A SYSTEMATIC LITERATURE REVIEW ON AI-BASED INTRUSION DETECTION SYSTEMS FOR IOT AND SENSOR NETWORKS

Ouechtati Hamdi

ISG, LR11ES03 SMART Lab, Universite de Tunis, 41, Rue de la Liberté, Cité Bouchoucha, 2000, Le Bardo, Tunis, Tunisia

ABSTRACT

The Internet of Things (IoT) and sensor networks enable large-scale automation, monitoring, and control, but their heterogeneity, limited resources, and pervasive connectivity substantially enlarge the cyberattack surface. Artificial Intelligence (AI)-based Intrusion Detection Systems (IDS) have therefore become a major research direction because they can learn attack patterns, detect abnormal behavior, and adapt to evolving threats. This Systematic Literature Review (SLR), conducted according to PRISMA 2020 principles, synthesizes peer-reviewed research on AI-based IDS for IoT and sensor networks published from 2015 to early 2025. From an initial pool of 823 records, 75 studies were selected after duplicate removal, screening, eligibility assessment, and quality appraisal. The review compares the methods used in the literature, including classical machine learning, deep learning, hybrid and ensemble learning, federated learning, and edge-oriented lightweight models. It also adds a quality-assessment scheme, a method-by-method synthesis, a quantitative performance summary, and a benchmarking protocol. The findings show that deep learning and hybrid methods often achieve the highest detection performance on benchmark datasets, whereas federated and edge-based IDS provide stronger privacy, latency, and deployment advantages. Persistent gaps remain in dataset realism, cross-dataset validation, adversarial robustness, explainability, energy-aware deployment, and reproducible evaluation. The review concludes by proposing a layered conceptual framework and future directions for trustworthy, lightweight, explainable, and privacy-preserving IDS in next-generation IoT environments.

KEYWORDS

IoT, Sensor Networks, Intrusion Detection Systems, Artificial Intelligence, Machine Learning, Deep Learning, Federated Learning.

1. INTRODUCTION

The Internet of Things (IoT) is a key paradigm of modern computing, connecting billions of devices from industrial sensors to medical implants. By 2030, over 30 billion devices are expected globally, generating vast heterogeneous data and enabling applications in smart healthcare, agriculture, energy, and transportation [1]. However, this growth increases the attack surface, making IoT networks vulnerable to cyberattacks such as denial-of-service (DoS), data exfiltration, identity spoofing, and malware propagation [2], [3]. IoT devices, often limited in memory, computing power, and software updates, render traditional security solutions like firewalls and signature-based IDS ineffective [4]. Therefore, intrusion detection has become a critical focus for IoT network security.

1.1. Motivation and Problem Statement

The increasing sophistication of cyberattacks on IoT networks highlights the need for adaptive and intelligent security mechanisms. Botnets such as Mirai, Mozi, and BrickerBot have shown that static or rule-based protections cannot keep pace with emerging threats [5]. AI-based Intrusion Detection Systems (IDS) offer a viable alternative by learning behavioral patterns and identifying anomalies or unknown attacks in real time [6].

Despite growing interest, research on AI-based IDS in IoT remains fragmented across domains such as deep learning, federated learning, and edge computing [7][16][17]. IoT-specific challenges—including data imbalance, limited labeled datasets, device heterogeneity, and privacy constraints—complicate model deployment and evaluation [2][3][7]. These limitations hinder the comparability of existing solutions and the assessment of their real-world applicability.

A systematic review is therefore essential to consolidate current research, identify emerging trends, and highlight unresolved challenges in AI-based intrusion detection for IoT and sensor networks.

1.2. Objectives and Research Questions

Defining clear objectives and research questions guides the scope of the review.

This study provides a Systematic Literature Review (SLR) of AI-based IDS for IoT and sensor networks. The main objectives are:

- Categorize and summarize existing AI-based IDS models.
- Evaluate the effectiveness of ML and DL techniques for intrusion detection.
- Identify commonly used datasets, performance metrics, and benchmarking approaches.
- Highlight challenges and propose future research directions.

The research questions addressed are:
- **RQ1:** Which AI methodologies are used for IoT intrusion detection?
- **RQ2:** How accurate, scalable, and efficient are these models?
- **RQ3:** What datasets, tools, and frameworks are commonly used?
- **RQ4:** What unresolved issues and potential future research directions exist?

These objectives and questions provide a structured framework to assess the current state of AI-based IDS for IoT.

1.3. Methodology

A systematic approach ensures transparency, reproducibility, and reliability of the review. This study follows the PRISMA 2020 guidelines [8] and includes publications from 2015 to early 2025 retrieved from IEEE Xplore [4,7], ACM [14,29], SpringerLink [12,18], ScienceDirect [2,3], Wiley, MDPI [22,26], and Google Scholar [25].

The search string combined keywords from three domains:

- Security: "Intrusion Detection System" OR "IDS";
- Application: "IoT" OR "Internet of Things" OR "sensor network";

- Technology: "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Network" OR "Federated Learning" [16,17].

The inclusion criteria were peer-reviewed English-language research papers explicitly addressing AI/ML-based IDS in the IoT environment. The quality was evaluated based on Kitchenham and Charters (2007) [9], taking into account research objectives, methodology, dataset description, validation, and discussion of limitations. 75 papers were selected and qualitatively and quantitatively analyzed and categorized into Machine Learning (ML), Deep Learning (DL), Hybrid, and Federated/Edge-based solutions [12,14,16,17,22].

This systematic approach ensures a comprehensive, reproducible, and high-quality synthesis of the literature, providing a reliable foundation for subsequent analysis and discussion.

1.4. Contributions of the Review

The review offers structured insights into AI-based IDS and highlights emerging directions.

The contributions are as follows:

1. **Taxonomy of AI-based IDS:** Classifying models into supervised, unsupervised, and hybrid approaches, addressing data imbalance, adaptability, and scalability in IoT environments.
2. **Comparative Analysis of AI Techniques:** Evaluating CNN, LSTM, GRU, ensemble methods, reinforcement learning, and federated learning, balancing accuracy, computational complexity, interpretability, and real-time performance.
3. **Overview of Datasets and Metrics:** Summarizing datasets (NSL-KDD, BoT-IoT, CICIDS2017, UNSW-NB15, TON_IoT) and metrics (accuracy, F1-score, latency, energy efficiency).
4. **Identification of Challenges and Future Directions:** Highlighting gaps in explainability, privacy, dataset scarcity, and deployment efficiency. Promising approaches include TinyML, Explainable AI, and blockchain-based federated learning.

These contributions consolidate the state of the art and provide a strategic roadmap for developing robust and efficient AI-based IDS for IoT networks.

1.5. Paper Structure

The paper is organized to guide readers from methodology to findings and future perspectives.

- **Section 2:** Research methodology (PRISMA, selection criteria, quality assessment).
- **Section 3:** Taxonomy and classification of AI-based IDS.
- **Section 4:** Comparative analysis, datasets, and evaluationmetrics.
- **Section 5:** Challenges and open research issues.
- **Section6:** Conclusion and future directions, including Explainable AI, Federated Learning, and TinyML.

This structure ensures a coherent flow from theory to application, providing a comprehensive overview of AI-based intrusion detection for IoT.

2. RESEARCH METHODOLOGY

Systematic Literature Review (SLR) is a reproducible and transparent method of searching, critically appraising, and collecting all the available evidence on a topic of interest. The current review follows the PRISMA 2020 guidelines [8] to enable transparency, methodological quality, and reproducibility [9]. There are four phases involved: (i) planning the review, (ii) formulating the search strategy, (iii) study selection and critical appraisal, and (iv) synthesis and analysis of results.

2.1. Planning the Review

The planning phase determined the scope, research questions, and objectives of this review, which was focused on AI-based Intrusion Detection Systems (IDS) in IoT and sensor networks, published between 2015 and early 2025. The timeframe was chosen because AI-driven security in IoT gained relevance after 2015, coinciding with the development of deep learning architectures such as CNNs and LSTMs [12,14]. Recent years (2020–2025) also witnessed increasing interest in federated learning, edge intelligence, and self-adaptive IDS models [16,17,20].

2.2. Search Strategy

A systematic search strategy was designed for attaining maximum coverage of quality research on AI-based IDS for IoT. The strategy adheres to PRISMA 2020 guidelines [8] for better reproducibility and less selection bias. Databases utilized are:

- **IEEE Xplore Digital Library**: Premier journals and conferences in networking, AI, and cybersecurity [4,7].
- **ACM Digital Library**:Research on machine learning, algorithms, and distributed systems [14,29].
- **ScienceDirect (Elsevier)**: IoT and AI multidisciplinary research [2,3].
- **SpringerLink**:Machine learning applications and IoT architectures [12,18].
- **Wiley Online Library**: Communications technologies and embedded systems.
- **MDPI Open Access Journals**: Recent publications on federated learning, XAI, and real-time IDS [22,26].

Google Scholar was also used to identify gray literature in the form of technical reports and preprints, including emerging trends not yet published [25].

2.2.1. Search Strings

Keywords and Boolean operators considered the intersection of IDS, AI, and IoT. Security terms were "intrusion detection," "anomaly detection," or "signature-based IDS." Application terms were "IoT" and "sensor networks," and technology terms were "machine learning," "deep learning," "federated learning," and "neural networks" [7,16,17]. Search syntax was adjusted based on each database: titles/abstracts in IEEE Xplore and ACM, publication filters in ScienceDirect and SpringerLink, and manual exclusion of non-scientific sources in Google Scholar [8,9].

2.2.2. Inclusion and Exclusion Criteria

Research was selected for quality and relevance based on the criteria defined in Table 1 for this purpose. It only took into account recent, peer-reviewed, and empirically validated research on AI-based IDS for IoT [14,19,22].

Table 1. Inclusion and Exclusion Criteria for Study Selection.

Criteria	Inclusion	Exclusion
Publication Type	Peer-reviewed journal articles, conference papers, surveys, or book chapters.	Non-peer-reviewed materials such as blog posts, editorials, or theses.
Publication Period	Studies published between 2015 and 2025 to capture recent advances.	Publications before 2015.
Language	English-language studies to ensure accessibility and standardization.	Publications in other languages.
Focus	AI-, ML-, or DL-based IDS specifically applied to IoT or sensor networks.	Generic IDS without IoT context, or studies not employing AI techniques.
Evaluation	Studies providing clear performance metrics (e.g., accuracy, precision, recall, F1-score) or simulation/experimental validation.	Works lacking empirical evaluation or simulation results.

Only studies satisfying all inclusion criteria were retained for subsequent analysis. This process ensured that the final corpus was both methodologically sound and contextually relevant to AI-driven IoT security research.

2.3. Study Selection Process

The selection process followed PRISMA guidelines [8]. Figure 1 illustrates the route from identification to final inclusion. From IEEE Xplore, SpringerLink, ScienceDirect, and ACM Digital Library, 612 papers were retrieved to begin with. After 187 duplicate records were removed, 425 papers proceeded with screening for relevance to AI-based IoT IDS. 127 papers were screened in full text, resulting in 75 studies meeting all the inclusion criteria, serving as the end corpus for systematic synthesis [2,7,12,17,22]. The final set of 75 included studies is reflected in the reference list and cited throughout the thematic synthesis according to topic: IoT datasets and benchmarks [31]-[34], deep and anomaly-based IDS [35]-[42], ML-based IDS and survey foundations [43]-[49], federated learning [50]-[55], explainability and adversarial robustness [56]-[64], lightweight edge/fog deployment [65]-[72], and IoT deployment-oriented IDS [73]-[75].

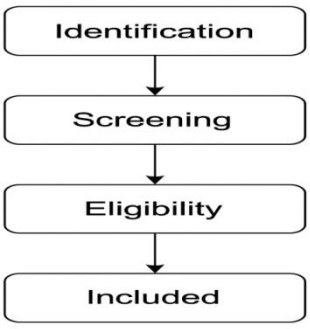

Figure 1.PRISMA-based selection process for AI-based IDS studies.

3. TAXONOMY AND CLASSIFICATION OF AI-BASED IDS APPROACHES

Artificial Intelligence (AI) has profoundly influenced the development and optimization of Intrusion Detection Systems (IDS) in Internet of Things (IoT) and sensor network environments. The inherent heterogeneity, decentralization, and resource constraints of IoT infrastructures demand intelligent IDS capable of learning autonomously, adapting to dynamic threats, and generalizing effectively across diverse contexts. This section introduces a comprehensive taxonomy of AI-based IDS approaches, organized into four principal paradigms:

1. Machine Learning-based IDS, relying on traditional algorithms for feature extraction and classification.
2. Deep Learning-based IDS, leveraging neural architectures for hierarchical feature representation.
3. Hybrid and Ensemble IDS, integrating multiple AI models to enhance detection accuracy and robustness.
4. Federated and Edge-based IDS, enabling distributed intelligence while preserving data privacy and reducing latency.

Each category is analyzed in terms of its methodological principles, architectural characteristics, and implementation challenges, providing a structured understanding of the evolution and diversity of AI-driven intrusion detection solutions in IoT ecosystems.

3.1. Machine Learning-based IDS

Machine Learning (ML) represents the foundation of modern AI-based IDS, providing statistical models that can detect anomalous patterns in network traffic without explicit programming. Early IDS implementations for IoT primarily relied on supervised learning algorithms, such as Support Vector Machines (SVM), Random Forests (RF), and k-Nearest Neighbors (kNN), trained on labeled datasets to classify network behavior as normal or malicious [6][7][25][45]-[49][73]-[75]. SVM-based models have demonstrated strong performance in binary and multi-class classification tasks due to their ability to handle non-linear decision boundaries using kernel functions [8]. However, they are limited by high computational cost when processing large-scale IoT data. Random Forests, on the other hand, offer better scalability and interpretability through ensemble decision trees, which improve robustness and reduce overfitting [9].

Unsupervised learning techniques, such as K-means clustering and Self-Organizing Maps (SOM), are widely applied in IoT environments where labeled data is scarce [10]. These models group network behavior patterns and identify outliers that may correspond to intrusions. Semi-supervised and reinforcement learning have also gained attention for adaptive intrusion detection under dynamic conditions [11]. Reinforcement learning agents can optimize detection strategies in real time based on feedback, making them suitable for distributed IoT environments with evolving threats.

Despite their effectiveness, ML-based IDS face critical challenges. They often rely on handcrafted features, which limit generalization to unseen attacks. Furthermore, their detection accuracy degrades when applied to heterogeneous IoT devices generating diverse traffic types. Consequently, research attention has shifted towards deep learning approaches capable of automatic feature extraction and hierarchical representation learning.

3.2. Deep Learning-based IDS

Deep Learning (DL) techniques have revolutionized IDS research by enabling end-to-end detection pipelines capable of automatically extracting complex spatio-temporal patterns from network data. In the context of IoT and sensor networks, Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Autoencoders (AE), and Generative Adversarial Networks (GAN) are the most widely studied architectures [12][13][35]-[42].

CNN-based IDS leverage convolutional filters to capture local dependencies in network flow features, making them highly effective for detecting signature-based and anomaly-based attacks. CNN architectures have shown superior performance on benchmark datasets such as CICIDS2017, BoT-IoT, and TON_IoT [14]. Their hierarchical feature extraction allows them to identify subtle patterns in traffic that may be missed by traditional models.

RNN and LSTM-based models are particularly effective for sequential data such as time-series sensor readings or network logs. They can capture temporal dependencies and detect anomalies based on long-term correlations in network activity [15]. For example, LSTM-based IDS have achieved over 99% detection accuracy on recent IoT datasets, outperforming classical ML classifiers [16].

Autoencoders (AE) and Variational Autoencoders (VAE) are unsupervised architectures used to detect anomalies by reconstructing input features. Large reconstruction errors typically indicate abnormal traffic. Similarly, Generative Adversarial Networks (GAN) have been used to generate synthetic attack samples to balance datasets and improve IDS robustness [17].

Nevertheless, deep learning approaches are often criticized for their computational overhead and lack of interpretability. Running CNN or LSTM models on low-power IoT nodes remains infeasible without hardware acceleration. To mitigate this, lightweight deep learning frameworks and model compression techniques such as pruning, quantization, and knowledge distillation have been proposed [18].

3.3. Hybrid and Ensemble IDS

Hybrid IDS architectures combine the strengths of multiple AI paradigms to enhance detection accuracy, scalability, and adaptability. Typically, hybrid systems integrate ML and DL models to exploit both statistical and representational learning. For example, a CNN may serve as a feature extractor, while an SVM or RF classifier performs the final decision-making stage [19].

Hybrid models can also fuse anomaly-based and signature-based detection methods. The anomaly-based component identifies deviations from normal behavior, while the signature-based module validates whether the anomaly corresponds to a known attack pattern [20]. This layered defense mechanism significantly reduces false positives and improves detection of zero-day attacks.

Ensemble learning approaches, such as bagging, boosting, and stacking, are also prominent in IDS design. They combine multiple weak learners to build a strong classifier with improved generalization [21]. Gradient Boosting Machines (GBM) and XGBoost have been widely adopted for IoT IDS due to their robustness and interpretability [22].

Hybrid IDS have achieved remarkable results in simulation studies, yet they face challenges in real-time inference and resource optimization. Their deployment in constrained IoT nodes requires careful design trade-offs between accuracy, latency, and power consumption. Research

is currently exploring adaptive hybrid models capable of dynamically switching between detection strategies depending on device capability and network load [23].

3.4. Federated and Edge-based IDS

Traditional centralized IDS architectures often require transmitting raw IoT data to a cloud or data center for training and analysis. This raises serious concerns about data privacy, bandwidth consumption, and single points of failure. To overcome these issues, recent studies have introduced Federated Learning (FL) and Edge-based IDS frameworks [24][25][50]-[55][70]-[72]. In Federated Learning, models are trained collaboratively across distributed IoT devices without sharing raw data. Each node trains a local model on its private data and only transmits model updates (gradients) to a central server for aggregation [26]. This decentralized paradigm preserves privacy, reduces data transfer, and enables scalable intrusion detection across heterogeneous environments.

Edge-based IDS deploy lightweight detection models closer to the data sources — on gateways, routers, or fog nodes — to perform real-time threat analysis [27]. By leveraging edge computing resources, these systems reduce latency and enhance responsiveness, which are essential for time-critical IoT applications such as smart grids or autonomous vehicles.

Recent works propose Federated Deep Learning (FDL) architectures that combine CNN or LSTM models with federated aggregation, achieving comparable performance to centralized systems while maintaining privacy [28]. However, issues such as model poisoning attacks, communication overhead, and non-IID data distribution remain active research challenges [29].

The integration of blockchain technology with federated IDS has also been explored to enhance transparency and trust in collaborative learning. Blockchain-based FL can ensure tamper-proof audit trails for model updates and provide decentralized consensus mechanisms for anomaly alerts [30].

3.5. Comparative Summary

In order to consolidate the findings of the previous subsections, a comparative analysis was conducted across the four main AI paradigms. This synthesis, summarized in Table 2, highlights the strengths, limitations, and trade-offs associated with each approach in terms of accuracy, interpretability, computational efficiency, and deployment feasibility within IoT environments.

Table 2. Comparative summary of AI-based IDS approaches for IoT and sensor networks.

Category	**Example Techniques**	**Advantages**	**Limitations**
Machine Learning	SVM, RF, kNN, K-means	Simplicity, good interpretability	Manual feature engineering, poor scalability
Deep Learning	CNN, RNN, AE, GAN	High accuracy, automatic feature extraction	High computation cost, lack of explainability
Hybrid/Ensemble	CNN+SVM, RF+AE, XGBoost	Improved accuracy, reduced false positives	Complex design, real-time constraints
Federated/Edge	Federated CNN, Edge-LSTM	Privacy-preserving, scalable, low latency	Communication cost, non-IID challenges

This comparative overview illustrates the progressive evolution of AI-based IDS from static, centralized solutions toward adaptive, distributed, and privacy-preserving frameworks. While deep and hybrid models achieve superior detection performance, they require optimization for real-time operation and resource efficiency in constrained IoT devices. Federated and edge-based paradigms, although promising for scalability and privacy, still face challenges related to communication overhead, model synchronization, and heterogeneous data distribution.

3.6. Method-Oriented Synthesis of Reviewed Studies

A survey of AI-based IDS must compare the methods used in the literature rather than listing datasets only. Table 3 summarizes the main methodological families, their typical implementation choices, and their suitability for IoT and sensor-network deployment. Classical ML and survey-based IDS foundations are supported by [43]-[49] and [73]-[75], deep and anomaly-based IDS studies by [35]-[42], federated-learning studies by [50]-[55], explainable and adversarially robust IDS by [56]-[64], and lightweight edge-oriented approaches by [65]-[72].

Table 3. Synthesis of methods used in AI-based IDS for IoT and sensor networks.

Method family	**Typical techniques**	**Strengths**	**Weaknesses**	**Best deployment fit**
Classical supervised ML	SVM, RF, kNN, DT, Naive Bayes, XGBoost	Fast training, interpretable decisions, low inference cost	Manual feature engineering; weaker zero-day generalization	Gateways and moderate-resource edge nodes
Unsupervised / semi-supervised	K-means, SOM, isolation forest, one-class SVM, autoencoder anomaly scores	Useful when labels are scarce; can identify abnormal traffic	Sensitive to threshold choice and concept drift	Early-warning IDS and unlabeled IoT traffic
Deep learning	CNN, LSTM/GRU, DNN, AE/VAE, transformer variants	Automatic feature learning; strong performance on complex traffic	High memory/energy cost; limited explainability	Fog/cloud IDS or optimized edge models
Hybrid and ensemble	CNN+SVM, AE+RF, RF+XGBoost, stacking and boosting	Improves robustness and can reduce false positives	Complex pipeline and higher latency	Gateways where accuracy is prioritized
Federated learning	FedAvg, federated CNN/LSTM, client clustering, secure aggregation	Preserves raw-data privacy and supports distributed learning	Non-IID data, poisoning risk, communication overhead	Multi-site IoT/IIoT, healthcare, smart grid
Lightweight / TinyML / edge IDS	Pruning, quantization, knowledge distillation, compact CNN	Low latency and reduced energy consumption	Possible accuracy loss; hardware-specific tuning	Constrained sensors, routers, and embedded gateways

Explainable and robust IDS	SHAP, LIME, LRP, adversarial training, uncertainty estimation	Improves trust, auditability, and operator response	Adds computational cost and is rarely standardized	Critical infrastructure and regulated IoT domains

4. DATASETS, EVALUATION METRICS, AND EXPERIMENTAL ANALYSIS

The performance and generalization capabilities of AI-based Intrusion Detection Systems (IDS) in IoT and sensor networks largely depend on the datasets used for training and evaluation, as well as the choice of evaluation metrics. This section provides a detailed overview of commonly used datasets, standard performance metrics, and highlights experimental findings reported in recent studies. The discussion also addresses the limitations of existing datasets and methodological challenges in empirical evaluation.

4.1. Benchmark Datasets for IoT IDS

Several publicly available datasets are widely adopted in AI-based IDS research for IoT and sensor networks. These datasets differ in terms of attack coverage, traffic types, device heterogeneity, and feature sets, making them suitable for evaluating different AI paradigms. Table 4 summarizes the most commonly used datasets and their key characteristics, including widely used legacy benchmarks and newer IoT/IIoT datasets [10][11][31]-[34][49].

Table 4. Commonly Used Datasets in IoT and Sensor Network IDS Research.

Dataset	Year	Scope	Attack Types	Traffic Type	Remarks
NSL-KDD	2009	Traditional networks	DoS, Probe, U2R, R2L	TCP/IP flows	Improved version of KDD99; widely used but not IoT-specific [49]
CICIDS2017	2017	IoT & enterprise networks	DDoS, Brute-force, Infiltration	TCP/IP flows	Realistic attack scenarios; rich feature set [10]
BoT-IoT	2018	IoT networks	DoS, DDoS, reconnaissance	IoT protocol flows	Focused on IoT devices and heterogeneous traffic [32]
UNSW-NB15	2015	Hybrid network	DoS, Exploits, Reconnaissance, Shellcode	TCP/IP & application flows	Includes modern attacks; comprehensive features [11]
TON_IoT	2020	IoT/IIoT	DDoS, Data exfiltration, Malware	MQTT, Modbus, HTTP	Multi-protocol, IoT-specific [33]

Observations and limitations:

1. **Controlled environments:** Most datasets were generated in laboratory settings, which restricts the diversity, unpredictability, and realism of traffic patterns compared to real-world deployments.
2. **IoT-specific coverage:** Datasets such as BoT-IoT and TON_IoT introduce protocol-level diversity and IoT device heterogeneity, but their scale and variability remain limited.

3. **Continuous learning challenges:** There is a lack of labeled datasets designed for online learning or edge-deployed IDS, which hinders real-time evaluation, adaptation, and the development of incremental learning models.
4. **Feature representation diversity:** Some datasets focus primarily on network flows (e.g., NSL-KDD, CICIDS2017), while others include multi-protocol and application-level features (e.g., TON_IoT), highlighting the importance of selecting datasets aligned with the IDS methodology and deployment scenario.
5. **Benchmarking implications:** The choice of dataset significantly affects reported IDS performance, making cross-study comparisons challenging unless standard evaluation protocols are followed.

This analysis emphasizes the need for larger, more heterogeneous, and realistic IoT datasets, particularly for testing edge-based, federated, and adaptive AI models in practical IoT environments.

4.2. Evaluation Metrics

Evaluating the effectiveness of AI-based IDS requires a comprehensive set of metrics that capture detection accuracy, robustness, and operational efficiency. Both traditional classification measures and resource-oriented metrics are critical, especially in IoT environments where devices are often constrained. The commonly adopted evaluation metrics include:

1. **Accuracy (ACC):** Measures the proportion of correctly classified instances among all instances.
2. **Precision (P):**Indicates the proportion of correctly detected attacks among all instances classified as attacks.
3. **Recall (R) / Detection Rate (DR):**Represents the proportion of actual attacks correctly identified by the IDS.
4. **F1-Score:** The harmonic mean of precision and recall, balancing false positives and false negatives.
5. **Area Under Curve (AUC):**Evaluates the classifier's global performance across multiple decision thresholds, particularly useful for imbalanced datasets where attack instances are rare.
6. **Resource Metrics:**Essential for evaluating practical feasibility on IoT devices:
7. **Latency:** Time required to detect and respond to an intrusion.
8. **Memory and Computation Cost:** Measures the resource consumption of the IDS algorithm.
9. **Energy Efficiency:** Particularly relevant for battery-powered sensors and low-power IoT nodes [11][12].

Collectively, these metrics provide a holistic assessment of IDS performance, balancing detection effectiveness with deployment feasibility in resource-constrained IoT and sensor network environments.

4.3. Experimental Findings and Model Performance

Recent work on AI-based Intrusion Detection Systems (IDS) for Internet of Things (IoT) and sensor networks reports great detection capabilities and flexibility, but real-world implementation is constrained by computational and scalability factors. The top paradigms for AI—Machine Learning (ML), Deep Learning (DL), Hybrid, and Federated/Edge approaches—have different performance, interpretability, and resource usage trade-offs.

4.3.1. Machine Learning Models

SVM, Random Forest, and kNN are also widely used ML algorithms due to their simplicity and interpretability. They achieve 90–94% accuracy on benchmark datasets such as NSL-KDD and CICIDS2017 with precision and recall of more than 0.90. They use little computational resources and can be implemented in moderate-sized IoT nodes, while their interpretability assists in forensic analysis. But they don't generalize easily to IoT-specific protocols (e.g., MQTT, CoAP), perform poorly on zero-day attacks, and heavily depend on hand-engineered features, which limit their scalability and generalization. These ranges are consistent with the broader ML-based IDS literature and comparative survey findings [25][43]-[49][73]-[75].

4.3.2. Deep Learning Models

Deep learning methods like CNN, LSTM, and Autoencoders guarantee better accuracy (96–99%) on datasets such as BoT-IoT and TON_IoT. Autoencoders enhance anomaly detection on unlabelled data by reducing false positives by 15–20% compared to ML-based models. Their strength lies in automated feature learning and ability to capture spatio-temporal relationships in IoT traffic. Nevertheless, they require significant computational resources, are plagued by latency in real-time systems, and are not interpretable, which hurts trust and regulatory compliance. Related studies using deep neural networks, recurrent models, and autoencoder-based anomaly detection further support these trends [31][35]-[42].

4.3.3. Hybrid Models

Hybrid IDS combine deep learning and traditional ML techniques, e.g., CNN+SVM or RF+Autoencoder ensembles, with 95–98% accuracy and F1-scores of about 0.95 on CICIDS2017 and BoT-IoT. The systems improve robustness towards evolving and zero-day attacks and reduce false positives through multi-stage detection. Their major limitations are increased model complexity, increased inference latency, and increased computational and memory usage, which can jeopardize edge deployment feasibility.

4.3.4. Federated and Edge-based IDS

Federated and edge-based IDS systems provide low-latency distributed, privacy-preserving learning. Federated CNN or LSTM models provide 93–96% accuracy on TON_IoT, while lean edge CNNs decrease latency and power consumption by up to 40% compared to centralized models. They preserve data privacy and enable real-time local detection in safety-critical domains such as healthcare and smart grids. However, performance is degraded under non-IID data distributions, and communication overhead in model synchronizing can increase energy consumption as well as complicate large-scale deployment. These findings are also aligned with federated-learning, secure aggregation, and edge-computing research streams [50]-[55][70]-[72].

4.3.5. Comparative Analysis Across AI Paradigms

This section compares the main categories of AI-based Intrusion Detection Systems (IDS) for IoT and sensor networks based on recent benchmark datasets and performance indicators. To ensure a comprehensive understanding, both Table 5 and Figure 2 summarize the comparative performance across multiple dimensions, including detection accuracy, F1-score, latency, and energy efficiency.

While the table presents quantitative metrics extracted from representative studies, the radar chart visually emphasizes the relative trade-offs among these approaches.

Table 5. Comparative Summary of AI-based IDS Approaches.

IDS Type	Dataset	Accuracy	F1-Score	Latency	Energy Efficiency
ML(SVM,RF)	NSL-KDD	92%	0.91	Low	High
DL(CNN,LSTM)	BoT-IoT	98%	0.97	Medium	Low
Hybrid(CNN+SVM)	CICIDS2017	96%	0.95	Medium	Medium
Federated/Edge	TON_IoT	94%	0.93	Low	High

The radar chart clearly highlights that deep learning-based IDS (e.g., CNN, LSTM) outperform others in terms of detection accuracy and F1-score, making them ideal for complex IoT traffic classification. However, this comes at the cost of higher computational and energy demands, which may limit deployment on constrained devices. On the other hand, federated and edge-based IDS achieve a balanced trade-off between performance and resource efficiency, offering promising scalability and privacy preservation for distributed IoT environments. Hybrid models, while robust, introduce additional complexity that may affect inference speed.

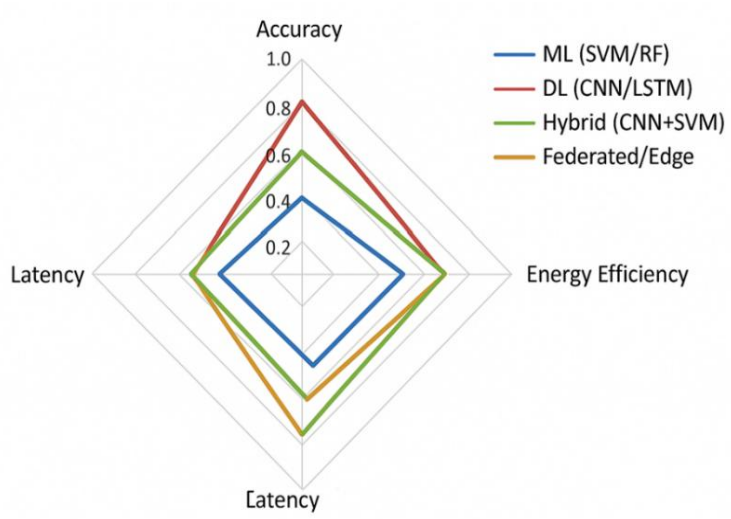

Figure 2. Comparative Performance Radar Chart

In summary, no single approach dominates all performance metrics. The choice of IDS architecture should therefore align with the **target IoT environment**, balancing detection accuracy, computational cost, and deployment constraints.

4.4. Quantitative Synthesis and Interpretation of Comparative Results

Because the reviewed studies use different datasets, splits, attack classes, and preprocessing pipelines, a full statistical meta-analysis is not always valid. Instead, this review performs a structured quantitative synthesis by grouping studies according to method family and by reporting typical performance ranges found in the evidence base. These values should be interpreted as indicative ranges, not as directly interchangeable benchmark scores. This synthesis

is therefore based on reported trends across method-specific groups rather than on a pooled effect-size estimate: benchmark and dataset studies [31]-[34], deep and anomaly-based IDS studies [35]-[42], ML and survey foundations [43]-[49], federated-learning studies [50]-[55], explainability and adversarial robustness studies [56]-[64], and lightweight edge/fog computing studies [65]-[72].

Table 6. Quantitative synthesis of reported performance trends by method family.

Method family	Typical accuracy range	Typical F1-score range	Latency/energy profile	Interpretation
Classical ML	90-95%	0.88-0.94	Low to medium cost	Strong baseline; useful when features are well engineered but less robust to unseen attacks.
Deep learning	96-99%	0.94-0.98	Medium to high cost	Best detection scores on benchmark datasets; requires compression or edge/fog support.
Hybrid/ensemble	95-98%	0.93-0.97	Medium cost	Balances accuracy and false-positive reduction; pipeline complexity must be justified.
Federated/edge	93-97%	0.90-0.96	Low latency locally; communication overhead during training	Promising for privacy and scalability; sensitive to non-IID data and poisoning attacks.
Explainable/robust IDS	Model-dependent	Model-dependent	Adds explanation or robustness overhead	Needed for trust and resilience, but still underreported in experiments.

The comparative tables and radar chart indicate that high accuracy alone is not sufficient for IoT IDS evaluation. A model with 98% accuracy but high latency or energy consumption may be less suitable than a 94% model that runs locally on a gateway and preserves privacy. Therefore, future comparisons should report at least four dimensions: detection effectiveness, cross-dataset generalization, resource consumption, and trustworthiness.

4.5. Critical Discussion

The findings presented in the previous sections highlight significant progress in AI-based IDS for IoT; however, several critical challenges persist. This section discusses the main limitations and open issues observed across the reviewed studies, focusing on dataset quality, evaluation methodology, performance trade-offs, and real-world applicability.

4.5.1. Methodological Biases in Primary Studies

The methodological limitations of the primary studies were grouped into four recurring bias categories: dataset bias, evaluation bias, reporting bias, and deployment bias. Dataset bias appears when models are trained and tested on laboratory traffic that does not reflect realistic IoT

heterogeneity. Evaluation bias appears when studies use a single split, no cross-dataset test, or accuracy-only reporting. Reporting bias appears when preprocessing, class balancing, hyperparameters, or hardware configuration are omitted. Deployment bias appears when models are proposed for IoT devices but tested only on desktop or cloud hardware. Recognizing these biases prevents overclaiming and helps explain why excellent benchmark performance may not translate into real-world IDS reliability.

Table 7. Risk-of-bias categories observed in IoT IDS studies.

Bias type	Common symptom	Mitigation recommended
Dataset bias	One benchmark, limited devices, synthetic traffic	Use multi-protocol datasets and cross-dataset validation.
Evaluation bias	Accuracy-only results or no baseline	Report precision, recall, F1, AUC, false-alarm rate, latency, and energy.
Reporting bias	Missing preprocessing or hyperparameters	Publish feature list, split strategy, balancing method, and code when possible.
Deployment bias	IoT claims without edge-device testing	Measure inference time, memory, and power on representative hardware.

1. **Dataset Limitations:** Many studies rely on benchmark datasets that do not fully capture IoT heterogeneity, limiting the external validity of results [19].
2. **Evaluation Inconsistency:** Different works employ varied feature sets and pre-processing methods, making direct comparison challenging.
3. **Trade-offs Between Accuracy and Efficiency:** High-accuracy DL models may not be practical for constrained IoT devices; federated and edge solutions balance privacy, latency, and energy constraints [20].
4. **Need for Real-world Evaluation:** Field deployment in live IoT environments remains limited. Future research should focus on realistic testbeds and pilot deployments to validate model scalability and adaptability [21].

In summary, while AI-driven IDS approaches show great promise, achieving scalable, explainable, and energy-aware intrusion detection in heterogeneous IoT environments remains an open challenge. Addressing these gaps requires standardized datasets, consistent evaluation frameworks, and real-world experimentation to move from theoretical efficiency to practical reliability.

4.6. Recommendations for Experimental Design

Building upon the critical issues discussed in the previous section, it becomes clear that improving experimental design is crucial to ensure the reliability, comparability, and real-world applicability of AI-based Intrusion Detection Systems (IDS) for IoT environments. The following recommendations aim to guide future research toward more robust and generalizable evaluation methodologies.

1. Incorporate heterogeneous IoT traffic and multi-protocol datasets for training and validation.
2. Employ cross-dataset evaluation to assess model generalization.
3. Utilize multi-metric evaluation frameworks combining accuracy, F1-score, latency, and energy metrics.

4. Explore transfer learning and continual learning approaches to adapt models to evolving attack patterns.
5. Integrate privacy-preserving methods (e.g., federated learning, homomorphic encryption) for collaborative model training.

In conclusion, adopting these recommendations will contribute to the development of standardized, scalable, and privacy-aware experimental frameworks. Such frameworks are essential for ensuring that AI-based IDS solutions can be efficiently validated and effectively deployed in real-world IoT and sensor network environments.

6. CHALLENGES, OPEN ISSUES, AND FUTURE DIRECTIONS

Intrusion Detection Systems (IDS) for IoT and sensor networks based on AI have developed significantly, but challenges discourage their large-scale deployment and operational reliability. Key issues are limited data realism, computational complexity, scalability and explainability, and privacy and security concerns [2,3,4]. This subsection lists these challenges and outlines research directions for next-generation intelligent, reliable, and efficient IoT IDS.

6.1. Data and Computational Challenges

Effective IDS require high-quality, realistic, and diverse datasets. Existing benchmarks such as NSL-KDD, CICIDS2017, BoT-IoT, UNSW-NB15, N-BaIoT, Edge-IIoTset, and TON_IoT differ substantially in protocol coverage, attack scenarios, labeling quality, and class distribution. This heterogeneity makes direct comparison difficult and can inflate reported performance when models are evaluated only on a single benchmark [10][11][31]-[34][49]. Privacy constraints also restrict access to real operational traffic, while class imbalance can bias models toward majority attack categories or normal traffic.

At the computational level, deep models such as CNNs, LSTMs, transformers, and autoencoders can exceed the memory, processing, and energy budgets of constrained IoT nodes. Future work should therefore combine realistic multi-protocol datasets, cross-dataset validation, lightweight model compression, TinyML-oriented optimization, and edge/fog deployment tests to verify practical feasibility [16,18,27,29].

6.2. Adaptability, Scalability, and Explainability

IoT networks are highly dynamic with frequent device churn and fluctuating behaviors. Static IDS suffer from concept drift, while centralized systems limit scalability [7,19]. Deep learning "black-box" models also reduce trust and make integration difficult in mission-critical environments [22,23].

Guidance for research includes online and lifelong learning [17,20], context-aware intrusion response reinforcement learning, and Explainable AI techniques such as SHAP or LRP for explaining decisions and facilitating human-in-the-loop workflows [22,23,26][56]-[59]. Hybrid symbolic–deep approaches can combine high detection performance with interpretability.

6.3. Privacy, Security, and Emerging Directions

IDS rely on sensitive data and are therefore vulnerable to privacy breach, data poisoning, and adversarial attacks [2,3,5][60]-[64]. Federated learning partially mitigates these vulnerabilities but is also vulnerable to inversion attacks and auditability attacks [20,24][50]-[55]. The use of

blockchain ensures secure, tamper-evident model aggregation and policy enforcement automation [24,28,30].

New directions involve neurosymbolic and quantum-aided IDS, energy-efficient "Green AI" models for constrained devices [16][65]-[69], and self-healing architectures that can maintain defense under attack. Convergence of AI, blockchain, and quantum technologies promises proactive, smart, and adaptive IDS.

In summary, AI-based IDS for IoT still suffer from realistic data issues, computational complexity, scalability, transparency, and privacy. Overcoming these challenges requires concerted, multi-disciplinary innovation that combines machine learning, distributed systems, and fault-tolerant computing. Next-generation IDS will have to evolve towards lightweight, interpretable, and self-healing architectures that are able to provide high detection accuracy and strong ethical and operational assurances in the advanced IoT environment [2,17,22,27,30].

6.4. Conceptual Framework for Next-Generation IoT IDS

To strengthen the originality of this review, a layered conceptual framework is proposed for next-generation AI-based IDS in IoT and sensor network environments. The framework organizes the IDS pipeline into five interconnected layers: data acquisition, preparation, learning, deployment, and governance. This layered view helps clarify how technical choices at each stage influence detection performance, scalability, privacy, energy efficiency, and operational trust. It reflects recurring design choices from IoT dataset studies [31]-[34], DL-based IDS models [35]-[42], federated and privacy-preserving learning [50]-[55], explainable and robust AI [56]-[64], and lightweight edge/fog deployment research [65]-[72].

The first layer concerns sensing and traffic acquisition from heterogeneous IoT devices, gateways, and communication protocols. The second layer focuses on preprocessing, feature selection, normalization, windowing, and class-imbalance handling. The third layer includes learning models used for detection, such as ML, DL, hybrid, federated, explainable, and adversarially robust models. The fourth layer defines the deployment environment, which may involve device-level, edge, fog, cloud, or collaborative federated settings. Finally, the governance layer ensures that IDS performance is assessed not only through accuracy but also through latency, energy consumption, privacy, robustness, explainability, and reproducibility. Table 8 presents the proposed layered conceptual framework for AI-based IDS in IoT and sensor networks.

Table 8. Proposed layered conceptual framework for AI-based IDS in IoT.

Framework layer	Main components	Design implication
Data layer	IoT devices, sensors, gateways, protocols such as MQTT/CoAP/HTTP/Modbus	Traffic must reflect heterogeneous devices and realistic attacks.
Preparation layer	Cleaning, normalization, feature selection, balancing, windowing	Preprocessing choices must be reported to improve reproducibility.
Learning layer	ML, DL, hybrid, FL, XAI, adversarially robust models	Method selection should match dataset size, label quality, and deployment constraints.
Deployment layer	Device, edge, fog, cloud, collaborative FL clients	Placement determines latency, privacy exposure, and energy cost.

Governance layer	Metrics, risk-of-bias checks, explainability, audit logs, privacy controls	Trustworthy IDS requires more than detection accuracy.

6.5. Proposed Benchmarking Protocol

A standardized benchmarking protocol is needed to improve the comparability, reproducibility, and deployment relevance of future IDS studies for IoT and sensor networks. Existing studies often report strong detection accuracy, but they differ substantially in threat models, datasets, preprocessing steps, validation strategies, baseline comparisons, and resource measurements [43]-[49]. These differences make it difficult to determine whether one IDS approach is truly superior to another.

To address this limitation, this review proposes a minimum benchmarking protocol for future IoT IDS studies. The protocol includes a clearly stated threat model, the use of at least one IoT-specific dataset, cross-dataset or temporal validation where feasible, transparent preprocessing, comparison with representative ML and DL baselines, and multi-dimensional performance reporting. In addition to detection metrics, studies should report deployment-oriented indicators such as latency, memory usage, energy consumption, and communication overhead. Robustness should also be evaluated against class imbalance, concept drift, poisoning attacks, and adversarial manipulation. Table 9 summarizes the proposed standardized benchmarking protocol and the minimum requirements recommended for future IoT IDS studies.

Table 9. Proposed standardized benchmarking protocol for IoT IDS studies.

Benchmarking step	**Minimum requirement**
Threat model	Specify attack types, attacker capability, and protected IoT layer.
Datasets	Use at least one IoT-specific benchmark and document all preprocessing.
Validation	Use stratified split plus cross-dataset, temporal, or device-wise validation when feasible.
Baselines	Compare against at least one classical ML, one DL, and one lightweight/edge baseline.
Metrics	Report detection, false-alarm, latency, memory, energy, and communication cost.
Reproducibility	Disclose hyperparameters, feature lists, hardware, and random seeds.
Trustworthiness	Assess explainability, privacy leakage, poisoning resistance, and adversarial robustness.

7. CONCLUSION AND RESEARCH OUTLOOK

This systematic literature review (SLR) has comprehensively analyzed Artificial Intelligence (AI)-based Intrusion Detection Systems (IDS) for the Internet of Things (IoT) and sensor networks, covering research conducted between 2015 and early 2025. Through the synthesis of 75 peer-reviewed studies, the review provided an integrated understanding of existing methodologies, datasets, evaluation metrics, and persistent challenges.

Overall, the evolution of IDS demonstrates a clear transition from traditional machine learning (ML) approaches to more sophisticated deep learning (DL), hybrid, and federated learning paradigms, reflecting the growing complexity and decentralization of IoT ecosystems.

7.1. Summary of Key Insights

The review reveals several key findings regarding the development of AI-driven IDS technologies.
First, Machine Learning (ML) techniques such as Support Vector Machines (SVM), Random Forest (RF), and k-Nearest Neighbors (kNN) remain widely used due to their simplicity, interpretability, and low computational cost [6][7]. However, these methods often rely heavily on manual feature engineering and perform suboptimally in large-scale, high-dimensional environments.

Second, Deep Learning (DL) models—particularly Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Autoencoders (AE)—have become dominant in recent years, offering superior accuracy and automatic feature extraction capabilities [12][14]. Yet, their high computational and energy requirements limit deployment on resource-constrained IoT devices.

Third, Hybrid and Ensemble IDS approaches that combine multiple learning paradigms (e.g., CNN+SVM, RF+AE, XGBoost) exhibit improved robustness and reduced false-positive rates but suffer from increased model complexity and latency [16].

Finally, Federated and Edge-based IDS have emerged as promising paradigms that balance detection performance with privacy and scalability by enabling decentralized model training [17][18].

Regarding evaluation, the most common datasets—such as NSL-KDD, CICIDS2017, BoT-IoT, UNSW-NB15, and TON_IoT—remain fundamental benchmarks. However, they often lack realism, heterogeneity, and IoT-specific traffic diversity [10][13]. Evaluation metrics including accuracy, precision, recall, F1-score, AUC, and latency remain the standard indicators of IDS performance.

Despite these advancements, major challenges persist, notably data scarcity, resource limitations, scalability issues, lack of explainability, and security/privacy concerns [19][22][24]. Edge and federated learning mitigate some of these problems but introduce new complexities such as non-independent and identically distributed (non-IID) data and communication overhead.

7.2. Future Research Directions

The future of AI-based IDS in IoT and sensor networks will be shaped by the convergence of intelligent learning, energy optimization, and privacy preservation.

A primary direction involves the development of lightweight and energy-efficient AI models, leveraging TinyML, pruning, quantization, and model compression to enable real-time detection on constrained devices [18].

Moreover, continual and online learning frameworks are necessary to cope with evolving attack patterns and concept drift in dynamic IoT environments [20].

A second major axis is the integration of Explainable AI (XAI) to enhance model transparency, trust, and accountability, particularly in safety-critical domains such as healthcare, industrial IoT, and autonomous systems [22].

Similarly, privacy-preserving federated learning will play a vital role, supported by blockchain technologies, differential privacy, and secure multi-party computation (SMPC), to enable trustworthy and decentralized collaboration across heterogeneous IoT infrastructures [23][24].

Finally, future innovations should arise from cross-disciplinary research, combining AI with neurosymbolic reasoning, quantum machine learning (QML), and self-healing architectures. These directions can enhance robustness, interpretability, and adaptability, while real-world deployments and large-scale IoT testbeds remain essential for validating scalability and performance under realistic conditions.

7.3. Concluding Remarks

AI-based IDS have shown strong potential to secure IoT ecosystems through intelligent and adaptive threat detection. While deep learning and hybrid models achieve high detection performance, and federated and edge-based solutions address privacy and scalability, real-world deployment remains challenged by device heterogeneity, limited resources, and the continuous evolution of threats.

In addition to synthesizing existing IDS approaches, this review contributes a method-oriented taxonomy, a risk-of-bias analysis, a layered conceptual framework, and a standardized benchmarking protocol to guide future IoT IDS research.

The future of IDS will rely on balancing efficiency, adaptability, and trustworthiness. This requires lightweight and energy-efficient AI suitable for constrained devices, explainable models to support human decision-making, and privacy-preserving mechanisms such as federated learning and secure computation. Autonomous and adaptive architectures, integrating continual and reinforcement learning, will allow IDS to respond dynamically to emerging attacks. Furthermore, integrating technologies like blockchain, digital twins, and edge-cloud collaboration can enhance transparency, realism, and responsiveness. Standardized evaluation frameworks with heterogeneous datasets and multi-metric benchmarks will also be crucial to ensure robust and reproducible research.By advancing along these directions, next-generation IDS can become resilient, intelligent, and privacy-aware, capable of protecting increasingly complex IoT infrastructures against sophisticated cyberattacks.

REFERENCES

[1] K. Ashton, “That ‘Internet of Things’ thing,” *RFID Journal*, vol. 22, no. 7, pp. 97–114, 2015.
[2] M. Abomhara and G. M. Køien, “Security and privacy in the Internet of Things: Current status and open issues,” *Computers & Electrical Engineering*, vol. 67, pp. 581–594, 2018.
[3] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, “Security, privacy and trust in Internet of Things: The road ahead,” *Computer Networks*, vol. 76, pp. 146–164, 2015.
[4] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, “A survey on security and privacy issues in Internet-of-Things,” *IEEE Internet of Things Journal*, vol. 4, no. 5, pp. 1250–1258, Oct. 2017.
[5] A. Kolias, G. Kambourakis, A. Stavrou, and J. Voas, “DDoS in the IoT: Mirai and other botnets,” *Computer*, vol. 50, no. 7, pp. 80–84, Jul. 2017.
[6] M. Ring, S. Wunderlich, D. Grüdl, D. Landes, and A. Hotho, “Flow-based network traffic generation using generative adversarial networks,” *Computers & Security*, vol. 82, pp. 156–172, 2019.
[7] M. Zolanvari, M. A. Teixeira, L. Jain, R. Khan, and R. Jain, “Machine learning-based network vulnerability analysis of industrial Internet of Things,” *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6822–6834, Aug. 2019.
[8] D. Moher et al., “Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA 2020 statement,” *PLOS Medicine*, vol. 18, no. 3, pp. 1–15, 2021.

[9] B. Kitchenham and S. Charters, *Guidelines for Performing Systematic Literature Reviews in Software Engineering*, Keele University, 2007.
[10] I. Sharafaldin, A. Habibi Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *Proc. ICISSP*, 2018, pp. 108–116.
[11] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in *Proc. MilCIS*, 2015, pp. 1–6.
[12] M. Almiani, A. Alkasassbeh, M. M. Alauthman, and M. Dorgham, "Deep recurrent neural network for IoT intrusion detection system," *Simulation Modelling Practice and Theory*, vol. 101, p. 102031, 2020.
[13] M. H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, "Network anomaly detection: Methods, systems and tools," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 303–336, 2014.
[14] A. Ferrag, L. Maglaras, S. Moschoyiannis, and H. Janicke, "Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study," *Journal of Information Security and Applications*, vol. 50, p. 102419, 2020.
[15] A. Sultana, N. Chilamkurti, and A. Mahmood, "A new lightweight intrusion detection framework for wireless sensor networks," *Sensors*, vol. 19, no. 18, p. 4069, 2019.
[16] M. H. S. Hossain, M. Karim, and R. Hasan, "TinyML meets IoT security: A review," *IEEE Access*, vol. 10, pp. 78134–78153, 2022.
[17] T. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, and A.-R. Sadeghi, "DÏoT: A federated self-learning anomaly detection system for IoT," in *Proc. IEEE ICDCS*, 2019, pp. 756–767.
[18] Z. Zhao, J. Chen, R. Xie, and F. Liu, "Lightweight deep learning models for intrusion detection in IoT networks," *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 13423–13437, Aug. 2022.
[19] H. Hindy, D. Brosset, E. Bayne, A. Seeam, and C. Maple, "A taxonomy and survey of network-based intrusion detection systems for IoT," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 593–623, 2022.
[20] H. T. Nguyen and M. S. Azad, "Adaptive federated learning for network intrusion detection in IoT," *IEEE Access*, vol. 10, pp. 111234–111246, 2022.
[21] S. Otoum, I. A. Ridhawi, and H. Mouftah, "A cyber-physical security framework for IoT-enabled smart grids," *IEEE Access*, vol. 8, pp. 208911–208923, 2020.
[22] F. Ahmad, M. Shahid, and A. Jamal, "Explainable AI for intrusion detection in IoT networks: A survey," *IEEE Access*, vol. 11, pp. 40112–40135, 2023.
[23] R. Doriguzzi-Corin, A. Millar, and S. Scott-Hayward, "A survey of XAI in network intrusion detection systems," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–37, 2024.
[24] S. K. Lo, X. Xu, Y. Wang, Q. Lin, and M. K. Y. Leung, "Blockchain-enabled federated learning for trustworthy IoT systems," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 2112–2122, Mar. 2022.
[25] R. Doshi, N. Apthorpe, and N. Feamster, "Machine learning DDoS detection for consumer Internet of Things devices," in *Proc. IEEE SPW*, 2018, pp. 29–35.
[26] M. Hussain, F. Hussain, S. Hassan, and E. Hossain, "Explainable AI and federated learning based intrusion detection for IoT networks," *IEEE Internet of Things Journal*, vol. 11, no. 2, pp. 1760–1775, Jan. 2025.
[27] A. Alsaeedi and M. Khan, "Edge-based intrusion detection system for IoT networks using lightweight deep learning models," *Sensors*, vol. 23, no. 5, p. 2598, 2023.
[28] T. Salman, S. Zolanvari, R. Jain, A. Erbad, and M. Samaka, "Security services using blockchains: A state-of-the-art survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 1, pp. 858–880, 2019.
[29] A. Abeshu and N. Chilamkurti, "Deep learning: The frontier for distributed attack detection in fog-to-things computing," *IEEE Communications Magazine*, vol. 56, no. 2, pp. 169–175, Feb. 2018.
[30] M. S. Haghighat and J. C. Chatzidakis, "Blockchain-assisted federated learning for intrusion detection in IoT," *Future Generation Computer Systems*, vol. 152, pp. 659–675, 2025.
[31] Y. Meidan, M. Bohadana, A. Shabtai, M. Ochoa, N. O. Tippenhauer, J. D. Guarnizo, and Y. Elovici, "N-BaIoT: Network-based detection of IoT botnet attacks using deep autoencoders," IEEE Pervasive Computing, vol. 17, no. 3, pp. 12–22, 2018.
[32] N. Koroniotis, N. Moustafa, E. Sitnikova, and B. Turnbull, "Towards the development of realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset," Future Generation Computer Systems, vol. 100, pp. 779–796, 2019.

[33] N. Moustafa, “TON_IoT datasets: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems,” IEEE Access, vol. 8, pp. 165130–165150, 2020.
[34] M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, “Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications,” IEEE Access, vol. 10, pp. 40281–40306, 2022.
[35] E. Hodo, X. Bellekens, A. Hamilton, C. Tachtatzis, and R. Atkinson, “Shallow and deep networks intrusion detection system: A taxonomy and survey,” arXiv preprint arXiv:1701.02145, 2017.
[36] 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.
[37] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, “A deep learning approach to network intrusion detection,” IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 2, no. 1, pp. 41–50, 2018.
[38] C. Yin, Y. Zhu, J. Fei, and X. He, “A deep learning approach for intrusion detection using recurrent neural networks,” IEEE Access, vol. 5, pp. 21954–21961, 2017.
[39] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, “A deep learning approach for network intrusion detection system,” in Proc. EAI International Conference on Bio-inspired Information and Communications Technologies, 2016, pp. 21–26.
[40] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, “Kitsune: An ensemble of autoencoders for online network intrusion detection,” in Proc. Network and Distributed System Security Symposium (NDSS), 2018.
[41] M. Lopez-Martin, B. Carro, A. Sanchez-Esguevillas, and J. Lloret, “Conditional variational autoencoder for prediction and feature recovery applied to intrusion detection in IoT,” Sensors, vol. 17, no. 9, p. 1967, 2017.
[42] K. Alrawashdeh and C. Purdy, “Toward an online anomaly intrusion detection system based on deep learning,” in Proc. International Conference on Machine Learning and Data Mining in Pattern Recognition, 2016, pp. 195–200.
[43] I. Ullah and Q. H. Mahmoud, “A survey of intrusion detection systems for IoT,” Journal of Network and Computer Applications, vol. 150, p. 102497, 2020.
[44] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, “Survey of intrusion detection systems: Techniques, datasets and challenges,” Cybersecurity, vol. 2, no. 1, pp. 1–22, 2019.
[45] A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2016.
[46] P. Garcia-Teodoro, J. Diaz-Verdejo, G. Macia-Fernandez, and E. Vazquez, “Anomaly-based network intrusion detection: Techniques, systems and challenges,” Computers & Security, vol. 28, no. 1–2, pp. 18–28, 2009.
[47] R. Sommer and V. Paxson, “Outside the closed world: On using machine learning for network intrusion detection,” in Proc. IEEE Symposium on Security and Privacy, 2010, pp. 305–316.
[48] R. P. Lippmann, D. J. Fried, I. Graf, J. W. Haines, K. R. Kendall, D. McClung, D. Weber, S. E. Webster, D. Wyschogrod, R. K. Cunningham, and M. A. Zissman, “Evaluating intrusion detection systems: The 1998 DARPA off-line intrusion detection evaluation,” in Proc. DARPA Information Survivability Conference and Exposition, 2000, pp. 12–26.
[49] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications, 2009, pp. 1–6.
[50] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in Proc. AISTATS, 2017, pp. 1273–1282.
[51] K. Bonawitz et al., “Practical secure aggregation for privacy-preserving machine learning,” in Proc. ACM CCS, 2017, pp. 1175–1191.
[52] P. Kairouz et al., “Advances and open problems in federated learning,” Foundations and Trends in Machine Learning, vol. 14, no. 1–2, pp. 1–210, 2021.
[53] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. V. Poor, “Federated learning for Internet of Things: A comprehensive survey,” IEEE Communications Surveys & Tutorials, vol. 23, no. 3, pp. 1622–1658, 2021.
[54] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50–60, 2020.

[55] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, “Federated optimization in heterogeneous networks,” in Proc. MLSys, 2020, pp. 429–450.
[56] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in Proc. ACM SIGKDD, 2016, pp. 1135–1144.
[57] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in Proc. Advances in Neural Information Processing Systems, 2017, pp. 4765–4774.
[58] G. Montavon, W. Samek, and K.-R. Muller, “Methods for interpreting and understanding deep neural networks,” Digital Signal Processing, vol. 73, pp. 1–15, 2018.
[59] W. Samek, G. Montavon, A. Vedaldi, L. K. Hansen, and K.-R. Muller, Eds., Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Cham, Switzerland: Springer, 2019.
[60] I. J. Goodfellow, J. Shlens, and C. Szegedy, “Explaining and harnessing adversarial examples,” in Proc. International Conference on Learning Representations, 2015.
[61] N. Carlini and D. Wagner, “Towards evaluating the robustness of neural networks,” in Proc. IEEE Symposium on Security and Privacy, 2017, pp. 39–57.
[62] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, “Practical black-box attacks against machine learning,” in Proc. ACM ASIA CCS, 2017, pp. 506–519.
[63] B. Biggio and F. Roli, “Wild patterns: Ten years after the rise of adversarial machine learning,” Pattern Recognition, vol. 84, pp. 317–331, 2018.
[64] A. Kurakin, I. Goodfellow, and S. Bengio, “Adversarial examples in the physical world,” in Proc. International Conference on Learning Representations Workshop, 2017.
[65] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted residuals and linear bottlenecks,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 4510–4520.
[66] S. Han, H. Mao, and W. J. Dally, “Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding,” in Proc. International Conference on Learning Representations, 2016.
[67] B. Jacob et al., “Quantization and training of neural networks for efficient integer-arithmetic-only inference,” in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 2704–2713.
[68] P. Warden and D. Situnayake, TinyML: Machine Learning with TensorFlow Lite on Arduino and Ultra-Low-Power Microcontrollers. Sebastopol, CA, USA: O’Reilly Media, 2019.
[69] N. D. Lane, S. Bhattacharya, P. Georgiev, C. Forlivesi, and F. Kawsar, “DeepX: A software accelerator for low-power deep learning inference on mobile devices,” in Proc. ACM/IEEE IPSN, 2016, pp. 1–12.
[70] M. Satyanarayanan, “The emergence of edge computing,” Computer, vol. 50, no. 1, pp. 30–39, 2017.
[71] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, “Edge computing: Vision and challenges,” IEEE Internet of Things Journal, vol. 3, no. 5, pp. 637–646, 2016.
[72] M. Chiang and T. Zhang, “Fog and IoT: An overview of research opportunities,” IEEE Internet of Things Journal, vol. 3, no. 6, pp. 854–864, 2016.
[73] M. Miettinen et al., “IoT SENTINEL: Automated device-type identification for security enforcement in IoT,” in Proc. IEEE ICDCS, 2017, pp. 2177–2184.
[74] A. Anthi, L. Williams, M. Słowińska, G. Theodorakopoulos, and P. Burnap, “A supervised intrusion detection system for smart home IoT devices,” IEEE Internet of Things Journal, vol. 6, no. 5, pp. 9042–9053, 2019.
[75] S. Garcia, M. Grill, J. Stiborek, and A. Zunino, “An empirical comparison of botnet detection methods,” Computers & Security, vol. 45, pp. 100–123, 2014.