Volume 12, No.1, January 2025
Journal of Global Research in Mathematical Archives
RESEARCH PAPER
Available online at https://www.jgrma.com
ISSN 2320 – 5822

Building Scalable Anomaly Identification Systems to IoT Threat Mitigation using Machine learning Techniques

Mani Gopalswamy
Independent Researcher,USA
manigopalsamy14@gmail.com

***Abstract*—The quick growth of IoT technologies has led to major cybersecurity issues for detecting abnormal signs that point to security risks or operational problems. The study offers a machine learning technique for identifying irregularities in Internet of Things networks by analyzing key performance indicators, such as packet loss, congestion in the lungs and bandwidth, and latency. The method includes data preprocessing as its first step, followed by SHAP-based feature importance analysis, then classification through Random Forest (RF) and Support Vector Machine (SVM). The analysis included 1000 entries before anomalies were found through the Tukey method and then classified. The experimental data shows SVM performs better than RF, producing accuracy at 96.5% with precision at 95.9% while recall reaches 96.2%, and the resulting F1-score comes out at 96.0%. SVM achieves effective anomaly detection in IoT environments according to comparison results obtained through Logistic Regression and Convolutional Neural Networks (CNN). The research shows that machine learning brings prospective improvements to IoT security because it enables preemptive anomaly detection, which results in better real-time defense capabilities against threats.**

Keywords*—IoT security, anomaly detection, machine learning, threat mitigation, scalable systems, cyber threats.

I. INTRODUCTION

Through Processing must take place very instantly since IoT sensors and smart items exchange data automatically without requiring human participation . To operate within the constrained computational budget, new procedures must be developed for any data analytics carried out via IoT. A type of data analysis called Outlier behavior classification or recognizing events is another name for the detection of anomalies, looks for unusual circumstances inside the system [1]. The techniques used to identify anomalies serve as inspections for incoming communication at several levels [2], above the lowest point to the data center of the IoT network. In the latter case, reliable detection is crucial for data cleaning and classification [3][4].

A behavior that deviates from the norm or anticipated behavior is called an anomaly. These are a few dataset patterns that don't follow their typical patterns [5][6]. One method for identifying the dataset's anomalous data pattern is anomaly detection [7]. Anomalies are referred to by a number of names for various application fields, including outliers, exceptions, surprises, unanticipated findings and oddities [8]. Applications such as credit card fraud, computer network intrusion detection systems, health care and insurance fraud detection, wireless banking sensors, and social networks all employ anomaly detection, etc [9].

Devices with limited resources cannot afford traditional security methods in the IoT [10], making security a tough task in and of itself. Perimeter security does not apply to distributed IoT networks [11], and current alternatives like the cloud have significant latency and centralization issues. Another factor contributing to this difficulty is that IoT device manufacturers frequently ignore security specifications since they have a rush-to-market mindset [12]. Additionally, the complexity of protecting IoT devices has increased due to the absence of security standards. The nature of IoT applications and these difficulties necessitate a monitoring system that can spot irregularities outside of organizational borders both at the network and device levels [13].

The IoT is an important technology that serves as the basis for a number of upcoming applications in the domains of intelligent manufacturing, transportation, and healthcare. The IoT uses many sensors detectors information about objects, people, including surroundings. Regular transmission of this data to the cloud server enables application administrators to increase the efficiency of their programs. Based on data analysis, AI technologies assist in implementing autonomous application control [14].

IoT infrastructure anomaly detection using ML and DL. The study focused on issues of privacy and security pertaining to data storage and sharing protocols in intelligent health applications [15]. The study conducted a study on ML approaches for IoMT authentication and anomaly detection systems. The paper discusses ML methods for finding irregularities to provide information on IoT network security. Additionally, emphasis is placed on Hadoop-based big data processing frameworks and identifying anomalies in IoT networks using ML techniques [16][17]. An additional explanation of how ML techniques, such as advanced ML strategies and intrusion detection systems, are used in the field of anomaly detection in the IoT [18].

A. Structure of paper

This is how remainder of the document is structured. The body of research on ML methods for IoT recognition of anomalies is reviewed in Section II. Models for classification, feature selection, preprocessing, and data collecting are all covered in Section III's approach. Section IV presents results analysis and discussion based on comparing performance across different ML models. Finally, Section V concludes the paper with potential improvements for IoT anomaly detection.

II. LITERATURE REVIEW

In this section, makes available the earlier studies on Scalable IoT Threat Anomaly Detection Systems. A comparative review of previous studies on Scalable Anomaly Detection Systems for IoT Threats is given in Table I below. It summarizes key aspects, including the methodology used, dataset employed, model performance, and limitations/future work recommendations.

Alrashdi et al. (2019) their proposed AD-IoT solution uses the RF machine learning method to identify anomalies intelligently and mitigate IoT cybersecurity vulnerabilities in smart cities. The rapid expansion of network traffic in smart cities through IoT systems generates fresh security threats because IoT devices directly link to sensors leading to cloud servers. The researchers tested their model with current data to verify its accuracy. When deployed the AD-IoT delivers a 99.34% accurate classification together with the lowest recorded false positive rate. According to their findings [19].

Gomez, Gutierrez Portela and Diaz Triana (2024) aim to develop an anomaly-based IDS specifically designed for IoT systems with low computational capacity in order to mitigate these security challenges. The results demonstrated that the Isolation Forest anomaly detection algorithm is the most suitable among those evaluated for separating attack network activity from regular traffic, with an AUC of 0.85. It is recommended to apply additional anomaly detection techniques to further optimize resources and enhance metrics. Numerous studies have attempted to address this issue using deep learning-based IDS, yet these systems remain unsuitable for low-power IoT devices [20].

Salem, Said and Nour (2024) One way to enhance IoT security is using an AI-Driven Anomaly Detection Framework. The framework improves IoT equipment reliability by minimizing operational disruptions boosting defense methods and sustaining technical performance of associated hardware. The identification of abnormal patterns through AI anomaly detection in IoT systems shows potential as a solution to deal with this issue. The research outcomes confirm how machine learning methods assist in anomaly detection capabilities through their effective operations [21].

Sana et al. (2024) focus on AD is a crucial method for spotting departures from typical system behavior that might point to IoT intrusions. With RF and Ensemble Bagged Tree, the used ML models showed exceptional training accuracy topping 99.90% accuracy, achieving balanced MCC of 99.78%, an F1-score, and an AUC of 1.00. The suggested ViT architecture greatly improved performance with 100% of all metrics, whereas the original DL LSTM model produced an accuracy of 99.97%. Additionally, it had perfect training accuracy and a validation accuracy of 78.70% [22].

Li et al. (2022) propose a organization that uses edge computing to detect anomalies in IoT networks and identify possible dangers. The LSTM autoencoder builds each detector in an unsupervised way, so it doesn't need any labeled attack data and can handle new zero-day attacks as they emerge. The test proves that ADR IoT is capable of efficiently and successfully detecting a range of assaults based on the IoT, suggesting that it might potentially help create an IoT ecosystem that is more secure [23].

Vajpayee and Hossain (2024) propose an anomaly-based method tailored for IoT, using anomaly scores, asset valuation, and controls to identify risks and offer mitigation suggestions. Effective detection relies on diverse ML techniques. Identifying high anomaly scores is crucial for informed decision-making. A tailored approach to assess security risks in IoT systems, focusing on anomaly detection, asset valuation, and risk quantification to strengthen IoT resilience against evolving threats [24].

TABLE I. COMPARATIVE TABLE FOR ANOMALY DETECTION SYSTEMS FOR IOT THREAT

Reference	Methodology	Dataset	Performance	Limitations & Future Work
Alrashdi et al. (2019)[19]	A RF-based AD-IoT solution for smart city IoT security threat detection. Focuses on analyzing network traffic anomalies introduced by IoT devices connected to large cloud servers.	Modern IoT network traffic dataset	Accuracy: 99.34%, Lowest false positive rate Effective in handling real-time IoT security threats.	Needs further validation on diverse IoT environments and scalability for large-scale smart city deployments. Performance on emerging threats like zero-day attacks is unclear.
Gomez, Gutierrez Portela and Diaz Triana (2024)[20]	Developed an anomaly-based IDS optimized for IoT systems with low computational capacity. Uses Isolation Forest algorithm to detect network traffic anomalies.	IoT network traffic dataset	AUC: 0.85 - Effective for distinguishing between normal and attack traffic.	The high computational cost of deep learning intrusion detection systems makes them inappropriate for low-power Internet of Things devices. To maximize efficiency and enhance detection precision, more anomaly detection methods are required.
Salem, Said and Nour (2024)[21]	The following Models are used in the deployment of interference uncovering systems based on ML: RF, ViT, LSTM, and Ensemble Bagged Tree. Tunes hyperparameters via Bayesian optimization. Focuses on reducing system downtime, improving security, and ensuring performance consistency.	IoT security datasets	Enhanced anomaly detection capability. Improves IoT system reliability and response time.	Lack of details on computational efficiency when deployed on resource-constrained IoT devices. - Requires benchmarking against deep learning models to assess comparative performance.
Sana et al. (2024)[22]	The following in order to develop machine learning-based intrusion detection systems, models are used: RF, ViT, LSTM, and Ensemble Bagged Tree. Tunes hyperparameters via Bayesian optimization.	IoT security datasets	Random Forest & Ensemble Bagged Tree: 99.90% accuracy, AUC = 1.00, F1-score = 99.78%. - LSTM model: 99.97% accuracy. - ViT model: 100% accuracy, 78.70% validation accuracy.	ViT model validation accuracy is lower than training accuracy, suggesting overfitting risks. - Needs real-world testing on large-scale IoT environments.
Li et al. (2022)[23]	LSTM Autoencoder-based anomaly detection framework with edge computing integration. Designed to	IoT network traffic dataset	Effectively detects previously unseen IoT attacks. Handles zero-day vulnerabilities and emerging threats.	Performance depends on high-quality training data. Requires further testing for large-scale IoT network deployments.

	detect zero-day attacks without requiring labeled data.			
Vajpayee and Hossain (2024)[24]	Anomaly-based detection method integrating asset valuation, risk quantification, and anomaly scores to assess IoT security threats. Uses multiple ML techniques to improve detection efficiency.	IoT security datasets	Identifies high anomaly scores crucial for security decision-making. Provides a tailored approach for risk assessment in IoT networks.	Needs validation on real-world IoT infrastructures. Effectiveness against adaptive cyber threats remains to be tested.

A. Research Gaps

Despite significant advancements in Scalable Anomaly Detection Systems for IoT Threats, several critical gaps remain unaddressed. First, most existing studies focus on improving detection accuracy but overlook the real-time computational efficiency required for resource-constrained IoT environments. While deep learning models like LSTM and ViT demonstrate high accuracy, their high computational cost makes them impractical for low-power IoT devices. Second, the fact that a lot of models are based on pre-labeled datasets limits their capacity to identify zero-day attacks and adjust to changing threats in actual IoT networks. Although unsupervised learning approaches such as LSTM autoencoders and Isolation Forest offer potential solutions, their detection capabilities still require optimization to reduce false positive rates and enhance scalability. Third, existing solutions often lack a comprehensive risk quantification mechanism that integrates anomaly scores with asset valuation to prioritize threats effectively. While some frameworks address risk assessment, they do not fully integrate with real-world IoT security infrastructures, limiting their practical implementation. Additionally, many studies do not explore the impact of adversarial attacks on anomaly detection models, leaving systems vulnerable to sophisticated cyber threats. Therefore, Future studies must concentrate on lightweight, flexible, and explainable AI-based intrusion detection systems that balance accuracy, efficiency, and real-time applicability while improving risk-based threat mitigation strategies.

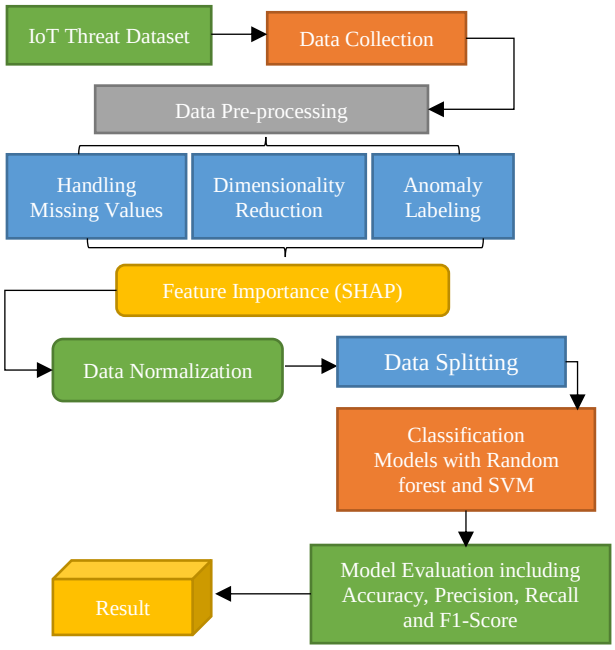

Fig. 1. Methodology Flow Diagram for Anomaly Detection in IoT

III. METHODOLOGY

The AD methodology in IoT threats proceeds through a systematic framework starting with IoT threat dataset accumulation followed by data preprocessing for value handling and dimension reduction and anomaly identification step. SHAP analytical methods determine which features hold the most prominent influence on the classification evaluation. In order to attain performance consistency and enhance model outcomes, the data is normalized. Due to their capacity to forecast across non-linear connections and analyze intricate patterns, RF and SVM are used for anomaly identification. In the final evaluation step, the models' effectiveness is assessed utilizing important assessment criteria like accuracy, precision, recall, and F1-score. As shown in Figure 1, the procedure concludes with a results analysis to evaluate the model's capacity to recognize IoT-based security concerns.

The data flow diagram's subsequent phases, which are illustrated below, need a thorough explanation:

A. Data Collection

This study utilized network data that researchers obtained from an experimental network system that duplicated genuine network situations. The virtual network topology emerged from using EVE-NG tool version 4.0.1-86-PRO to connect various nodes such as routers and servers. A customized network comprised the data collection system. Network metrics obtained from simulations were organized into a CSV file structure within the dataset. Media encoding at different bitrates through FFmpeg version 4.4.2 created video streams that were transmitted between servers to establish diverse network simulations for anomaly detection purposes.

B. Data Pre-processing

Data pre-processing operations were performed to enhance the dataset's quality and reliability prior to training, resulting in several modifications to the acquired data. The steps included:

- **Handling Missing Values:** The analysis removed instances containing data gaps and filled missing or inconsistent values through statistical methods. Network congestion, along with packet loss and throughput variations, formed the basis for the main features which underwent analysis.
- **Dimensionality Reduction:** The application of PCA resulted in dimension reduction of the dataset through which researchers maintained the most critical data variations.
- **Anomaly Labeling:** The Tukey approach examined the data to identify global anomalies before supervised learning models could effectively classify anomalies.

C. Feature Importance

In Feature reputation to determine the most relevant features influencing anomaly detection, SHapley Additive exPlanations (SHAP) were used. In ML, feature significance is a technique used to assess how each input information

affects the model's predictions. It helps improve model interpretability, optimize performance, reduce dimensionality, and detect biases.

D. Data Normalization

Normalizing data is important since network metrics had varying scales and units [25], data normalization was performed to guarantee consistency and enhance model functionality. The Min-Max Scaling technique was applied, transforming all numerical values into a range between 0 and 1. The Normalization is calculated as Equation (1):

$$Xnorm = X - XminXmax - XminX_{\{norm\}} = frac\{X - X_{\{min\}}\}\{X_{\{max\}} - X_{\{min\}}\} \quad (1)$$

This standardization helped eliminate bias caused by differing feature scales, facilitating more accurate anomaly detection.

E. Data Splitting

The dataset was divided into two parts: 20% for testing and 80% for training. While the testing data was used to assess the machine learning models' performance on unseen data, the training data was used to build the models.

F. Classification Models

Several ML models were employed to classify network anomalies, each offering unique strengths and trade-offs in terms of precision, comprehensibility, and mathematical effectiveness:

1) Random Forest

In order to enhance prediction performance, RF builds many decision trees and aggregates their results using an ensemble learning technique. It works essentially like training many trees on different subsets of the dataset, incorporating randomness in feature selection along with sampling the data. As this ensemble technique reduces the risk of overfitting, it's far more robust to noisy data and higher-dimensional feature spaces [26]. In anomaly detection, RF is able to learn decision boundaries based on historical data and thus can distinguish between normal and anomalous network behavior seen in Figure 2. Each of the several decision trees that make up the model structure was trained using a bootstrapped sample of the dataset. To increase variety among trees and lower bias and variance, a haphazard subset of traits is taken into consideration at each split. In inference, each tree classifies the input independently, and the final prediction is achieved through Regression tasks use averaging, whereas assignments involving classification use voting with the majority. In addition, this mechanism is effective for IoT anomaly detection since it captures non-linear relationships and detects subtle variations in network metrics. One of the contributing factors to RF effectiveness is its robustness against high dimensional data as well as its feature selection ability in removing noisy or irrelevant features. In comparison to real data is less prone to overfit than its conventional analogue, DT, and strikes a reasonable balance between recall and precision. In this way, it would be able to avoid getting false positives for many of the anomalies, thereby increasing its reliability. RF will be a perfect choice for such real-time IoT security applications that exigency for efficiency in computation and accuracy in detecting are necessary.

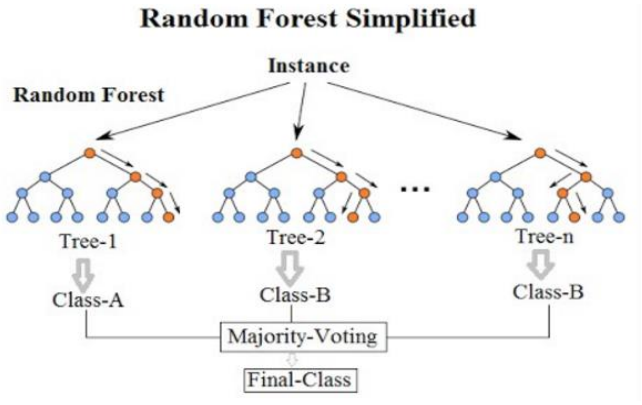

Fig. 2. The Architecture of Random Forest

2) Support Vector Machine (SVM):

SVM is a powerful prototypical that has an ideal hyperplane for data classification. As an alternative to conventional models like RF or DT, SVM models use a kernel function to change the features of inputs into higher dimensions for its powerful ability to handle complex, non-linearly separable data shown in Figure 3. This aspect proves it to be very useful in the practice of anomaly detection since malicious patterns are usually quite intricate and cannot be detected by simpler models.

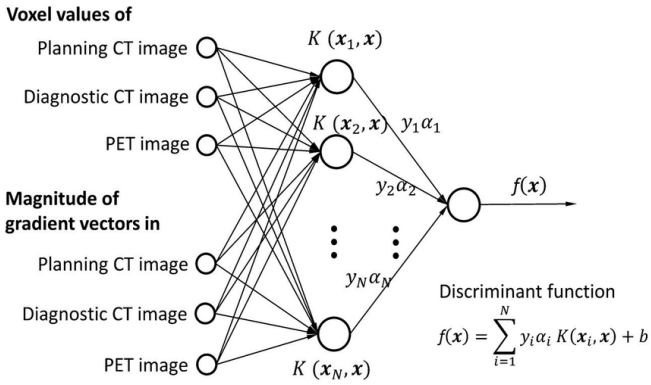

Fig. 3. The Structure of SVM Model

An SVM model's structure includes a hyperplane that divides normal and anomalous instances and Support vectors are important pieces of information that define the decision boundary. This hyperplane's location is established by optimizing the margin, guaranteeing that the model performs effectively when applied to previously unknown data. Different kernel functions, for instance, the linear, polynomial, and RBF, enable SVM to adapt to different data distributions, and this is why it is considered a much more advanced concept than traditional classification models. The training process involves solving a quadratic optimization problem, which becomes computationally expensive as the dataset size increases. These are challenges despite its great utility in IoT anomaly detection security issues, particularly deep pattern recognition needs and high-accuracy classification scenarios.

G. Performance Metrics and Model Evaluation

To select the evaluation metric, to properly analyze the model, it is essential to comprehend how each metric is measured. The objective was to evaluate the effectiveness of ML techniques by examining each of these capabilities metrics: accuracy score, precision, recall, and F1 score.

1) Accuracy

The percentage of cases the accuracy of The categorization of the representation is the total error in class prediction. However, performance may be misrepresented by biased data. A classifier may correctly anticipate instances of the majority class while incorrectly categorizing cases of the minority class. The accuracy is calculated as Equation (2):

$$Acuracy = \frac{TP+TN}{TP+TN+FN+FP} \quad (2)$$

2) Precision

Once all of the data has been categorized, precision is the proportion of instances that are correctly assigned to a class. In this instance, it shows the proportion of corona cases that actually are corona cases. The precision is calculated as Equation (3):

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

3) Recall

The number of instances that are accurately classified into a class is determined by recall or sensitivity.. This context measures the proportion of properly represented instances by the classifier among all carriers of the illness. The recall is calculated as Equation (4):

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

4) F1-Score

The F1-score, often called the F-measure, is the weighted choral mean of accurately and recall. This metric is best suited for usage when the dataset is significantly unbalanced. The F1-score is calculated as Equation (5):

$$F1 - score = 2 * \frac{precision*recall}{precision+recall} \quad (5)$$

Where,

- **TP (True Positive):** The frequency with which a positive class was accurately anticipated by the mathematical framework.
- **TN (True Negative):** This graphic shows how often The model projected the negative class accurately.
- **FP (False Positive):** The model failed to accurately predict the positive class in some cases.
- **FN (False Negative):** The quantity of false forecasts made by the model that anticipated the negative class.

IV. Results And Discussion

In this segment, the simulated results of anomaly detection of IoT risks using ML techniques are discussed. The findings of the dataset evaluation conducted for this study are represented in this part, which also includes performance metrics, classifier statistics, results, and a description of the dataset.

A. Dataset Description

There were about 1000 entries in the dataset, which were key features of throughput, congestion, packet loss, latency, and jitter. Each network state is represented as a data point recorded in a specific condition, including base scenarios, single video streaming, dual video streaming, or adaptive video streaming. The performance thresholds at which the network is tested span a wide range of data points to protect from anomalies. Experimental tests showed that packet loss and latency experienced increased with increasing congestion values, greatly affecting the network's performance.

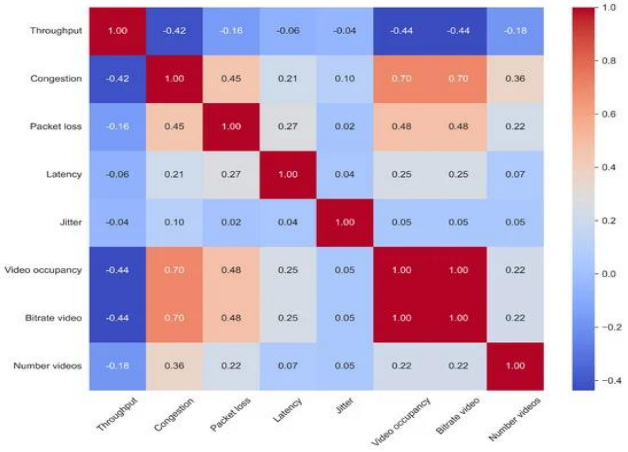

Fig. 4. Correlation Heatmap of Network Parameters in IoT Systems

As seen in Figure 4, several network parameters like throughput, congestion, packet loss, latency, jitter, video occupancy, bitrate video, and number of videos are correlated with each other in an IoT system. Red displays is a strong positive correlation, while blue shows a strong negative correlation. For example, video occupancy and congestion (0.70) is quite a high positive correlation indicating that higher congestion is related to greater video occupancy. On the contrary, throughput (-0.42) and congestion are negatively related in that as congestion increases, throughput decreases.

B. Experiment Results

The findings in this unit are the usage of ML in scalable DL for IoT using ML applied to a large dataset by using the RF and SVM models. As the classification models were trained and evaluated using the anomalies effectively labeled by the Tukey method, anomalies could be effectively labeled by the Tukey method.

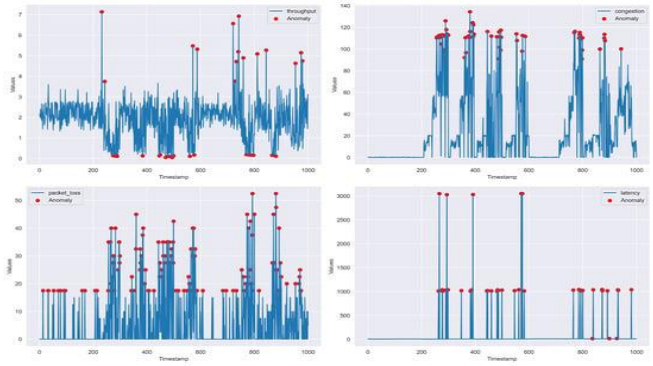

Fig. 5. Anomaly Detection in IoT Networks Using Time-Series Data Analysis

In Figure 5, the throughput, congestion, and packet loss of the IoT network, and delay are displayed in four time-series charts that demonstrate anomaly identification. The normal data variations in time are represented by the blue lines, red dots being the detected anomalies. Most of these anomalies occur when the network is behaving in an unusual manner, potentially an indication of a threat or inefficiency in the network. This finds such visualizations crucial for discovering irregular patterns in the IoT traffic to help with securing its traffic proactively and optimizing networks.

TABLE II. RANDOM FOREST AND SVM MODELS PERFORMANCE MATRICES FOR ANOMALY DETECTION FOR IOT THREAT

Model	Accuracy	Precision	Recall	F1-Score
Random Forest (RF)	94.3	93.8	94.0	93.9
Support Vector Machine (SVM)	96.5	95.9	96.2	96.0

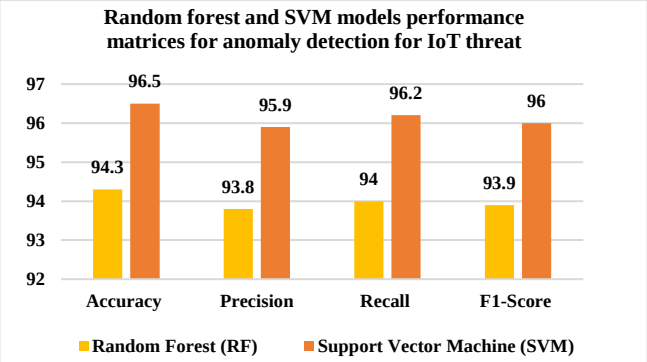

Fig. 6. Random forest and SVM models performance matrices for anomaly detection for IoT threat

Table II and Figure 6 provide a comparison of SVM and RF models for identifying irregularities related to IoT security risks. With an accuracy of 96.5%, precision of 95.9%, recall of 96.2%, and F1 score of 96.0%, the SVM model is shown to perform better than the RF model, which has an accuracy of 94.3% and lower accuracy in other metrics. The results indicate that SVM is better at finding the anomalies obtaining higher detection accuracy and reliability. Although SVM has a high performance, RF is still a competitor with less computational demands, pleasingly suitable for real time IoT, as shown in Figure 6.

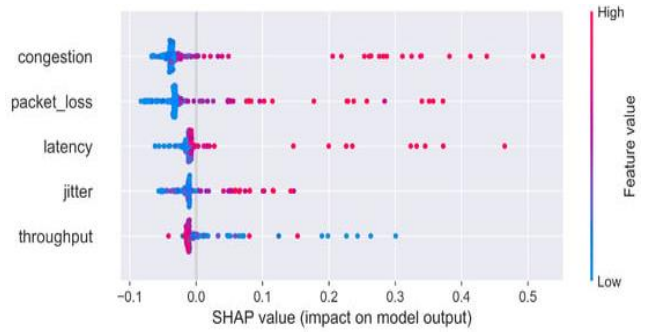

Fig. 7. SHAP Value Analysis for IoT Network Performance Metrics in Machine Learning Models

Figure 7 visualizes the impact of different network performance metrics on an ML model's prediction. The following Throughput along with congestion along with packet loss and delay and jitter are presented on the vertical axis beneath the x-axis that shows the SHAP values which represent each feature's effect on the predicted model results. The colored dots on the plot represent data points whose attributes match the feature value color; low values are blue, while high values are red. The prediction is more heavily influenced by features with higher SHAP values, which can make it easier to understand how the model makes decisions.

C. Comparative Analysis and Discussion

The classification Among the key performance metrics utilized to Accuracy and precision together with recall and F1-score were used to evaluate the models. A summary of the LR, CNN, RF, and SVM models' machine learning performance in the study on anomaly detection for IoT risks is shown in Table III below.

TABLE III. COMPARISON BETWEEN VARIOUS MODELS FOR ANOMALY DETECTION FOR IOT THREAT

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression (LR) [27]	68.97	75.07	62.76	67.86
CNN [28]	94.2	-	91.7	93.0
Random Forest (RF)	94.3	93.8	94.0	93.9
Support Vector Machine (SVM)	96.5	95.9	96.2	96.0

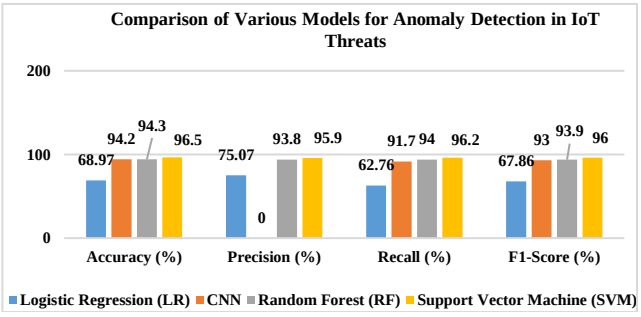

Fig. 8. Comparison of Various Models for Anomaly Detection in IoT Threats.

The comparison of many ML models used to identify irregularities in IoT security is displayed in Figure 8. Additionally, the examined models include CNN, SVM, LR, and RF. It has 96.0% F1 score, 96.2% recall, 95.9% precision, and 96.5% accuracy. CNN achieves a high accuracy of 94.2% without any precision data, and similar metrics to RF are used. With its lowest accuracy of 68.97, LR is a simpler model that is ineffective in identifying abnormalities in the IoT. According to the results, SVM and RF are the best models for detecting anomalies in IoT scenarios where threat identification requires high precision and recall.

V. CONCLUSION AND FUTURE SCOPE

Incorporated into this problem are the security challenges of IoT devices rapidly proliferating and, therefore, require robust and scalable anomaly detection systems. This introduced an ML-based method that was characterized by accuracy, scalability, and real-time processing to identify and reduce possible IoT hazards. The suggested technique improved security overall and successfully identified abnormalities at the cost of false positives. By optimizing ML algorithms and IoT infrastructure, the plant's operating efficiency was significantly increased. In order to decrease unplanned downtime and improve resource utilization, accurate anomaly detection algorithms were adjusted and optimized with lower false favorable rates. The study's findings are consistent with other The researcher conducts investigations related to IoT security combined with anomaly detection techniques in industrial environments. Early anomaly detection and improved security in IoT systems section for IoT with the use of ML-based security solutions worked well.

Future research on DL models may be applied to improve detection accuracy and flexibility in response to evolving threats. In addition, using federated learning adds privacy, while providing the capability of joint threat intelligence. Although the deployment of lightweight models designed for IoT devices with limited resources is an important problem, this should be done in a way that enables real-time threat

mitigation without sacrificing performance. Also, better decision-making can be achieved with trusted and explainable AI in anomaly detection systems.

REFERENCES

[1] A. Chatterjee and B. S. Ahmed, "IoT anomaly detection methods and applications: A survey," 2022. doi: 10.1016/j.iot.2022.100568.

[2] V. Pillai, "Anomaly Detection for Innovators: Transforming Data into Breakthroughs," *Lib. Media Priv. Ltd.*, 2022.

[3] H. Elazhary, "Internet of Things (IoT), mobile cloud, cloudlet, mobile IoT, IoT cloud, fog, mobile edge, and edge emerging computing paradigms: Disambiguation and research directions," 2019. doi: 10.1016/j.jnca.2018.10.021.

[4] M. S. Akaash Vishal Hazarika, "Blockchain-based Distributed AI Models: Trust in AI model sharing," *Int. J. Sci. Res. Arch.*, vol. 13, no. 2, pp. 3493–3498, 2024.

[5] S. Chatterjee, "Mitigating Supply Chain Malware Risks in Operational Technology : Challenges and Solutions for the Oil and Gas Industry," *J. Adv. Dev. Res.*, vol. 12, no. 2, pp. 1–12, 2021.

[6] Rajarshi Tarafdar, "AI-Powered Cybersecurity Threat Detection in Cloud," *Int. J. Comput. Eng. Technol.*, p. 266, 2025.

[7] A. V. Hazarika, M. Shah, S. Patil, and N. Carolina, "Risk Management for Distributed Arbitrage Systems : Integrating Artificial Intelligence," *arXiv Prepr. arXiv2503.18265*, 2025, doi: https://doi.org/10.48550/arXiv.2503.18265.

[8] V. Prakash, O. Odedina, A. Kumar, L. Garg, and S. Bawa, *A secure framework for the Internet of Things anomalies using machine learning*, vol. 4, no. 1. Springer International Publishing, 2024. doi: 10.1007/s43926-024-00088-z.

[9] A. Immadisetty, "Machine Learning for Real-Time Anomaly Detection," *Int. J. Multidiscip. Res.*, vol. 6, no. 6, 2022.

[10] S. Duary, P. Choudhury, S. Mishra, V. Sharma, D. D. Rao, and A. Paul Aderemi, "Cybersecurity Threats Detection in Intelligent Networks using Predictive Analytics Approaches," in *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, IEEE, Feb. 2024, pp. 1–5. doi: 10.1109/ICIPTM59628.2024.10563348.

[11] P. Piyush, N. S. Gill, P. Gulia, D. D. Rao, Y. Mandiga, and P. K. Pareek, "Systematic Analysis of threats, Machine Learning solutions and Challenges for Securing IoT environment," *J. Cybersecurity Inf. Manag.*, vol. 14, no. 2, pp. 367–382, 2024, doi: 10.54216/JCIM.140227.

[12] N. Malali and S. R. Praveen Madugula, "Robustness and Adversarial Resilience of Actuarial AI/ML Models in the Face of Evolving Threats," *Int. J. Innov. Sci. Res. Technol.*, vol. 10, no. 3, pp. 910–916, Mar. 2025, doi: 10.38124/ijisrt/25mar1287.

[13] A. Diro, N. Chilamkurti, V. D. Nguyen, and W. Heyne, "A comprehensive study of anomaly detection schemes in iot networks using machine learning algorithms," 2021. doi: 10.3390/s21248320.

[14] S. S. S. Neeli, "Critical Cybersecurity Strategies for Database Protection against Cyber Attacks," *J. Artif. Intell. Mach. Learn. Data Sci.*, vol. 1, no. 1, p. 5, 2023.

[15] M. I. Khan, A. Arif, and A. R. A. Khan, "AI-Driven Threat Detection: A Brief Overview of AI Techniques in Cybersecurity," *BIN Bull. Informatics*, vol. 2, no. 2, pp. 248–261, 2024.

[16] S. H. Rafique, A. Abdallah, N. S. Musa, and T. Murugan, "Machine Learning and Deep Learning Tework Anomaly Detection—Current Research Trendschniques for Internet of Things Net," *Sensors*, vol. 24, no. 6, 2024, doi: 10.3390/s24061968.

[17] M. K. A Arif, A Khan, "Role of AI in Predicting and Mitigating Threats: A Comprehensive Review," *JURIHUM J. Inov. dan Hum.*, vol. 2, no. 3, pp. 297–311, 2024.

[18] P. Schummer, A. del Rio, J. Serrano, D. Jimenez, G. Sánchez, and Á. Llorente, "Machine Learning-Based Network Anomaly Detection: Design, Implementation, and Evaluation," *AI*, vol. 5, no. 4, pp. 2967–2983, 2024, doi: 10.3390/ai5040143.

[19] I. Alrashdi, A. Alqazzaz, E. Aloufi, R. Alharthi, M. Zohdy, and H. Ming, "AD-IoT: Anomaly detection of IoT cyberattacks in smart city using machine learning," in *2019 IEEE 9th Annual Computing and Communication Workshop and Conference, CCWC 2019*, 2019. doi: 10.1109/CCWC.2019.8666450.

[20] A. S. Gomez, F. Gutierrez Portela, and O. A. Diaz Triana, "Intrusion Detection System for IoT using Anomaly Detection Techniques," *7th Int. Congr. Ambient Intell. Softw. Eng. e-Health Mob. Heal. AmITIC 2024*, pp. 2024–2025, 2024, doi: 10.1109/AmITIC62658.2024.10747648.

[21] S. A. Salem, S. A. Said, and S. M. Nour, "AI-Driven Anomaly Detection Framework for Improving IoT System Reliability," *2024 IEEE Glob. Conf. Artif. Intell. Internet Things, GCAIoT 2024*, no. Ml, pp. 6–8, 2024, doi: 10.1109/GCAIOT63427.2024.10833531.

[22] L. Sana *et al.*, "Securing the IoT Cyber Environment: Enhancing Intrusion Anomaly Detection With Vision Transformers," *IEEE Access*, vol. 12, pp. 82443–82468, 2024, doi: 10.1109/ACCESS.2024.3404778.

[23] R. Li, Q. Li, J. Zhou, and Y. Jiang, "ADRIoT: An Edge-Assisted Anomaly Detection Framework Against IoT-Based Network Attacks," *IEEE Internet Things J.*, 2022, doi: 10.1109/JIOT.2021.3122148.

[24] P. Vajpayee and G. Hossain, "Risk Assessment of Cybersecurity IoT Anomalies Through Cyber Value at Risk (CVaR)," *2024 IEEE 5th World AI IoT Congr. AIIoT 2024*, pp. 77–83, 2024, doi: 10.1109/AIIoT61789.2024.10578956.

[25] B. Boddu, "Scaling Data Processing with Amazon Redshift Dba Best Practices for Heavy Loads," *Int. J. Core Eng. Manag.*, vol. 7, no. 7, 2023.

[26] Y. H. Rajarshi Tarafdar, "Finding Majority for Integer Elements," *J. Comput. Sci. Coll.*, vol. 33, no. 5, pp. 187–191, 2018.

[27] A. Jaramillo-Alcazar, J. Govea, and W. Villegas-Ch, "Anomaly Detection in a Smart Industrial Machinery Plant Using IoT and Machine Learning," *Sensors*, 2023, doi: 10.3390/s23198286.

[28] I. Priyadarshini, "Anomaly Detection of IoT Cyberattacks in Smart Cities Using Federated Learning and Split Learning," *Big Data Cogn. Comput.*, vol. 8, no. 3, 2024, doi: 10.3390/bdcc8030021.

