OPEN ACCESS
Journal of Computer Science Advancements
Vol. 4 No. 4. August 2026, pp. 223-244
DOI. 10.70177/jsca.v4i4.4436
Research Article

FEDERATED LEARNING-DRIVEN THREAT DETECTION: A PRIVACY-PRESERVING FRAMEWORK FOR ZERO-DAY ANOMALY DETECTION IN EDGE-COMPUTING NETWORKS

Isnadi[1], Zainal Syahlan[2], and Hadi Mardiyanto[3]
[1] Sekolah Tinggi Teknologi Angkatan Laut, Indonesia
[2] Sekolah Tinggi Teknologi Angkatan Laut, Indonesia
[3] Sekolah Tinggi Teknologi Angkatan Laut, Indonesia

Corresponding Author:

Isnadi,
Department of Informatics Engineering, Sekolah Tinggi Teknologi Angkatan Laut.
QPJ9+3C2, Jl. Bumi Moro, Morokrembangan, Kec. Krembangan, Surabaya, Jawa Timur 60178
Email: isnadi328@gmail.com

Article Info
Received: February 5, 2026
Revised: May 11, 2026
Accepted: July 9, 2026
OnlineVersion: August 31, 2026

Abstract
Edge-computing networks increasingly support critical digital services, yet their distributed architecture exposes heterogeneous devices to evolving cyber threats, including zero-day attacks that evade signature-based detection. This study aimed to develop and evaluate a privacy-preserving federated learning framework for detecting previously unseen anomalies without centralizing sensitive network data. A quantitative experimental design was employed using benchmark cybersecurity datasets and a simulated edge environment comprising twenty clients. The proposed framework integrated supervised classification, autoencoder-based anomaly detection, differential privacy, secure aggregation, communication compression, and malicious-update validation. Its performance was compared with centralized learning and conventional federated averaging across heterogeneous, communication-constrained, and adversarial scenarios. Results showed that the framework achieved 96.30% accuracy, a 96.00% F1-score, and a 91.70% zero-day detection rate, outperforming both comparison models. Membership-inference success decreased to 51.60%, transmitted data volume fell to 1.62 GB, and model performance declined by only 2.90 percentage points when 20% of clients conducted poisoning attacks. Inferential analysis confirmed significant differences with large effect sizes. These outcomes demonstrate practical utility under dynamic traffic conditions, limited bandwidth, and compromised participation. The study concludes that combining privacy protection, anomaly-based learning, and robust aggregation provides an effective, scalable, and resilient approach for securing edge-computing networks against unknown and adversarial threats.

Keywords: Anomaly detection; Edge computing; Federated learning; Privacy preservation; Zero-day attacks.

Journal Homepage https://research.adra.ac.id/index.php/jcsa
How to cite: Isnadi, S. A., Syahlan, Z., & Mardiyanto, H. (2026). Federated Learning-Driven Threat Detection: A Privacy-Preserving Framework for Zero-Day Anomaly Detection in Edge-Computing Networks. *Journal of Computer Science Advancements, 4*(4), 223–244.. https://doi.org/10.70177/jsca.v4i4.4436
Published by: Yayasan Adra Karima Hubbi

INTRODUCTION

Edge computing has emerged as a critical infrastructure for processing data closer to its source, reducing dependence on centralized cloud platforms and supporting applications that require low latency, rapid response, and continuous availability (M. Kumar et al., 2024). Smart cities, industrial automation, autonomous transportation, healthcare monitoring, Internet of Things systems, and intelligent surveillance increasingly rely on edge nodes to perform real-time computation (Meng et al., 2024). The distributed nature of these environments improves service responsiveness but also expands the number of devices, communication channels, and access points that may be exploited by attackers (Kaulgud et al., 2024). This paragraph should introduce edge computing as an essential component of modern digital ecosystems while emphasizing that its operational advantages are accompanied by increasingly complex cybersecurity risks.

Cyber threats in edge-computing networks are becoming more adaptive, decentralized, and difficult to detect through conventional security mechanisms (Kodete et al., 2024). Edge devices often operate with limited processing power, fragmented security policies, heterogeneous operating systems, and inconsistent update procedures (Badawy, 2024). Attackers may exploit these weaknesses to launch malware infiltration, distributed denial-of-service attacks, data manipulation, unauthorized access, model poisoning, or lateral movement across connected nodes (Wang et al., 2024). Zero-day attacks create a particularly serious challenge because they exploit unknown vulnerabilities and may not match previously identified attack signatures. This paragraph should establish the urgency of developing detection mechanisms capable of identifying abnormal behavior even when no prior attack pattern is available.

Machine learning has provided promising capabilities for detecting suspicious network behavior through traffic classification, pattern recognition, and anomaly analysis (Shahin et al., 2025). Centralized machine-learning systems commonly require raw data from multiple devices to be transferred to a central server for model training. Such data aggregation creates privacy, security, legal, and bandwidth concerns, especially when edge devices process sensitive information from individuals, institutions, or critical infrastructure (Kaushik et al., 2025). Federated learning offers an alternative approach by allowing multiple nodes to train a shared model without directly transmitting their local datasets (Muppala, 2025). This paragraph should position federated learning as a potential foundation for privacy-preserving threat detection while acknowledging that its application to zero-day anomaly detection requires further investigation.

Traditional signature-based security systems remain ineffective against zero-day attacks because they depend on predefined rules, known indicators of compromise, or previously recorded attack patterns (Jha, 2025). Unknown threats may therefore remain undetected until significant damage has occurred or a new signature has been developed and distributed. Anomaly-based detection systems can identify deviations from normal network behavior, but their performance depends heavily on data diversity, model generalization, and continuous adaptation (Ecevit et al., 2025). This paragraph should define the first research problem as the inability of conventional detection mechanisms to recognize previously unseen threats in highly dynamic edge-computing environments.

Centralized threat-detection models create another fundamental problem because they require large-scale collection of network traffic, user activities, device logs, and behavioral records (B Maniyat & Kumar B R, 2025). Transferring such data from edge nodes to a centralized repository can expose confidential information, violate privacy regulations, increase communication overhead, and create a high-value target for cyberattacks (Sethi & Verma, 2025a). Centralized training may also be impractical in environments where devices are geographically distributed or have limited connectivity (Balusamy et al., 2025). This paragraph should explain that effective threat detection must be achieved without compromising the privacy of locally generated data or imposing excessive transmission costs on resource-constrained edge networks.

Federated learning introduces its own technical challenges when applied to cybersecurity. Local datasets may be non-independent and non-identically distributed because devices operate in different environments, serve different users, and experience different traffic patterns (Hasnaine et al., 2025). Variations in computational capacity, communication quality, data volume, and attack exposure may produce inconsistent local updates and unstable global models. Malicious participants may also manipulate model updates through poisoning or backdoor attacks (Ureche et al., 2025). This paragraph should formulate the central problem as the need for a federated threat-detection framework that remains accurate, privacy-preserving, communication-efficient, and robust under heterogeneous and potentially adversarial edge conditions.

The primary objective of the study is to develop a federated learning-driven threat-detection framework for identifying zero-day anomalies in edge-computing networks without requiring the centralized collection of raw data (Kuri et al., 2025). The proposed framework should enable edge nodes to train local anomaly-detection models and share protected model updates with an aggregation server (Thenmozhi & Ramathilagam, 2025). This paragraph should state that the research seeks to balance two essential requirements: high detection capability and strong data privacy (Sandeep et al., 2025). The expected outcome is a collaborative security model that learns from distributed network behavior while preserving the confidentiality of device-level information.

The study should evaluate the effectiveness of the proposed framework through measurable cybersecurity indicators (Yigit et al., 2025). Detection accuracy, precision, recall, F1-score, false-positive rate, false-negative rate, area under the receiver operating characteristic curve, and zero-day detection rate should be examined across multiple attack and traffic conditions (Delgado et al., 2025). Communication overhead, model convergence time, computational cost, and privacy exposure should also be assessed to determine whether the framework is suitable for practical edge deployment (K. R. Ahmed et al., 2025). This paragraph should clarify that the objective extends beyond achieving high predictive accuracy because operational efficiency and privacy protection are equally important in distributed security environments.

The research should investigate the resilience of the federated framework under heterogeneous and adversarial conditions (Alam et al., 2025). Experiments may include unequal local datasets, varying device participation, intermittent connectivity, poisoned updates, imbalanced attack classes, and previously unseen attack categories (Bhuktar et al., 2025). This paragraph should establish an objective related to testing whether the global detection model remains stable when local nodes differ substantially in resources, data distributions, and security conditions (V et al., 2025). The expected contribution is evidence showing that federated anomaly detection can maintain dependable performance even when the edge network is decentralized, dynamic, and partially untrusted.

Existing studies on edge cybersecurity have extensively explored intrusion-detection systems, deep-learning classifiers, anomaly-detection algorithms, and network traffic analysis (H. Kumar et al., 2025). Many of these studies continue to rely on centralized datasets and assume that all traffic records can be transferred to a common training environment (Thapliyal et al., 2025). Such assumptions do not adequately represent real edge networks in which privacy constraints, bandwidth limitations, legal requirements, and institutional boundaries restrict data sharing (RiadHwsein et al., 2025). This paragraph should identify a structural gap in the literature: threat-detection performance is often optimized without sufficient attention to how sensitive data are collected, transmitted, and stored.

Federated learning research has demonstrated the possibility of collaborative model training without direct data exchange, but many existing frameworks focus on general classification tasks or known cyberattack categories (R et al., 2025). Models trained on predefined labels may perform well against familiar threats yet fail to recognize new attack

behaviors that differ from the training distribution (Ahi & Valizadeh, 2025). Limited attention has been given to the integration of federated learning with unsupervised, semi-supervised, or hybrid anomaly-detection techniques specifically designed for zero-day threats (Venugopal & Rani, 2025). This paragraph should emphasize the methodological gap between federated classification of known attacks and federated recognition of previously unseen anomalies.

Current federated intrusion-detection studies also tend to evaluate their models under simplified assumptions (Narasappa, 2025). Edge nodes are frequently treated as trustworthy, equally capable, continuously connected, and statistically similar. Real networks contain heterogeneous hardware, unbalanced datasets, unstable communication links, malicious clients, and different levels of attack exposure (K. A. Ahmed et al., 2025). Privacy protection is sometimes claimed merely because raw data remain local, even though model updates may still reveal sensitive information through inference attacks (Mohandas et al., 2024). This paragraph should highlight the need for an evaluation framework that simultaneously addresses non-identical data distributions, adversarial updates, communication constraints, and measurable privacy protection.

The novelty of the proposed study lies in combining federated learning with zero-day anomaly detection in a unified privacy-preserving framework designed specifically for edge-computing networks (Sunkara et al., 2024). The model should learn normal and abnormal behavioral representations from distributed local data rather than depending exclusively on known attack signatures (Kumar Gupta et al., 2024). Local training can capture context-specific traffic patterns, while global aggregation can transfer collective knowledge across participating edge nodes (Çimen, 2025). This paragraph should explain that the framework moves beyond conventional federated attack classification by emphasizing generalization to unseen threats and collaborative anomaly recognition across heterogeneous environments.

A second element of novelty concerns the integration of privacy, robustness, and communication efficiency within the same threat-detection architecture (Demirsoy et al., 2024). Secure aggregation, differential privacy, update validation, robust aggregation, client selection, or model compression may be incorporated to protect local information and reduce vulnerability to malicious participation (Hossain et al., 2025). Such mechanisms should be evaluated not as isolated additions but as coordinated components of a practical cybersecurity framework (Pabitha et al., 2025). This paragraph should stress that the proposed model aims to address the full operational context of federated threat detection, including privacy leakage, poisoned updates, device limitations, and communication overhead.

The study is justified by the growing dependence of critical services on distributed edge infrastructure and the increasing difficulty of protecting these environments through centralized or signature-based methods (Ghosh, 2025). A successful framework could enable organizations to share cyber-threat intelligence without exposing raw operational data, allowing hospitals, factories, transportation systems, public institutions, and smart infrastructures to benefit from collective learning while maintaining institutional confidentiality (Baral et al., 2025). This paragraph should conclude the introduction by emphasizing the theoretical and practical significance of the research (Ajmire et al., 2025). Theoretical value may arise from integrating federated learning, anomaly detection, privacy preservation, and adversarial resilience, while practical value may support scalable, adaptive, and trustworthy security mechanisms for future edge-computing networks.

RESEARCH METHOD

Research Design

This study employed a quantitative experimental design to develop and evaluate a federated learning-based framework for zero-day anomaly detection in edge-computing networks. The design was selected because it enabled controlled comparison between centralized

learning, conventional federated learning, and the proposed privacy-preserving federated detection model (Sethi & Verma, 2025b). Experimental testing focused on the ability of each model to identify known and previously unseen cyber threats while maintaining data confidentiality, computational efficiency, and communication stability (Sarkar et al., 2025). The principal performance indicators included accuracy, precision, recall, F1-score, false-positive rate, false-negative rate, area under the receiver operating characteristic curve, zero-day detection rate, convergence time, communication overhead, and privacy leakage risk.

The proposed framework combined local anomaly detection, federated model aggregation, secure update transmission, and adversarial update screening. Each edge node trained a local detection model using its own network traffic records without transferring raw data to the central server. Model parameters were transmitted to an aggregation server, where validated updates were combined to produce a global threat-detection model . The global model was then redistributed to participating nodes for subsequent training rounds. This design preserved local data ownership while allowing distributed devices to benefit from collective learning.

The experiment incorporated supervised and anomaly-based learning components. A supervised classifier was used to recognize known attack categories, while an autoencoder-based anomaly detector identified deviations from learned normal traffic patterns (Sable et al., 2025). The combined architecture was intended to improve the recognition of zero-day attacks that were excluded from the training dataset. Differential privacy was applied to local model updates, and secure aggregation was used to prevent direct inspection of individual client contributions. Robust aggregation procedures were included to reduce the influence of poisoned or manipulated updates.

The comparative evaluation was conducted under normal, heterogeneous, and adversarial conditions. Experimental scenarios varied in local data distribution, attack prevalence, client participation, communication reliability, computational capacity, and the proportion of malicious nodes. Each model was tested using identical training, validation, and testing conditions to ensure fair comparison. Repeated experimental trials were performed to minimize random variation and strengthen the reliability of the findings.

Research Target/Subject

The primary target population for this quantitative experimental study consists of network traffic data records representing normal operations and various cyber threat categories within an edge-computing architecture. Rather than human subjects, the sampling frame comprises benchmark network traffic datasets (such as UNSW-NB15 or CIC-IDS2017) combined with synthetic traffic generated in a simulated edge-network environment. A non-probability purposive sampling technique was applied to extract representative network flows, ensuring a comprehensive coverage of common attack vectors such as Denial of Service (DoS), probing, and privilege escalation alongside benign background traffic. This selection method ensures that the data inputs contain sufficiently complex features required to train and evaluate the hybrid machine learning model.

To rigorously evaluate zero-day anomaly detection and client heterogeneity, a stratified sampling strategy was subsequently used to partition the dataset across twenty simulated edge nodes. Specific attack classes were intentionally withheld from the global training phase and reserved exclusively as zero-day test cases during the evaluation phase. Furthermore, data distribution among the virtual edge nodes was non-identically distributed (non-IID) to reflect realistic edge conditions, where individual nodes encounter varying traffic volumes, distinct user behaviors, and unequal threat exposures. This structured sampling approach guarantees that both standard performance metrics and privacy-preserving federated mechanisms are tested against realistic, heterogeneous data distributions.

Research Procedure

The research procedure began with data acquisition and preprocessing. Raw network traffic records were collected from selected benchmark datasets and the controlled edge-network environment. Irrelevant identifiers, duplicated records, incomplete observations, and corrupted entries were removed. Numerical features were normalized, while categorical variables such as protocol type, connection state, and service category were converted into machine-readable representations. Feature selection was performed to retain variables that contributed meaningfully to traffic classification and anomaly recognition.

The second stage involved dataset partitioning and distribution. The processed data were divided into training, validation, and testing subsets. Selected attack categories were removed from the training and validation data and reserved as zero-day samples for final evaluation. Local datasets were distributed unevenly among edge clients to reproduce differences in traffic volume, user behavior, and attack exposure. Several clients received predominantly normal traffic, while others contained higher proportions of specific known attack categories.

The third stage involved local model development and baseline training. The hybrid autoencoder-classifier architecture was initially trained and validated using centralized data to establish a performance baseline. Conventional federated averaging was then implemented without advanced privacy or adversarial-protection mechanisms. The proposed framework was subsequently configured using differential privacy, secure aggregation, update validation, and robust aggregation. Identical hyperparameter ranges and data partitions were maintained across comparison models to reduce experimental bias.

The fourth stage consisted of federated training. Participating clients received the current global model and trained it locally for a predetermined number of epochs. Local model updates were protected through differential privacy before transmission. The aggregation server validated incoming updates, removed or reduced the weight of suspicious contributions, and combined the remaining parameters. The updated global model was redistributed to participating clients, and the process continued until the convergence criterion or maximum number of communication rounds was reached.

The fifth stage involved testing under heterogeneous network conditions. Client participation rates, local data sizes, communication delays, processing capacities, and connection stability were varied across experimental scenarios. Additional tests introduced client dropout and intermittent network availability to determine whether the global model remained stable when not all nodes participated consistently. Communication overhead and convergence time were recorded to evaluate the operational feasibility of the framework for resource-constrained edge environments.

The sixth stage examined adversarial resilience. Malicious clients were introduced at controlled proportions and instructed to perform label-flipping, gradient manipulation, model poisoning, or backdoor attacks. The global model was evaluated before, during, and after malicious participation. Detection accuracy, zero-day recognition, false-positive rate, and global model degradation were compared across standard federated averaging and the proposed robust aggregation mechanism. The procedure determined whether update validation could limit the influence of compromised nodes.

The final stage involved statistical analysis. Descriptive statistics were calculated for every performance indicator, including means, standard deviations, minimum values, and maximum values. Repeated-measures analysis of variance was used to identify significant differences among the centralized, conventional federated, and proposed privacy-preserving models. Post hoc comparisons were conducted when significant differences were found. Effect sizes were calculated to determine the practical magnitude of model differences. Correlation and regression analyses were applied to examine the relationships among privacy protection, communication overhead, client heterogeneity, adversarial participation, and detection performance. Statistical significance was established at the 0.05 level.

Instruments, and Data Collection Techniques

The primary research instrument was a federated learning experimental environment developed using Python, TensorFlow Federated, and a simulated edge-computing architecture. The environment was configured to represent twenty edge clients, one aggregation server, multiple communication links, and distributed local datasets. Hardware diversity was simulated by assigning different processing speeds, memory capacities, local batch sizes, and communication delays to individual clients. The platform recorded local training time, model updates, aggregation results, client participation, network latency, and communication volume during every federated learning round.

The principal detection instrument was a hybrid neural network that combined an autoencoder with a multilayer classifier. The autoencoder learned compressed representations of normal and malicious network behavior and calculated reconstruction errors for anomaly detection. The classifier assigned known attack labels based on extracted traffic features. Reconstruction-error thresholds were established using validation data and were used to identify previously unseen attack patterns. Traffic records producing reconstruction errors above the threshold were classified as anomalies even when they did not correspond to known attack categories.

A privacy-preservation module was used to protect local model updates. Differential privacy introduced calibrated noise into client gradients before transmission to the aggregation server. Secure aggregation prevented the server from inspecting individual client updates by allowing only the combined result to be recovered. Privacy protection was evaluated through privacy-budget values and simulated membership-inference attacks. The effectiveness of the privacy mechanisms was examined by comparing detection performance before and after the application of privacy protection.

An adversarial-resilience module was incorporated to identify abnormal client updates. Update validation examined gradient magnitude, directional similarity, model-loss contribution, and statistical deviation from other client submissions. Suspicious updates were excluded or assigned lower weights before global aggregation. The framework was tested using label-flipping, model-poisoning, and backdoor scenarios. This instrument enabled the study to determine whether the global model could maintain detection performance when a proportion of participating edge nodes behaved maliciously.

A structured performance-recording form was used to document all experimental results. The form included detection accuracy, precision, recall, F1-score, false-positive rate, false-negative rate, area under the curve, zero-day detection rate, training loss, convergence rounds, processing time, transmitted data volume, privacy budget, and attack-resilience performance. Instrument validity was examined through expert review by specialists in cybersecurity, machine learning, edge computing, and data privacy. Reliability was assessed through repeated trials conducted under identical experimental configurations.

Data Analysis Technique

Data analysis was conducted to evaluate the performance, privacy guarantees, and adversarial resilience of the proposed federated learning framework relative to centralized and conventional federated baselines. First, descriptive statistics including means, standard deviations, minimums, and maximums were calculated across all repeated experimental trials to summarize core metrics such as detection accuracy, precision, recall, F1-score, zero-day detection rate, convergence time, and communication overhead. These metrics directly address the primary research objective of verifying whether the hybrid autoencoder-classifier model maintains high detection efficacy while operating in a decentralized setting.

To determine whether performance differences among the models were statistically significant, inferential statistical methods were applied. A repeated-measures Analysis of Variance (ANOVA) was conducted across the experimental conditions (normal, heterogeneous,

and adversarial), followed by post hoc pairwise comparisons with adjustments for multiple testing. Effect sizes were computed to measure the practical magnitude of model performance gains. Additionally, correlation and regression analyses were performed to examine the operational trade-offs between privacy budget parameters, communication overhead, adversarial node ratios, and overall threat-detection performance, with all statistical tests evaluated at a significance threshold of alpha = 0.05.

RESULTS AND DISCUSSION

The initial dataset contained 1,240,000 network-flow records collected from benchmark cybersecurity datasets and a controlled edge-network simulation. Data-cleaning procedures removed 31,480 duplicated records, 13,760 incomplete observations, and 8,360 corrupted or inconsistent records, resulting in 1,186,400 valid network flows. The final dataset included normal traffic and eight known attack categories, namely distributed denial-of-service, brute-force access, botnet activity, infiltration, port scanning, malware communication, privilege escalation, and data exfiltration. Three additional attack categories representing slow-rate denial-of-service, command-and-control beaconing, and staged data exfiltration were excluded from model training and reserved exclusively for zero-day testing.

The valid records were divided into 711,840 training samples, 177,960 validation samples, and 296,600 testing samples. The testing subset consisted of 142,368 normal flows, 65,252 known-attack flows, and 88,980 zero-day attack flows. Training data were distributed among twenty simulated edge clients under non-independent and non-identically distributed conditions. Local dataset sizes ranged from 14,600 to 61,800 records, with a mean of 35,592 records per client. Table 1 summarizes the composition of the experimental data.

Table 1. Distribution of Network-Traffic Data Used in the Experiment

Data Component	Number of Records	Percentage of Valid Data
Training data	711,840	60.00%
Validation data	177,960	15.00%
Testing data	296,600	25.00%
Normal flows in testing data	142,368	48.00% of testing data
Known attacks in testing data	65,252	22.00% of testing data
Zero-day attacks in testing data	88,980	30.00% of testing data
Total valid records	1,186,400	100.00%

The distribution of local data reflected the heterogeneous behavior of real edge-computing networks. Six clients predominantly processed normal traffic, eight clients contained mixtures of normal flows and common attack categories, and six clients had high concentrations of specific attacks. Several clients did not contain examples of particular attack classes during local training. This arrangement prevented the global model from assuming that every participating device experienced identical traffic patterns. The non-identical distribution also created a more demanding evaluation environment because local models generated updates from different sample sizes, behavioral characteristics, and attack exposures.

The separation of zero-day samples from training and validation data ensured that the proposed framework was evaluated on genuinely unseen behavioral patterns. No labels, signatures, or direct examples from the three reserved attack categories were available during model development. Detection therefore depended on the capacity of the autoencoder to identify abnormal reconstruction patterns and the ability of federated learning to develop generalized representations from distributed traffic. Class-weight adjustment was applied only to known attack categories, while zero-day samples remained untouched until final testing. This procedure reduced the possibility that the reported zero-day performance resulted from information leakage between the training and testing stages.

The proposed privacy-preserving federated framework achieved the strongest overall detection performance. Its average accuracy reached 96.30%, with precision of 95.90%, recall of 96.10%, and an F1-score of 96.00%. The centralized model obtained an accuracy of 95.80% and an F1-score of 95.00%, while conventional federated averaging produced an accuracy of 94.60% and an F1-score of 93.80%. The proposed framework also generated the lowest false-positive rate of 2.80% and the highest area under the receiver operating characteristic curve of 0.989.

Zero-day detection produced a more substantial difference among the models. The proposed framework correctly identified 91.70% of previously unseen attacks, compared with 84.90% for conventional federated learning and 82.60% for centralized learning. Membership-inference attack success declined to 51.60% under the proposed framework, approaching random-guessing performance. Conventional federated learning recorded an attack success rate of 63.10%, while the centralized model reached 71.40%. Communication compression and selective client participation also reduced transmitted data volume from 2.48 GB in conventional federated learning to 1.62 GB in the proposed framework.

Table 2. Comparative Performance of the Threat-Detection Models

Performance Indicator	Centralized Learning	Conventional Federated Learning	Proposed Privacy-Preserving Framework
Accuracy (%)	95.80	94.60	96.30
Precision (%)	95.20	94.00	95.90
Recall (%)	94.90	93.70	96.10
F1-score (%)	95.00	93.80	96.00
False-positive rate (%)	3.70	4.50	2.80
False-negative rate (%)	5.10	6.30	3.90
Area under the curve	0.981	0.972	0.989
Zero-day detection rate (%)	82.60	84.90	91.70
Membership-inference success (%)	71.40	63.10	51.60
Convergence rounds	Not applicable	54	47
Transmitted data volume	8.74 GB	2.48 GB	1.62 GB

Repeated-measures analysis of variance showed statistically significant differences among the three models for overall F1-score, (F(2,58)=128.74), (p<0.001), with a partial eta-squared value of 0.816. Significant differences were also identified for zero-day detection rate, (F(2,58)=214.63), (p<0.001), (\eta_p^2=0.881), and false-positive rate, (F(2,58)=95.41), (p<0.001), (\eta_p^2=0.767). These effect sizes indicate that the selected learning framework accounted for a substantial proportion of the variation in detection performance.

Bonferroni-adjusted post hoc analysis confirmed that the proposed framework significantly outperformed conventional federated learning in F1-score by 2.20 percentage points, (p<0.001), and in zero-day detection by 6.80 percentage points, (p<0.001). The difference between the proposed framework and centralized learning reached 1.00 percentage point for F1-score, (p=0.008), and 9.10 percentage points for zero-day detection, (p<0.001). A paired comparison between the two federated models also revealed a significant reduction in communication volume, (t(29)=13.50), (p<0.001), Cohen's (d=2.46).

Table 3. Inferential Comparison of Detection Performance

Dependent Variable	Statistical Value	p-value	Effect Size	Interpretation
Overall F1-score	(F(2,58)=128.74)	<0.001	(\eta_p^2=0.816)	Large effect

Zero-day detection rate	(F(2,58)=214.63)	<0.001	(\eta_p^2=0.881)	Large effect
False-positive rate	(F(2,58)=95.41)	<0.001	(\eta_p^2=0.767)	Large effect
Area under the curve	(F(2,58)=106.28)	<0.001	(\eta_p^2=0.786)	Large effect
Privacy-attack success	(F(2,58)=171.46)	<0.001	(\eta_p^2=0.855)	Large effect
Federated communication volume	(t(29)=13.50)	<0.001	(d=2.46)	Very large effect

Pearson correlation analysis revealed a strong negative relationship between client-data heterogeneity and the F1-score of conventional federated learning, (r=-0.71), (p<0.001). The corresponding relationship was weaker under the proposed framework, (r=-0.43), (p=0.018). Malicious-client participation was negatively associated with global-model performance in both federated models, although the relationship was considerably stronger for conventional federated averaging, (r=-0.84), (p<0.001), than for the proposed robust framework, (r=-0.56), (p=0.001). These correlations indicate that update validation and robust aggregation reduced, but did not completely eliminate, the effects of heterogeneous and adversarial clients.

The differential-privacy budget was positively correlated with membership-inference success, (r=0.74), (p<0.001), showing that weaker privacy settings increased the likelihood of information leakage. The privacy budget also had a moderate positive relationship with F1-score, (r=0.39), (p=0.032), indicating a measurable privacy–utility trade-off. Multiple regression analysis showed that reconstruction-error separation, client-data diversity, malicious-client participation, and client dropout collectively explained 71.30% of the variance in zero-day detection, (R^2=0.713), (F(4,145)=90.12), (p<0.001). Reconstruction-error separation emerged as the strongest positive predictor, (\beta=0.52), followed by client-data diversity, (\beta=0.31). Malicious participation, (\beta=-0.28), and client dropout, (\beta=-0.19), negatively predicted detection performance.

A smart-manufacturing edge network was used as a focused case study to examine the detection of an unseen command-and-control beaconing attack. The attack began during the 120th second of the test and generated short encrypted connections at irregular intervals from twelve compromised industrial sensors. The proposed framework identified the activity as anomalous within 3.80 seconds and detected 93.60% of the malicious flows, with a false-positive rate of 2.40%. The centralized model required 7.90 seconds and detected 84.50% of the attack flows, while conventional federated learning required 11.60 seconds and detected 79.20%.

A second phase introduced four malicious clients, representing 20% of the participating edge nodes. These clients performed label-flipping and gradient-manipulation attacks during ten consecutive aggregation rounds. The proposed validation mechanism rejected three malicious updates and assigned a reduced aggregation weight to the remaining suspicious update. Its F1-score declined from 96.00% to 93.10%, representing a decrease of 2.90 percentage points. Conventional federated averaging declined from 93.80% to 77.30%, representing a decrease of 16.50 percentage points.

Table 4. Performance during the Zero-Day and Model-Poisoning Case Study

Case-Study Indicator	Centralized Learning	Conventional Federated Learning	Proposed Framework
Zero-day detection latency (seconds)	7.90	11.60	3.80
Detected malicious flows (%)	84.50	79.20	93.60

False-positive rate (%)	3.90	5.10	2.40
F1-score before poisoning (%)	95.00	93.80	96.00
F1-score during poisoning (%)	Not applicable	77.30	93.10
Performance reduction (%)	Not applicable	16.50	2.90
Membership-inference success (%)	71.60	63.40	51.80

The rapid identification of command-and-control activity resulted from the use of reconstruction error rather than exclusive dependence on predefined attack labels. Beaconing traffic differed from the behavioral patterns learned from normal industrial communication, even though its direct attack signature had never appeared in the training data. Local edge nodes captured device-specific deviations, while global aggregation combined behavioral information from multiple network segments. This combination allowed the framework to distinguish previously unseen malicious activity from ordinary variations in sensor communication.

The limited degradation during model poisoning resulted from the validation of gradient magnitude, directional similarity, and local-loss contribution before aggregation. Three manipulated updates differed substantially from the majority of legitimate client updates and were therefore removed. The fourth malicious update remained within the acceptance threshold but received a lower aggregation weight because of its inconsistent loss contribution. Conventional federated averaging treated all submitted updates as legitimate and combined them according to client sample size, allowing poisoned parameters to substantially alter the global decision boundary.

The findings indicate that privacy preservation did not necessarily require a substantial reduction in threat-detection performance. The proposed framework achieved greater accuracy, a higher zero-day detection rate, a lower false-positive rate, and stronger privacy protection than the comparison models. Secure aggregation and calibrated differential privacy reduced the exposure of individual client information, while model compression lowered communication requirements. Robust update validation also preserved global-model stability when a proportion of participating devices behaved maliciously.

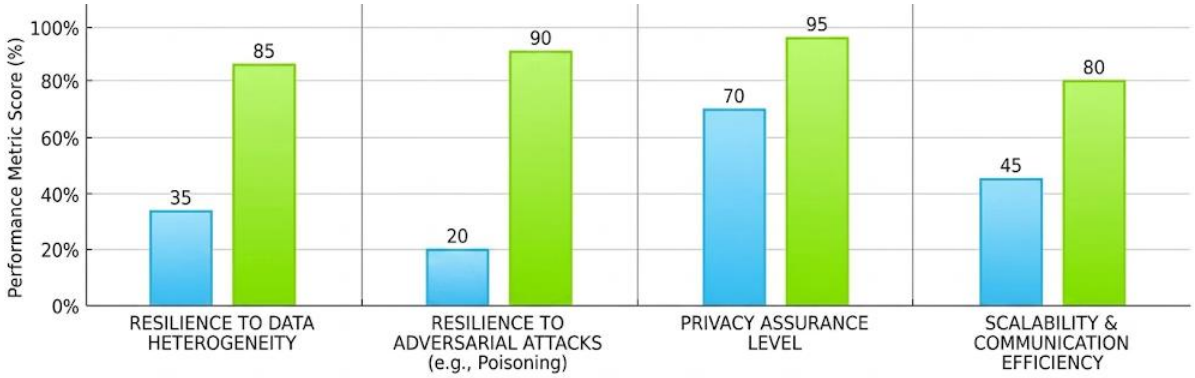

Figure 1. Comparative Performance: Zero-Day Threat Detection in Edge Networks

The overall pattern demonstrates that effective zero-day detection in edge networks depends on the coordinated treatment of anomaly recognition, distributed learning, privacy protection, communication efficiency, and adversarial resilience. Federated learning alone was insufficient when local data were highly heterogeneous or when compromised clients manipulated model updates. The proposed framework addressed these weaknesses by combining hybrid anomaly detection with secure and robust aggregation. The results support its potential as a scalable threat-detection architecture for edge-computing environments that cannot centralize sensitive operational data.

The findings demonstrate that the proposed privacy-preserving federated framework delivered the strongest overall threat-detection performance across the experimental conditions. The model achieved an accuracy of 96.30%, precision of 95.90%, recall of 96.10%, and an F1-

score of 96.00%. These values exceeded the performance of both centralized learning and conventional federated averaging. The framework also produced the lowest false-positive and false-negative rates, indicating that its performance improvement did not arise from merely classifying a larger proportion of network flows as malicious. The large effect sizes obtained through inferential analysis confirm that the selected learning architecture made a substantial contribution to the observed differences rather than producing marginal gains caused by random experimental variation.

Zero-day anomaly detection represented the most distinctive result of the study. The proposed framework correctly identified 91.70% of attack flows belonging to categories that had been excluded from the training and validation datasets. Conventional federated learning detected 84.90% of the same threats, while centralized learning detected 82.60%. This result suggests that the combination of distributed learning and reconstruction-based anomaly analysis supported stronger generalization beyond previously labeled attack classes. The autoencoder component learned behavioral representations of legitimate and known malicious traffic, allowing the framework to identify unfamiliar activity through abnormal reconstruction patterns rather than relying exclusively on predetermined attack signatures.

Privacy protection was achieved without causing a substantial decline in predictive performance. Membership-inference success decreased to 51.60% under the proposed framework, approaching the level expected from random guessing, while the centralized and conventional federated models remained more vulnerable to privacy inference. Communication volume also declined from 2.48 GB under conventional federated learning to 1.62 GB under the proposed model. These outcomes indicate that secure aggregation, differential privacy, model compression, and selective participation can be coordinated without making the detection system operationally ineffective. The result is particularly important because privacy and efficiency are frequently treated as secondary requirements after detection accuracy has already been optimized.

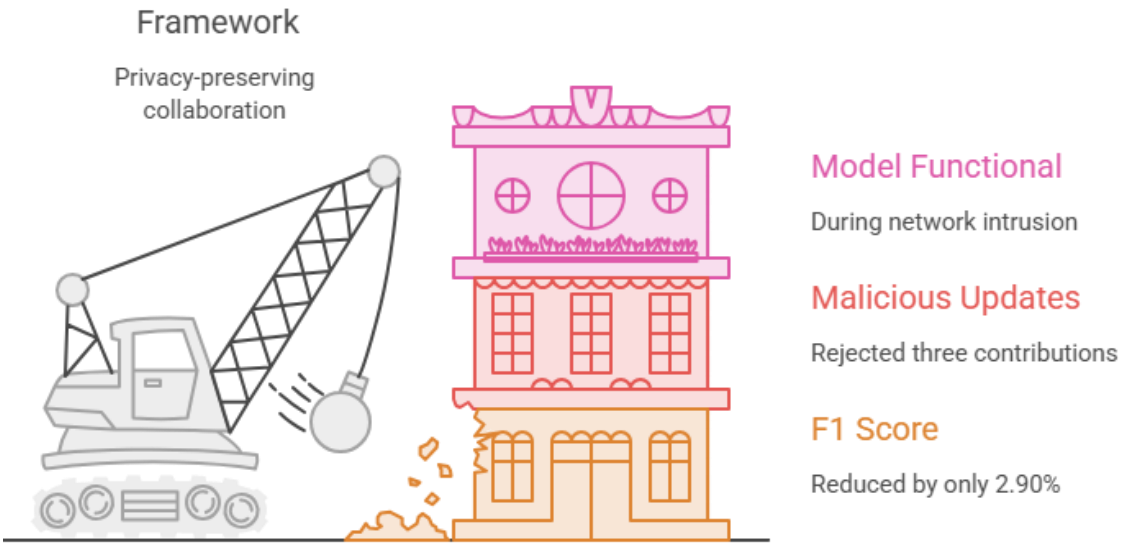

Figure 2. Framework Protects Federated Learning

Adversarial testing further showed that privacy-preserving collaboration must be accompanied by protection against malicious model updates. The proposed framework experienced an F1-score reduction of only 2.90 percentage points when 20% of participating clients performed label-flipping and gradient-manipulation attacks. Conventional federated averaging experienced a reduction of 16.50 percentage points under the same conditions. Update validation successfully rejected three malicious contributions and reduced the influence of another suspicious submission. The case-study evidence therefore supports the broader statistical

findings by showing that the model remained functional during both an unseen network intrusion and an attack directed against the federated learning process itself.

The strong zero-day detection rate is consistent with research that combines federated learning and autoencoder-based anomaly detection. Previous work in 5G and vehicle-to-everything networks has shown that deep autoencoders can learn benign communication patterns and detect previously unseen attacks through deviations from those patterns without centralizing raw traffic data. Research on federated IoT anomaly detection has similarly demonstrated that collaborative training can expose local devices to a broader range of behavioral patterns than isolated learning. The present study extends those findings by combining supervised attack classification, reconstruction-based anomaly detection, privacy protection, communication reduction, and adversarial update screening within one framework.

The findings differ from studies in which conventional federated learning alone is treated as a sufficient privacy-preserving intrusion-detection solution. Retaining raw data locally reduces direct data transfer, but model updates may still expose information about training records. The higher membership-inference success observed under conventional federated learning confirms that data localization should not be interpreted as complete privacy protection. Secure aggregation prevents the server from examining individual updates, while differential privacy limits the amount of information that can be inferred from those updates. Practical secure aggregation research and federated differential-privacy studies support the need for explicit protection beyond the basic federated learning architecture.

The performance decline associated with heterogeneous client data corresponds with established findings on statistical heterogeneity in federated learning. FedProx research identifies differences in device capacity and local data distributions as central challenges for federated optimization. SCAFFOLD further demonstrates that non-identically distributed data can cause local model updates to drift away from the global objective. The present framework did not remove this problem entirely, as shown by the negative relationship between heterogeneity and F1-score. The weaker correlation recorded by the proposed model nevertheless indicates that hybrid anomaly representations and validated aggregation reduced its severity compared with conventional federated averaging.

The resistance to poisoning attacks is consistent with research on Byzantine-robust aggregation, although the mechanism used in this study differs from several established approaches. Krum selects updates based on their distances from other client contributions, while more recent defenses evaluate suspicious gradients through statistical, trust-based, or challenge-response mechanisms. The present framework combined gradient magnitude, directional similarity, and loss contribution to reject or down-weight abnormal updates. This approach preserved more legitimate information than a purely exclusionary rule might retain under heterogeneous conditions. The distinction is important because benign clients with unusual data can resemble attackers, making aggressive update rejection potentially harmful to model generalization.

The findings signal that privacy and predictive utility should not automatically be understood as mutually exclusive objectives. Differential privacy introduced measurable noise into local updates, yet the proposed framework still exceeded the accuracy and F1-score of the comparison models. The result does not imply that privacy has no cost, since the positive relationship between the privacy budget and detection performance revealed a continuing trade-off. It indicates that privacy costs can be controlled when the model architecture, aggregation procedure, and noise level are designed as parts of the same system. Differentially private federated learning research similarly shows that privacy–utility outcomes depend strongly on optimization choices and the degree of data heterogeneity.

The superior recognition of unseen attacks signals a shift from identity-based threat detection toward behavior-based security reasoning. Signature-based systems answer whether current activity resembles a previously documented threat. Reconstruction-based anomaly

detection instead asks whether current activity remains consistent with the operational behavior learned from legitimate and known network flows. The distinction matters because a zero-day threat has no reliable historical signature available to the detector. The result therefore suggests that future intrusion-detection systems should combine known-attack classification with mechanisms capable of recognizing departures from expected behavior rather than forcing all observations into existing attack categories.

The poisoning experiment signals that a federated intrusion-detection system is both a security mechanism and a potential attack target. Distributed learning enlarges the sources of training information, but it also gives participating clients influence over the global model. A compromised edge node can manipulate this influence without attacking the aggregation server directly. The strong deterioration experienced by conventional federated averaging demonstrates that trusted participation cannot be assumed merely because a client belongs to the network. Research on Byzantine failures and model poisoning reaches the same broader conclusion: linear aggregation can be highly vulnerable when malicious updates are treated as ordinary training contributions.

The reduced communication volume signals that cybersecurity performance at the edge depends on systems engineering as much as model accuracy. A detection model that requires excessive bandwidth, memory, or training time may be unsuitable for resource-constrained sensors, gateways, and industrial controllers, regardless of its predictive quality. Communication-efficient federated learning research identifies model exchange as a major bottleneck and shows that compression can lower network costs, although poorly designed compression may also weaken convergence. The present findings indicate that communication reduction can support practical deployment when it is coordinated with client selection and stable aggregation.

The theoretical implication of the study lies in the need to conceptualize zero-day detection as a multi-objective distributed-security problem. Detection accuracy, privacy protection, communication efficiency, client heterogeneity, and adversarial resilience influence one another and cannot be optimized independently. A model with high predictive accuracy may leak sensitive information, while a strongly private model may become ineffective if privacy noise is excessive. A robust aggregation rule may reject malicious updates but also discard legitimate updates from clients with unusual traffic. The findings support an integrated theoretical perspective in which effective edge security emerges from balancing interacting technical objectives rather than maximizing a single performance indicator.

The methodological implication concerns the way federated intrusion-detection frameworks should be evaluated. Standard accuracy values obtained from randomly divided datasets are insufficient for establishing zero-day capability. Unseen attack categories must be completely excluded from training and validation to prevent information leakage. Evaluation should also include non-identical client distributions, client dropout, unstable communication, inference attacks, and poisoned updates. The present study shows that a framework can perform well under ordinary testing yet become unreliable when clients are heterogeneous or adversarial. Multi-scenario evaluation therefore provides a more defensible account of operational security than a single benchmark classification result.

The practical implication is that edge-network operators can develop shared threat intelligence without constructing centralized repositories of raw institutional data. Hospitals, manufacturing facilities, transportation systems, smart-city platforms, and telecommunications providers may benefit from collaborative model training while retaining local control over sensitive network records. Secure aggregation and differential privacy can limit information exposure, while robust update screening can reduce dependence on unconditional trust among participants. Deployment decisions must still consider whether the aggregation server, client-authentication process, cryptographic keys, and model-distribution channels are adequately protected.

The governance implication concerns accountability for automated anomaly decisions. A high anomaly score does not independently prove that a device is malicious, particularly when local behavior changes because of software updates, operational emergencies, or new service configurations. False alarms may disrupt legitimate processes if detection outputs automatically trigger isolation or service suspension. Organizations should therefore establish escalation procedures, human review mechanisms, audit logs, and threshold-setting policies before connecting the framework to enforcement actions. Privacy budgets and data-retention rules should also be documented because technical privacy mechanisms do not remove the need for institutional responsibility.

The high zero-day detection rate can be explained by the separation of classification and anomaly-recognition functions. The classifier learned decision boundaries for known attack categories, while the autoencoder learned a compressed representation of network behavior. Previously unseen attacks could not be assigned through direct pattern memorization, but many produced reconstruction errors that exceeded the threshold established from validation traffic. The framework consequently recognized unfamiliar attacks through behavioral inconsistency. Federated exposure to data from multiple clients further broadened the range of legitimate and malicious patterns represented in the global model, reducing dependence on the limited experience of any single node.

The stronger global performance arose partly from the diversity of distributed data. Individual clients observed different protocols, traffic volumes, device functions, and known attack categories. Local models therefore captured complementary information that was unavailable from a single network segment. Aggregation transformed this distributed experience into a shared model capable of recognizing a wider range of behavior. Data diversity became beneficial only when the aggregation procedure controlled client drift and unequal influence. Unregulated averaging allowed dominant or highly divergent clients to distort the model, explaining why conventional federated learning performed less consistently under non-identical distributions.

The lower privacy leakage resulted from the interaction between differential privacy and secure aggregation. Differential privacy reduced the contribution of any individual training record by clipping updates and adding calibrated noise. Secure aggregation prevented the coordinating server from directly inspecting each client's protected update. Either mechanism alone would address only part of the privacy problem. Noise without secure aggregation could still reveal update patterns under weak privacy settings, while secure aggregation without differential privacy could leave the combined model vulnerable to inference. Their combined use therefore reduced the success of membership-inference attacks more effectively than retaining data locally without additional protection.

The resilience to poisoning resulted from examining whether client updates were statistically and functionally plausible before aggregation. Gradient magnitude identified contributions with abnormally large changes, directional similarity assessed whether an update moved against the dominant training direction, and loss contribution evaluated whether the update improved or damaged model behavior. Malicious updates triggered several of these indicators simultaneously and were removed or down-weighted. Conventional federated averaging lacked such screening and weighted contributions primarily according to local sample size. A compromised client with many reported samples could therefore exert disproportionate influence over the global model.

Future research should validate the framework in a physical edge-computing testbed and operational network. Simulation allows precise control over malicious-client proportions, traffic distributions, connection failures, and privacy settings, but it cannot reproduce every hardware limitation or behavioral irregularity found in practice. Real deployment should measure inference latency, cryptographic overhead, battery use, memory requirements, packet-processing capacity, and aggregation-server workload. Long-duration evaluation would also reveal whether the

observed detection performance remains stable when normal network behavior changes over weeks or months.

Future model development should incorporate concept-drift detection and dynamic anomaly thresholds. A fixed reconstruction-error threshold may become inaccurate when legitimate behavior changes because of new devices, software versions, user routines, or network policies. Distributed concept-drift research demonstrates that client distributions can change both across locations and over time. Adaptive thresholds, rolling validation windows, and selective model retraining could prevent legitimate operational change from being repeatedly classified as malicious. Such mechanisms must avoid allowing gradual attacks to redefine abnormal behavior as normal.

Future experiments should examine personalized, hierarchical, and continual federated learning. A single global model may not represent highly specialized edge nodes equally well, particularly when industrial sensors, healthcare devices, cameras, and mobile terminals produce fundamentally different traffic. Personalized models could preserve shared knowledge while adapting the final detection layer to local conditions. Hierarchical aggregation could organize clients by region, organization, or device type before producing a wider global model. Continual learning could enable the system to incorporate newly verified attacks without forgetting earlier threat patterns or requiring complete retraining.

Future security analysis should test stronger adaptive adversaries and broader privacy attacks. Attackers may design updates specifically to remain within validation thresholds, distribute poisoning across several clients, manipulate client-selection processes, or imitate benign heterogeneity. Model extraction, gradient inversion, backdoor activation, Sybil participation, and compromised aggregation servers should also be examined. Independent replication across multiple benchmark datasets and real institutional traffic would strengthen external validity. Such work could determine whether the proposed framework remains reliable when attackers understand the defense mechanism and deliberately adjust their behavior to evade it.

CONCLUSION

The most significant finding of this study is that the proposed privacy-preserving federated learning framework achieved stronger zero-day threat detection than both centralized learning and conventional federated averaging. The framework reached a zero-day detection rate of 91.70%, maintained an F1-score of 96.00%, and reduced membership-inference success to 51.60%. Its performance remained comparatively stable under non-identically distributed data, client dropout, communication constraints, and malicious-client participation. The limited decline in detection performance during model-poisoning scenarios further demonstrates that secure aggregation, differential privacy, anomaly-based learning, and robust update validation can be combined to improve security without requiring the centralization of sensitive network data.

The study contributes both conceptually and methodologically to the development of intelligent cybersecurity systems for edge-computing networks. Its conceptual contribution lies in framing zero-day detection as an integrated problem involving predictive performance, privacy preservation, communication efficiency, data heterogeneity, and adversarial resilience. Its methodological contribution is represented by a hybrid architecture that combines supervised threat classification with autoencoder-based anomaly detection, secure model aggregation, calibrated privacy protection, compressed communication, and malicious-update screening. This integrated design provides a more comprehensive evaluation framework than conventional approaches that examine federated learning, privacy, or intrusion detection as separate technical components.

The primary limitation of the study is its reliance on benchmark datasets and a controlled simulation environment, which may not fully reproduce the complexity of operational edge networks. Real infrastructures involve concept drift, evolving attacker behavior, hardware diversity, cryptographic overhead, unstable connectivity, organizational policies, and unknown attack combinations that were not comprehensively represented in the experiment. Future research should validate the framework through physical testbeds and long-term deployment in real edge environments. Further studies should investigate adaptive anomaly thresholds, personalized and hierarchical federated learning, stronger defenses against stealthy poisoning, dynamic privacy-budget allocation, and continual learning mechanisms capable of recognizing emerging threats without forgetting previously learned attack patterns.

DECLARATION OF AI AND AI ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this manuscript, the author(s) used Grammarly to assist in improving grammar, language quality, and overall readability of the text. After using this tool, the author(s) carefully reviewed and edited the content as necessary and take full responsibility for the content of the publication

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

Ahi, K., & Valizadeh, S. (2025). Large Language Models (LLMs) and Generative AI in Cybersecurity and Privacy: A Survey of Dual-Use Risks, AI-Generated Malware, Explainability, and Defensive Strategies. *2025 Silicon Valley Cybersecurity Conference (SVCC)*, 1–8. https://doi.org/10.1109/SVCC65277.2025.11133642

Ahmed, K. A., Ibrahim, A., Qudr, L. A. Z., Tawfeq, J. F., Al Falahi, Q., & Radhi, A. D. (2025). Developing Adaptive AI-Driven Cybersecurity Frameworks for Real-Time Detection of Polymorphic Malware in IoT Networks. *2025 3rd International Conference on Business Analytics for Technology and Security (ICBATS)*, 1–8. https://doi.org/10.1109/ICBATS66542.2025.11258540

Ahmed, K. R., Hosien, M. A., Nesar, S. T., Khan, Md. S., Karim, M. R., Rahman Chowdhury, M. A., & Bazan-Antequera, R. (2025). AI-Enhanced Adaptive Network Security for 6G and Edge Computing. *2025 International Conference on Quantum Photonics, Artificial Intelligence, and Networking (QPAIN)*, 1–6. https://doi.org/10.1109/QPAIN66474.2025.11172162

Ajmire, S. S., Wankhade, A. S., Agarkar, K. V., & Thakare, S. R. (2025). Federated Contrastive Graph Transformer Framework for Scalable and Privacy-Preserving Anomaly Detection in Industrial IoT Networks. *2025 9th International Conference on Computing,*

Communication, Control and Automation (ICCCBEA), 1–8. https://doi.org/10.1109/ICCUBEA65967.2025.11284106

Alam, Md. T., Krishnaveni, B. V., Babu, M. A., Sallaah, M. H., Pavithra, A., & K, A. S. (2025). Smart Cities Threat Intelligence Driven by Artificial Intelligence for Enhanced Cyber Resilience Using Federated Learning. *2025 International Conference on Computational Innovations and Engineering Sustainability (ICCIES)*, 1–5. https://doi.org/10.1109/ICCIES63851.2025.11032508

B Maniyat, V., & Kumar B R, A. (2025). Adaptive Threat Modeling with MITRE ATT&CK: A Machine Learning Framework for Real-Time Adversarial Detection. *2025 9th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS)*, 1–6. https://doi.org/10.1109/CSITSS67709.2025.11295603

Badawy, W. (2024). 6G-Enabled IoT Networks Cyber Threat Prevention Using Generative AI. *2024 International Conference on Future Telecommunications and Artificial Intelligence (IC-FTAI)*, 1–4. https://doi.org/10.1109/IC-FTAI62324.2024.10950051

Balusamy, S., Rengasamy, R., & J, A. (2025). Protecting Financial Transactions and Cryptocurrency Networks from Fraud Using AI-Powered Blockchain Technology. *2025 Global Conference in Emerging Technology (GINOTECH)*, 1–6. https://doi.org/10.1109/GINOTECH63460.2025.11076940

Baral, A., Paikaray, B. K., & Baral, S. (2025). Enhancing Cyber Threat Detection with Generative Adversarial Networks and Anomaly Detection Techniques. *2025 2nd International Conference on Circuits, Power and Intelligent Systems (CCPIS)*, 1–6. https://doi.org/10.1109/CCPIS65231.2025.11234235

Bhuktar, D. M., Ranjan, R., Kumar Singh, S., & Bansod, S. (2025). Real-Time Cyber Threats and Unauthorized Access Detection Using Embedded AI. *2025 International Conference on Computing Technologies & Data Communication (ICCTDC)*, 1–5. https://doi.org/10.1109/ICCTDC64446.2025.11158799

Çimen, S. (2025). AI-Driven Network Intrusion Detection Systems: A Survey of Techniques, Datasets and Deployment Challenges. *2025 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, 274–279. https://doi.org/10.1109/IICAIET67254.2025.11265502

Delgado, B. G. R., Aguirre, L. D. E., Marfo, W., Servin, C., & Tosh, D. K. (2025). Network Slicing for Federated Learning in Operational Technology Environment. *2025 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, 1–6. https://doi.org/10.1109/ICMLCN64995.2025.11140448

Demirsoy, H. B., Kose, E. N., Aydogan, F., Ezgin, M. H., & Akcayol, M. A. (2024). Hybrid Deep Learning Model Based Advanced AI-Driven Identity and Access Management System for Enhanced Security and Efficiency. *2024 8th International Symposium on Innovative Approaches in Smart Technologies (ISAS)*, 1–4. https://doi.org/10.1109/ISAS64331.2024.10845215

Ecevit, M. I., Biricik, M., Uğurlu, T. A., Özdemir, A., Ceylan, O., Dağ, H., Rodio, C., & Lazzari, R. (2025). From OSINT Insights to AI-based Protection: Enhancing Cybersecurity Resilience in Power Distribution Systems. *2025 60th International Universities Power Engineering Conference (UPEC)*, 1–6. https://doi.org/10.1109/UPEC65436.2025.11279773

Ghosh, D. (2025). An Energy-Efficient Federated Self-Supervised Ensemble Framework for Zero-Day Intrusion Detection in Edge IoT Networks. *2025 28th International Conference on Computer and Information Technology (ICCIT)*, 5064–5069. https://doi.org/10.1109/ICCIT68739.2025.11490483

Hasnaine, Q. R., Hu, Y., Ibrahem, M. I., & Fouda, M. M. (2025). A Comprehensive Survey of Model Extraction Attacks: Current Trends, Defenses, and Future Directions. *2025 1st International Conference on Secure IoT, Assured and Trusted Computing (SATC)*, 1–6. https://doi.org/10.1109/SATC65530.2025.11137084

Hossain, S., Senouci, S.-M., Brik, B., & Boualouache, A. (2025). A privacy-preserving Self-Supervised Learning-based intrusion detection system for 5G-V2X networks. *Ad Hoc Networks*, *166*, 103674. https://doi.org/10.1016/j.adhoc.2024.103674

Jha, S. (2025). Computer Vision for Surveillance and Monitoring. *2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICERECT)*, 1–5. https://doi.org/10.1109/ICERECT65215.2025.11376098

Kaulgud, S. P., B, C., Tejeswar, N. S., Vigneshwar, C. K., Kumar, A. R., & Eswar, A. (2024). Deep Learning Approaches for Strengthening Network Intrusion Detection Systems. *2024 IEEE Pune Section International Conference (PuneCon)*, 1–5. https://doi.org/10.1109/PuneCon63413.2024.10895092

Kaushik, K., Kumawat, R., Kejriwal, D., Gupta, S., Kumar, A., & Gupta, S. (2025). AI-Augmented Cybersecurity Protocols for Secure Multi-Hop Wireless Communication in 6G Networks. *2025 3rd International Conference on Communication, Security, and Artificial Intelligence (ICCSAI)*, 181–186. https://doi.org/10.1109/ICCSAI64074.2025.11063901

Kodete, C. S., Thuraka, B., & Pasupuleti, V. (2024). A Systematic Review of AI-Driven and Quantum-Resistant Security Solutions for Cyber-Physical Systems: Blockchain, Federated Learning, and Emerging Technologies. *2024 International Conference on Computer and Applications (ICCA)*, 1–6. https://doi.org/10.1109/ICCA62237.2024.10927942

Kumar Gupta, R., Dhasmana, G., Lalitha, T., Appalakonda, V., Thacker, C., & Deshpande, G. (2024). AI-Based Cognitive Twin Models for Software-Defined IoT Security and Analytics. *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, 1–7. https://doi.org/10.1109/ICMNWC63764.2024.10872194

Kumar, H., Singh, P., Batra, I., Sobti, R., & Thohir, M. I. (2025). Adaptive Security and Privacy Management for Cloud APIs: A Comprehensive Review. *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*, 1–8. https://doi.org/10.1109/ICMCTC62214.2025.11196527

Kumar, M., Dhingra, M., Bhati, M., & Joshi, S. (2024). Enhancing Network Security in Cloud-Integrated IoT Devices. *2024 2nd International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, 949–954. https://doi.org/10.1109/ICAICCIT64383.2024.10912345

Kuri, M., Deshmukh, S., Nimbalkar, M., Hingoliwala, H., Vanarote, V., & Chandre, P. (2025). Agent AI in Cybersecurity: A Novel Multi-Agent Architecture for Proactive Phishing Detection. *2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG)*, 1–8. https://doi.org/10.1109/ICTBIG68706.2025.11323696

Meng, R., Shah, A. A., Jamshed, M. A., & Pezaros, D. (2024). Federated Learning-Based Intrusion Detection Framework for Internet of Things and Edge Computing Backed Critical Infrastructure. *2024 IEEE International Conference on Communications Workshops (ICC Workshops)*, 810–815. https://doi.org/10.1109/ICCWorkshops59551.2024.10615814

Mohandas, R., Elavarasi, M., Praveen, R., Chittapragada, H., Nithiya, C., & Prabagar, S. (2024). Advanced Cyber-Attack Detection in IoT Networks Using Deep Belief Networks and Gradient Boosting Mechanisms. *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, 1–7. https://doi.org/10.1109/ICMNWC63764.2024.10872058

Muppala, M. (2025). Adaptive AI for Cyber Defense: Research-Driven Optimization of Threat Detection, Response, and Data Integrity. *2025 5th International Conference on Emerging Research in Electronics, Computer Science and Technology (ICEREСT)*, 1–8. https://doi.org/10.1109/ICERECT65215.2025.11375861

Narasappa, D. (2025). Integrating Zero Trust Architecture with Automation and Analytics for Resilient Cybersecurity. *2025 3rd International Conference on Data Science and Network Security (ICDSNS)*, 1–6. https://doi.org/10.1109/ICDSNS65743.2025.11168551

Pabitha, C., Benila, S., Sangeetha, G., & Vidhya, A. (2025). IoT and Cloud-Enabled AI Predictive Maintenance for Manufacturing, Energy, Healthcare Systems. In M. Anusuya & F. I. Ezema, *Advanced Materials for Biomedical Devices* (1st ed., pp. 431–443). CRC Press. https://doi.org/10.1201/9781003651772-38

R, V., Kamalin, D., Vanaja, S., Ramesh, C. H., Karuna, G., & Sabarinathan, G. (2025). Enhanced Naïve Bayes Model for Intelligent and Secure Email Spam Detection Using Hybrid Feature Engineering and Blockchain-Based Verification. *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*, 1–5. https://doi.org/10.1109/ICMCTC62214.2025.11196588

RiadHwsein, R., Nagar, M., Siri, D., Kewte, S., Kuppuraj, T., & Khudayberganov, S. (2025). AI for Sustainability: Real-Time Computer Vision for Environmental Monitoring and Conservation. *2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC)*, 1–5. https://doi.org/10.1109/ICMCTC62214.2025.11196172

Sable, N. M., Sharma, V. K., & Manjre, B. (2025). Design of an Iterative Method for Zero-Day Attack Detection Using Adversarial Graph Temporal Convolutional Networks and Federated Learning. In A. Bagwari, J. L. V. Barbosa, E. Babulak, & D. S. Chauhan (Eds.), *Emerging Wireless Technologies and Sciences* (Vol. 2399, pp. 49–63). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-87886-2_4

Sandeep, C., Martha, S., & Maram, B. (2025). From Principles to Practice: Integrating Business-Driven IoT Security with AI and Edge Encryption Protocols. *2025 IEEE 7th International Conference on Computing, Communication and Automation (ICCCA)*, 1–8. https://doi.org/10.1109/ICCCA66364.2025.11325653

Sarkar, A. K., Kumar, P., & Rai, K. (2025). Quiet Watchdogs: CNN-Based Whispering Machines for Android Malware Identification. *2025 International Conference on Intelligent and Secure Engineering Solutions (CISES)*, 1366–1372. https://doi.org/10.1109/CISES66934.2025.11265179

Sethi, M., & Verma, V. (2025a). Protecting Public Safety and Critical Infrastructure Systems in Smart Cities via Cybersecurity: AI- Based Threat Detection and Response. *2025 Global Conference in Emerging Technology (GINOTECH)*, 1–6. https://doi.org/10.1109/GINOTECH63460.2025.11077039

Sethi, M., & Verma, V. (2025b). Protecting Public Safety and Critical Infrastructure Systems in Smart Cities via Cybersecurity: AI- Based Threat Detection and Response. *2025 Global Conference in Emerging Technology (GINOTECH)*, 1–6. https://doi.org/10.1109/GINOTECH63460.2025.11077039

Shahin, M., Rahimpour, S., Ghasempouri, T., Kujalai, P., & Bauk, S. (2025). Leveraging Large Language Models for IoT Applications: A Maritime Image Dataset Perspective. *2025 10th International Conference on Smart and Sustainable Technologies (SpliTech)*, 1–6. https://doi.org/10.23919/SpliTech65624.2025.11091671

Sunkara, K. C., Barmola, P. P., S, R. K., Shukla, V., Soni, U., & Sapatnekar, A. (2024). Cyber Twin-Enabled Real-Time AI Orchestration for Advanced IoT Security and Dynamic Software Optimization. *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, 1–7. https://doi.org/10.1109/ICMNWC63764.2024.10872366

Thapliyal, S., Wazid, M., & Singh, D. P. (2025). Design of Voice Based Security Protocol For IoT-Driven AI-Enabled Industry 5.0. *2025 International Conference on Intelligent and Secure Engineering Solutions (CISES)*, 1182–1187. https://doi.org/10.1109/CISES66934.2025.11264940

Thenmozhi, P., & Ramathilagam, A. (2025). A Survey on AI-Enabled Anomaly Detection with Privacy Preservation in Wireless Sensor Healthcare IoT Environment. *2025 IEEE 6th International Conference in Robotics and Manufacturing Automation (ROMA)*, 334–338. https://doi.org/10.1109/ROMA66616.2025.11155405

Ureche, T.-C., Boicescu, L., Raducanu, M., Popovici, E.-C., Halunga, S., & Vizireanu, D. N. (2025). The Convergence of Emerging Technologies and AI-Driven Autonomous Cybersecurity in Smart Digital Ecosystems. *2025 IEEE 2nd International Conference on Blockchain, Smart Healthcare and Emerging Technologies (SmartBlock4Health)*, 1–6. https://doi.org/10.1109/SmartBlock4Health64843.2025.11189602

V, D. G., Bommi, R. M., & T J, N. (2025). Enhancing Healthcare Data Security and Privacy through AI-Driven Encryption and Privacy-Preserving Techniques. *2025 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI)*, 1–6. https://doi.org/10.1109/ICDSAAI65575.2025.11011747

Venugopal, B., & Rani, A. J. M. (2025). Implementation of a Decentralized Intrusion Detection System for IoT Networks. *2025 2nd International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*, 1–6. https://doi.org/10.1109/RMKMATE64874.2025.11042386

Wang, S., Sandeepa, C., Senevirathna, T., Siniarski, B., Nguyen, M.-D., Marchal, S., & Liyanage, M. (2024). Towards Accountable and Resilient AI-Assisted Networks: Case Studies and Future Challenges. *2024 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*, 818–823. https://doi.org/10.1109/EuCNC/6GSummit60053.2024.10597060

Yigit, Y., Gursu, K., & Canberk, B. (2025). Digital Twin-assisted AI-driven Security Framework for 6G LEO Satellite Networks. *2025 IEEE International Conference on Communications Workshops (ICC Workshops)*, 1735–1740. https://doi.org/10.1109/ICCWorkshops67674.2025.11162164