, 2009, 28, (1), pp. 18-28

[10] Chen, T., and Sun, M.: 'Network intrusion detection system', in Editor (Ed.)^(Eds.): 'Book Network intrusion detection system' (Google Patents, 2009, edn.), pp.

[11] de Boer, P., and Pels, M.: 'Host-based intrusion detection systems', Amsterdam University, 2005

[12] Vokorokos, L., and Baláž, A.: 'Host-based intrusion detection system', in Editor (Ed.)^(Eds.): 'Book Hostbased intrusion detection system' (IEEE Press, 2010, edn.), pp. 32-36

[13] Kothari, S., Parmar, H., Das, E., Panda, N., Ahmed, A., and Marchang, J.: 'Host Based Intrusion Detection System', in Editor (Ed.)^(Eds.): 'Book Host Based Intrusion Detection System' (ASME Press, 2011, edn.), pp.

[14] Kruegel, C., and Toth, T.: 'Using decision trees to improve signature-based intrusion detection', in Editor (Ed.)^(Eds.): 'Book Using decision trees to improve signature-based intrusion detection' (Springer, 2003, edn.), pp. 173-191

[15] Depren, O., Topallar, M., Anarim, E., and Ciliz, M.K.: 'An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks', Expert systems with Applications, 2005, 29, (4), pp. 713722

[16] Whitman, M., and Mattord, H.: 'Principles of information security' (Cengage Learning, 2011. 2011)

[17] Axelsson, S.: 'Intrusion detection systems: A survey and taxonomy', in Editor (Ed.)^(Eds.): 'Book Intrusion detection systems: A survey and taxonomy' (Technical report, 2000, edn.), pp.

[18] Kalekar, A., Kshatriya, N., Chakranarayan, S., and Wadekar, S.: 'Real Time Intrusion Detection System using Machine Learning', in Editor (Ed.)^(Eds.): 'Book Real Time Intrusion Detection System using Machine Learning' (ESRSA Publications, 2014, edn.), pp.

[19] Münz, G., Li, S., and Carle, G.: 'Traffic anomaly detection using k-means clustering', in Editor (Ed.)^(Eds.): 'Book Traffic anomaly detection using k-means clustering' (2007, edn.), pp.

[20] Dönmez, P.: 'Introduction to Machine Learning, by Ethem Alpaydın. Cambridge, MA: The MIT Press 2010. ISBN: 978-0-262-01243-0. $54/£ 39.95+ 584 pages', Natural Language Engineering, 2013, 19, (02), pp. 285288

[21] Aickelin, U., Greensmith, J., and Twycross, J.: ‘Immune system approaches to intrusion detection–a review’: ‘Artificial Immune Systems’ (Springer, 2004), pp. 316-329

[22] Abadeh, M.S., Habibi, J., and Lucas, C.: ‘Intrusion detection using a fuzzy genetics-based learning algorithm’, Journal of Network and Computer Applications, 2007, 30, (1), pp. 414-428

[23] Tsai, C.-F., Hsu, Y.-F., Lin, C.-Y., and Lin, W.-Y.: ‘Intrusion detection by machine learning: A review’, Expert Systems with Applications, 2009, 36, (10), pp. 11994-12000

[24] Abraham, A., and Grosan, C.: ‘Evolving intrusion detection systems’: ‘Genetic Systems Programming: Theory and Experiences’ (Springer, 2006), pp. 57-79

[25] Amiri, F., Yousefi, M.R., Lucas, C., Shakery, A., and Yazdani, N.: ‘Mutual information-based feature selection for intrusion detection systems’, Journal of network and computer applications, 2011, 34, (4), pp. 1184-1199

[26] Lei, J.Z., and Ghorbani, A.A.: ‘Improved competitive learning neural networks for network intrusion and fraud detection’, Neurocomputing, 2012, 75, (1), pp. 135-145

[27] Kala, T.S., and Christy, A.: ‘An intrusion detection system using opposition based particle swarm optimization algorithm and PNN’, in Editor (Ed.)^(Eds.): ‘Book An intrusion detection system using opposition based particle swarm optimization algorithm and PNN’ (IEEE, 2019, edn.), pp. 184-188

[28] Kshirsagar, D., and Kumar, S.: ‘A feature reduction based reflected and exploited DDoS attacks detection system’, Journal of Ambient Intelligence and Humanized Computing, 2022, 13, (1), pp. 393-405

[29] Thakkar, A., and Lohiya, R.: ‘A review of the advancement in intrusion detection datasets’, Procedia Computer Science, 2020, 167, pp. 636-645

[30] Alves da Silva, A.P., and Falcão, D.M.: ‘Fundamentals of genetic algorithms’, Modern Heuristic Optimization Techniques: Theory and Applications to Power Systems, 2008, pp. 25-42