



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** V **Month of publication:** May 2026

DOI: <https://doi.org/10.22214/ijraset.2026.82464>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

A Survey on Blockchain Enabled Explainable Artificial Intelligence Framework for Secure Intrusion Detection in Internet of Things Environments

Prof. Dr.Kamalakshi Naganna, Preethi R, Rakshitha B, Shrusti M, Syed Ashika N

Dept. of CSE Sapthagiri College of Engineering

Abstract—TherapidexpansionoftheInternetofThings(IoT)hasled to the interconnection of a large number of devices, which has significantly increased exposure to cybersecurity threats Examples include unauthorized entry, data misuse, malicious software, and traffic flooding attacks breaches, and denial-of-service attacks. Traditional intrusion detection systems (IDS) commonly rely on centralized structures and complex machine learning models, which notonlyraiseconcernsaboutdataprivacybutalsolacktransparency indecision-making.Thesechallengeshighlightthenecessityformore secure, efficient, and interpretable intrusion detection solutions. This paper presents a comprehensive survey of a Blockchain-enabled Explainable Artificial Intelligence (XAI) framework aimed at enhancing intrusion detection in IoT environments. The framework explored in this survey unifies three complementary paradigms: This survey synthesizes three intersecting research directions: privacy- preserving model training via federated learning, tamper-evident logging via distributed ledger technology, and post-hoc decision attribution via explainability methods — and evaluates how their combinationaddressesgapsleftbysingle-paradigmapproaches.This combinationis motivatedbypractical gapswe identify across reviewed literature. The survey analyzes existing methods by considering key factors such as detection performance, scalability, and real-world applicability, while also identifying ongoing challenges including computational complexity and data heterogeneity. The findings indicate that the integration of blockchain, federated learning, and explainableAI can greatly enhance the overall performance security, transparency, and reliability of security monitoring systems for detecting attacks in IoT networks. However, further research is mandatory to develop lightweight and scalable solutions suitable for practical deployment.

Keyword: IoT Security, Intrusion Detection, Explainable AI, Blockchain, Federated Learning, Deep Learning.

I. INTRODUCTION

The Internet of Things (IoT) has rapidly evolved into a key technology that connects everyday devices, sensors, and systems to enable intelligent communication and automation. IoT adoption has grown fastest in sectors where continuous sensing is operationally valuable — factory floors, hospital wards,trafficsystems,andresidentialenergymanagement.The very densityofdeploymentthatmakesIoTvaluablealsowidens theattacksurface:eachendpointisapotentialingresspoint,and compromise of one node can cascade across a mesh of interdependent devices.

These threats include unauthorized access, data manipulation, malwareinjection,anddistributeddenial-of-serviceattacks, which can severely impact system reliability and user privacy. To protect IoT systems from such attacks, Intrusion Detection Systems(IDS)arewidelyused.IDSmonitornetworktrafficand systembehaviortoidentify suspiciousactivities.Traditional IDS techniques, including signature-based and anomaly-based methods,have showneffectivenessindetectingknownattacks. However, these approachesface limitationswhen dealing with evolvingandunknownthreats.Moreover,Whenamodelraises asecurityalertwithoutsurfacingthefeaturesorlogicthat triggeredit,analystsfaceverificationproblemthey cannot determinewhetherthedetectionisreliableorasuspiciouspattern learnedfromtrainingdata.Inhigh-stakesenvironments such as hospitals or power grids, unverifiable alerts are operationally unusable.

Inrecentyears,ExplainableArtificialIntelligence(XAI) has become a potential solution to enhance the transparency and understanding of machine learning model decisions. XAI techniques provide clear insights into how and why a model makes a particular decision, which is especially important in cybersecurity applications.

At the same time, Federated Learning (FL) This approach has attracted significant interest as an effective solution. decentralized learning approach that allows multiple devices to collaboratively train models without sharing raw data, thereby preserving privacy. In addition to these advancements, blockchain technology offers a secure and decentralized mechanism for data storage and communication. By using distributed ledgers, blockchain ensures that data cannot be altered or tampered with, improving trust and security in IoT networks. This survey focuses on recent developments in IoT intrusion detection by analyzing approaches that integrate deep learning, explainable AI, federated learning, and blockchain technology. The study aims to identify the strengths, limitations, research gaps in existing methods and highlights the need for lightweight, scalable, and real-time solutions suitable for practical deployment.

II. BACKGROUND AND MOTIVATION

A. Need for Intelligent Intrusion Detection Systems in IoT

The global IoT device base has expanded rapidly across industrial, clinical, and residential domains, with projections indicating tens of billions of connected endpoints by the end of this decade [cite primary market source], creating an attack surface that scales with deployment density. clinical monitoring, traffic infrastructure, and residential energy systems. Unlike enterprise IT environments—where endpoints are homogeneous and centrally managed—IoT deployments are heterogeneous, resource-constrained, and operated across incompatible communication protocols. This structural diversity is precisely what makes security difficult: there is no universal traffic baseline, and compromise of one device can cascade silently across an entire mesh. Conventional security approaches are frequently insufficient to handle the dynamic and distributed nature of IoT networks. Consequently, Intrusion Detection Systems (IDS) have become crucial for observing network traffic and identifying malicious activities. With the increasing complexity of modern network environments and volume of IoT data, manual analysis becomes inefficient and error-prone. This creates a need for intelligent intrusion detection systems that can effectively identify and respond to cyber threats. automatically analyze network behavior and detect potential threats in real time. Techniques based on machine learning and deep learning are widely used for detecting complex attack patterns have shown great potential in improving intrusion detection performance. These models are capable of identifying intricate patterns within large datasets and accurately classify network traffic as normal or malicious. However, as these systems become more widely used, there is an increasing demand not only for high detection accuracy but also for secure, scalable, and interpretable solutions.

B. Limitations of Conventional Intrusion Detection Approaches

Although intrusion detection systems powered by deep learning provide high accuracy, they have multiple constraints that reduce their practical deployment in IoT environments. The opacity problem in IDS is not merely aesthetic: when a classifier cannot surface the packet-level or flow-level features that triggered a given alert, the analyst's only recourse is either to trust the model unconditionally or to ignore its output. Neither option is operationally acceptable in high-stakes infrastructure, where unverified alerts carry both false-positive costs (wasted response effort) and false-negative risks (undetected breaches). Traditional centralized learning approaches require collecting data from multiple devices into a central server, which increases the risk of data leakage and unauthorized access. This is a critical concern in IoT environments where sensitive information is continuously generated. Without proper protection mechanisms, stored records can be modified or falsified, corrupting forensic analysis and undermining response decisions. Additionally, most deep learning models demand computational resources—memory, processing cycles, and power—that exceed the capacity of typical IoT edge devices [1][2]. The four limitations identified across reviewed systems—lack of transparency, centralized data exposure, insecure storage, and per-device computational overhead—have not been solved simultaneously by any single production-ready deployment [5][7][9]. This gap motivates the integrated framework surveyed in Section III. Scalability is also a concern, as IoT networks continue to expand rapidly. Existing approaches struggle to handle large-scale and distributed environments efficiently. These limitations highlight the indicate a demand for improved attack detection systems that can address issues related to transparency, privacy, security, and scalability simultaneously.

C. Role of Explainable AI, Federated Learning, and Blockchain in IoT Security

To overcome the challenges of existing intrusion detection systems, advanced technologies such as Explainable Artificial Intelligence (XAI), Federated Learning (FL), and Blockchain have gained significant attention.

Rather than restating the three technologies abstractly, this subsection positions them functionally within the IDS pipeline. XAI techniques provide insights into how models make decisions by identifying the most important features influencing predictions. This helps security analysts understand the reasoning behind detected threats and increases trust in the system.

XAI operates at the output layer to validate each detection; FL replaces centralized data aggregation at the training layer; and blockchain acts as a tamper-evident ledger at the storage layer. Instead of sharing raw data, each IoT device trains the model locally and shares only model updates. This ensures that sensitive information remains secure while still allowing collaborative learning across devices.

The key insight motivating this survey is that all three layers must be addressed together — securing one while leaving the others open creates residual vulnerabilities. Combining these technologies can create a powerful framework for identifying attacks within IoT system environments. Critically, addressing only one or two of these layers is insufficient: a privacy-preserving FL system with no explainability still fails the analyst verification problem, while an XAI-capable system with centralized training still exposes raw data. The key design principle driving this survey is that all three layers — training, storage, and inference — must be secured and made transparent together.

III. LITERATURE SURVEY

- 1) *Manlaibaatar Tserenkhuu et al.* [1] propose an SDN- a unified intrusion detection framework that applies, The Tserenkhuu et al. framework [1] is notable for its dual use of explainability: feature importance rankings derived from SHAP serve not merely as post-hoc justifications but as an active filter that prunes non-discriminative inputs prior to training — a design choice that contributes to near-ceiling benchmark accuracy while also improving model interpretability. The trade-off is deployment feasibility: the compound architecture's memory footprint is incompatible with resource-constrained edge hardware. and the authors do not test the framework on physically constrained hardware. The system therefore demonstrates what is achievable under favourable computational conditions rather than in a realistic edge deployment.
- 2) *Ahwar Khan et al.* [2] directly confront the resource problem left unresolved by high-complexity systems such as [1]. Their framework pairs a Bi-LSTM network with two complementary efficiency mechanisms: a genetic algorithm that eliminates low-signal input features before training, and post-training quantization that reduces parameter precision to shrink the deployed model's memory footprint. Explainability is handled through LIME, which generates locally faithful for individual predictions instead of global feature rankings — making each alert's reasoning inspectable by an analyst reviewing a specific traffic record. Evaluated on CICIDS2017 and UNSW-NB15, the system achieves accuracy comparable to heavier architectures while consuming substantially fewer resources. The limitation the authors themselves identify is meaningful: genetic feature selection optimises for the traffic distribution present during training, and there is no guarantee that the selected feature subset remains discriminative as network conditions evolve or novel attack variants emerge. The system's behaviour in large-scale, non-stationary IoT environments remains an open empirical question.
- 3) *Manuel J. Cabral S.* [3] takes an unconventional approach to the scalability problem in federated IoT intrusion detection by replacing full local dataset training with property testing — a statistical inference technique that draws conclusions about a dataset's distributional properties from a carefully chosen sample rather than from exhaustive scan. This architectural choice yields two concrete advantages: per-device computation drops significantly, and communication rounds become cheaper because local updates are derived from smaller data passes. The system integrates these efficiency gains within a federated edge AI framework, keeping raw traffic on-device and achieving the highest scalability rating among the reviewed systems. The cost is detection coverage. Statistical sampling that reduces computational burden inevitably reduces exposure to infrequent traffic patterns, and low-frequency attack classes — precisely the kind of novel threat that signature-based systems already miss — are the most likely casualties of this trade-off. The paper does not characterise detection performance specifically on rare attack types, which limits the conclusions that can be drawn about the system's practical reliability in environments where novel attacks are the primary concern.
- 4) *Nguyen et al* identify communication overhead as the primary scalability bottleneck in federated IoT intrusion detection and design their systems specifically to reduce it. Their approach applies dimensionality reduction to the feature space before local GRU models are trained, decreasing the volume and complexity of model updates that must be transmitted per federated round. GRU is chosen over LSTM for its simpler gating structure, which reduces per-device computation while retaining the ability to model sequential dependencies in network traffic. The system achieves high detection accuracy on IoT benchmark datasets, and its communication efficiency makes it well-suited to bandwidth-constrained IoT networks where frequent large model updates would be impractical. The tension the authors identify — and do not fully resolve — is between compression and coverage: dimensionality reduction that lowers communication cost does so by discarding features deemed less important under the training distribution, but features that appear low-importance in normal traffic may carry significant signal in rare or novel attack scenarios.

No explainability mechanism is included, so analysts have no visibility into which features were retained, which were discarded, or how those decisions affect specific classification outcomes.

- 5) *Fatemaetal* proposed FEDXAIIDS, a federated explainable AI-powered network attack detection system. The model combines privacy-preserving federated learning combined SHAP-based explanation methods. Each IoT device in the system performs local training, sharing only the learned updates instead of raw data. SHAP is used to show which features have the greatest influence on attack classification [7] (2025). This improves trust and helps security experts understand why a particular traffic record is identified as malicious. The limitation of this method is that combining federated learning and explainability may increase system complexity.
- 6) *Taherietal* build one of only two systems in this review that simultaneously target both privacy and interpretability. Their framework implements federated learning in the conventional sense — each participating IoT device develops a local model based on its own traffic data and contributes only parameter updates to a centralized aggregator, so original data remains within the device. SHAP explanations are generated at inference time to attribute each classification decision to specific input features, giving security analysts a basis for independently verifying alerts rather than accepting model outputs on trust. The authors explicitly frame this verification capability as an operational necessity rather than a feature enhancement, which reflects an important shift in how the field is beginning to think about XAI in high-stakes deployments. The system achieves high accuracy on benchmark datasets. Its primary unresolved cost is the coordination overhead introduced by running SHAP alongside federated aggregation: synchronising explanation generation with model update rounds adds latency that the authors acknowledge but do not measure in concrete terms, leaving the system's viability for time-sensitive detection scenarios undemonstrated.
- 7) *Adel Alabbadietal* developed an intrusion detection system for IoT data streams using explainable artificial intelligence. Their work uses deep learning architectures including CNN, DNN, and TabNet to detect attacks from continuous IoT traffic. SHAP and LIME are applied to explain the model decisions [4] (2025). This helps users understand the reason behind each classification result. The system is useful for real-time intrusion detection, especially when continuous monitoring is required. However, deep learning models may need high computational power for practical deployment.
- 8) *Olanrewaju-George et al* address a detection coverage problem that purely supervised federated systems leave open: labelled training data captures known attack patterns, but IoT environments routinely encounter traffic that does not match any previously documented signature. Their system combines supervised learning — which classifies traffic against labelled attack categories — with unsupervised learning, which flags statistical anomalies in unlabelled traffic without requiring prior knowledge of the attack type. Both methods operate within a federated framework evaluated on the N-BaIoT botnet dataset, keeping raw device data local while enabling collaborative model improvement across distributed nodes. This dual-method design makes the system more resilient to novel threats than purely supervised alternatives. The operational challenges are convergence speed and device heterogeneity: federated systems trained across devices with substantially different hardware capabilities and traffic volumes converge more slowly than centralised equivalents [9], and the paper does not quantify how much this affects detection latency in practice. Interpretability is absent entirely — the system provides no mechanism for analysts to understand why a specific traffic record was classified as malicious.

IV. COMPARATIVE ANALYSIS

A. Trade-off Between Accuracy and Efficiency

The accuracy-efficiency trade-off is most clearly illustrated by the contrast between [1] and [2]. Tserenkhuetal. [1] achieve ~99% detection accuracy using CNN+LSTM but explicitly note that their model is unsuitable for memory-constrained devices. Khan et al. [2] address this directly through model quantization and genetic feature selection, reducing computational footprint — but acknowledge that performance may degrade in large dynamic networks. No reviewed system resolves this tension fully: the highest-accuracy systems [1][4] are resource-intensive, while the most efficient systems [3][8] sacrifice either accuracy or interpretability.

B. Gap Between Theoretical Performance and Real-World Deployment

A consistent pattern across all eight systems is strong benchmark performance paired with untested real-world applicability. All evaluations use clean, labelled datasets (NSL-KDD, UNSW-NB15, CICIoT2023, N-BaIoT) that do not reflect the noise, label imbalance, or adversarial traffic found in live deployments. None of the reviewed papers reports performance under adversarial inputs, device failure conditions, or heterogeneous network configurations. This gap is not incidental — it reflects a field-wide absence of standardized real-world evaluation protocols for IoT intrusion detection [18][19].

C. Lack of Unified and Integrated Solutions

Another important issue refers to the absence of a unified framework capable of address multiple security requirements simultaneously. Existing approaches typically concentrate on one objective, such as improving detection accuracy, ensuring data privacy, or enhancing interpretability. While these methods provide improvements in specific areas, they fail to offer a comprehensive solution that balances all critical factors. This lack of integration limits the overall effectiveness of intrusion detection systems in complex IoT environments.

D. Challenges in Scalability and Adaptability

Scalability constraints manifest differently across the reviewed systems. The federated approaches [3, 6, 8] show higher scalability ratings in Table 2, but this comes at the cost of convergence speed [6] or detection coverage for rare attacks [3]. None of the eight systems was evaluated on a network topology larger than its training dataset, leaving open the question of how detection performance degrades as the number of participating devices grows.

E. Increasing System Complexity and Implementation Challenges

The two systems that combine FL and XAI [5, 7] carry the highest design complexity among those reviewed. Fatema et al. [7] acknowledge that synchronising federated model updates with SHAP computation introduces both latency and coordination overhead. This trade-off — transparency at the cost of simplicity — is not quantified in either paper, leaving practitioners without a concrete deployment cost estimate. Therefore, there is a need to develop solutions.

F. Dependence on Dataset Quality and Diversity

The efficiency of intrusion detection mechanisms is highly influenced by the quality and diversity of the datasets used for training and evaluation. All eight reviewed systems evaluate on labelled benchmark datasets (NSL-KDD, UNSW-NB15, CICIoT2023, N-BaIoT, TON-IoT—see Table 1). None reports performance on unlabelled live traffic or under class-imbalanced conditions. Given that real IoT deployments generate predominantly normal traffic with rare attack events, the accuracy figures in Table 1 may substantially overstate field performance [18][19]. As a result, models trained on such datasets may perform well in testing but struggle when exposed to noisy, incomplete, or highly dynamic real-time data. This highlights the need for more realistic and diverse datasets for model development.

G. Communication Overhead in Distributed Systems

Approaches that rely on distributed learning, particularly federated learning, introduce additional communication overhead. Frequent exchange of model updates between devices and central servers can lead to increased network traffic, latency, and synchronization issues. In large-scale IoT environments, this overhead may affect system performance and efficiency, especially in networks with limited bandwidth.

H. Energy Consumption and Resource Constraints

Many IoT devices operate with limited power and computational resources. High-performance intrusion detection models, especially those based on deep learning, may consume significant energy and processing capacity. This can reduce device lifespan and limit continuous monitoring capabilities. Energy efficiency is therefore an important factor that is often not fully addressed in existing approaches.

I. Robustness Against Evolving and Adversarial Attacks

Cyber threats are continuously evolving, and attackers may use advanced techniques to bypass detection systems. Some intrusion detection models are not robust enough to handle adversarial inputs or rapidly changing attack patterns. This reduces their reliability in real-world scenarios and highlights the need for adaptive and resilient detection mechanisms.

J. Challenges in Real-Time Response and Alert Accuracy

An important limitation observed in many intrusion detection systems is their inability to provide timely and accurate responses. While several models can recognize anomalies, there may be delays in generating alerts or difficulties in distinguishing between legitimate and harmful activities.

Alert latency is the least-studied gap across the reviewed papers. None of the eight systems report end-to-end response times from detection event to analyst notification. Given that some IoT attacks (e.g., DDoS amplification) escalate within seconds, response latency is not an implementation detail but a primary design constraint one that future work should explicitly measure and report. Handling communication delays in distributed environments can impact the timely delivery of alerts, reducing the system's efficiency and responsiveness. Another challenge is the lack of efficient mechanisms to prioritize alerts based on severity. In many cases, a large number of alerts are generated, making it difficult for security analysts to identify critical threats quickly. This can lead to delayed responses and potential system vulnerabilities. Therefore, there is a need for intelligent alert management strategies that can improve both the speed and reliability of intrusion detection systems in IoT environments.

K. Handling of Imbalanced Data

Many intrusion detection datasets suffer from class imbalance, where normal traffic data is significantly higher than attack data. This disparity may bias models toward predicting normal behavior, reducing their capability to identify rare yet critical attacks. Existing approaches do not always mitigate this problem effectively, which limits their reliability in practical scenarios.

L. Dependency on Continuous Model Training

Intrusion detection systems require frequent updates to remain effective against new attack patterns. However, continuous retraining of models can be resource-intensive and time-consuming, especially in distributed IoT environments. This creates challenges in maintaining up-to-date systems without affecting performance.

M. Limited Explainability for Non-Expert Users

Although explainable AI techniques improve transparency, the explanations provided are often complex and difficult for non-technical users to understand. In real-world applications, security analysts or system administrators may require simpler and more intuitive explanations for effective decision-making.

N. Integration Challenges with Existing Systems

IoT environments are characterized by continuously changing network traffic patterns, influenced by factors such as user behavior, device interactions, application requirements, and environmental conditions. These variations can result in significant differences in data distribution over time. Many existing intrusion detection models are trained on static datasets and may not adapt effectively to such dynamic changes. As a result, their detection capability can become inconsistent when deployed in real-world scenarios.

Table 1: Techniques and Performance Comparison of Existing Approach

SL. No	Technique	Dataset	Accuracy	Privacy	Explainability
1	Deep Learning (CNN, DNN, LSTM)+XAI (SHAP, RF) [1]	NSL-KDD, X-IIoTID	High (~99%)	No	Yes
2	Lightweight Deep Learning+Bi-LSTM+LIME [2]	CICIDS2017, UNSW-NB15	High (~99%)	No	Yes
3	Property Testing+Federated Learning+Edge AI [3]	IoT datasets	Moderate (~92%)	Yes	No
4	Deep Learning (CNN, DNN, TabNet)+XAI [4]	TON-IoT	High	No	Yes
5	Federated Learning+Deep Learning+SHAP [5]	IoT datasets	High	Yes	Yes
6	Federated Learning+Deep Learning (AE, CNN, LSTM) [6]	N-BaIoT	High	Yes	No

7	Federated Learning+XAI(SHAP)+Neural Networks [7]	CICIoT2023	Moderate	Yes	Yes
8	Federated Learning+GRU+Feature Reduction [8]	IoT datasets	High	Yes	No

Table 2: Advantages, Limitations, and Scalability Analysis

SL. No	Advantages	Limitations	Scalability
1	High detection accuracy and effective feature selection	High computational cost	Moderate
2	Lightweight and efficient for IoT devices	Limited scalability in large environments	Moderate
3	Fast processing and reduced computation	May fail to detect rare attacks	High
4	Improved interpretability and strong detection performance	Resource-intensive	Moderate
5	Ensures privacy and enhances transparency	Communication overhead and delay	Moderate
6	Supports decentralized learning and protects data	Slower convergence	High
7	Increases trust through explainability	High system complexity	Moderate
8	Reduces communication and computation cost	Possible loss of important features	High

V. CONCLUSION

This survey examined eight recently published intrusion detection systems for IoT environments, each integrating one or more of three complementary technologies: deep learning, federated learning, and explainable AI. The review was motivated by a practical observation: despite strong individual progress in each of these areas, no production-ready system has simultaneously resolved the four core deficiencies that make existing IDS deployments inadequate for real IoT infrastructure: model opacity, centralised data exposure, tamper-vulnerable audit records, and per-device computational overhead.

The comparative analysis in Section IV reveals a consistent pattern across all eight systems. Detection accuracy on standard benchmarks is high, ranging from approximately 92% to 99%, and the algorithmic foundations—SHAP-based explainability, federated aggregation protocols, and deep learning classifiers—are sufficiently mature. The bottleneck is not algorithmic. It is systems-level. The evaluation gaps identified in Section IV—clean benchmarks, no adversarial stress, no hardware profiling—is not a property of any individual system but a structural feature of the entire field as of 2025. Closing it will require new evaluation protocols, not new algorithms, or heterogeneous network configurations representative of live deployments. None measures end-to-end alert latency—the time from detection event to analyst notification—despite the fact that fast-escalating attacks such as DDoS amplification can cause irreversible damage within seconds of onset. None quantifies per-device energy consumption, which is a hard deployment constraint on battery-powered IoT endpoints.

Among the reviewed systems, only Fatema et al. [7] and Taheri et al. [5] simultaneously address privacy, detection accuracy, and interpretability within a single framework. This makes them the closest existing approximations to the integrated architecture this survey advocates. However, both systems validate exclusively on benchmark data and neither report hardware-level performance metrics. Their complexity arising from the coordination overhead of combining federated aggregation with SHAP computation—remains unquantified in deployment terms, leaving practitioners without the information needed to assess feasibility on constrained devices.

The integration of blockchain as a tamper-evident audit layer, surveyed in related work [12][13][16][30][32], addresses the storage-layer vulnerability that federated and XAI-capable systems leave open.

Current evidences suggest that lightweight blockchain schemes can be implemented on resource-constrained IoT nodes [16], but no reviewed paper combines all four layers — federated training, blockchain audit, XAI inference, and edge deployment — within a single validated system. Closing this gap is the most consequential direction for future work in this domain. Three specific research directions follow directly from this survey's findings. First, the community needs standardised evaluation protocols that measure detection performance, alert latency, explainability, fidelity, and energy consumption jointly on physical IoT hardware under live or adversarially-perturbed traffic — not in isolation on clean benchmarks. Second, lightweight model architectures that preserve the accuracy of systems like [1] and [4] while operating within the memory and power budgets of constrained edge devices remain an open design problem; the quantization approach in [2] is a partial step, but generalisation to large dynamic networks has not been demonstrated. Third, federated aggregation schemes must be hardened against adversarial participants — poisoning attacks on the shared model are a known vulnerability [9] that none of the eight reviewed systems explicitly defends against.

In summary, this survey identifies the primary barrier to practical IoT intrusion detection not as a shortage of accurate models, but as an absence of systems-level evaluation frameworks that treat latency, energy, explainability, and adversarial robustness as equal design constraints alongside detection accuracy. Addressing this gap will require closer collaboration between the machine learning, systems security, and IoT hardware communities — and it represents the most impactful open problem in this field as of the time of writing.

REFERENCES

- [1] M. Tserenkhuu, M. D. Hossain, Y. Taenaka, and Y. Kadobayashi, "Deep Learning-Based Intrusion Detection with Explainable Feature Selection for SDN-Enabled IoT Environments," *IEEE Access*, vol. 13, pp. 136864–136880, 2025.
- [2] A. Khan, M. A. Hussain, and F. Anwer, "Lightweight Hybrid Deep Learning for IoT Intrusion Detection Combining Bi-LSTM, Genetic Feature Selection, and Explainable AI," *IEEE Access*, vol. 13, pp. 192451–192470, 2025.
- [3] M. J. C. S. Reis, "Privacy-Preserving and Scalable IoT Intrusion Detection via Property Testing and Federated Edge Intelligence," *IEEE Access*, vol. 13, pp. 153244–153260, 2025.
- [4] A. Alabbadi and F. Bajaber, "XAI-Driven Intrusion Detection over Continuous IoT Data Streams Using CNN, DNN, and TabNet," *Sensors*, vol. 25, no. 3, p. 847, 2025.
- [5] S. Taheri et al., "Towards Transparent IoT Security: A Federated Learning Framework with Built-In Explainability for Intrusion Detection," *IEEE Internet of Things Journal*, 2025.
- [6] T. Olanrewaju-George and B. Pranggono, "Decentralized Intrusion Detection for IoT Using Federated Learning with Supervised and Unsupervised Techniques," *Journal of Information Security and Applications*, vol. 75, 2024.
- [7] N. Fatema et al., "FEDXAIIDS: A Privacy-Preserving Intrusion Detection Architecture Unifying Federated Learning and SHAP-Based Explainability," *IEEE Access*, 2025.
- [8] Q. Zhang, C. Yue, X. Dong, G. Du, and D. Wang, "Blockchain-Powered LSTM-Attention Hybrid Model for Device Situation Awareness and On-Chain Anomaly Detection in IIoT," *Sensors*, vol. 25, no. 15, p. 4663, Jul. 2025.
- [9] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and D. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in *Proc. 20th Int. Conf. Artif. Intell. Stat. (AISTATS)*, Fort Lauderdale, FL, USA, Apr. 2017, pp. 1273–1282.
- [11] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, pp. 4765–4774.
- [12] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You?: Explaining the Predictions of Any Classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, San Francisco, CA, USA, Aug. 2016, pp. 1135–1144.
- [13] Z. A. El Houda, H. Moudoud, B. Brik, and L. Khoukhi, "Securing Federated Learning Through Blockchain and Explainable AI for Robust Intrusion Detection in IoT Networks," in *Proc. IEEE INFOCOM Workshops*, Hoboken, NJ, USA, May 2023, pp. 1–6.
- [14] K. Begum, M. A. I. Mozumder, M.-I. Joo, and H.-C. Kim, "BFLIDS: Blockchain-Driven Federated Learning for Intrusion Detection in IoT Networks," *Sensors*, vol. 24, no. 14, p. 4591, Jul. 2024.
- [15] R. W. Anwar, M. Abrar, A. Salam, and F. Ullah, "Federated Learning with LSTM for Intrusion Detection in IoT-Based Wireless Sensor Networks: A Multi-Dataset Analysis," *PeerJ Computer Science*, vol. 11, p. e2751, Mar. 2025.
- [16] R. Baidar, S. Maric, and R. Abbas, "Hybrid Deep Learning–Federated Learning Powered Intrusion Detection System for IoT/5G Advanced Edge Computing Networks," *arXiv preprint arXiv:2509.15555*, 2025.
- [17] Y. Mirsky, T. Golomb, and Y. Elovici, "Lightweight Collaborative Anomaly Detection for the IoT Using Blockchain," *Journal of Parallel and Distributed Computing*, Elsevier, Jun. 2020.
- [18] A. A. Diro and N. Chilamkurti, "Distributed Attack Detection Scheme Using Deep Learning Approach for Internet of Things," *Future Generation Computer Systems*, vol. 82, pp. 761–768, 2018.
- [19] N. Moustafa and J. Slay, "UNSW-NB15: A Comprehensive Data Set for Network Intrusion Detection Systems," in *Proc. Military Communications and Information Systems Conf. (MilCIS)*, Canberra, Australia, Nov. 2015, pp. 1–6.
- [20] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," in *Proc. 4th Int. Conf. Information Systems Security and Privacy (ICISSP)*, Funchal, Madeira, Jan. 2018, pp. 108–116.



- [21] M.A.Ferrag,L.Maglaras,A.Ahmim,M.Derdour,andH. Janicke, "RDTIDS: Rules and Decision Tree-Based Intrusion Detection System for Internet-of-Things Networks," *Future Internet*, vol. 12, no. 3, p. 44, 2020.
- [22] Y. Meidan et al., "N-BaIoT: Network-Based Detection of IoTBotnetAttacks Using DeepAutoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 12–22, Jul.–Sep. 2018.
- [23] P.Porambage,J.Okwuibe,M.Liyanage,M.Ylianttila,and T.Taleb,"SurveyonMulti-AccessEdgeComputingforInternet of Things Realization," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 2961–2991, 2018.
- [24] A. B. Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward ResponsibleAI," *Information Fusion*, vol. 58, pp. 82– 115, Jun. 2020.
- [25] S.Wang,T.Tuor,T.Salonidis,K.K.Leung,C.Makaya,T. He, and K. Chan, "Adaptive Federated Learning in Resource ConstrainedEdgeComputing,"*IEEEJournalonSelectedAreas in Communications*, vol. 37, no. 6, pp. 1205–1221, Jun. 2019.
- [26] M. Zhao, F. Ge, T. Zhang, and Z. Yuan, "AntiFraud: A System for Online Credit Card Transaction Fraud Detection Using Machine Learning," in *Proc. IEEE Int. Conf. Big Data*, Washington, DC, USA, Dec. 2019, pp. 5–7.
- [27] A.Yazdinejad,R.M. Parizi,A.Dehghantanha, Q.Zhang, and K.-K. R. Choo, "An Energy-Efficient SDN Controller Architecture for IoT Networks With Blockchain-Based Security," *IEEE Transactions on Services Computing*, vol. 13, no. 4, pp. 625–638, Jul.–Aug. 2020.
- [28] M.A.Ferrag and L. Maglaras, "DeepCoin:ANovel Deep LearningandBlockchain-BasedEnergyExchangeFramework for Smart Grids," *IEEE Transactions on Engineering Management*, vol. 67, no. 4, pp. 1285–1297, Nov. 2020.
- [29] X. Zheng, Z. Cai, andY. Li, "Data Linkage in Smart IoT Systems:A Consideration From a Privacy Perspective," *IEEE CommunicationsMagazine*,vol.56,no.9,pp.55–61,Sep. 2018.
- [30] F.A.Alaba,M.Othman,I.A.T.Hashem,andF.Alotaibi, "InternetofThingsSecurity:ASurvey,"*JournalofNetworkand Computer Applications*, vol. 88, pp. 10–28, Jun. 2017.
- [31] O.A. H. Gwasssi, O. N. Uçan, and E.A. Navarro, "Cyber- XAI-Block: An End-to-End Cyber Threat Detection and FL- Based Risk Assessment Framework for IoT-Enabled Smart Organizations Using XAI and Blockchain Technologies," *Scientific Reports*, 2025.
- [32] S. Sun, L. Zhou, Z.Wang, and L. Han, "Robust Intrusion Detection Based on Personalized Federated Learning for IoT Environments," *Computers & Security*, 2025.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)