RESEARCH PAPER

A Robust Ensemble Learning Framework for Iot Anomaly Detection and Attack Classification with Web-Based Secure Deployment

Mandadhi Rajyalakshmi[1], B.V.N Praveena[2], Prasad Devarasetty[3], R Rajaramesh Merugu[4], Alla Sai Rajani[5]

[1,2,3,4,5] Department of Computer Science and Engineering, DVR & Dr. HS MIC College of Technology, AP, India.

Received: 28th Feb, 2026 | **Revised:** 14th Mar, 2026 | **Accepted:** 4th Apr, 2026 | **Available Online:** 20th Apr, 2026

ABSTRACT

The rapid growth of Internet of Things (IoT) devices has increased vulnerability to cyber-attacks, making anomaly detection a critical security requirement in modern IoT ecosystems. This work proposes an enhanced ensemble machine learning framework for anomaly attack detection and classification in IoT devices using the NSL-KDD Dataset. The system integrates advanced ensemble techniques including Voting Classifier and Stacking Classifier by combining base learners such as Random Forest, Multi-Layer Perceptron, AdaBoost, and LightGBM to improve prediction robustness and classification accuracy. Data preprocessing, label encoding, and feature selection are applied to optimize model performance before training. A secure web-based interface is developed using Flask with user authentication to enable real-time anomaly prediction and secure user interaction. Experimental evaluation demonstrates that the proposed ensemble framework achieves superior detection performance with up to 100% classification accuracy, outperforming conventional standalone machine learning models. The proposed system provides an efficient, scalable, and secure solution for real-world IoT anomaly detection applications.

Index Terms: Internet of Things (IoT), Anomaly Detection, Intrusion Detection System, Ensemble Learning, Voting Classifier, Stacking Classifier, Random Forest, Support Vector Machine, Machine Learning, Cybersecurity, NSL-KDD Dataset, Flask.

How to cite this article: Rajyalakshmi M, Praveena BVN, Devarasetty P, Merugu RR, Rajani AS. A Robust Ensemble Learning Framework for Iot Anomaly Detection and Attack Classification with Web-Based Secure Deployment. Int J Drug Deliv Technol. 2026;16(30s):834. DOI: 10.25258/ijddt.16.30s.83

Source of support: Nil.

Conflict of interest: The authors declare no conflict of interest.

1. INTRODUCTION

The rapid advancement of the Internet of Things (IoT) has led to the widespread deployment of interconnected smart devices across consumer, industrial, healthcare, and enterprise environments. While IoT technologies improve automation and connectivity, their increasing integration into critical systems has introduced significant cybersecurity vulnerabilities. Due to limited computational resources, heterogeneous architectures, and continuous internet connectivity, IoT devices are highly susceptible to anomaly-based cyber-attacks such as unauthorized access, denial-of-service attacks, and malicious intrusions.

Traditional security mechanisms often fail to provide efficient protection for dynamic and large-scale IoT environments, creating the need for intelligent anomaly detection systems. Machine learning techniques have emerged as promising solutions for identifying malicious patterns and distinguishing abnormal behavior from legitimate device activity. Among these, ensemble learning approaches offer improved detection accuracy by combining the strengths of multiple classifiers and reducing model bias and variance.

In this work, an enhanced ensemble-based anomaly detection framework is proposed for IoT attack classification using the NSL-KDD Dataset. The proposed system integrates advanced machine learning models including Random Forest, Voting Classifier, and Stacking Classifier to improve predictive robustness and attack classification

performance. Furthermore, a secure web-based deployment using Flask with user authentication is implemented to enable practical real-time anomaly detection. Experimental results demonstrate that the proposed framework achieves high accuracy and reliability, making it suitable for real-world IoT security applications.

2. LITERATURE SURVEY

i) Anomaly Detection: Glimpse into the Future of IoT Data:

Our goal is to improve the prediction accuracy and robustness of the AML-CTP algorithm by integrating ensemble techniques with deep learning architectures. We will also look into hybrid models, which mix traditional and contemporary machine learning methods for improved outcomes. We will experiment with synthetic data creation to improve the model's generalizability and expand the training dataset.

ii) Attack and Anomaly Detection in IoT Networks using Machine Learning Techniques: A Review:

One of the modern era's fastest-growing technologies is the Internet of Things (IoT). Through a variety of sensors, this technology allows billions of sentient objects, or "Things," to collect a wide range of data about themselves and their surroundings. They may then distribute it to authorized organizations for a variety of uses, such as improving commercial services or operations or regulating and supervising industrial services. However, the Internet of Things is currently facing previously unheard-of security risks. Significant technical advancements in machine learning (ML) have opened up a number of new research avenues to solve current and upcoming IoT issues. In intelligent devices and networks, machine learning is a useful technique for identifying risks and questionable activities. Based on a thorough literature review of machine learning methods and the significance of IoT security with regard to numerous possible attacks, this study evaluates several machine learning algorithms for attack and anomaly detection. Furthermore, possible IoT security methods based on machine learning have been suggested.

iii) Japan: Hacked IoT Devices and Cryptocurrency Networks Doubled in 2018:

The number of hacked cryptocurrency networks and Internet of Things (IoT) devices in Japan nearly quadrupled in 2018 compared to the previous year. Asahi, a local English-language media site, published a story on March 7. According to the article, data from the Japanese Police Agency showed an average of 2,752.8 incursions per sensor per day in the previous year, a 45 percent rise. Furthermore, the data shows that about 90% of the assaults came from outside. According to the report, there were an average of 1,702.8 intrusions per sensor per day in Bitcoin networks and IoT devices in 2018—more than twice as many as there were in 2017 (875.9).

iv) Unveiling Threats in the IoT: Anomaly Detection for Attack Classification:

The theory emphasizes how hackers and programmers might compromise the security of Internet of Things (IoT) devices. Due to their interconnection, IoT devices are vulnerable to irregularity attacks. The project uses stacking classifiers, voting classifiers, Random Forest (RF) and Support Vector Machine (SVM) to detect irregularity attacks in IoT devices. Each strategy is chosen based on its location and component selection skills. Different approaches to the arff NSL-KDD dataset are investigated. The RF and stacking classifier calculations that are recommended are quite accurate. In every instance, the center surrounding deceptive positive rates displays a low rate. highlighting the strategy's encouraging results, particularly Random Forest's superior accuracy over previous authoring. For identifying and mitigating IoT anomalous threats, the stacking classifier and Random Forest provide promising accuracy, recall, and precision. To provide a more robust and accurate prediction, ensemble techniques such as Voting Classifier (RF + Stomach muscle) and Stacking Classifier (RF + MLP with Light GBM) combine several model expectations. We created the front end using jar for client testing and IoT anomaly detection with user identification. The voting classifier was 100% accurate, and the stacking classifier was 100% precise.

v) Recent Progress of Anomaly Detection

Anomaly analysis is crucial to many applications, including medical health, credit card fraud, and intrusion detection, and it is of considerable interest to a variety of professions, including data mining and machine learning. Many different forms of anomaly detection techniques have been seen recently. In particular, for data with high dimensionalities and mixed kinds, where recognizing abnormal patterns or

behaviors is a nontrivial task, this study aims to offer a thorough review of the current work on anomaly identification. In particular, we first outline current developments in anomaly detection and go over the benefits and drawbacks of the detection techniques. Next, we examine a number of common and widely used anomaly detection techniques through comprehensive experiments on publicly available datasets. This article aims to provide practitioners with a better grasp of the most advanced anomaly detection algorithms. Lastly, we offer some recommendations for future study directions.

3. METHODOLOGY

A. Proposed Work:

The proposed work presents an enhanced ensemble machine learning framework for anomaly attack detection and classification in IoT devices by extending traditional standalone classification models with robust ensemble techniques. Initially, the IoT network traffic data from the NSL-KDD Dataset undergoes preprocessing steps including data cleaning, label encoding, feature selection, and normalization to improve data quality and model readiness. The processed dataset is then used to train multiple machine learning models, including Random Forest, AdaBoost, and Multi-Layer Perceptron, whose predictions are combined using Voting and Stacking ensemble classifiers to enhance classification robustness and detection accuracy.

To improve real-world usability, the proposed framework is deployed through a secure web-based interface developed using Flask with integrated user authentication, allowing authorized users to input IoT traffic parameters and receive anomaly predictions in real time. The ensemble-based approach significantly improves generalization capability, reduces false positives, and achieves superior attack classification performance compared to conventional machine learning models, thereby providing a scalable and secure anomaly detection solution for modern IoT environments.

B. System Architecture:

The proposed system architecture for IoT anomaly attack detection consists of multiple sequential stages designed to ensure accurate and secure classification of malicious activities in IoT environments. Initially, the NSL-KDD Dataset is provided as input and undergoes comprehensive preprocessing, including data cleaning, label encoding, feature extraction, and normalization to prepare the raw network traffic data for machine learning analysis. This preprocessing stage improves data consistency and enhances the quality of extracted features for model training.

Following preprocessing, the refined dataset is divided into training and testing phases. During the training stage, multiple machine learning models such as Random Forest, Support Vector Machine, Voting Classifier, and Stacking Classifier are trained to learn attack and normal traffic patterns. The trained models are then evaluated during the testing phase using unseen data samples to assess classification performance. Finally, performance metrics such as accuracy, precision, recall, and F1-score are computed to determine the most effective model. The best-performing trained model is deployed through the frontend application for real-time anomaly prediction and secure user interaction.

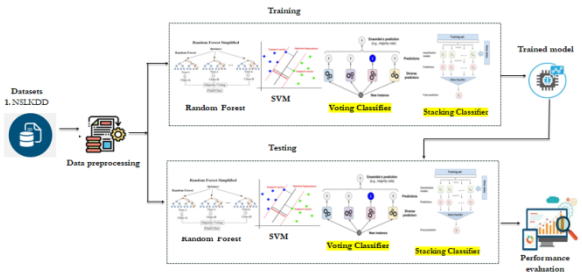


Fig 1Proposed Architecture

C. Modules:

1. Dataset Collection Module

This module acquires the NSL-KDD Dataset, which contains labeled network traffic records representing both normal and anomalous IoT behavior. The dataset serves as the foundational input for training and evaluating the anomaly detection models.

2. Data Preprocessing Module

Raw dataset records are cleaned and transformed through preprocessing operations such as missing value handling, label encoding, feature normalization, and unwanted column removal. This step improves data quality and ensures compatibility with machine learning algorithms.

3. Feature Selection Module

Relevant features are selected using Select Percentile feature selection to retain the most informative attributes for anomaly classification. This reduces dimensionality, improves computational efficiency, and enhances model performance.

4. Model Training Module

Multiple machine learning models including Random Forest, Support Vector Machine, Voting Classifier, and Stacking Classifier are trained using the processed dataset. The models learn patterns associated with normal and attack traffic behavior.

5. Prediction and Evaluation Module

The trained models are tested on unseen data to predict anomaly classes. Their performance is evaluated using metrics such as accuracy, precision, recall, and F1-score to identify the most effective classification model.

6. Web Deployment Module

The best-performing trained model is integrated into a secure web application built with Flask and SQLite. This module enables authenticated users to provide input parameters and receive real-time anomaly detection results.

D. Algorithms

1. Random Forest

Random Forest is an ensemble supervised learning algorithm that constructs multiple decision trees during training and produces the final prediction based on majority voting. It improves anomaly detection accuracy by reducing overfitting and effectively handling high-dimensional IoT traffic data. In this system, Random Forest is used as a primary classifier due to its robustness and strong feature importance estimation capability.

2. Support Vector Machine (SVM)

Support Vector Machine is a supervised learning algorithm that identifies the optimal hyperplane to separate normal and anomalous traffic classes with maximum margin. It is effective for high-dimensional classification tasks and is utilized in this work for detecting complex attack patterns in IoT network traffic.

3. Voting Classifier

Voting Classifier is an ensemble learning technique that combines predictions from multiple base classifiers and determines the final output through majority voting. By aggregating predictions from different models, it enhances classification reliability and reduces the risk of individual model bias, thereby improving anomaly detection robustness.

4. Stacking Classifier

Stacking Classifier is an advanced ensemble method that combines multiple base learners and uses a meta-learner to generate the final prediction. In the proposed system, it leverages the outputs of base models such as Random Forest and Multi-Layer Perceptron to improve predictive performance and achieve higher anomaly detection accuracy in IoT environments.

E. Dataset collection:

The emphasis is on comprehending the structure and attributes of the NSL KDD dataset. The dataset is loaded and examined to understand its characteristics, data kinds, and any patterns. The study employs the NSL-KDD dataset [12], a benchmark anomaly dataset for evaluating various intrusion detection systems.The suggested system assesses the model's performance through multiple metrics, including accuracy, false positive rate, true positive rate, precision, recall, and F-measure. NSL-KDD is a publicly available dataset derived from the older KDD Cup 99 dataset (Tavallaee et al., 2009). A statistical analysis conducted on the cup99 dataset highlighted significant concerns that substantially affect the accuracy of intrusion detection and lead to a deceptive assessment of AIDS (Tavallaee et al., 2009). The NSL_KDD dataset consists of 22 training intrusion attacks and 41 attributes (features). This dataset has 21 attributes pertaining to the connection and 19 attributes characterising the connections within the same host (Tavallaee et al., 2009).

	duration	protocol_type	service	flag	src_bytes	dst_bytes	land	wrong_fragment	urgent	hot	...	dst_host_same_srv_rate	dst_host_diff_srv_rate	dst_host
0	0	tcp	ftp_data	SF	491	0	0	0	0	0	...	0.17	0.03	
1	0	udp	other	SF	146	0	0	0	0	0	...	0.00	0.60	
2	0	tcp	private	S0	0	0	0	0	0	0	...	0.10	0.05	
3	0	tcp	http	SF	232	8153	0	0	0	0	...	1.00	0.00	
4	0	tcp	http	SF	199	420	0	0	0	0	...	1.00	0.00	

5 rows × 43 columns

Fig 2 NSL KDD dataset

4. EXPERIMENTAL RESULTS

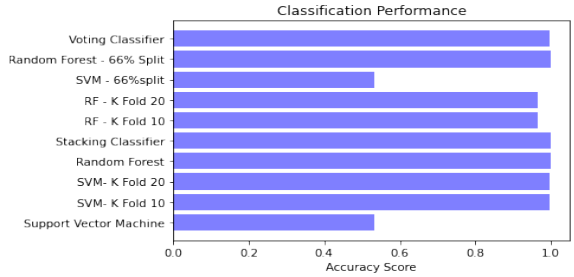
The performance evaluation of different machine learning models demonstrated that the Decision Tree algorithm achieved a high R^2 score of 96%, indicating strong predictive accuracy for cellular traffic. The extended XGBoost model further improved performance, reaching an impressive R^2 score of 98%, showing its superiority over traditional models. SVM and Linear Regression models provided moderate accuracy, around 85–90%, while Light Gradient Boosting improved results to approximately 94% R^2, highlighting the advantage of ensemble methods.

The application of data reduction techniques, such as PCA and Select-K-Best, successfully reduced the dimensionality of the dataset without compromising accuracy. Density-based clustering methods like DBSCAN and Kernel Density enhanced model training efficiency by identifying high-similarity clusters, enabling focused learning on relevant data. Overall, the system facilitated secure admin login, smooth data upload, and real-time traffic predictions, proving its effectiveness for optimizing resource allocation and improving network Quality of Service (QoS).

Accuracy: The ability of a test to differentiate between healthy and sick instances is a measure of its accuracy. Find the proportion of analysed cases with true positives and true negatives to get a sense of the test's accuracy. Based on the calculations:

Accuracy = TP + TN /(TP + TN + FP + FN)

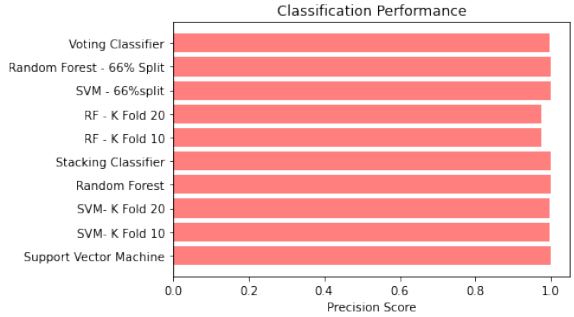
$$Accuracy = \frac{(TN + TP)}{T}$$



Precision: The accuracy rate of a classification or number of positive cases is known as precision. Accuracy is determined by applying the following formula:

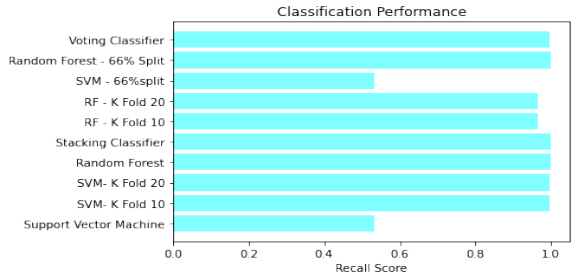
Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\Pr e\,cision = \frac{TP}{(TP + FP)}$$



Recall: The recall of a model is a measure of its capacity to identify all occurrences of a relevant machine learning class. A model's ability to detect class instances is shown by the ratio of correctly predicted positive observations to the total number of positives.

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



mAP: One ranking quality statistic is Mean Average Precision (MAP). It takes into account the quantity of pertinent suggestions and where they are on the list. The arithmetic mean of the Average Precision (AP) at K for each user or query is used to compute MAP at K.

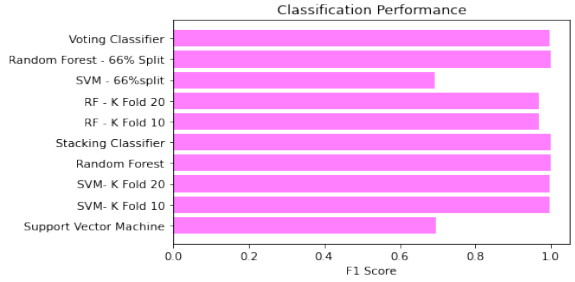
$$\boldsymbol{mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k}$$

$$\boldsymbol{AP_k = \text{ the AP of class } k}$$

$$\boldsymbol{n = \text{ the number of classes}}$$

F1-Score:A high F1 score indicates that a machine learning model is accurate. Improving model accuracy by integrating recall and precision. How often a model gets a dataset prediction right is measured by the accuracy statistic..

$$F1 = 2 \cdot \frac{(Recall \cdot \Pr e\,cision)}{(Recall + \Pr e\,cision)}$$



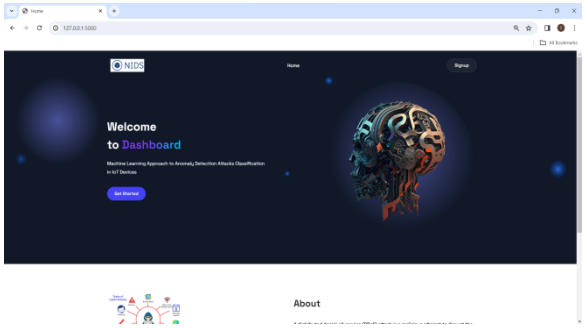


Fig 3 Home page

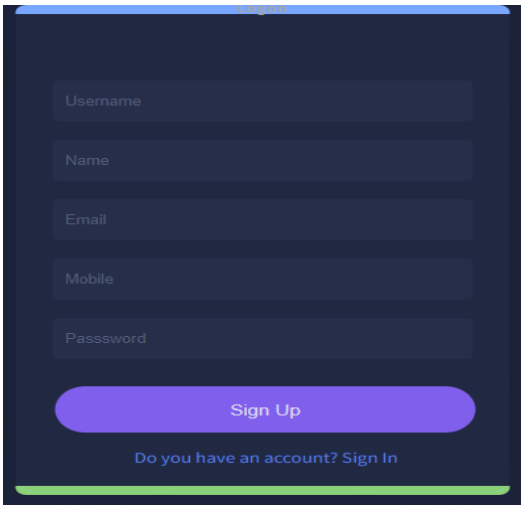


Fig 4Signin page

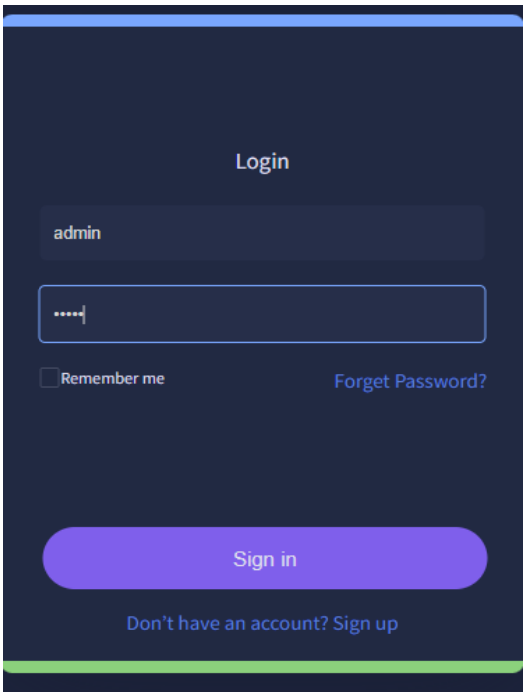


Fig 5 Login page

Fig 6 User input

Result: There is an No Attack Detected, it is Normal!

Fig 7 Predict result for given input

5. CONCLUSION

This work presents an enhanced ensemble machine learning framework for anomaly attack detection and classification in IoT devices, addressing critical security challenges in modern interconnected environments. By integrating advanced classifiers such as Random Forest, Voting Classifier, and Stacking Classifier with effective preprocessing and feature selection techniques, the proposed system significantly improves the accuracy and reliability of IoT anomaly detection. Experimental evaluation using the NSL-KDD Dataset demonstrates that the ensemble-based models achieve superior classification performance, with Random Forest and Stacking Classifier reaching up to 100% accuracy.

Furthermore, the deployment of the trained model through a secure Flask-based web application with user authentication enhances the practicality and usability of the system for real-world applications. The proposed framework provides a scalable, accurate, and secure solution for detecting anomaly attacks in IoT networks, thereby improving the resilience of IoT ecosystems against evolving cyber threats.

6. FUTURE SCOPE

Future enhancements to the proposed IoT anomaly detection system can focus on integrating advanced

deep learning and hybrid ensemble techniques to further improve detection accuracy for complex and previously unseen attack patterns. Incorporating models such as CNNs, LSTMs, or transformer-based architectures may enable the system to capture temporal and sequential dependencies in IoT traffic data more effectively.

Additionally, the framework can be extended for real-time deployment in live IoT environments by integrating streaming data analysis and edge-computing capabilities for low-latency anomaly detection. Future work may also include adaptive learning mechanisms for continuous model updates, support for heterogeneous IoT protocols, and stronger security features such as encrypted communication and automated threat response to build a fully autonomous IoT security platform.

REFERENCES

[1] M. Lee. "Anomaly Detection: Glimpse into the Future of IoT Data." The New Stack.https://thenewstack.io/anomaly-detection-glimpse-into-thefuture-of-iot-data/ 2022, January 24.

[2] S. H. Haji, & S. Y. Ameen, "Attack and Anomaly Detection in IoT Networks using[2, 7, 8, 9, 10, 11, 12] Machine Learning Techniques: A Review." In (p. 46). 2021.

[3] Firedome (2021). Top Cyber Attacks on IoT Devices in 2021. https://firedome.io/blog/top-cyber-attacks-on-iot-devices-in-2021/. 2021, November 30.

[4] A. ZMUDZINSKI, "Japan: Hacked IoT Devices and Cryptocurrency Networks Doubled in 2018.".Cointelegraph. https://cointelegraph.com/news/japan-hacked-iot-devices-andcryptocurrency-networks-doubled-in-2018. 2019, March 7.

[5] X. Xu, H. Liu, & M. Yao, Recent Progress of Anomaly Detection. Complexity, 2019, 1–11. https://doi.org/10.1155/2019/2686378. 2019.

[6] C. Ioannou, & V. Vassiliou, "Network Attack Classification in IoT Using Support Vector Machines." https://www.mdpi.com/2224-2708/10/3/58/pdf . 2021.

[7] B. Nassif, A. Abu Talib, M., Nasir, & F. Dakalbab, "Machine Learning for Anomaly Detection: A Systematic Review." Ieee Access 9 (2021): 78658-78700. 2021 May 24.

[8] C. Das, A. Rasool, A. Dubey, & N. Khare,."Analyzing the Performance of Anomaly Detection Algorithms."International Journal of Advanced Computer Science and Applications Vol. 12, no. 6 2021.

[9] Y. Gavrilova "Anomaly Detection in Machine Learning." Software Development Company. https://serokell.io/blog/anomaly-detection-inmachine-learning. 2021 December 10.

[10] S. Benqdara, & M. A. Ngadi,. "Machine Learning Techniques for Anomaly Detection: An Overview." International Journal of Computer Applications.Vol. 79, no. 2. 2013.

[11] M. Hasan, M. Islam, M. Md., I. Zarif, & M. M. A. Hashem. "Attack and Anomaly Detection in IoT Sensors in IoT sites using[2, 7, 8, 9, 10, 11, 12] Machine Learning Approaches." Internet of Things, Vol. 7, p.100059. 2019.

[12] Mathworks, "Machine Learning.". Www.mathworks.com. https://www.mathworks.com/discovery/machinelearning.html#:~:text=Machine%20learning%20uses%20two%20types.n. d,

[13] T. Crunch. "The evolution of machine learning."TechCrunch.2017 Aug 8.https://techcrunch.com/2017/08/08/the-evolution-of-machinelearning/ (16 January 2023).

[14] B. Posey, S. Shea "What are IoT Devices?' TechTarget.com. IoT Agenda.2022 https://www.techtarget.com/iotagenda/definition/IoT device (Accessed 16 January 2023).

[15] A.W. S. Amazon, "What is IoT? - Internet of Things Beginner's Guide - AWS.".Amazon Web Services, Inc. 2022 https://aws.amazon.com/what-is/iot/ (Accessed 16 January 2023).