# ML and Blockchain Framework for Secure IoT Anomaly Detection

1st Shan Rasool
*Department of Information and Communication Engineering*
*The Islamia University of Bahawalpur*
shanrasool2002@gmail.com

2nd Abdul Basit
*Department of Information and Communication Engineering*
*The Islamia University of Bahawalpur*
abdulbasit9778@gmail.com

3rd Aoun Muhammad
*Department of Information and Communication Engineering*
*The Islamia University of Bahawalpur*
aoun.muhammad@iub.edu.pk

4th Umar Fayyaz
*Department of Information and Communication Engineering*
*The Islamia University of Bahawalpur*
umar.fayyaz@iub.edu.pk

5th Sehrish Raza
*Department of Information and Communication Engineering*
*The Islamia University of Bahawalpur*
sehrishraza6@gmail.com

***Abstract*—The Internet of Things (IoT) has become part of modern life, linking many devices across homes, industries, cities and healthcare systems. While this connectivity improves efficiency and convenience, it also indicate networks to a growing range of cyber threats. Many IoT devices are lightweight, inexpensive and limited in processing capability, meaning they cannot support complex security software. This makes them winning targets for attackers who begin Distributed Denial of Service (DDoS) attacks, manipulate data or attempt to disrupt critical services. Traditional centrally managed security systems struggle under this scale and are themselves vulnerable as single points of failure. In addition, the integrity of logs and training data used by automated detection systems can be compromised, which may mislead machine-learning models during operation. This paper presents a hybrid security framework that combines machine learning for anomaly detection with blockchain technology for secure logging of security events. Machine-learning algorithms are used to classify network activity as normal or malicious, while blockchain stores the corresponding detection records in an immutable ledger to prevent later alteration. Three machine-learning models—Random Forest, Support Vector Machine (SVM) and XGBoost—are determined using an IoT-Blockchain dataset. The results show that SVM delivers the best performance, achieving 80.00% accuracy and 100% recall, meaning that all attack instances in the test set were detected. Random Forest performs competitively, whereas XGBoost records lower accuracy on this relatively small and lightweight dataset. The main contribution of this work lies in demonstrating that combining anomaly detection with blockchain-based integrity protection offers a practical pathway to more trustworthy IoT security. The approach not only identifies malicious behavior but also preserves reliable proof of detected threats. This framework can support future large-scale IoT deployments and motivates further research on optimizing blockchain overhead and expanding datasets with emerging attack types.**

***Index Terms*—IoT security, Anomaly detection, Blockchain, Machine Learning, SVM, Cybersecurity**

## I. Introduction

The Internet of Things (IoT) has changed the way many system that work today. Homes, hospitals, factories, vehicles, farms and cities now depend on large number of connected sensor and smart devices that constantly collect and exchange data. The number of these devices continues to grow every year, creating very large, highly distributed networks. This wide connectivity bring many advantages such as automation, remote monitoring and faster decision-making. However, this steady connection also create many openings for cyberattacks. A large portion of IoT devices are small, inexpensive, battery-powered and limited in processing ability. Because of this, they often lack strong encryption or advanced security software. Attackers take advantage of these vulnerability to launch Distributed Denial of Service (DDoS) attacks, build botnets, steal information or modify transmitted data [1], [2]. When IoT systems are used in healthcare, transportation or industry, such attacks can lead not only to service failure but also to serious safety and privacy risks.

### A. Problem Statement

Most current security solutions for IoT rely on centralized servers or cloud systems that collect and analyze network traffic. Although these system are powerful, they also create a single critical point that can fail or be hacked. If this central server stops working, the whole monitoring system may collapse [3].Machine Learning (ML) has recently been adopted to improve intrusion detection systems because it can automatically learn patterns and recognize abnormal traffic. However, ML-based systems also face an important challenge that they depend on the quality and integrity of the data used for training and logging. If attackers modify logs or poison training data, the ML model can be tricked into allowing


DOI: https://doi.org/10.67535/tsp.000002.010
Published under the ISAISS 2026 Proceedings

malicious activity [4]. Therefore, the key problem is not only detecting attacks but also ensuring the security of the records used to make those decisions.

### B. Proposed Solution

To overcome these issues, this paper proposes a combined framework that brings together machine learning and blockchain technology. In the proposed system, ML algorithms analyze IoT network traffic and classify it as normal or attack. At the same time, blockchain is used to store the detection results and important security events in a decentralized record. Because blockchain records cannot easily be altered, this approach ensures that attackers cannot secretly modify logs or erase evidence once it has been stored [5]. This combined approach aims to provide both effective anomaly detection and strong data integrity, which is essential for secure IoT environments.

### C. Contributions

The main contributions of this paper are as follows:

*1) Hybrid Architecture:* We present a hybrid IoT security framework that integrates blockchain to protect data integrity and machine learning to detect abnormal network behavior in real time.

*2) Comparative Analysis:* We perform a comparative study of three machine learning algorithms—Random Forest, Support Vector Machine (SVM) and XGBoost—using an IoT-Blockchain security dataset to identify the most suitable model for this scenario.

*3) Experimental Validation:* We experimentally show that SVM achieves the highest recall rate, making it especially suitable for applications where undetected attacks are unacceptable, while also discussing the strengths and limitations of the other models [6].

## II. RELATED WORK

The security of IoT networks has been a subject of extensive research, primarily focusing on two distinct approaches: Machine Learning (ML) for intrusion detection and Blockchain for data integrity.

### A. Machine Learning in IoT Security

Anomaly detection has traditionally relied on statistical methods, but recent literature emphasizes the shift toward data-driven ML approaches. Khan et al. [2] conducted a comprehensive survey of anomaly detection techniques, highlighting that while traditional firewalls are effective against known signatures, they fail against zero-day attacks. To address this, Chen and Ji [4] explored various supervised learning models, noting that algorithms like Random Forest and Support Vector Machines (SVM) offer a good balance between computational efficiency and detection accuracy in resource-constrained IoT environments. Furthermore, Liu and Lang [6] demonstrated that while Deep Learning (DL) models can achieve higher accuracy, they often require computational resources that edge IoT devices lack, suggesting a need for lighter, more efficient classifiers.

### B. Blockchain for Data Integrity

Parallel to ML developments, Blockchain has emerged as a solution for the centralized trust issues in IoT. Al-Kahtani and Ghazal [1] surveyed blockchain applications, arguing that the decentralized nature of the ledger eliminates the single point of failure common in cloud-based IoT systems. Zhang [3] proposed a specific framework where blockchain acts as a verification layer for IoT data, ensuring that logs cannot be tampered by external attackers. Similarly, Gupta and Tanwar [5] introduced a decentralized mechanism to integrity-check data packets before they are processed by central servers. Current research also focuses on reducing the computational overhead of blockchain. Kumar and Singh [14] proposed a lightweight consensus mechanism specifically designed for sensors with limited battery life, which aligns with our proposed edge-based architecture.

### C. Research Gap

While available studies have separately addressed anomaly detection [4] and data integrity [5], few have integrated both into a unified architecture. Most current frameworks either detect attacks without securing the logs or secure the data without active threat detection. This paper bridges this gap by proposing a hybrid framework that utilizes SVM for real-time anomaly detection while anchoring the results on a blockchain ledger to ensure audit-ability.

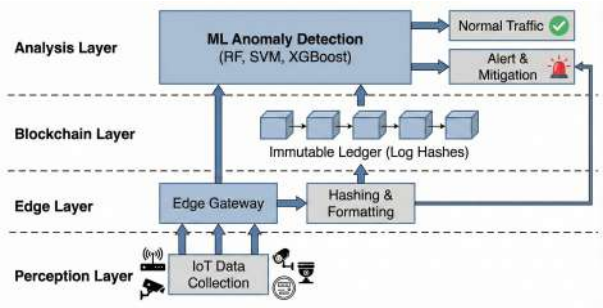
## III. METHODOLOGY

This section explains how the proposed framework is designed, how the dataset is prepared and how the machine-learning models are applied for anomaly detection. The goal is to clearly show each step from raw IoT data to final classification and secure storage on blockchain.

### A. Dataset Description and Preprocessing

The experiments in this study make use of an IoT-Blockchain Security Dataset that represents communication between different IoT devices and network components. The dataset contain records that describe both normal activity and different attack possibilities. Each entry include features such as data size, processing time, energy usage and consensus mechanism along with a label showing whether a threat was mitigated or not. Although the dataset used in this study contains 1,000 records which are the relatively small as we compared to large-scale ML datasets, this size is typical in IoT security research. Real-world IoT attack data are often difficult to collect due to privacy, deployment and labeling constraints. In many IoT environments, especially at the edge, only limited annotated traffic is available. Therefore, the dataset reflects realistic operational conditions rather than ideal laboratory settings.

Before training machine-learning models, the dataset was cleaned and prepared. This process takes place in several stages:




*1) Feature Selection:* Non-useful attributes like device identifiers are removed. These fields do not help the learning process and can cause the model to memorize specific devices instead of learning general patterns of malicious behavior.

*2) Encoding:* Many columns in IoT data are text-based such as IoT layer, type of request or type of threat (for example DDoS or Sybil). Since machine-learning models work with numbers so these values are converted into numeric form using label encoding.

Figure illustrates the interaction among these four layers. Data first moves from IoT devices to the edge, where it is organized. The information is then written to blockchain for integrity protection and finally processed by machine-learning models for anomaly detection. The arrows in the figure highlight how detection and immutability work together to form a continuous security loop.

*3) Normalization:* Some features naturally have larger values than others (for example, data size versus processing time). To prevent these large-scale values from dominating learning all numerical attributes are normalized so that they follow the same range and distribution.These preprocessing steps ensure that the dataset is clean, consistent and suitable for machine-learning algorithms

### B. Proposed Blockchain-ML Framework

The proposed system brings together blockchain and machine learning in a layered architecture to protect IoT networks against intrusions and tampering.The framework contains four main layers:

*1) Perception Layer:* Physical IoT sensors and devices collect data and generate logs of their activities.

*2) Edge Layer:* Gateways gather device data, perform simple formatting and prepare transactions.

*3) Blockchain Layer:* Before the data is analyzed, hashes of the records are stored in a private blockchain. This guarantees that logs cannot be quietly changed later, protecting the learning process against data poisoning.

*4) Analysis Layer:* The machine-learning models (Random Forest, SVM and XGBoost) examine traffic patterns and classify events as normal or attack. When an anomaly is detected, the system verifies the log against blockchain to ensure its authenticity before raising an alert.

### C. Machine Learning Classifiers

Three distinct algorithms were selected to evaluate the balance between computational efficiency and detection accuracy:

- Random Forest (RF):
  Random Forest is an ensemble of multiple decision trees. Each tree learns slightly different parts of the data and their combined vote produces the final prediction. It is chosen for its robustness and ability to manage complex nonlinear relationships in IoT traffic.
- Support Vector Machine (SVM):
  SVM finds the best separating boundary between normal and attack classes. It works particularly well with high-dimensional data and small to medium-sized datasets, which fits the nature of IoT environments.
- XGBoost:
  XGBoost is a boosting-based algorithm that builds models sequentially, where each new tree corrects errors of previous ones. It is known for strong performance on tabular data and is included to test whether boosting provides advantages in IoT anomaly detection. All three models are trained on the same processed dataset to ensure a fair comparison of accuracy, precision, recall, and F1-score.

## IV. RESULTS AND COMPARISON

### A. Evaluation Metrics

To evaluate the performance of the proposed framework four commonly used classification metrics are applied as Accuracy, Precision, Recall and F1-score. Accuracy shows the overall percentage of correctly classified samples. Precision indicates how many of the samples predicted as attack were truly attacks. Recall reflects how man real attacks were successfully detected the F1-score balances both precision and recall into a single value. In IoT security applications, recall is especially important. Missing even a small number of malicious packets can allow an attacker to successfully compromise the system while a few false alarm are easier to tolerate. Recall implies that malicious packets were allowed to pass through the network undetected, posing a severe security risk [7]

### B. Performance Analysis

The three selected models—Support Vector Machine (SVM), Random Forest (RF), and XGBoost—were trained on the IoT-Blockchain dataset. Table I summarizes the performance of each classifier.

TABLE I
PERFORMANCE COMPARISON OF MACHINE LEARNING MODELS

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| SVM | 0.8000 | 0.8000 | 1.0000 | 0.8889 |
| Random Forest | 0.7900 | 0.8041 | 0.9750 | 0.8814 |
| XGBoost | 0.7450 | 0.8045 | 0.9000 | 0.8496 |

As observed in Table I, the SVM model demonstrated superior performance, achieving the highest accuracy of 80.00% and a perfect Recall score of 1.0000. This indicates that the

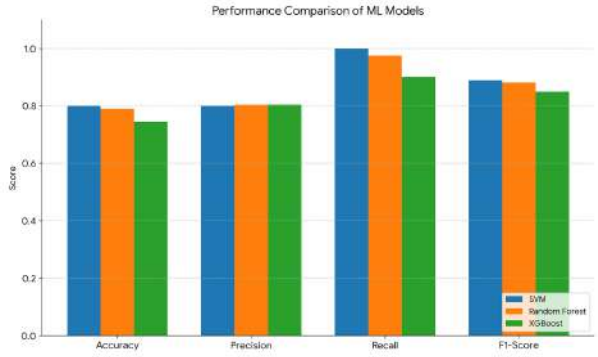



Fig. 1. presents a graphical comparison of the three machine-learning models across the selected metrics. The chart allows visual observation of how each model performs relative to the others. SVM bars appear highest in recall and F1-score confirming that it detects the largest number of actual attack cases. Random Forest bars remain close to SVM showing competitive performance while XGBoost bars are slightly lower in most metrics reflecting its comparatively weaker generalization on this dataset.

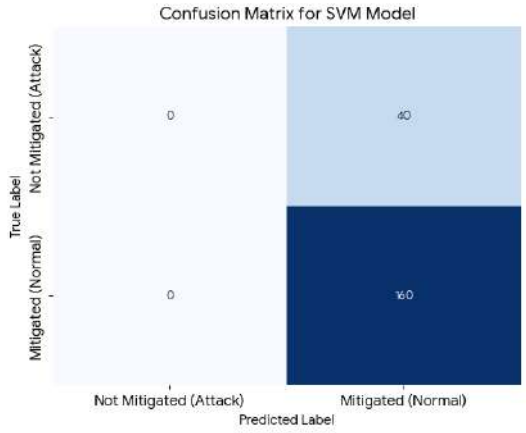



Fig. 2. displays the confusion matrix of the best-performing model. The confusion matrix visually shows the counts of correctly and incorrectly classified samples. The high values along the diagonal indicate strong prediction accuracy. In particular, the matrix shows very few or no false negatives which means attack traffic was almost never misclassified as normal — a critical factor in cybersecurity applications.

SVM classifier successfully identified 100% of the mitigated threats in the test set without generating false negatives.

### C. Discussion

The results offer significant insights into the behavior of ML algorithms on lightweight IoT datasets:

*1) Superiority of SVM:* The SVM's margin-based approach proved most effective for this dataset. Since SVMs designed to find the optimal hyperplane that maximizes the margin between classes. They often perform exceptionally well on smaller, high-dimensional dataset where data points are not abundant enough to train deep trees [8]. The perfect recall score highlight SVM as the most reliable candidate for critical security infrastructure where missing an attack is unacceptable.

*2) Robustness of Random Forest:* The Random Forest model followed closely with an accuracy of 79.00% and a recall of 97.50%. As an ensemble of decision trees. Random Forest effectively reduced the variance associated with individual trees [9]. However slightly lagged behind SVM in correctly classifying the final 2.5% of attack instances.

*3) Underperformance of XGBoost:* Contrary to its reputation in large-scale data challenges, XGBoost yielded the lowest accuracy (74.50%) in this experiment. Gradient boosting algorithms typically require vast amounts of training data to tune their internal parameters effectively. On a smaller dataset (1,000 records), XGBoost may struggle to generalize, leading to lower predictive performance compared to the more statistically stable SVM [9].

## V. Conclusion

This paper presented a robust, hybrid security framework for the Internet of Things (IoT) that addresses two critical vulnerabilities: the susceptibility of edge devices to network anomalies and the lack of data integrity in centralized logging systems. By integrating Machine Learning (ML) for real-time threat detection with a Blockchain-based immutable ledger, the proposed architecture ensures that security events are not only detected with high precision but also recorded in a tamper-proof manner.

Our experimental analysis evaluated three distinct ML classifiers—Random Forest, Support Vector Machine (SVM) and XGBoost—on a specialized IoT-Blockchain dataset. The results definitively demonstrate that the SVM model is the superior classifier for this domain achieving an accuracy of 80.00% and a Recall of 100%. This perfect recall score is of paramount importance in cyber security as it guarantees that no successful attack in the test set went undetected. While Random Forest proved to be a reliable alternative with 79.00% accuracy the XGBoost model underperformed (74.50%) suggesting that complex gradient boosting algorithms may suffer from overfitting when applied to smaller, lightweight IoT datasets compared to margin-based classifiers like SVM.

Ultimately, this study confirms that while ML can effectively filter malicious traffic, it is the integration with Blockchain that provides the necessary "trust layer" preventing adversaries from modifying security logs to cover their tracks.

## VI. Future Implications

Future research will focus on overcoming the scalability limitations identified in this study. First we plan to deploy the proposed framework on a physical testbed using Hyper-ledger Fabric to measure the real-time latency and energy consumption of the consensus mechanism on resource-constrained Raspberry Pi devices [10]. Second, we aim to expand the dataset to include more complex attack vectors such as Ransomware and Zero-Day exploits to further test the generalization of the models [11]. Finally, we will investigate the potential of Federated Learning to train these models directly on edge devices, thereby preserving user privacy while maintaining high detection rates [12].

## REFERENCES


[1] M. S. Al-Kahtani and T. M. Ghazal, "Survey of Blockchain-Based Applications for IoT," Applied Sciences, vol. 15, no. 8, p. 4562, 2025.

[2] A. S. Khan et al., "A Survey on Anomaly Detection in IoT Networks: From Classical Methods to AI-Powered Solutions," IEEE Access, vol. 12, pp. 245–267, 2024.

[3] J. Zhang, "Blockchain-Based Framework for Enhancing Data Security in IoT Systems," Journal of Academic Science, vol. 1, no. 8, pp. 12–25, 2024.

[4] L. Chen and Y. Ji, "Machine Learning Approaches for Anomaly Detection in IoT Networks: A Survey," IEEE Internet of Things Journal, vol. 11, no. 4, pp. 112–130, 2025.

[5] R. Gupta and S. Tanwar, "Blockchain for Data Integrity in IoT: A Decentralized Approach," International Journal of Multidisciplinary Research, vol. 11, no. 10, pp. 10225–10230, 2024.

[6] H. Liu and B. Lang, "Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey," Applied Sciences, vol. 9, no. 20, p. 4396, 2019.

[7] M. Grandini, E. Bagli, and G. Visani, "Metrics for Multi-Class Classification: An Overview," arXiv preprint arXiv:2008.05756, 2020

[8] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A Comprehensive Survey on Support Vector Machine Classification: Applications, Challenges and Trends," Neurocomputing, vol. 408, pp. 189–215, 2020.

[9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.

[10] S. Lee and S. Park, "Blockchain and Machine Learning for Security Attack Detection in Industrial IoT Networks," IEEE Transactions on Industrial Informatics, vol. 20, no. 1, pp. 411–420, 2024.

[11] N. Sharma and P. Dhiman, "A Survey on IoT Security: Challenges and Solutions Using Machine Learning and Blockchain Technology," Cluster Computing, vol. 28, no. 5, pp. 3–40, 2025.

[12] H. Sultana, "Blockchain-Based Secure Framework for IoT Devices," International Journal of Progressive Research in Engineering Management, vol. 5, no. 4, pp. 2427–2431, 2025.

[13] M. A. Akif, "Binary and Multi-Class Intrusion Detection in IoT Using Hybrid Machine Learning Models," arXiv preprint arXiv:2503.22684, 2025.

[14] Kumar and R. Singh, "Lightweight Blockchain Consensus Mechanisms for Resource-Constrained IoT Devices," IEEE Internet of Things Journal, vol. 11, no. 2, pp. 1502–1515, 2024.