Journal of Basic Physical Research ISSN: 2141-8403 Prints, 2141-8411 Online
Vol.14, Issue 1, March, 2025, pp. 150 – 163, Available online at https://journals.unizik.edu.ng/jbpr/

HYBRIDIZATION OF GENETIC ALGORITHM AND ARTIFICIAL NEURAL NETWORK FOR THE DETECTION OF ADVANCED PERSISTENT THREATS

Lateef Caleb Umoru[1], Joseph Adebayo Ojeniyi[2], Christopher Ubaka Ebelogu[3], Moses Olabode Esan[4]
[1&2]Department of Computer Science, Confluence University of Science and Technology, Osara, Nigeria.
[3&4]Department of Computer Science, University of Abuja, Abuja-FCT, Nigeria.
umorulc@custech.edu.ng, 0jeniyija@futminna.edu.ng, christopher@uniabuja.edu.ng, astoundbme@hotmail.com

Abstract
*Advanced Persistent Threat (*APT*) is defined as an attack targeted on organizations for the main purpose of stealing data that are of important in the organization or to cause a particular damage. As the name implies, it is advanced i.e. APT uses different forms of vulnerabilities that are identified within the organization. Attackers are capable of detecting the attacks that have been previously known and therefore the efficiency of these systems is more than the efficiency of the APT detection system. Hence, the need for several artificial intelligence methods to be worked on and proven predictions for the detection of APTs. The paper aims to develop a hybridized technique using Genetic Algorithm (*GA*) and Artificial Neural Network (*ANN*) in the detection of APT. The study imports technical indicators inform of datasets of which is represented by 21 input variables based on 1781 URL of past time spans of different lengths and is collected before the day of prediction of APT. It is used to generate more diverse subsets of input which is then culled down to a manageable number of effective ones by Genetic Algorithm (GA) and passed onto Artificial Neural Network (ANN) to make prediction. At the end, the results show that the highest rate to detect APT is achieved by GA with ANN in comparison to Modified Mutual Information based Feature Selection (MMIFS), Learning Fuzzy Classifier System (LCFS) and Firefly Swarm Algorithm (FFSA) techniques.*

Keywords*: Algorithm, Artificial Neural Network, Genetic algorithm, Hybridized algorithm, APT.*

Introduction

Most of the everyday activities in the modern society show the use of electronic devices and communication networks in form of cyberspace. In this format, both hardware and software suffer attacks or threats which compromise these activities that ranges from provision of public administration services to works that are related to economic aspects, access to information, education, company activities, etc (Rahman., 2024). Due to the high-level attacks that is in constant evolution, and also the evolved technologies and tools being used, today we have a very high sophisticated threats that exploit vulnerabilities in the code, the protocols, the computer systems, the communication networks, etc (Jimmy, 2024). These forms of attacks are called Advanced Persistent Threats (APT).

An advanced persistent threat (APT) is defined as any attack in which an unauthorized user gains access to a system or network and remains there for along period of time without being detected. Advanced persistent threats are highly dangerous fororganizations, as attackers have access to sensitive dataof the organization.

APT attackers use methods that are familiar toenter into any available and target entity's network, but the tools used to enter such targets are not familiar. As the term implies,

these tools that are used by attackers are advanced, needs to be because the attacks are persistent in the network and also for a longer period of time. They remain undetected, slowly expanding their foothold from one system to another in the organization's network and gain meaningful information as they export it to their command and control center in a highly organized fashion (Buchta *et al*., 2024).

Below are some examples or indicators to show an organization is under an APT attack

- ✓ When log-ins late at night are at an increase, or employees cannot access the network.
- ✓ Discovery of a widespread backdoor Trojans. These are used by such attackers to ensure access can be retained, even when a user's login credentials is compromised and breached due to changes in his or her credentials.
- ✓ Flows of data are large and unexpected. These flows should be different from an organization typical baseline.
- ✓ Unexpected data bundles are discovered. This is as a result of aggregate data inside the network in an attempt to move data outside of the network.
- ✓ Cracking of password/ password hashing is detected. The attacks steal password hashes from password-hash-storage databases and create new authentication periods which are not always used in advanced persistent threats.

APTs are performed usually by well-funded attackers that possess resources needed to perform such attacks, and as long as the funding for the work keeps coming, the organization will continue to be affected. The attack can only stop when not detected by the companyor when the funding organization gets all the information required for use. In whichever form it is, considerable damage would have been done to the organization that is a victim of any APT attack, and usually, these attacks causes damage that are irreparable, which is one of the ways the attackers ensure the attacks willnot be detected until all the organization's data falls into the wrong hands. The organization that falls victim of these APT attacks are most often questioned on their failure to detect such attacks and also the fact that they have a knowledge of security measures which include strong intrusion detection and prevention systems that are being implemented in the cyberspace daily. The major aim of an APT attack is not limited to gathering of a target entity's data, but also to remain undetected until when the attack has been removed (Ahmed, 2022).
This paper aims at developing and evaluating the hybridized algorithm using Genetic algorithm/Artificial Neural Network using benign and malware dataset for predictions of APT. Also other techniques for detection of APTs result were used to compare with the results from this paperwork.

Literature Review

Advance Persistent Threat

APT can be defined using the three words for which its coined from and they are:
Advanced: The attackers of APT are most often well- funded and have access totools and methods that are advanced which are required to create an APT attack [Sharma *et al*., 2023].

Persistent: APT attackers are often determined and persistent and will not give up easily. once they enter into the system, their aim is to stay in the system for as long as possible till they achieve their purpose. Their plan is to use several undetected techniques tobypass detection by their target's intrusion detection systems. They work with a "low and slow" approach toensure their success rate is increased.
Threat: APT attacks threats are often on sensitive data loss or infringement on organization critical components or mission. These threats are on a rise to many nation, companies, entities, and organizations that uses the protection systems which are highly advance in protecting their missions and/or data.

National Institute of Standards and Technology (NIST) (Ahmed, 2022) reviewed APT attackers as a group of people who:

(i) are after repeated objectives over along period of time;
(ii) adapts to defenders' or victim's efforts to resist it; and
(iii) is determined to keep the level of interaction constant which is required to execute its purpose.

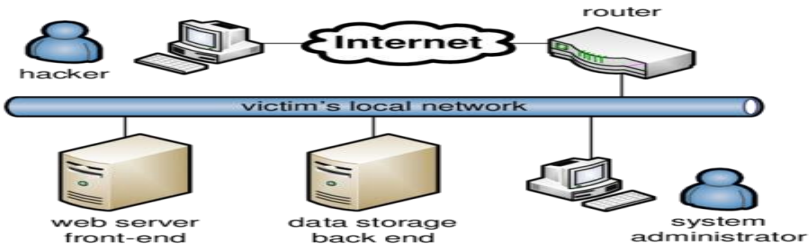


Fig 2.1: Network Topology of an Attacker

Al-Sada (2024) identifies several generic steps in ATPs.

- ✓ The first step is identification of the target and reconnaissance after which social engineering and malware are used to gain access.
- ✓ After this the attacker starts to explore the target network and will extend its access until the attacker has reached his goal.

Al-Sada (2024) shows a similar order of events in steps of advanced persistent threat. The eight steps defined by TNO are more detailed and are more in line with attack related literature. These eight steps will therefore be used as general steps of APTs.
The eight steps defined by TNO each have a different purpose in an APT. Some of the steps are overlapping and steps can be very similar to earlier steps having only a distinctly different location. A general description of activities in the eight steps is given below. In this description some of the steps are combined because they often occur simultaneously. They do however have different purposes and are therefore still separate steps.

Step 1; **reconnaissance**: This ensures that the target delivers to the attacker information about the ICT environment and about the people working at the target. (Roy *et al.*, 2022).

Step 2; **gaining access**: The information from step one helps obtains a foothold in the target system through the use of a weakness in the target system and attacker malware strength.

Step 3 and 4; **internal reconnaissance and expanding access**: an attacker obtains a foothold and continues by searching for more information about the internal infrastructureof the target system.

Step 5; **gathering information**: immediately the attacker has acomprehensive outlook of the network structure and locations of the wanted proprietary information, the attacker starts gathering useful information to a location where it can be stored until extraction. (Kim *et al*., 2022).

Step 6; **information extraction**: information gathered in step 5 can be extracted from the network to locations on the internet.

Step 7&8; **control and erasing tracks**: these are done immediately the attacker has gotten the required information and it can be done through the use of controlled malware through a botnet infrastructureor limited detection ability by the attackers.

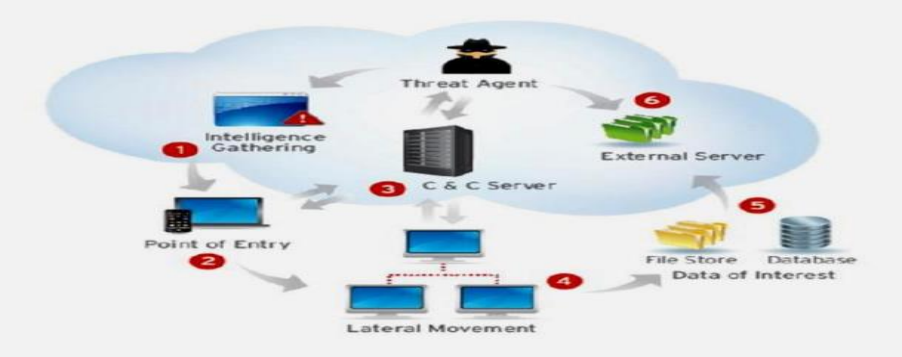


Fig 2.2 Generic steps involved in an APT attack

Machine Learning Language

Machine learning is an area of artificial intelligence (AI) that involves the development and of algorithms and techniques which aid the computers to "learn". The sole aim of Machine Learning research is to extract vital information from data in an automated form using computational and statistical methods. It is very well related to data mining and statistics (Garouani, 2022).

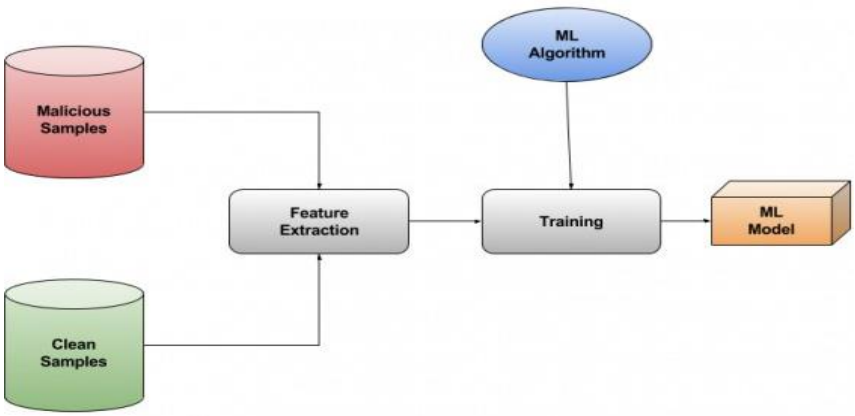


Fig 2.3: Machine Learning Training Cycle

The methods of Machine learning language are of great practical value which can be applied in different areas of application domains and found in areas where its highly not practicable tomanually extract information from data.

In the field of machine learning language, the steps involve the use of subset or setof data selected in a related feature in order to make a model of solution is called a feature selection. When this feature is being used, it is assumed that the data includes some irrelevant and redundant information. Therefore, when it comes tothis type of artificial intelligence toovercoming any situation, the algorithm of feature selection is applied to select relevant information (Halim *et al*., 2021).

Genetic Algorithm as a Machine Language Approach

Genetic algorithm is popular and majorly used in the field of machine learning language. Genetic algorithm is highly exploratory as well as adaptive for work and search that has not been based on natural genetics and evolutions works. The main populationof an individual is extracted by GA in a level of individuals' quality. In addition, a solution can be represented by each one of these individuals for the problem (Alhijawi and Awajan, 2024). Genetic Alogrithm is a parallel algorithm and can find solution to a problem with multi subsets, it is considered to be suitable for IDS. In a way, GA is able to find and search for different subsets problems simultaneously find its solutions. In addition, there is no mathematical derivation in GA and it can reach to the suitable set of solutions for problems. Besides, GA is capable of proposing solution and every single solution with an optimal result. one other capability of the GA is that it has the ability toidentify any new attacks/data from theother attacks that were prreivously performed ones a proper method for Intrusion Detection System, mostly for the detection of attacks which are humanbehavioral based (Liu *et al*., 2021).

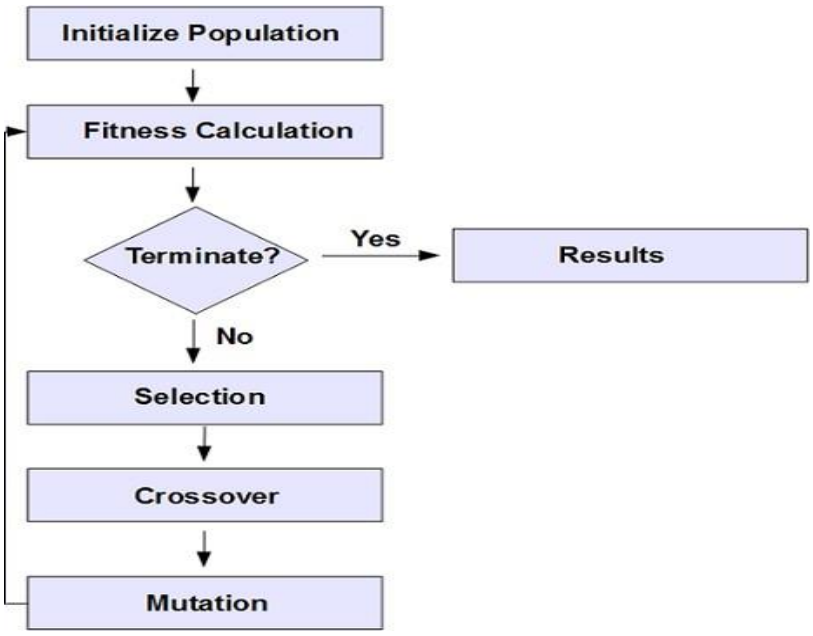


Fig 2.4: Genetic Algorithm Working Model

Artificial Neural Network (ANN) in Machine Learning Techniques

Artificial Neural Network (ANN) is another popular machine learning language that can be used to solve problems on regression and classificationof data. The artificial neural network model shows computation which are inspired by an animal central nervous systems, which has machine learning language ability and patternsrecognition. For instance, in recognition of a neural network, the input data feature may activate a set of input neurons to represent a normal activity or an attack. In general, they are introduced as the systems with neurons that have interconnections and have the ability to compute values out of the input data by feeding the information via the network. There are several advantages for the ANN, but one of them is considered as the most popular one on them and that is its ability to learn from data set observation (Khrisat and Alqadi., 2023). Any task of these tools is to assist the estimation of the methods with the most ideality and the most cost effectiveness for reaching the solutions while they define the distributions of computing or functions of computing. Instead of the entire set of data, a data sample is taken by ANN for reaching the solutions. The mathematical methods considered by the ANN are fairly simple and that is for enhancing the technologies of the existing data analysis. There are three interconnected levels in ANNs. The input neurons are in the first layer. The data is sent by these neurons to the next layer which is the 2nd one and in turn, the 2nd layer will send the outcome neurons to the 3rd layer (Adem, 2022).

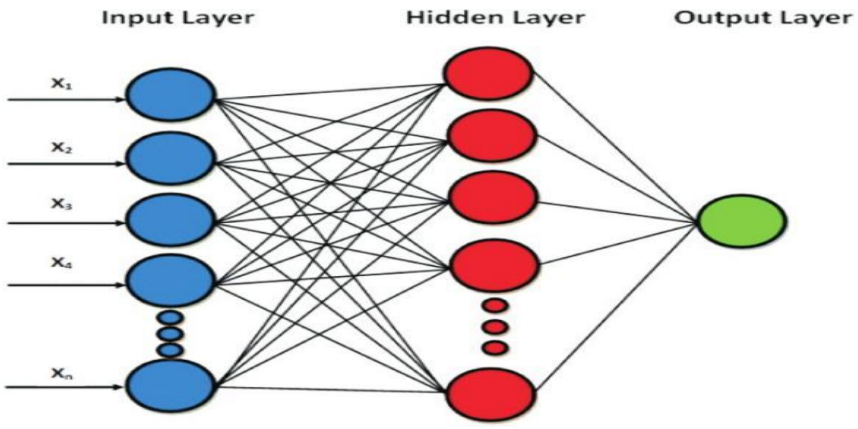


Fig 2.5 A Feed-forward Three Layered Artificial Neural Network Model

Review on Hybridization of two Machine Learning Language in Mitigating Advanced Persistent Threat

The main aim of any machine learning language technique is to discover, learn and adapt to any issue that will require any change over time and can improve machine performance. In the areaof threat detection, any input can be used on the machine learning language algorithm so the computer can "learn" any patterns of attacks (Khrisat and Alqadi., 2023). Then the algorithms are deployed on the input attacks that have been previously unseen in order to perform the actual process of detection. Aside from the ability of recognizing the new patterns of attacks, these algorithms have another capability. It is to sanitize the dataset with the redundant and irrelevant features, thus the dataset will be containing only a few numbers of key features and the process of detection will be optimized. There are some known methods of machine learning such as regression analysis, fuzzy networks, probably approximately correct (PAC)

learning, self-organizing map (SoM), support vector machine (SVM), Bayesian statistics, Genetic algorithm (GA), and artificial neural network (ANN) (Alhijawi and Awajan, 2024).

Theoretical Frameworks

Over the decade, malware and botnets have shown great level of increase in attackers world to become a key reason of the majority of the Denial-of-Service (DoS) activities (Sutheekshan, 2024), direct attacks (Fritsch *et al*., 2022), spear phishing (Baig *et al*., 2021) and scanning, which takes place through the Internet. Botnets are networks created through" enslaving" host computers, called bots. These bots can be controlled by any attacker (botmasters), and its intention is to carry out malicious activities (Kolomeets, M., & Chechulin, 2021). Bots can also be any malicious codes which runs on a host computer and allows the botmasters remotely control the host computer and perform any actions that is created to perform (Choi *et al*., 2023).

A serious change in motivation has been on the notice, ranging from curiosity, fame seeking and excitement- seeking togainof illegal finances, marked by arising sophisticationof malicious software evolution (Choi *et al*., 2023). Moreover, the easy-to-use toolkits made available in building any malware will probably keep malware as a threat tobusiness owners, consumers and governmental organizations in the future. Furthermore, monitoring of network approaches (Fritsch *et al*., 2022) are highly effective and needed in the modern complicated networks.

A major typeof cyber-attacks is the Advanced Persistent Threat (APT) (Sharma *et al*., 2023), and it targets a particularorganisation can be carried out through through several steps. APT aims at exporting and data ex-filtration. APT is hereby considered as a new improved and complex version of many-step attack (Fritsch *et al*., 2022). APTs has created a breach in the current detection due to its advance techniques e.g. social engineering (Choi *et al*., 2023). Damages caused by a successful APT attack can be very expensive and cause serious damages. The costs of APT attacks are major and huge for the investments in intrusion detection system and this makes APTs a current andone of the major threats to companies and governmental organization (Choi *et al*., 2023).

Fritsch *et al*., (2022), proposed a new approach which undergoes two main phases: first, it detects the 8 generic steps and techniques commonly used in APT lifecycle. For this reason, the eight detection modules are highlighted from disguised exe file detection, malicious file hash detection, malicious domain name detection (Sharma *et al*., 2023), malicious IP address detection, SSL certificate detection, domain flux detection (Sharma *et al*., 2023), scan detection, and Tor connection detection (Fritsch *et al*., 2022). After the 8 generic detection modules, the correlation framework is created and links the outputs from the detection modules. In the APT lifecycle, communication continuation between infected hosts and the Command and Control servers are preserved toguide and instruct the machines which are compromised. These communications are protected mainly by Secure Sockets Layer (SSL) encryption, causing it to be more difficult to identify if traffic directed tosuch sites are malicious.

Empirical Framew0rk

Advanced Persistent Threat Detecti0n

Lee *et al*., 2023 presented a classification model for APT detection which is based on machine learning language algorithms. It works on the data traffic for normal users analysed and its aim is toextract CPU usage, open ports, memory usage, open files number in the system 32 folder.

Sakthivelu and Vinoth, (2023) introduces and described TerminAPT or, an APT detector. This detector allows flow of information tracking tocreate the links on elementary attacks, triggered by the APT life cycle. TerminAPT has it dependency on an agent, that can be an intrusion detection system standardized to detect any elementary attack.

Xuan and Nguyen (2024) developed a statistical APT detector, which is similar to TerminAPT detectorand puts into consideration that APT undergoes five stages and they are: delivery, exploit, installation, Command and Control domain and actions; and these activities are taken into each stages. The generated events in each state are correlated in a statistical manner.

Shang *et al*., (2021) introduced an APT detection system that is based on Command and Control domains detection. It analyses Command and Control communication which is based on observation and access to Command and Control domains. This domain is independent, while the access to legal domains is correlated.

Al-Hamar, (2021) explored the APT detection based on spear-phishing detections. This approach relieson computation and mathematical analysis to filter spam emails. outputs that act as a group of words and characters are being defined for the detection algorithm to separate legitimate and spam emails.
Li *et al*., (2022) suggested the use of an active- learning based framework for malicious PDFs detection.

Gupta *et al*., (2022) proposed a Data Leakage Prevention (DLP) approaches and it focuses on detecting the last step of an APT attack which is data ex-filtration. ADL Palgorith misused to process the data traffic which detects leaks in data and generate “fingerprints” from the leak features.

Shou *et al*, (2025) explained in the context-based framework for APT detection a framework based on modelling APT as an attack pyramid and on the pyramid topwhich is the attack goal is highlighted, and the later planes shows the environments is applied in the APT lifecycle.

Krishnapriya and Singh, (2024) presented an in-depth analysis of Duqu when European organizations weretargeted by attackers using Duqumal ware to steal data. He proposed the Duqu detector toolkit that has six investigation tools to detect the Duqumal ware. The outputs of all those tools are then written into a specific log for a possible correlation of the findings. The authors admit that the empirical results show a high rate off also negatives. Moreover, the detection output needs to be carefully investigated by the network security team. Besides, the detection tools are developed to detect the APT

attacks particularly performed using the Duqu malware, this means, the attack cannot be detected when using a different piece of malware.

Attack Pyramid proposed by Shichkina, Y. A., & Fatkieva, (2021) shows an inspired model by the proposed attack tree concept. Attack Pyramid shows the pyramid shape as a model of an APT attack in which the apex represents APTobjectives, while its faces represent the paths and barriers that can be used toovercome such threat.
Goldblum *et al*., (2022) suggest the framework that creates a particular model that depends on the attack and using dataset. This approach shows the effectiveness of the model by combining datasets generated in the organization without a previous knowledge of their structures.

Alomari *et al*., (2021) proposed the combination of an automatic system that is created from big data and machine learning language. The authors used datasets which are generated using web servers to detect attacks in their Internet services and their firewalls datasets to analyse for any possible data ex-filtration. Three models were combined in this work and include: Matrix Decomposition based, Replicator Neural Networks and Density-based out lier analysis.

Hybridization ofApproaches of Machine Learning Languages
Almazrua and Alshamlan. (2022) applied correlation-based feature selection (CFS) with genetic search approach to identify the ideal features and implemented Artificial Immune System as the classifier. In their work, a total of eight features were selected based on its importance, from the total of 41 features in the KDD99 dataset. The proposed system is able toachieve 98.6354% of correct classified instances with mean absolute error value of 0.0068.

Al Mamun *et al*. (2024) applied genetic algorithm (GA) in their proposed APT attack detection system. The implemented datasets and evolution process for GA. This proposed method uses information theory in the filtering of traffic data and can in turn reduce complexity by using a linear structure rule in classifying network behaviors to form normal and abnormal behaviors. Implemented with the GALIB java library and KDD CUP 99 data set used to train and test the system classifier. The result shows that the proposed system is able to achieve a detection rate of uptoa 99.87% and a low false positive rate of 0.003%.

Wang *et al*. (2023) used the genetic algorithm for optimize the support vector machine (SVM) based on feature selection in propose of improving rate of detection. By the using 41 feature, achieve the almost 99% of detection and try to compare the result of GA-SVM with other machine learning algorithm to show the improving of detection rate.

Materials and Methods
The objectives of researches were also addressed under the research process as shown in Fig 3.1. The first objective includes hybridizing both techniques together to form a new technique which is achieved by the mathematical model as well as designing an algorithm for it. After which the algorithm is being implemented using MATLAB and tested using benign and malware datasets. The second objective is evaluation of the new hybridized algorithm with performance metrics. A comparative analysis is carried out with the results obtained in objective one and then compared with the results obtained from other related works, with the aim of determining the performance of the

new hybridized algorithms.

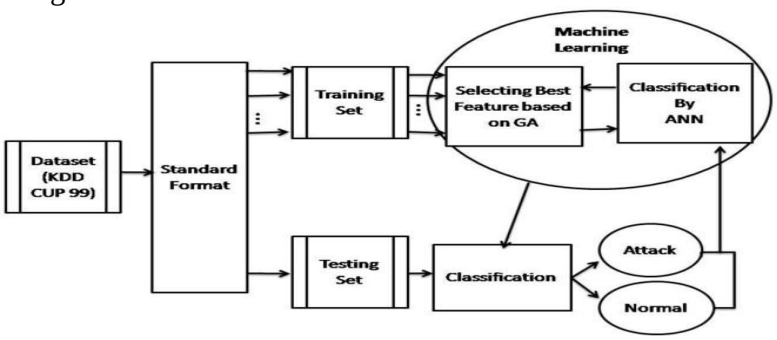


Fig 3.1: Research Process Showing Overall Method used in this Research

Fig 3.2 shows the overall method in designing the hybridized algorithm. First of all, this method will divide the dataset through a pattern that is random and divided into two categories, testing set and training set. In the training phase, machine learning algorithm is used in the first task to learn and select the most appropriate features and in testing phase, the knowledge of the algorithm used is tested by the machine leaning method and each feature that had been selected in the phase of training are tested as well and then the data is classified into normal and attacks categories.

In the process of machine learning, the data is received by GA and then the features are made and selected for the classification of ANN. The classification of ANN is used for preventing the detection rate and the problem of over fitting from n tests, which their average is for receiving a value for fitness.

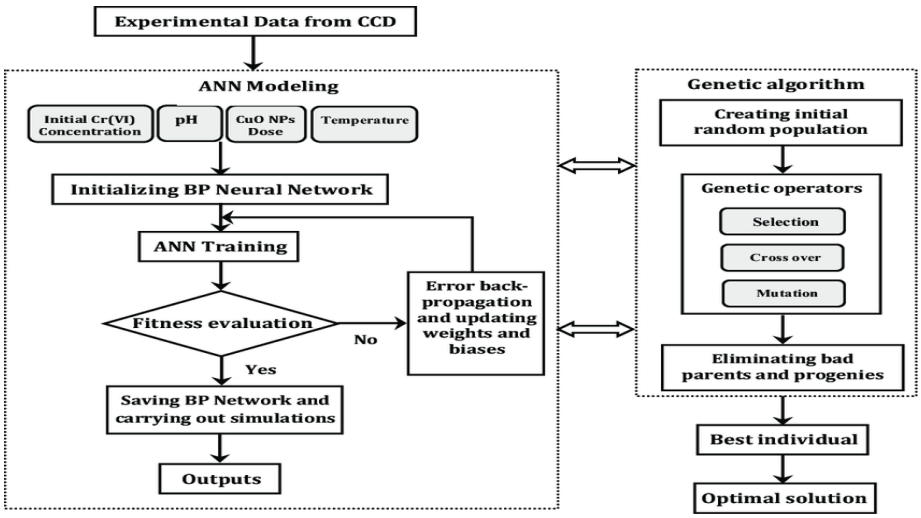


Fig 3.2: A Direct Model of the ANN and GA Hybrid Intelligence.

Table 4.1 Results of the GA and ANN Algorithm

Activation Function Pair	Accuracy (%)	Error Rate (%)	Inconclusive Rate (%)	Classified Rate (%)	Sensitivity (%)	Specif (%)
PPPP	89.21	10.79	0	100	55.2	55.4
PPPT	88.33	11.67	0	100	55.2	53.7
PPPL	92.67	7.33	0	100	60.2	62.2
TTTP	**99.76**	**0.24**	**0**	**100**	**98.19**	**98.1**
TTTT	95.92	4.08	0	100	64.3	62.5
TTTL	90.11	9.89	0	100	58.9	60.3
LLLP	89.99	10.01	0	100	67.78	68.87
LLLT	87.83	12.17	0	100	66.76	68.13
LLLL	88.43	11.57	0	100	60.67	63.33

Findings

This paper provides a hybridized algorithm for detection of advanced persistent threat. It also enhanced the previous accuracy and detection rate of other APT detection system and algorithms by hybridizing GA (genetic algorithm) for optimization and ANN (artificial neural network) for classification. Below are the architectural designs and results:

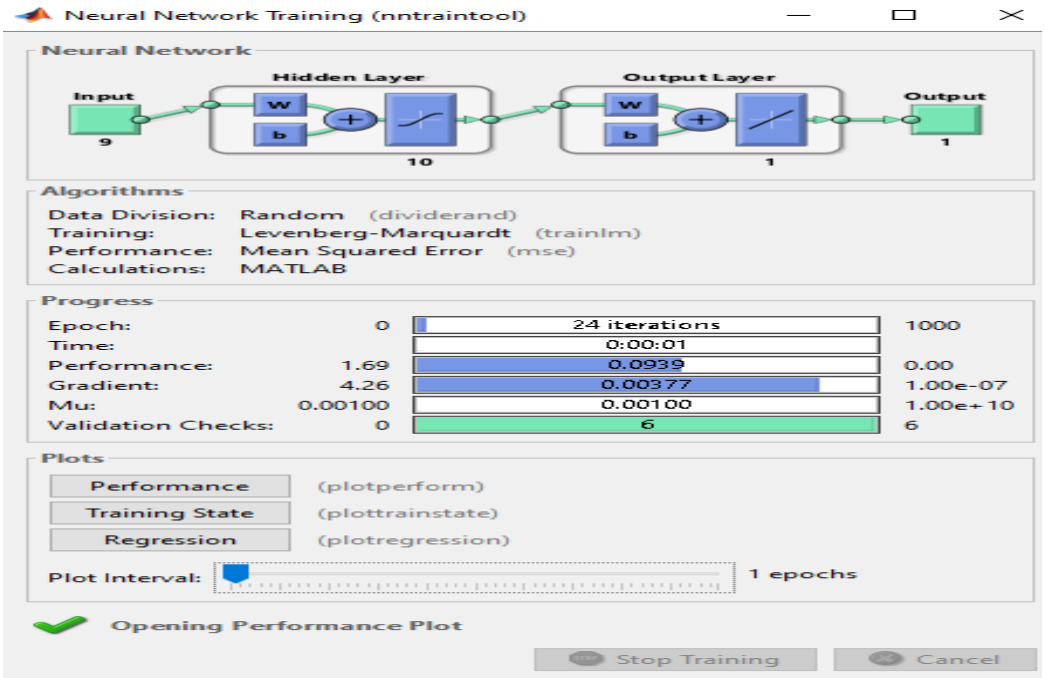


Fig 4.1: Genetic Algorithm and Artificial Neural Network Architecture

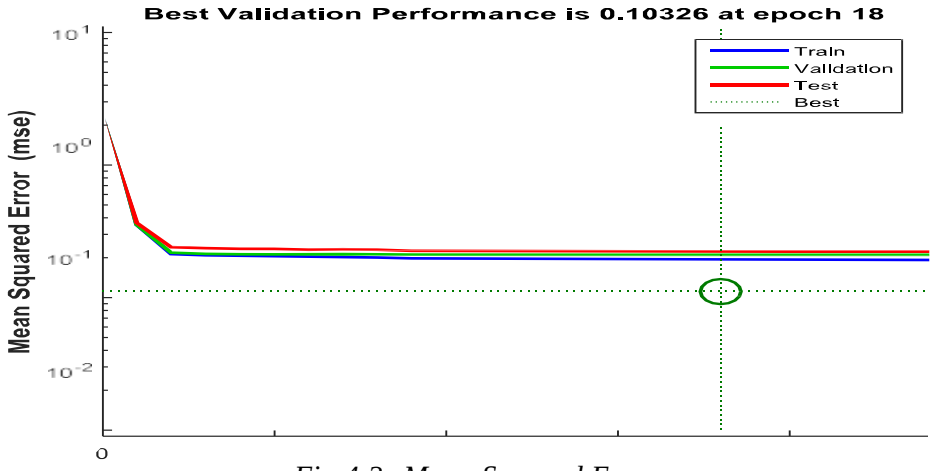


Fig 4.2 Mean Squared Error

The closer to zero, the better the value of Mean Squared Error (MSE). From the graphs, the performances of training, validation, and testing remain stable at close values to zero.

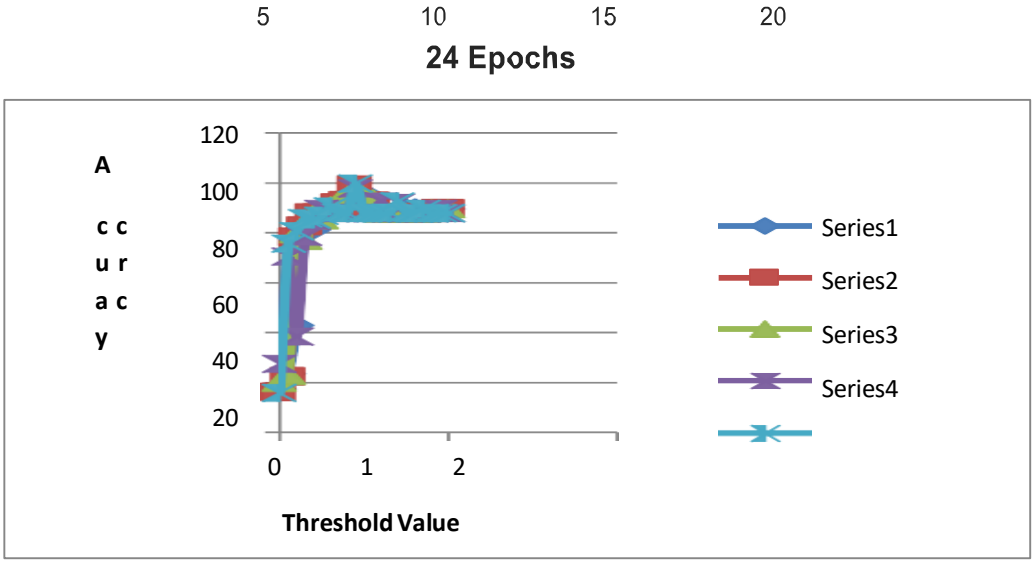


Fig 4.3: Accuracy Versus Threshold Values
Highest mean accuracy of 99.76% was achieved at the threshold value of 0.45.

Performance Evaluation

The table shows the performance of the hybridized model across varying activation functions and processing neurons. The best performance is at TTTP which means activation function combinations: tansig – Tansig – Tansig – Purelin. This implies that the Tansig-activation functions for the hidden layers while the purelin-activation function is for the output layer

Conclusion

In order to produce features for detection and improve accuracy in APT detection, hybridization and implementation of two algorithms - GA and ANN has been proposed in this research. The ANN is applied as classifier detection system. The findings show the highest rate for detection is achieved by GA with ANN when compared to the

Modified Mutual Information based Feature Selection (MMIFS), Learning Fuzzy Classifier System (LCFS) and Firefly Swarm Algorithm (FFSA). This study conducted a series of experiments by using the dataset of Benign and Malicious files from Kaggle for detection of network attacks categories. The feature optimization with GA and the classification of ANN indicates better rates of detection in the proposed Advanced Persistent Threat detection system. The detection rate results of GA when compared with Modified Mutual Information based Feature Selection (MMIFS), Learning Fuzzy Classifier System (LCFS), Firefly Swarm Algorithm (FFSA) and Artificial Neural Network (ANN) in APT detection and with an accuracy of 99.76% shows the highest detection rate.

References

Adem, K. (2022). *Impact of activation functions and number of layers on detection of exudates using circular Hough transform and convolutional neural networks*. Expert Systems with Applications, 203, 117583.

Ahmed, K. (2022). *Pattern Extraction and Behavior Analysis of Self Created HTTP Based Advanced Persistent Threat (APT) for Better Detection (Doctoral dissertation, College of Signals*, National University of Sciences and Technology).

Al-Hamar, Y., Kolivand, H., Tajdini, M., Saba, T., & Ramachandran, V. (2021). *Enterprise Credential Spear-phishing attack detection*. Computers & Electrical Engineering, 94, 107363.

Almazrua, H. and Alshamlan, H. (2022). *A comprehensive survey of recent hybrid feature selection methods in cancer microarray gene expression data*. IEEE Access, 10, 71427-71449.

Al Mamun, A., Al-Sahaf, H., Welch, I., Mansoori, M. and Camtepe, S. (2024). *Detection of advanced persistent threat: A genetic programming approach.* Applied Soft Computing, 167, 112447.

Alomari, E., Katib, I., Albeshri, A., Yigitcanlar, T. and Mehmood, R. (2021). Iktishaf+: a big data tool with automatic labeling for road traffic social sensing and event detection using distributed machine learning. Sensors, 21(9), 2993.

Baig, M. S., Ahmed, F., and Memon, A. M. (2021). *Spear-Phishing campaigns: Link Vulnerability leads to phishing attacks, Spear-Phishing electronic/UAV communication-scam targeted.* In 2021 4th International Conference on Computing & Information Sciences (ICCIS) (pp. 1-6). IEEE.

Buchta, R., Gkoktsis, G., Heine, F., & Kleiner, C. (2024). A*dvanced Persistent Threat Attack Detection Systems: A Review of Approaches, Challenges, and Trends*. Digital Threats: Research and Practice, 5(4), 1-37.

Choi, K. S., Lee, C. S., & Merizalde, J. (2023). *Spreading viruses and malicious codes*. In Handbook on Crime and Technology (pp. 232-250). Edward Elgar Publishing.

Fritsch, L., Jaber, A., & Yazidi, A. (2022). *An overview of artificial intelligence used in malware*. In Symposium of the Norwegian AI Society (pp. 41-51). Cham: Springer International Publishing.

Garouani, M. (2022). *Towards efficient and explainable automated machine learning pipelines design: Application to industry 4.0 data* (Doctoral dissertation, Université du Littoral Côte d'Opale; Université Hassan II (Casablanca, Maroc).

Goldblum, M., Tsipras, D., Xie, C., Chen, X., Schwarzschild, A., Song, D., and Goldstein, T. (2022). *Dataset security for machine learning: Data poisoning, backdoor attacks, and defenses*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(2), 1563-1580.

Gupta, I., Mittal, S., Tiwari, A., Agarwal, P. and Singh, A. K. (2022). *Tidf-dlpm: Term and inverse document frequency based data leakage prevention model.* arXiv preprint arXiv:2203.05367.

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

Jimmy, F. N. U. (2024). Cyber security vulnerabilities and remediation through cloud security tools. *Journal of Artificial Intelligence General science* (JAIGS) ISSN: 3006-4023, 2(1), 129-171.

Khrisat, M. and Alqadi, Z. (2023). *Performance evaluation of ANN models for prediction.* Acadlore Trans. Mach. Learn, 2(1), 13-20.

Kim, S., Park, K. J. and Lu, C. (2022). *A survey on network security for cyber–physical systems: From threats to resilient design*. IEEE Communications Surveys & Tutorials, 24(3), 1534-1573.

Kolomeets, M. and Chechulin, A. (2021). *Analysis of the malicious bots market*. In 2021 29th conference of open innovationsassociation (FRUCT) (pp. 199-205). IEEE.

Krishnapriya, S. and Singh, S. (2024). *A Comprehensive Survey on Advanced Persistent Threat (APT) Detection Techniques.* Computers, Materials & Continua, 80(2)

Lee, K., Lee, J. and Yim, K. (2023). *Classification and analysis of malicious code detection techniques based on the APT attack.*. Applied Sciences, 13(5), 2894.

Li, Y., Wang, X., Shi, Z., Zhang, R., Xue, J. and Wang, Z. (2022). Boosting training for PDF malware classifier via active learning. *International journal of intelligent systems*, 37(4), 2803-282

Liu, Q., Hagenmeyer, V. and Keller, H. B. (2021). *A review of rule learning-based intrusion detection systems and their prospects in smart grids*. Ieee Access, 9, 57542-57564.

Rahman, M. H. (2024). *A Comprehensive Survey on Hardware-Software co-Protection against Invasive, Non-Invasive and Interactive Security Threats*. Cryptology ePrint Archive.

Roy, S., Sharmin, N., Acosta, J. C., Kiekintveld, C., and Laszka, A. (2022). Survey and taxonomy of adversarial reconnaissance techniques. *ACM Computing Surveys*, 55(6), 1-38

Sakthivelu, U. and Vinoth Kumar, C. N. S. (2023). Advanced Persistent Threat Detection and Mitigation Using Machine Learning Model. *Intelligent Automation & Soft Computing*, 36(3).

Shang, L., Guo, D., Ji, Y. and Li, Q. (2021). *Discovering unknown advanced persistent threat using shared features mined by neural networks*. Computer Networks, 189, 107937.

Sharma, A., Gupta, B. B., Singh, A. K. and Saraswat, V. K. (2023). Advanced persistent threats (apt): evolution, anatomy, attribution and countermeasures. *Journal of Ambient Intelligence and Humanized Computing*, 14(7), 9355-9381.

Shichkina, Y. A. and Fatkieva, R. R. (2021). *Detection of network attacks using of growing pyramid networks.* In 2021 10th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-4). IEEE.

Shou, Z., Di, Y. B., Ma, X., Xu, R. C., Chai, H. Q. and Yin, L. (2025). The APT family classification system based on APT call sequences and attention mechanism. International Journal of Information and Computer Security, 26(1-2), 22-40.

Sutheekshan, B., Basheer, S., Thangavel, G. and Sharma, O. P. (2024). Evolution of malware targeting IoT devices and botnet formation. In 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT) (Vol. 5, pp. 1415-1422). IEEE.

Wang, F., Xie, K., Han, L., Han, M. and Wang, Z. (2023). Research on support vector machine optimization based on improved quantum genetic algorithm. Quantum Information Processing, 22(10), 380).

Xuan, C. D. and Nguyen, T. T. (2024). A novel approach for APT attack detection based on an advanced computing. Scientific Reports, 14(1), 22223.