Explainable Hybrid Machine Learning Framework for Cyber Threat Detection in Smart Home IoT Environments

Mr. Maulik K. Shah
Research Scholar, Faculty of Computer Science, BKNM University, Junagadh, Gujarat, India.

Abstract:
While smart homes are becoming more popular by the day, they are also becoming massive, highly vulnerable targets for cyberattacks. This is largely because the IoT ecosystem is incredibly messy – devices are wildly diverse, they often lack the processing power needed to run heavy security software, and they rely on inherently weak communication protocols. To make matters worse, traditional intrusion detection systems usually miss brand-new, sophisticated types of attacks. Even when they do catch something, they function like a "black box," offering absolutely no explanation as to why they flagged a threat. We wanted to solve both of these problems at once, so we developed the Explainable Hybrid Machine Learning Framework (EHMLF). Our approach combines the predictive strengths of Random Forest (RF) and Extreme Gradient Boosting (XG Boost) models using a stacking ensemble architecture to maximize detection accuracy. To remove the mystery behind the algorithm's decisions, we integrated SHAP and LIME to provide clear, understandable reasons for every alert. The system breaks down to five straightforward steps: gathering data, cleaning it, engineering features, building the hybrid model, and generating the final explanations. Ultimately, this framework bridges the gap between high-tech AI and human trust, giving security analysts and everyday users the accurate, transparent, and trustworthy security tools they actually need to protect next-generation smart homes.

Keywords: Internet of Things, Cyber Threat Detection, Explainable Artificial Intelligence, Hybrid Machine Learning, Smart Home Security

1. Introduction

The Internet of Things (IoT) has transformed modern living by enabling communication among interconnected smart devices such as surveillance cameras, smart televisions, voice assistants, thermostats, sensors, and smart appliances. According to recent industry reports, billions of IoT devices are expected to be deployed worldwide over the next decade [1].
Despite numerous advantages, IoT ecosystems have become attractive targets for cyber attackers due to their heterogeneous architecture, resource constraints, weak authentication mechanisms, and outdated firmware [2]. Cyberattacks such as Distributed Denial of Service (DDoS), botnets, malware, ransomware, spoofing, and data exfiltration have increased considerably in recent years [3]. Traditional signature-based Intrusion Detection Systems (IDS) are incapable of detecting unknown threats and zero-day attacks [4]. Machine Learning (ML)-based IDS solutions have demonstrated superior performance in identifying malicious activities [5]. However, standalone ML algorithms often suffer from overfitting, reduced generalizability, and limited interpretability [6]. Explainable Artificial Intelligence (XAI) techniques such as SHAP and LIME have gained popularity because they allow users to understand the decision-making process of complex machine learning models [7]. This paper proposes an Explainable Hybrid Machine Learning Framework for cyber threat detection in smart home IoT environments by integrating RF, XGBoost, SHAP, and LIME.

The objectives of this research are:

- To develop a hybrid machine learning framework for IoT cyber threat detection.
- To integrate explainability mechanisms into the detection process.
- To improve trust and transparency in security decisions.
- To support practical deployment in smart home environments.

2. Literature Review

Recent studies have extensively explored ML-based cyber threat detection for IoT environments. Meidan et al. [8] introduced the N-BaIoT dataset and demonstrated the effectiveness of deep autoencoders for botnet detection. However, the model required device-specific training. Khraisat et al. [6] reported that ensemble methods outperform individual machine learning algorithms for intrusion detection applications. Fan et al. [9] highlighted the importance of Explainable AI in cybersecurity and emphasized the need for transparent security systems. Rjoub G et al. [10] presented a comprehensive survey on Explainable Artificial Intelligence for cybersecurity and highlighted the importance of transparent, trustworthy, and interpretable AI models for security applications. Rey et al. [12] introduced federated learning for IoT security but did not incorporate explainability mechanisms. Although substantial progress has been made, existing systems still face challenges regarding transparency, computational efficiency, and real-world deployment. Also, AI-based intrusion detection systems specifically designed for smart home environments have demonstrated promising results. Recent studies have also explored AI-based intrusion detection systems tailored for smart home environments. Shah [11] proposed an AI-based Intrusion Detection System (IDS) using Random Forest and Support Vector Machine (SVM) algorithms to enhance anomaly detection accuracy and reduce false positive rates. The study emphasized the importance of dataset-driven learning and real-time threat detection mechanisms for improving smart home IoT security.

Table 1: Summary of Existing Studies

Author	Technique	Dataset	Limitation
Meidan et al. [8]	Autoencoder	N-BaIoT	Device-specific
Khraisat et al. [6]	Ensemble IDS	Multiple	Explainability absent
Yan et al. [9]	XAI Survey	Cybersecurity	No IoT implementation
Rjoub G et al. [10]	XAI Survey	Cybersecurity	User-centric explanations missing
Rey et al. [12]	Federated Learning	IoT Malware	No explainability

3. Research Gap

Existing IoT intrusion detection systems primarily focus on improving detection accuracy; however, several limitations remain. Most studies rely on standalone machine learning algorithms and fail to provide transparent explanations for detected threats. In addition, user-centric explainability, real-time deployment capability, and integration of hybrid models are insufficiently explored. Therefore, an explainable hybrid machine learning framework is required to improve detection performance, trustworthiness, and practical applicability in smart home IoT environments.

Table 2: Research Gaps and Proposed Solutions

Existing Limitation	Proposed Solution
Lack of explainability	Integrate SHAP and LIME
Standalone ML models	Hybrid RF + XG Boost framework
Low user trust	Generate explainable security alerts
Limited deployment capability	Smart home-oriented architecture
High false positive rates	Ensemble-based learning

4. Proposed Explainable Hybrid Machine Learning Framework

The proposed framework introduces a structured, five-tier pipeline designed to systematically ingest raw network traffic, optimize it for classification, detect anomalies, and interpret the model's underlying logic.

Phase 1: Ingestion of Network Benchmarks

The pipeline starts by pulling malicious and normal interaction signatures from three verified public IoT environments: N-BaIoT, ToN_IoT and UNSW-NB15.

Phase 2: Data Preprocessing

Data cleaning remove anomalies introduced during capture. The sequence executes missing value resolution, data deduplication, min-max scaling/normalization, structural label encoding, and minority class oversampling to stabilize dataset balances.

Phase 3: Network Parameter Selection

The system narrows its scope down to highly predictive features, prioritizing: packet count, total flow duration, network port origins and targets, transport protocol types, individual packet sizes, packet inter-arrival times, and overall connection frequencies.

Phase 4: Hybrid Model Development

The model achieves enhanced precision by mapping inputs to a stacked meta-classifier composed of Random Forest and XG Boost [13]. The RF ensemble evaluates predictions across a forest network via:

$$\left[RF(x) = \frac{1}{T}\sum_{i=1}^{T} h_i\,(x)\right] \qquad (1)$$

Where:

- (T) = number of trees
- $(h_i(x))$ = prediction of individual tree

The XG Boost algorithm adds a gradient boosting approach to minimize residual losses:

$$[Fm(x) = Fm{-}1(x) + \eta hm(x)] \qquad (2)$$

Where:

- (η) = learning rate
- $(hm(x))$ = weak learner

Phase 5: Interpretability Implementation

To explain model decisions, the final layer applies dual-focus local and global metrics. SHAP assesses individual feature contributions against standard baseline outputs using:

$$SHAP\left[f(x) = E[f(X)] + \sum_{i=1}^{n} \phi_i\right] \qquad (3)$$

Where:

- $(E[f(X)])$ = expected model output
- (ϕi) = contribution of feature i

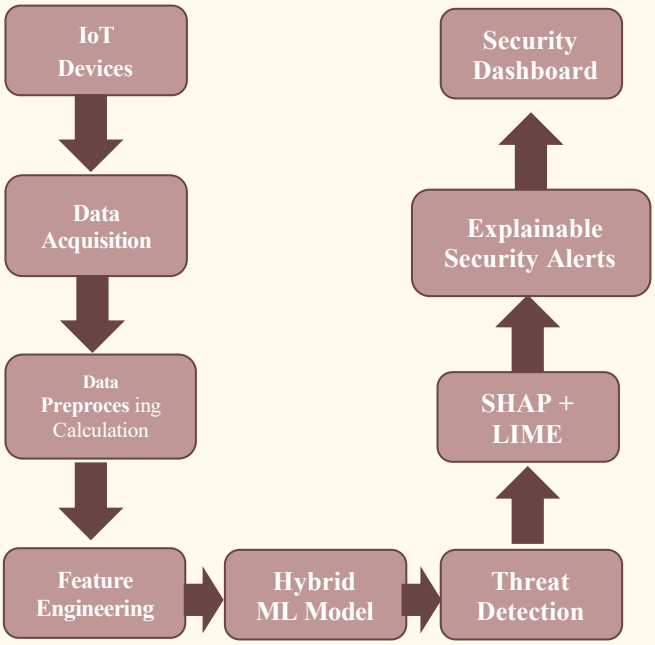
LIME solves localized decision boundaries by matching an explicit surrogate model g to the black-box predictor f while optimizing for local fidelity loss L and a structural complexity penalty Ω.

$$\left[\underset{g\in G}{argmin}; L(f, g, \pi_x) + \Omega(g)\right] \qquad (4)$$

Where:

- (f) = original moÞl
- (g) = surrogate model
- (L) = fidelity loss
- (Ω) = complexity penalty

5. Proposed Framework Architecture



6. Expected Outcomes

The proposed framework is expected to:

- Improve cyber threat detection accuracy.
- Reduce false positive rates.
- Enhance model transparency.
- Increase user trust.
- Provide actionable security recommendations. Support practical deployment in smart homes.

7. Research Challenges

Several challenges still exist:

- Heterogeneous IoT devices
- Resource-constrained environments
- Real-time detection requirements
- Zero-day attack detection
- Explainability overhead
- Privacy concerns

These challenges motivate the development of efficient and explainable hybrid security models.

8. Future Research Directions

Future work may focus on:

- Federated learning integration
- Edge computing deployment
- Real-time adaptive learning
- Large Language Model (LLM)-based explanations
- Adversarial attack resistance
- Continual learning techniques

9. Conclusion

This paper proposed an Explainable Hybrid Machine Learning Framework for cyber threat detection in smart home IoT environments. The framework integrates Random Forest and XGBoost with SHAP and LIME to provide accurate, transparent, and trustworthy cyber threat detection. The proposed approach addresses the limitations of conventional intrusion detection systems and supports intelligent security management for future smart homes. The integration of explainable AI is expected to improve user trust and enable practical adoption of machine learning-based cybersecurity solutions.

References:

1. Chen, T., and C. Guestrin. "XGBoost: A Scalable Tree Boosting System." Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2016, pp. 785–794. https://doi.org/10.1002/int.23088.
2. García-Teodoro, P., et al. "Anomaly-Based Network Intrusion Detection: Techniques, Systems and Challenges." Computer Security, vol. 28, no. 1–2, Feb. 2009, pp. 18–28. https://doi.org/10.1016/j.cose.2008.08.003.
3. Hassija, V., et al. "A Survey on IoT Security: Application Areas, Security Threats, and Solution Architectures." IEEE Access, vol. 7, 2019, pp. 82721–82743. https://doi.org/10.1109/ACCESS.2019.2924045.
4. Khraisat, A., et al. "Survey of Intrusion Detection Systems: Techniques, Datasets and Challenges." Cybersecurity, vol. 2, no. 1, Jul. 2019, p. 20. https://doi.org/10.1186/s42400-019-0038-7.
5. Lundberg, S. M., and S. I. Lee. "A Unified Approach to Interpreting Model Predictions." Advances in Neural Information Processing Systems, May 2017, pp. 4766–4775. arXiv, https://arxiv.org/pdf/1705.07874.
6. Meidan, Y., et al. "N-BaIoT—Network-Based Detection of IoT Botnet Attacks Using Deep Autoencoders." IEEE Pervasive Computing, vol. 17, no. 3, Jul. 2018, pp. 12–22. https://doi.org/10.1109/MPRV.2018.03367731.
7. Ning, H. S., and H. Liu. "Cyber-Physical-Social-Thinking Space Based Science and Technology Framework for the Internet of Things." Science China Information Sciences, vol. 58, no. 3, Jan. 2015, pp. 1–19. https://doi.org/10.1007/s11432-014-5209-2.
8. Rey, V., et al. "Federated Learning for Malware Detection in IoT Devices." Computer Networks, vol. 204, Feb. 2022, p. 108693. https://doi.org/10.1016/j.comnet.2021.108693.
9. Rjoub, G., et al. "A Survey on Explainable Artificial Intelligence for Cybersecurity." IEEE Transactions on Network and Service Management, vol. 20, no. 4, Dec. 2023, pp. 5115–5140. https://doi.org/10.1109/TNSM.2023.3282740.
10. Shah, Maulik K. "AI-Based Intrusion Detection System for Smart Homes." Journal of Emerging Technologies and Innovative Research, vol. 12, no. 7, Jul. 2025, pp. d513–d515.

11. Sicari, S., et al. "Security, Privacy and Trust in Internet of Things: The Road Ahead." Computer Networks, vol. 76, Jan. 2015, pp. 146–164. https://doi.org/10.1016/j.comnet.2014.11.008.
12. Vinayakumar, R., et al. "Deep Learning Approach for Intelligent Intrusion Detection System." IEEE Access, vol. 7, 2019, pp. 41525–41550. https://doi.org/10.1109/ACCESS.2019.2895334.
13. Yan, F., et al. "Explainable Machine Learning in Cybersecurity: A Survey." International Journal of Intelligent Systems, vol. 37, no. 12, Dec. 2022, pp. 12305–12334. https://doi.org/10.1002/int.23088.