Ilorin Journal of Computer Science and Information Technology
Vol. 5, No. 1 (2022)

ISSN: 2141-3959 *(print)*

A Metaheuristic Approach to Network Intrusion Detection

Modinat A. Mabayoje [1*], Kayode S. Adewole[2], Omonigho E. Ekeruvwe[3], Jumoke F. Ajao[4], Akinyemi O. Akinrotimi[5], Abdullateef O. Balogun[6].

[1.2.3,5.6]*Department of Computer Science, University of Ilorin, Ilorin, Nigeria. mabayoje.ma@unilorin.edu.ng, adewole.ks@unilorin.edu.ng, ekeruvweomonigho@gmail.com, akinrotimi2015@gmail.com, balogun.ao1@unilorin.edu.ng*
[4]*Kwara State University, Malete, Nigeria. jumoke.ajao@kwasu.edu.ng*

*Corresponding author

Abstract

Content: *It is crucial to avoid intrusion in networks; hence, a developing and intrusion detection system that used a strong mechanism for detecting intrusions is important. Several studies have been conducted in the domain of intrusion detections. However, some of them suffer from high false alarms, in terms of the use of a raw dataset with redundancy.* ***Objective:*** *This paper, therefore, proposes a multi-level dimensionality reduction framework that is based on meta-heuristic optimization and Principal Component Analysis (PCA).* ***Method:*** *In this research, PCA was applied for feature extraction. Genetic Algorithm and Particle Swarm Optimization, that is GA-PSO, algorithms were utilized for feature selection to extract the most discriminative features to develop intrusion detection model. In the classification phase, both Artificial Neural Network (ANN) and Support Vector Machine (SVM) algorithms were used to develop intrusion detection, using kddcup.data_10_percent dataset.* ***Result:*** *Experimental results reveal that the proposed framework brought about an accuracy of 99.7% and ROC of 99.9%, while the time required building model is 0.23 seconds.* ***Conclusion:*** *To a very high extent, incidences of high false alarm are allayed through the GA-PSO induced feature selection method.*

keywords*: Genetic Algorithm, Principal Component Analysis, Particle Swarm Optimization, Support Vector Machine, Intrusion Detection System.*

1. Introduction

Wireless Sensor Networks (WSNs) are defenseless against several security threats because of reasons like their weak nature, dynamism, fault-tolerance, scalability, and their scattered nature. Nowadays WSNs are increasingly playing important roles in the economy and as such become prime targets of intruders (Rezaeipanah, Mojarad & Sechin Matoori, 2021). In recent researches about security threat against WSNs, Principal Component Analysis (PCA) has been utilized for feature selection in which features are chosen at some level consisting of the main principal components. There exists the tendency to omit a few cogent characteristics and to incorporate insignificant features in the feature subset amid this procedure. The technique described may not be viable because some important characteristics which may be more delicate and useful for classification (Xue et al., 2012). An adaptation is, thus, necessary with respect to this issue. Particle Swarm Optimization (PSO) is proposed and executed for ideal element choice. Artificial Neural Network **(**ANN), which simulates the way information is analyzed and analyzed the brain and Support Vector Machine (SVM), (which performs classification, regression, and outer detection), given its proven effectiveness in detecting diverse types of anomalies and intrusions (Mohammadi, 2021; Kocher & Kumar, 2021) is utilized for grouping. PSO is a powerful and productive worldwide search method.

A fitting algorithm that addresses feature selection issues. It is simpler to develop because only few parameters are required to search for large spaces. Also, it is less expensive computationally and provides better representation.

Private communication on the internet and some other system is usually exposed to threat of intrusion and abuses. As such, intrusion detection systems have turned into an essential part of computer resources and system security. Different methodologies are being used in intrusion detection. However, none of the systems so far is without its faults.

A lot of previous works on intrusion detection and prevention have not been able to seriously address the issue of multi-level feature selection. Existing techniques suffer a few issues like high computational cost, the complexity of classifier architecture, and higher memory use (Kumar, 2021). Researches have shown that PCA has been utilized for feature selection in which features are chosen at some level of the most crucial components. It is possible to miss a few important features and to incorporate insignificant features in feature subset amid this procedure. This technique may not be viable because of the negligence of certain features which may be more delicate and important to the classifier.

PSO, ANN and SVM were utilized for grouping. PSO is a powerful and productive worldwide search method which is simpler to use because only few parameters are needed. It is capable of searching spaces that are very wide. It is also computationally inexpensive and has better representation. In recent times PSO is being used in developing intrusion detection mechanisms for Internet of Things (IoT) applications so as to foster security, integrity and reliability (Liu et al., 2021; Kan et al., 2021). Genetic algorithm is connected to search the principal space for the feature subset selection (Metawa et al., 2017; Mirjalili, 2019). This strategy by-passes previous approaches by which genetic algorithm yet has minor shortcomings like incapability of solving variant problems, and the inability to discover global optimum. Hence, optimal feature selection based on the meta-heuristics approach is considered important to enhance the classifier performance.

The remaining part of this paper is organized as follows: Section 2 discusses related works in network intrusion detection. Section 3 outlines the methodology, while section 4 discusses the outcomes of the research, and section 5 presents the conclusion.

2. Related Works

Intrusion is known as an arrangement of activities that endeavors to compromise the honesty, privacy, or accessibility of computer system resources (Zamboni, 2001). An intruder, therefore, can be characterized as a system, program or individual who attempts to break into an computer to carry out an illegal activity (Graham, 2000). The act of identifying intrusion; i.e, activities that endeavour to reduce or totally eliminate, the confidentiality, and integrity of the information on a system is often called intrusion detection (Zamboni, 2001). This can be handled through the deployment of an intrusion detection system – that is a gadget or programming application that monitors networks and/or system activities for malicious actions and produces reports (Scarfone et al., 2007).

Gong, Zhong, Yu and Hu (2018) used a genetic algorithm to distinguish strange system practices considering data hypothesis. Some of these practices are related to organized attacks while considering common data between arranged highlights and sort of intrusions; after which these highlights are utilized to determine a direct structure lead and a Generic

Algorithm (GA). The approach of using common data and coming about straight run can be deemed exceptionally compelling due to the lessened multifaceted nature and greater identification rate. However the challenge is that it is seen as just the discrete highlights. Gong, Zulkernine, and Abolmaesumi, (2005) exhibited an execution of GA based technique to deal with Network intrusion detection and indicated programming usage. The method determined an arrangement of grouping rules and uses a help certainty system to judge wellness work. A yet another effort is Abdullah et al. (2009) attempt which demonstrated a genetic algorithm-based execution assessment calculation to network intrusion recognition. The approach utilizes data hypothesis for sifting the activity information.

The exertion of utilizing a genetic algorithm for detecting intrusion can be traced back to Crosbie and Spafford (1995) which connected the different operator innovation and GP to distinguish network abnormalities. The scientists used GP (Genetic Programming) to create bizarre system practices and every operator can screen one parameter of the system review information. The proposed philosophy has a acceptable position when many self-sufficient operators are utilized. However, it has an issue when conveying among the specialists and the procedure can be very difficult if the operators are not appropriately introduced.

In Kolias et al. (2011) a point by point correlation of a few intrusion detection approaches considering swarm insight is exhibited. The fundamental spotlight was on the investigation of the proficiency of each technique in the territory of intrusion recognition.

Anand et al. (2012), proposed a lead-based component determination calculation to evacuate repetitive ascribes, and to choose a touchy list of capabilities that is significant for meddling investigation motor in remote sensor systems. The concentration was just disavowal of administration attacks. Further, the multiclass- Support Vector Machine (SVM) is stretched out for the change of characterization precision. In Hassanzadeh et al. (2014), a checking strategy was used for intrusion recognition in Wireless Mesh Networks (WMN) which depends on two classes: movement freethinker and clever and activity mindful and creative. The outcome shows ideal execution in the rate of intrusion detection and asset utilization in WMN. In (Fawzy et al, 2013) tended to exception detection issues in remote sensor systems and they proposed anomaly detection and grouping instrument in a sensor network. The outcomes indicated a change in the arrangement procedure.

Staudemeyer et al. (2014) offered an element determination instrument which depended on custom component preprocessing. This strategy takes a shot at the difference. The features with higher fluctuation are chosen, while the ones with less change were overlooked. This procedure may miss numerous essential features. Othman et al. (2014), however, proposed a features choice calculation considering record to record travel while Support Vector Machine is connected for the characterization. In the same vein, Halim et al. (2021) proposed a technique which preserves many important information which are related to a dataset being analyzed with few number of bring about a solution to problems of network security and intrusion detection. The study utilized an enhanced Genetic Algorithm (GA)-based feature selection method, named as GA-based Feature Selection (GbFS), in increasing the classifiers' accuracy. It introduces parameter tuning for the GA-based feature selection along with a novel fitness function. The improved GA-based feature selection technique resulting from the study was tested over three benchmark network traffic datasets. A comparison is also performed with standard feature selection methods. Results show that the accuracies improved using GbFS by giving a maximum accuracy of 99.80%.

Onah et al. (2021) developed a Genetic Algorithm Wrapper-Based feature selection and Nave Bayes for Anomaly Detection Model (GANBADM) in a Fog Environment that eliminates extraneous attributes to reduce time complexity while also developing an enhanced model that can predict results with greater accuracy using the Security Laboratory Knowledge Discovery Dataset (NSL-KDD). The results showed that the developed system has a higher overall performance of 99.73% accuracy, with a false positive rate as low as 0.6%. The results reveal that the proposed GANBADM approach performs better than similar approaches captured in the researchers' literature.

Gamal, Abbas, & Sadek (2020) proposed the use of Genetic Algorithm (GA) along with Reinforcement Learning (RL) and threat intelligence to overcome XSS attacks. For validation, the proposed approach is applied on a real dataset of XSS attacks. Results show

better performance of our proposed approach when compared to the approaches reported in the literature. In addition to better performance, our method is not only flexible to changes in XSS payloads, but the results are also more understandable to end users. Moreover, our approach shows improvement when the number of attacks is increased. Alimi et al. (2021) developed an IDS model which relies on the hybridization of particle swarm optimization (PSO) and back-propagation neural network (BPNN) to classify intrusions in water system infrastructure. The PSO is used to optimize the parameters for the BPNN, increasing the efficiency of classification. The iTrust Lab's secure water treatment dataset was for validating the procedure. The experimental outcome revealed that: using prominent classification metrics, the precision results achieved (97% accuracy and 98.7%) using the developed BPNN-PSO model is better compared to those obtained using other methods including models from related works. The proposed model is therefore deemed to be able to meet the requirements of cyberattacks and intrusions detection in practical water distribution infrastructure.

2.1 Theoretical Background

This section looks at the genetic algorithm and other related concepts applied to intrusion detection, as well as the parameters in genetic algorithm.

2.1.1 Genetic Algorithm

A Genetic Algorithm (GA) is a programming procedure that imitates natural development as a critical thinking technique. It depends on Darwinian's guideline of advancement and survival of the fittest to streamline a populace of hopeful arrangements towards predefined wellness. GA utilizes a development and common choice that uses a chromosome-like information structure and advance the chromosomes utilizing choice, recombination and transformation administrators. It is a class of computational models given standards of advancement and characteristic determination. These calculations change over the issue in a particular space into a model by utilizing a chromosome-like information structure and develop the chromosomes utilizing determination, recombination, and transformation administrators. The scope of the applications that can make utilization of the genetic algorithm is very wide.

2.1.2 Principal Component Analysis

Principal Component Analysis (PCA) is a measurement method that is ordinarily utilized for information examination. It is an extremely helpful strategy for feature selection. The PCA is connected to change raw features into foremost features to ensure that the features are all the more unmistakable and their significance envisioned. This method has been utilized in various areas. In this system, the features are chosen based on eigenvalues. The features with greater eigenvalues are chosen, while the features with low eigenvalues are disregarded.

2.1.3 Particle Swarm Optimization

Particle Swarm Optimization (PSO) is an algorithm for stochastic optimization method that is based on swarm. It was developed by Eberhat and Kenedy (1995). As a populace based system, PSO simulates animals' social behaviour. The main focus of PSO algorithm is closely related to two kinds of research, including an evolutionary related algorithm which, like PSO, makes use of swarm mode by which it simultaneously searches wide region within the solution space of the optimized objective function. Likewise, the artificial life-related algorithm as discussed in the next subsection below.

2.1.4 Artificial Neural Network

Artificial Neural Network (ANN), also simply referred to as neural networks, are systems of computing that are inspired by the biological neural network that constitutes the brains of animals (van den Bergh, 2001) (Basheer, and Hajmeer, 2020). In practice, ANN refers to the part of artificial intelligence which simulates the functioning of the human brain. Processing units make up ANNs which also consist of inputs and outputs.

2.1.5 Support Vector Machine

The *support vector machine* (*SVM*) is an administered *machine* learning design which uses classification algorithms for two-group classification problems. After giving SVM model sets of labelled training data for each category, they can group new text, so that a text classification problem is being worked on. More specifically, an SVM builds a hyperplane or set of hyperplanes in an infinite-dimensional space, which can be used for classification, or other functions like detection of outliers. Naturally, a good separation is derived by the hyperplane with the largest distance to the nearest training-data point of any class, also known as functional margin. This is so because generally, the generalization error of the classifier gets lower as much as the margin gets larger (Trevor et al., 2008).

3. Methodology

Figure 1 shows the methodology employed in this paper. Each of the components is discussed in the subsequent sections.

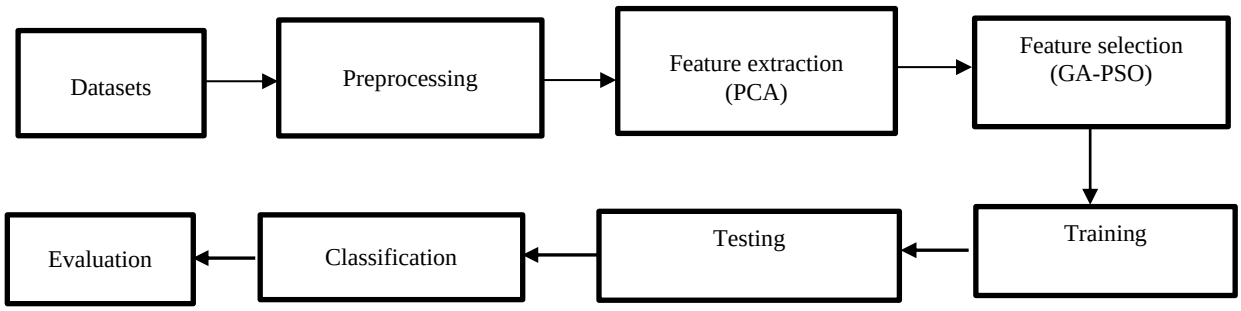

Figure 1: Architecture of proposed system

3.1 Dataset

Ten percent of the KDD 99 intrusion detection datasets were utilized for the implementation of our algorithm which depends on the 1998 DARPA activity. This in turn gives originators of intrusion detection systems (IDS) with a benchmark upon which assess distinctive systems lies. The dataset can be downloaded from this link https://www.kdd.org/kdd-cup/view/kdd-cup-1999/Data.

3.2 Preprocessing

Preprocessing of raw features is vital since they confound the classifier which brings about false cautions. Also, a couple of representatives feature increment computational and memory assets and are unexploited for the grouping methods. The raw list of capabilities from the KDD Cup dataset is communicated by:

$rf=\{rx_1,rx_2,rx_3,rx_4,\ldots,rx_l\}$, (1) where r=2, which shows that there are 2 features in the raw dataset. The emblematic features are disposed of from the raw list of capabilities as these features increment overheads with no advantages in the learning process.

3.3 Feature Extraction using Principal Component Analysis (PCA)

PCA is an approach that is ordinarily used for information examination. It is an extremely helpful strategy for feature selection. The PCA is connected to change raw features into foremost features with the goal that the features are all the more unmistakable with envisioned significance. This method has been utilized in various areas. In this system, the features are chosen based on eigenvalues. The features with the higher eigenvalues are chosen and the features with bringing down eigenvalues are disregarded. The Flow of PCA for feature extraction is illustrated with Figure 2.

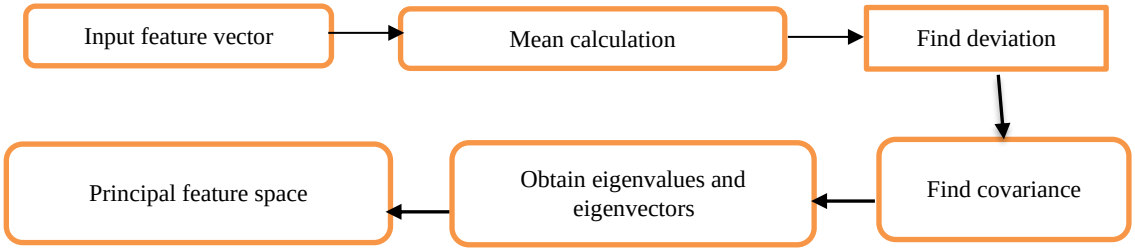

Figure 2: Flow of PCA for feature extraction

3.4 Feature Selection using Genetic Algorithms

Genetic algorithms (GA) are a piece of transformative figuring, which is a quickly developing region of man-made reasoning. GAs seeks calculations in light of the standards of characteristic choice and hereditary qualities. In GAs, a populace of chromosomes shows hopeful arrangements of the issue. Every chromosome is spoken to, with settled length bits. The essential populace of chromosomes is made by conveying 0 s discretionarily. This circulation depended on and large task of 0 s. In this encoding plan, each chromosome is a touch of strings (0 s) whose length is figured by the number of foremost segments in the essential space. The bit with 1 is chosen and a bit with 0 isn't chosen in this encoding plan. Every chromosome demonstrates a competitor arrangement or a subset of the main parts. The populace develops via looking through the ideal arrangement utilizing hereditary administrators. The

GAs has two major problems, which are: neighbourhood optima and high computational cost (Hamamoto, 2018). The stream of GA for include determination appears in Figure 3. The calculation connected is portrayed.

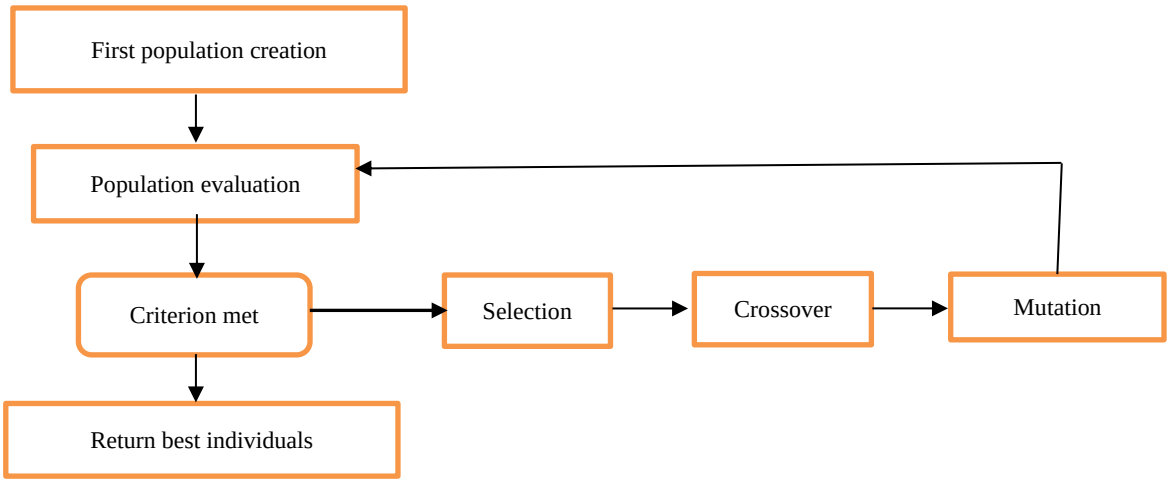

Figure 3: Flow of GA for feature selection

3.5 Feature Selection using Particle Swarm Optimization (PSO)

The Particle Swarm Optimization (PSO) is a populace-based system created by Eberhart and Kennedy(1995). PSO is an effective and esteemed worldwide hunt method. It is an appropriate calculation to address determination issues due to the accompanying reasons like simple encoding of highlight, worldwide hunt office, being sensible computationally, fewer parameters, and less demanding execution. The PSO is connected to include determination due to the previously mentioned reasons. Its stream for the feature selection process is established in Figure 4 below.

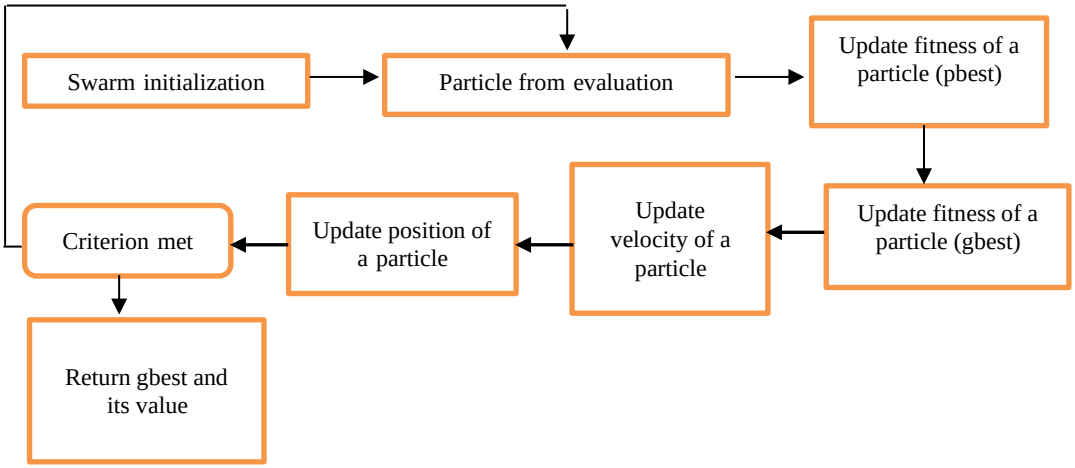

Figure 4: Flow of PSO for feature selection

The vital space is the inquiry space through which a subset of important segments or vital features were investigated and chosen using PSO. The particles speak to applicant arrangements in the pursuit of space particles and frame a populace which is otherwise called a swarm. The swarm of the molecule is created by dispersing 0 s arbitrarily. For each molecule, if the foremost segment is 1, it is chosen and the main part with 0 is disregarded. Along these lines, each molecule demonstrates an alternate subset of central

parts. The particles swarm is introduced haphazardly and afterwards, it moved in the hunt space or central space to look through the ideal subset of features by refreshing its position and speed.

3.6 Classification

This section outlines the classification method which is a technique of grouping a given set of data into classes. This technique can be performed on both structured and unstructured data. It begins with predicting the class of given data points.

3.6.1 Artificial Neural Network

An artificial neural network is made of numerous neurons that are connected as per particular system design (Mukherjee, 1994). The target of the neural system is to change the contributions to significant yields. The outcome is dictated by the attributes of the hubs and the weights related to the interconnections among neurons. By changing the associations between the hubs, the system can adjust to the coveted yields. It looks like the mind in two regards:

1) Knowledge is obtained by the system from its condition through a learning procedure.

2) Interneuron association qualities, known as synaptic weights, are utilized to store the gained information.

3.6.2 Support Vector Machine

The support vector machine utilizes a segment of the information to prepare the system, finding a few help vectors that speak to the preparation information. These help vectors will frame an SVM show. As indicated by this model, the SVM will work with PSO for advancement and highlight subset determination. What's more, it enhances the SVM display. After that SVM is utilized to group a given obscure dataset.

3.6.3 Detection Rate

Detection rate (DR) is calculated as the ratio between the number of correctly detected intrusions and the total number of intrusions, that is:

$$\text{DR} = \frac{\#TruePositive}{\#FalseNegative+\#TruePositive} \quad (2)$$

3.6.4 False Positive Rate

False positive rate (FP) is calculated as the ratio between the numbers of normal connections that are incorrectly classified as intrusions and the total number of normal connections, that is:

$$\text{DR} = \frac{\#FalsePositive}{\#TrueNegative+\#FalsePositive} \quad (3)$$

4. Results Discussion

In the testing stage, for each of the test information, the underlying populace is made to utilize the information and happening change in various features. This populace is contrasted, and every chromosome arranged in the preparation stage. Part of the populace, which is more approximately related to all preparation information than others, is expelled. Hybrid and change happen in the rest of the populace which turns into the number of inhabitants in a new age. The procedure keeps running up to the point that the age estimate reduces to one. The gathering of the chromosome which is the nearest relative of a just surviving chromosome of test information is returned as the anticipated sort.

4.1 Results Based on All Features

This testing work utilized "kddcup.data_10_percent" as preparing dataset and "redressed" as testing dataset. For this situation, the training set comprises of 44,104records among which 3,232 are connection records which are normal and 40,872 are attack connection records, while the test set consists of 55,894 records which include 55036 are normal connection records and 858 are attack connection records. The table below reveals the distribution of each of the intrusion type in the training and testing sets.

Table 1: Classification results based on all features

Class	TP Rate	FP Rate	F-Measure	MCC	Precision	Recall	ROC Area	PRC Area
Normal	0.985	0.073	0.964	0.918	0.945	0.985	0.959	0.928
Attack	0.927	0.015	0.952	0.918	0.979	0.927	0.958	0.961
Weighted Avg	0.959	0.048	0.959	0.918	0.960	0.959	0.959	0.943

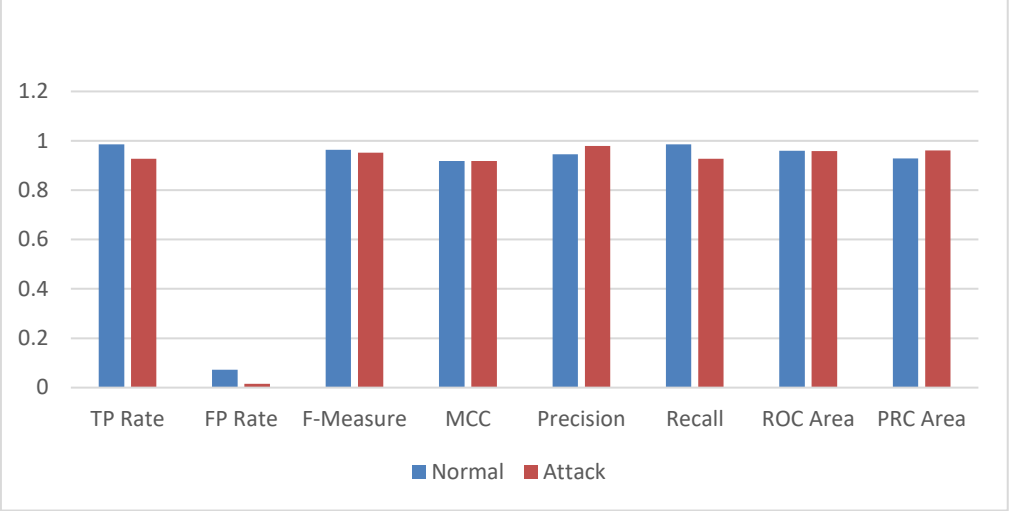

Figure 5: Classification results based on all features

In this classification, performance of different features that could produce a high level of accuracy and also some error to diagnose the dataset is examined. The aggregate time that is required to develop this model is also an important parameter in the comparison of the classification algorithm which is 0.23 seconds.

Table 2: Standard metrics for system evaluation

		Predicted Label	
		Instances	**Percentage (%)**
Actual Class	**Normal**	95908	95.9099 %
	Intrusion	4090	4.0901%
Kappa statistic		0.9166	
Mean absolute error		0.0412	
Root mean squared error		0.201	
Relative absolute error		8.3484%	
Root relative squared error		40.4829%	
Total Number of Instances		99998	

4.2 Results Based on Principal Component Analysis

This testing work utilized "kddcup.data_10_percent" as the preparation dataset, and "redressed" as the testing dataset. For this situation, the training set comprises of 44,104 records among which 419 are normal connection records and 43,685 are attack connection records, while the test set contains 55,894 records among which 55,617 are normal connection records and 277 are attack connection records. The table below shows the distribution of each intrusion type in the training and testing set.

Table 3: Classification results based on PCA using multi-class classifier

Class	**TP Rate**	**FP Rate**	**F-Measure**	**MCC**	**Precision**	**Recall**	**ROC Area**	**PRC Area**
Attack	0.990	0.005	0.992	0.986	0.994	0.990	0.997	0.996
Normal	0.995	0.010	0.994	0.986	0.993	0.995	0.997	0.996
Weighted Avg	0.993	0.007	0.993	0.986	0.993	0.993	0.997	0.996

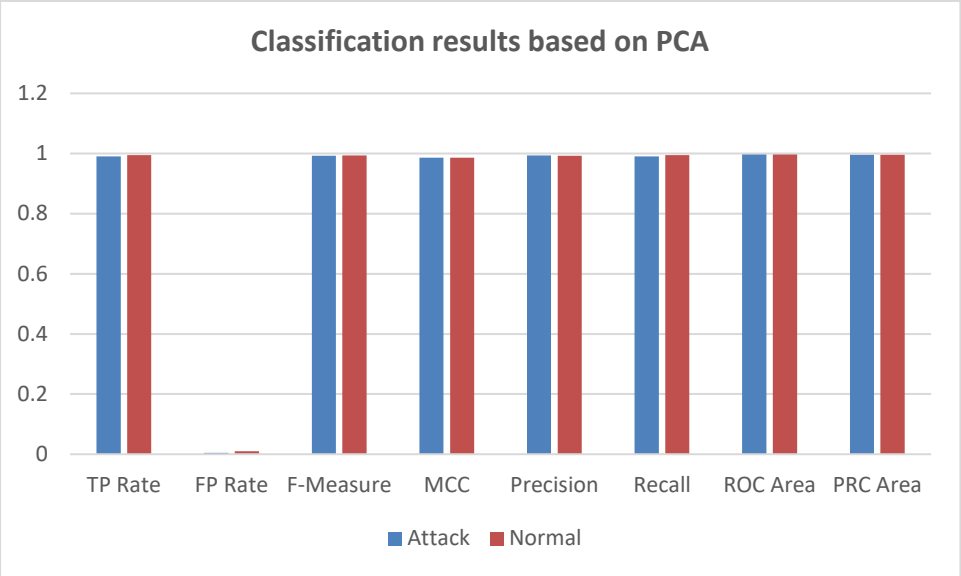

Figure 6: Classification results based on principal component analysis

In this classification, we examine the performance of principal component using multi-class identifier that could generate accuracy and some error to diagnosis the data set. The total time required to build this model is also a crucial parameter in comparing the classification algorithm which is 19.16 seconds.

Table 4: Standard metrics for system evaluation

		Predicted Label	
		Instances	**Percentage (%)**
Actual Class	**Normal**	99302	99.304 %
	Intrusion	696	0.696 %
Kappa statistic		0.9859	
Mean absolute error		0.021	
Root mean squared error		0.0858	
Relative absolute error		4.2648 %	
Root relative squared error		17.2826 %	
Total Number of Instances		99998	

4.3 Results Based on Genetic Algorithm

This testing work utilized "kddcup.data_10_percent" as preparing dataset, and "redressed" as testing dataset. For this situation, the training set comprises of 44,104 records among of which 303 are normal connection records and 43,801 are attack connection records. The test set contains 55,894 records among of which 55,855 are normal connection records and 39 are attack connection records. The table below reveals the distribution of each intrusion type in the training and testing set.

Table 5: Classification results based on GA using Bagging

Class	**TP Rate**	**FP Rate**	**F-Measure**	**MCC**	**Precision**	**Recall**	**ROC Area**	**PRC Area**
Attack	0.993	0.001	0.996	0.993	0.999	0.993	0.999	0.999
Normal	0.999	0.007	0.994	0.986	0.995	0.999	0.999	0.999
Weighted Avg	0.997	0.004	0.997	0.993	0.997	0.997	0.999	0.999

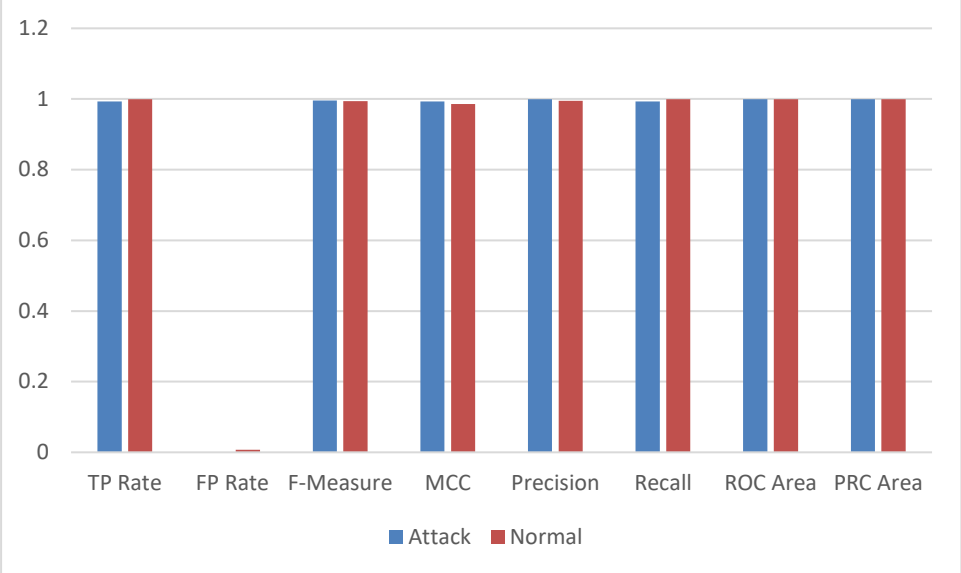

Figure 7: Classification results based on GA

In the classification, the performance of a genetic algorithm using bagging that could generate accuracy and some error to diagnose the data set were examined. The total time required to build this model is also a crucial parameter in comparing the classification algorithm which is 10.08 seconds.

Table 6: Standard metrics for system evaluation

		Predicted Label	
		Instances	**Percentage (%)**
Actual Class	**Normal**	99656	99.658 %
	Intrusion	342	0.342 %
Kappa statistic		0.9931	
Mean absolute error		0.0062	
Root mean squared error		0.0556	
Relative absolute error		1.2497 %	
Root relative squared error		11.1959 %	
Total Number of Instances		99998	

4.4 Results Based on Generic Algorithm and Particle Swarm Optimization (GA-PSO)

This testing work utilized "kddcup.data_10_percent" as preparing dataset and "redressed" as testing dataset. For this situation, the training set comprises of 44,104 records, among of which 305 are normal connection records and 43,799 are attack connection records. The test set contains 55,894 records, among of which 55,831 are normal connection records and 63 are attack connection records. The table below reveals the distribution of each intrusion type in the training and testing set.

Table 7: Classification results based on Generic Algorithm and Particle Swarm Optimization (GA-PSO)

Class	**TP Rate**	**FP Rate**	**F-Measure**	**MCC**	**Precision**	**Recall**	**ROC Area**	**PRC Area**
Attack	0.993	0.001	0.996	0.993	0.999	0.993	0.999	0.998
Normal	0.999	0.007	0.997	0.993	0.995	0.999	0.999	0.996
Weighted Avg	0.996	0.004	0.996	0.993	0.996	0.996	0.999	0.997

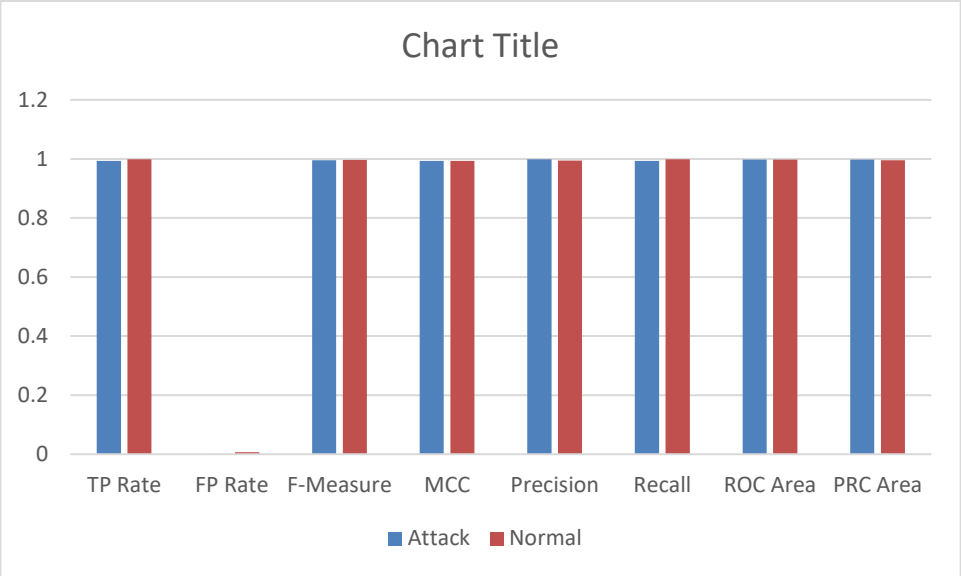

Figure 7: Classification results based on GA-PSO

In this classification, the performance of a genetic algorithm using bagging that could generate accuracy and some error to diagnose the data set were examined. The total time required to build this model is also a crucial parameter in comparing the classification algorithm which is 2.31 seconds.

Table 8: Standard metrics for system evaluation

		Predicted Label	
		Instances	**Percentage (%)**
Actual Class	**Normal**	99630	99.632 %
	Intrusion	368	0.368 %
Kappa statistic		0.9925	
Mean absolute error		0.0061	
Root mean squared error		0.0588	
Relative absolute error		1.2366 %	
Root relative squared error		11.8515%	
Total Number of Instances		99998	

4.5 Models Comparison

In this study, the performance of different classification methods that could generate accuracy and some error to diagnose the data set were examined. In order to make the evaluation of the system simple, besides the classical accuracy measure, the two standard metrics of detection rate and false positive rate were employed for network intrusions detection evaluations.

According to Table 7 above, which shows that the proposed framework achieved an accuracy of 99.7% and ROC of 99.9%, the time taken to build model is 0.23 seconds based on GA model. The total time required to build the model is also a crucial parameter in comparing the classification algorithm.

5. Conclusion

This paper talked about a strategy of utilizing genetic algorithm in developing intrusion detection systems. A short diagram of Intrusion Detection System (IDS), genetic algorithm and related recognition strategies were examined. To actualize and measure the execution of the new system, this study utilized the standard KDD99 benchmark dataset and got sensible detection rate. To quantify the wellness of a chromosome, the study utilized the standard deviation condition with remove. On the off chance, where we can utilize a superior

condition or heuristic in this detection procedure, we trust that the identification rate and process will enhance an awesome degree. Particularly, false positive rate will, most likely, be much lower. Subsequently, the designed intrusion detection system can be enhanced with the assistance of more measurable investigation with better and more intricate conditions. The system design is additionally presented. The variables influencing the GA are attended to in detail. This usage of genetic algorithm is extraordinary as it thinks about both fleeting and spatial data of system associations amid the encoding

of the issue; hence, it ought to be more useful for intrusion detection. The results returned by this experiment on the proposed framework show that the accuracy is 0.946 and its ROC area is 0.959. Hence, the conclusion is drawn from the results obtained that the framework designed is efficient compared to existing techniques.

Future work is expected to combine the making of a standard test information collection for the genetic algorithm proposed in this paper, and then apply it to a test situation. The details of parameters to be considered for genetic algorithm ought to be resolved amongst the trials. Consolidating learning from various security sensors into a standard govern base is another promising aspect in this study.

References

Ahmad, I. (2014). Enhancing MLP performance in intrusion detection using optimal feature subset selection based on genetic principal components Applied Mathematics & Information Sciences.

Alimi O., Ouahada K., Abu-Mahfouz A., Rimer S. & Alimi K., (2021)."Intrusion Detection for Water Distribution Systems based on an Hybrid Particle Swarm Optimization with Back Propagation Neural Network," 2021 IEEE AFRICON, 2021, pp. 1-5, doi: 10.1109/AFRICON51333.2021.9570951.

Basheer, I. and Hajmeer, M.N.(2020). Artificial Neural Networks: Fundamentals,

Bobor, V. (2006). "Efficient Intrusion Detection System Architecture Based on Neural Networks and Genetic Algorithms. Department of Computer and Systems Sciences, Stockholm University/Royal Institute of Technology, KTH/DSV.

Brad, L.M., Shaw, M.J. (1996). "Genetic Algorithms with Dynamic Niche Sharing for Multimodal Function Optimization." In Proceedings of IEEE international conference on evolutionary computation (pp. 786-791). IEEE.
Computing, Design, and Application, Journal of microbiological methods, 10.1016/S0167-7012(00)00201-3

Crosbie, M., Spafford, E. (1995). "Applying Genetic Programming to Intrusion Detection". In Working Notes for the AAAI Symposium on Genetic Programming (pp. 1-8). Cambridge, MA: MIT Press.
GAbased approach to network intrusion detection,
http://www.cse.msu.edu/~cse848/2011/Student_papers/Tavon_Pourboghrat.pdf.

Gamal, M., Abbas, H., & Sadek, R. (2020, April). Hybrid approach for improving intrusion detection based on deep learning and machine learning techniques. In The International Conference on Artificial Intelligence and Computer Vision (pp. 225-236). Springer, Cham.

Gong, R. H.,Zulkernine, M. and Abolmaesumi, P., (2005): A software implementation of Gong, Z., Zhong, P., Yu Y., and Hu, W. (2018), ''Diversity-promoting deep structural metric

Graham, R. (2000). FAQ: Network Intrusion Detection Systems. Retrieved from http://www.
robertgraham.com/pubs/network-intrusion-detection. html.

Halim, Z., Yousaf, M. N., Waqas, M., Sulaiman, M., Abbas, G., Hussain, M., ... & Hanif, M. (2021). An effective genetic algorithm-based feature selection method for intrusion detection systems. Computers & Security, 110, 102448.

Hamamoto, A. H., Carvalho, L. F., Sampaio, L. D. H., Abrão, T., &Proença Jr, M. L. (2018). Network anomaly detection system using genetic algorithm and fuzzy logic. Expert Systems with Applications, 92, 390-402.

Iftikhar A. (2015) Feature Selection Using Particle Swarm Optimization in Intrusion Detection, International Journal of Distributed Sensor Networks, article 9. Pp 9, 'Hindawi Limited'

Kan, X., Fan, Y., Fang, Z., Cao, L., Xiong, N., Yang, D., & Li, X. (2021). A novel IoT network intrusion detection approach based on adaptive particle swarm optimization convolutional neural network. Information Sciences, 568, 147-162.

Kennedy, J., Eberhart, R. C., Shi, Y. (1995). Swarm Intelligence. San Francisco, Calif, USA.

Kocher, G. & Kumar, G. (2021). Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges. Soft Comput 25, 9731–9763 (2021). https://doi.org/10.1007/s00500-021-05893-0
learning for remote sensing scene classification,'' IEEE Trans. Geosci. Remote Sens., vol. 56.

Metawa, N., Hassan, M. K., &Elhoseny, M. (2017). Genetic algorithm based model for optimizing bank lending decisions. *Expert Systems with Applications, 80,* 75-82.

‘Mirjalili, S. (2019). Genetic algorithm. In *Evolutionary algorithms and neural networks* (pp. 43-55). Springer, Cham.’

Mohammadi, M., Rashid, T., Karim, S.., Aldalwie, A., Tho, Q.., Bidaki, M., & Hosseinzadeh, M. (2021). A comprehensive survey and taxonomy of the SVM-based intrusion detection systems. Journal of Network and Computer Applications, 178, 102983.
Mukherjee, B., Heberlein, L.T., Levitt, K.N. (1994). “Network Intrusion Detection”. IEEE network, 8(3), 26-41.

Onah, J. O., Abdullahi, M., Hassan, I. H., & Al-Ghusham, A. (2021). Genetic Algorithm based feature selection and Naïve Bayes for anomaly detection in fog computing environment. Machine Learning with Applications, 6, 100156.

Rezaeipanah, A., Mojarad, M., & Sechin Matoori, S. (2021). Intrusion Detection in Computer Networks Through Combining Particle Swarm Optimization and Decision Tree Algorithms. Journal of Business Data Science Research, 1(1), 14-22.

Scarfone, K., Mell,P. (2007). Guide to Intrusion Detection and Prevention Systems (IDPS). (No. NIST Special Publication (SP) 800-94 Rev. 1 (Draft)). National Institute of Standards and Technology.

Sinclair, C., Pierce, L., Matzner, S. (1999). “An Application of Machine Learning to network intrusion detection”. In Proceedings 15th Annual Computer Security Applications Conference (ACSAC'99) (pp. 371-377). IEEE.

Staudemeyer, R., Omlin, C. (2014). Extracting salient features for network intrusion detection using machine learning methods. South African computer journal, 52(1), 82-96.

Wang, L., Xiao, Y. (2006). A survey of energy-efficient scheduling mechanisms in sensor networks Mobile Networks and Applications. Mobile Networks and Applications, 11(5), 723-740.

Xue, B., Zhang, M., Browne, W. N. (2013). Particle swarm optimization for feature selection in classification: a multi-objective approach IEEE Transactions on Cybernetics. IEEE transactions on cybernetics, 43(6), 1656-1671.

Zamboni, D. (2001). Using Internal Sensors for Computer Intrusion Detection. Center for Education and Research in Information Assurance and Security, Purdue University.