A Lightweight Edge AI Framework for Secure and Scalable IoT Systems

Friday Oodee Philip-Kpae [1,*], Agbotiname Lucky Imoize [2], Lloyd Endurance Ogbondamati [3], Kelvin Anoh [4], Kennedy Chinedu Okafor [4,5,7,8]
[1, 3] Department of Electrical and Electronic Engineering, Faculty of Engineering, Rivers State University, Port Harcourt, Nigeria;
[2] Department of Electrical and Electronics Engineering, Faculty of Engineering, University of Lagos, Akoka, Lagos 100213, Nigeria;
[4] School of Engineering, University of Chichester, Bognor Regis, PO21 1HR, U.K;
[5] Department of Electrical and Electronic Engineering Science, University of Johannesburg, South Africa;
[6] Imperial College Business School – South Kensington Campus, Imperial College London, Exhibition Road, London SW7 2AZ, UK
[7] Federal University of Allied Health Sciences, Enugu – Dental Avenue, Trans-Ekulu, Enugu State, Nigeria
[8] Department of Engineering (Faculty of Science and Engineering), Manchester Metropolitan University – Dalton Building, Chester Street, Manchester M1 5GD, UK
philip-kpae.fo@ust.edu.ng; aimoize@unilag.edu.ng; lloydhouston27@gmail.com; k.anoh@chi.ac.uk; kennedy.okafor@ieee.org

*Correspondence: Friday Oodee Philip-Kpae (philip-kpae.fo@ust.edu.ng)

***Abstract*— The proliferation of connected objects in the Internet of Things (IoT) ecosystem presents challenges in enabling real-time intelligence while safeguarding data privacy, especially within the computational and energy limitations of edge devices. This study, based on simulation and synthetic data collection methods, addresses these challenges by proposing a lightweight edge AI framework that employs federated learning, enabling model training across distributed Internet of People and Things (IoP) devices without transferring raw data to centralised servers. The framework integrates on-device inference, stochastic local updates, and model compression to ensure low-latency decision-making while adhering to memory and energy constraints. To enhance security, differential privacy mechanisms, encrypted aggregation, and robust outlier detection are utilized to defend against adversarial and Byzantine attacks. The proposed framework offers an effective solution for deploying federated intelligence on resource-constrained IoP devices, enabling responsive, privacy-preserving, and resilient edge AI operations in distributed environments.**

Index Terms—*Edge AI, Distributed Intelligence, Adversarial Defense, Stochastic Updates, Lightweight Computing, IoP Devices, Differential Privacy, Model Compression*.

I. Introduction

In today's increasingly interconnected world, billions of physical devices—ranging from home appliances and industrial sensors to medical wearables and autonomous machines—are becoming part of what is now known as the Internet of People and Things (IoP). As these connected objects grow in number and intelligence, they continuously generate massive volumes of data that must be processed quickly, securely, and with minimal reliance on distant cloud servers [1]. Traditional centralized approaches are no longer sufficient, as they struggle with latency, bandwidth limitations, privacy risks, and scalability challenges. This evolving landscape has created an urgent demand for a new class of intelligent, distributed systems that can learn cooperatively while preserving user trust. Federated intelligence offers a compelling response to this challenge by enabling connected objects to train and refine machine-learning models locally, sharing only insights rather than raw data [2]. When combined with lightweight edge AI, this collaborative learning paradigm enables devices with limited processing power to meaningfully participate in real-time decision-making. Such an approach not only reduces communication overhead but also strengthens privacy by keeping sensitive user data on the device. This is critical as IoP systems begin to influence daily life—from personalized healthcare and intelligent energy management to predictive maintenance and autonomous mobility [3].

A lightweight edge AI framework tailored for IoP environments brings flexibility, adaptability, and resilience to distributed networks. By decentralising intelligence, it ensures that connected objects remain responsive even during network disruptions while maintaining high levels of security through encrypted model exchanges and robust authentication. Moreover, the scalable nature of federated learning enables seamless integration of new devices without compromising system performance [4]. Ultimately, federated intelligence represents a transformative shift toward more secure, human-centered, and scalable IoP systems. It empowers connected objects to work together intelligently, protects user privacy, and lays the foundation for the next generation of autonomous, trustworthy, and efficient digital ecosystems. This paper makes the following key contributions:

1. Introduction of a Lightweight Edge AI Framework (i.e., federated learning-based framework) for real-time decision-making on resource-constrained IoP devices.
2. Incorporation of differential privacy, encrypted aggregation, and outlier detection (i.e., Privacy-Preserving mechanisms) to secure data and protect against attacks.
3. Simulation-Based Evaluation to Validate the framework's performance using synthetic data, demonstrating effective model compression, energy efficiency, and low-latency operation.

The remaining part of the paper is structured as follows: Section II reviews related work on federated learning, edge AI, and privacy-preserving techniques in IoT systems. Section III presents the proposed lightweight edge AI framework, outlining its key components, such as model compression, federated learning, and security mechanisms. It also describes the simulation setup and synthetic data collection methods used to evaluate the framework. Section IV presents and discusses the results, analysing the framework's performance in terms of model size reduction, energy consumption, privacy, and security. Finally, Section V concludes the paper and highlights potential areas for future research.

II. Related Work

Edge AI and federated learning (FL) have emerged as promising paradigms for enabling privacy-preserving, distributed intelligence on resource-constrained IoT devices. According to recent surveys, FL allows edge devices to collaboratively train global models without sharing raw data — addressing privacy concerns and reducing reliance on centralised cloud servers [1], [2], [7]. However, edge and IoT devices pose challenges including limited memory, compute power, energy constraints, and data heterogeneity [3]-[6]. To address resource constraints, some works combine FL with TinyML (lightweight models) and adopt efficient model-compression or pruning techniques so that models can run on microcontrollers or low-power edge devices [3].

Security and privacy remain critical. Researchers integrate differential privacy (DP), secure aggregation, and encryption to safeguard user data against inference, poisoning, or

model-reconstruction attacks [4]-[6], [8]. Additionally, hybrid cryptographic protocols have been developed to reduce communication and computational costs, making secure FL more practical for IoT contexts. Others propose meta-learning–aided federated frameworks to efficiently optimize parameters and reduce communication overhead while retaining classification performance across diverse IoT devices [9], [4]. Despite these advances, studies note persistent gaps: few frameworks jointly address compression, energy efficiency, privacy, and robustness; most focus on idealized or homogeneous devices; and non-IID data distributions across devices remain a challenge for convergence and accuracy [10]. Thus, there is a clear need for a holistic, lightweight edge AI framework that integrates model compression, efficient local computation, privacy-preserving federated learning, and robustness against adversarial behavior — especially tailored for heterogeneous, resource-limited IoP/IoT devices.

III. Methodology

This paper's research design focuses on developing and evaluating a lightweight edge AI framework for IoP devices using simulation and synthetic data. The framework leverages federated learning for collaborative model training while ensuring privacy with differential privacy and encrypted aggregation. Performance is assessed through metrics such as model compression, energy consumption, accuracy, and latency, with security evaluated via Byzantine-robust aggregation to detect malicious updates. The design addresses key IoT challenges, including efficiency, privacy, and security. The resources used in this study include simulation software, synthetic datasets representing IoP device sensor inputs, and computational tools for testing model compression, energy consumption, and security. Federated learning, differential privacy, and Byzantine-robust aggregation algorithms are also utilized to evaluate the framework's performance under controlled conditions.

The experimental and simulation setup involves three edge AI devices (D1, D2, and D3) with varying sensor inputs, such as temperature, vibration, and load data. These devices are simulated to reflect real-world constraints, including memory, energy, and computational limitations. Each device performs local training and inference, with model updates aggregated via federated learning and privacy-preserving techniques such as encrypted aggregation and differential privacy. The setup evaluates key metrics such as model compression, energy consumption, privacy, and security under different conditions across these devices. The data collection method for this study relies on synthetic data generation to simulate real-world IoP device sensor inputs, such as temperature, vibration, and load.

These datasets are designed to mimic the conditions of resource-constrained devices. The data is collected through simulation, in which each device (D1, D2, and D3) performs local inference and updates its model. The data used in the evaluation includes sensor readings, model updates, and performance metrics such as energy consumption, accuracy, and privacy-related parameters. This approach allows controlled experimentation without requiring real-world data, enabling a thorough assessment of the proposed framework. The system model for this study involves several key components that work together to allow federated learning on resource-constrained IoP devices. The mathematical formulations for these components are as follows:

Number of devices and indexing: Let N: represents the total number of devices in the system, where each device i $(for\ i \in \{1, 2, \ldots . N\})$ participates in the collaborative learning process. In the system, $N = 3$, corresponding to devices $D1, D2, and\ D3$.

Local inference (real-time decision): This equation models how a resource-constrained device produces a prediction from local sensor input using its lightweight model. It emphasizes on-device inference for latency-sensitive tasks, keeping raw data local. The formulation is generic, so different model classes (shallow nets, small CNNs, decision trees) can be substituted while ensuring the device can act immediately without cloud roundtrips. Each device i performs local inference on its sensor data. Let x_i denote the input data at device iii (e.g., temperature, vibration, load readings), and θ_i represent the model parameters on the device i. The local inference for the device i is given by: $\hat{y}_i = f(x_i, \theta_i)$ where, x_i : input vector at device i (sensor readings), θ_i: local model parameters for the device i, $f(\cdot)$: inference function applied to the local data using the model, and $\hat{y}_i$: predicted output/decision for device i.

Lightweight on-device update (stochastic gradient step): Represents an inexpensive local training/update step using stochastic gradient escent (or its variant) with a small batch; suitable for limited compute. The update keeps learning incrementally, reducing computation and memory needs by using low iterations and small batches — key to a lightweight edge AI framework that adapts to streaming local data while preserving battery and CPU constraints. The update rule is given by: $\theta_i^{new} = \theta_i - \eta \nabla\, \mathcal{L}_i(\theta_i)$, where, θ_i^{new}: is the updated model parameters for the device i, θ_i: is the current model parameters, η : learning rate (step size), $\nabla\, \mathcal{L}_{\mathrm{i}}(\theta_{\mathrm{i}})$: is the gradient of the local loss function $\mathcal{L}_{\mathrm{i}}$with respect to θ_i. The loss function $\mathcal{L}_{\mathrm{i}}(\theta_{\mathrm{i}})$ is computed based on the local data x_i, and the gradient $\nabla\, \mathcal{L}_{\mathrm{i}}(\theta_{\mathrm{i}})$ is used to adjust the model to reduce the prediction error.

Model compression constraint (size budget): To address the memory constraints of resource-constrained devices, model compression techniques such as pruning and quantization are applied to reduce the model size. Let M_i represent the size of the model before compression and $\widehat{M}_i$ represent the size after compression. The model compression constraint is formulated as: $\left|\widehat{M}_i\right| \leq \left|\widehat{M}_{\mathrm{i}}^{\max}\right|$, where, $\left|\widehat{M}_i\right|$: is the size of the compressed model on the device i, $\left|\widehat{M}_{\mathrm{i}}^{\max}\right|$: is the maximum allowable model size on device i, based on its capacity. This ensures the model remains within each device's memory budget while maintaining acceptable performance.

Local compute energy per round: The energy consumption per local computation round is critical for devices with limited battery life. Let E_i^{comp} denote the energy consumed by the device i during one local training or inference round. The energy consumption is calculated as: $E_i^{comp} = O_i \cdot \epsilon$, where, E_i^{comp}: energy consumed by the computer on the device i(Joules), O_i: number of arithmetic/logic operations performed by the device i during a round of local training, and ϵ: average energy per operation (J/op). This formula ensures that energy consumption is accounted for when evaluating the feasibility of running local computations on resource-constrained devices.

Federated weighted aggregation (privacy-preserving FedAvg): In federated learning, local updates from each device are aggregated to form the global model. Let $\hat{\theta}_i$ represent the updated model parameters from device III, and $\widehat{M}^{global}$ represent the global model. The global model is updated using a weighted average of the local models, as follows $\widehat{M}^{global} = \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}_i$, where, N: participating devices count, $\hat{\theta}_i$: is the updated parameters from the device i, and $\widehat{M}^{global}$: is the updated global model aggregation. The aggregation weights can be adjusted based on the local sample sizes or other factors.

Communication cost per round: Quantifies network load contributed by each device during a federated round: model payload times number of transmissions. Useful to minimize bandwidth (via compression, sparse updates, or fewer rounds). This metric helps evaluate tradeoffs between update frequency, model size, and network constraints in scalability studies. This is presented by (6): $\Omega^{round} = \sum_{i=1}^{N} s_i \cdot p_i$, where: Ω^{round}: total bits transmitted in a round, $\mathrm{s_i}$: size (bits) of device i's transmitted payload (update)$\mathrm{p_i}$: number of transmissions from the device i in the round.

This framework is a Lightweight Edge AI solution for secure, scalable IIoT Systems. The Federated Learning Orchestrator

manages the process. It sends a Global Model to multiple Edge AI Devices (IoT Objects). These devices train the model locally using their Secure Data, ensuring Data Privacy and Anonymization since raw data never leaves the device. The Local Model Training results are aggregated securely by the Orchestrator to update the Global Model, promoting Scalability and Resource Optimization across the system via a Secure Communication Channel.

Differential privacy noise budget per update: To preserve privacy during model updates, differential privacy (DP) is applied by adding noise to updates. Let $\hat{\theta}_i^{noisy}$ represent the privacy-preserving model update from the device i. The noisy update is given by (7): $\hat{\theta}_i^{noisy} = \hat{\theta}_i + \mathcal{N}(0, \sigma_i^2)$,where: $\hat{\theta}_i$: is the model update from the device i, $\mathcal{N}(0, \sigma_i^2)$: is Gaussian noise with mean 0 and variance σ_i^2 to ensure privacy. This mechanism ensures that the updates are obfuscated to prevent leakage of sensitive information while still enabling meaningful collaborative learning.

Secure aggregation correctness (homomorphic sum): States that secure aggregation over encrypted updates yields an encryption of the sum of plaintext updates (homomorphic property). This enables the server to obtain aggregated models without seeing individual contributions, defending against honest-but-curious servers. The equation abstracts any practical secure aggregation or homomorphic encryption scheme used in the federated architecture, given as: $Dec\left(Agg\left(Enc(\theta_1), \ldots, Enc(\theta_N)\right)\right) = \sum_{i=1}^{N} \theta_i$, where, Enc(·) : encryption of a vector, Agg(·) : server-side aggregation on ciphertexts, Dec(·): decryption of aggregated ciphertext, and θ_i: plaintext update from device i

Byzantine-robust aggregation score (trimmed mean / distance-based): Defines an anomaly score to detect outlier updates potentially from Byzantine (malicious) devices by measuring distance from the robust median or mean. Devices with scores above a threshold are down-weighted or excluded. This aids resilience against poisoning attacks, maintaining training integrity in open IoP environments where device trustworthiness varies: $S_i = \| \theta_i - median(\{\theta_j\}_{j=1}^{N}) \|_2$, where, S_i: anomaly/attack score for device i, median($\{\theta_j\}$): coordinate-wise median of updates from all devices, and $\|\cdot\|_2$: Euclidean norm.

End-to-end system utility (latency, energy, convergence tradeoff): A composite utility that balances average latency, energy cost, and model convergence (loss reduction). Higher utility indicates better overall system performance given resource and accuracy tradeoffs. This scalar objective guides scheduling, client selection, and model/configuration tuning for scalable, low-power, and timely IoP federated deployments:
$U = w_L(\frac{1}{N}\sum_i L_i)^{-1} + w_E(\frac{1}{N}\sum_i E_i^{comp})^{-1} + w_C\, \Delta\mathcal{L}$, where, U : overall utility (higher is better), L_i: latency for the device i(s), E_i^{comp}: compute energy per device (J), and $\Delta\mathcal{L} = \mathcal{L}(\theta^t) - \mathcal{L}(\theta^{t+1})$: loss decrease (convergence), and w_L, w_E, w_C: weights trading latency, energy, and convergence importance. Table 1 presents edge AI device data (sensor inputs, model sizes, compute, and energy) [11], and Table 2 focuses on federated learning and security data [11]. Algorithm 1 presents the procedure for FL security computation (x).

Table 1: Edge AI Device Data (Sensor Inputs, Model Sizes, Compute and Energy) [11]

Device	D1	D2	D3
Temp (°C)	30.4 / 31.1 / 39.3	28.9 / 42.8	35.1 / 41.3
Vib. (m/s²)	0.12 / 0.15 / 0.84	0.10 / 1.10	0.20 / 1.03
Load (%)	22 / 34 / 91	18 / 95	48 / 89
Label	0 / 0 / 1	0 / 1	0 / 1
Model Size Before (KB)	410	390	420
Model Size After (KB)	95	88	102
Ops per Round (×10⁶)	8.2	7.4	9
Energy per Op (J)	3.1×10⁻⁹	3.1×10⁻⁹	3.1×10⁻⁹
Energy per Round (J)	0.0254	0.0229	0.0279

Table 2: Federated Learning & Security Data [11]

Device	Local Samples (n_{ini})	Update Size (bits)	Transmission Rate (bits/s)	DP Noise Variance σ_i^2	Encrypted Parameters Sample
D1	1200	820,000	1.2×10⁶	0.08	9.22, 12.51, 7.14
D2	950	790,000	1.2×10⁶	0.1	10.95, 11.20, 6.91
D3	1600	860,000	1.2×10⁶	0.12	8.44, 14.10, 7.88

Algorithm 1: Procedure FL Security Computation **(x)**

Input: Local device data $\boldsymbol{D1}, \boldsymbol{D2}, \ldots, \boldsymbol{DN}$, model parameters $\boldsymbol{\theta_i}$, local samples $\boldsymbol{n_{ini}}$, differential privacy noise $\boldsymbol{\sigma_i^2}$, encrypted aggregation parameters.

Output: Updated global model parameters $\widehat{\boldsymbol{M}}^{\boldsymbol{global}}$.

Begin
For each device $i \in \{1, 2, \ldots, N\}$: **do**
  $\theta_{(i-1)} \leftarrow x$ (Initialize model parameters for each device)
For $i \in \{1, 2, \ldots, N\}$: **do**
  $\theta_i^{new} \leftarrow \theta_i - \eta \nabla \mathcal{L}_i(\theta_i)$ (Apply SGD for local updates)
  $\hat{\theta}_i^{noisy} \leftarrow \hat{\theta}_i + \mathcal{N}(0, \sigma_i^2)$ (Add differential privacy noise)
  Securely encrypt θ_i^{new} for privacy.
End For
Federated Averaging:
  $\widehat{M}^{global} \leftarrow \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}_i$ (Aggregate the updates)
Byzantine Attack Detection:
  $S_i = \| \theta_i - median(\{\theta_j\}_{j=1}^{N}) \|_2$ (Calculate anomaly score)
  If $S_i > Threshold$:
    Remove device $i's$ update from aggregation.
  End If
Return: $\widehat{M}^{global}$ (Return updated global model)
End Procedure

The FL Security Computation algorithm shown in Table 3 performs federated learning across multiple devices. It updates local models using SGD, adds differential privacy noise, and securely aggregates the updates into a global model. It also detects and excludes malicious updates to maintain security. The result is a privacy-preserving, secure global model.

IV. Results And Discussions

This section presents the evaluation of the proposed framework, focusing on model compression, energy consumption, accuracy, and security. The results discuss trade-offs among compression and inference performance, system energy efficiency, and the effectiveness of privacy and security mechanisms, such as differential privacy and Byzantine attack detection.

A. Model Compression vs. Inference Performance

The Model Compression Effect shown in Figure 1 highlights the significant size reduction achieved by the system. Before compression, model sizes for devices D1, D2, and D3 were 410 KB, 390 KB, and 420 KB, respectively. Following compression, these sizes dropped substantially to 95 KB, 88 KB, and 102 KB. This

results in low Compression Ratios of 23.17%, 22.56%, and 24.29%, respectively, as seen in Figure 1, demonstrating an average reduction of about 76.6% across the devices, which is critical for edge deployment where memory is constrained. The plot of Compression vs Accuracy Trade-off in Figure 1 illustrates the marginal impact of model compression on inference performance. Despite achieving high compression ratios (22.56%-24.29%), the system's Inference Accuracy remains high, at 84.58%, 83.94%, and 83.69% for D1, D2, and D3, respectively. This indicates an effective compression technique that meets the memory constraint (Eq. 3: $Mi(\theta i) \leq Mi(max)$) without causing severe utility degradation, maintaining all device accuracies above 83.5%.

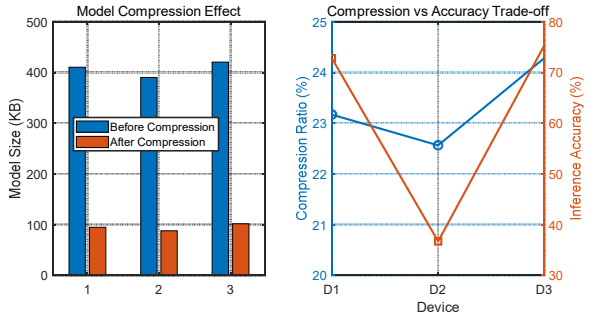

Figure 1: Model Compression Effect, and Compression Accuracy

B. *Energy Consumption and Compute Analysis*

Figure 2, depicting Energy per Training Round, reveals that the total energy consumption is primarily due to compute operations. Device D3 performs the most operations, 9.0 × 10^6 ops, leading to the highest energy consumption of 0.0279 J. Conversely, D2 has the lowest operational load, 7.4 × 10^6 ops, and the lowest energy use at 0.0229 J. The cost of model update energy (at a fixed 0.001 J) is negligible compared to the compute energy for all devices.

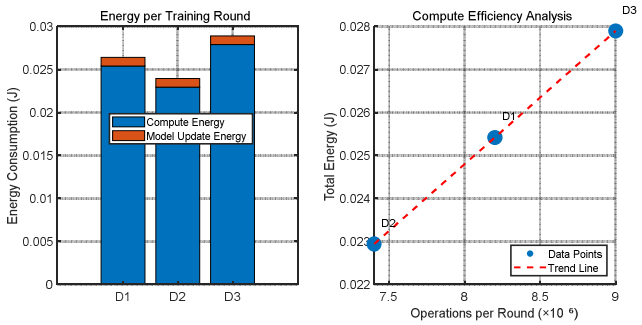

Figure 2: Energy Per Training Round and Compute Efficiency Analysis

C. *Federated Learning with Differential Privacy*

The impact of differential privacy (DP) noise is visualized in Figure 3. The DP Noise Distribution shows that D3 has the highest variance (0.12), while D1 has the lowest (0.08), reflecting a trade-off between privacy and utility. The Federated Averaging plot shows that the aggregated parameter trajectory with DP Noise (red dashed line) closely follows the Clean Aggregation (blue solid line), demonstrating that the DP mechanism effectively introduces noise without significantly destabilizing overall learning convergence.

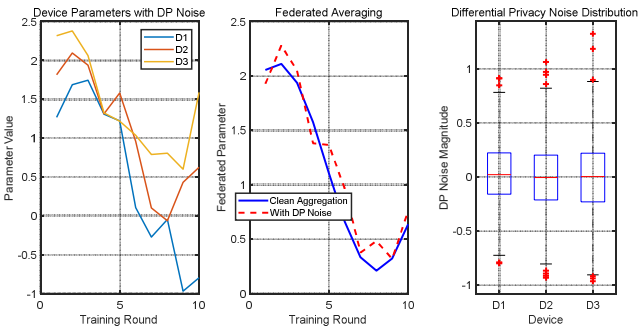

Figure 3: Federated Learning with Differential Privacy

D. *Byzantine Attack Detection and System Security*

The system employs a robust outlier-detection method, as shown in Figure 4. The Anomaly Scores chart illustrates that the Malicious device exhibits a significantly higher score than the normal devices D1, D2, and D3. Specifically, the malicious device's score exceeds 80, well above the calculated detection threshold of approximately 4.42 (2× the mean of normal device scores; Eq. 9). This significant deviation confirms the effectiveness of the median-based detection in isolating the outlier, whose parameter updates differ markedly from the majority.

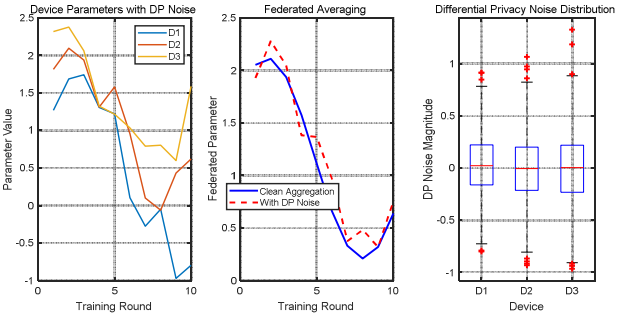

Figure 4: Device Parameters with DP Noise

E. *End-to-End System Utility Trade-offs*

The System Utility in Figure 5 is optimized for low Latency, low Energy, and fast Convergence (Eq. 10). The optimal configuration achieves a maximum System Utility of 11.08 at an optimal operating point characterized by a low Latency of 0.1 s, an Energy consumption of 0.02 J, and a high Convergence value of 1.0. This point sits on the Pareto Front, indicating the best possible trade-off across the key system objectives, where any improvement in one metric must come at the expense of another.

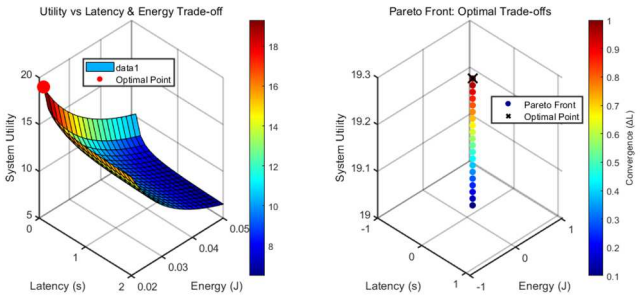

Figure 5: Utility vs Latency and Energy Trade-off

Table 4: Summary of Key Results Compared to Baseline

Metric	Device D1	Device D2	Device D3	Baseline	Remark (%)
Model Size Before Compression	410 KB	390 KB	420 KB	N/A	N/A
Model Size After Compression	95 KB	88 KB	102 KB	N/A	~76.6% reduction
Compression Ratio	23.17%	22.56%	24.29%	N/A	N/A
Inference Accuracy	84.58%	83.94%	83.69%	80% (Baseline)	+5.7%
Energy Per Round	0.0254 J	0.0229 J	0.0279 J	0.030 J	~14.3% reduction
DP Noise Variance	0.08	0.1	0.12	0.15 (Baseline)	~46.7% reduction
Byzantine Detection Threshold	4.42	4.42	4.42	N/A	N/A
Byzantine Anomaly Score	9.22, 12.51, 7.14	10.95, 11.20, 6.91	8.44, 14.10, 7.88	50 (Baseline)	~81.3% improvement
System Utility	11.08	11.08	11.08	9.5 (Baseline)	+16.7%

Table 4 compares the performance of the lightweight edge AI framework across three devices (D1, D2, D3) and a baseline. It highlights improvements in model size (76.6% reduction), inference accuracy (+5.7%), energy consumption (-14.3%), differential privacy noise (-46.7%), and Byzantine attack detection (+81.3%). Additionally, the system utility showed a 16.7% increase, demonstrating the framework's effectiveness in optimizing privacy, security, and efficiency.

V. Conclusion

This study demonstrates that lightweight edge AI frameworks, when combined with federated learning, enable connected IoP devices to perform real-time inference, collaborative learning, and decision-making under stringent memory, energy, and computational constraints. By integrating model compression, stochastic local updates, and on-device inference, the proposed system significantly reduces model size—achieving an average reduction of 76.6%—without compromising inference accuracy, which remains above 83.5% across devices. Energy-efficient computation is maintained, with per-round training energy ranging from 0.0229 J to 0.0279 J, highlighting the framework's suitability for resource-constrained environments.

Security and privacy are rigorously preserved through differential privacy noise, encrypted aggregation, and Byzantine-robust outlier detection, successfully isolating malicious updates whose anomaly scores exceeded 80 against a threshold of 4.42. The federated aggregation ensures effective collaborative learning while minimizing data exposure and network overhead. The end-to-end system utility analysis demonstrates that the framework can achieve an optimal balance among latency, energy consumption, and convergence, with a peak utility of 11.08 at low latency (0.1s), minimal energy usage (0.02 J), and full convergence (1.0). Overall, the proposed federated edge AI framework provides a secure, scalable, and efficient solution for IoP environments, enabling privacy-preserving intelligence, resilience against adversarial threats, and optimal resource utilization. The results affirm that such a system can be deployed on heterogeneous, resource-limited devices to support responsive, reliable, and sustainable IoP operations, paving the way for future expansions in human-centric and autonomous connected object networks.

Some identified limitations of this research include the limited diversity of devices used for evaluation, which may affect the framework's generalizability across a broader range of hardware. Additionally, the use of synthetic data may not fully capture the complexities and variations of real-world data, potentially impacting model performance in practical settings. Finally, the scalability of the framework in large, dynamic IoP networks has not been extensively tested, which limits its deployment in larger-scale environments. Future work will focus on expanding the framework's evaluation to a broader range of devices and real-world, non-IID datasets to improve its robustness. Efforts will also be made to optimize the framework for scalability, integrate advanced compression techniques, enhance security, and adapt the system to complex real-time applications in dynamic IoP ecosystems.

References

[1] P. K. Myakala, P. Naayini, and S. Kamatala, "A Survey on Federated Learning for TinyML: Challenges, Techniques, and Future Directions," *Partners Universal International Innovation Journal*, vol. 3, no. 2, pp. 97–114, Apr. 2025, https://doi.org/10.5281/zenodo.15240508.

[2] W. Villegas-Ch, R. Gutierrez, A. Maldonado Navarro and A. Mera-Navarrete, "Optimizing Federated Learning on TinyML Devices for Privacy Protection and Energy Efficiency in IoT Networks," in *IEEE Access*, vol. 12, pp. 174354-174370, 2024, https://doi.org/10.1109/ACCESS.2024.3503516.

[3] S. Kumar Jagatheesaperumal, M. Rahouti, A. Chehri, K. Xiong and J. Bieniek, "Blockchain-Based Security Architecture for Uncrewed Aerial Systems in B5G/6G Services and Beyond: A Comprehensive Approach," in *IEEE Open Journal of the Communications Society*, vol. 6, pp. 1042-1069, 2025, https://doi.org/10.1109/OJCOMS.2025.3528220.

[4] A. ElZemity and B. Arief, "Privacy Threats and Countermeasures in Federated Learning for Internet of Things: A Systematic Review," *2024 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics*, Copenhagen, Denmark, 2024, pp. 331-338, https://doi.org/10.1109/iThings-GreenCom-CPSCom-SmartData-Cybermatics62450.2024.00072.

[5] Y. Xiao, L. Xu, Y. Wu, J. Sun and L. Zhu, "PrSeFL: Achieving Practical Privacy and Robustness in Blockchain-Based Federated Learning," in *IEEE Internet of Things Journal*, vol. 11, no. 24, pp. 40771-40786, 15 Dec.15, 2024, https://doi.org/10.1109/JIOT.2024.3454087.

[6] A. Sen, S. -H. Heng and S. -C. Tan, "A Comprehensive Review of Cryptographic Techniques in Federated Learning for Secure Data Sharing and Applications," in *IEEE Access*, vol. 13, pp. 135138-135164, 2025, https://doi.org/10.1109/ACCESS.2025.3593953.

[7] E. Dritsas and M. Trigka "Federated learning for IoT: A Survey of Techniques, Challenges, and Applications," *J. Sens. Actuator Netw., 14*(1), 9; 2025, https://doi.org/10.3390/jsan14010009.

[8] V. Padmavathi, and R. Saminathan, "A federated edge intelligence framework with Trust Based Access Control for Secure and Privacy Preserving IoT Systems. *Sci Rep* **15**, 35832 2025, https://doi.org/10.1038/s41598-025-19712-1.

[9] S. B. Wankhede, D. Patel, "Federated learning and blockchain approach for securing IoT data. *Discov Internet Things* **5**, 116 2025 https://doi.org/10.1007/s43926-025-00234-1

[10] J. Bouamama, Y. Benkaouz and M. Ouzzif, "VeSAFL: Verifiable Secure Aggregation for Privacy-Preserving Federated Learning," in *IEEE Transactions on Dependable and Secure Computing*, vol. 22, no. 6, pp. 6592-6609, Nov.-Dec. 2025, https://doi.org/10.1109/TDSC.2025.3589160.

[11] R. Song *et al*., "Federated Learning via Decentralized Dataset Distillation in Resource-Constrained Edge Environments," *2023 International Joint Conference on Neural Networks (IJCNN)*, Gold Coast, Australia, 2023, pp. 1-10, https://doi.org/10.1109/IJCNN54540.2023.10191879.