

Lightweight Federated Intrusion Detection with Secure Aggregation for Healthcare IoT

Rakesh V S¹, Shankara Gowda S R², Anupama Vijaykumar³, Mithun B N⁴, Aparna K S⁵ and Alpha Vijayan⁶

¹Cambridge Institute of Technology, K R Puram, Bengaluru, India

Email: rakesh.tech102@gmail.com

²Don Bosco Institute of Technology, Bengaluru, India

Email: shankargowdasr@gmail.com

³Dayananda Sagar College of Engineering, Bengaluru, India

Email: anupamarevadi@gmail.com

⁴A P S College of Engineering, Bengaluru, India

Email: mithunbn.apsce@gmail.com

⁵Global Academy of Technology, Bengaluru, India

Email: aparna.vastrad@gmail.com

⁶Christ University, Bengaluru, India

Email: alphavijayan@gmail.com

Abstract— Healthcare IoT environments, including wearables, bedside monitors, and medical sensors, generate sensitive and diverse data streams that are highly vulnerable to cyberattacks. Centralized intrusion detection methods increase privacy risks and cause significant communication overhead, making them unsuitable for resource-limited medical systems. This work introduces a Lightweight Federated Intrusion Detection System (LF-IDS) that combines Convolutional Neural Networks (CNN) for feature extraction with an Isolation Forest for anomaly detection. Hospitals train models locally and only share encrypted updates, which are then aggregated using secure protocols to protect patient privacy. Experiments on the CICIOT2023 dataset show that LF-IDS achieves 96.2% accuracy with a 36% reduction in communication costs, while remaining effective under non-IID conditions, making it a practical and privacy-preserving solution for healthcare IoT security.

Index Terms— Anomaly Detection, Federated Learning, Healthcare IoT, Intrusion Detection, and Privacy-Preserving.

I. INTRODUCTION

The integration of Healthcare Internet of Things (IoT) devices—such as wearable sensors, bedside monitors, and interconnected hospital systems—has transformed medical practice by enabling continuous patient monitoring and real-time clinical decision support. However, the highly connected nature of these systems has also introduced new avenues for cyberattacks, threatening both service availability and the confidentiality of sensitive medical records. Conventional intrusion detection systems (IDS), which depend on centralized data collection, are increasingly unsuitable in this context because they expose patient data to privacy risks and place heavy communication demands on bandwidth-limited medical IoT networks.

Federated learning (FL) has emerged as a promising paradigm to overcome these limitations by enabling institutions to collaboratively train models without sharing raw data, thereby preserving privacy in sensitive healthcare environments [1]. Studies in smart healthcare settings show that FL can significantly enhance privacy while supporting secure knowledge transfer across medical centers. In parallel, IDS frameworks based on FL have been tested for IoT applications, demonstrating that distributed training can reduce communication overhead while sustaining high detection accuracy [2].

Despite these advances, several challenges remain unaddressed. Healthcare IoT infrastructures continue to be attractive targets for sophisticated cyberattacks, and existing IDS solutions often fail to capture the diversity of attack patterns unique to medical environments [3]. Moreover, many IoT devices in hospitals operate with strict resource constraints, making it impractical to deploy heavy deep learning architectures. Recent research into lightweight intrusion detection models confirms that efficient neural designs can deliver strong accuracy while avoiding excessive computational costs [4]. Motivated by these gaps, this paper proposes a lightweight, privacy-preserving intrusion detection framework that combines Convolutional Neural Networks (CNN) and Isolation Forests within a federated learning setting, strengthened by secure aggregation to protect confidentiality during collaboration.

II. RELATED WORKS

Research in recent years has increasingly focused on intrusion detection in IoT environments using Federated Learning (FL), lightweight machine learning models, or hybrid privacy-aware frameworks. These studies examine trade-offs among accuracy, latency, communication costs, model complexity, and privacy, especially considering the resource constraints of IoT devices and the heterogeneity of network data. Table I summarizes the existing work.

TABLE I. COMPARISON OF RECENT IDS APPROACHES USING FEDERATED AND LIGHTWEIGHT MODELS

Ref. No.	Technique	Strength	Limitation
[5]	Semi-decentralized FL with BiLSTM	Improves accuracy under heterogeneous data; supports distributed IoT clients	BiLSTM requires higher resources, which is not ideal for very constrained devices.
[6]	FL with CNN/BiLSTM tested on edge hardware	Demonstrates feasibility on Raspberry Pi; balances accuracy and cost	Deep models require increased computation and memory, while privacy aspects are limited.
[7]	FL with clustering and DL classifiers	Handles non-IID data; better global convergence	Complex architecture; latency and energy cost are higher.
[8]	Decision tree-based lightweight IDS	Low memory use; high accuracy; real device validation	Simpler models may not capture advanced attack patterns.
[9]	FL with knowledge distillation	Reduces model size; improves generalization across clients	Limited real-world testing; additional steps add complexity.
[10]	Lightweight FL updating only weight changes	Cuts communication overhead; faster model updates	Update security and poisoning resistance, which were not deeply addressed.
[11]	Hybrid feature selection + FL	Detects multiple attack types; optimized feature set lowers latency	Feature selection adds local computation load.
[12]	FL-enabled lightweight IDS	Real-time and resource-efficient; suitable for IoT networks	Evaluated on limited datasets; lacks strong, secure aggregation.

Recent studies identify several common gaps. Many depend on deep models, such as BiLSTMs, which are impractical for resource-limited IoT and wearable medical devices. Although federated learning has been well researched, only a few studies have combined it with secure aggregation or other cryptographic methods to protect privacy in healthcare settings. Real-world testing on medical IoT hardware is limited, with most results coming from simulations. Lastly, while the issues of non-IID and heterogeneous data are recognized, current approaches seldom strike a balance between strong privacy protections and lightweight efficiency.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

The proposed methodology introduces a lightweight, privacy-preserving intrusion detection framework tailored for healthcare IoT environments. The framework, as illustrated in Fig. 1, adopts a layered design, comprising a data layer for local traffic monitoring and alerting, a federated learning layer for collaborative training, and a control and trust layer for security and governance. This setup ensures a balance between detection accuracy, efficiency, and patient data confidentiality, making it suitable for resource-constrained medical IoT systems.

A. Local Model Training in Data Plane

At each hospital, IoT devices (e.g., wearables, infusion pumps, bedside monitors) generate traffic traces. Let the traffic window be observed at the hospital be represented as a matrix as given in (1).

$$X_i \in R^{T \times d} \quad (1)$$

where T is the window length and d is the number of features (e.g., flow size, packet counts, timing).

Preprocessing: Each feature is normalized as shown in (2).

$$\tilde{X}_i(t, j) = \frac{X_i(t, j) - \mu_j}{\sigma_j + \epsilon} \quad (2)$$

where μ_j , σ_j are the local mean and standard deviation of feature j , and ϵ avoids division by zero.

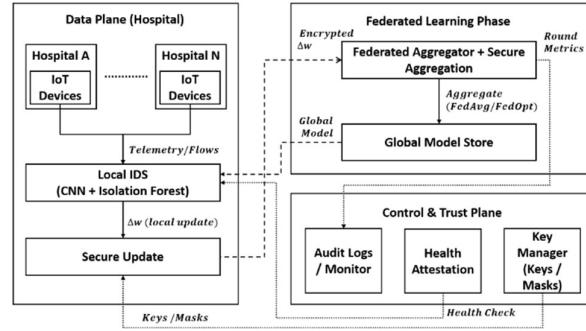


Figure 1. System Architecture of Lightweight Federated IDS

Feature Extraction (CNN): A convolutional filter $W^{(l)}$ of size K slides over the window to extract temporal-spatial correlations, which is as represented in (3).

$$H_c^{(l)}(t) = \sigma \left(\sum_{j=1}^d \sum_{\tau=0}^{K-1} W_{c,j,r}^{(l)} \tilde{X}_i(t - \tau, j) + b_c^{(l)} \right) \quad (3)$$

where $\sigma(\cdot)$ is a non-linear activation. After several convolutional/pooling layers, the CNN produces a latent representation as given in (4).

$$z_i = f_\theta(\tilde{X}_i) \in \mathbb{R}^p \quad (4)$$

with θ denoting CNN parameters.

Anomaly Detection (Isolation Forest): Each embedding z_i is evaluated by an Isolation Forest. For a dataset of size n , the anomaly score is calculated using (5).

$$s(z_i, n) = 2^{-\frac{E[h(z_i)]}{c(n)}}, \quad c(n) = 2H_{n-1} - \frac{2(n-1)}{n} \quad (5)$$

where $h(z_i)$ is the path length of z_i in a tree, and H_n is the n -th harmonic number. Predictions are made by thresholding, given in (6).

$$\hat{y}_i = 1\{s(z_i, n) > \tau\} \quad (6)$$

where τ is a detection threshold.

Secure Update Formation: Instead of transmitting raw traffic, each hospital trains locally and computes a model update vector as given in (7).

$$\Delta w_i^{(t)} = w_i^{(t)} - w^{(t-1)} \quad (7)$$

where $w^{(t-1)}$ is the global model from the previous round and $w_i^{(t)}$ is the locally updated model. This ensures privacy, as only parameter differences are shared.

B. Secure Model Update Sharing in Federated Learning Plane

Once local models are trained, hospitals share only model updates (parameters/gradients) rather than raw datasets. The secure updates $\Delta w_i^{(t)}$ are protected before transmission. Two cryptographic strategies are supported.

Additive Masking (Secure Aggregation): Each client masks its update with a random vector m_i as shown in (8).

$$\Delta w_i^{(t)} = \Delta w_i^{(t)} + m_i, \quad \sum_{i=1}^N m_i = 0 \quad (8)$$

so that only the aggregated sum is revealed to the server.

Homomorphic Encryption (HE): Using an additively homomorphic scheme (e.g., Paillier), updates are encrypted as given in (9).

$$\Delta w_i^{(t)} = \text{Enc}_{pk}(\Delta w_i^{(t)}), \quad \sum_{i=1}^N \Delta w_i^{(t)} = \text{Enc}_{pk}\left(\sum_i \Delta w_i^{(t)}\right) \quad (9)$$

The Federated Aggregator receives all encrypted/masked updates and applies a standard optimization rule, e.g., FedAvg, which is computed using (10).

$$w^{(t)} = w^{(t-1)} + \eta_t \frac{1}{N} \sum_{i=1}^N \Delta w_i^{(t)} \quad (10)$$

where η_t is the learning rate and N is the number of participating hospitals. The updated global model $w^{(t)}$ is redistributed to clients for the next training round.

C. Global Model Aggregation in Control and Trust Plane

The central aggregator receives encrypted updates from participating hospitals. Through secure aggregation, the updates are combined into a single global model. This process not only protects privacy but also reduces the risk of model inversion attacks.

Key Manager: Generates and distributes cryptographic keys or random masks for secure aggregation as shown in (11).

$$k_i \leftarrow \text{PRF}(s, id_i) \quad (11)$$

where PRF does a secret seed a pseudorandom function s .

Health Attestation: Each client submits integrity proofs, modelled as a verification function given in (12).

$$\text{Attest}(M_i^{(t)}) \in \{0, 1\} \quad (12)$$

where 1 denotes a valid update and 0 rejects compromised clients.

Audit Logs / Monitor: All training rounds are logged as tuples $L(i, t, \Delta w_i^{(t)})$, enabling detection of anomalies (e.g., poisoning or unusual update magnitudes).

IV. EXPERIMENTAL SETUP

The experimental setup defines the datasets, system parameters, and evaluation criteria adopted to assess the effectiveness, efficiency, and privacy guarantees of the proposed LF-IDS model. Table II summarizes the dataset partitions and experimental parameters considered during evaluation.

Dataset: The evaluation utilizes the CICIOT2023 dataset, which comprises benign IoT traffic and various attack categories, including DDoS, brute force, infiltration, and data theft. To emulate heterogeneous healthcare

environments, the dataset was partitioned into non-identically distributed (non-IID) subsets, each representing a simulated hospital client.

Features and Preprocessing: Raw PCAP/NetFlow traces were converted into flow-level and statistical features, including packet/byte counts, flow duration, inter-arrival times, protocol flags, and packet size statistics. All features were normalized using z-score scaling and segmented into fixed-length traffic windows ($T = 100$ packets).

Federated Setup: The system was configured with 10 simulated hospitals as federated clients. Training was performed for 50 communication rounds, with five local epochs per round and a batch size of 64. The Adam optimizer ($\eta = 0.001$) was used for local updates, and the FedAvg algorithm aggregated encrypted updates at the server.

Secure aggregation was realized through additive masking, ensuring that only the aggregated model update was visible to the server. Two baselines were considered: a standalone IDS trained per client and a centralized IDS trained on pooled data.

Model: Each client deployed a hybrid model combining a CNN (comprising three convolutional blocks) for feature extraction with an Isolation Forest for anomaly detection. CNN weights were refined locally, while the Isolation Forest operated on learned embeddings.

Environment: Experiments were conducted on a workstation equipped with an Intel i7-11700 CPU, 32 GB of RAM, and an NVIDIA RTX 3080 GPU, running Ubuntu 22.04, Python 3.10, and PyTorch 2.1. Federated learning was implemented using the Flower framework, and communication security was enforced through TLS channels.

Evaluation Metrics: Performance was assessed using Accuracy, Precision, Recall, F1-score, AUC, and Communication Overhead (average client update size per round).

TABLE II. KEY EXPERIMENTAL SETTINGS

Aspect	Setting
Dataset	CICIoT2023; non-IID partitions for clients
Features	Flow statistics + timing; z-score normalization; window length = 100
Clients / Rounds	10 simulated clients / 50 rounds
Local Training	5 epochs per round; batch size = 64; Adam ($\eta=0.001$)
Aggregation	FedAvg; secure aggregation via additive masking
Model	CNN (3 convolutional layers) + Isolation Forest
Environment	Intel i7-11700, 32 GB RAM, RTX 3080; Ubuntu 22.04; PyTorch 2.1; Flower
Evaluation Metrics	Accuracy, Precision, Recall, F1-score, AUC, Communication Overhead

V. RESULTS AND DISCUSSION

The results demonstrate how the proposed LF-IDS achieves near-centralized detection accuracy while significantly reducing communication overhead and ensuring privacy, as discussed in the following subsections.

A. Detection Performance

The comparative results across different IDS approaches are presented in Table 3. As shown, the proposed LF-IDS achieves 96.2% accuracy, closely matching the centralized IDS (96.8%) while outperforming both the standalone IDS (91.2%) and FL with BiLSTM (95.1%).

TABLE III. COMPARISON OF IDS PERFORMANCE

Method	Accuracy (%)	Precision	Recall	F1-score	AUC
Standalone IDS	91.2	0.89	0.9	0.89	0.92
Centralized IDS	96.8	0.96	0.95	0.96	0.97
FL with BiLSTM	95.1	0.94	0.93	0.94	0.95
Proposed LF-IDS	96.2	0.96	0.94	0.95	0.97

B. Convergence Behavior

The convergence trends across 50 federated rounds are illustrated in Fig. 2. LF-IDS achieves ~96% accuracy by the 30th round, whereas the BiLSTM baseline requires nearly 50 rounds.

The standalone IDS appears as a flat line at ~91%, reflecting the ceiling of isolated local training without federation.

The centralized IDS is omitted, as it trains on aggregated data in a conventional manner and does not involve communication rounds.

C. Communication and Computation Overhead

As illustrated in Fig. 3, LF-IDS reduces the average update size per client to 16 MB per round, representing a 36% reduction compared to BiLSTM (25 MB). This demonstrates the communication efficiency of the lightweight design.

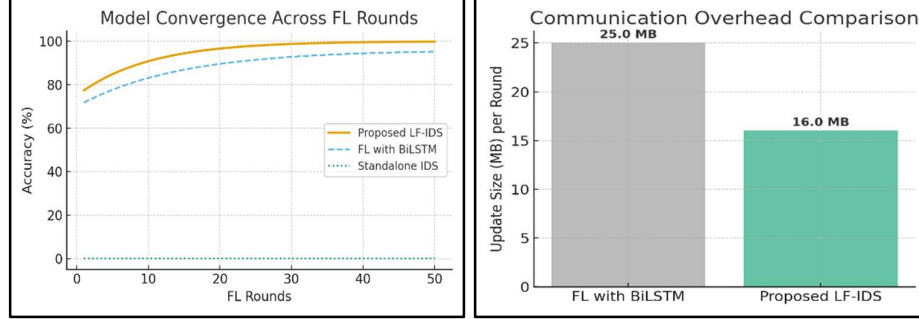


Figure 2: Accuracy vs. FL Rounds

Figure 3: Communication Overhead Comparison

D. Impact of Non-IID Data

The effect of skewed client distributions on detection accuracy is presented in Figure 4. Even under an 80% data imbalance, LF-IDS maintains an accuracy above 93%, whereas BiLSTM and standalone IDS drop below 87% and 77%, respectively.

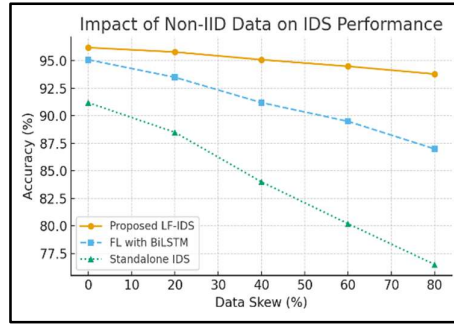


Figure 4: Non-IID Robustness

III. CONCLUSIONS

The proposed LF-IDS demonstrated near-centralized accuracy while significantly decreasing communication overhead and maintaining privacy, making it ideal for healthcare IoT environments. By combining CNN-based feature extraction with Isolation Forest detection, the framework achieved strong performance even under non-IID client distributions. Secure aggregation further provided robust protection against data leakage and model inversion attacks. Future work may focus on expanding the approach to real-world hospital deployments, adding adversarial defense mechanisms, and improving energy efficiency on edge devices. Broader evaluations on cross-domain IoT datasets could also confirm the model's scalability and adaptability.

REFERENCES

- [1] Abbas, Syed Raza, Zeeshan Abbas, Arifa Zahir, and Seung Won Lee. 2024. "Federated Learning in Smart Healthcare: A Comprehensive Review on Privacy, Security, and Predictive Analytics with IoT Integration" *Healthcare* 12, no. 24: 2587. <https://doi.org/10.3390/healthcare12242587>.
- [2] Devine, Mark, Saeid Pourroostaei Ardakani, Mohammed Al-Khafajiy, and Yvonne James. 2025. "Federated Machine Learning to Enable Intrusion Detection Systems in IoT Networks" *Electronics* 14, no. 6: 1176. <https://doi.org/10.3390/electronics14061176>.
- [3] Doménech, Jordi, Olga León, Muhammad Shuaib Siddiqui, and Josep Peguerols. 2025. "Evaluating and Enhancing Intrusion Detection Systems in IoMT: The Importance of Domain-Specific Datasets." *Internet of Things* 32: 101631. <https://doi.org/10.1016/j.iot.2025.101631>.

- [4] Ullah, S., Wu, J., Kamal, M.M. et al. KronNet a lightweight Kronecker enhanced feed forward neural network for efficient IoT intrusion detection. *Sci Rep* 15, 20850 (2025). <https://doi.org/10.1038/s41598-025-08921-3>.
- [5] Alsaleh, Shuroog; Menai, M. E. B.; Al-Ahmadi, S. 2025. "A Heterogeneity-Aware Semi-Decentralized Model for a Lightweight Intrusion Detection System for IoT Networks Based on Federated Learning and BiLSTM." *Sensors*, 25(4): 1039. <https://doi.org/10.3390/s25041039>.
- [6] Albanbay, Nurtay, Yerlan Tursynbek, Kalman Graffi, Raissa Uskenbayeva, Zhuldyz Kalpeyeva, Zhastalap Abilkaiyr, and Yerlan Ayapov. 2025. "Federated Learning-Based Intrusion Detection in IoT Networks: Performance Evaluation and Data Scaling Study" *Journal of Sensor and Actuator Networks* 14, no. 4: 78. <https://doi.org/10.3390/jsan14040078>.
- [7] Karunamurthy, A., Vijayan, K., Kshirsagar, P.R. et al. An optimal federated learning-based intrusion detection for IoT environment. *Sci Rep* 15, 8696 (2025). <https://doi.org/10.1038/s41598-025-93501-8>.
- [8] Pandey, V.K., Sahu, D., Prakash, S. et al. A lightweight framework to secure IoT devices with limited resources in cloud environments. *Sci Rep* 15, 26009 (2025). <https://doi.org/10.1038/s41598-025-09885-0>.
- [9] Peng H, Wu C, Xiao Y. FD-IDS: Federated Learning with Knowledge Distillation for Intrusion Detection in Non-IID IoT Environments. *Sensors (Basel)*. 2025 Jul 10;25(14):4309. doi: 10.3390/s25144309. PMID: 40732440; PMCID: PMC12300861.
- [10] Lee, Sun-Jin, and Il-Gu Lee. 2025. "Lightweight Federated Learning-Based Intrusion Detection System for Industrial Internet of Things." *ICT Express* 11, no. 4: 690–695. <https://doi.org/10.1016/j.ict.2025.05.002>
- [11] Prathap Mani, Arthi D, Periyakaruppan K and Surenderkumar S, "A Lightweight and Federated Machine Learning Based Intrusion Detection System for Multi Attack Detection in IoT Networks", *Journal of Machine and Computing*, vol.5, no.1, pp. 421-429, January 2025, doi: 10.53759/7669/jmc202505033.
- [12] Devi, Manu, Priyanka Nandal, and Harkesh Schrawat. 2025. "Federated Learning-Enabled Lightweight Intrusion Detection System for Wireless Sensor Networks: A Cybersecurity Approach against DDoS Attacks in Smart City Environments." *Intelligent Systems with Applications* 27: 200553. <https://doi.org/10.1016/j.iswa.2025.200553>.