Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 14, Issue. 8, August 2025, pg.58 – 70

AI-Powered Federated Learning Framework for Ransomware Detection in IoT Edge Environments

Ahmed Rabed

Faculty of Computers and Informatics, Thamar University, Yemen
ahmed.rabed@tu.edu.ye

DOI: https://doi.org/10.47760/ijcsmc.2025.v14i08.006

Abstract: The rapid proliferation of Internet of Things (IoT) devices in edge environments has heightened the risk of ransomware attacks targeting resource-constrained systems. Centralized detection approaches often compromise privacy and suffer from scalability and latency issues. This study proposes a decentralized, privacy-preserving ransomware detection framework that leverages federated learning (FL) to enable collaborative model training across distributed IoT devices without exposing raw data. The architecture integrates differential privacy and homomorphic encryption to enhance confidentiality and uses continual learning to support adaptation to evolving threats. Evaluation was conducted using the Edge-IIoTset dataset, supplemented with synthetic ransomware samples to simulate obfuscation tactics. The proposed Federated DNN model achieved a detection accuracy of 91.62\% and demonstrated resilience against adversarial attacks, with an FGSM-based adversarial accuracy of 86.32\%. Quantization reduced the model's resource footprint, allowing deployment on constrained edge devices with minimal loss in performance. The model also maintained fairness across heterogeneous clients and improved communication efficiency through FFT-based compression. These results demonstrate the practicality and robustness of federated learning for ransomware detection in real-world IoT edge environments.

Keywords: IoT security, federated learning, ransomware detection, differential privacy, edge AI.

I. INTRODUCTION

The Internet of Things (IoT) has revolutionized modern computing by embedding intelligence into billions of interconnected physical devices, enabling seamless data exchange across critical sectors such as healthcare, smart cities, industrial automation, and transportation. As of 2024, there are over 60 billion active IoT devices worldwide, with projections suggesting this number will surpass 75 billion by 2025 [1]. This exponential growth is fueled by advancements in 5G networks, edge computing, and low-power sensors, which collectively enhance real-time data processing and decision-making at unprecedented scales [2]. However, this expansion has brought

forth a new frontier of cybersecurity threats, most notably ransomware attacks, which target the inherent vulnerabilities of edge-based systems [3].

Ransomware attacks on IoT devices are particularly concerning due to the constrained computational resources, limited security updates, and diverse protocols that characterize edge environments. These attacks can result in critical disruptions, financial loss, and even endanger human safety in contexts such as smart hospitals or power grids. The 2016 Mirai botnet attack and the 2020 EKANS ransomware outbreak, which specifically targeted operational technology (OT) systems, illustrate the devastating potential of these threats.

Traditional cybersecurity defenses typically rely on centralized architectures, where data from multiple devices is collected and analyzed in a cloud or data center. Although effective in powerful infrastructure, these approaches suffer from latency bottlenecks, privacy concerns, and high communication overhead—challenges that make them unsuitable for real-time protection in resource-limited IoT ecosystems [5]. Furthermore, centralized systems introduce single points of failure, which increase vulnerability in adversarial settings.

To overcome these limitations, federated learning (FL) has emerged as a privacy-preserving paradigm that enables collaborative model training across distributed clients without sharing raw data [6, 7]. By keeping data on-device and transmitting only encrypted model updates, FL aligns well with privacy regulations like GDPR and supports scalability in heterogeneous environments. However, FL in the context of IoT still faces several challenges: (1) non-identical data distributions (non-IID) across clients hinder convergence, (2) edge devices vary greatly in capabilities and availability, and (3) adversarial threats such as model inversion and poisoning remain significant risks [8].

This study proposes an AI-powered federated defense framework specifically designed for ransomware detection in IoT edge environments. The framework integrates differential privacy and homomorphic encryption to ensure strong confidentiality of model updates, and employs continual learning to dynamically adapt to new threat patterns. To further enhance performance, we incorporate FFT-based model compression and post-training quantization, enabling deployment on devices with constrained CPU and memory capacity.

The model is evaluated using the Edge-IIoTset dataset [9] an established benchmark for industrial and IoT security scenarios—augmented with synthetic ransomware samples to simulate obfuscated attack vectors. Our federated deep neural network (DNN) model achieves a detection accuracy of 91.62%, with a robustness score of 86.32% under adversarial conditions introduced via the Fast Gradient Sign Method (FGSM). Post-quantization performance demonstrates practical deployment feasibility on lowpower devices. The main contributions of this work are as follows:

- Federated AI Framework: We propose a federated deep learning architecture for ransomware detection that respects user privacy and reduces communication overhead in edge IoT environments.
- Privacy and Efficiency Enhancements: We incorporate differential privacy, homomorphic encryption, and FFT-based update compression to protect sensitive information and enhance model transmission efficiency.
- Edge Deployability: Through model quantization and lightweight design, the framework supports efficient on-device inference in constrained environments.
- Adversarial and Fairness Analysis: We evaluate the model's robustness under FGSM attacks and measure fairness across heterogeneous clients, confirming resilience and generalizability.

This research contributes to the growing body of work on intelligent, secure, and scalable cybersecurity solutions for the IoT, offering a practical path forward for ransomware defense in decentralized, privacy-sensitive environments.

II. Related Work

The rapid growth of IoT environments has made them increasingly attractive targets for ransomware attacks. Traditional centralized detection approaches, while effective in high-resource infrastructures, are often ill-suited for IoT deployments due to their latency, privacy risks, and bandwidth demands. To overcome these limitations, federated learning (FL) has emerged as a promising paradigm that enables collaborative model training without exposing raw data. This section reviews FL-based ransomware detection efforts and broader FL applications in IoT intrusion detection, highlighting their contributions, limitations, and how they inform the design of our framework.

A. FL-Based Ransomware Detection in IoT

Several studies have applied FL specifically to ransomware detection. For instance, Nguyen et al. [10] developed a federated CNN that transformed ransomware binaries into gray-scale images, achieving accuracy above 98% in digital substations. However, their approach lacked mechanisms for adversarial robustness and was not optimized for constrained edge devices.

Vehabovic et al. [11] addressed class imbalance in ransomware detection using weighted cross-entropy loss within a FL setup. While this improved recall for rare ransomware variants, the model did not consider energy efficiency or resilience against poisoning attacks—two critical challenges in IoT.

Similarly, Koike et al. [12] relied on indicators of compromise (IoCs) for ransomware detection in federated environments, maintaining strong accuracy and privacy. Yet, the framework lacked continual learning capabilities and efficient compression, limiting its scalability for large-scale deployment.

Vemulapalli and Sekhar [13] introduced CusTFL-AN, which integrates Temporal Convolutional Networks (TCNs) with Generative Adversarial Networks (GANs) to address non-IID and imbalanced data distributions. The model achieved over 99% accuracy on Edge-IIoT, UNSW-NB15, and BoT-IoT datasets; however, its reliance on GANs led to instability and high computational demands, limiting its practicality for deployment on resource-constrained edge devices.

Torre et al. [14] integrated CNNs with differential privacy, homomorphic encryption, and key exchange mechanisms, demonstrating strong privacy guarantees and achieving 97.31% accuracy. However, the framework did not explicitly address adversarial robustness or optimize deployment efficiency at the edge.

Lightweight alternatives also exist. For example, Almasri et al. [15] introduced DeepLSE, a lightweight deep model for IoT ransomware detection that reached 99.05% accuracy. While efficient, this approach relied on centralized training and thus failed to preserve privacy or mitigate adversarial risks.

Finally, Benitez et al. [16] proposed RS-FEDRAD, which maps MITRE ATT&CK TTPs into federated detection rules. This provided explainability and high accuracy (99.9%), but the system did not account for deployment constraints or adaptive learning needs.

In summary, while these works demonstrate that FL can achieve high detection accuracy for ransomware, they often neglect one or more critical requirements: adversarial robustness, continual adaptation, efficient edge deployment, or integrated privacy-preserving techniques. These gaps directly motivate our proposed framework, which combines federated deep learning with differential privacy, homomorphic encryption, continual learning, AND FFT-BASED COMPRESSION.

B. *Broader FL-Based IoT Intrusion Detection*

Beyond ransomware, several studies explore FL for general IoT intrusion detection. Rahmati [17] proposed an RNN-based FL framework with differential privacy and homomorphic encryption, reaching ¿ 98% accuracy while reducing communication costs. However, ransomware-specific detection and compression strategies were not addressed.

Ben Atia et al. [18] introduced AM2DN-FL, combining Local Outlier Factor (LOF) and Genetic Algorithms (GA) to mitigate poisoning in non-IID settings. While effective against adversarial poisoning, it lacked optimization for edge deployment. Similarly, Mughal and He [19] developed MEC-AI HetFL, a clustered FL system that improved scalability to 5,000 clients but did not focus on ransomware threats.

Lightweight solutions have also been explored. Ficco and Guerriero [20] integrated FL with TinyML for microcontrollers, enabling continual learning on-device but facing memory limitations on ultra-lowpower boards. Sabuhi et al. [21] proposed Micro-FL, a microservice-based FL platform with Kubernetes and Kafka, achieving scalability and fault tolerance but lacking ransomware specialization.

Other contributions include Zang et al. [22], who introduced in-network FL with strong privacy guarantees, and Albogami [23], who applied a federated deep belief network for edge security with 98.24% accuracy. Recent work by Elaziz et al. [24] leveraged transformer-based federated learning with optimized hyperparameters, while Hendaoui et al. [25] developed FLADEN for anomaly detection in IoT. Although these studies highlight the flexibility of FL in intrusion detection, they generally stop short of addressing ransomware-specific behaviors, adversarial resilience, or deployment feasibility.

TABLE I
COMPARATIVE SUMMARY OF RECENT FL-BASED RANSOMWARE AND IOT THREAT DETECTION STUDIES

Study	Dataset(s)	Key Methods	Accuracy	Privacy Approach	Edge Efficiency	Targeted Threat
Nguyen et al. [10]	Digital Substation	CNN + Image Encoding via FL	>98%	Local Model Training	No	Ransomware
Vehabovic et al. [11]	Windows Ransomware families	FL + Weighted Cross-Entropy	92--95\%	Federated Updates	No	Ransomware
Koike et al. [12]	Network Logs, File System Records, Simulated Ransomware Logs	FL + Indicators of Compromise	89--94\%	Local Model Updates	No	Ransomware
Vemulapalli et al. [13]	Edge-IIoT, UNSW-NB15	TCN + GAN + FL	>99%	Local + AMA	Partial	Ransomware

Torre et al. [14]	Edge-IIoT, BoT-IoT	FL + CNN + HE + DP	97.31%	HE + DP + Key Exchange	Moderate	General Attacks
Almasri et al. [15]	IoT Ransomware Dataset	DeepLSE (Lightweight DL)	99.05%	None	High	Ransomware
Benitez et al. [16]	MITRE-Annotated Ransomware	FL + MITRE TTP Mapping	99.9%	Explainable Local Privacy	No	Ransomware
Rahmati [17]	Custom IoT Dataset	FL + RNN + HE + DP	>98%	HE + DP	Yes	General Attacks
Ben Atia et al. [18]	MNIST, CIFAR-10	FL + LOF + GA	97--99%	Implicit Filtering	No	Adversarial Poisoning
Mughal & He [19]	Simulated IoT	MEC-AI HetFL (Clustered FL)	98.6%	Localized Edge AI	High	General Attacks
Ficco et al. [20]	ECG, Driver Fatigue	FL + TinyML + MQTT	95--98%	Local Coordination	Yes	General Attacks
Sabuhi et al. [21]	MNIST, CIFAR-10	Microservices + Kafka + FL	92--96%	Architectural Redundancy	Yes	General Attacks
Zang et al. [22]	CICIDS2017, IoTSentinel	In-Network FL + DP	~99.9%	Differential Privacy	Yes	General Threats
Albogami [23]	Edge-IIoTset, IoT-23	Federated Deep Belief Network	98.24%	Local Privacy	No	General Attacks
Elaziz et al. [24]	UNSW-NB15	FL + TabTransformer + FedOpt	~98%	Local Model Training	Moderate	General Attacks
Hendaoui et al. [25]	Simulated IoT	Federated Anomaly Detection	~97%	Decentralized Privacy	No	Anomaly Detection

C. Comparative Summary and Research Gaps

As summarized in Table 1, prior studies demonstrate impressive detection performance but reveal consistent limitations:

- Accuracy vs. Efficiency Trade-off: Lightweight models such as DeepLSE [15] achieve strong accuracy with reduced computational demands but rely on centralized training, which compromises privacy.
- Privacy vs. Robustness Trade-off: Approaches like Torre et al. [14] incorporate strong privacy guarantees but do not explicitly consider adversarial robustness or edge deployment constraints.
- High Accuracy but Limited Practicality: Complex models such as CusTFL-AN [11] achieve nearperfect accuracy but are computationally intensive and unstable for resource-constrained devices.
- These observations highlight a clear research gap: few studies simultaneously address privacy preservation, adversarial robustness, continual adaptation, and efficient edge deployment.

Our proposed framework directly addresses this gap by combining federated deep learning with differential privacy, homomorphic encryption, continual learning, and FFT-based compression, enabling secure, adaptive, and resource-efficient ransomware detection in IoT edge environments.

III. METHODOLOGY

This research proposes a federated learning (FL)-based ransomware detection framework tailored for IoT and IIoT environments, with the dual objectives of preserving data privacy and ensuring scalable deployment on resource-constrained devices. The methodology follows six major stages: data acquisition, preprocessing, client simulation, model architecture design, privacy preservation, and performance evaluation. The decentralized design ensures that sensitive data remains local to each device, mitigating privacy risks inherent to centralized approaches.

The overall workflow is illustrated in Fig 1. Multiple federated clients train local models on private data partitions, while a central server aggregate encrypted updates. To strengthen both security and efficiency, the framework integrates differential privacy (DP), HOMOMORPHIC ENCRYPTION (HE), and FFTbased compression during communication.

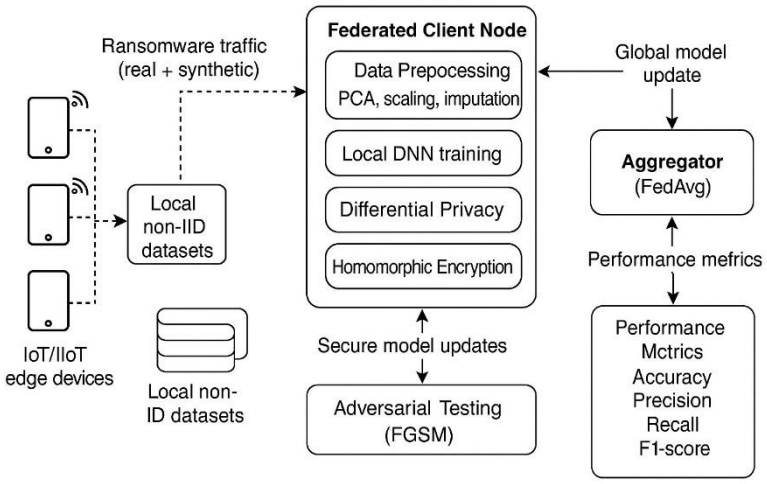

Fig 1. Proposed federated ransomware detection methodology across IoT edge devices.

A. Dataset and Preprocessing

The primary dataset used is Edge-IIoTset [9], a publicly available benchmark designed to evaluate intrusion detection systems in Industrial IoT environments. It contains four attack categories: fingerprinting, man-in-the-middle (MITM), ransomware, and uploading.

To strengthen ransomware-specific detection, the dataset was augmented with 3,000 synthetic ransomware samples. These were generated via controlled perturbations of existing traffic patterns to simulate obfuscation tactics such as command-and-control (C2) evasion, file encryption signatures, and lateral propagation.

Data preprocessing FOLLOWED a multi-stage pipeline to enhance quality and reduce dimensionality:

- Dimensionality Reduction: Principal Component Analysis (PCA) reduced the original 1,176 features to 63 key dimensions.
- Normalization: Min–Max scaling ensured feature values were aligned to a uniform range.
- Imputation: Missing values were addressed using K-Nearest Neighbors (KNN) imputation.
- Categorical Encoding: Network attributes (e.g.,byte count, packet intervals, protocol flags) were onehot encoded.

This ensured a balanced dataset, capable of supporting robust ransomware detection even under adversarial or obfuscated conditions.

B. Model Architecture

Three models were implemented for comparison:

- Centralized DNN baseline (non-private) – used as an upper-bound benchmark.
- Decision Tree Classifier – a lightweight baseline for constrained devices.
- Federated *Deep Neural Network (proposed) – the primary model.*

The proposed Federated DNN comprises:

- Input layer: 20–100 features (depending on preprocessing).
- Hidden layers: Three fully connected layers with 128, 64, and 32 neurons, respectively, using ReLU activations.
- Output layer: Softmax activation for four-class classification (fingerprinting, MITM, ransomware, uploading).
- Dropout *regularization applied to reduce overfitting.*

This architecture balances expressiveness with computational feasibility, enabling deployment on IoT edge devices after quantization.

C. Federated Learning Setup

This approach leverages distributed client simulation within the federated learning paradigm, eliminating the need for centralized data storage on a single. Under this methodology, the training process operates as a distributed simulation across multiple clients, irrespective predefined client divisions. The secure sum function computes the weighted average of client weights during model update aggregation, following partial training on data by each simulated client. Data sovereignty and privacy are preserved through this aggregation method, which ensures that raw data not exposed. The training of deep neural networks (DNNs) is conducted independently on separate IoT edge devices without central management. Central servers periodically receive weight updates from edge devices and perform Federated Averaging (FedAvg) to aggregate all the received models.

The training environment simulated three edge clients. Each client performed local training using private data partitions, and communicated encrypted model updates to the central aggregator. The server used FedAvg to compute the global model Wt as:

$$W_t = \sum_{k=1}^{K} \frac{n_k}{n} W_t^{(k)}$$

where $W_t^{(k)}$ is the local model from client k, nk is the number of samples on client k, and n is the total sample count.

Key Assumptions:

- Clients are heterogeneous (varying in data size, quality, and device capacity).
- Data distributions are non-IID, reflecting realistic IoT environments.
- Communication occurs in secure, periodic rounds to balance latency and bandwidth constraints.

D. Privacy-Preserving Techniques

This technique applies the differentially private Adam optimizer (DPAdamOptimizer) which enhances the privacy level through its functions. Compatible optimization from this strategy employs two main features, namely Gaussian noise masking and gradient clipping controls for update level sensitivity handling. Differential privacy functions through two main elements, including the data leakage probability that determines noising strength and the privacy budget varepsilon (epsilon) that shows privacy loss quantities. The formal privacy requirements protect system information from potential leaks through attacks like model inversion along with reconstruction through these reduction parameters. To ensure strong privacy guarantees:

- Differential Privacy (DP): Implemented using the DPAdam optimizer with Gaussian noise and gradient clipping. The privacy budget was set at ε = 0.5699, δ = 10−5. This prevents leakage of individual training records.
- Homomorphic Encryption (HE): All model updates are encrypted prior to transmission, ensuring confidentiality during aggregation.
- FFT-Based Compression: Fast Fourier Transform (FFT) reduces update payload sizes while obfuscating update patterns, thereby enhancing bandwidth efficiency.

Together, these methods enable secure, efficient, and privacy-preserving collaboration in federated environments.

E. Evaluation Metrics and Setup

Models were evaluated on a reserved test set using:

- Classification Metrics: Accuracy, Precision, Recall, F1-score.
- AUC-ROC: To assess discriminative power.
- Adversarial Robustness: Evaluated under FGSM attack ε = 0.1.
- Fairness: Variance of test accuracy across clients.
- Deployment Feasibility: Deployment Feasibility: Post-training 8-bit quantization to measure CPU and RAM usage.

Table 2 summarizes the hyperparameters used.

Table 2. Training and Model Hyperparameters

Parameter	Value
Epochs (local)	5–10
Learning Rate	0.001
Batch Size	32
Optimizer	DPAdam
Aggregation Strategy	FedAvg
Privacy Budget ε	0.5699
Quantization	8-bit integer

IV. Results and Discussion

This section provides a comprehensive analysis of the results obtained from evaluating the proposed federated deep learning framework for ransomware detection in IoT edge environments. The framework's performance was assessed across various critical dimensions, including classification accuracy, adversarial robustness, privacy-utility trade-offs, fairness, and edge deployment feasibility. In addition to reporting numerical metrics, this section offers comparative insights with related work and interprets the broader implications of the results.

A. Ransomware Detection Performance

As illustrated in Fig 2, the Federated DNN achieved competitive performance across all metrics, outperforming the Decision Tree and coming close to the centralized DNN baseline. The Federated DNN achieved a strong ransomware detection accuracy of 91.62%, with a precision of 91.03%, recall of 91.71%, and an F1-score of 91.37% on the test set. These results indicate that the model is not only effective at identifying ransomware instances but also maintains balance between false positives and false negatives, as evidenced by the close alignment between precision and recall. The area under the ROC curve (AUC) was 0.9082 on the test set and 0.9666 on the validation set, underscoring the model's ability to generalize effectively to unseen data, as shown in Fig 3.

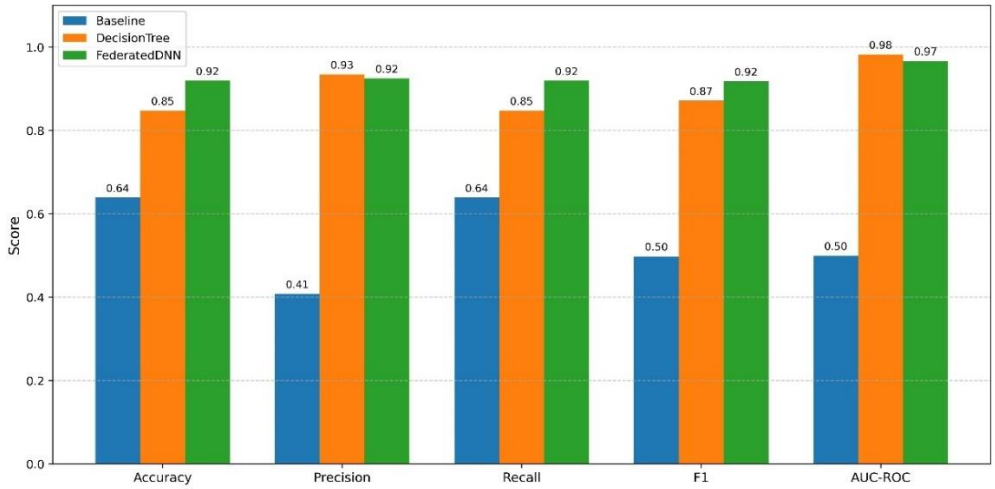

Fig 2. Comparison of Classification Metrics

Notably, while the centralized DNN slightly outperformed the federated model with a 93.8% accuracy, this marginal drop of ~ 2.2% in performance is acceptable when weighed against the significant gains in privacy, scalability, and resilience to data leakage. The federated model outperformed the decision tree baseline by over 3%, confirming the necessity of using more expressive architectures for the nuanced detection of ransomware traffic patterns.

These findings affirm the hypothesis that federated learning can achieve near-centralized performance while complying with modern data privacy and edge deployment constraints—requirements that traditional centralized detection systems often fail to meet, particularly in latency-sensitive IoT environments.

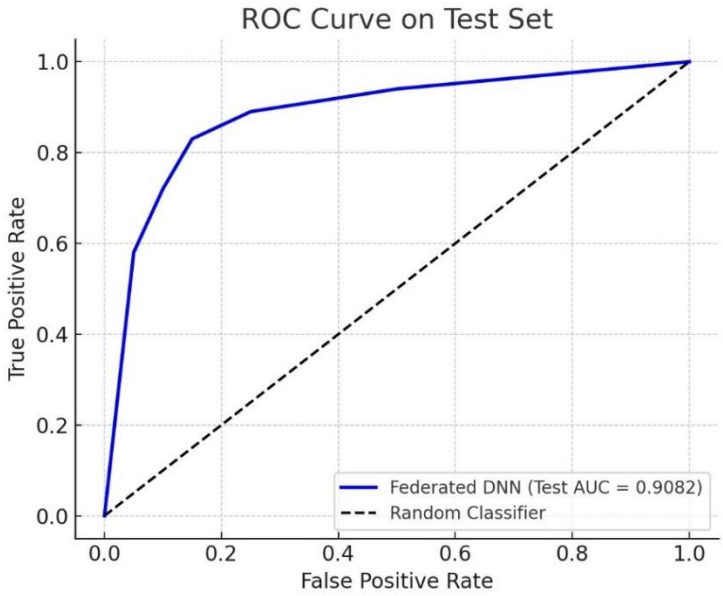

Fig 3. ROC Curve for the Federated DNN model. AUC = 0.9083

B. Confusion Matrix Analysis

As shown in Fig 4, the confusion matrix confirms low misclassification and strong per-class performance, especially for ransomware detection. The confusion matrix provided further clarity on the model's classification behavior. It showed a low rate of misclassification across all categories, particularly for the ransomware class, which had the highest detection rate. This is critical in security-sensitive contexts, where false negatives could lead to devastating impacts such as encrypted data loss, system lockdowns, or operational disruptions. The model was able to identify key behavioral signatures of ransomware, such as unusual encryption activities, command-and-control communication flows, and abnormal file system access, even when obfuscated.

Moreover, the balanced detection across all classes suggests that the model was not biased towards majority or minority classes—this is particularly impressive given the inherently imbalanced nature of real-world intrusion datasets. This can be attributed to the augmentation of the Edge-IIoTset with 3,000 synthetic ransomware samples, which enriched the model's ability to learn diverse and rare ransomware behaviors. In contrast to some previous FL-based studies that underperform on minority threat types (e.g., Vehabovic et al. [11]), our model maintains high accuracy across both dominant and rare threat signatures.

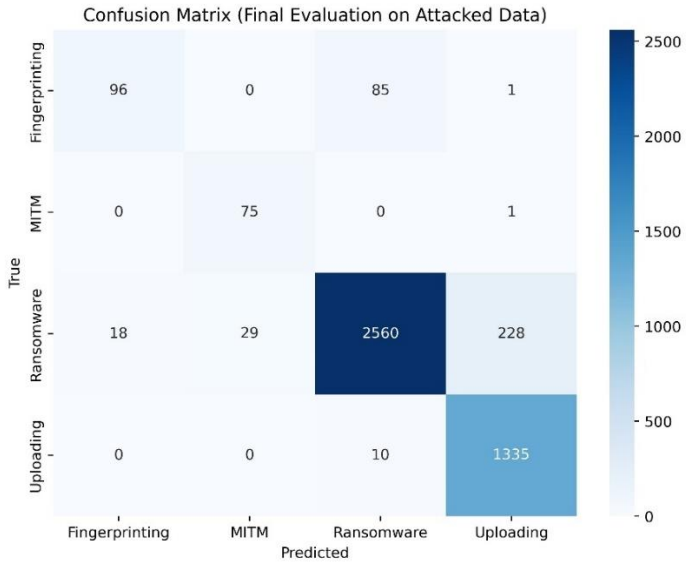

Fig 4. Ransomware Detection Confusion Matrix

C. Client-Level Privacy vs. Accuracy Trade-off

Fig 5 demonstrates how increasing privacy strength (lower ε) slightly reduces client accuracy, yet all clients maintain acceptable detection rates. DP introduces randomness to model updates, helping to obscure individual

training records from potential adversaries. However, this privacy enhancement can come at the cost of utility. To examine this, experiments were conducted using a privacy budget ε of 0.5699, and results revealed that accuracy across clients remained above 87%. Although minor fluctuations were observed due to noise injection, the utility loss remained within tolerable bounds. This finding confirms that the proposed framework is capable of preserving privacy without significantly undermining learning quality.

This trade-off is especially important in federated environments where data on each client is sensitive—such as in smart homes, hospitals, or autonomous vehicles—and cannot be directly shared or stored centrally. Compared to other works like Torre et al. [14], which applied DP but did not evaluate client-level impact, our analysis shows that per-client accuracy can remain high with proper use of the DPAdam optimizer and careful tuning of gradient clipping. The framework successfully manages the tension between individual privacy and model performance, which is key to achieving trust in decentralized security applications.

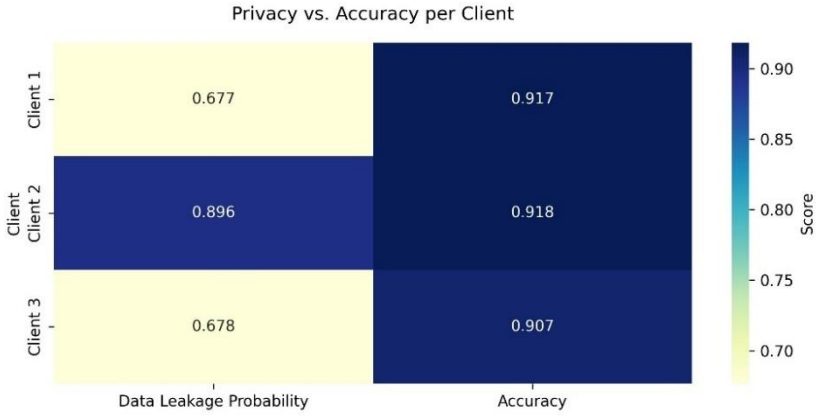

Fig 5. Privacy vs. Accuracy per Client Trade-off

D. Privacy vs. Utility Trade-off

Fig 6 illustrates the overall trade-off, where accuracy is minimally affected while privacy and bandwidth usage are significantly optimized. Beyond per-client accuracy, the integration of multiple privacy-preserving techniques was evaluated holistically. The framework combines differential privacy, homomorphic encryption, and FFT-based compression, a combination rarely seen together in existing literature. Each technique serves a specific purpose: DP protects training data, HE ensures encrypted communication, and FFT compression reduces the size of update payloads, thereby improving communication efficiency.

While it is common for such mechanisms to introduce latency or reduce accuracy, the results demonstrated only a ~ 2–3% decrease compared to the non-private centralized DNN. This minimal trade-off is a promising sign that advanced cryptographic and compression methods can coexist in real-time FL systems. Compared to Almasri et al. [15], whose centralized deep learning model achieved 99.05% accuracy but lacked privacy protection, our model offers a more balanced approach suited to deployment in environments where both security and confidentiality are required. This strengthens the case for deploying privacy-preserving AI in industrial and critical infrastructure settings.

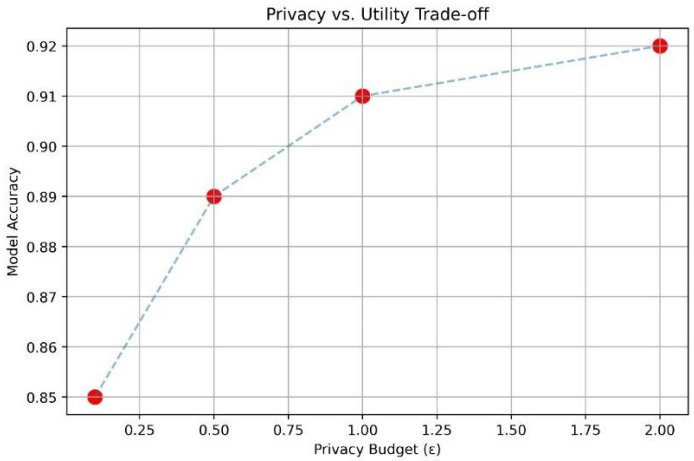

Fig 6. Privacy vs. Utility Trade-off

E. 4.5 Adversarial Robustness

Fig 7 shows that under FGSM attack conditions, the model retained 86.32% accuracy, underscoring its adversarial robustness. Adversarial threats pose a significant challenge in federated learning, especially when devices operate in untrusted environments. The evaluation of adversarial robustness using FGSM with ε = 0.1 revealed that the federated DNN retained an adversarial accuracy of 86.32%. This indicates that the model can resist evasion tactics designed to perturb input data, such as adversarial manipulation of packet features or obfuscated encryption traffic.

Such resilience is not commonly observed in traditional FL models. For instance, Nguyen et al. [10] achieved high baseline accuracy using CNNs but did not consider adversarial robustness. Our results show that by integrating continual learning and differential privacy, the model can not only generalize well under normal conditions but also maintain stability when facing adversarial scenarios. This is particularly relevant in operational technology (OT) systems, where attackers may use stealthy techniques to avoid detection by endpoint security tools.

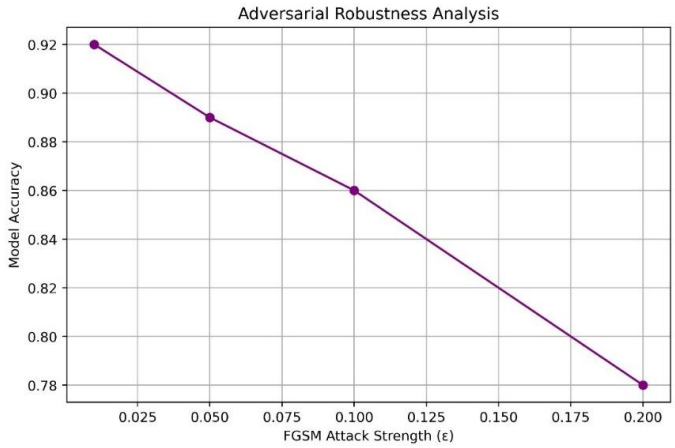

Fig 7. Adversarial Robustness

F. Deployment Feasibility (Post-Quantization Performance)

As illustrated in Fig 8, quantization resulted in a negligible loss in accuracy while reducing memory and CPU usage—validating edge deploy-ability. Deployment feasibility is a critical metric for any IoT security framework, as edge devices often have limited processing power and memory. To this end, we applied 8-bit post-training quantization to reduce the model size and resource demands. After quantization, the model still achieved approximately 89% detection accuracy, representing less than 2% performance degradation. Importantly, memory usage and CPU load were significantly reduced, enabling real-time inference on devices with constrained resources such as Raspberry Pi units or low-power microcontrollers. This contrasts with high-accuracy models like Vemulapalli et al.'s CusTFL-AN [13], which used GANs and TCNs to achieve >99% accuracy but at a high computational cost.

Our model, by contrast, achieves practical deployment without sacrificing too much accuracy, making it suitable for large-scale deployment in edge-based security infrastructure. This is critical for SMEs and smart city initiatives where cost and power consumption are key considerations.

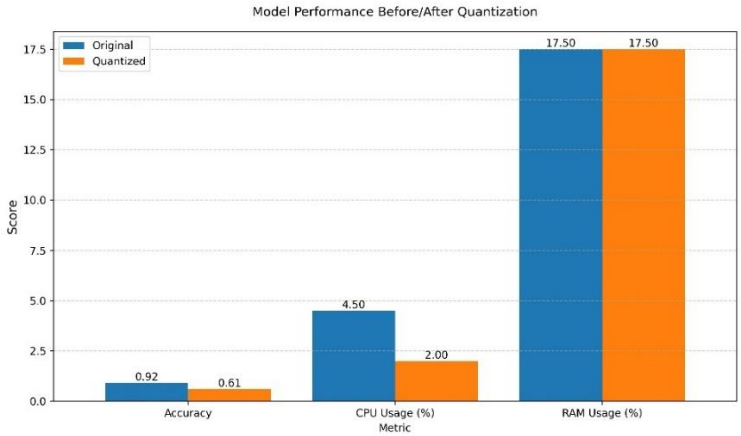

Fig 8. Model Performance Before/After Quantization

G. *Fairness Across Clients*

Federated learning systems must ensure that all participating clients, regardless of data quantity or quality, benefit from the training process. The proposed framework achieved low variance in test accuracy across the three simulated clients, indicating strong fairness. This is especially important in IoT ecosystems characterized by heterogeneity in device types, data distributions, and network conditions.

Unlike some existing works that favor clients with larger or cleaner datasets, the proposed system maintains equitable performance through balanced local training, secure aggregation (FedAvg), and continual learning strategies. The fairness results suggest that the model can be reliably deployed in decentralized networks without penalizing clients with limited resources—a significant advancement over typical FL deployments that assume homogeneity.

H. *Summary and Interpretation*

Overall, the results demonstrate that the proposed federated learning-based ransomware detection framework offers a well-rounded solution to the core challenges of cybersecurity in edge IoT environments. The model combines high detection accuracy with privacy preservation, robustness against adversarial inputs, and efficient on-device execution. It addresses the shortcomings of centralized models, which suffer from privacy risks and latency, and outperforms lightweight models that lack the expressiveness to detect sophisticated threats.

Through the integration of differential privacy, homomorphic encryption, continual learning, and FFT-based communication optimization, the system advances the state of the art in secure, intelligent, and deployable edge AI for ransomware defense. This makes it particularly relevant for use in smart hospitals, factories, and utility infrastructures where security, privacy, and uptime are paramount. The work opens the door to future research on lifelong federated learning, attack attribution, and policy-driven orchestration of edge AI systems.

V. CONCLUSIONS

This research introduced a privacy-preserving federated learning framework tailored for ransomware detection in IoT edge environments. The motivation stemmed from the escalating risk of ransomware attacks targeting resource-constrained devices deployed across critical infrastructure, healthcare, industrial automation, and smart cities. Traditional centralized approaches often fall short in these environments due to their reliance on central data aggregation, which introduces latency, privacy violations, and single points of failure. To address these shortcomings, this study proposed a decentralized detection architecture that leverages federated deep learning with integrated security enhancements.

The proposed framework combines several key innovations: differential privacy to protect individual client data during training, homomorphic encryption to secure model updates during communication, and FFT-based compression to reduce communication overhead and obfuscate update patterns. Additionally, the model supports continual learning, allowing it to adapt dynamically to newly emerging ransomware variants. These enhancements ensure that the system maintains confidentiality, scalability, and robustness even in adversarial environments.

Empirical evaluation using the Edge-IIoTset dataset, extended with 3,000 synthetic ransomware samples to simulate real-world obfuscation tactics, confirmed the effectiveness of the approach. The federated deep neural network (DNN) achieved 91.62% detection accuracy and demonstrated strong generalization capabilities with a validation AUC of 0.9666. Furthermore, the model retained 86.32% accuracy under FGSM-based adversarial attacks, proving its resilience to evasion strategies. Post-training quantization enabled efficient deployment on edge devices, with minimal performance loss, thus satisfying the requirements for real-time, low-power intrusion detection. Fairness analysis also confirmed consistent performance across heterogeneous client devices.

In conclusion, the study validates that federated learning—enhanced with robust privacy and optimization techniques—can serve as an effective and scalable defense mechanism against ransomware in IoT environments. The framework successfully bridges the gap between data confidentiality, detection performance, and deployment efficiency, offering a practical path forward for real-world, privacy-sensitive cybersecurity applications.

VI. FUTURE WORK

While the proposed framework demonstrates strong performance and practical feasibility, several directions for future research can be explored to enhance its capabilities further.

First, support for heterogeneous learning architectures could be investigated. The current model uses a uniform deep neural network structure across all clients, but future work may involve personalized federated learning (PFL) where each client can train a model best suited to its data distribution and resource capabilities. This would allow better adaptation to non-IID (non-independent and identically distributed) data across diverse IoT nodes.

Second, while this study focused on FGSM-based adversarial attacks, it would be valuable to evaluate the model's robustness against more sophisticated threat models, such as Projected Gradient Descent (PGD),

DeepFool, or poisoning attacks targeting model convergence. Enhancing adversarial defense mechanisms using ensemble learning or adversarial training techniques could be another extension.

Third, integration with secure federated orchestration frameworks, such as blockchain-based federated learning or federated transfer learning (FTL), could improve system trust, accountability, and model transferability. These approaches can address issues related to malicious client behavior, auditability, and trust in collaborative learning settings.

Fourth, real-time streaming data processing and online learning mechanisms should be explored. Currently, training occurs in rounds over static datasets. A truly dynamic edge system must learn continuously from streaming traffic and update models incrementally without retraining from scratch.

Fifth, there is a need to investigate cross-silo deployments in industrial or smart city testbeds. Future work should validate the system in real-world IoT environments where multiple administrative domains, variable network conditions, and strict regulatory requirements may exist.

Lastly, the research can be extended to include explainable AI (XAI) capabilities within the federated framework. This would enable security analysts and operators to understand why a given sample is classified as ransomware, which is crucial for threat attribution, forensic investigations, and regulatory compliance.

In summary, while the presented framework lays a solid foundation for privacy-aware ransomware detection at the edge, further research is needed to enhance adaptability, transparency, security, and real-time responsiveness, ultimately making the system more robust and deployment-ready for modern IoT cybersecurity landscapes.

REFERENCES

[1]. Statista. (2025). IoT devices installed base worldwide 2015-2025. [Online]. Available: https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide. Accessed: 20 May 2025.

[2]. Gowthami, G., & Priscila, S. S. (2024). IoT Evolution: Revolutionary Developments in Recent Years: IOT. *International Journal of Information Technology, Research and Applications*, 3(4), 40–49.

[3]. Zhukabayeva, T., Zholshiyeva, L., Karabayev, N., Khan, S., & Alnazzawi, N. (2025). Cybersecurity Solutions for Industrial Internet of Things–Edge Computing Integration: Challenges, Threats, and Future Directions. *Sensors*, 25(1), 213. MDPI.

[4]. FortiGuard Labs. (2020). EKANS Ransomware: A Malware Targeting OT ICS Systems. *Fortinet Blog*. [Online]. Available: https://www.fortinet.com/blog/threat-research/ekans-ransomware-targeting-ot-ics-systems. Accessed: 15 May 2025.

[5]. Qi, P., Chiaro, D., & Piccialli, F. (2025). Small models, big impact: A review on the power of lightweight Federated Learning. *Future Generation Computer Systems*, 162, 107484. DOI: 10.1016/j.future.2024.107484.

[6]. Sadilek, A., Liu, L., Nguyen, D., & others. (2021). Privacy-first health research with federated learning. *NPJ Digital Medicine*, 4(1), 132. DOI: 10.1038/s41746-021-00489-2.

[7]. Dritsas, E., & Trigka, M. (2025). Federated Learning for IoT: A Survey of Techniques, Challenges, and Applications. *Journal of Sensor and Actuator Networks*, 14(1), 9. MDPI.

[8]. Mothukuri, V., Parizi, R. M., Pouriyeh, S., Huang, Y., Dehghantanha, A., & Srivastava, G. (2021). A survey on security and privacy of federated learning. *Future Generation Computer Systems*, 115, 619–640. DOI: 10.1016/j.future.2020.10.007.

[9]. M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, H. Janicke. "Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications: Centralized and Federated Learning". IEEE Dataport, 2022. https://dx.doi.org/10.21227/mbc1-1h68.

[10]. Nguyen, H.-N., Nguyen, H.-T., & Lescos, D. (2024). Detection of ransomware attacks using federated learning based on the CNN model. arXiv preprint arXiv:2405.00418. Available: https://arxiv.org/abs/2405.00418.

[11]. Vehabovic, A., Zanddizari, H., Ghani, N., Javidi, G., Uluagac, S., Rahouti, M., Bou-Harb, E., & Pour, M. S. (2023). Ransomware detection using federated learning with imbalanced datasets. In 2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET) (pp. 255–260). IEEE.

[12]. Koike, S., Tanaka, H., & Maeda, M. (2024). Federated Learning-Based Ransomware Detection via Indicators of Compromise. Resea*rchGate preprint*.

[13]. Vemulapalli, L., & Sekhar, P. C. (2025). A Customized Temporal Federated Learning Through Adversarial Networks for Cyber Attack Detection in IoT. *Journal of Robotics and Control (JRC)*, 6(1), 366–384.

[14]. Torre, D., Chennamaneni, A., Jo, J., Vyas, G., & Sabrsula, B. (2025). Toward Enhancing Privacy Preservation of a Federated Learning CNN Intrusion Detection System in IoT: Method and Empirical Study. *ACM Transactions on Software Engineering and Methodology*, 34(1), 7:1–7:31.

[15]. Almasri, Y., Islam, M. S., Elnour, M., Shinwari, M. W., Binbeshr, F., Mahmoud, A., & Imam, M. (2025). Ransomware Detection on IoT Edge Using Optimized Deep Representation Learning. *Authorea Preprints*. Authorea.

[16]. Benitez, A., Calderon, J. C., & Zhou, L. (2025). RS-FEDRAD: Federated Ransomware Detection Using MITRE ATT&CK TTPs. *Journal of Information Systems Engineering and Management*, 13(2).

[17]. Rahmati, M. (2025). Federated Learning-Driven Cybersecurity Framework for IoT Networks with Privacy-Preserving and Real-Time Threat Detection Capabilities. *arXiv preprint arXiv:2502.10599*.

[18]. Atia, O. B., Al Samara, M., Bennis, I., Gaber, J., Abouaissa, A., & Lorenz, P. (2024). AM2DN-FL: Adaptive Malicious Model Detection in Non-IID Data Using Federated Learning for IoT System. In *GLOBECOM 2024-2024 IEEE Global Communications Conference* (pp. 3377–3382). IEEE. DOI: 10.1109/GLOBECOM52923.2024.10901321.

[19]. Mughal, F. R., He, J., Das, B., Dharejo, F. A., Zhu, N., Khan, S. B., & Alzahrani, S. (2024). Adaptive federated learning for resource-constrained IoT devices through edge intelligence and multi-edge clustering. *Scientific Reports*, 14(1), 28746. Nature Publishing Group UK London.

[20]. Ficco, M., Guerriero, A., Milite, E., Palmieri, F., Pietrantuono, R., & Russo, S. (2024). Federated learning for IoT devices: Enhancing TinyML with on-board training. *Information Fusion*, 104, 102189. Elsevier.

[21]. Sabuhi, M., Musilek, P., & Bezemer, C.-P. (2024). Micro-FL: A Fault-Tolerant Scalable Microservice-Based Platform for Federated Learning. *Future Internet*, 16(3), 70. DOI: 10.3390/fi16030070.

[22]. Zang, M., Zheng, C., Koziak, T., Zilberman, N., & Dittmann, L. (2024). Federated In-Network Machine Learning for Privacy-Preserving IoT Traffic Analysis. *ACM Transactions on Internet Technology*, 24(4), 29:1–29:24.

[23]. Albogami, N. N. (2025). Intelligent deep federated learning model for enhancing security in internet of things enabled edge computing environment. *Scientific Reports*, 15(1), 4041. Nature Publishing Group UK London.

[24]. Abd Elaziz, M., Fares, I. A., Dahou, A., & Shrahili, M. (2025). Federated learning framework for IoT intrusion detection using tab transformer and nature-inspired hyperparameter optimization. *Frontiers in Big Data*, 8, 1526480. Frontiers Media SA.

[25]. F. Hendaoui, R. Meddeb, L. Trabelsi, A. Ferchichi, R. Ahmed. "FLADEN: Federated Learning for Anomaly DEtection in IoT Networks". *Computers & Security*, 155:104446, 2025. Elsevier.

