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

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.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 systembehaviortoidentifysuspiciousactivities.Traditional IDS techniques, including signature-based and anomaly-based methods,haveshowneffectivenessindetectingknownattacks. However, these approachesface limitationswhen dealing with evolvingandunknownthreats.Moreover,Whenamodelraises asecurityalertwithoutsurfacingthefeaturesorlogicthat triggeredit,analystsfaceaverificationproblemtheycannot determinewhetherthedetectionisreliableoraspuriouspattern learnedfromtrainingdata.Inhigh-stakesenvironmentssuchas 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)Thisapproachhasattractedsignificantinterestas an effective solution. decentralized learning approach that allowsmultipledevicestocollaborativelytrainmodelswithout sharing raw data, thereby preserving privacy .In addition to theseadvancements,blockchaintechnologyoffersasecureand decentralized mechanism for data storage and communication. By using distributed ledgers, blockchain ensures that data cannotbealteredortamperedwith,improvingtrustandsecurity in IoT networks.
Thissurveyfocusesonrecentdevelopmentsin IoTintrusion detectionbyanalyzingapproachesthatintegratedeeplearning, explainableAI, federatedlearning, andblockchaintechnology. 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. NeedforIntelligentIntrusionDetectionSystemsin IoT

The global IoT device base has expanded rapidly across industrial, clinical, and residential domains, with projections indicatingtensofbillionsofconnectedendpointsbytheendof this decade [cite primary market source], creating an attack surface that scales with deployment density .clinical monitoring, traffic infrastructure, and residential energy systems.UnlikeenterpriseITenvironments—whereendpoints 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 anddistributednatureofIoTnetworks.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. LimitationsofConventionalIntrusionDetectionApproaches

Although intrusion detection systems powered by deep learning provide high accuracy, they have multiple constraints thatreducetheirpracticaldeploymentinIoTenvironments.The opacity problem in IDS is not merely aesthetic: when a classifiercannotsurfacethepacket-levelorflow-levelfeatures thattriggered a given alert,the analyst'sonly recourse iseither to trust the model unconditionally or to ignore its output. Neither option is operationally acceptable in high-stakes infrastructure,whereunverifiedalertscarrybothfalse-positive costs (wasted response effort) and false-negative risks (undetected breaches).Traditional centralized learning approachesrequirecollectingdatafrommultipledevicesintoa 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, mostdeeplearningmodelsdemandcomputationalresources— memory, processing cycles, and power — that exceed the capacityoftypicalIoTedgedevices[1][2].Thefourlimitations 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. RoleofExplainableAI,FederatedLearning,andBlockchain in IoT Security

To overcome the challenges of existing intrusion detection systems, advanced technologies such as ExplainableArtificial Intelligence (XAI), Federated Learning (FL), and Blockchain have gained significant attention.
Rather than restating the three technologies abstractly, this subsectionpositionsthemfunctionallywithintheIDSpipeline XAI techniques provide insights into how models make decisions by identifying the most important features influencingpredictions.Thishelpssecurityanalystsunderstand the reasoningbehind detectedthreats 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 modellocallyandsharesonlymodelupdates.Thisensuresthat sensitive information remains secure while still allowing collaborative learning across devices.

Thekeyinsightmotivatingthissurveyisthatallthreelayers must be addressed together — securing one while leaving the others open creates residual vulnerabilities. Combining these technologiescan createsapowerful framework foridentifying attackswithinIoTsystemsenvironmentsCritically,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 withcentralizedtrainingstillexposesrawdata.Thekeydesign principledrivingthissurvey isthat all three layers — training, storage,andinference—mustbesecuredandmadetransparent together.

III. LITERATURE SURVEY

1) *ManlaibaatarTserenkhuu et al*. [1] propose an SDN- a unified intrusion detection framework that applies, TheTserenkhuu et al. framework [1] is notable for itsdualuse 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 compoundarchitecture'smemoryfootprintisincompatiblewith resource-constrainededgehardware.andtheauthorsdonottest theframeworkonphysicallyconstrainedhardware.Thesystem therefore demonstrates what is achievable under favourable computational conditions rather than in a realistic edge deployment.
2) *AhwarKhanetal.*[2]directlyconfrontthesourceproblem left unresolved by high-complexity systems such as [1]. Their frameworkpairsaBi-LSTMnetworkwithtwocomplementary efficiencymechanisms:ageneticalgorithmthateliminateslow- 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'sreasoning inspectableby an analystreviewinga 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 conditionsevolveornovelattackvariantsemerge.Thesystem'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 fulllocaldataset training with property testing — astatisticalinferencetechniquethatdrawsconclusionsabouta dataset's distributional properties from a carefully chosen sample rather than from exhaustive scan. This architectural choiceyieldstwoconcreteadvantages:per-devicecomputation drops significantly, and communication rounds become cheaper becauselocalupdatesarederivedfromsmallerdatapasses.The systemintegratestheseefficiencygainswithinafederatededge AIframework,keepingrawtrafficon-deviceandachievingthe highestscalabilityratingamongthereviewedsystems.Thecost is detection coverage. Statistical sampling that reduces computationalburdeninevitablyreducesexposuretoinfrequent traffic patterns, and low-frequency attack classes — precisely thekindofnovelthreatsthatsignature-basedsystemsalready miss—arethemostlikelycasualtiesofthistrade-off.Thepaper doesnotcharacterisedetectionperformancespecificallyonrare 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 detectionanddesigntheirsystemspecificallytoreduceit.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 frequentlarge model updateswould be impractical.The tension the authors identify — and do not fully resolve — is between compression and coverage: dimensionality reduction that lowerscommunication costdoesso by discarding features deemed less important under the training distribution, but featuresthatappearlow-importanceinnormaltrafficmaycarry 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*proposedFEDXAIIDS,afederatedexplainable AI-powered network attack detection system. The model combines privacy-preserving federated learning combined SHAP-basedexplanationmethods.EachIoTdevicethesystem performslocaltraining,sharingonlythelearnedupdatesshared insteadofrawdata.SHAPisusedtoshowwhichfeatureshave the greatest influence on attack classification *[7] (2025)*. This improves trust and helps security experts understand why a particular traffic record isidentified as malicious.The limitation of this method is that combining federated learning and explainability may increase system complexity.

6) *Taherietal*buildoneofonlytwosystemsinthisreviewthat 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,whichreflectsanimportantshiftinhowthefield 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 SHAPalongside 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) *AdelAlabbadietal* developedanintrusiondetectionsystem for IoT data streams using explainable artificial intelligence. Their work uses deep learning architectures including CNN, DNN,andTabNettodetectattacksfromcontinuousIoTtraffic. SHAPandLIMEareappliedtoexplainthemodeldecisions*[4] (2025)*. This helps users understand the reason behind each classificationresult.Thesystemisusefulforreal-timeintrusion 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-BaIoTbotnet dataset, keeping raw device data local while enabling collaborative modelimprovementacrossdistributednodes.Thisdual-method design makes the system more resilient to novel threats than purely supervised alternatives. The operational challenges are convergencespeedanddeviceheterogeneity:federatedsystems trained across devices with substantially different hardware capabilities and traffic volumes converge more slowly than centralisedequivalents[9],andthepaperdoesnotquantifyhow 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-offBetweenAccuracyandEfficiency*

The accuracy-efficiency trade-off ismost clearly illustratedby thecontrastbetween[1]and[2]. Tserenkhuuetal.[1]achieve~99%detectionaccuracyusingCNN+LSTMbutexplicitlynote that their model isunsuitablefor memory-constrained devices. Khan et al. [2] address this directly through model quantizationand genetic feature selection, reducing computational footprint — but acknowledge that performance maydegradeinlargedynamicnetworks.Noreviewedsystem 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.Allevaluationsuseclean,labelleddatasets(NSL- KDD,UNSW-NB15,CICIoT2023,N-BaIoT)thatdonotreflect the noise, label imbalance, or adversarial traffic found in live deployments.Noneofthereviewedpapersreportsperformance 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. LackofUnifiedandIntegratedSolutions

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. ChallengesinScalabilityandAdaptability

Scalabilityconstraintsmanifestdifferentlyacrossthereviewed systems. The federated approaches [3, 6, 8] show higher scalability ratings in Table 2, but this comes at the cost of convergencespeed[6]ordetectioncoverageforrareattacks[3]. Noneoftheeightsystemswasevaluatedonanetworktopology largerthanitstrainingdataset,leavingopenthequestionofhow detection performancedegrades as the number ofparticipating devices grows

E. Increasing System Complexity and Implementation Challenges

The two systems that cmbine FL and XAI [5, 7] carry the highestdesigncomplexityamongthosereviewed.Fatemaetal. [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. DependenceonDatasetQualityandDiversity

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—seeTable1).Nonereports performance on unlabelled live traffic or under class- imbalanced conditions. Given that real IoT deployments generate predominantly normal traffic with rare attack events, theaccuracyfiguresinTable1maysubstantiallyoverstatefield performance [18][19].As a result, models trained on such datasetsmayperformwellintestingbutstrugglewhenexposed to noisy, incomplete, or highly dynamic real-time data. This highlights the need for more realistic and diverse datasets for model development.

G. CommunicationOverheadinDistributedSystems

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 ConsumptionandResourceConstraints

Many IoT devices operate with limited power and computationalresources.High-performanceintrusiondetection models,especiallythosebasedondeeplearning,mayconsume 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. RobustnessAgainstEvolvingandAdversarialAttacks

Cyber threats are continuously evolving, and attackersmay 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. ChallengesinReal-TimeResponseandAlertAccuracy

An important limitation observed in many intrusion detection systemsistheirabilitytoprovidetimelyandaccurateresponses. 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-studiedgap acrossthereviewedpapers.Noneoftheeight systems report end-to-end response times from detection event toanalystnotification.GiventhatsomeIoTattacks(e.g.,DDoS amplification) escalate within seconds, response latency is not an implementation detail but a primary design constraint one thatfutureworkshouldexplicitlymeasureandreport.Handling communication delays in distributed environments can impact thetimelydeliveryofalerts,reducingthesystem'sefficiency responsiveness. Another challenge is the lack of efficient mechanismstoprioritizealertsbasedonseverity.Inmanycases, 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. HandlingofImbalanced Data

Many intrusion detection datasets suffer from class imbalance, wherenormaltrafficdataissignificantlyhigherthanattackdata. 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. DependencyonContinuousModelTraining

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. LimitedExplainabilityforNon-ExpertUsers

Although explainableAI 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. IntegrationChallengeswithExistingSystems

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 existingintrusiondetectionmodelsaretrainedonstaticdatasets and may not adapt effectively to such dynamic changes.As a result, their detection capability can become inconsistentwhen deployed in real-world scenarios.

Table1: TechniquesandPerformanceComparisonofExistingApproach

SL. No	Technique	Dataset	Accuracy	Privacy	Explainability
1	DeepLearning (CNN,DNN, LSTM)+XAI(SHAP, RF) [1]	NSL-KDD,X-IIoTID	High (~99%)	No	Yes
2	LightweightDeepLearning+Bi-LSTM+LIME[2]	CICIDS2017, UNSW-NB15	High (~99%)	No	Yes
3	PropertyTesting+FederatedLearning+EdgeAI[3]	IoTdatasets	Moderate(~92%)	Yes	No
4	DeepLearning(CNN,DNN,TabNet)+XAI[4]	TON-IoT	High	No	Yes
5	FederatedLearning+DeepLearning+SHAP[5]	IoTdatasets	High	Yes	Yes
6	FederatedLearning+DeepLearning(AE,CNN,LSTM)[6]	N-BaIoT	High	Yes	No

7	FederatedLearning+XAI(SHAP)+NeuralNetworks [7]	CICIoT2023	Moderate	Yes	Yes
8	FederatedLearning+GRU+FeatureReduction[8]	IoTdatasets	High	Yes	No

Table2:Advantages,Limitations,andScalabilityAnalysis

SL. No	Advantages	Limitations	Scalability
1	Highdetectionaccuracyandeffectivefeature selection	Highcomputationalcost	Moderate
2	LightweightandefficientforIoTdevices	Limited scalability in large environments	Moderate
3	Fastprocessingandreducedcomputation	Mayfailtodetectrareattacks	High
4	Improvedinterpretabilityandstrongdetection performance	Resource-intensive	Moderate
5	Ensuresprivacyandenhancestransparency	Communicationoverheadand delay	Moderate
6	Supportsdecentralizedlearningandprotectsdata	Slowerconvergence	High
7	Increasestrustthroughexplainability	Highsystemcomplexity	Moderate
8	Reducescommunicationandcomputationcost	Possible loss of important features	High

V. CONCLUSION

This survey examined eight recently published intrusion detectionsystemsforIoTenvironments,eachintegratingoneor more of three complementary technologies: deep learning, federated learning, and explainable AI. The review was motivated by a practical observation: despite strong individual progressineachofthesеareas,noproduction-readysystemhas simultaneously resolved the four core deficiencies that make existingIDSdeploymentsinadequateforrealIoTinfrastructure modelopacity,centraliseddataexposure, tamper-vulnerable audit records, and per-device computational overhead.

The comparative analysis in Section IV reveals a consistent patternacrossalleightsystems.Detectionaccuracyonstandard benchmarks is high, ranging from approximately 92% to 99%, andthealgorithmicfoundations—SHAP-basedexplainability, federatedaggregationprotocols,anddeeplearningclassifiers aresufficientlymature.Thebottleneckisnotalgorithmic.It issystems-level.TheevaluationgapidentifiedinSectionIV— cleanbenchmarks,noadversarialstress,nohardwareprofiling is not a property of any individual system but a structural featureoftheentirefieldasof2025.Closingitwillrequirenew 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 reportshardware-levelperformancemetrics.Theircomplexity arising from the coordination overhead of combining federated aggregation with SHAP computation — remains unquantifiedindeploymentterms,leavingpractitionerswithout the information needed to assess feasibility on constrained devices.

Theintegrationofblockchainasatamper-evidentauditlayer, surveyed in related work [12][13][16][30][32], addresses the storage-layer vulnerability that federated and XAI-capable systemsleaveopen.

Currentevidencesuggeststhatlightweight blockchain schemes can be implemented on resource-constrainedIoTnodes[16],butnoreviewedpapercombinesall four layers — federated training, blockchain audit, XAI inference, and edge deployment — within a single validated system.Closingthisgapisthemostconsequentialdirectionfor future work in this domain.

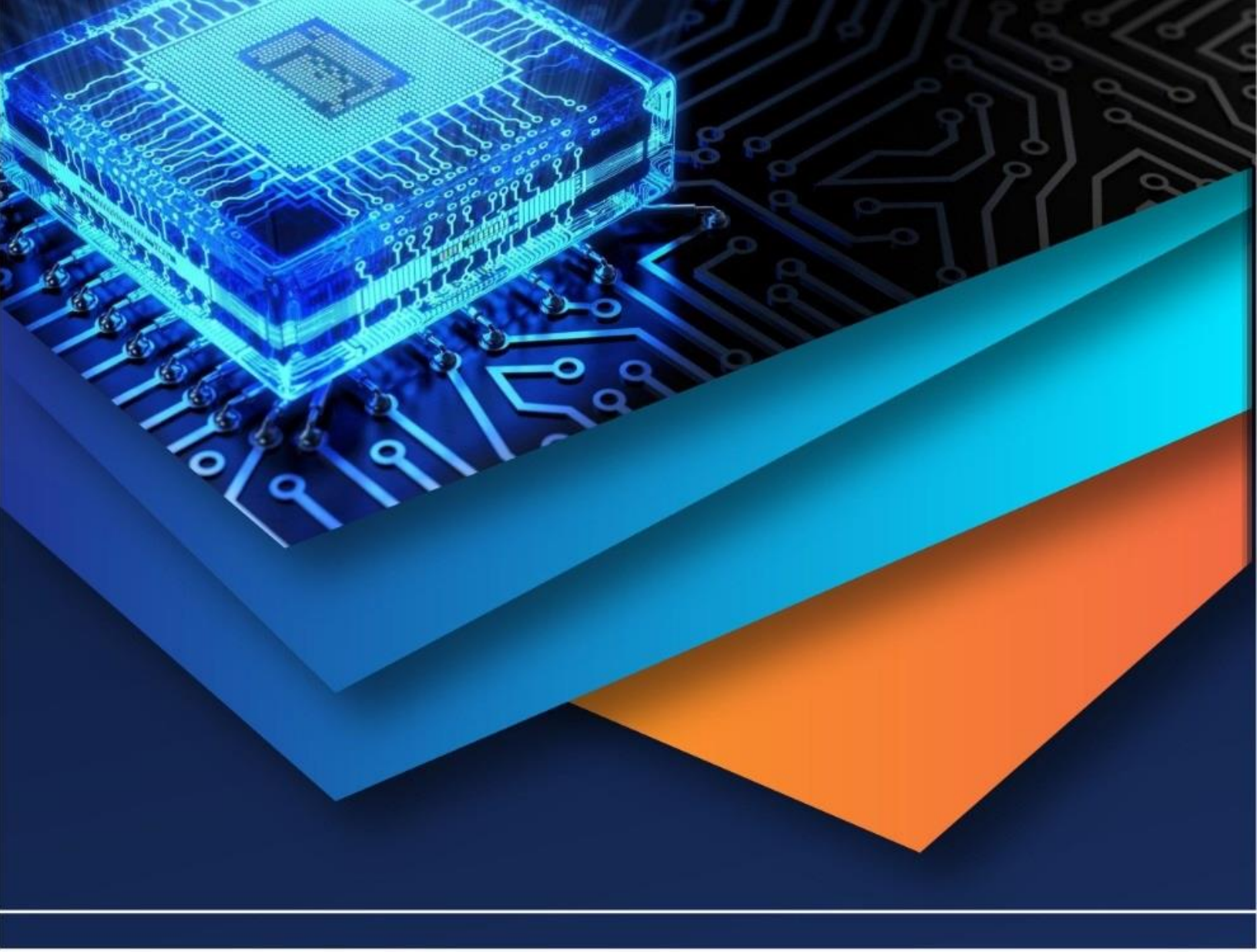
Three specific research directions follow directly from this survey's findings. First, the community needs standardised evaluation protocols that measure detection performance, alert latency,explainabilityfidelity,andenergyconsumptionjointly onphysicalIoThardwareunderliveoradversarially-perturbed traffic — not in isolation on clean benchmarks. Second, lightweight model architectures that preserve the accuracy of systemslike[1]and[4]whileoperatingwithinthememoryand power budgets of constrained edge devices remain an open design problem; the quantization approach in [2] is a partial step,butgeneralisationtolargedynamicnetworkshasnotbeen 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 adversarialrobustnessasco-equaldesignconstraintsalongside 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,andF.Anwer,"LightweightHybrid Deep Learning for IoT Intrusion Detection Combining Bi- LSTM, Genetic Feature Selection, and ExplainableAI," 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,"IEEEAccess,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 FederatedLearningFrameworkwithBuilt-InExplainabilityfor 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 DetectionArchitecture Unifying Federated Learning and SHAP-Based Explainability," IEEE Access, 2025.
[8] Q. Zhang, C. Yue, X. Dong, G. Du, and D. Wang, "Blockchain-PoweredLSTM-AttentionHybridModelfor DeviceSituationAwarenessandOn-ChainAnomalyDetection in IIoT," Sensors, vol. 25, no. 15, p. 4663, Jul. 2025.
[9] H.B.McMahan,E.Moore,D.Ramage,S.Hampson,and
[10] B. A. y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proc. 20th Int. Conf. Artif.Intell.Stat.(AISTATS),FortLauderdale,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 ofAny Classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery andDataMining(KDD),SanFrancisco,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,andH.-C.Kim, "BFLIDS:Blockchain-DrivenFederatedLearningforIntrusion DetectioninIoMTNetworks, "Sensors,vol.24,no.14, p.4591, Jul. 2024.
[15] R.W.Anwar,M.Abrar,A.Salam,andF.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 DetectionSchemeUsingDeepLearningApproachforInternet ofThings," Future Generation Computer Systems, vol. 82, pp. 761–768, 2018.
[19] N.MoustafaandJ.Slay,"UNSW-NB15:AComprehensive 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. Gwassi, 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