International Journal of Engineering Research-Online
A Peer Reviewed International Journal
ISSN: 2321-7758 http://www.ijoer.in editorijoer@gmail.com
Vol.13., Issue.2, 2025
April-June

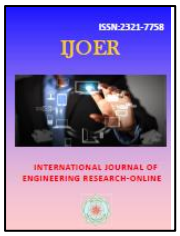
REVIEW ARTICLE

ISSN: 2321-7758

Artificial Intelligence-Enabled Cybersecurity in IoT and Edge Computing: A Review of Techniques, Applications, and Research Challenges

Jayram A. Patel
Lecturer, Department of Electrical Engineering and IT, Polytechnic,
The M.S. University of Baroda, Vadodara, Gujarat, India

DOI: 10.33329/ijoer.13.2.26



Abstract

Internet of Things (IoT) environments connect large numbers of heterogeneous, resource-constrained devices that continuously collect, process and exchange data. This connectivity improves automation and intelligent services but also expands the cybersecurity attack surface. Traditional security mechanisms remain essential, yet the scale, heterogeneity and dynamic behavior of IoT and edge networks create a need for adaptive, data-driven security techniques.

Artificial intelligence (AI), including machine learning (ML), deep learning (DL), federated learning, anomaly detection and explainable AI, has therefore become an important research direction for IoT cybersecurity. This review examines AI-enabled cybersecurity across IoT and edge-computing environments, with particular attention to intrusion detection, anomaly detection, malware and attack classification, privacy-preserving learning, distributed intelligence, explainability and resource-aware deployment. Published research through 30 April 2025 is synthesized to compare centralized, edge-based and federated approaches. The review shows that AI-based intrusion and anomaly detection dominates the literature, while federated learning provides a promising mechanism for collaborative learning without centralizing raw data.

Edge deployment can reduce latency and communication overhead, but introduces constraints involving memory, computation, energy, model updates and device security. The review also identifies persistent problems including class imbalance, non-IID data, outdated or synthetic datasets, limited reproducibility, lack of explainability, adversarial manipulation of AI models and insufficient reporting of real-world deployment characteristics. A conceptual architecture is proposed in which IoT devices, edge nodes and higher-level services cooperate through data collection, local preprocessing, AI-based detection, risk scoring, response orchestration and secure model management. The paper concludes that future IoT cybersecurity should move from accuracy-centered experimentation toward lightweight, explainable, privacy-preserving and continuously evaluated defense architectures.

Keywords: Artificial Intelligence, Cybersecurity, Internet of Things, Edge Computing, Intrusion Detection, Anomaly Detection, Machine Learning, Deep Learning, Federated Learning, Explainable AI.

1. Introduction

Internet of Things technology has transformed conventional computing systems into interconnected cyber-physical environments. Devices such as sensors, cameras, controllers, gateways, appliances, vehicles and industrial machines increasingly collect data and communicate with other devices or cloud services. The resulting infrastructure supports smart homes, healthcare, industrial automation, transportation, agriculture, energy management and many other applications. The same connectivity, however, creates additional opportunities for unauthorized access, malware propagation, denial-of-service attacks, data tampering, credential compromise and exploitation of vulnerable devices.

IoT cybersecurity is particularly difficult because the devices are heterogeneous and frequently resource constrained. Many endpoints have limited memory, processing capability, storage and battery capacity. Devices may use different protocols, operating systems and communication technologies, and they may remain deployed for long periods. Security mechanisms designed for conventional servers cannot always be transferred directly to embedded devices. NIST guidance emphasizes that IoT products should be considered within broader system risk management and that organizations should establish explicit device cybersecurity requirements rather than assuming that every device provides sufficient security by default [1–3].

Artificial intelligence provides a complementary approach. Instead of relying only on predefined signatures or static rules, AI models can learn patterns associated with normal and abnormal behavior. Supervised learning can classify known attack patterns, unsupervised or semi-supervised learning can identify deviations from expected behavior, and deep-learning models can learn complex representations from high-dimensional network or device data. These approaches are particularly attractive for intrusion detection systems (IDSs), where large volumes of network and telemetry data must be analyzed [4–6].

The integration of AI with edge computing further changes the cybersecurity architecture [4, 6, 7, 8]. Cloud-centric security can provide large computational resources and centralized analytics, but transferring all IoT data to remote servers introduces latency, communication cost and privacy concerns. Edge computing places computation closer to data sources, enabling faster local analysis and reducing the amount of raw data that must traverse the network. The resulting edge intelligence can be valuable for time-sensitive detection and response [4, 6, 7, 8].

AI itself is not automatically secure. Machine-learning models can be affected by poor-quality training data, class imbalance, distribution shifts, adversarial inputs, poisoning, model theft and privacy leakage. In federated learning, for example, raw data can remain local, but the training process can introduce new attack surfaces such as malicious updates or model poisoning. Therefore, AI-enabled cybersecurity must address both the threats against IoT systems and the security of the AI mechanisms used to defend them [9–13, 4, 5, 6].

This review focuses on the intersection of AI, cybersecurity, IoT and edge computing. It examines the principal attack categories, AI techniques used for detection and protection, edge-aware deployment strategies, federated and privacy-preserving learning, explainability, datasets and evaluation practices. The objective is to identify the limitations of existing approaches and derive a research direction for

practical, resource-aware and adaptive IoT cybersecurity [9–11, 4, 6, 7, 8, 5, 6].

2. Literature Review

2.1 IoT Cybersecurity Threat Landscape

IoT attacks can target devices, communication links, applications, management interfaces or the data generated by the system. Common network-level threats include denial-of-service and distributed denial-of-service attacks, scanning, spoofing, man-in-the-middle activity and unauthorized access. At the device level, attackers may exploit weak credentials, vulnerable firmware, exposed services or insecure update mechanisms. Malware can use compromised devices as entry points into larger networks or as participants in coordinated botnets [14–16].

Data-related attacks are also important because IoT systems often support decisions in physical environments. Modification of sensor data can influence control decisions even when the underlying device appears operational. Replay attacks can provide old but apparently valid observations, while injection attacks can introduce false measurements. In industrial environments, such manipulation may affect production or safety-related processes [17, 18].

The diversity of IoT devices makes universal protection difficult. A smart camera, medical sensor, industrial controller and low-power environmental node can have very different operating constraints. Cybersecurity mechanisms must therefore consider device capabilities, communication patterns, application context and the consequences of compromise [19, 20].

2.2 Conventional Cybersecurity and the Need for AI

Traditional cybersecurity uses mechanisms such as authentication, authorization, encryption, secure configuration, access control, firewalls, antivirus tools, signature-based intrusion detection and rule-based monitoring. These mechanisms remain fundamental and should not be replaced simply because AI is available [4–6]. However, signature-based systems are inherently limited when an attack is previously unknown or changes its observable characteristics.

AI-based detection attempts to learn patterns from data. A supervised IDS can learn a mapping between features and known attack classes, whereas an anomaly detector attempts to characterize normal behavior and flag significant deviations. This distinction is important because IoT environments can experience previously unseen attacks, device failures and unusual but legitimate behavior [5, 6, 21, 4, 6].

The 2024 review literature shows strong interest in AI-enhanced intrusion detection for IoT [4–6]. Recent reviews classify approaches according to ML and DL algorithms and identify persistent issues involving feature selection, data imbalance, interpretability and adaptation to new threats. AI should therefore be viewed as an additional analytical layer within a defense-in-depth architecture rather than as a complete cybersecurity solution [4–6].

2.3 Machine Learning for IoT Intrusion Detection

ML algorithms remain widely used because they can provide good performance with comparatively moderate computational requirements. Decision trees, random forests, support vector machines, k-nearest neighbors, naive Bayes, logistic regression and ensemble methods have all been applied to network and IoT security tasks. Their advantages include comparatively simple training, interpretable features in some models and the possibility of deployment on constrained platforms [4–6].

Feature engineering is important in classical ML. Network-flow statistics, packet counts, protocol information, timing features, connection duration, byte rates and device behavior indicators can be used to characterize traffic. Feature selection can reduce computation and remove redundant variables, which is especially useful for edge deployment [4, 6, 7, 8].

However, classical supervised models depend on labelled data. Attack datasets may be imbalanced because benign traffic is usually much more abundant than individual attack classes. A classifier can therefore achieve high overall accuracy while performing poorly on rare but important attacks. Precision, recall, F1-score, false-positive rate and class-specific results are consequently more informative than accuracy alone [4, 5, 6].

2.4 Deep Learning and Representation Learning

Deep learning can automatically learn hierarchical representations from raw or minimally processed data. Convolutional neural networks (CNNs) have been used to extract patterns from structured traffic representations, while recurrent architectures such as LSTM and GRU networks can represent temporal dependencies. Hybrid architectures combine spatial and temporal learning to capture multiple characteristics of network behavior [6, 7].

The advantage of DL is its ability to model complex relationships without relying entirely on manually designed features. However, this benefit comes with higher computational and data requirements. Large models can be difficult to deploy on small IoT devices and may require optimization before edge execution. Recent research also investigates lightweight architectures, feature reduction, pruning, quantization and knowledge distillation. These methods attempt to preserve useful detection capability while reducing memory, computation and energy requirements. Such optimization is central to the practical use of AI-based IDS at the edge [4–6].

2.5 Anomaly Detection and AI of Things

Anomaly detection is particularly relevant to IoT because not every future attack can be represented in a labelled dataset. Anomaly-based systems attempt to learn normal behavior and identify deviations. Methods include clustering, autoencoders, one-class learning, statistical techniques and other unsupervised or semi-supervised approaches [5, 6, 21, 4, 6].

A 2024 systematic literature mapping of anomaly detection on resource-constrained edge devices found growing use of TinyML and microcontroller-based implementations. The review identified classification as a common algorithmic category and CNN-based methods as frequent approaches, while also emphasizing the constraints associated with embedded deployment. This indicates that AI-enabled security is moving beyond cloud servers toward smaller devices [5, 6, 21, 7].

Anomaly detection has a major practical difficulty: not every anomaly is an attack. Firmware updates, changes in user behavior, maintenance, network congestion and device failures can all appear abnormal. A useful IDS should therefore combine anomaly scores with contextual information and response policies rather than automatically treating every deviation as malicious [5, 6, 21, 4, 6].

2.6 Federated Learning for Privacy-Preserving Security

Federated learning (FL) enables multiple devices or edge nodes to train a shared model without transferring their raw training data to a central server. This is attractive for IoT because organizations or devices may not be willing or legally able to share sensitive telemetry. Local training can preserve data locality while allowing a global model to benefit from information distributed across participants [9–11, 4, 6, 7, 8].

The 2024 survey by Li, Ge and Tian reviewed federated learning for data security and privacy in edge-IoT environments and highlighted its relevance to heterogeneous, distributed data. Federated learning can reduce the need for centralized data collection, but it does not automatically guarantee privacy or security. Model updates themselves can contain sensitive information, and malicious participants may attempt poisoning or other attacks [9–11, 12, 13].

Federated intrusion detection is also challenged by non-independent and non-identically distributed data. Different IoT devices may observe different traffic patterns and different attack types. A global model trained under an assumption of homogeneous data may therefore perform poorly for particular clients. Personalized federated learning, robust aggregation and client-selection strategies are active research directions [9–11].

2.7 Explainable AI for Cybersecurity

AI-based IDS models can be difficult to interpret, particularly when deep neural networks are used. Security operators need more than an attack label; they often need to understand why a flow or device was considered suspicious and which features contributed to the decision. Explainable AI (XAI) methods such as SHAP and LIME can provide post-hoc explanations for individual predictions [4–6, 22, 23, 24].

Explainability can support incident response, debugging, model validation and operator trust. For example, an explanation may identify unusual traffic volume, connection frequency or protocol behavior as important factors in an alert. This information can help an analyst determine whether the alert is plausible.

However, explanations must themselves be evaluated. A visually convincing explanation does not necessarily mean that the model reasoning is causally correct. XAI can also introduce computational overhead. The appropriate level of explanation should therefore depend on the operational context, device resources and consequence of the security decision [22–24].

2.8 Edge Computing and Distributed Cybersecurity

Edge computing places processing closer to IoT data sources. In cybersecurity, this can enable local traffic analysis, device profiling, anomaly detection and preliminary response. Edge nodes can aggregate data from nearby devices and perform more computationally intensive analysis than individual microcontrollers while still remaining closer to the operational environment than a remote cloud [5, 6, 21, 4, 7, 8].

The principal benefit is reduced latency. A suspicious communication pattern can be detected locally without waiting for a remote round trip. Edge processing can also reduce bandwidth consumption by transmitting alerts, summaries or selected features instead of complete raw streams.

Edge computing introduces new security boundaries. An edge node becomes a valuable target because it may aggregate information from many devices. Compromise of the edge layer can therefore have a larger impact than compromise of a single endpoint. Authentication, secure boot, protected model storage, secure updates, access control and monitoring of edge nodes are required [4, 6, 7, 8].

Resource management is another challenge. Edge nodes are more capable than many endpoints but remain more constrained than cloud servers. AI models must therefore be selected according to memory, CPU/GPU/NPU capability, energy and latency requirements. Deployment decisions should consider the complete pipeline rather than model accuracy alone [4, 6, 7, 8].

3. Comparative Analysis of AI-Enabled IoT Cybersecurity Approaches

3.1 Datasets, Evaluation and Reproducibility

Datasets strongly influence the reported performance of AI-based cybersecurity systems. Public intrusion-detection datasets are useful because they allow researchers to compare methods, but they may not represent current IoT traffic or the diversity of modern deployments. Reviews of IoT IDS research repeatedly identify concerns about outdated datasets, class imbalance and limited representation of real-world conditions [4–6, 21].

Evaluation should therefore include more than overall accuracy. Precision, recall, F1-score, false-positive rate, detection latency and computational cost are important. For edge systems, memory footprint, energy consumption, model size, inference time and communication overhead should also be reported. A model that improves F1-score slightly but requires several times the computational resources may be unsuitable for an embedded deployment.

Reproducibility is another concern. Studies should report preprocessing, feature selection, train/test splitting, class balancing, hyperparameters, model architecture and hardware environment. Leakage between training and testing data can produce unrealistically high results, especially when traffic from the same device or time period appears in both sets.

Real-world evaluation should include changes in network conditions, device heterogeneity, new devices, previously unseen attacks and benign anomalies. Continuous deployment also requires monitoring of model drift and mechanisms for updating models safely.

3.2 AI Security: Adversarial Attacks and Model Protection

Using AI for cybersecurity creates a second security problem: the AI system itself can become a target. Attackers may attempt to manipulate inputs so that malicious traffic is classified as benign. In training-time attacks, poisoned data or malicious updates can alter model behavior. Other threats include model extraction, inference and exploitation of weaknesses in model deployment [12, 13].

These threats are especially relevant to distributed IoT environments. A compromised device participating in federated learning could send malicious updates. A local IDS could be targeted by carefully crafted traffic intended to evade detection [9–11, 4, 6].

Edge nodes could also be attacked because they host models and aggregate data from multiple devices [4, 6, 7, 8]. Defenses include robust aggregation, secure update validation, anomaly detection on model updates, adversarial training, input validation, access control and secure model storage. The appropriate defense depends on the threat model. AI security should therefore be documented as explicitly as network security rather than treated as an implicit property of the algorithm [5, 6, 12, 13].

3.3 Research Gap Identified

The literature demonstrates substantial progress in AI-based IoT intrusion detection, but several gaps remain. First, many studies optimize predictive performance using benchmark datasets while providing limited evidence about actual deployment on constrained hardware. Accuracy therefore often receives more attention than latency, memory, energy, communication cost and update reliability [4–6, 21].

Second, many approaches are evaluated in relatively controlled settings. Real IoT environments are heterogeneous, dynamic and non-IID. Different devices generate different traffic distributions, and benign behavior can change over time. A security model must therefore adapt without becoming unstable or excessively sensitive to normal variation [9–11].

Third, privacy-preserving approaches such as federated learning solve only part of the data-sharing problem. The federated training process introduces its own security risks and communication overhead. Practical systems require protection against poisoning, malicious clients, model leakage and unreliable participants [9–11, 12, 13].

Fourth, explainability is still unevenly integrated into security workflows. A high-performing black-box detector may generate alerts that are difficult for an operator to trust or investigate. Explainable and lightweight models

are therefore needed, but the explanation itself must be validated [22–24].

Finally, many research papers focus on detection while giving less attention to response and recovery. A complete cybersecurity system should connect detection to risk scoring, containment, logging, recovery and secure model management. The key research opportunity is thus an integrated, edge-aware cybersecurity architecture that combines detection, privacy, explainability and response under realistic resource constraints.

3.4 Proposed Conceptual Framework for AI-Enabled Edge Cybersecurity

The proposed framework is organized as a layered defense architecture spanning IoT devices, edge nodes and optional cloud or enterprise services. At the device layer, secure configuration, authentication and basic telemetry collection provide the first security boundary. Lightweight local models can perform initial anomaly or behavior checks where device resources permit [5, 6, 21, 4, 7, 8].

The edge layer receives selected telemetry from multiple devices and performs more capable analysis. It can maintain device profiles, calculate anomaly scores, classify suspicious flows and correlate events across neighboring devices. Because the edge node is closer to the devices, it can also support low-latency response actions such as blocking a connection, isolating a device or increasing monitoring intensity, subject to authorized policy [5, 6, 21, 4, 7, 8].

A model-management layer handles training, validation, deployment and updates. In a centralized architecture, the edge node can receive a validated model from a trusted server. In a federated architecture, multiple edge nodes can train locally and share model updates. In both cases, update validation, version control and rollback are required [9–11, 4, 6, 7, 8].

An explainability layer can provide interpretable information about important features or behavioral deviations. The objective is to assist human analysts and support auditing, not to expose internal model details unnecessarily.

The response layer converts detection results into risk-aware actions. Low-risk anomalies may be logged for observation, while higher-risk events can trigger additional verification, containment or escalation. This layered approach avoids treating a single AI score as an automatic security decision [4–6].

4. Methodology

This paper follows a structured narrative review and engineering-oriented synthesis. The review questions are: (1) which AI techniques are most commonly used for IoT cybersecurity; (2) how are AI-based security functions distributed between endpoints, edge nodes and cloud services; (3) what are the principal privacy, explainability and adversarial risks; (4) which datasets and evaluation metrics dominate the literature; and (5) what architecture can support practical, resource-aware and adaptive cybersecurity [4, 6, 7, 8, 12, 13, 21, 6]?

The literature scope was limited to publications available through 30 April 2025. Priority was given to peer-reviewed reviews, systematic literature reviews, surveys, representative research studies and authoritative cybersecurity guidance. The synthesis was organized around threats, AI techniques, edge deployment, federated learning, explainability, evaluation and research gaps [9–11, 4, 6, 7, 8].

The paper does not claim a new experimental intrusion-detection result. The Results and Observations section therefore reports synthesized findings from the reviewed literature and converts them into engineering requirements for a future integrated architecture. This distinction is maintained to avoid presenting literature-derived observations as original experimental data.

5. Results and Observations

Observation 1 — AI-based intrusion and anomaly detection dominates current research. Reviews published in 2024 and early 2025 show strong concentration on IDS applications using classical ML, DL and anomaly-detection methods. This confirms the importance of detection but also suggests that other security functions, including access control, secure configuration and response, receive comparatively less attention in AI research [5, 6, 21, 4, 6].

Observation 2 — Edge deployment is technically promising but underreported. Moving inference closer to IoT devices can reduce latency and communication overhead, but studies do not always report memory, energy, model size or sustained inference behavior. These properties should become standard evaluation dimensions for edge cybersecurity [4, 6, 7, 8].

Observation 3 — Federated learning can reduce raw-data sharing but introduces new security challenges. Distributed training is well suited to privacy-sensitive IoT environments, yet non-IID data, malicious clients, poisoning and communication overhead remain important barriers [9–11, 12, 13].

Observation 4 — Explainability is increasingly important. Security operators need interpretable evidence for alerts, especially when AI models influence containment decisions. XAI methods can improve transparency, but explanation fidelity and computational cost require evaluation [22–24].

Observation 5 — Dataset limitations constrain generalization. Public datasets support comparison but may be outdated, imbalanced or insufficiently representative of modern heterogeneous IoT networks. Realistic evaluation should include temporal variation, device diversity and previously unseen attack conditions [4, 5, 6].

Observation 6 — AI security must be treated as part of cybersecurity. Models can be attacked through evasion, poisoning or other manipulation. A secure IDS therefore needs protected training, validated updates, robust inference and secure model storage [4–6, 12, 13].

Observation 7 — Detection should connect to response. Identifying an anomaly is only one step in cybersecurity. A practical system should transform detection results into risk-aware actions such as logging, additional verification, containment, escalation and recovery [5, 6, 21].

Observation 8 — Hybrid architectures provide a practical direction. Combining lightweight local detection, more capable edge analytics, privacy-preserving collaboration and centralized governance can balance responsiveness, resource constraints and security management.

6. Discussion

The central finding is that AI-enabled IoT cybersecurity should move beyond classifier accuracy toward system-level security engineering. A model can achieve high benchmark performance and still be unsuitable for deployment if it is too large, too slow, energy intensive, difficult to update or vulnerable to manipulation.

A layered edge architecture provides a practical balance. Endpoints can perform lightweight checks, edge nodes can perform more complex correlation and inference, and centralized services can provide governance, threat intelligence and model lifecycle management. This distribution reduces unnecessary data movement while preserving access to higher-level resources when needed [4, 6, 7, 8].

Federated learning is particularly relevant when organizations cannot centralize raw security telemetry. However, privacy should not be equated with federated learning alone. Secure aggregation, participant authentication, robust aggregation and monitoring for malicious updates are needed to protect the collaborative training process [9–11].

Explainability should also be integrated with operations. A security analyst may need to

know which features or behaviors caused an alert, but explanations should be concise and actionable. Future systems should evaluate whether XAI actually improves analyst decision making rather than simply generating visually appealing explanations [22–24].

The most important methodological improvement is realistic deployment evaluation. Researchers should report the hardware platform, inference time, memory footprint, energy consumption and communication overhead. They should also test the system against temporal changes and unseen attacks. Such reporting would make it possible to compare research prototypes on practical criteria rather than benchmark accuracy alone.

The framework should finally distinguish detection from response. AI can provide a probability or risk estimate, but automated containment may have consequences for legitimate devices. Response policies should therefore incorporate confidence, asset criticality, contextual evidence and authorization. High-impact actions may require human confirmation unless the policy has been validated for automatic operation.

8. Reliability, Security of the AI Pipeline and Lifecycle Management

A cybersecurity AI system should itself be treated as a protected asset. Training data, feature pipelines, model files, configuration policies and update mechanisms can all become attack targets. The system should maintain integrity across the full machine-learning lifecycle, from data acquisition to model deployment and monitoring.

Secure model management requires version control and authenticated updates. Every deployed model should have an identifiable version and validation status. If a new model performs poorly or is compromised, rollback to a previously validated version should be possible. The update process should also prevent unauthorized parties from replacing the model.

Data pipelines require equivalent attention. Attackers may attempt to inject false telemetry or manipulate labels during training. Data-quality checks, provenance tracking and access controls can reduce these risks. In federated environments, participating clients should be authenticated and their updates should be evaluated for abnormal behavior [9–11].

Operational monitoring should continue after deployment. Changes in traffic patterns may indicate either a new attack or legitimate changes in system behavior. Monitoring model confidence, false-alert rates and distribution changes can help identify model drift and trigger retraining or human review.

These requirements reinforce the concept of cybersecurity as a lifecycle rather than a single detection algorithm. Design, deployment, monitoring, update, incident response and retirement should all be considered in the system architecture.

9. Future Scope and Research Roadmap

Future research should first establish standardized edge-aware benchmarks. Public datasets should be complemented by controlled experiments and, where possible, realistic operational traces. Evaluation should report both cybersecurity metrics and resource metrics so that accuracy and deployability can be considered together [4, 5, 6].

The next stage should focus on lightweight and adaptive models. Compression, quantization, pruning, knowledge distillation and efficient architectures can make AI more suitable for constrained devices. The challenge is to reduce resource consumption without creating unacceptable degradation in detection capability.

Personalized and federated security models are another important direction. Devices in different environments have different normal behavior, so a single global model may not be optimal. Federated and personalized learning can allow local adaptation while retaining a collaborative

component, provided that model updates are securely managed [9–11].

Explainable and human-in-the-loop cybersecurity should also receive greater attention. Future systems can combine automated detection with analyst feedback, using human review for ambiguous or high-impact cases. This approach can improve trust and provide valuable information for subsequent model refinement [22–24].

Finally, AI-enabled cybersecurity should be integrated with response and recovery. Future edge systems should not stop at detecting an attack; they should support policy-driven containment, evidence collection, secure recovery and post-incident learning. Such capabilities would transform AI-based IDS research into broader intelligent cybersecurity systems [4–6].

10. Limitations

This paper is an engineering-oriented narrative review and conceptual synthesis rather than a formal meta-analysis. The review does not reproduce quantitative results from individual studies and does not claim a new intrusion-detection benchmark [4–6].

The literature cutoff of 30 April 2025 was selected to maintain a defined historical scope suitable for a manuscript associated with the March–April 2025 period. Later publications are intentionally excluded from the evidence synthesis.

Research in AI-enabled cybersecurity evolves rapidly, and differences in datasets, attack taxonomies, preprocessing and evaluation protocols make direct comparison difficult. Conclusions should therefore be interpreted as research-direction findings rather than universal performance claims [4, 5, 6].

11. Conclusion

IoT and edge computing create substantial opportunities for intelligent services but also expand the cybersecurity attack surface. AI-based techniques provide useful capabilities for intrusion detection, anomaly detection, classification, privacy-preserving learning and security analytics. The literature demonstrates significant progress, particularly in ML/DL-based IDS, federated learning and edge intelligence [9–11, 5, 6, 21, 4, 6, 7, 8].

At the same time, practical deployment remains constrained by heterogeneous devices, limited resources, non-IID data, class imbalance, outdated datasets, limited explainability and attacks against the AI pipeline itself. These issues demonstrate that benchmark accuracy alone is insufficient to establish a secure and deployable solution [9–11, 4, 5, 6].

A future AI-enabled IoT cybersecurity architecture should therefore combine lightweight local intelligence, edge-level correlation, privacy-aware collaborative learning, explainable detection, secure model management and policy-driven response. Evaluation should include security effectiveness, latency, memory, energy, communication overhead, robustness, privacy and operational reliability [22–24].

The convergence of AI, IoT and edge computing is consequently best understood as a systems-engineering problem. The strongest research direction is not a single new classifier but an integrated and continuously evaluated security framework capable of detecting changing threats while respecting the constraints and security requirements of distributed IoT environments [4, 6, 7, 8].

Table 1. Comparison of AI-enabled IoT cybersecurity approaches

Approach	Main Function	Strengths	Limitations	Deployment	Security Role
Rule/signature IDS	Known attack matching	Simple, fast, predictable	Weak for unknown attacks	Endpoint/edge/cloud	Baseline defense
Classical ML	Classification/anomaly detection	Moderate cost; effective with engineered features	Needs suitable data/features	Edge/cloud	Adaptive detection
Deep learning	Complex pattern learning	Strong representation learning	Higher resource/data needs	Edge/cloud	Advanced detection
Federated learning	Distributed model training	Data locality; collaboration	Non-IID data; poisoning risk	Edge/federated edge	Privacy-aware learning
XAI	Decision explanation	Improves interpretability	Extra computation; explanation limits	Edge/cloud	Trust and analysis
Hybrid Edge-AI	Local detection + coordination	Low latency; reduced raw-data transfer	System complexity	Endpoint + edge + cloud	Integrated defense

Table 2. AI-IoT cybersecurity evaluation requirements

Dimension	**Key Measures**	**Why Important**	**Recommended Evidence**
Detection	Precision, recall, F1, FPR, detection rate	Measures security effectiveness	Per-class and overall results
Latency	Inference and alert delay	Critical for real-time edge response	Hardware-based timing
Resource use	Memory, CPU/GPU, energy, model size	Determines deployability	Device/edge profiling
Robustness	Noise, drift, unseen attacks	Measures resilience	Stress and temporal tests
Privacy	Data exposure, update leakage	Protects sensitive telemetry	Threat-model analysis
Explainability	Fidelity, usefulness, overhead	Supports trust and analysis	XAI validation
Operations	Update, rollback, recovery	Ensures long-term reliability	Lifecycle testing

References

[1]. Mallidi SKR, Ramisetty RR. Advancements in training and deployment strategies for AI-based intrusion detection systems in IoT: a systematic literature review. Discover Internet Things. 2025;5:8. doi:10.1007/s43926-025-00099-4.

[2]. Li H, Ge L, Tian L. Survey: federated learning data security and privacy-preserving in edge-Internet of Things. Artif Intell Rev. 2024;57:130. doi:10.1007/s10462-024-10774-7.

[3]. Vyas A, Lin PC, Hwang RH, Tripathi M. Privacy-preserving federated learning for intrusion detection in IoT environments: a survey. IEEE Access. 2024;12:127018-127050. doi:10.1109/ACCESS.2024.3454211.

[4]. Adhikari D, Jiang W, Zhan J, Rawat DB, Bhattarai A. Recent advances in anomaly detection in Internet of Things: status, challenges, and perspectives. Comput Sci Rev. 2024;54:100665. doi:10.1016/j.cosrev.2024.100665.

[5]. Anomaly detection based on Artificial Intelligence of Things: a systematic literature mapping. Internet Things. 2024;25:101063. doi:10.1016/j.iot.2024.101063.

[6]. Review of artificial intelligence for enhancing intrusion detection in the Internet of Things. Eng Appl Artif Intell. 2024;127(Pt A):107231. doi:10.1016/j.engappai.2023.107231.

[7]. Recent endeavors in machine learning-powered intrusion detection systems for the Internet of Things. J Netw Comput Appl. 2024;229:103925. doi:10.1016/j.jnca.2024.103925.

[8]. A survey on intrusion detection system in IoT networks. Complex Intell Syst. 2024. doi:10.1016/j.csa.2024.100082.

[9]. Ali S, Li Q, Yousafzai A. Blockchain and federated learning-based intrusion detection approaches for edge-enabled industrial IoT networks: a survey. Ad Hoc Netw. 2024;152:103320. doi:10.1016/j.adhoc.2023.103320.

[10]. Sharma B, Sharma L, Lal C, Roy S. Explainable artificial intelligence for intrusion detection in IoT networks: a deep learning based approach. Expert Syst Appl. 2024;238:121751. doi:10.1016/j.eswa.2023.121751.

[11]. A survey on explainable artificial intelligence for Internet traffic classification and prediction, and intrusion detection. IEEE Commun Surv Tutor. 2024. doi:10.1109/COMST.2024.3504955.

[12]. A survey on anomaly detection in IoT and cloud computing security. In: 2024 8th International Conference on I-SMAC. IEEE; 2024. doi:10.1109/I-SMAC61858.2024.10714750.

[13]. Roadmap of adversarial machine learning in Internet of Things-enabled security systems. Sensors (Basel). 2024;24(16):5150. doi:10.3390/s24165150.

[14]. A survey of machine learning in edge computing: techniques, frameworks, applications, issues, and research directions. Technologies (Basel). 2024;12(6):81. doi:10.3390/technologies12060081.

[15]. Fagan M, Megas KN, Marron J, Brady KG, Cuthill BB, Herold R, et al. IoT device cybersecurity guidance for the federal government: establishing IoT device cybersecurity requirements. NIST SP 800-213. Gaithersburg (MD): National Institute of Standards and Technology; 2021. doi:10.6028/NIST.SP.800-213.

[16]. Megas KN, Fagan M, Marron J, Brady KG, Cuthill BB, Herold R, et al. IoT device cybersecurity guidance for the federal government: IoT device cybersecurity requirement catalog. NIST SP 800-213A. Gaithersburg (MD): National Institute of Standards and Technology; 2021. doi:10.6028/NIST.SP.800-213A.

[17]. Boeckl K, Fagan M, Fisher B, Lefkovitz N, Megas KN, Nadeau EM, et al. Considerations for managing Internet of Things (IoT) cybersecurity and privacy risks. NISTIR 8228. Gaithersburg (MD): National Institute of Standards and Technology; 2019.

[18]. Internet of intelligent things: a convergence of embedded systems, edge computing and machine learning. Internet Things. 2024;26:101153. doi:10.1016/j.iot.2024.101153.

[19]. Cyber attacks management in IoT networks using deep learning and edge computing. Procedia Comput Sci. 2024;251:721-726. doi:10.1016/j.procs.2024.11.175.

[20]. Khan N, Ahmad K, Al Tamimi A, Alani MM, Bermak A, Khalil I. Explainable AI-based intrusion detection system for Industry 5.0: an overview of the literature, associated challenges, the existing solutions, and potential research directions. arXiv. 2024. arXiv:2408.03335.