Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 44 No. 8
August 2023

A Study on Artificial Intelligence and Machine Learning Based Intrusion Detection Systems for Detecting Cyber Attacks

K.Shanthi[1], R.Maruthi[2]

[1] Department of Computer Science, Ponnaiyah Ramajayam Institute of Science & Technology, Thanjavur, Tamil Nadu, India

[2] Department of Computer Applications, Hindustan Institute of Technology and Science, Chennai, Tamil Nadu, India)

Abstract - Intrusion detection Systems (IDS) is software and hardware that examines the network traffic and tries to find possible attacks and intrusions. The usage of technologies and services has increased tremendously on the internet and at the same time the numbers of cyber attacks in various forms and means have also increased. Many techniques and methods are available to detect malicious attacks in the networked environment. Each of those methods has some shortfalls and failures to identify complex and dynamic attacks. Artificial Intelligence (AI) and Machine Learning (ML) techniques were used to overcome the issues with the existing intrusion detection systems. This work focuses on some of the AI and ML based for IDS and those methods were evaluated using KDD99 dataset. It is found that the decision tree algorithm shows efficient results in terms of precision, F-Score, False alarm rate and accuracy .

Keywords - artificial intelligence, cyber attacks, intrusion detection systems, machine learning

1. Introduction

Cyber Security is the major concern in today's online service world. Due to the advancements of technologies, protocols and services, a lot of security issues arise. AI and ML based approaches were used in all the sectors to enhance the accuracy and efficiency of the task. AI based methods collect the information and try to draw the relationships between the threats such as malicious data, files, spoofed IP addresses etc.. An IDS is a device or a piece of that monitors networks for harmful activity. The two categories of intrusion detection systems are host-based intrusion detection systems (HIDS) and network-based intrusion detection systems (NIDS). IDS techniques can be divided into three main categories: anomaly-based, signature-based, and specification-based. intrusion detection systems. In anomaly based IDS, the malicious activities are identified by differentiating the valid activities in the network traffic. In case of signature based IDS, it is detected using the patterns in the malicious attacks and any deviation on the malicious activity is detected using specification based IDS.

The implementation of appropriate detection systems is a challenging task and it depends on the network and other components in the network infrastructure like protocols, routers, firewalls etc,.[Vanlalruata and Jamal 2013,Talaei and Kaabouch 2023]. AI and ML in cyber security create a safe environment for users and organizations. In cyber security, the ML techniques are used to analyze the previous cyber attacks and suggest suitable defense mechanisms. The IDS uses ML algorithms to identify anomalous patterns by learning the patterns from the large datasets of normal and abnormal behaviors in the network traffic. Some of the most often used ML methods are K-nearest neighbor (KNN) algorithms, Support Vector Machines (SVM), Logistic Regression (LR) models and Decision Trees to detect anomalies and attacks in cyberspace. In order to identify cyberattacks and by attackers ,ML based methods such as, Bayesian network, Decision tree, Naive Bayes Classifier, Artificial Neural Network (ANN), random tree, random Decision forest and Decision table were employed. The performance of these techniques was evaluated using precision, f1-score, recall, and accuracy [Alqahtani and Hamed et. al 2020].

The remaining sections of the paper are organized as follows; the literature survey of the existing AI and ML based methods for IDS are presented in section 2. Section 3 discusses the various AI and ML based methods for IDS, followed by performance evaluation of IDS in section 4, and then followed by conclusion in section 5.

2. Related Works

Nowadays researchers are focusing on designing intelligent intrusion detection systems using AI and ML techniques. The machine learning based methods use binary classifiers and multiclass classifiers to identify the attacks. Some of the ML based IDS found in the literature are as follows, Neural network based threat prediction system has been proposed by Jay Kumar Jain and Waoo in 2023 using widely used NSL-KDD data set. Adversarial machine learning (AML) strategies and defense mechanisms to reduce the attacks were discussed by Alotaibi, Rassam 2023. A Stacked Long Short Term Memory model configuration to improve the performance of AIDS has been suggested by Figueiredo and Serrao et.al 2023 with the preprocessed CICIDS2017 data set to achieve highest accuracy. ML is used in IDS for IoT applications using VGG-16 and DenseNet, K-nearest neighbors, random forest and SVM were studied and VGG-16 stacked model shows the highest accuracy of 98.3%(Musleh and Alotaibi 2023). Controller Area Network (CAN) in vehicle communication protocol lacks message authentication and encryption schemes to protect the data and the AI-based IDS has been suggested as a countermeasure against automotive cyber attacks(Rajapaksha and Kalutarage et.al 2023). A Hybrid model that merge ML and Deep Learning (DL) has been devised by Talukder and Hasan et.al 2023. using SMOTE for preprocessing and XGBoost for feature selection to increase the detection rates and it is tested with KDDCUP99 and CIC MAlMem-2022 datasets with an accuracy of 99.99% and 100% respectively. An IDS technique based on deep learning and Swarm Intelligence using CNN- based feature selection method called Capuchin Search Algorithm(CapSA) optimizer is studied by Abd Elaziz and Al-qaness et.al 2023.The proposed method is evaluated with metrics like average accuracy, average recall, average precision and performance indication Rate (PIR) in the NSL-KDD, BoT-IoT ,KDDcup-99 and CICIDS-2017 datasets. ML-based IDS for IoT using stacking ensemble model as a most optimal classifier and it is evaluated using Mathews correlation coefficient of 0.9971 and 0.9909 in the binary classification and the multi-class classification respectively (Guo and Pan et.al 2023)

SCADA (Supervisory Control and Data Acquisition) Systems protection is important for national and international security. A novel European framework - 7 project CockpitCI has been introduced by Yasakethu and Jiang 2013 using intelligent intrusion detection methods like rule-base approach, hidden markov model and support vector machines to protect the SCADA systems. AI and Computational Intelligence (CI) methods like Genetic Algorithm (GA), ANN, Fuzzy logic and Artificial Immune Systems (AIS) were discussed by Zamani and Mahdi 2013 to build efficient IDS. Widely used supervised and unsupervised ML methods were explored by Maseer and Kamil et. al 2021 using CICIDS2017 dataset in real world network attacks were studied. An unsupervised technique using temporal convolutional networks and edge computing is proposed using generative adversarial networks (GAN) is proposed by Filho and Naili et.al 2023 for edge servers and the proposed approach is 3.8 times faster than other ML based methods and it is found to be more accurate. An Improved Long Short-Term Memory (LSTM) network-based Secured Automatic Two-Level Intrusion Detection System (SATIDS) for IoT and Software Defined Networks (SDN) has been presented by Elsayed and Rania et. al 2023. For the ToN-IoT dataset, the two level IDS attains 96.35% accuracy, 96% detection rate, and 98.4% precision, while for the InSDN dataset, it obtains 99.73% accuracy, 98.6% detection rate, and 98.9% precision.

A work by Djenna and Bouridane et.al 2023 proposes a systematic approach that combines behavior-based deep learning and heuristic-based approaches to identify modern malware like Adware, rootkit Radware, SMS malware and ransomware and the proposed method outperforms the other deep learning methods. A method suggested by Belhadi and Djenouri et.al 2023 detects a group of intrusions which combines both deep learning and decomposition methods and it is evaluated using IDS 2018 and LUFlow data set. The Concept of decision trees is used as predictive models to detect cyber attacks and reduces the computational complexity (Al-Omari and Rawashdeh et.al 2021). It is clearly found from the literature that the AI and ML based techniques for IDS produces high detection rate and low false alarm rates compared to other traditional approaches.

3. Materials and Methods

An IDS using AI and ML approaches are used to find

whether the host or network is under attack or not. These systems continuously monitor the network traffic or system activities and try to detect the anomalies if any i.e it detects unknown malware by analyzing the behavior. The three main phases in any AI and ML based methods are data preprocessing, training and testing phases. The real-time network traffic data is first recorded and gathered in a database. Then the collected data is preprocessed to generate a training data set. The data preprocessing is very much needed to get an efficient solution for any machine learning algorithms. The test data is acquired from the real time scenario and it is fed into AI or ML based methods. The overall structure of AI and ML based IDS is shown in Figure:1 The AI or ML based methods detect the abnormal patterns using the training and testing dataset. If an abnormal pattern appears, it sets an alarm and takes appropriate action; otherwise it allows the traffic to reach the destination.

AI and ML based methods attempt to identify anomalies in the network traffic using the chosen features and detecting them as normal or abnormal from the behavioral patterns with machine learning algorithms. The following list includes some of the ML algorithms employed in this study and the algorithms were evaluated using the benchmark dataset (KDD99)

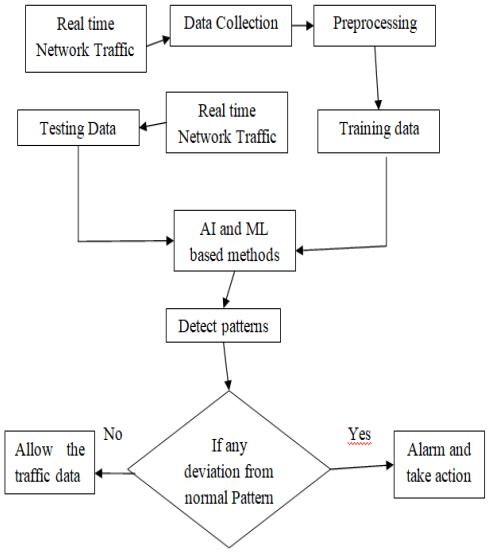

Figure-1 Overall structure of AI and ML based IDS

3.1 Decision Tree Algorithm

Decision tree algorithm is a supervised learning method used for classification problems. The two nodes in the hierarchical tree classifier are decision nodes and leaf nodes. Leaf nodes are the output of the decision and do not have any sub nodes. CART (Classification and Regression Tree Algorithm) is used to construct a decision tree. A decision tree gives yes or no based on the question and its further split into sub trees and it is shown in figure-2.

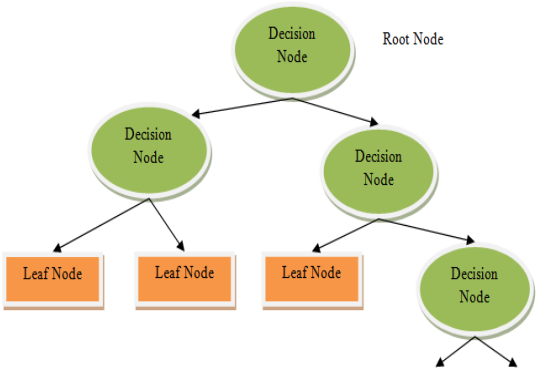

Figure-2: Decision tree

The most relevant feature selection is important to make the classifier work efficiently for more frequent values. The decision tree works as follows

- The root node of the tree contains all of the data in the dataset.
- A suitable attribute or feature is selected from the data set and the root node is divided into subsets.
- Then the decision tree node is generated with the best attribute.
- It is a recursive process and new decision trees are generated using the subset of the data set until a stage is reached.

A example decision tree in network traffic is shown in Figure:3

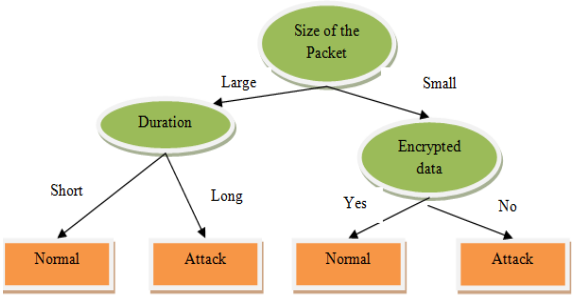

Figure-3 Decision tree for Network Traffic

The decision tree requires little data preprocessing and handles both numerical and categorical data and also multi output problems, but it becomes unstable if there is minimal change in the data resulting in the major change in the structure of the decision tree. Due of its intricacy and duration, decision tree algorithm is relatively expensive.

3.2 K-Nearest Neighbor (KNN) Algorithm

The network traffic is classified as either normal or intrusive using the KNN supervised classifier. In Figure 4, similar data points that are near to one another are displayed. The data points are determined by calculating the distance between the points in the graph. The KNN classifier works as follow

- The training and testing Data is loaded
- A new data point K is chosen from the neighboring points
- For each data, the distance is calculated between the chosen K and the current data
- It should be indexed in the ordered collection, and the indexes should be sorted by distances from smallest to greatest.
- Choose the first K entries from the sorted collection, then retrieve the K entries' labels.

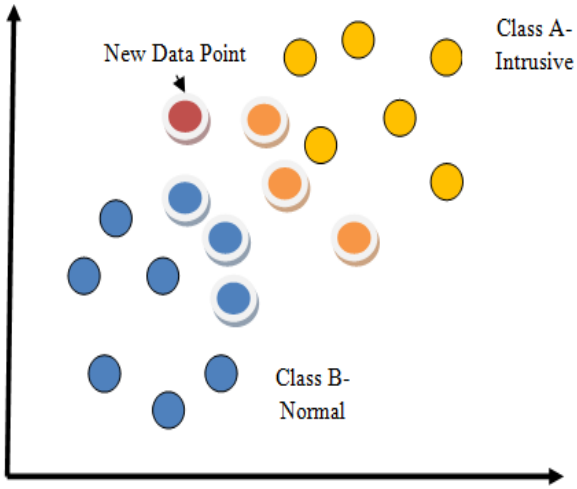

Figure-4 KNN classifier

The KNN algorithm is simple to use, however the parameters must be adjusted. At the same time, it becomes slower with more data.

3.3 Artificial Neural Networks (ANN)

ANN consists of neurons and it is responsible for creating layers and the neurons are also referred as tuned parameters. One layer's output serves as the input for the following layer's and each layer has different non-linear activation functions to help the learning process and to derive the output from each layer. The term "terminal neurons" also refers to the output layer. The same input is processed through the layers to give different outputs. This is represented with a multilayer perceptron. In IDS the input for the neural networks is the well defined features and output is the class variable (0 or 1) where 0 represents the normal data in the network traffic and 0 represents intru

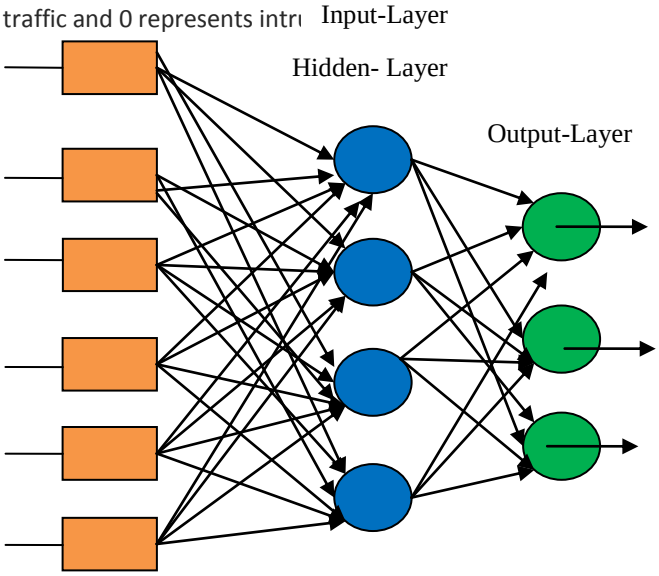

Figure-5: ANN for IDS

The neural network's initial layer will include a number of features. The parameters for model training and predictions are then set. The learning rate is balanced by an optimizer, which cycles through different weight sets. The loss function (binary cross entropy) is used to estimate the weights for the different layers. The neural networks are mostly suitable for large datasets.

3.4 Support Vector Machine (SVM)

SVM is a supervised learning technique that can be employed for classification and regression problems. Both linear and non-linear data can be handled by it. It can be used for anomaly detection and outlier detection. It draws a boundary between any two classes in order to classify them. It determines the best decision boundary also called a hyperplane between the data that belong to the specified group and data that does not belong to the group. It chooses extreme points or data to create a hyperplane called support vectors. These are the locations that are most near the hyperplane. It tries to maximize the nearest data points of all the classes. The figure-6 shows two classes of data or vectors that are classified using the decision boundary. The decision boundary created by SVM is called the maximum margin classifier. SVM has better classification rates than ANN and it works efficiently on smaller datasets and it is not suitable for large datasets. But it works effectively , if the number of features is more than the number of data points. In a

machine learning algorithm the dataset is loaded and it gets pre-processed and then it splits the data into attributes and labels. Next, the data is divided into training set and testing sets, and the SVM algorithm is taught to generate predictions.

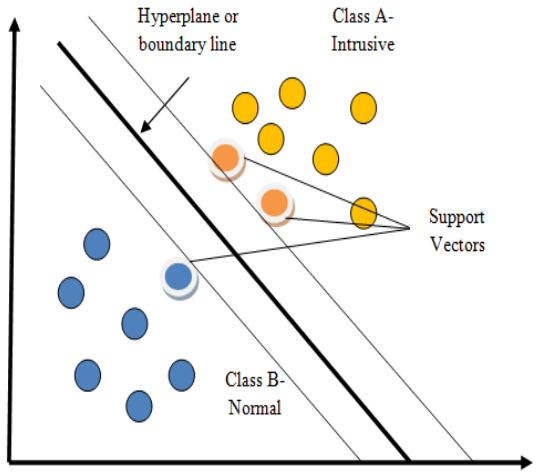

Figure-6 Support Vector Machine-classification

3.5 K-Means Clustering Algorithm

K means clustering is an unsupervised classification technique that categorizes data or objects based on their characteristics. The unlabelled data set is grouped into K different clusters. For example, for K=2, the clusters created are two and for K=3, three clusters will be created and so on. K-Means Clustering is a centroid-based algorithm, in which each cluster is linked with a centroid. The K-Means Clustering algorithm works as follows,.

- Input is the unlabeled data set as shown in Figure-7(a)
- Divide the dataset into k-number of clusters. In IDS the values of K are 2, to create two clusters, say normal and intrusive.
- Determine K random center points (centroid) for each cluster from the input dataset.
- Assign each data point closer to the chosen closest centroid to form a cluster by calculating the distance between the points
- Draw a median between the divisions as shown in Figure-7(b) to show two clusters. In Figure-7(c) , the circles with red indicate one group of cluster(intrusive data) and the circles with green indicate another cluster(normal traffic data)
- Then a new centroid is chosen to find the closest cluster and this process will be repeated until there are no dissimilar data points between the two clusters. The

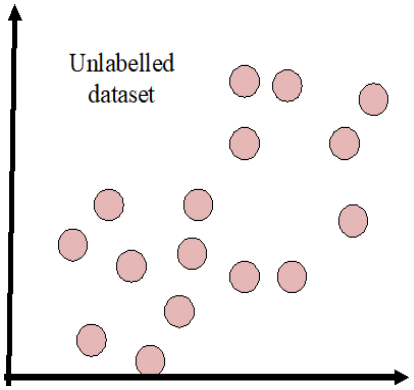

Figure-7(a) Unlabelled Input dataset

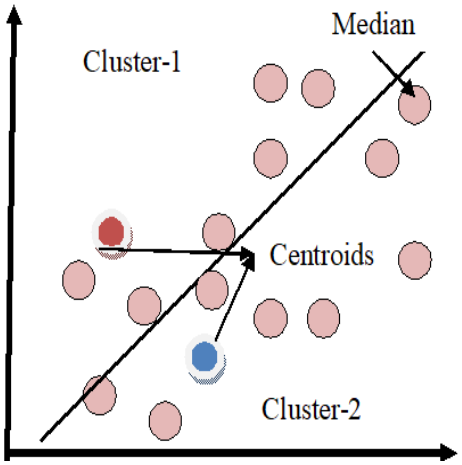

Figure-7 (b) Choosing two centroids

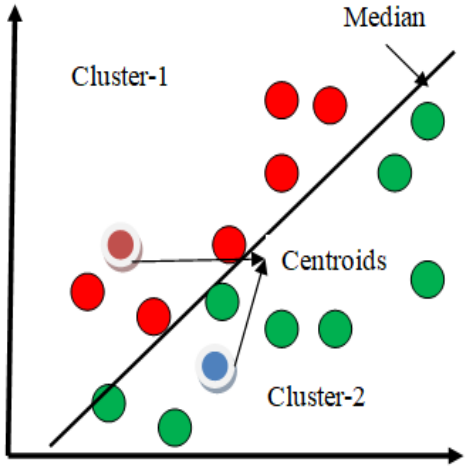

Figure-7(c)Centroids with two clusters

The Figures in 7(d) to 7(f) show the repeated process of finding the new centroids and new data points for the clusters. The figure-7(f) shows the clusters in which the median is drawn the other way than the previously formed clusters. The final two different clusters were shown in Figure-7(f).

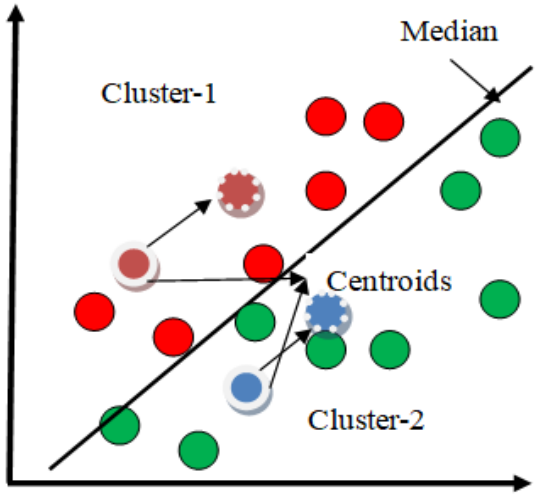

Figure-7(d)New Centroids

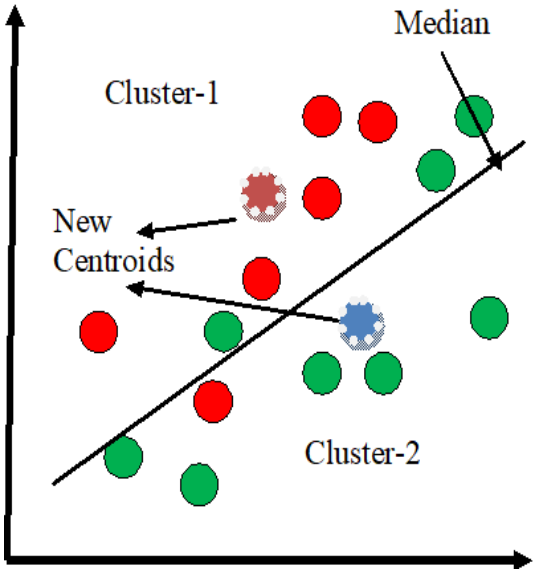

Figure-7(e)New Centroids with data points

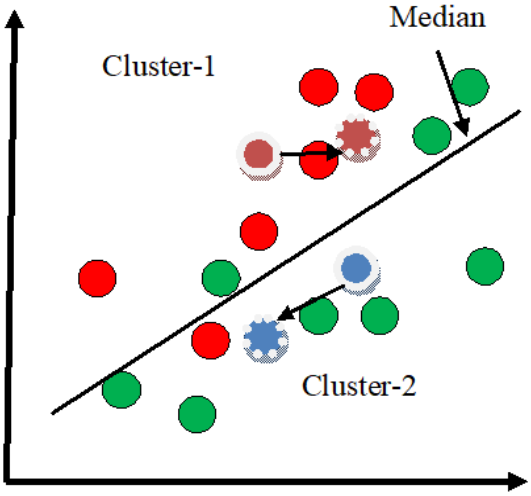

Figure-7(f) Finding New Centroids

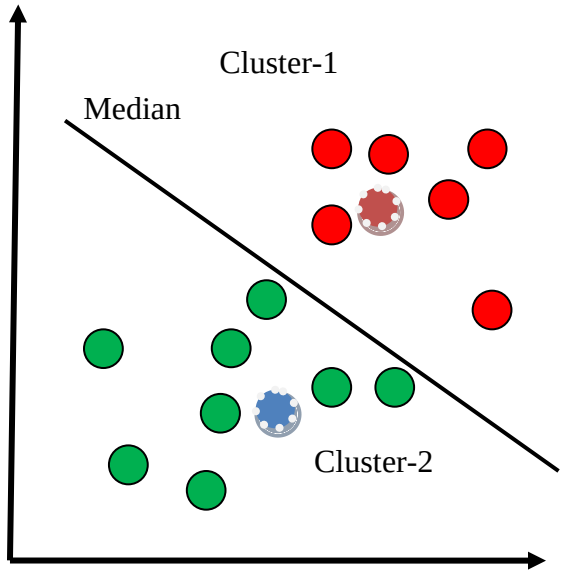

Figure-7(e)New Centroids with new data points

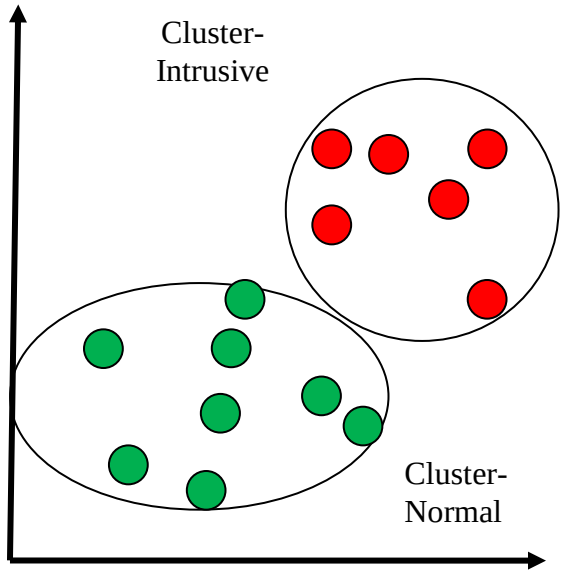

Figure-7(f) Final Clusters

4.Performance Evaluation

Performance Evaluation of intrusion detection systems is carried out using the commonly used evaluation metrics as follows.

a) Precision is defined as the ratio of correctly predicted attacks to all predicted attacks

Precision = $\frac{TP}{(TP+FP)}$ ----Equation-1

In the above formula, TP is true positives and FP is false positives

b)Recall is defined as the ratio of between the number of correctly predicted positive samples to the total number of positive samples , also referred as detection rate.

Recall=$\frac{TP}{(TP+FN)}$ ------Equation -2

In the above formula, FN is false negatives

c)F-Score/Measure is a statistical technique and it is defined as follows using the precision and recall

F-Score/Measure= $2*\frac{(Precsion*Recall)}{(Precision+Recall)}$-------Equation-3

d)Accuracy is the ratio of correctly predicted attacks to the total number data's also called as detection accuracy

Accuracy= $\frac{(TP+TN)}{(TP+TN+FP+FN)}$--------Equation-4

e)False Alarm Rate: referred s false positive rates and is defined as the ratio of incorrectly predicted attack to normal

False Alarm Rate= $\frac{FP}{(FP+TN)}$-------Equation-5

The parameters used in the above formula's were derived from the confusion matrix. It is a two dimensional matrix with information about actual and predicted categories. In a binary classification, the learning outcomes are as follows and it is represented in a confusion matrix and it contains only four possibilities as follows

- True Positive: The class of network traffic identified correctly as attacks
- True Negative: The class of network traffic that are identified as normal correctly
- False Positive: The class of network traffic that are not identified as attacks
- False Negative: The class of network traffic those are normal but identified as attacks

Table 1. Precision and Accuracy

lgorithm/Metrics	recision	ccuracy
ecision tree	6.3%	8%
NN	4.3%	6.5%
NN	5%	5.6%
VM	4.5%	4.6%
-Means Clustering	3%	3.2%

Table I1. F-Score and False Alarm Rate

lgorithm/Metrics	-Score	alse Alarm Rate
ecision tree	6%	8.5%
NN	3.2%	7%
NN	4.6%	7.5%
VM	3%	6%
-Means Clustering	1.7%	4%

The evaluation of AI and ML based algorithms for detecting the malicious activities in the networks has been done and it is tabulated in the Table-I& II. It is found that the decision tree algorithms outperform the other methods used in this study.

5. Conclusion

This work analyzed the AI and ML based methods like Decision tree, KNN, ANN, SVM and K-means clustering for intrusion detection systems. It is clear that the decision tree algorithm outperforms the other methods used in this study. It is also evident from the study that AI and ML based intrusion detection methods are used more widely as an efficient technique for predicting the attacks than the other traditional intrusion detection tools. The AI and ML based methods for IDS improves the performance and efficiency in terms of accuracy, precision, F-score and false alarm rate

Conflicts of Interest

All authors declare that they have no conflicts of interest in the submitted manuscript.

References

[1] Abd Elaziz, M., Al-qaness, M. A., Dahou, A., Ibrahim, R. A., & Abd El-Latif, A. A. (2023). Intrusion detection approach for cloud and IoT environments using deep learning and Capuchin Search Algorithm. Advances in Engineering Software, 103402.

[2] Al-Omari, M., Rawashdeh, M., Qutaishat, F., Alshira'H, M., & Ababneh, N. (2021). An intelligent tree-based intrusion detection model for cyber security. Journal of Network and Systems Management, 29, 1-18.

[3] Alotaibi, A.; Rassam, M.A. Adversarial Machine Learning Attacks against Intrusion Detection Systems: A Survey on Strategies and Defense. *Future Internet* **2023**, *15*, 62. https://doi.org/10.3390/fi15020062

[4] Alqahtani, Hamed, et al. "Cyber intrusion detection using machine learning classification techniques." Computing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020,

Revised Selected Papers 1. Springer Singapore, 2020.

[5] Belhadi, A., Djenouri, Y., Djenouri, D. et al. Group intrusion detection in the Internet of Things using a hybrid recurrent neural network. Cluster Comput 26, 1147–1158 (2023). https://doi.org/10.1007/s10586-022-03779

[6] Djenna, A., Bouridane, A., Rubab, S., & Marou, I. M. (2023). Artificial Intelligence-Based Malware Detection, Analysis, and Mitigation. Symmetry, 15(3), 677.

[7] Elsayed, Rania A., et al. "Securing IoT and SDN systems using deep-learning based automatic intrusion detection." Ain Shams Engineering Journal (2023)102211.

[8] Figueiredo, J.; Serrão, C.; de Almeida, A.M. Deep Learning Model Transposition for Network Intrusion Detection Systems. Electronics 2023, 12, 293. https://doi.org/10.3390/electronics12020293

[9] Guo, G., Pan, X., Liu, H., Li, F., Pei, L., & Hu, K. (2023, March). An IoT Intrusion Detection System Based on TON IoT Network Dataset. In 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC) (pp. 0333-0338). IEEE.

[10] Jay Kumar Jain, & Waoo, A. A. . (2023). An Artificial Neural Network Technique for Prediction of Cyber-Attack using Intrusion Detection System. Journal of Artificial Intelligence,Machine Learning and Neural Network (JAIMLNN) ISSN: 2799-1172, 3(02), 33–42. https://doi.org/10.55529/jaimlnn.32.33.4.

[11] Maseer, Ziadoon Kamil, et al. "Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset." IEEE access 9 (2021): 22351-22370.

[12] Musleh, D.; Alotaibi, M.; Alhaidari, F.; Rahman, A.; Mohammad, R.M. Intrusion Detection System Using Feature Extraction with Machine Learning Algorithms in IoT. J. Sens. Actuator Netw. 2023, 12, 29. https://doi.org/10.3390/jsan12020029

[13] P. F. de Araujo-Filho, M. Naili, G. Kaddoum, E. T. Fapi and Z. Zhu, "Unsupervised GAN-Based Intrusion Detection System Using Temporal Convolutional Networks and Self-Attention," in IEEE Transactions on Network and Service Management, doi: 10.1109/TNSM.2023.3260039.

[14] Rajapaksha, S., Kalutarage, H., Al-Kadri, M. O., Petrovski, A., Madzudzo, G., & Cheah, M. (2023). Ai-based intrusion detection systems for in-vehicle networks: A survey. ACM Computing Surveys, 55(11), 1-40.

[15] Talaei Khoei, T.; Kaabouch, N. A Comparative Analysis of Supervised and Unsupervised Models for Detecting Attacks on the Intrusion Detection Systems. Information 2023, 14, 103. https://doi.org/10.3390/info14020103

[16] Talukder, M. A., Hasan, K. F., Islam, M. M., Uddin, M. A., Akhter, A., Yousuf, M. A & Moni, M. A. (2023). A dependable hybrid machine learning model for network intrusion detection. Journal of Information Security and Applications, 72, 103405.

[17] Vanlalruata Hnamte, Jamal Hussain,DCNNBiLSTM - An Efficient Hybrid Deep Learning-Based Intrusion Detection System, Telematics and Informatics Reports, Volume 10,2023,100053,ISSN 2772-5030,https://doi.org/10.1016/j.teler.2023.100053.

[18] Yasakethu, S. L. P., and J. Jiang. "Intrusion detection via machine learning for SCADA system protection." 1st International Symposium for ICS & SCADA Cyber Security Research 2013 (ICS-CSR 2013) 1. 2013.

[19] Zamani, Mahdi, and Mahnush Movahedi."Machine learning techniques for intrusion detection." arXiv preprint arXiv:1312.2177 (2013).