Vokasi Unesa Bulletin of Engineering, Technology and Applied Science (VUBETA)
https://journal.unesa.ac.id/index.php/vubeta
Vol. 3, No. 3, 2026, pp. 445~460 DOI: 10.26740/vubeta.v3i3.46474
ISSN: 3064-0768

Privacy-Enhanced Federated Learning Model for Secure Internet of Things Applications: A Systematic Review

Mohammed Ajuji[1*], Yusuf Musa Malgwi[2], Asabe Sandra Ahmadu[3]
[1]Department of Computer Science, Faculty of Science, Gombe State University, Gombe State, Nigeria
[2,3]Department of Computer Science, Faculty of Computing, Modibbo Adama University, Adamawa State, Nigeria

Article Info

Article history:

Received January 05, 2026
Revised March 05, 2026
Accepted June 10, 2026

Keywords:

Federated Learning
Internet of Things
Privacy Enhancing
Cybersecurity
Edge Intelligence

ABSTRACT

The rapid growth of Internet of Things (IoT) ecosystems has increased security and privacy risks due to the large volume of sensitive and distributed data generated by connected devices. Traditional machine learning approaches for IoT security rely on centralized data aggregation, which raises privacy concerns, scalability limitations, and regulatory challenges. This study presents a systematic literature review of privacy-enhanced federated learning (PEFL) for secure IoT applications. The review analyzes federated learning architectures, optimization strategies, privacy-preserving mechanisms, and cybersecurity use cases in heterogeneous IoT environments. Findings indicate that federated learning improves privacy by keeping data localized at edge devices, while techniques such as differential privacy and secure aggregation enhance confidentiality. However, these methods introduce trade-offs in accuracy, communication efficiency, and computational overhead. Hierarchical and adaptive FL architectures improve scalability and robustness, particularly for intrusion detection systems. Persistent challenges include non-IID data, device heterogeneity, limited edge resources, and the lack of standardized benchmarks and real-world deployments. PEFL offers a promising framework for securing IoT systems, though further research is needed to address deployment and performance constraints in practical environments.

1. INTRODUCTION

The speedy growth of the Internet of Things (IoT) has fundamentally transformed critical sectors such as healthcare, smart cities, industrial automation, and intelligent transportation systems by enabling pervasive sensing, large-scale data collection, and data-driven decision-making [1][2]. Despite these benefits, IoT expansion has significantly intensified security and privacy challenges because IoT devices are distributed, heterogeneous, and resource-constrained [1][3]. IoT risk can be broadly classified into device-level, network-level, application-level, and cloud-level risks, each mapping to distinct attack surfaces across the IoT stack. Device-level risks arise from insecure firmware, weak authentication, and hardware tampering, which adversaries exploit to gain local control [4][39], while network-level risks such as eavesdropping, spoofing, and distributed denial-of-service (DDoS) attacks target communication channels between devices, edge, and cloud nodes [5][31][43]. Application- and cloud-level risks involve exploiting APIs, misconfigured services, and centralized data stores, leading to data breaches and integrity loss and underscoring the need for layered, context-aware security controls throughout the IoT ecosystem [6][7].

*Corresponding Author
Email: majuji@gsu.edu.ng

Many IoT nodes operate with limited computational power, memory, and energy, making them highly susceptible to cyber threats such as malware injection, spoofing, man-in-the-middle attacks, data breaches, and denial-of-service attacks [4][8]. Moreover, IoT systems continuously generate sensitive data, including personal identifiers, medical records, behavioral patterns, and operational telemetry, raising serious concerns about data confidentiality, integrity, availability, and compliance with data protection regulations such as GDPR and HIPAA [1][5].

To address these security concerns, machine learning (ML) techniques have been widely employed for intrusion detection, anomaly detection, and threat classification in IoT networks. Traditional IoT security solutions predominantly rely on centralized ML architectures, where raw data collected from distributed IoT devices is transmitted to a central server for training and analysis. While centralized ML can achieve high detection accuracy in controlled environments, it introduces several limitations when deployed in large-scale, real-world IoT systems. Centralized data aggregation significantly increases the risk of privacy leakage, creates single points of failure, incurs high communication overhead, and often conflicts with regulatory requirements that restrict cross-border or third-party data sharing [3][9]. Furthermore, centralized models struggle to scale efficiently and adapt to highly dynamic IoT environments characterized by device churn, evolving attack patterns, and non-stationary data distributions [10][11].

Federated learning (FL) has emerged as a promising decentralized alternative that enables collaborative model training across distributed clients without requiring raw data to leave local devices [12]. In FL, participating IoT nodes train models locally using their private data and share only model updates or gradients with a coordinating server for aggregation [12][13]. This paradigm reduces communication costs, preserves data locality, and mitigates several privacy risks associated with centralized learning [14]. As a result, FL has been increasingly applied to IoT security applications, including distributed intrusion detection systems, malware detection, and anomaly detection at the network and host levels [6][15][16]. However, recent studies have demonstrated that standard FL implementations remain vulnerable to a variety of security and privacy threats, including gradient leakage attacks, membership inference, model poisoning, Byzantine behavior, and backdoor injections [17]-[19]. However, recent studies have demonstrated that standard federated learning (FL) implementations remain vulnerable to several security and privacy threats. These threats are typically analyzed under a semi-honest (honest-but-curious) or malicious adversarial threat model, depending on whether the attacker follows the protocol correctly while attempting to extract information or actively manipulates the training process [18][22].

In gradient leakage attacks, the adversary exploits shared gradients or model updates to reconstruct sensitive training data from a client. Although raw data is never transmitted in FL, gradients can encode sufficient information to enable partial or even near-exact recovery of input samples, particularly in deep neural networks. Under a semi-honest server model, an honest-but-curious central aggregator can analyze received gradients to perform reconstruction attacks, thereby violating data confidentiality [18]. In cross-device FL, this risk is amplified when batch sizes are small or when model updates are shared without encryption or secure aggregation.

Inference attacks, including membership inference and property inference attacks, operate by analyzing model parameters or outputs to determine whether a specific data record was part of a client's training dataset or to infer sensitive attributes about that dataset. In membership inference attacks, the adversary queries the trained global model and exploits confidence scores or prediction behavior to infer data participation. In property inference attacks, the adversary analyzes aggregated updates to deduce global patterns, such as whether a hospital dataset contains patients with a specific condition. These attacks are particularly feasible under centralized aggregation without formal privacy guarantees such as differential privacy [51].

Beyond passive threats, model poisoning attacks assume a malicious client adversary that can submit manipulated model updates to bias the global model or embed backdoors [8]. In heterogeneous IoT environments, where client verification is weak and data is non-IID, distinguishing malicious updates from benign statistical divergence becomes challenging [24].

In general, the FL threat model typically considers three adversarial positions: a malicious or curious server; malicious clients; and external eavesdroppers intercepting communications. Mitigating gradient leakage and inference attacks requires integrating secure aggregation, differential privacy, encryption mechanisms, and robust aggregation strategies into the FL pipeline [17][51]. Without such protections, the assumption that "data never leaves the device" does not guarantee privacy in federated IoT systems. Additionally, non-independent and identically distributed (non-IID) data across heterogeneous IoT devices can severely degrade model convergence, accuracy, and generalization performance [10][20].

To overcome these limitations, privacy-enhanced federated learning has gained increasing attention in recent literature. Privacy-enhanced FL integrates advanced privacy-preserving mechanisms such as differential privacy, secure aggregation, homomorphic encryption, and secure multiparty computation to strengthen confidentiality guarantees throughout the federated learning lifecycle [21][24]. Differential privacy introduces

carefully calibrated noise to model updates to limit information leakage, while secure aggregation protocols ensure that individual client updates remain hidden even from the aggregation server [21][25]. Cryptographic techniques, including homomorphic encryption and secure multiparty computation, further enable computation on encrypted data, albeit with additional computational and communication overhead [23][26]. These techniques are particularly critical in IoT environments, where devices exhibit diverse capabilities, operate under strict energy and bandwidth constraints, and interact continuously with adversarial environments [27][28].

Privacy-preserving solutions in federated learning can be classified into three main categories: differential privacy (DP), which adds calibrated noise to local updates to mitigate inference and gradient leakage attacks; cryptographic approaches, including homomorphic encryption (HE) and secure multi-party computation (SMPC), which protect model updates from inspection during transmission and aggregation; and robust aggregation mechanisms, which defend against model poisoning and Byzantine attacks by filtering or down-weighting malicious updates [17][22][28]. While DP provides formal statistical privacy guarantees, HE and SMPC ensure cryptographic confidentiality, and robust aggregation enhances integrity; hybrid combinations often balance privacy, security, computational overhead, and model performance in IoT environments [24][54].

Despite significant research progress, privacy-enhanced FL for IoT security remains fragmented. Existing studies often focus on isolated aspects such as optimization efficiency, privacy mechanisms, or specific attack models, with limited systematic consolidation of architectures, threat modeling frameworks, datasets, and evaluation methodologies [9][11][29]. Moreover, many proposed solutions are validated primarily through simulations or static datasets, with few real-world deployments on heterogeneous IoT hardware platforms [20], [30]. These gaps highlight the need for comprehensive reviews and unified analytical frameworks that synthesize existing knowledge, identify research challenges, and guide the development of robust, adaptive, and trustworthy federated learning systems for next-generation IoT security.

This paper presents a systematic review of privacy-enhanced federated learning for secure IoT applications. The main contributions of this review are threefold: (i) a comprehensive taxonomy of FL architectures, algorithms, and privacy-preserving techniques tailored to IoT security; (ii) a critical analysis of existing FL-based cybersecurity solutions, including datasets, evaluation metrics, and experimental platforms; and (iii) the identification of key research gaps and future directions for developing robust, adaptive, and trustworthy FL systems in heterogeneous IoT environments. The remainder of this paper is organized as follows: Section 2 describes the review of relevant studies; Section 3 reviews the methodology of federated learning fundamentals and privacy-enhancing techniques; Section 4 presents the results and discussion, discusses FL applications in IoT security, and highlights some open challenges and future research directions; and Section 6 concludes the paper.

2. LITERATURE REVIEW

2.1 Internet of Things (IoT) Architecture and Security Challenges

The Internet of Things (IoT) architecture is commonly organized into multiple layers, namely the device, edge, fog, and cloud layers, to support scalable data acquisition, processing, and intelligent decision-making. The device layer comprises sensors, actuators, and embedded systems responsible for data generation. The severe computational, memory, and energy constraints at the device layer make it impractical to deploy centralized data processing or heavy cryptographic protections directly on IoT nodes, thereby increasing exposure to data leakage and localized attacks [7][54]. These limitations necessitate a distributed learning and processing approach, where computation is offloaded to edge, fog, or cloud layers while keeping raw data localized, enabling scalable analytics, reduced communication overhead, and improved privacy preservation across the IoT architecture [22][26].

Despite these architectural advantages, IoT systems face a broad and evolving network-level, data-level, and device-level threat landscape. Common security threats include malware propagation, spoofing, data injection, botnet attacks, denial-of-service (DoS), unauthorized access, and privacy leakage [16][23][39]. The distributed and heterogeneous nature of IoT networks further complicates security management, as devices operate under diverse hardware configurations, communication protocols, and trust assumptions. Attackers can exploit weak authentication mechanisms, insecure communication channels, and limited patching capabilities to compromise IoT nodes at scale, making robust, adaptive, and scalable security solutions a critical requirement for modern IoT deployments [31][39][54].

2.2 Machine Learning in IoT Security

Machine learning (ML) plays a central role in IoT security, particularly in intrusion detection and anomaly detection systems, where it enables automated identification of malicious patterns in large-scale network traffic and device telemetry [16][25][29]. The selection of ML techniques is closely tied to IoT data characteristics, which are typically high-volume, heterogeneous, streaming, and often non-IID across distributed devices. Supervised learning methods such as SVMs, random forests, gradient boosting, and deep neural networks are

effective when labeled attack datasets are available, supporting accurate classification of known threats [25][41]. In contrast, unsupervised and semi-supervised approaches, including clustering and autoencoders, are better suited to IoT environments with limited labeled data and evolving attack patterns, enabling detection of anomalies and zero-day threats [29][22].

However, practical deployment of ML in IoT environments is constrained by device-level resource limitations, communication overhead, and dynamic data distributions. IoT devices often lack sufficient computational power, memory, and energy to support complex models, while centralized data aggregation introduces privacy risks, scalability challenges, and single points of failure [7][39][54]. Additionally, non-IID data and concept drift degrade model generalization and convergence, particularly in distributed settings [22][43]. These constraints require lightweight, adaptive, and privacy-aware ML frameworks that operate efficiently across heterogeneous IoT architectures.

2.3 Federated Learning Fundamentals

Federated learning (FL) is a distributed machine learning paradigm that enables collaborative model training across multiple clients without sharing raw data [21][22]. In a typical FL workflow, a central coordinator initializes a global model and distributes it to participating clients. Each client trains the model locally using its private data and transmits only model updates, such as gradients or weights, to the server. The server then aggregates these updates, commonly using the Federated Averaging (FedAvg) algorithm, to produce an improved global model that is iteratively refined over multiple communication rounds [21].

Compared to centralized learning, FL offers significant advantages in terms of data privacy, regulatory compliance, and communication efficiency, making it particularly well-suited for IoT environments [22][26]. However, FL also introduces new challenges, including handling non-independent and identically distributed (non-IID) data, system heterogeneity, unreliable client participation, and vulnerability to adversarial attacks such as model poisoning, backdoor insertion, and inference attacks [8][18][30]. These limitations have motivated extensive research into privacy-enhancing and robustness mechanisms to strengthen FL pipelines for security-critical IoT applications [12][53].

Federated learning systems are generally categorized into cross-device FL and cross-silo FL. Cross-device FL involves a large number of highly distributed, resource-constrained, and intermittently connected clients, such as IoT sensors and edge devices. In contrast, cross-silo FL typically involves fewer reliable, resource-rich participants, such as organizations or data centers. Most IoT security applications align with the cross-device FL paradigm, where scalability, adaptivity, and privacy preservation are essential design considerations [22][43]. Federated learning is conditionally suitable for IoT because its benefits depend heavily on architectural design choices and environmental constraints.

2.4 Taxonomy of Federated Learning for IoT Applications

2.4.1 Federated Learning System Architectures

Flat FL architectures are particularly problematic in IoT environments because they do not align well with inherent IoT characteristics such as massive scale, high client churn, and device heterogeneity. In large-scale IoT networks consisting of thousands or millions of intermittently connected sensors, direct communication between every client and a single central server creates significant uplink congestion, increased latency, and synchronization overhead [43]. Frequent client dropouts caused by mobility, power-saving modes, or unstable connectivity further destabilize training rounds, leading to slow convergence and biased aggregation.

Additionally, IoT devices exhibit substantial heterogeneity in computational capability, energy availability, and data distributions (non-IID). A flat topology forces all clients into uniform participation schedules and aggregation cycles, disadvantaging low-resource devices and amplifying straggler effects. This mismatch between architectural simplicity and IoT operational realities makes flat FL less efficient and less robust for large, dynamic, and resource-constrained IoT ecosystems, motivating hierarchical or clustered FL designs better aligned with edge-fog-cloud structures [22][43][54].

Figure 1 illustrates three major federated learning (FL) architectures commonly adopted in Internet of Things (IoT) environments, namely flat, hierarchical, and hybrid federated learning. These architectures differ significantly in terms of coordination structure, communication overhead, scalability, and suitability for heterogeneous IoT deployments.

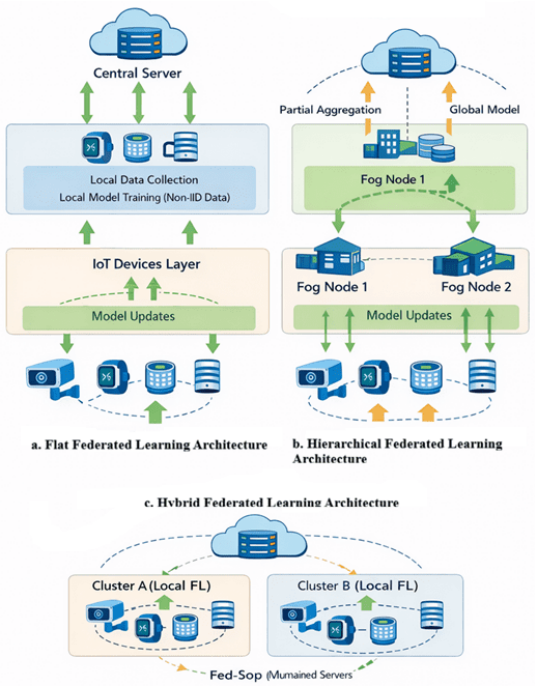

Figure 1. Federated Learning Architectures for IoT Applications
Source: Adapted from [43][45][51]

The flat federated learning architecture (Figure 1(a)) represents the classical FL paradigm originally introduced by [21]. In this architecture, a central server coordinates the learning process by directly aggregating model updates from all participating IoT devices. Each client trains a local model using its private data and transmits either model parameters or gradients to the central server for aggregation. While this approach is conceptually simple and effective in small-scale or relatively homogeneous environments, it exhibits notable limitations in large-scale IoT systems. Specifically, flat FL assumes reliable connectivity and relatively uniform client capabilities, assumptions that rarely hold in real-world IoT deployments characterized by non-IID data distributions, intermittent connectivity, and resource-constrained devices. Furthermore, direct communication between a large number of devices and a central server creates substantial communication overhead and increases susceptibility to single points of failure and denial-of-service attacks [24][43].

The hierarchical federated learning architecture (Figure 1(b)) addresses these limitations by introducing intermediate aggregation layers, such as edge or fog nodes, between IoT devices and the central server. In this model, local updates are first aggregated at edge or fog nodes to produce partial models, which are then forwarded to the cloud for global aggregation. This multi-tier design significantly reduces communication costs, improves scalability, and enhances responsiveness by moving computation closer to data sources. Hierarchical FL aligns well with large-scale IoT environments, including smart cities and industrial IoT systems, where devices are naturally organized based on geographical proximity or functional roles [26][43]. However, including intermediate aggregators introduces additional security and trust challenges, as compromised edge or fog nodes may manipulate aggregated updates or leak sensitive information if appropriate protection mechanisms are not enforced [18][54].

The hybrid federated learning architecture (Figure 1(c)) combines elements of both flat and hierarchical models by enabling localized federated learning within device clusters while maintaining coordination with a global server. In this approach, IoT devices are grouped into clusters based on proximity, task similarity, or network conditions, and FL is performed locally within each cluster. Periodically, cluster-level models are synchronized with a global model maintained in the cloud. This architecture offers increased flexibility and adaptability, making it particularly suitable for highly heterogeneous and dynamic IoT environments. Hybrid FL supports model personalization, reduces communication overhead, and improves robustness against both system and statistical heterogeneity [45][51]. Nevertheless, its increased architectural complexity introduces challenges related to orchestration, synchronization, and security management, necessitating advanced coordination mechanisms and adaptive aggregation strategies [22][54].

The comparison highlights that no single FL architecture is universally optimal for all IoT scenarios. Flat FL is best suited for small, controlled deployments; hierarchical FL excels in large-scale, latency-sensitive systems; and hybrid FL provides a balanced solution for complex, heterogeneous environments. Consequently, modern privacy-enhanced and security-aware FL systems increasingly adopt adaptive hybrid architectures,

integrating hierarchical aggregation, secure communication, and privacy-preserving mechanisms to meet the stringent requirements of real-world IoT applications [22][26][43].

2.4.2 Data Distribution and Heterogeneity

Data heterogeneity represents a fundamental challenge in federated learning–based IoT intrusion detection systems (IDS). Although many FL algorithms assume independent and identically distributed (IID) data, real-world IoT deployments generate highly non-IID traffic patterns due to differences in device roles, geographic location, firmware versions, user behavior, and attack exposure profiles [22][30]. For IDS, this means some clients may predominantly observe benign traffic, while others experience localized attack types (e.g., botnets, DDoS, spoofing). When such skewed distributions are aggregated without compensation, the global model becomes biased toward dominant traffic patterns, reducing detection rates for minority or emerging attacks, increasing false negatives, and degrading generalization performance [45][46].

Beyond statistical heterogeneity, device and system heterogeneity further affect IDS accuracy and reliability. Resource-constrained IoT devices may train shallow or truncated local models due to memory and energy limits, leading to inconsistent gradient quality and unstable aggregation. Intermittent connectivity and varying data volumes create uneven client contributions, which can distort class balance and impair convergence stability. In security-critical contexts, these effects translate directly into reduced attack-detection sensitivity, delayed response to zero-day threats, and increased false positives. Therefore, effective FL-based IDS for IoT must incorporate adaptive aggregation, personalization mechanisms, and heterogeneity-aware optimization strategies to preserve detection robustness under real-world operational variability [43][54].

2.4.3 Federated Optimization Algorithms

Federated optimization algorithms govern how local model updates are combined to form a global model. Federated Averaging (FedAvg) is the foundational FL algorithm, which aggregates client updates using a weighted average of local parameters. Although computationally efficient, FedAvg assumes relatively balanced data distributions and stable participation, making it less robust in heterogeneous IoT environments [21].

To address these limitations, FedProx extends FedAvg by adding a proximal term to the local objective function, constraining local updates and reducing divergence caused by system and data heterogeneity [24]. SCAFFOLD further improves convergence by employing control variates to correct client drift, particularly under highly non-IID data distributions [45]. In addition, personalized FL methods have emerged to allow clients to maintain customized models tailored to local data characteristics while still benefiting from shared global knowledge. These approaches are especially relevant for IoT applications, where device contexts and operating conditions vary widely [51].

2.4.4 Privacy-Preserving Mechanisms

Although federated learning reduces the need to share raw data, it does not inherently guarantee privacy. Consequently, several privacy-preserving mechanisms have been integrated into FL pipelines. Differential Privacy (DP) introduces carefully calibrated noise to model updates or gradients to limit the risk of inferring sensitive information about individual data points. While DP provides formal privacy guarantees, it introduces a trade-off between privacy strength and model accuracy, which is particularly critical in resource-constrained IoT environments [17][51].

Secure aggregation protocols ensure that the central server can access only aggregated model updates, preventing inspection of individual client contributions during training [17]. Cryptographic techniques such as Homomorphic Encryption (HE) and Secure Multi-Party Computation (SMPC) further enhance confidentiality by enabling computation over encrypted data. Although these approaches offer strong privacy guarantees, their computational and communication overhead must be carefully managed to remain feasible for large-scale IoT deployments [18][44][56].

2.4.5 Security Threat Models

Beyond privacy risks, federated learning (FL) systems face a wide range of security threats that can compromise model integrity and system reliability. Model poisoning attacks occur when malicious clients deliberately inject manipulated or corrupted updates during training to degrade global model performance or bias predictions toward adversarial objectives. A more sophisticated variant, backdoor attacks, embeds hidden triggers into the global model, causing it to behave maliciously under specific conditions while remaining benign during standard evaluation, making detection particularly challenging [8].

Gradient leakage attacks exploit shared gradients or model updates to infer sensitive information about local training data, undermining FL's privacy guarantees even without raw data sharing. Additionally, Byzantine attacks involve adversarial clients that behave arbitrarily or inconsistently, sending malformed or

strategically crafted updates to disrupt convergence or mislead the aggregation process. These threats are particularly amplified by the inherent openness of IoT ecosystems characterized by large-scale device enrollment, dynamic client churn, weak identity management, remote accessibility, and heterogeneous administrative domains where client trustworthiness cannot be strictly enforced. In such open IoT environments, compromised edge devices, counterfeit nodes, or externally accessible sensors can easily participate in training rounds, increasing the attack surface and making integrity violations more difficult to detect. Addressing these risks requires robust aggregation mechanisms, anomaly detection techniques, secure update validation, and trust-aware FL frameworks that can operate under adversarial and partially trusted IoT conditions [18][56].

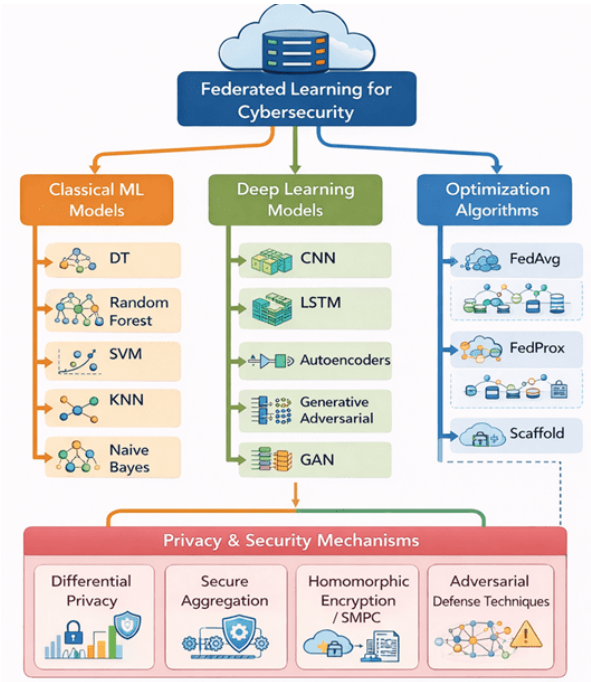

Figure 2: Taxonomy of ML and FL models in Cybersecurity
Source: Adapted from [22][24][44][51][54]

Figure 2 presents a comprehensive taxonomy of machine learning (ML) and federated learning (FL) models widely employed in cybersecurity research, with a particular focus on IoT environments. The taxonomy captures both algorithmic diversity and the increasing emphasis on privacy-preserving distributed learning.

At the foundation of the taxonomy are classical machine learning models, including Decision Trees (DT), Random Forests (RF), Support Vector Machines (SVM), k-Nearest Neighbors (KNN), and Naïve Bayes classifiers. These models have been widely used for intrusion and anomaly detection because of their interpretability, low computational complexity, and effectiveness on structured network traffic data [25][29]. In federated IoT settings, classical ML models are often preferred for deployment on resource-constrained devices because they require fewer parameters and incur lower communication overhead than deep learning alternatives [45].

The taxonomy further incorporates deep learning models, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, autoencoders, and Generative Adversarial Networks (GANs). These architectures are particularly effective in capturing complex spatial and temporal patterns in IoT traffic, enabling the detection of sophisticated cyberattacks, including distributed denial-of-service (DDoS) and advanced persistent threats [16][41]. Despite their higher communication and computation costs, deep models are increasingly adopted in FL systems, supported by techniques such as model compression, sparse updates, and partial aggregation to improve scalability [22].

A critical layer of the taxonomy comprises federated optimization algorithms, including FedAvg, FedProx, and SCAFFOLD. FedAvg remains the most widely used aggregation strategy due to its simplicity and efficiency; however, it assumes IID data distributions, which rarely hold in IoT environments [21]. FedProx mitigates this limitation by introducing proximal regularization to address system and statistical heterogeneity, while SCAFFOLD employs control variates to correct client drift under non-IID data conditions [24][45]. These optimization techniques are essential for achieving stable convergence and robust performance in heterogeneous IoT networks.

At the top of the taxonomy, privacy and security mechanisms are integrated across all learning paradigms. Differential Privacy (DP), Secure Aggregation, Homomorphic Encryption (HE), and Secure Multi-Party Computation (SMPC) are highlighted as key defenses against inference attacks, gradient leakage, and model inversion threats [17][51][44]. Their inclusion underscores a shift toward privacy-by-design federated learning, where security safeguards are embedded throughout the learning lifecycle rather than treated as post-hoc enhancements.

This taxonomy demonstrates that effective FL-based IoT cybersecurity requires a synergistic integration of learning models, optimization algorithms, and privacy-preserving techniques. By consolidating these dimensions into a unified framework, the taxonomy provides a structured lens for analyzing existing literature and identifying research gaps, such as the limited adoption of adaptive optimization and cryptographic protections in real-world deployments, thereby motivating the proposed privacy-enhanced FL architecture presented in this study.

2.5 Privacy-Enhanced Federated Learning in IoT Security

2.5.1 Federated Learning–Based Intrusion Detection Systems

Federated learning has gained significant attention as an enabling technology for intrusion detection systems (IDS) in IoT environments because it can learn collaboratively from distributed data while preserving privacy. Network-based IDS (NIDS) leverages FL to analyze traffic patterns across multiple IoT nodes or gateways, enabling collaborative detection of attacks such as DDoS, scanning, and botnet activity without centralizing sensitive traffic data [12]. By sharing only model updates, FL-based NIDS reduce privacy risks and communication overhead while improving detection robustness across diverse network segments.

Host-based IDS (HIDS) focuses on device-level behaviors, including system calls, resource utilization, and firmware activities. FL-enabled HIDS allow resource-constrained IoT devices to contribute to a global detection model without exposing sensitive local logs, which is particularly important in domains such as healthcare and industrial IoT. Hybrid IDS, combining network- and host-level features, represents an emerging trend in FL-based IoT security. These systems achieve higher detection accuracy by capturing both global traffic patterns and local behavioral anomalies while maintaining strong privacy guarantees through decentralized learning [63][64].

2.5.2 Proactive vs. Reactive Cybersecurity Models

Traditional IDS solutions largely operate under a reactive cybersecurity paradigm, where threats are detected only after malicious activity has occurred. While effective against known attack signatures, reactive IDS struggles with zero-day attacks, evolving threat patterns, and fast-moving adversaries commonly observed in IoT ecosystems [39]. These limitations are further amplified in centralized ML-based systems due to delayed data aggregation and analysis.

In contrast, proactive cybersecurity models aim to anticipate and prevent attacks by continuously learning from evolving system behavior. Privacy-enhanced FL supports this shift by enabling predictive, adaptive learning, where models are updated dynamically across distributed IoT nodes. Adaptive FL mechanisms can adjust client participation, aggregation frequency, and personalization strategies based on detected risk levels and environmental context, thereby enhancing resilience against emerging threats and enabling near real-time response [26][43].

2.5.3 Threat Modeling Frameworks

Formal threat modeling frameworks provide a structured approach for analyzing adversarial behavior and evaluating the effectiveness of FL-based security defenses. The MITRE ATT&CK framework offers a comprehensive knowledge base of real-world adversary tactics, techniques, and procedures (TTPs). In FL-enabled IoT security, MITRE ATT&CK can model attack scenarios such as data poisoning, model inversion, and command-and-control communication, supporting systematic evaluation of IDS and defensive mechanisms [65].

The STRIDE model, which categorizes threats into Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege, provides a complementary system-centric perspective. Mapping FL defenses such as differential privacy, secure aggregation, homomorphic encryption, and robust aggregation to STRIDE categories enables comprehensive risk assessment and reinforces security-by-design principles. Together, MITRE ATT&CK and STRIDE provide a robust foundation for analyzing both adversarial behavior and architectural vulnerabilities in privacy-enhanced federated learning systems for IoT security [56].

3. METHOD

This study adopts a systematic literature review (SLR) methodology to comprehensively examine privacy-enhanced federated learning techniques for secure Internet of Things (IoT) applications. The methodology is designed to ensure rigor, transparency, and reproducibility in identifying, analyzing, and synthesizing existing research. A structured review protocol was followed, encompassing literature selection, screening, classification, and comparative analysis of relevant studies.

3.1 Review Methodology and Data Sources

3.1.1 Literature Selection Criteria

Relevant literature was collected from well-established and authoritative digital libraries, including IEEE Xplore, Elsevier ScienceDirect, SpringerLink, ACM Digital Library, and MDPI. The search focused on peer-reviewed journal articles, conference proceedings, and high-impact survey papers, most published between 2020 and 2025, reflecting the rapid evolution of federated learning and IoT security research.

We used a set of carefully defined keywords and Boolean search strings, including "federated learning," "privacy-enhancing/preserving machine learning," "IoT security," "intrusion detection," "distributed learning," "secure aggregation," and "differential privacy." Studies were included if they: (i) addressed federated learning or privacy-enhanced machine learning, (ii) focused on IoT, cyber-physical systems, or edge computing environments, (iii) proposed or evaluated security and privacy mechanisms, and (iv) provided experimental validation or systematic analysis. Articles were excluded if they were non-peer-reviewed, lacked technical depth, or were not directly related to security or privacy in federated learning contexts.

3.1.2 Classification and Analysis Approach

Following selection, the identified studies were systematically classified using a multi-dimensional taxonomy. The classification dimensions included FL system architecture (flat, hierarchical, hybrid), data distribution characteristics (IID vs non-IID), federated optimization algorithms (e.g., FedAvg, FedProx, SCAFFOLD), privacy-preserving mechanisms (differential privacy, secure aggregation, cryptographic techniques), and security threat models such as model poisoning, backdoor attacks, and gradient leakage.

A qualitative and comparative analysis was then conducted to evaluate the strengths, limitations, and applicability of existing approaches in heterogeneous IoT environments. Particular emphasis was placed on performance metrics, scalability, privacy–utility trade-offs, and robustness against adversarial threats. The findings were synthesized to identify recurring challenges, research gaps, and emerging trends, which subsequently informed the discussion of open issues and future research directions in privacy-enhanced federated learning for IoT security.

3.2 Datasets, Tools, and Experimental Platforms

3.2.1 IoT Cybersecurity Datasets

Among the available benchmarks, UNSW-NB15 has emerged as dominant in the FL-based IoT security literature, primarily because of its balanced attack taxonomy, high-dimensional feature set, and manageable size, which collectively facilitate controlled experimental partitioning across simulated clients [23][25]. Unlike older datasets such as KDD-based corpora, UNSW-NB15 includes modern attack behaviors and well-engineered flow features, enabling researchers to construct statistically heterogeneous client subsets (e.g., by attack class, traffic volume, or temporal segmentation). This makes it particularly suitable for evaluating federated convergence behavior, robustness under non-IID distributions, and the impact of aggregation strategies, core concerns in FL research [22][57].

From a federated suitability perspective, UNSW-NB15 supports reproducible data partitioning schemes that emulate cross-device IoT environments without requiring real distributed infrastructure. Its structured labeling allows simulation of client imbalance, skewed class distributions, and adversarial participation conditions frequently studied in FL-based intrusion detection systems. Moreover, its widespread adoption promotes comparability across studies, enabling consistent benchmarking of privacy-preserving techniques, communication efficiency, and model accuracy trade-offs in federated settings. Thus, its dominance is less about IoT specificity and more about experimental controllability, reproducibility, and alignment with FL evaluation requirements.

3.2.2 Federated Learning Frameworks and Libraries

A variety of open-source frameworks support federated learning experimentation and deployment. TensorFlow Federated (TFF) is a production-grade framework designed for implementing and simulating FL algorithms using TensorFlow. It provides high-level abstractions for federated computation and supports large-scale simulations across virtual clients, making it suitable for evaluating FL algorithms under controlled experimental settings [1].

PySyft, developed by OpenMined, focuses on privacy-preserving machine learning and supports federated learning, differential privacy, and encrypted computation. Its flexible design enables researchers to experiment with advanced privacy mechanisms such as secure aggregation and secure multi-party computation, making it particularly relevant for IoT applications that require strong privacy guarantees [17][56]. FedML is a research-oriented FL platform that emphasizes scalability and heterogeneity awareness. It supports cross-device and cross-silo FL scenarios, non-IID data distributions, and diverse system configurations, making it well-suited for studying FL in heterogeneous IoT environments [19].

3.2.3 Simulation and Hardware Platforms

Simulation tools are widely used to model communication behavior, network dynamics, and scalability issues in FL-enabled IoT systems. NS-3 and Mininet enable detailed simulation of network protocols, latency, packet loss, and bandwidth constraints, which are critical for evaluating the communication overhead and performance of federated learning in distributed IoT networks [9]. These tools allow researchers to test FL algorithms under realistic networking conditions without the cost of large-scale physical deployments.

For application-level communication and prototyping, Node-RED and MQTT brokers such as Mosquitto are commonly employed. Node-RED provides a visual programming environment for wiring IoT devices, services, and data flows, while MQTT enables lightweight publish/subscribe messaging suitable for resource-constrained IoT devices [58]. To validate real-world feasibility, hardware platforms such as Raspberry Pi and NVIDIA Jetson Nano are frequently used as edge clients in FL experiments. These devices enable practical evaluation of computational overhead, energy consumption, and network integration, bridging the gap between simulation-based studies and real-world IoT deployments [38].

Table 1: Summary of IoT Cybersecurity Datasets Used in the Literature

Dataset	Year	Domain / Traffic Type	Key Attack Types	Common ML / FL Usage	Key Limitations	Representative References
KDD Cup 99	1999	Network traffic	DoS, Probe, R2L, U2R	Classical ML baseline, IDS benchmarking	Outdated attacks, redundancy, unrealistic traffic	[41][16]
NSL-KDD	2009	Network traffic	DoS, Probe, R2L, U2R	ML/FL IDS evaluation	Limited realism, not IoT-specific	[41]
UNSW-NB15	2015	Enterprise & IoT-like traffic	DoS, Fuzzers, Backdoors, Exploits	ML, DL, FL-based IDS	High dimensionality, centralized labeling	[25][16]
CIC-IDS2017	2017	Network flow traffic	DDoS, Botnet, Brute-force	Deep learning IDS	Limited IoT device diversity	[29][42]
CIC-DDoS2019	2019	Network traffic	Volumetric & application-layer DDoS	DDoS-focused IDS	Single-attack focus	[29][42]
Bot-IoT	2018	IoT network traffic	DDoS, Reconnaissance, Data exfiltration	IoT-specific IDS, FL-based studies	Synthetic environment	[23][10]
TON_IoT	2020	IoT + Edge + Cloud telemetry	DDoS, Injection, Backdoor, MITM	FL-based IDS, CPS security	Large-scale complexity	[57][10]
IoT-23	2020	IoT malware traffic	Botnets, C&C, Scanning	Malware detection, IDS	Limited labeled benign data	[23]
Edge-IIoTset	2022	Industrial IoT	DoS, Spoofing, Malware	FL-based industrial IDS	Domain-specific constraints	[10][54]
Synthetic FL Datasets	–	Partitioned network traffic	Varies	FL benchmarking under non-IID data	Lack of realism	[22][30][45]

The reviewed literature shows that IoT security research relies heavily on a small set of benchmark datasets, as shown in Table 1, with UNSW-NB15, Bot-IoT, and TON_IoT being the most frequently adopted in federated learning–based intrusion detection systems [10][23][42]. Earlier datasets such as KDD Cup 99 and NSL-KDD continue to appear in comparative studies due to their historical significance; however, their limited realism and outdated attack profiles reduce their suitability for modern IoT environments [41][16].

Recent datasets such as TON_IoT and Edge-IIoTset represent a significant advancement by incorporating heterogeneous IoT devices, edge computing, and cloud interactions, making them more aligned with contemporary IoT–FL deployment scenarios [57][54]. These datasets are particularly valuable for evaluating non-IID data distributions, system heterogeneity, and cross-device federated learning, which are central challenges identified in FL research [22][30].

Despite these advances, the literature consistently highlights a lack of standardized federated learning benchmarks. Most FL-based studies rely on synthetic data partitioning to emulate federated settings, which

limits reproducibility and real-world validity [22][45]. Furthermore, few datasets explicitly support privacy attack evaluation, such as gradient leakage or model inversion, underscoring a critical gap in current benchmarking practices [18][42]. This review indicates a pressing need for large-scale, real-world, privacy-aware IoT datasets explicitly designed for federated learning evaluation, incorporating realistic adversarial behaviors, heterogeneous devices, and regulatory constraints.

3.3 Evaluation Metrics and Performance Analysis

The effectiveness of privacy-enhanced federated learning (FL) models for IoT security must be assessed using a comprehensive set of evaluation metrics that capture not only detection accuracy but also system efficiency, privacy protection, and scalability. This section discusses the key performance metrics commonly employed in the literature to evaluate FL-based intrusion detection and cybersecurity systems in heterogeneous IoT environments.

3.3.1 Detection Performance Metrics

Detection performance metrics are fundamental for assessing the effectiveness of FL-based models in identifying malicious activities within IoT systems. Accuracy measures the proportion of correctly classified instances among all predictions and provides a general overview of model performance. However, in IoT security contexts where benign traffic significantly outweighs attack samples, accuracy can be misleading, as a model may achieve high accuracy while failing to detect rare but critical attacks [29]. From a security perspective, overreliance on accuracy may obscure vulnerabilities to stealthy or low-frequency threats.

To address this limitation, precision, recall, and F1-score offer more security-relevant insights. Precision reflects the model's ability to minimize false positives, which is essential in IoT deployments where excessive false alarms can overwhelm limited administrative resources and degrade system trust. Recall (detection rate) measures the proportion of actual attacks correctly identified, directly relating to the system's capability to prevent compromise and limit attack propagation. The F1-score balances these two objectives, making it particularly suitable for intrusion detection under class imbalance and heterogeneous data conditions common in federated IoT environments [6][52]. Collectively, these metrics ensure that FL-based IDS is evaluated not only for statistical correctness but also for operational security impact and real-world deployability.

3.3.2 Communication Overhead

Communication overhead is a critical performance metric in federated learning, especially in bandwidth-constrained IoT environments. It refers to the amount of data exchanged between clients and the central aggregator during model training, typically measured in terms of transmitted bytes per training round or total communication cost. Excessive communication can increase latency, energy consumption, and network congestion, undermining the practicality of FL deployment in large-scale IoT systems [22].

Evaluating communication overhead involves analyzing factors such as model size, update frequency, client participation rate, and aggregation strategy. Techniques such as model compression, sparse updates, and hierarchical aggregation are often assessed based on their ability to reduce communication cost without significantly degrading detection performance [43].

3.3.3 Computation Cost

Computation cost captures the processing burden imposed on IoT devices during local model training and inference. This metric is particularly important for resource-constrained edge devices with limited CPU, memory, and energy resources. Computation cost is typically evaluated in terms of training time, CPU utilization, energy consumption, or number of floating-point operations (FLOPs) per training round [38].

In privacy-enhanced FL settings, computation cost may increase due to the integration of privacy-preserving mechanisms such as differential privacy, secure aggregation, or cryptographic operations. Therefore, evaluating computation cost helps assess the feasibility of deploying FL-based security models on real-world IoT hardware and guides the selection of lightweight model architectures and optimization techniques [54].

3.3.4 Privacy Leakage

Although federated learning mitigates direct data sharing, it remains susceptible to privacy inference attacks. Privacy leakage metrics quantify how much sensitive information about local training data can be inferred from shared model updates. Common evaluation approaches include measuring the success rate of membership inference attacks, gradient inversion attacks, or model reconstruction attacks under different privacy configurations [17][8].

The effectiveness of privacy-enhancing techniques such as differential privacy and secure aggregation is evaluated by analyzing the trade-off between privacy guarantees and model utility. Metrics such as privacy

budget (ε) in differential privacy or attack success probability are commonly reported to demonstrate the robustness of FL systems against information leakage [51].

3.3.5 Scalability and Convergence

Scalability and convergence metrics assess FL systems' ability to maintain stable, efficient learning as the number of participating IoT devices increases. Scalability is evaluated by analyzing performance trends as client population size, data volume, or network complexity grows. Convergence metrics focus on the number of communication rounds or training iterations required for the global model to reach a stable performance level [24].

In heterogeneous IoT environments, non-IID data distributions and intermittent client participation can significantly affect convergence behavior. Therefore, evaluating scalability and convergence is essential for validating adaptive FL algorithms such as FedProx and SCAFFOLD, which are designed to improve stability and learning efficiency under realistic deployment conditions [43][45].

4. RESULTS AND DISCUSSION

4.1 Comparative Analysis of FL Architectures and Algorithms

The reviewed studies consistently demonstrate that hierarchical and hybrid FL architectures outperform flat FL deployments in large-scale and heterogeneous IoT systems [12][18][27][41]. By introducing intermediate aggregation layers at the edge or fog level, hierarchical FL significantly reduces communication latency and server bottlenecks while improving scalability and convergence stability. Hybrid architectures further enhance performance by dynamically balancing local autonomy and global coordination, particularly in dense or mission-critical IoT environments [21][35].

Regarding optimization strategies, FedAvg remains the most widely adopted baseline due to its simplicity and low computational overhead [6][14]. However, multiple studies report notable performance degradation under non-IID data distributions, which are typical in IoT settings [22][30]. Advanced optimization methods such as FedProx and SCAFFOLD demonstrate improved convergence behavior and higher accuracy by explicitly addressing client drift and statistical heterogeneity [25][33][44]. Personalized FL variants also show measurable gains in local model performance, especially for context-aware applications such as smart healthcare and industrial IoT [38][52].

Heterogeneity emerges as a central challenge affecting FL effectiveness across nearly all reviewed works [9][17][29]. Device heterogeneity, driven by disparities in processing power, memory, energy constraints, and network reliability, leads to uneven participation and training instability. Statistical heterogeneity, driven by non-IID and context-specific data distributions, further exacerbates model bias and slows convergence [24][31][46]. The literature emphasizes that heterogeneity-aware client selection, adaptive aggregation, and algorithmic personalization are essential for sustaining FL performance in realistic IoT deployments [36][48][57].

4.2 Effectiveness of Privacy-Preserving Mechanisms

Privacy-enhancing mechanisms, including Differential Privacy (DP), Secure Aggregation, Homomorphic Encryption, and Secure Multi-Party Computation (SMPC), are widely adopted to mitigate information leakage in FL [11][19][28][40]. However, the literature consistently reports an inherent privacy–utility trade-off. DP-based approaches provide formal privacy guarantees but often degrade model accuracy due to injected noise, particularly in low-data or resource-constrained IoT nodes [20][34]. Secure aggregation techniques preserve learning accuracy but increase communication overhead and protocol complexity [26][43].

The reviewed studies show that privacy-enhanced FL significantly improves resilience against model poisoning, backdoor attacks, and gradient inference attacks [15][23][39][51]. Secure aggregation and encrypted computation limit an adversary’s visibility into individual updates, while robust aggregation rules reduce the influence of malicious clients. Nevertheless, several works acknowledge that no single defense is sufficient when attackers control a fraction of participating devices, highlighting the need for layered defenses combining privacy mechanisms with anomaly detection and trust-aware aggregation [47][54][60].

4.3 Discussion of Open Challenges

Despite substantial progress, the literature reveals persistent challenges. Resource constraints remain a major barrier, as many IoT devices lack the computational and energy capacity required for cryptographic privacy mechanisms and deep learning models [8][32][49]. Additionally, real-world deployments are scarce, with most studies relying on simulations or static datasets, limiting insight into long-term behavior under device churn, evolving threats, and dynamic network conditions [37][55]. The absence of standardized benchmarks and evaluation protocols further complicates cross-study comparison and reproducibility [45][58][62].

4.4 Implications for Secure IoT System Design

Collectively, the reviewed evidence suggests that effective FL-based IoT security systems should adopt hierarchical or hybrid architectures, integrate privacy-by-design principles, and employ heterogeneity-aware and adaptive optimization strategies [16][42][53]. Security mechanisms must be embedded throughout the FL lifecycle rather than treated as post hoc additions. These findings underscore the need to design scalable, adaptive, and trustworthy federated learning frameworks that support real-world IoT cybersecurity requirements.

5. CONCLUSION AND LIMITATION

This systematic review examined the state of the art in privacy-enhanced federated learning (FL) for securing Internet of Things (IoT) applications. The review synthesized existing research on FL architectures, optimization algorithms, privacy-preserving mechanisms, and security threat models, with particular emphasis on their application to intrusion detection and anomaly detection in heterogeneous IoT environments. The findings highlight that federated learning offers a promising alternative to centralized machine learning by enabling collaborative model training while minimizing raw data sharing, thereby reducing privacy risks and regulatory concerns. However, the review also reveals persistent challenges related to non-IID data distributions, system heterogeneity, adversarial robustness, and the privacy–utility trade-off. From a practical perspective, the reviewed studies demonstrate that privacy-enhanced FL can significantly improve the feasibility of deploying intelligent security solutions in real-world IoT systems. Techniques such as differential privacy, secure aggregation, and cryptographic protections enable compliance with data protection requirements while maintaining acceptable detection performance. Moreover, hierarchical and edge-aware FL architectures reduce communication overhead and support scalability in resource-constrained environments. These insights provide valuable guidance for practitioners and system designers seeking to implement secure, distributed, and privacy-aware IoT security frameworks. Lastly, privacy-enhanced federated learning represents a critical enabler for next-generation IoT cybersecurity, bridging the gap between intelligent threat detection and stringent privacy requirements. While substantial progress has been made, further research is needed to advance adaptive, explainable, and trust-aware FL systems that can operate autonomously in dynamic and large-scale IoT deployments. Addressing these challenges will be essential for translating federated learning from experimental settings into robust, real-world IoT security solutions.

Future research on privacy-enhanced federated learning (FL) for secure Internet of Things (IoT) applications should focus on advancing adaptability, transparency, trustworthiness, and autonomy to meet the demands of highly dynamic and large-scale IoT ecosystems.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Tertiary Education Trust Fund (TETFund), Nigeria, for the provision of a PhD scholarship.

REFERENCES

[1] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, et al., "TensorFlow: Large-scale machine learning on heterogeneous systems," *Proc. 12th USENIX Symp. Operating Systems Design and Implementation (OSDI)*, 2016.

[2] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, vol. 9, no. 1, pp. 1400–1421, 2022. https://doi.org/10.1109/JIOT.2021.3097904

[3] S. R. Abbas, Z. Abbas, A. Zahir, and S. W. Lee, "Federated learning in smart healthcare: A comprehensive review on privacy, security, and predictive analytics with IoT integration," *Healthcare*, vol. 12, no. 24, p. 2587, 2025. https://doi.org/10.3390/healthcare12242587

[4] M. Ajuji, M. Dawaki, A. Mohammed, and A. Ahmad, "Estimating residential natural gas demand and consumption: A hybrid ensemble machine learning approach," *Vokasi Unesa Bulletin of Engineering, Technology and Applied Science*, vol. 2, no. 3, pp. 549–557, 2025. https://doi.org/10.26740/vubeta.v2i3.40135

[5] M. N. Alatawi, "SAFEL-IoT: Secure adaptive federated learning with explainability for anomaly detection in 6G-enabled smart industry 5.0," *Electronics*, vol. 14, no. 11, p. 2153, 2025. https://doi.org/10.3390/electronics14112153

[6] A. Alazab, A. Khraisat, S. Singh, and T. Jan, "Enhancing privacy-preserving intrusion detection through federated learning," *Electronics*, vol. 12, no. 16, p. 3382, 2023. https://doi.org/10.3390/electronics12163382

[7] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surv. Tutor.*, vol. 17, no. 4, pp. 2347–2376, 2015. https://doi.org/10.1109/COMST.2015.2444095

[8] E. Bagdasaryan, A. Veit, Y. Hua, D. Estrin, and V. Shmatikov, "How to backdoor federated learning," *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, 2020.

[9] D. Bankov, E. Khorov, and A. Lyakhov, "On the limits of LoRaWAN channel access," *Proc. Int. Conf. Eng. Telecommun. (EnT)*, pp. 10–14, 2018. https://doi.org/10.1109/EnT.2016.011.

[10] E. M. Campos *et al.*, "Evaluating federated learning for intrusion detection in Internet of Things: Review and challenges," *Comput. Netw.*, vol. 203, p. 108661, 2022.
[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, pp. 785–794, 2016. https://doi.org/10.1145/2939672.2939785
[12] X. Chen *et al.*, "Trustworthy federated learning: Privacy, security, and beyond," *Knowl. Inf. Syst.*, early access, 2024. https://doi.org/10.1007/s10115-024-02285-2
[13] Cisco, "Annual Internet Report," 2023. [Online]. Available: Cisco. Accessed: May 12, 2025.\
[14] J. W. Creswell and J. D. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed. Thousand Oaks, CA, USA: SAGE, 2018.
[15] J. W. Creswell and V. L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 3rd ed. Thousand Oaks, CA, USA: SAGE, 2017.
[16] S. García, M. Grill, J. Stiborek, and A. Zunino, "An empirical comparison of botnet detection methods," *Comput. Secur.*, vol. 45, pp. 100–123, 2014. https://doi.org/10.1016/j.cose.2014.05.011
[17] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client-level perspective," *arXiv preprint arXiv:1712.07557*, 2017.
[18] R. Gosselin *et al.*, "Privacy and security in federated learning: A survey," *Appl. Sci.*, vol. 12, no. 19, p. 9901, 2022. https://doi.org/10.3390/app12199901
[19] C. He *et al.*, "FedML: A research library and benchmark for federated machine learning," *arXiv preprint arXiv:2007.13518*, 2020.
[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735
[21] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, 2017, pp. 1273–1282.
[22] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, *et al.*, "Advances and open problems in federated learning," *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021. https://doi.org/10.1561/2200000083
[23] N. Koroniotis *et al.*, "Towards the development of realistic botnet dataset in the Internet of Things," *Future Gener. Comput. Syst.*, vol. 100, pp. 779–796, 2019. https://doi.org/10.1016/j.future.2019.05.041
[24] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process. Mag.*, vol. 37, no. 3, pp. 50–60, 2020. https://doi.org/10.1109/MSP.2020.2975749
[25] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems," in *Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS)*, pp. 1–6, 2015. https://doi.org/10.1109/MilCIS.2015.7348942
[26] D. C. Nguyen *et al.*, "Federated learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surv. Tutor.*, vol. 23, no. 3, pp. 1622–1658, 2021. https://doi.org/10.1109/COMST.2021.3075439
[27] A. Paszke *et al.*, "PyTorch: An imperative style, high-performance deep learning library," in *Adv. Neural Inf. Process. Syst.*, vol. 32, pp. 8026–8037, 2019.
[28] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[29] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset," in *Proc. ICISSP*, pp. 108–116, 2018. https://doi.org/10.5220/0006639801080116
[30] H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-IID data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, 2021. https://doi.org/10.1016/j.neucom.2021.01.119
[31] K. Peng, M. Li, H. Huang, C. Wang, S. Wan, and K. K. R. Choo, "Security challenges and opportunities for smart contracts in Internet of Things: A survey," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 12004–12020, 2021. https://doi.org/10.1109/JIOT.2021.3074544
[32] G. F. Riley and T. R. Henderson, "The ns-3 network simulator," in *Modeling and Tools for Network Simulation*. Berlin, Germany: Springer, pp. 15–34, 2010. https://doi.org/10.1007/978-3-642-12331-3_2
[33] P. Ruzafa-Alcázar *et al.*, "Intrusion detection based on privacy-preserving federated IDS using differential privacy," *IEEE Access*, vol. 10, pp. 62098–62113, 2022. https://doi.org/10.1109/ACCESS.2022.3178054
[34] T. Ryffel *et al.*, "A generic framework for privacy-preserving deep learning," *J. Mach. Learn. Res.*, vol. 21, no. 1, pp. 1–46, 2020.
[35] M. Sarhan, W. W. Lo, S. Layeghy, and M. Portmann, "HBFL: A hierarchical blockchain-based federated learning framework for collaborative IoT intrusion detection," *Comput. Electr. Eng.*, vol. 103, p. 108379, 2022. https://doi.org/10.1016/j.compeleceng.2022.108379
[36] J. Shen *et al.*, "Effective intrusion detection in heterogeneous Internet-of-Things networks via ensemble knowledge distillation-based federated learning," *arXiv preprint arXiv:2401.11968*, 2024. https://doi.org/10.1109/ICC51166.2024.10622262

[37] A. Shostack, *Threat Modeling: Designing for Security*. Hoboken, NJ, USA: Wiley, 2014.
[38] B. Shubyn *et al.*, "Resource consumption of federated learning approach applied on edge IoT devices in the AGV environment," *Proc. Int. Conf. Comput. Sci. (ICCS)*, 2023. https://doi.org/10.1007/978-3-031-36030-5_39
[39] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in IoT: Challenges and solutions," *Comput. Netw.*, vol. 190, p. 107859, 2022. https://doi.org/10.1016/j.comnet.2021.107859
[40] H. Tabrizchi and M. Aghasi, "Cyber security intelligent systems based on federated learning," *Adv. Multimedia and Ubiquitous Engineering: FutureTech 2024*, C. Figueroa, T. H. Kim, and K. Choo, Eds. Singapore: Springer, pp. 33–43, 2024. https://doi.org/10.1007/978-3-031-86592-3_4
[41] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," *Proc. IEEE Symp. Comput. Intell. Secur. Def. Appl.*, pp. 1–6, 2009. https://doi.org/10.1109/CISDA.2009.5356528
[42] A. Vyas, P.-C. Lin, R.-H. Hwang, and M. Tripathi, "Privacy-preserving federated learning for intrusion detection in IoT environments: A survey," *IEEE Access*, vol. 12, pp. 127018–127050, 2024. https://doi.org/10.1109/ACCESS.2024.3454211
[43] T. Xie, J. Liu, C. Zhang, and M. Chen, "Adaptive federated learning for heterogeneous IoT environments: A survey," *IEEE Internet Things J.*, vol. 10, no. 5, pp. 3212–3228, 2023. https://doi.org/10.1109/JIOT.2022.3194719
[44] Y. Zhang, R. Yu, and S. Xie, "Privacy-preserving machine learning for healthcare: A survey," *IEEE Trans. Ind. Informatics*, vol. 17, no. 6, pp. 3772–3784, 2021.
[45] R. Zhao, H. Chen, Y. Li, Y. Yin, J. Xu, Z. Xu, and Q. Yang, "Federated learning with non-IID data in edge computing: A survey," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4475–4497, Mar. 2021. https://doi.org/10.1109/JIOT.2020.3033424
[46] H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-IID data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, 2021. https://doi.org/10.1016/j.neucom.2021.01.119
[47] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
[48] X. Huang, J. Liu, Y. Lai, B. Mao, and H. Lyu, "EEFED: Personalized federated learning of execution and evaluation dual network for CPS intrusion detection," *IEEE Trans. Inf. Forensics Secur.*, 2023. https://doi.org/10.1109/TIFS.2023.3327481
[49] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 30, pp. 3146–3154, 2017.
[50] L. U. Khan, W. Saad, Z. Han, E. Hossain, and C. S. Hong, "Federated learning for Internet of Things: Recent advances, taxonomy, and open challenges," *arXiv preprint arXiv:2009.13012*, 2020.
[51] L. Lyu, H. Yu, and Q. Yang, "Towards fair and privacy-preserving federated deep models," *IEEE Trans. Mobile Comput.*, vol. 21, no. 3, pp. 838–851, 2022. https://doi.org/10.1109/TMC.2020.2981310.
[52] S. A. Mahmud, N. Islam, Z. Islam, Z. Rahman, and S. T. Mehedi, "Privacy-preserving federated learning-based intrusion detection technique for cyber-physical systems," *Mathematics*, vol. 12, no. 20, p. 3194, 2024. https://doi.org/10.3390/math12203194
[53] H. U. Manzoor, A. Shabbir, A. Chen, D. Flynn, and A. Zoha, "A survey of security strategies in federated learning: Defending models, data, and privacy," *Future Internet*, vol. 16, no. 10, p. 374, 2024. https://doi.org/10.3390/fi16100374
[54] T. M. Mengistu, T. Kim, and J.-W. Lin, "A survey on heterogeneity taxonomy, security, and privacy preservation in the integration of IoT, wireless sensor networks, and federated learning," *Sensors*, vol. 24, no. 3, p. 968, 2024. https://doi.org/10.3390/s24030968
[55] MITRE, "MITRE ATT&CK: Adversarial tactics, techniques, and common knowledge," 2023. [Online]. Available: https://attack.mitre.org/
[56] F. Mosaiyebzadeh *et al.*, "Privacy-enhancing technologies in federated learning for the Internet of Healthcare Things: A survey," *Electronics*, vol. 12, no. 12, p. 2703, 2023. https://doi.org/10.3390/electronics12122703
[57] N. Moustafa, "A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets," *Sustain. Cities Soc.*, vol. 72, p. 102994, 2021. https://doi.org/10.1016/j.scs.2021.102994
[58] N. Naik, "Choice of effective messaging protocols for IoT systems: MQTT, CoAP, AMQP and HTTP," in *Proc. IEEE Int. Syst. Eng. Symp. (ISSE)*, pp. 1–7, 2017. https://doi.org/10.1109/SysEng.2017.8088251
[59] D. C. Nguyen, M. Ding, P. N. Pathirana, and A. Seneviratne, "Federated learning for smart cities: Recent advances and applications," *IEEE Commun. Surv. Tutor.*, vol. 25, no. 1, pp. 734–766, 2023.
[60] R. Ryffel *et al.*, "A generic framework for privacy-preserving deep learning," *J. Mach. Learn. Res.*, vol. 21, no. 1, pp. 1–46, 2020.
[61] J. Zhang, R. Yu, and S. Xie, "Privacy-preserving machine learning for healthcare: A survey," *IEEE Trans. Ind. Informatics*, vol. 17, no. 6, pp. 3772–3784, 2021.
[62] H. Zhu, J. Xu, S. Liu, and Y. Jin, "Federated learning on non-IID data: A survey," *Neurocomputing*, vol. 465, pp. 371–390, 2021. https://doi.org/10.1016/j.neucom.2021.01.119

[63] D. Bestari and A. Wibowo, "IoT Based Real-Time Weather Monitoring System Using Telegram Bot and Thingsboard Platform", *International Journal of Interactive Mobile Technologies (iJIM)*, vol. 17, no. 06, pp. 4-19, 2023. https://doi.org/10.3991/ijim.v17i06.34129

[64] R. D. Handayani, Z. Jamal, M. Alkahfiansyah, L. Rosmalia, N. H Sudibyo, and R. Herwanto, "Intelligent and Secure Package Receiver System Utilizing Internet of Things (IoT) Technology*", Vokasi Unesa Bull. Eng. Technol. Appl. Sci.*, vol. 2, no. 3, pp. 581–592, Sep. 2025. https://doi.org/10.26740/vubeta.v2i3.38254

[65] M. M. Mansour, A. M. lafta, A. M. Salman, and H. S. Salman, "Exploring the Impact of AI and IoT on Production Efficiency, Quality Precision, and Environmental Sustainability in Manufacturing", *Vokasi Unesa Bull. Eng. Technol. Appl. Sci.*, vol. 2, no. 2, pp. 342–355, Jun. 2025. https://doi.org/10.26740/vubeta.v2i2.38200

BIOGRAPHIES OF AUTHORS

Mohammed Ajuji is a lecturer in the Department of Computer Science, Gombe State University, Gombe, Nigeria. He received his B.Sc. in Computer Science from Gombe State University in 2012 and his M.Sc. in Computer Science from Abubakar Tafawa Balewa University, Bauchi in 2023. He is currently pursuing a Ph.D. in Computer Science at Modibbo Adama University, Yola, Nigeria. His research interests include artificial intelligence, machine learning, deep learning, and computer vision, with applications in smart systems and sustainable development. He can be contacted at email: majuji@gsu.edu.ng.

Yusuf Musa Malgwi is a Senior Lecturer and Head of the Department of Computer Science at Modibbo Adama University, Yola, Nigeria. He obtained his B.Tech. in Computer Science from the Federal University of Technology, Yola (now Modibbo Adama University) in 2006 and his M.Sc. in Computer Science from Adamawa State University, Mubi in 2014. He earned his Ph.D. in Computer Science from Modibbo Adama University, Yola in 2019. His research interests include machine learning, artificial intelligence, medical informatics, and computational intelligence, with applications in healthcare systems, cybersecurity, and intelligent decision support. He can be contacted at email: yusuf.malgwi@mautech.edu.ng.

Asabe Sandra Ahmadu is a Professor in the Department of Computer Science, Modibbo Adama University, Yola, Nigeria. She obtained her B.Tech. in Computer Science from Federal University of Technology, Yola (now Modibbo Adama University), and her M.Sc. and Ph.D. in Computer Science from Abubakar Tafawa Balewa University, Bauchi. Her research interests include data mining, machine learning, artificial intelligence, and image processing, with applications in education, health informatics, and remote sensing. She can be contacted at email: asabe.ahmadu@mautech.edu.ng.