Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.082021

ARTICLE

Resilient Federated Ensemble Learning for IoT Intrusion Detection in Adversarial and Imbalanced Environments

Arvind Prasad[1,*], Ibrahim Aljubayri[2], Mohammad Zubair Khan[3,*] and Abdulfattah Noorwali[4]

[1]Department of Computer Engineering & Applications, GLA University, Mathura, India
[2]Department of Computer Science and Information, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia
[3]Faculty of Computer and Information Systems, Islamic University of Madinah, Medina, Saudi Arabia
[4]Electrical Engineering Department, Umm Al-Qura University, Makkah, Saudi Arabia

*Corresponding Authors: Arvind Prasad. Email: arvind.prasad@gla.ac.in; Mohammad Zubair Khan. Email: zubair.762001@gmail.com

Received: 11 March 2026; Accepted: 18 May 2026; Published: 30 June 2026

ABSTRACT: Intrusion detection in large-scale IoT deployments becomes particularly challenging during ongoing attack scenarios, where malicious traffic may temporarily dominate benign traffic. In such conditions, streaming network data exhibits severe class imbalance in favor of attack traffic, while device behavior remains heterogeneous, non-identically distributed (non-IID), and temporally evolving. Within federated learning environments, this imbalance can destabilize early aggregation rounds, dominant attack gradients bias the global model, distort decision boundaries, and degrade reliable discrimination of residual benign behavior. Since the server has no access to raw data, these effects can persist across communication rounds if not addressed at initialization. This article addresses the problem of distributed intrusion detection in streaming IoT networks under severe class imbalance and partial adversarial behavior. In ongoing attack scenarios, malicious traffic may dominate streaming observations, creating an inverted imbalance where benign behavior becomes underrepresented. To mitigate these issues, we propose a resilient federated ensemble framework with three key components: a similarity-guided balancing phase that selects a structurally diverse subset of majority-class samples to form a balanced initialization dataset, an incremental ensemble composed of Bernoulli Naïve Bayes, Passive–Aggressive, and SGD classifiers for pre-training over this data to produce a warm-start global model, and an accuracy-gating mechanism that accepts only performance-preserving local updates during online training. The proposed approach is evaluated on flow-level data from the CICIoT2023 benchmark that demonstrated stable client-wise convergence with high accuracy and consistent minority-class discrimination, even under skewed attack-dominant distributions. Under complete label-flipping poisoning, corrupted updates are systematically rejected, preventing global degradation. Ablation analysis confirms that removing structured initialization significantly increases early instability. The results indicate that balanced global conditioning and selective federation are critical for maintaining detection reliability in streaming IoT systems operating under sustained attack conditions.

KEYWORDS: Federated learning; IoT intrusion detection; streaming analytics; class imbalance; adversarial robustness

1 Introduction

The Internet of Things (IoT) is now pervasive, with billions of devices and sensors operating in smart homes, industry, healthcare, and cities [1]. In the past few years alone, the number of IoT devices has increased considerably, resulting in a large growth of cyber attacks. Many IoT endpoints are low-cost and under-protected, so they enlarge the attack surface (botnets, malware, ransomware, etc.). IoT network traffic is

extremely imbalanced [2]. Client data are heterogeneous and non-identically distributed (non-IID). One smart camera's traffic (e.g., home surveillance) can be very different from a temperature sensor. Clients collect different volumes of data and different feature patterns. In practice, this means a single global model must reconcile many local distributions. In this context, intrusion detection systems (IDS) are essential. IDS has acquired a key role in defending IoT environments [3]. Monitoring the high-volume streams of IoT data is critical because the continuous increase in cyber attacks makes automated defense a necessity [4–6].

Traditional IDS designs assume stable traffic and centralized visibility, whereas IoT environments are highly heterogeneous and continuously evolving [7]. Device behavior varies across deployment contexts, firmware states, workloads, and attack conditions, resulting in non-IID and temporally drifting traffic distributions.

Streaming IoT traffic further complicates intrusion detection because data arrives continuously under low-latency constraints [8]. Models must therefore support incremental adaptation while remaining robust to concept drift, evolving device behavior, and emerging attacks.

A major challenge in sustained IoT attack scenarios is inverted class imbalance, where malicious traffic temporarily dominates benign communication (attack >> benign) [9]. During large-scale attacks, streaming data becomes attack-heavy, causing optimization to be dominated by malicious gradients and biasing decision boundaries toward attack patterns. In federated environments, these effects are amplified because the server aggregates model updates without direct visibility into class distributions [10,11].

In sustained attack scenarios, the local class prior may satisfy $P(y = 1) \gg P(y = 0)$, where $y = 1$ denotes malicious traffic and $y = 0$ benign traffic. Under such skew, the learning objective becomes dominated by gradients from the majority attack class, shifting the decision boundary toward malicious regions and weakening recognition of minority benign behavior. In federated settings, this effect can be amplified because the server observes model updates rather than the underlying class distributions.

Federated Learning (FL) is often suggested for IoT to save bandwidth and protect privacy [12–14]. In FL, the devices keep their raw data and only send model updates to a central server. This follows privacy laws and saves data. However, since the server never sees the raw data, it cannot see if the data distribution is wrong. The server just combines updates from many different, and maybe even malicious, devices. Also, communication is limited, and some devices might go offline.

In federated IoT intrusion detection, global initialization strongly influences early-round stability and convergence behavior [15,16]. Under attack-heavy distributions, poorly conditioned initialization may delay reliable discrimination of sparse benign traffic and amplify instability across subsequent communication rounds.

The interaction between streaming non-IID traffic, local imbalance, federated aggregation constraints, and adversarial behavior creates a cascading instability effect during deployment [17–19]. The overall attack and instability pathway is summarized in Fig. 1.

As shown in Fig. 1, heterogeneous IoT clients generate skewed local updates under class imbalance and temporal drift. These updates are aggregated without access to raw data, which can amplify majority-class dominance during early rounds. The resulting unstable initialization suppresses rare attack gradients, reinforces bias across clients, and delays detection of zero-day activity, thereby expanding the security exposure window during ramp-up.

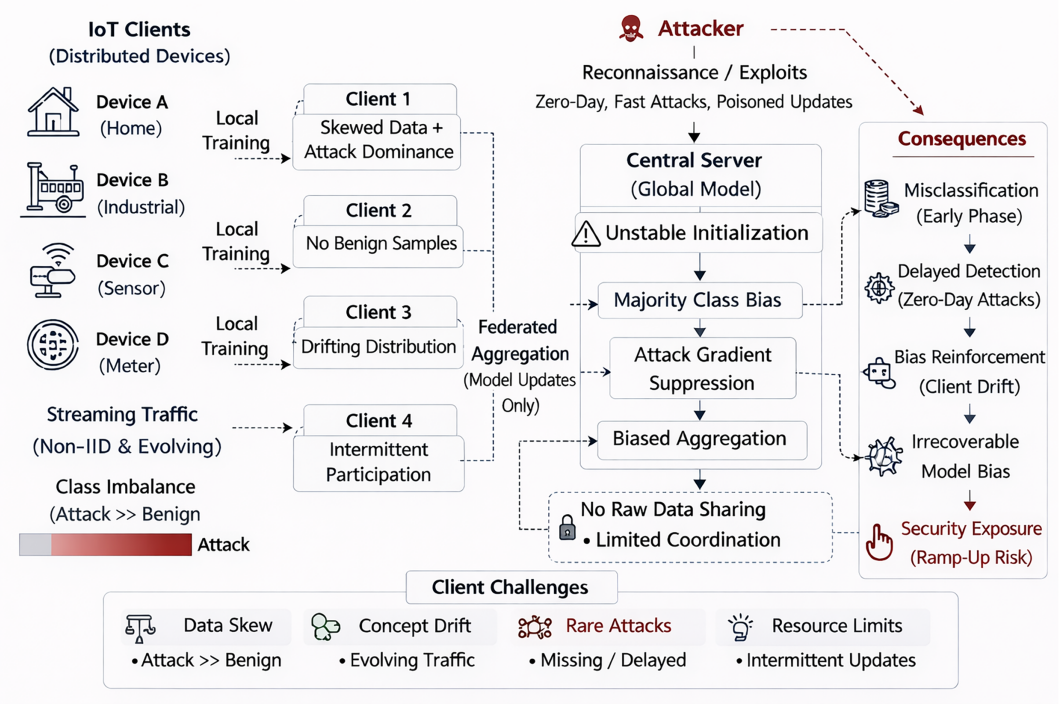

Figure 1: Detailed attack propagation and instability mechanisms in federated IoT intrusion detection under streaming, non-IID, and imbalanced conditions.

These challenges become more severe in the presence of adversarial behavior such as poisoning or label-flipping attacks. Existing defenses typically assume a reasonably stable initial model [20]. Consequently, poorly conditioned initialization can amplify both natural non-IID effects and malicious client influence during early federated rounds [21–24].

In this work, we propose a resilient federated ensemble learning method for IoT. We focus on cases with attacks and imbalanced data. We do not try to solve every problem, but we focus on how streaming data, imbalance, and limited coordination affect the stability of the early model. Our goal is to make the system work in the real world, not just in theory.

The main contributions of this paper are summarized as follows:

- A similarity-guided balancing mechanism that selects structurally diverse dominant-class samples to construct a balanced and information-rich initialization dataset.
- A cold-start-aware federated initialization strategy that improves early-round stability under attack-heavy IoT traffic.
- A streaming ensemble learning design based on incremental classifiers for continuous adaptation without full retraining.
- An accuracy-gated client update mechanism that rejects poisoned, noisy, or performance-degrading local updates before aggregation.
- A unified framework that jointly addresses imbalance, non-IID streaming learning, and adversarial robustness in federated IoT intrusion detection.
- Experimental validation through client-wise analysis, sensitivity studies, poisoning experiments, FGSM rejection tests, and ablation results.

The remainder of this paper is organized as follows. Section 2 reviews existing literature on federated intrusion detection, class imbalance mitigation, streaming learning, and adversarial robustness in IoT environments. Section 3 presents the proposed resilient federated intrusion detection framework. Section 4 describes the experimental setup and evaluates the proposed approach through client-wise performance analysis, poisoning attack experiments, and ablation studies. Finally, Section 5 concludes the paper and outlines directions for future research.

2 Related Work

Recent research has explored multiple directions, including federated learning, class imbalance mitigation, streaming adaptation, and adversarial robustness. While several studies report high detection accuracy, their assumptions, optimization objectives, and deployment settings vary significantly. It is therefore necessary to examine these works in a structured manner to understand the practical limitations that remain unresolved, particularly under streaming and attack-dominant traffic conditions. The following subsections categorize the literature accordingly.

2.1 Federated Learning for IDS on CICIoT2023

Abid et al. [25] proposed a hybrid intrusion detection system for IoT environments combining Grey Wolf Optimization (GWO), LightGBM, and a CNN classifier. GWO performs wrapper-based feature selection to reduce dimensionality. LightGBM generates leaf-index representations through gradient-boosted trees, which are embedded and processed by a Conv1D-based CNN for final classification. The system is evaluated on the CICIoT2023 and CICIoMT2024 datasets. The authors report improved detection accuracy, reduced false alarm rates, and enhanced computational efficiency compared to standalone RF, SVM, DNN, LSTM, CNN, and attention-based models. The approach is designed for deployment in fog-layer IoT architectures. The authors achieved an average accuracy of 87.93%. A wrapper-based dimensionality reduction using GWO and efficient leaf-wise boosting from LightGBM was implemented, which achieved high performance on a reduced feature space (18 features for CICIoT2023, 5 for CICIoMT2024) and achieved improved computational efficiency for fog-layer deployment.

Torre et al. [26] proposed a federated learning-based intrusion detection system (IDS) for IoT environments implementing a 1D CNN, with explicit integration of three privacy-preserving (PP) mechanisms: Differential Privacy (DP), Diffie–Hellman (DH) key exchange, and Fully Homomorphic Encryption (FHE). The proposed approach is evaluated on TON-IoT, IoT-23, BoT-IoT, CIC IoT 2023, CIC IoMT 2024, RT-IoT 2022, and EdgeIIoT. Empirical results indicate strong predictive performance with limited computational overhead when PP layers are enabled. The study emphasizes operational feasibility rather than purely theoretical guarantees, and includes convergence, timing, and comparative analyses. The rich feature sets used Network traffic features extracted from CSV/PCAP-derived flow records, including protocol-based, behavioral, telemetry, and IoT device communication attributes. The main operational advantage is layered privacy integration within FL. The simultaneous use of DP, DH exchange, and FHE is uncommon in IDS literature. The approach reduces exposure of raw gradients and model parameters during aggregation. Empirically, privacy enforcement adds only 10% computational overhead, which is relatively modest given the cryptographic operations involved. Cross-dataset validation improves external validity.

Okey et al. [27] presented RAID-KL, a hybrid KL–JS knowledge distillation framework for resource-aware IoT intrusion detection using 1D CNN teacher–student models, adaptive temperature scaling, SMOTE balancing, and SHAP explanations on CICIoT2023, CICIoMT2024, and . The formulation is mathematically coherent, and the experimental scope is reasonably broad. However, several core elements need tightening:

hyperparameter justification is weak, resource measurements lack methodological clarity, and ablation depth is insufficient to isolate the hybrid loss contribution.

2.2 Class Imbalance in Intrusion Detection on CICIoT2023

Wahab et al. [28] proposed an optimized CNN, DNN, and Transformer architectures that mitigate class imbalance issues through SMOTE-based oversampling applied strictly to training folds to avoid leakage. Log normalization and Min–Max scaling stabilize feature variance before balancing. Empirically, balancing significantly improved recall for minority classes, especially in the 12-class setting. Without rebalancing, preliminary experiments showed biased decision boundaries favoring high-frequency DDoS floods. However, synthetic oversampling introduces risks of overfitting, particularly in highly sparse minority distributions. Lightweight CNN/DNN configurations adapted for IoT resource constraints. The authors selected feature categories such as statistical network flow features, time-based traffic features, and packet-based characteristics. The authors reduced the features from the original 46 features to 37 meaningful features after preprocessing.

Mallikarjun and Patro [29] addressed imbalance issues in the CICIoT2023 dataset by introducing ZeroDefense, focusing on detecting emerging zero-day threats. The authors implemented SMOTE during supervised training that limits overfitting to minority data and preserves sensitivity to minority behavior, which, in practice, often resembles zero-day activity. The authors identified 46 statistical network flow features extracted from IoT traffic (rate, timing, and protocol-level statistics) for detection. Robust handling of extreme class imbalance without collapsing minority-class recall, achieved through a cautious separation of anomaly screening and SMOTE-balanced supervised learning.

Wakili and Bakkali [30] addressed class imbalance in the CICIoT2023 dataset by proposing a SVM Weights-based Synthetic sampling (SVWS) mechanism that generates boundary-aware synthetic samples rather than uniform oversampling. This approach concentrates on minority augmentation near decision margins, reducing overfitting risks observed with SMOTE-like methods. Combined with an MLP probability feature extractor and categorical boosting classifier, the framework stabilizes learning under skewed distributions. Integrated probabilistic feature transformation (MLPP) improves categorical boosting discrimination. The statistically significant improvements were validated via a paired t-test and the Wilcoxon signed-rank test. The authors used features categories such as Flow-based features (flow duration, rate, srate, drate), TCP flag statistics (syn, ack, fin, rst, urg counts), Protocol indicators (HTTP, HTTPS, DNS, TCP, UDP, etc.), and Statistical features (mean, variance, covariance, IAT, magnitude, total size, total sum).

2.3 Streaming and Incremental Learning for Network Security on CICIoT2023

Wang et al. [31] proposed CTWA, an incremental learning framework integrating a Convolutional Autoencoder (CAE) and a Temporal Convolutional Network (TCN) for IoT intrusion detection under streaming conditions. On CICIoT2023, initial classes (Benign, DDoS-ACK, DNS, DDoS-ICMP) are trained first, followed by incremental addition of new DDoS categories. A Gaussian-based task discriminator and Weight Alignment (WA) mitigate catastrophic forgetting. Feature fusion combines spatial (CAE) and temporal (TCN) representations. Label smoothing is integrated with cross-entropy to stabilize updates. Empirically, CTWA achieves 0.9643 accuracy after 100 epochs while maintaining the old task performance, though with non-trivial runtime overhead (789.58 s) The integration of CAE (spatial compression) with TCN (causal dilated temporal modeling) allows simultaneous structural and sequential feature extraction. Residual connections reduce gradient degradation in deeper CAE layers. The Gaussian task discriminator provides a probabilistic routing mechanism, which in practice reduces interference between old and new classes. WA partially corrects classifier bias across task heads. Empirically, forgetting is controlled

without explicit exemplar replay. The types of features used for detection are 46-dimensional flow-level statistical features.

Mahdi et al. [32] proposed an end-to-end intrusion detection framework integrating streaming incremental learning with secure data lifecycle management. An SGD-based incremental classifier is initially trained on CICIoT2023 and subsequently updated using traffic data extracted from a blockchain-secured storage layer. The model supports partial_fit retraining, avoiding full retraining cycles. Data collected from live traffic is encrypted (hybrid AES/RSA + signature), stored via Proof-of-Work blockchain, then periodically retrieved for incremental updates on a central server. Retraining occurs in multiple rounds, improving cross-validation accuracy from 97.9% to 99.8%. The approach addresses resource constraints in IoT devices while maintaining model adaptability to evolving attack distributions.

2.4 Adversarial Robustness in Federated Learning on CICIoT2023

Doménech et al. [33] evaluated adversarial robustness through cross-domain generalization between CICIoT2023 and CICIoMT2024. Models trained on CICIoT2023 show strong baseline performance, but substantial degradation when evaluated on IoMT traffic, even under equivalent attack categories. The proposed approach demonstrates systematic cross-domain generalization analysis between IoT and IoMT datasets. The experimental result shows that minor IoMT augmentation significantly restores detection capability. The authors achieved 99.85% accuracy after a controlled optimization pipeline. The results suggest CICIoT2023-trained models are not inherently adversarially robust to domain shifts in healthcare environments.

Saxena et al. [34] proposed GNN-IDS, a Graph Neural Network-based IDS that was experimented on the CICIoT2023 dataset. The proposed approach emphasizes on structural modeling of network traffic. Unlike traditional ML approaches that operate on independent flows, the proposed method constructs a graph using node and edge attributes, allowing relational reasoning across IoT entities. The proposed approach's advantage lies in explicit graph construction using both node and edge attributes. The GraphSAGE-based architecture leverages relational dependencies across IoT entities, mitigating the independence assumption common in flow-based classifiers. Binary accuracy (99.95%) exceeds baseline models (FedFK, RF, MLP-Mimic). The model demonstrates high precision across most attack categories, particularly DoS and DDoS subclasses.

To better understand how existing studies address intrusion detection in IoT environments, Table 1 provides a structured comparison of recent works evaluated on CICIoT2023 and related datasets. The table summarizes their feature selection strategies, learning paradigms, key strengths, and practical limitations—particularly in terms of federated learning, streaming capability, class imbalance handling, and robustness. This comparison helps highlight the gaps that motivate the proposed framework.

Table 1: Comparative analysis of recent IDS research on CICIoT2023.

Authors (Year)	Feature Selection	Model/Learning Paradigm	Key Advantages	Critical Limitations (W.r.t. FL, Streaming, Imbalance, Robustness)
Abid et al. [25]	GWO (Wrapper-based)	Hybrid LightGBM + CNN	Significant dimensionality reduction (43 → 18); low false alarm rate; strong detection for dominant attacks	No federated learning; offline evaluation only; degraded minority attack detection; no streaming capability; no adversarial robustness analysis

(Continued)

Table 1 (continued)

Authors (Year)	Feature Selection	Model/Learning Paradigm	Key Advantages	Critical Limitations (W.r.t. FL, Streaming, Imbalance, Robustness)
Torre et al. [26]	None	1D-CNN + Federated Averaging + DP + FHE	Privacy-preserving federated framework; asynchronous FL; evaluated on multiple large-scale datasets	No imbalance handling; no poisoning defense; no streaming incremental updates; computational overhead (10%)
Okey et al. [27]	Firefly Algorithm	1D-CNN + Knowledge Distillation (Teacher–Student)	91.24% model compression; resource-aware deployment; SMOTE + weighted loss; explainability integration	Centralized training only; no federated learning; no streaming setting; no adversarial robustness evaluation
Wahab et al. [28]	Variance-based pruning	DNN/CNN/ Transformer	High binary accuracy; SMOTE balancing; multi-class evaluation	Evaluated on 3M subset (not full dataset); no federated learning; performance drop in 12-class setting; no robustness analysis
Mallikarjun & Patro [29]	Implicit (CatBoost importance)	MLPP-CB + SVWS balancing	Effective severe imbalance handling; very high attack recall; stable cross-validation performance	Binary classification only; centralized setting; no streaming learning; no federated paradigm; no adversarial resilience
Wakili & Bakkali [30]	None (implicit feature learning)	Hybrid anomaly-first cascade (IF + AE + LOF + RF/XGB)	Zero-day detection capability; macro-F1 improvement; edge-deployable (<100 ms inference)	Centralized architecture; no federated learning; no streaming adaptation; no poisoning defense
Wang et al. [31]	Automatic (CAE-TCN feature extraction)	Incremental Deep Learning (CTWA)	Mitigates catastrophic forgetting; supports unknown attack detection; hybrid temporal-spatial modeling	No federated learning; high computational runtime; evaluated on relatively balanced subset; no adversarial study

(Continued)

Table 1 (continued)

Authors (Year)	Feature Selection	Model/Learning Paradigm	Key Advantages	Critical Limitations (W.r.t. FL, Streaming, Imbalance, Robustness)
Mahdi et al. [32]	ML-based feature selection (unspecified)	Incremental SGD + Blockchain	Continuous retraining; secure data storage; validation on unseen datasets	Central retraining only; no federated learning; no explicit imbalance strategy; no adversarial robustness evaluation
Doménech et al. [33]	None	Random Forest + SMOTE	Quantified dataset shift (F1 drop up to 66.87%); optimized preprocessing; domain-aware evaluation	No federated learning; no streaming; no adversarial robustness; limited cross-dataset scope
Saxena et al. [34]	None	GraphSAGE + MLP (GNN-based IDS)	Captures structural dependencies; improved topology-aware intrusion detection	Weak minority-class precision; no imbalance analysis; no federated learning; no streaming capability

In summary, existing studies generally optimize one aspect of IoT intrusion detection, such as privacy-preserving federation, imbalance handling, streaming adaptation, or adversarial defense. Limited attention has been given to their joint interaction under sustained attack-dominant and non-IID streaming conditions. This gap motivates the integrated framework proposed in this work.

3 Proposed Approach

This section presents the proposed federated intrusion detection framework for IoT environments operating under adversarial and imbalanced conditions.

In this study, resilience denotes the ability of the intrusion detection framework to preserve stable and reliable performance under non-IID streaming traffic, severe class imbalance, and malicious or degraded client updates.

The overall workflow consists of feature selection, similarity-guided initialization, and accuracy-gated federated ensemble learning, as illustrated in Fig. 2.

Similarity-Guided Data Balancing: Majority-class samples are analyzed for redundancy, and a structurally diverse subset is selected to construct a balanced initialization dataset.

Robust Global Ensemble Initialization: An incremental ensemble model is trained on the balanced dataset to obtain a stable, cold-start global model.

Streaming Federated Refinement with Accuracy Gating: Clients process local data in streaming batches, perform conditional incremental updates, and transmit only performance-preserving updates. The server aggregates accepted updates using performance-weighted aggregation under secure communication.

Fig. 2 summarizes the complete workflow of the proposed framework, integrating feature reduction, similarity-guided initialization, and accuracy-gated federated refinement under streaming IoT conditions.

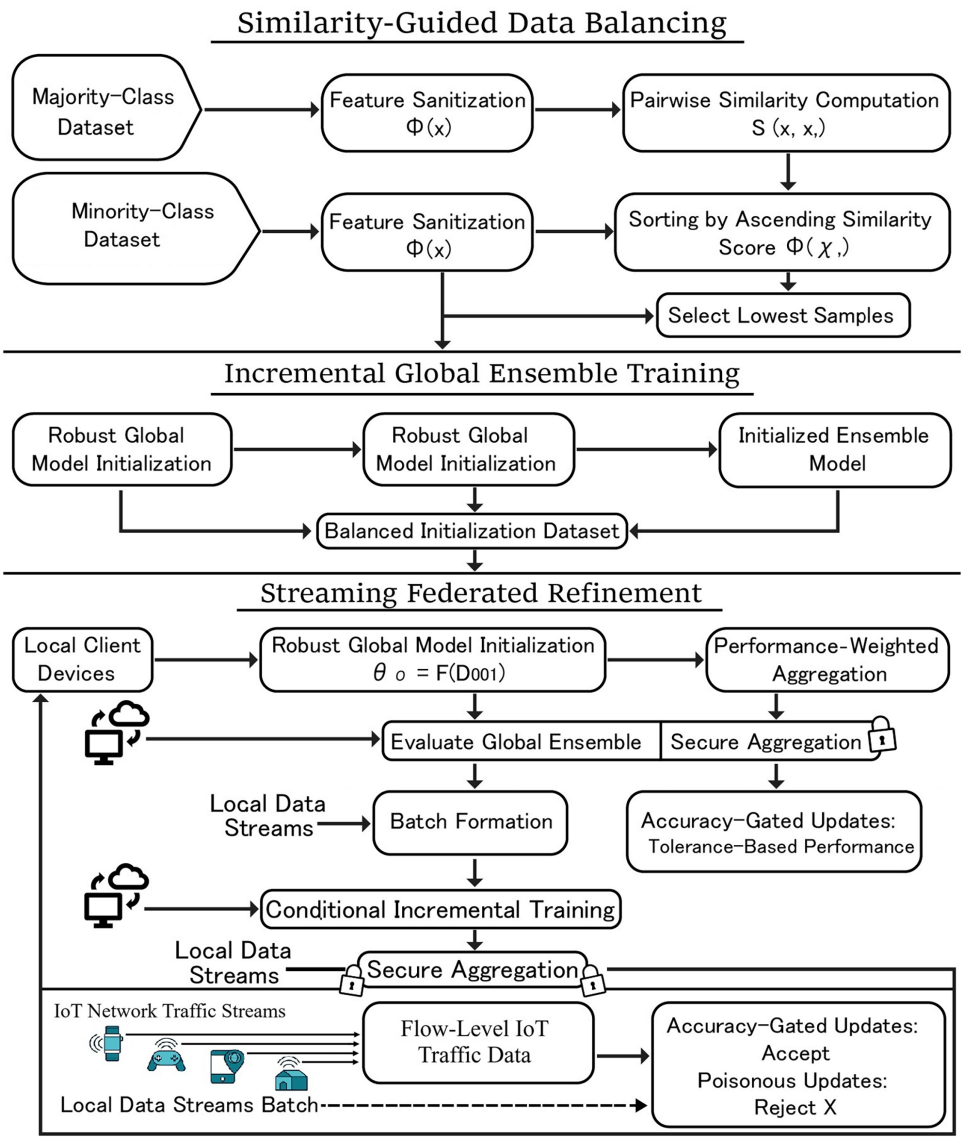

Figure 2: Overall process flow of the proposed resilient streaming federated IoT intrusion detection framework.

3.1 mRMR-Based Feature Selection and Importance Analysis

Feature selection plays a critical role in intrusion detection systems for IoT environments, where traffic characteristics are high-dimensional and exhibit strong statistical dependencies [35–37]. To identify features that are both informative with respect to class behavior and minimally redundant, this work employs a Maximum Relevance Minimum Redundancy (mRMR)-based feature selection framework on the CICIoT2023 dataset [38]. The original CICIoT2023 dataset contains 46 input features (excluding the label attribute). The objective is not to rely solely on a greedy ranking order, but to obtain a compact and discriminative subset of features while preserving the semantic relevance of the original traffic descriptors [39,40].

Let $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ denote the preprocessed dataset, where $\mathbf{x}_i \in \mathbb{R}^d$ represents a d-dimensional vector of normalized network-flow features and $y_i \in \{0,1\}$ corresponds to benign and malicious IoT traffic, respectively. Feature relevance is quantified using mutual information between an individual feature f_j and the class label Y,

$$\mathcal{I}(f_j; Y) = \sum_{f_j, y} p(f_j, y) \log \frac{p(f_j, y)}{p(f_j)p(y)}, \tag{1}$$

which captures the expected reduction in uncertainty about Y provided by f_j.

To account for the pronounced correlation structure typical of IoT traffic statistics, feature redundancy is modeled as the mean absolute Pearson correlation between f_j and all remaining features,

$$\mathcal{R}(f_j) = \frac{1}{d-1} \sum_{\substack{k=1 \\ k \neq j}}^{d} \left| \rho(f_j, f_k) \right|. \tag{2}$$

The raw mRMR feature importance score is then computed as

$$\text{mRMR}(f_j) = \mathcal{I}(f_j; Y) - \mathcal{R}(f_j). \tag{3}$$

Since redundancy may dominate relevance in highly correlated feature spaces, raw mRMR values can assume negative magnitudes. To enhance interpretability while preserving relative importance relationships, all scores are normalized using min–max scaling into the interval $[0, 1]$.

While performance-based weighting improves contribution from reliable updates, high local accuracy may not always correspond to greater global usefulness under heterogeneous non-IID client distributions. In particular, clients with comparatively easier local data may receive stronger influence than clients representing more challenging but important distributions. This trade-off is accepted here for computational simplicity and lightweight deployment.

The analysis is performed independently for benign and malicious traffic to capture class-specific behavioral characteristics. Fig. 3 illustrates the normalized mRMR feature importance for class 0 (benign IoT traffic). Several temporal and protocol-related features exhibit consistently high importance, indicating their role in characterizing stable and regular communication patterns typical of legitimate IoT behavior. In contrast, statistical dispersion features, such as minimum, maximum, and variance-based descriptors, demonstrate comparatively lower importance, suggesting limited discriminative value for benign traffic modeling.

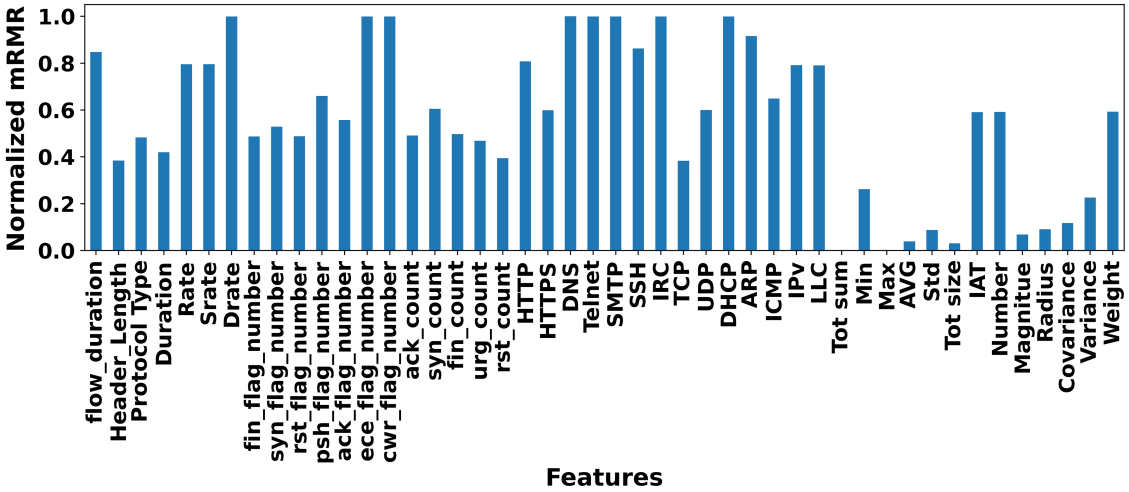

Figure 3: Normalized mRMR feature importance for class 0 (benign IoT traffic) in the CICIoT2023 dataset.

Fig. 4 presents the corresponding feature importance profile for class 1 (malicious IoT traffic). In this case, a broader set of features achieves elevated importance values, reflecting the heterogeneous and irregular nature of attack traffic. Notably, features associated with protocol usage, control flags, and traffic intensity display stronger relevance, indicating their effectiveness in capturing deviations from normal communication behavior. This divergence in importance patterns highlights the asymmetric information content between benign and malicious traffic classes.

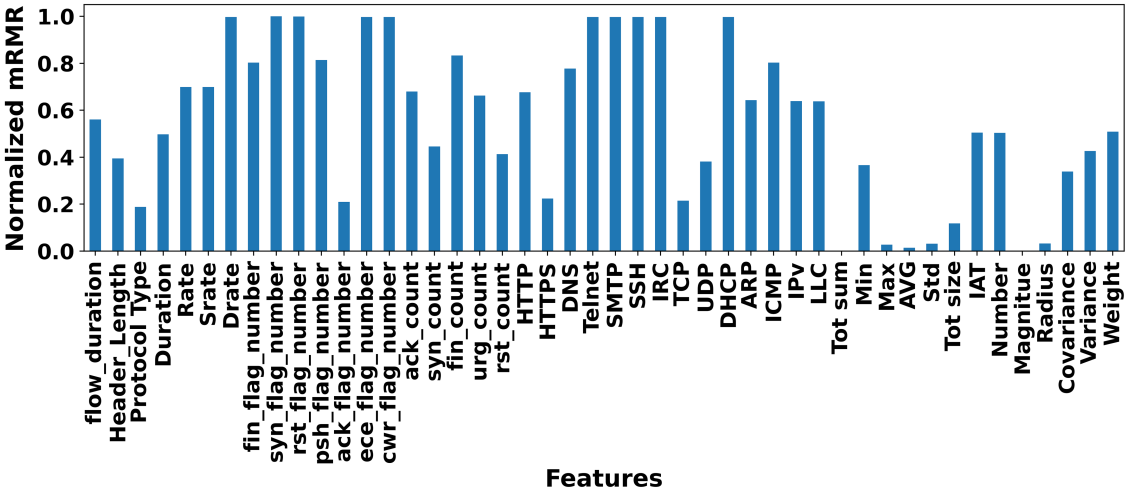

Figure 4: Normalized mRMR feature importance for class 1 (malicious IoT traffic) in the CICIoT2023 dataset.

To identify features that remain consistently informative across both traffic classes, a combined feature importance profile is constructed by averaging the raw mRMR scores from class 0 and class 1 prior to normalization. The resulting distribution, shown in Fig. 5, emphasizes features that contribute robustly to discrimination regardless of class context. These features are particularly well-suited for intrusion detection models intended for deployment in dynamic IoT environments, where attack patterns and benign behaviors may evolve over time.

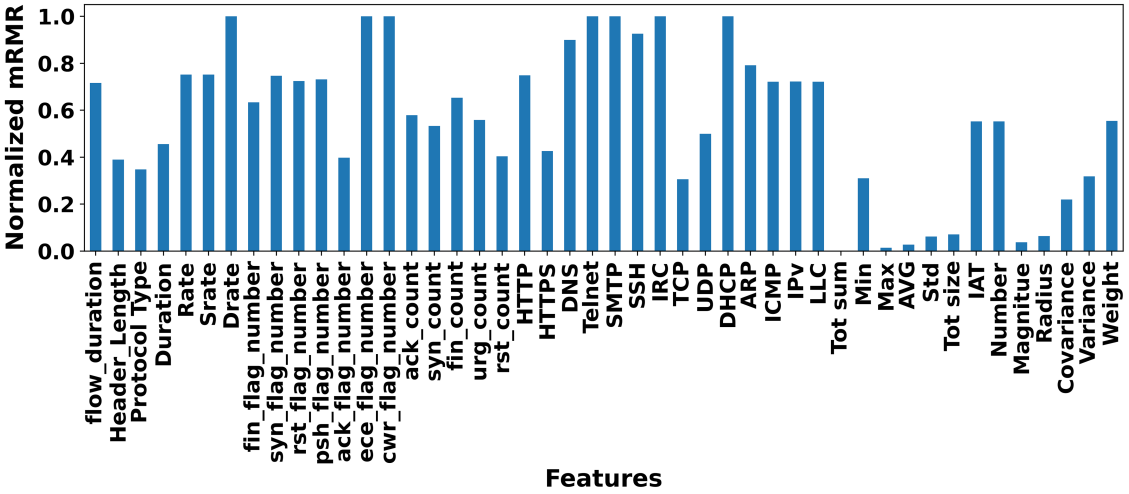

Figure 5: Combined normalized mRMR feature importance across benign and malicious traffic classes in the CICIoT2023 dataset.

For subset construction, the combined normalized mRMR scores were sorted in descending order, and a fixed retention ratio of 80% was applied. Based on this criterion, the top 80% highest-scoring attributes were retained for subsequent experiments, while lower-ranked features were discarded. This process reduced the original 46-dimensional feature space to 31 selected features, improving computational efficiency while preserving discriminative information relevant to IoT intrusion detection.

Fig. 6 shows the final selected feature subset used for federated model training. The retained attributes include protocol indicators (e.g., ARP, DHCP, DNS, SSH, Telnet, SMTP), traffic-rate descriptors (e.g., Rate, Srate), temporal flow properties (e.g., Duration, flow_duration), and TCP control-flag statistics (e.g., ack, syn, rst, fin, cwr, ece). This distribution indicates that both protocol semantics and traffic behavior are important for reliable detection in dynamic IoT environments.

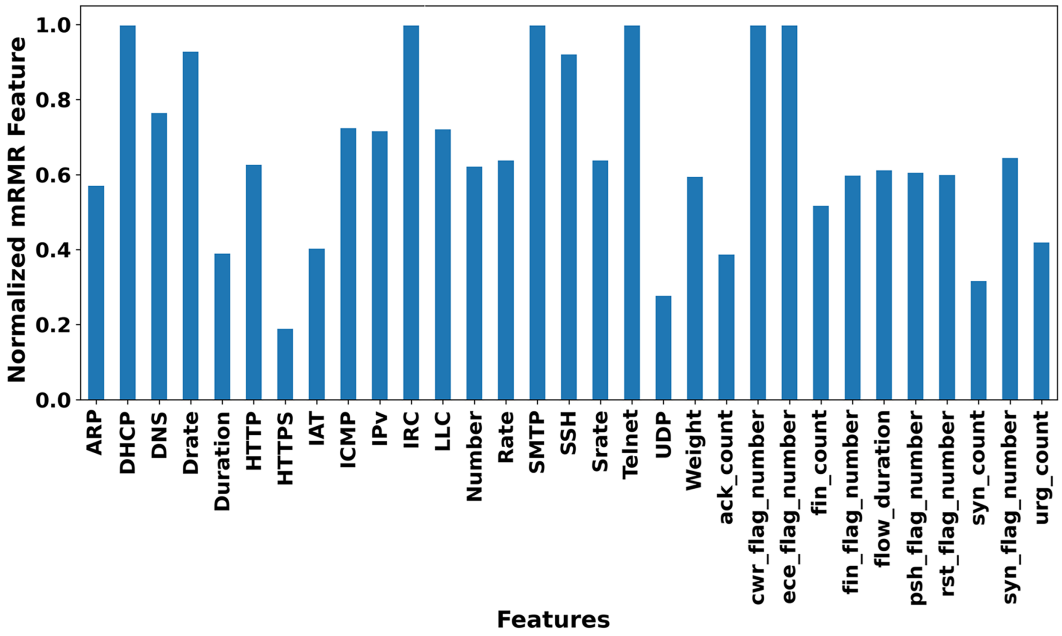

Figure 6: Final selected feature subset with normalized mRMR scores used in the proposed federated intrusion detection framework.

The mRMR framework was therefore used not only for feature importance interpretation, but also to guide feature reduction and construct a consistent input space shared across all clients and all experiments.

3.2 Similarity-Guided Data Balancing for Robust Global Initialization

A fundamental challenge in federated and streaming intrusion detection systems lies in the simultaneous presence of severe class imbalance and cold-start instability during global model initialization [41,42]. In large-scale network telemetry datasets, malicious traffic (class 1) often dominates benign instances (class 0) by several orders of magnitude, biasing both centralized and federated learners toward majority-class decision boundaries. When such an imbalance is propagated into federated environments, the initial global model becomes poorly conditioned, leading to unstable early updates, delayed convergence, and degraded detection accuracy [43]. To address this issue, we propose a similarity-guided balancing strategy that performs principled majority-class reduction while preserving behavioral diversity, thereby enabling robust global initialization and mitigating cold-start effects.

To formalize the imbalance setting considered in this study, the dominant and minority traffic distributions are defined as follows.

Let $\mathcal{D}_1 = \{\mathbf{x}_i^{(1)}\}_{i=1}^{N_1}$ denote the set of dominant-class samples (class 1, malicious traffic), where each $\mathbf{x}_i^{(1)} \in \mathbb{R}^d$ represents a d-dimensional feature vector derived from network traffic statistics. Similarly, let $\mathcal{D}_0 = \{\mathbf{x}_j^{(0)}\}_{j=1}^{N_0}$ denote the underrepresented-class dataset (class 0, benign traffic), with $N_1 \gg N_0$.

In ongoing attack scenarios, streaming traffic becomes attack-heavy, causing benign communication patterns to appear sparsely [44]. Under such an inverted imbalance, gradient-based optimization is dominated by malicious gradients, suppressing benign structure and biasing early decision boundaries. To prevent benign-class underrepresentation during initialization, the objective is to construct a reduced subset $\tilde{\mathcal{D}}_1 \subset \mathcal{D}_1$ such that $|\tilde{\mathcal{D}}_1| = N_0$, yielding a balanced dataset

$$\mathcal{D}_{bal} = \tilde{\mathcal{D}}_1 \cup \mathcal{D}_0,$$

while preserving structurally diverse malicious behaviors.

To achieve this, we quantify the intrinsic redundancy of dominant-class (malicious) samples through a similarity-based functional. Let $\phi(\mathbf{x}) \in \mathbb{R}^d$ denote the normalized representation of a sample after numeric feature sanitization. For any two dominant-class samples $\mathbf{x}_i^{(1)}$ and $\mathbf{x}_j^{(1)}$, we define a similarity kernel

$$\mathcal{S}(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(1)}) = 1 - \frac{\|\phi(\mathbf{x}_i^{(1)}) - \phi(\mathbf{x}_j^{(1)})\|_2}{\max\limits_{k,l} \|\phi(\mathbf{x}_k^{(1)}) - \phi(\mathbf{x}_l^{(1)})\|_2}, \tag{4}$$

which maps pairwise Euclidean distances into a bounded similarity score in $[0,1]$. Higher values of $\mathcal{S}$ indicate greater redundancy within the attack manifold, while lower values correspond to structurally distinct malicious behaviors.

For each dominant-class instance $\mathbf{x}_i^{(1)}$, we compute an aggregate similarity score with respect to the entire dominant-class distribution:

$$\psi(\mathbf{x}_i^{(1)}) = \frac{1}{N_1 - 1} \sum_{\substack{j=1 \\ j \neq i}}^{N_1} \mathcal{S}(\mathbf{x}_i^{(1)}, \mathbf{x}_j^{(1)}). \tag{5}$$

The scalar $\psi(\mathbf{x}_i^{(1)})$ serves as a redundancy measure: samples with high ψ lie in dense regions of the malicious manifold, whereas samples with low ψ capture diverse attack behaviors that are more informative for discriminative learning.

The pairwise similarity computation has quadratic complexity with respect to the dominant-class sample count, i.e., $\mathcal{O}(N_1^2)$. However, this operation is performed only once during the server-side initialization stage and is not repeated during subsequent federated communication rounds. Therefore, it does not contribute to the recurring online training overhead.

In practical deployments, the computation can be accelerated through vectorized matrix operations, multi-core or GPU-based parallelization, and block-wise distance evaluation. For large-scale datasets, approximate strategies such as random subsampling, clustering-based prototype generation, nearest-neighbor sparsification, or locality-sensitive hashing can be adopted to reduce computational cost while preserving representative structural diversity within the dominant class.

The dominant-class dataset $\mathcal{D}_1$ is then sorted in ascending order of $\psi(\cdot)$, and only the lowest N_0 samples are retained:

$$\tilde{\mathcal{D}}_1 = \arg \min_{A \subset \mathcal{D}_1, |A| = N_0} \sum_{\mathbf{x} \in A} \psi(\mathbf{x}). \tag{6}$$

This selection criterion ensures that malicious samples exhibiting maximal structural diversity are preserved, while highly redundant attack instances are systematically discarded. In the conducted experiments, this procedure reduced the original dominant-class dataset to match the cardinality of the benign class, yielding a balanced initialization dataset. Since only the reduced initialization subset is required, exact evaluation over the full dominant-class set is not mandatory, and approximate candidate selection methods remain suitable in large deployments.

Beyond class balance, the proposed strategy plays a critical role in stabilizing global model initialization. Let θ_0 denote the parameters of the initial global model trained on $\mathcal{D}_{bal}$. Since $\mathcal{D}_{bal}$ is both balanced and diversity-preserving, θ_0 provides a well-conditioned approximation of the joint class distribution:

$$\theta_0 \approx \arg \min_{\theta} \mathbb{E}_{(\mathbf{x},y)\sim \mathcal{D}_{bal}} \left[\mathcal{L}(f_\theta(\mathbf{x}), y) \right], \tag{7}$$

where $\mathcal{L}(\cdot)$ denotes the model-specific classification loss (log-likelihood for Bernoulli Naïve Bayes, hinge loss for Passive–Aggressive, and logistic loss for SGD).

This initialization mitigates the cold-start instability commonly observed in federated streaming learning, wherein early updates would otherwise be dominated by attack-heavy gradients that suppress benign discrimination.

The proposed initialization stage reduces attack-dominant bias by constructing a balanced and diversity-preserving training subset prior to federated optimization.

Empirically, the similarity-guided balancing approach produces a reduced yet information-rich training set that improves convergence speed, stabilizes early federated updates, and preserves benign-class sensitivity under attack-dominant streaming conditions. By embedding balancing directly into global initialization through a mathematically grounded similarity functional, the proposed method establishes a principled link between dataset conditioning and federated robustness in adversarial IoT environments.

3.3 Proposed Streaming Federated Learning Framework with Robust Global Initialization

This section presents the proposed federated intrusion detection framework for handling attack-dominant imbalance, cold-start instability, federated privacy constraints, and unreliable client updates in streaming IoT environments.

The proposed streaming federated framework is designed for IoT environments characterized by non-IID traffic distributions, attack-dominant imbalance, and unreliable client updates during continuous operation.

Problem Setting and Motivation

Let $\mathcal{D}_k = \{(\mathbf{x}_i^{(k)}, y_i^{(k)})\}_{i=1}^{N_k}$ denote the local dataset of client k, where $\mathbf{x}_i^{(k)} \in \mathbb{R}^d$ represents numeric IoT traffic features and $y_i^{(k)} \in \{0,1\}$ denotes benign (0) or attack (1) traffic. In sustained attack conditions, the local class distribution typically satisfies $N_k^{(1)} \gg N_k^{(0)}$, meaning malicious instances dominate streaming observations while benign traffic becomes underrepresented.

Under an attack-dominant imbalance, early optimization becomes biased toward malicious gradients, weakening benign-class discrimination. Since federated servers do not access raw data, this bias may propagate across communication rounds if initialization is poorly conditioned.

Classical federated learning approaches assume reasonably balanced distributions and reliable client updates [45]. However, in attack-heavy streaming IoT environments, poorly conditioned initial models amplify benign-class suppression, while corrupted or mislabeled local data—whether accidental or adversarial—can further destabilize aggregation. These effects motivate a federated design that is *cold-start aware*, *incremental by construction*, and *selective in accepting client updates*.

Similarity-Guided Global Initialization

To mitigate cold-start instability, the global model is initialized using a balanced and diversity-preserving dataset constructed via the previously introduced *Similarity-Guided Data Balancing* mechanism.

Let $\mathcal{D}_1$ and $\mathcal{D}_0$ denote the majority (malicious) and minority (benign) class datasets, respectively. A similarity functional $\psi(\mathbf{x})$ is computed for each $\mathbf{x} \in \mathcal{D}_0$, capturing redundancy within the benign feature manifold. The balanced initialization set is defined as

$$\mathcal{D}_{\text{init}} = \tilde{\mathcal{D}}_0 \cup \mathcal{D}_1, \quad \text{where } |\tilde{\mathcal{D}}_0| = |\mathcal{D}_1|.$$

This construction preserves structural diversity while eliminating redundant benign observations, yielding a well-conditioned dataset for initializing the global model.

Incremental Learning as a First-Class Design Choice

To support streaming IoT traffic under federated constraints, the framework adopts incremental learning through partial_fit-based updates, enabling mini-batch adaptation without full retraining.

Based on extensive empirical evaluation, three incremental classifiers were selected due to their complementary inductive biases and superior performance on the CICIoT2023 datasets:

- **Bernoulli Naïve Bayes (BNB):** robust to sparse, binary, and thresholded traffic indicators.
- **Passive–Aggressive Classifier (PA):** margin-based learner capable of rapid adaptation to emerging attack patterns.
- **Stochastic Gradient Descent Classifier (SGD):** scalable linear learner with probabilistic loss modeling.

Streaming Federated Learning Protocol

The federated learning process unfolds in two operational phases.

Phase I: Server-Side Global Initialization

The federated server initializes an ensemble $\mathcal{E}_0$ by incrementally training each base learner on $\mathcal{D}_{\text{init}}$:

$$\theta_m^{(0)} \leftarrow \text{partial\_fit}(\theta_m^{(0)}, \mathcal{D}_{\text{init}}), \quad m \in \{1, 2, 3\}.$$

Model quality is evaluated on a disjoint evaluation dataset using stratified sampling, allocating 20% for validation and 20% for testing while preserving the original class distribution. The initialization dataset D_{init} is never reused for validation or testing.

Phase II: Accuracy-Gated Streaming Federated Refinement

When a client k connects, it retrieves the current global ensemble $\mathcal{E}_t$ over a secure HTTPS channel. Local data are processed in streaming batches of size $B = 1000$:

$$\mathcal{D}_k^{(b)} = \{(\mathbf{x}_i^{(k)}, y_i^{(k)})\}_{i=1}^{B}.$$

Before performing any local update, the client evaluates the current global ensemble on $\mathcal{D}_k^{(b)}$ and obtains a baseline accuracy $\mathrm{Acc}_{\mathrm{global}}^{(b)}$. The client then tentatively updates its local copy using incremental learning:

$$\theta_{m,k}^{(t+1)} \leftarrow \mathrm{partial\_fit}(\theta_m^{(t)}, \mathcal{D}_k^{(b)}).$$

The updated local ensemble is re-evaluated on the same batch, yielding $\mathrm{Acc}_{\mathrm{local}}^{(b)}$. The update is accepted *only if*

$$\mathrm{Acc}_{\mathrm{local}}^{(b)} \geq \mathrm{Acc}_{\mathrm{global}}^{(b)} - \delta,$$

where δ is a tolerance margin (set to 5% in experiments). If this condition is violated, the batch update is discarded, and a diagnostic message is generated.

This accuracy-gated mechanism prevents harmful updates arising from label-flipped, noisy, or adversarial data, while still permitting benign distributional variation.

Algorithm 1 summarizes the client-side streaming federated procedure, including batch evaluation, temporary incremental learning, and the accuracy-gated update acceptance mechanism under non-IID and potentially corrupted local data.

The algorithm explicitly separates tentative local updates from committed parameter replacement, thereby preventing propagation of performance-degrading gradients to the federation.

Performance-Weighted Ensemble Aggregation

Accepted client updates are transmitted to the server as model parameters accompanied by performance statistics. Model parameters are aggregated using symmetric parameter averaging, while ensemble voting weights are updated in a performance-proportional manner. This hybrid design preserves the stability of parameter fusion while allowing reliable client contributions to influence ensemble decision strength. Let $f_m(\mathbf{x})$ denote the prediction of base learner m. The ensemble decision is defined as

$$\hat{y}(\mathbf{x}) = \mathbb{I}\left(\sum_{m=1}^{M} w_m f_m(\mathbf{x}) \geq \frac{1}{2}\right),$$

where the weights are updated as

$$w_m^{(t+1)} = w_m^{(t)} + \frac{Acc_k^{(b)}}{2},$$

followed by normalization:

$$w_m^{(t+1)} \leftarrow \frac{w_m^{(t+1)}}{\sum_{j=1}^{M} w_j^{(t+1)}}.$$

Weight normalization prevents unbounded growth across federated rounds and preserves the numerical stability of the ensemble decision threshold.

In the experimental evaluation, the default client-side tolerance parameter was set to 0.95 relative to the current server accuracy, corresponding to $\delta = 0.05$, while the server-side acceptance threshold was set to 0.90. Additional sensitivity analysis was conducted using alternative client-side thresholds to examine the trade-off between robustness and update acceptance. This asymmetric design enforces stricter local filtering prior to a comparatively tolerant global aggregation stage.

Algorithm 1: Accuracy-gated client-side streaming federated update

Require: Global ensemble $\mathcal{E}_g = \{(M_m, w_m)\}_{m=1}^{M}$, server accuracy Acc_g, local dataset D_k, batch size B, tolerance δ

Ensure: Accepted local model updates transmitted to server

Receive $(\mathcal{E}_g, Acc_g)$ from server
Partition D_k into batches $\{D_k^{(b)}\}_{b=1}^{T}$ of size B
for each batch $D_k^{(b)} = (X_b, y_b)$ **do**
    // Step 1: Evaluate current global ensemble
    Compute weighted ensemble predictions:

$$\hat{y}_b^{global} = \mathbb{I}\left(\sum_{m=1}^{M} w_m f_m(X_b) \geq \frac{1}{2}\right)$$

    Compute Acc_b^{global} on (X_b, y_b)
    // Step 2: Tentative local incremental update
    for each base learner $M_m \in \mathcal{E}_g$ **do**
        $\theta_m^{temp} \leftarrow$ partial_fit(M_m, X_b, y_b)
    end for
    Compute updated ensemble predictions:

$$\hat{y}_b^{local} = \mathbb{I}\left(\sum_{m=1}^{M} w_m f_m^{temp}(X_b) \geq \frac{1}{2}\right)$$

    Compute Acc_b^{local}
    // Step 3: Accuracy-Gated Acceptance Rule
    if $Acc_b^{local} \geq Acc_b^{global} - \delta$ **then**
        Accept update
        Replace $M_m \leftarrow \theta_m^{temp}$ for all m
        Transmit updated parameters and metrics to server
        Receive updated global accuracy Acc_g
    else
        Reject batch update
        Discard temporary parameters
    end if
end for

To operationalize the performance-weighted aggregation mechanism described above, the complete server-side procedure is summarized in Algorithm 2. The server performs robust global initialization using the similarity-balanced dataset, followed by selective aggregation of client updates based on accuracy thresholds. Only updates that satisfy the global acceptance criterion contribute to parameter refinement and ensemble weight adjustment, thereby preventing propagation of unreliable gradients into the global model.

Algorithm 2: Server-side performance-weighted federated aggregation with accuracy-based rejection

Require: Balanced initialization dataset D_{init}, evaluation dataset D_{eval}, tolerance γ

Ensure: Updated global ensemble $\mathcal{E}_g$

Phase I: Robust Global Initialization
Train base learners $\{M_m\}_{m=1}^{M}$ on D_{init} using incremental `partial_fit`
Evaluate ensemble on validation subset of D_{eval}

(Continued)

Algorithm 2 (continued)

Compute initial global accuracy Acc_g
Initialize performance weights $\{w_m\}_{m=1}^{M}$
Distribute $(\mathcal{E}_g, Acc_g)$ to clients
Phase II: Accuracy-Based Federated Aggregation
Upon receiving client update $(\mathcal{E}_c, Acc_c)$
if $Acc_c < \gamma \cdot Acc_g$ **then**
    Reject update
    Return current Acc_g
else
    for each base learner M_m **do**
        Update model parameters: $\theta_m \leftarrow \frac{\theta_m + \theta_m^{(c)}}{2}$
        Update ensemble weight: $w_m \leftarrow w_m + \frac{Acc_c}{2}$
    end for
    $Acc_g \leftarrow \max(Acc_g, Acc_c)$
    Store updated global ensemble
    Return updated Acc_g
end if

4 Results and Discussion

In this section, we present and analyze the experimental findings of the proposed resilient federated intrusion detection framework. The objective is to examine how the system behaves under streaming, non-IID, and adversarial conditions that reflect realistic IoT deployments. We evaluate client-wise stability, batch-level performance, and robustness against label-flipping attacks. In addition, ablation experiments are conducted to understand the contribution of structured global initialization. Rather than only reporting high-level accuracy, we focus on convergence behavior, minority-class sensitivity, and resistance to unstable updates. The discussion aims to connect observed empirical trends with the design choices introduced in the proposed methodology.

4.1 Experimental Setup and Reproducibility

All experiments were implemented in Python using standard scientific and machine learning libraries, including NumPy, Pandas, Scikit-learn, and Flask for client-server communication. The evaluation was conducted on a workstation equipped with a multi-core CPU and sufficient main memory for streaming batch execution.

The CICIoT2023 dataset was preprocessed using label encoding, missing-value handling, and min–max normalization. The processed data were partitioned into three client datasets to simulate distributed non-IID federated learning conditions.

Each client received streaming batches sequentially. For every batch, local evaluation was first performed using the current global model. Updates satisfying the acceptance criterion were used for incremental local training and then transmitted to the server for weighted aggregation. Final reported metrics were computed by averaging client-wise results over the full streaming process.

To reduce numerical instability and experimental bias, normalized features were used throughout training, identical preprocessing was applied across all clients, deterministic data partitions were maintained during evaluation, and multiple complementary metrics were reported, including Accuracy, Precision, Sensitivity, F1-score, MCC, Youden's J, and FMI.

4.2 Computational Complexity and Runtime Considerations

The proposed framework introduces an additional preprocessing stage during global initialization through similarity-guided balancing. Let N_1 denote the number of dominant-class samples. The pairwise similarity computation is defined in Eqs. (4) and (5) require $\mathcal{O}(N_1^2)$ operations in the worst case due to all-pairs distance evaluation within the dominant class. However, this computation is executed only once during server-side initialization and is not repeated during subsequent federated communication rounds.

Compared with conventional federated learning approaches that directly initialize from randomly sampled or locally available data, the proposed method therefore incurs additional initialization cost but does not increase recurring online communication overhead. During federated operation, clients continue to perform lightweight incremental partial_fit updates on streaming mini-batches and transmit only model parameters, preserving communication behavior comparable to standard incremental FL settings.

The additional initialization overhead is compensated by improved early-round convergence stability and reduced propagation of unstable updates under attack-dominant traffic distributions. Furthermore, for large-scale IoT deployments, the preprocessing stage can be accelerated using approximate nearest-neighbor search, clustering-based prototype selection, locality-sensitive hashing, or parallel block-wise similarity computation.

4.3 Client-Wise Federated Performance Analysis

This subsection analyzes the client-wise behavior of the proposed streaming federated intrusion detection framework in relation to its underlying design principles, namely similarity-guided global initialization, incremental ensemble learning, and accuracy-gated update acceptance. Rather than focusing solely on aggregate performance, the analysis examines batch-wise dynamics across individual clients to assess stability, robustness, and consistency under non-IID streaming conditions.

Fig. 7 presents the federated accuracy progression for three clients over successive streaming batches. Across all clients, accuracy remains tightly bound within a narrow high-performance band, predominantly between 0.97 and 0.99. Importantly, no client exhibits sustained accuracy degradation as training progresses. This behavior is non-trivial in streaming federated environments, where early instability and concept drift often amplify local errors. The observed stability can be attributed to the similarity-guided global initialization strategy, which produces a well-conditioned starting model by preserving structurally diverse benign samples while eliminating majority-class redundancy. As a result, early federated updates are not dominated by biased decision boundaries, allowing subsequent incremental updates to refine rather than correct the global model.

Short-lived accuracy fluctuations are visible at isolated batches for individual clients. These variations are expected in streaming IoT settings due to localized traffic bursts, transient attack patterns, and client-specific distribution shifts. Crucially, such fluctuations do not accumulate across batches, indicating that unreliable local updates are effectively prevented from influencing the global ensemble. This behavior directly reflects the impact of the accuracy-gated update acceptance mechanism, which selectively filters batch updates that would otherwise degrade global performance. Consequently, the federation exhibits resilience to noisy or inconsistent local batches without suppressing benign adaptation.

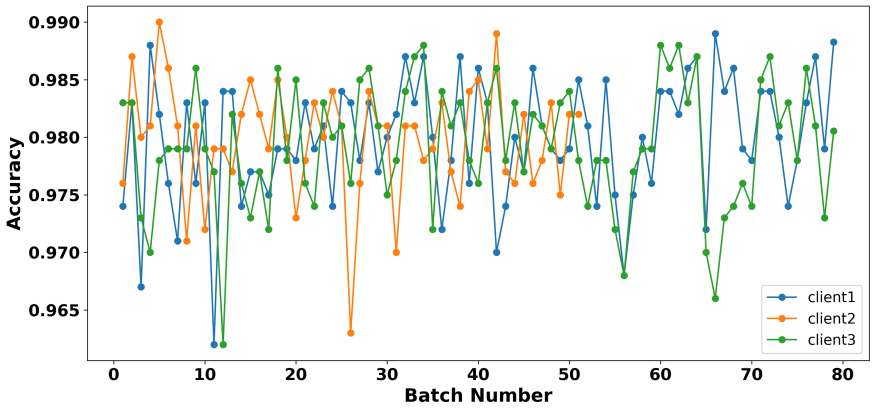

Figure 7: Federated accuracy progression per client across streaming batches.

Fig. 8 complements the accuracy analysis by reporting the Fowlkes–Mallows Index (FMI) progression. Unlike accuracy, FMI jointly captures precision–recall agreement for the positive (attack) class and is therefore more sensitive to class imbalance effects common in IoT intrusion detection. Across all clients, FMI remains consistently high and closely tracks the accuracy curves, largely exceeding 0.97 throughout the streaming process. The absence of divergence between accuracy and FMI trends indicates that high overall accuracy is not achieved by biasing predictions toward the dominant class. Instead, the model preserves reliable attack-class discrimination across clients and batches.

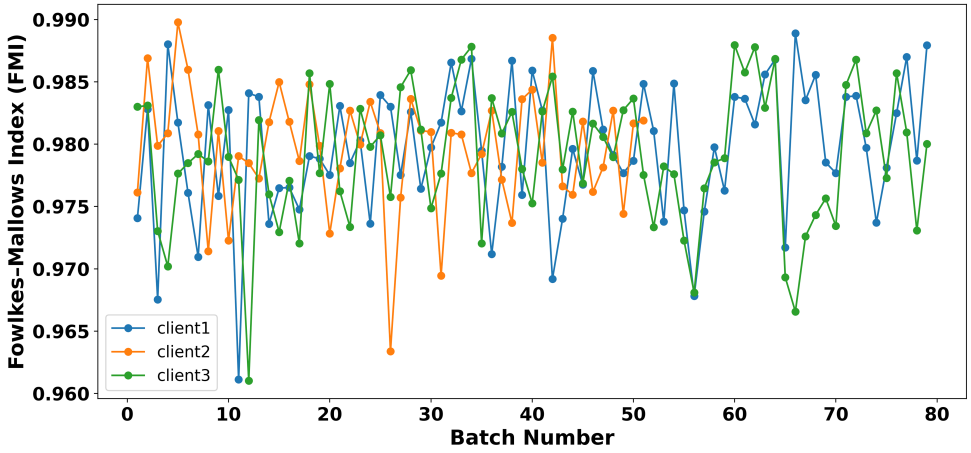

Figure 8: Federated Fowlkes–Mallows Index (FMI) progression per client across streaming batches.

The consistency between accuracy and FMI trends indicates that the ensemble maintains stable minority-class discrimination throughout streaming adaptation.

Table 2 summarizes the aggregate federated performance across all participating clients. The achieved accuracy of 0.99 and precision of 1.00 indicate a strong overall detection capability under distributed operation without centralized access to raw data. This aggregate evaluation complements the client-wise analysis by providing a global view of the final federated model. However, in imbalanced intrusion detection settings, accuracy alone can overstate effectiveness because the dominant class contributes more heavily to the final score.

Table 2: Average federated performance metrics aggregated across all clients.

Evaluation Stage	Accuracy	Precision	Sensitivity	F1-Score	MCC	Markedness	YoudenJ	FMI
Clients Average	0.99	1.00	0.8636	0.7925	0.7870	0.9881	0.8278	0.9898

Although the overall accuracy is high, performance should be interpreted together with minority-sensitive metrics such as Sensitivity, F1-score, MCC, and FMI. These measures are particularly important in imbalanced settings, where accuracy alone may mask bias toward the dominant attack class.

The sensitivity of 0.8636 and F1-score of 0.7925 indicate that minority-class detection remains more challenging under attack-dominant and non-IID streaming conditions, where class skew and heterogeneous client distributions can affect recall and precision for underrepresented patterns.

At the same time, the MCC, Markedness, Youden's J, and FMI values remain strong, indicating that the model preserves meaningful discriminative performance and does not rely solely on majority-class predictions. Together, the client-level trends and aggregate metrics indicate that the proposed framework achieves both stable local adaptation and strong global federated performance through similarity-guided initialization, incremental ensemble learning, and accuracy-aware federation.

4.4 Robustness against Label-Flipping Poisoning Attacks

This section evaluates the behavior of the proposed streaming federated learning framework under complete label-flipping poisoning, where all client-side labels are inverted prior to local processing. The analysis is restricted to client-level Accuracy, as shown in Fig. 9, and focuses on update rejection behavior induced by the accuracy-gated mechanism.

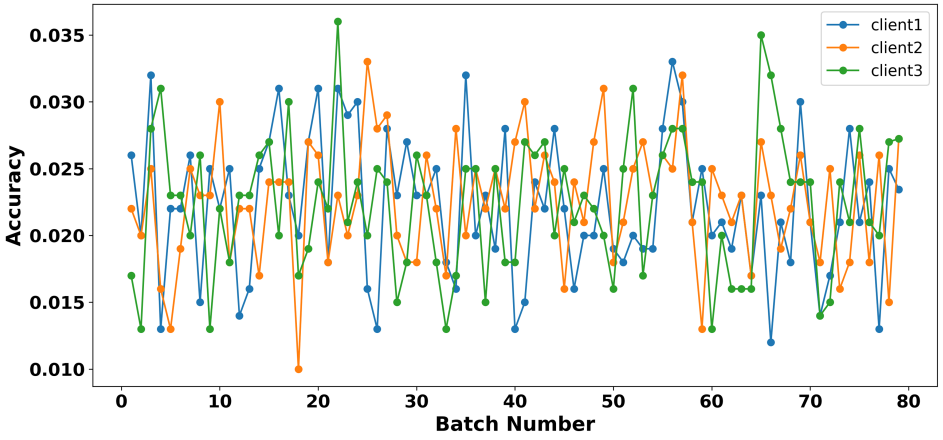

Figure 9: Client-level batch-wise accuracy under complete label-flipping poisoning. Across all clients and batches, accuracy remains persistently low, fluctuating approximately between 0.01 and 0.036, with no sustained upward trend. All batches fall below the acceptance criterion, resulting in systematic rejection of local updates.

Fig. 9 shows the batch-wise accuracy trajectories for three clients under complete label inversion. For all clients, accuracy remains consistently low throughout the streaming process, fluctuating within a narrow range close to random-guess performance (approximately 0.01–0.036). No client exhibits sustained improvement or stabilization across batches, and no batch reaches accuracy levels associated with benign operation.

As a direct consequence of these low accuracy values, all local batches fail to satisfy the accuracy-gated update acceptance criterion and are rejected prior to server aggregation. This behavior demonstrates

deterministic suppression of learning under maximal label corruption. Since no client updates are accepted during this experiment, the results do not permit analysis of convergence, recovery, or adaptation dynamics under adversarial influence.

These findings are limited to an extreme boundary-condition scenario in which all client labels are flipped. Accordingly, the results should be interpreted strictly as empirical evidence of fail-safe update rejection under complete label-flipping poisoning, rather than as robustness to partial, probabilistic, or adaptive poisoning strategies.

4.5 Adversarial Batch Rejection under FGSM Perturbation

To extend the robustness evaluation beyond complete label-flipping attacks, an additional experiment was conducted using adversarially perturbed client datasets generated with the Fast Gradient Sign Method (FGSM). Three independent adversarial datasets were created in separate runs using randomized seeds and assigned to three federated clients for streaming evaluation under the standard training and communication protocol.

Adversarial samples were generated using a feed-forward neural network surrogate model with two hidden layers of 128 neurons each and ReLU activation. The surrogate model was trained for five epochs on the initialization dataset. FGSM perturbations were then applied in the normalized feature space using randomly selected perturbation magnitudes $\epsilon \in [0.05, 0.25]$ for each run. After perturbation, the generated samples were mapped back to the original feature scale while preserving their original class labels. Each adversarial dataset contained 200,000 samples.

Each client processed 200 streaming batches, resulting in a total of 600 batch-update attempts across the federation. For every batch, the client first evaluated the current global ensemble and then applied the proposed accuracy-gating acceptance rule before any local update could be transmitted to the server.

The outcome was consistent across all three clients. In every case, the batch-level accuracy of the adversarial data remained below the required acceptance threshold, and all attempted updates were rejected locally. Consequently, no adversarial update was forwarded for server-side aggregation.

As shown in Table 3, all FGSM-based adversarial batch updates were rejected across the three federated clients. These results indicate that strongly degraded adversarial inputs can be effectively filtered before entering the federated aggregation stage. Although FGSM perturbation does not cover all poisoning scenarios or adaptive adversarial strategies, the experiment provides additional evidence that the proposed gating mechanism can block low-quality or manipulated client updates under adversarial conditions.

Table 3: FGSM-based adversarial batch rejection results across three federated clients.

Client	Total Batches	Accepted Updates	Rejected Updates
Client 1	200	0	200
Client 2	200	0	200
Client 3	200	0	200
Total	600	0	600

4.6 Ablation Study: Effect of Removing Global Initialization

To isolate the contribution of structured global initialization, we disable the similarity-guided balancing phase and skip the initial incremental training over D_{init}. The federated process, therefore, begins with

randomly initialized BNB, PA, and SGD classifiers. The ensemble aggregation, client-side 0.95 accuracy gating, and server-side 0.90 acceptance threshold remain unchanged. No other hyperparameters are modified.

Cold-start behavior under this ablated configuration is illustrated in Fig. 10. The degradation is immediate. Sensitivity drops to 0.6582, while overall accuracy settles at 0.8159. At first glance, this accuracy may appear moderate. However, precision increases to 0.9616, indicating that the model is strongly biased toward predicting benign traffic during early batches.

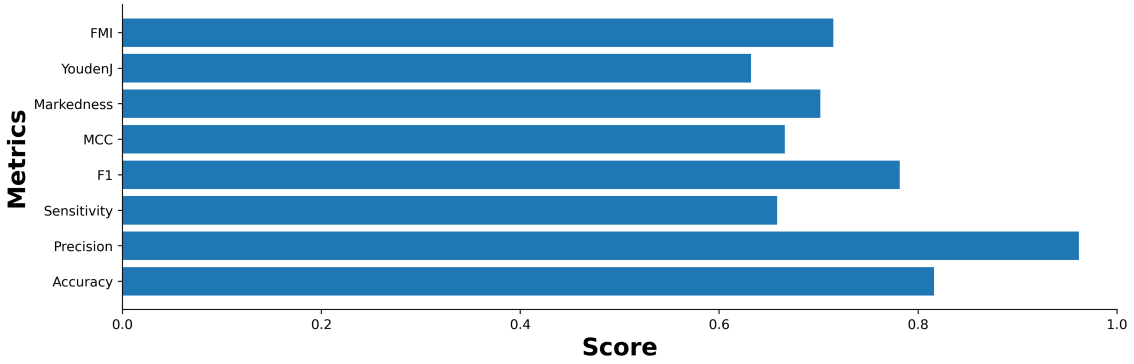

Figure 10: Cold-start performance without structured global initialization.

The imbalance propagates across other performance measures. The F1 score decreases to 0.7815, reflecting the recall loss. The Matthews Correlation Coefficient (MCC) falls to 0.6659, indicating weakened agreement between predictions and ground truth. Similarly, the Fowlkes–Mallows Index (FMI) reduces to 0.7146. Markedness (0.7017) and Youden's J statistic (0.6319) further confirm that the decision boundary is poorly conditioned at initialization.

Across repeated runs, random initialization consistently produced unstable early optimization behavior and weaker minority-class sensitivity during initial communication rounds.

In contrast, the initialized configuration discussed earlier stabilizes near the 0.97–0.99 accuracy range while maintaining substantially higher minority recall. The divergence between the two settings is most visible during the first communication rounds. Later batches show partial recovery in the ablated case, but the early vulnerability window remains extended. In streaming IoT deployments, that early phase corresponds to live traffic conditions and therefore carries operational significance.

We do not claim that structured initialization yields theoretical optimality. The experiment simply demonstrates that removing principled global conditioning degrades early stability and minority detection in practice. For streaming federated intrusion detection, this degradation is non-trivial.

4.7 Sensitivity Analysis of Accuracy-Gating Tolerance

To evaluate the effect of the client-side acceptance tolerance, additional experiments were conducted on three federated clients using multiple values of δ. The acceptance rule rejects a local update when its batch accuracy falls below $(1 - \delta)$ times the current server accuracy. Four operating points were examined: $\delta = \{0.01, 0.05, 0.10, 0.15\}$, corresponding to acceptance thresholds of 99%, 95%, 90%, and 85%, respectively.

Table 4 summarizes the results in terms of average accepted-update accuracy and the total number of accepted packets/connections (accepted batch updates). A stricter threshold yields higher average update quality but accepts fewer updates. In contrast, relaxed thresholds improve adaptation by accepting more updates, with a small reduction in average accuracy. The 95% threshold provides a balanced trade-off between robustness and update acceptance.

Table 4: Sensitivity analysis of client-side accuracy-gating tolerance under different acceptance thresholds.

Acceptance Threshold	Average Accuracy	Accepted Batch Updates
99% (δ = 0.01)	0.9838	88
95% (δ = 0.05)	0.9811	137
90% (δ = 0.10)	0.9796	210
85% (δ = 0.15)	0.9794	238

4.8 Comparison with State-of-the-Art

Table 5 summarizes quantitative performance and deployment characteristics of recent intrusion detection studies evaluated on CICIoT2023 and related settings. While several centralized approaches ([28–30,33]) report very high accuracy values, many are limited to offline evaluation. Federated learning is explored in [26], but without poisoning resilience or streaming adaptation. Incremental approaches ([31,32]) remain centrally coordinated. As observed, high accuracy alone does not imply robustness under attack-dominated imbalance or distributed updates. The proposed framework differs primarily in its integration of streaming federated learning with structured initialization and selective update acceptance under imbalanced IoT traffic conditions.

Table 5: Quantitative comparison with state-of-the-art on CICIoT2023.

Ref	Learning Type	Accuracy	Precision	Recall	F1	Special Mechanism	Real-World Deployment Consideration
Abid et al. [25]	Centralized	95.24%	95.22%	86.39%	95.09%	GWO and CNN-LGBM	Offline only
Torre et al. [26]	Federated	97.31%	95.59%	92.43%	92.69%	DP + FHE	10% overhead; no poisoning defense
Okey et al. [27]	Centralized	98%	High	High	High	Model compression	Two-stage training
Wahab et al. [28]	Centralized	99.20%	94%	93%	93%	SMOTE	3M subset only
Mallikarjun & Patro [29]	Centralized	99.73%	99.73%	99.73%	99.73%	SVWS balancing	Binary only
Wakili & Bakkali [30]	Centralized Hybrid	99.94%	95.53%	95.87%	95.64%	Anomaly-first cascade	CPU-only edge deployable
Wang et al. [31]	Incremental (Centralized)	96.43%	96.59%	96.43%	96.45%	Weight Alignment	High runtime (789s)
Mahdi et al. [32]	Incremental + Blockchain	99.89%	99.86%	99.92%	99.89%	Secure retraining	Central retraining only
Doménech et al. [33]	Centralized	99.85%	96.91%	97.12%	97.00%	SMOTE + optimized splits	Shows dataset shift issue
Saxena et al. [34]	Centralized (GNN)	90.69%	0.90 (weighted)	–	0.90	Graph modeling	Weak minority class precision
Proposed	Federated	99%	100%	86.36%	79.25%	Similarity-guided data balancing strategy	Cold-start-aware initialization; streaming federated learning

4.9 Communication Cost, Computational Efficiency, and Scalability

Beyond predictive performance, practical federated IoT deployment requires attention to communication cost and computational efficiency. The proposed framework uses incremental linear classifiers, which update model parameters batch-wise without repeated full retraining, reducing client-side processing overhead. In addition, mRMR-based feature selection reduces the original 46-dimensional feature space to 31 features, lowering both training cost and data handling complexity.

Communication overhead is also reduced through the proposed accuracy-gating mechanism. Only clients whose local batch performance satisfies the acceptance criterion transmit updated models to the server. This avoids unnecessary communication from low-quality or corrupted updates and can reduce uplink traffic under unstable client conditions.

The similarity-guided balancing stage is executed once during server-side initialization and therefore does not add recurring communication cost during federation. Its quadratic worst-case complexity may become expensive for very large datasets, but this step can be accelerated using approximate nearest-neighbor search, block-wise similarity computation, parallel processing, or sampled initialization subsets.

For larger federated deployments, the framework can be extended through client subsampling, asynchronous participation, or hierarchical aggregation across edge gateways. These strategies can improve scalability while preserving the proposed update filtering mechanism.

4.10 Threats to Validity

As the present study is based on simulation using the CICIoT2023 dataset, the reported findings should be interpreted with appropriate validity considerations.

Internal Validity: Internal validity may be affected by preprocessing choices, client data partitioning, hyperparameter settings, and assumptions used in the simulated federated environment. To reduce these effects, consistent preprocessing was applied across all clients, fixed evaluation settings were used, and multiple complementary metrics were reported instead of relying on a single performance measure.

External Validity: External validity concerns the generalization of the results beyond the current dataset and simulation setting. Real IoT deployments may exhibit different traffic distributions, larger numbers of clients, changing network conditions, device constraints, and more adaptive attack behaviors than those modeled here. Therefore, performance may vary under other datasets or operational environments.

Future work will include validation on additional datasets, larger-scale federated settings, and real-world deployment scenarios.

4.11 Effect of Initialization on Early Federated Stability

To further examine the effect of early federated initialization, an additional experiment was conducted using three alternative starting conditions for the global model: random balanced initialization, attack-dominant skewed initialization, and the proposed similarity-guided initialization. In all cases, the same client datasets, streaming protocol, and aggregation settings were maintained so that only the initialization strategy varied.

The random setting used an equal number of benign and malicious samples selected uniformly at random. The skewed setting used an attack-dominant initialization set with a 90:10 malicious-to-benign ratio. The proposed setting used the similarity-guided balanced initialization described in Section 3.2.

Table 6 reports the final aggregate performance. The proposed initialization achieved the highest Accuracy (0.979596) and FMI (0.979392), followed by the skewed and random settings. These results suggest

that initialization quality influences downstream federated learning performance, and that a balanced yet structurally diverse starting model provides improved convergence reliability.

Table 6: Impact of initialization strategy on federated performance.

Experiment	Accuracy	FMI
Random Initialization	0.975008	0.975414
Skewed Initialization	0.978662	0.978391
Proposed Initialization	0.979596	0.979392

5 Conclusions

This work investigated federated intrusion detection under sustained IoT attack conditions, where malicious traffic may dominate the data stream and create severe imbalance during deployment. Unlike conventional settings in which benign traffic overwhelms rare attacks, ongoing attack scenarios reverse this distribution and introduce a different form of instability. When attack-heavy gradients dominate early federated rounds, the global model can overfit malicious patterns, weakening discrimination of residual benign traffic and increasing false alarms or misclassification after attack intensity subsides.

To address this challenge, we introduced a similarity-guided balancing mechanism that preserves structurally diverse samples from the dominant class while constructing a stable and class-conditioned initialization dataset. The balancing stage is performed once during server-side initialization and can be accelerated through parallel, block-wise, or approximate similarity computation in large-scale deployments. The framework further integrates mRMR-based feature selection, reducing the original 46-dimensional CICIoT2023 feature space to 31 informative features through an 80% retention criterion based on combined relevance-redundancy scores. This compact representation improves efficiency while preserving discriminative traffic characteristics. In addition, the proposed framework employs incremental ensemble learning for continuous adaptation and an accuracy-gated update mechanism that selectively filters unreliable or corrupted client updates.

Experimental evaluation demonstrates stable convergence across clients, strong agreement between accuracy and minority-sensitive metrics, and deterministic rejection of updates under complete label-flipping poisoning. The ablation study confirms that removing structured initialization increases early-stage bias and extends the vulnerability window. An additional initialization study further confirmed that the proposed similarity-guided starting model yields stronger federated performance than random or attack-dominant initialization baselines. These results indicate that robust initialization, selective federation, and principled feature reduction jointly improve detection reliability in streaming IoT environments operating under adversarial and imbalanced conditions.

Few limitations remain. First, although the similarity-based redundancy computation is an offline one-time operation, its quadratic worst-case complexity may still require approximation for extremely large IoT telemetry streams. Second, adversarial evaluation was limited to complete label-flipping and strongly degraded adversarial batches; more subtle, adaptive, partial, stealthy, or collusive poisoning strategies were not examined. Third, the use of lightweight linear incremental classifiers may restrict expressive capacity under highly nonlinear traffic dynamics or severe cross-domain distribution shifts. Fourth, the current performance-weighted aggregation relies primarily on local accuracy and may favor clients with trivially high-performing local data. Fifth, a formal convergence guarantee under non-IID streaming data with

dynamic client acceptance and adversarial behavior was not derived in this study due to non-stationary distributions, incremental updates, and time-varying participation sets.

Future work will address these limitations by developing approximate or distributed similarity estimation techniques for scalable initialization, incorporating adaptive trust or reputation-aware aggregation mechanisms to counter stealthy adversaries, exploring lightweight nonlinear incremental models with domain-adaptive capabilities, and investigating theoretical convergence under selective participation with robustness bounds against poisoning attacks. These extensions aim to strengthen robustness in prolonged real-world attack scenarios while preserving the privacy, communication, and streaming constraints inherent to federated IoT intrusion detection.

Acknowledgement: The authors extend their appreciation to Umm Al-Qura University, Saudi Arabia, for funding this research work through grant number: 26UQU4290235GSSR01.

Funding Statement: This research work was funded by Umm Al-Qura University, Saudi Arabia under grant number: 26UQU4290235GSSR01.

Author Contributions: Arvind Prasad contributed to conceptualization, methodology, experimental design, data curation, formal analysis, software implementation, writing of the original draft, and manuscript revision. Ibrahim Aljubayri contributed through visualization, expert consultation, and professional guidance. Mohammad Zubair Khan contributed to supervision, validation, visualization, and critical manuscript review and editing. Abdulfattah Noorwali contributed to funding acquisition, investigation, and visualization. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The CICIoT2023 dataset used in this study is publicly available at https://www.unb.ca/cic/datasets/iotdataset-2023.html.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Al-Quayed F. AI-powered anomaly detection and cybersecurity in healthcare IoT with fog-edge. Comput Model Eng Sci. 2026;146(1):1–10. doi:10.32604/cmes.2025.074799.
2. Abdulganiyu OH, Fadi O, Moukafih Y, Ait Tchakoucht T, Saheed YK, Chukwuere JE, et al. Explainable attention based few shot LSTM for intrusion detection in imbalanced cyber physical system networks. Sci Rep. 2026;16(1):7217. doi:10.1038/s41598-026-38668-4.
3. Hazman C, Guezzaz A, Benkirane S, Azrour M. A smart model integrating LSTM and XGBoost for improving IoT-enabled smart cities security. Clust Comput. 2024;28(1):70. doi:10.1007/s10586-024-04780-1.
4. Singh S, Tyagi V, Malik A, Kumar R, Ankur, Kumar N. Intelligent energy-aware routing via protozoa behavior in IoT-enabled WSNs. IEEE Trans Netw Serv Manage. 2026;23:1960–9. doi:10.1109/TNSM.2025.3636202.
5. Prasad A, Chandra S. BotDefender: a collaborative defense framework against botnet attacks using network traffic analysis and machine learning. Arab J Sci Eng. 2024;49(3):3313–29. doi:10.1007/s13369-023-08016-z.
6. Khasawneh M, Azab A, Alrabaee S, Sakkal H, Bakhit HH. Convergence of IoT and cognitive radio networks: a survey of applications, techniques, and challenges. IEEE Access. 2023;11:71097–112. doi:10.1109/ACCESS.2023.3294091.
7. Al-Bakhrani AA, Li M, Obaidat MS, Amran GA. MOALF-UAV-MEC: adaptive multiobjective optimization for UAV-assisted mobile edge computing in dynamic IoT environments. IEEE Internet Things J. 2025;12(12):20736–56. doi:10.1109/JIOT.2025.3544624.
8. Karmous N, Jlassi W, Aoueileyine MO, Filali I, Bouallegue R. A new dataset for network flooding attacks in SDN-based IoT environments. Comput Model Eng Sci. 2025;145(3):4363–93. doi:10.32604/cmes.2025.074178.

9. Joe M, Kim M, Kwon M. Contrastive learning based network attack classifier for imbalanced data. J Commun Netw. 2026;28(1):86–97. doi:10.23919/JCN.2025.000082.
10. Kim MG, Kim H. Anomaly detection in imbalanced encrypted traffic with few packet metadata-based feature extraction. Comput Model Eng Sci. 2024;141(1):585–607. doi:10.32604/cmes.2024.051221.
11. Saba T, Sadad T, Rehman A, Mehmood Z, Javaid Q. Intrusion detection system through advance machine learning for the Internet of Things networks. IT Prof. 2021;23(2):58–64. doi:10.1109/MITP.2020.2992710.
12. Li C, Li J, Liu M, Wang H. Entropy-driven personalized privacy and dynamic federated learning inIoTSystems. Secur Priv. 2026;9(1):e70180. doi:10.1002/spy2.70180.
13. Patnaik LM, Wang W. AI fairness–from machine learning to federated learning. Comput Model Eng Sci. 2024;139(2):1203–15. doi:10.32604/cmes.2023.029451.
14. Khan AY, Latif R, Latif S, Tahir S, Batool G, Saba T. Malicious insider attack detection in IoTs using data analytics. IEEE Access. 2020;8:11743–53. doi:10.1109/ACCESS.2019.2959047.
15. Prasad A, Mohammad Alenazy W, Ahmad N, Ali G, Abdallah HA, Ahmad S. Optimizing IoT intrusion detection with cosine similarity based dataset balancing and hybrid deep learning. Sci Rep. 2025;15(1):30939. doi:10.1038/s41598-025-15631-3.
16. Sánchez PMS, Celdrán AH, Schenk T, Iten ALB, Bovet G, Pérez GM, et al. Studying the robustness of anti-adversarial federated learning models detecting cyberattacks in IoT spectrum sensors. IEEE Trans Dependable Secure Comput. 2024;21(2):573–84. doi:10.1109/TDSC.2022.3204535.
17. Lei T. Securing Fog-enabled IoT: federated learning and generative adversarial networks for intrusion detection. Telecommun Syst. 2025;88(1):11. doi:10.1007/s11235-024-01237-z.
18. Pathak J, Mundra P, Sejpal Y, Mahapatra T, Rajput AS. Early round detection protocols: strategies against untargeted adversarial attacks in federated learning network. Comput Netw. 2026;279:112098. doi:10.1016/j.comnet.2026.112098.
19. Prasad A, Chandra S, Uddin M, Al-Shehari T, Alsadhan NA, Sajid Ullah S. PermGuard: a scalable framework for Android malware detection using permission-to-exploitation mapping. IEEE Access. 2025;13:507–28. doi:10.1109/ACCESS.2024.3523629.
20. Alshehri MS, Saidani O, Al Malwi W, Asiri F, Latif S, Ahmad Khattak A, et al. A hybrid Wasserstein GAN and autoencoder model for robust intrusion detection in IoT. Comput Model Eng Sci. 2025;143(3):3899–920. doi:10.32604/cmes.2025.064874.
21. Tyagi V, Singh S, Wu H, Gill SS. Load balancing in SDN-enabled WSNs toward 6G IoE: partial cluster migration approach. IEEE Internet Things J. 2024;11(18):29557–68. doi:10.1109/JIOT.2024.3402266.
22. Malik A, Tyagi V, Singh S, Kumar R, Wu H, Gill SS. Optimizing secure data transmission in 6G-enabled IoMT using blockchain integration. IEEE Trans Consumer Electron. 2025;71(2):4534–43. doi:10.1109/TCE.2024.3510812.
23. Grispos G, Studiawan H, Alrabaee S. Internet of Things (IoT) forensics and incident response: the good, the bad, and the unaddressed. Forensic Sci Int Digit Investig. 2024;48(3):301671. doi:10.1016/j.fsidi.2023.301671.
24. Prasad A, Chandra S, Alenazy WM, Ali G, Shah S, ElAffendi M. AndroMD: an Android malware detection framework based on source code analysis and permission scanning. Results Eng. 2025;28(1):107050. doi:10.1016/j.rineng.2025.107050.
25. Abid T, Ahmim A, Maazouzi F, Chefrour D, Ullah I, Ahmim M, et al. A novel IoT threat detection using GWO feature selection and CNN-enhanced LightGBM. J Cloud Comput. 2025;14(1):72. doi:10.1186/s13677-025-00785-2.
26. Torre D, Chennamaneni A, Jo J, Vyas G, Sabrsula B. Toward enhancing privacy preservation of a federated learning CNN intrusion detection system in IoT: method and empirical study. ACM Trans Softw Eng Methodol. 2025;34(2):1–48. doi:10.1145/3695998.
27. Okey OD, Dadkhah S, Rodríguez DZ, Kleinschmidt JH. Explainable resource-Aware IoT security model via knowledge distillation and adaptive loss function optimization. Expert Syst Appl. 2026;302(7):130460. doi:10.1016/j.eswa.2025.130460.
28. Wahab SA, Sultana S, Tariq N, Mujahid M, Ali Khan J, Mylonas A. A multi-class intrusion detection system for DDoS attacks in IoT networks using deep learning and transformers. Sensors. 2025;25(15):4845. doi:10.3390/s25154845.

29. Mallikarjun A, Patro P. Hybrid data balancing with MLP probabilities-based categorical boosting model for robust intrusion detection system in IoT environment. Syst Soft Comput. 2026;8(16):200443. doi:10.1016/j.sasc.2026.200443.
30. Wakili A, Bakkali S. ZeroDefense: an adaptive hybrid fusion-based intrusion detection system for zero-day threat detection in IoT networks. J Electron Sci Technol. 2026;24(1):100345. doi:10.1016/j.jnlest.2026.100345.
31. Wang H, Yang Y, Tan P. CTWA: a novel incremental deep learning-based intrusion detection method for the Internet of Things. Artif Intell Rev. 2025;58(12):374. doi:10.1007/s10462-025-11358-9.
32. Mahdi ZS, Zaki RM, Alzubaidi L. A secure and adaptive framework for enhancing intrusion detection in IoT networks using incremental learning and blockchain. Secur Priv. 2025;8(4):e70071. doi:10.1002/spy2.70071.
33. Doménech J, León O, Siddiqui MS, Pegueroles J. Evaluating and enhancing intrusion detection systems in IoMT: the importance of domain-specific datasets. Internet Things. 2025;32(1):101631. doi:10.1016/j.iot.2025.101631.
34. Saxena S, Grover J, Singhal S. Exploring graph neural networks for robust network intrusion detection. Procedia Comput Sci. 2025;258(10):3630–9. doi:10.1016/j.procs.2025.04.618.
35. Farhan S, Mubashir J, Haq YU, Mahmood T, Rehman A. Enhancing network security: an intrusion detection system using residual network-based convolutional neural network. Clust Comput. 2025;28(4):251. doi:10.1007/s10586-025-05156-9.
36. Alrayes FS, Amin SU, Hakami N. An adaptive framework for intrusion detection in IoT security using MAML (model-agnostic meta-learning). Sensors. 2025;25(8):2487. doi:10.3390/s25082487.
37. Prasad A, Chandra S. PhiUSIIL: a diverse security profile empowered phishing URL detection framework based on similarity index and incremental learning. Comput Secur. 2024;136(4):103545. doi:10.1016/j.cose.2023.103545.
38. Neto ECP, Dadkhah S, Ferreira R, Zohourian A, Lu R, Ghorbani AA. CICIoT2023: a real-time dataset and benchmark for large-scale attacks in IoT environment. Sensors. 2023;23(13):5941. doi:10.3390/s23135941.
39. Khalil A, Farman H, Nasralla MM, Jan B, Ahmad J. Artificial Intelligence-based intrusion detection system for V2V communication in vehicular adhoc networks. Ain Shams Eng J. 2024;15(4):102616. doi:10.1016/j.asej.2023.102616.
40. Prasad A, Yadav V, Solanki C, Goswami H, Jha T, Nagal D. StealthPhisher: a defensive framework against phishing attack using hybrid deep learning and GenAI. Expert Syst Appl. 2026;299(4):130205. doi:10.1016/j.eswa.2025.130205.
41. Zhao Z, Zheng M, Ma F, Li T. A novel federated adaptive hybrid sampling framework for imbalanced data classification. Appl Soft Comput. 2026;193(8):114866. doi:10.1016/j.asoc.2026.114866.
42. Zhao X, Wu Y, Chaddad A, Daqqaq T, Kateb R. Federated vision transformer with adaptive focal loss for medical image classification. Knowl Based Syst. 2026;338(Suppl 3):115474. doi:10.1016/j.knosys.2026.115474.
43. Alnajar O, Barnawi A. A novel clustered distributed federated learning architecture for tactile Internet of Things applications in 6G environment. Comput Model Eng Sci. 2025;143(3):3861–97. doi:10.32604/cmes.2025.065833.
44. Martinez-Lopez F, Santana L, Rahouti M, Chehri A, Al-Maliki S, Jeon G. Learning in multiple spaces: prototypical few-shot learning with metric fusion for next-generation network security. IEEE Trans Netw Serv Manage. 2026;23:3156–65. doi:10.1109/TNSM.2026.3665647.
45. Krouka M, Ben Issaid C, Bennis M. Distributionally robust federated learning with client drift minimization. Trans Mach Learn Comm Netw. 2026;4:438–56. doi:10.1109/TMLCN.2026.3658026.