AI

MDPI

Article

Developing a Federated, Lightweight, Interpretable, and Specialized AI-Based Intrusion Detection System for Medical Internet of Things (IoMT) Environments

Jehad M. Hamamreh [1,2,*], Aman M. Araf [3], Ahmad M. Jaradat [4], Adnan Daraghmeh [5], Moamin Abughazala [6,7] and Henry Muccini [7]

[1] Department of Computer Science and Cybersecurity, An-Najah National University, Nablus P.O. Box 7, Palestine
[2] Electrical and Computer Engineering Department, Antalya Bilim University, 07190 Antalya, Turkey
[3] Department of Cybersecurity, An-Najah National University, Nablus P.O. Box 7, Palestine
[4] Department of Electrical and Computer Engineering and Computer Science, University of Detroit Mercy, Detroit, MI 48221-3038, USA
[5] Department of Mathmatics, An-Najah National University, Nablus P.O. Box 7, Palestine
[6] Department of Computer Science Apprenticeship, An-Najah University, Nablus P.O. Box 7, Palestine; m.abughazaleh@najah.edu
[7] Department of Information Engineering, Computer Science and Mathematics, University of L'Aquila, 67100 L'Aquila, Italy; henry.muccini@univaq.it
* Correspondence: jehad.hamamreh@najah.edu or jehad.hamamreh@antalya.edu.tr

Abstract

The rapid expansion of the Internet of Medical Things (IoMT) has enhanced healthcare services, but it has also exposed these systems to a growing range of cyberattacks. Addressing this challenge requires effective intrusion detection solutions. Machine learning-based intrusion detection systems (IDSs) offer a promising approach. In this research, we leverage IoMT-TrafficData, a recently published dataset specific to IoMT environments, to develop intrusion detection models based on federated learning (FL), which preserves data privacy while enabling the use of lightweight learning algorithms suited to resource-constrained devices. In addition, SHAP-based interpretability methods are employed to explain model decisions, thereby improving reliability and transparency. Compared with existing IDS solutions for IoMT, this research proposes an approach that balances performance, privacy, and interpretability while accounting for resource limitations, resulting in three unified, lightweight, and interpretable detection models tailored to different data types in IoMT environments. This work thus provides a solid foundation for future research on secure medical IoT systems.

Keywords: IoMT-TrafficData; federated learning; lightweight; explainable AI; SHAP; IDS; medical internet of things; IoMT; IoT industry; ICT; health and social care services

Academic Editors: Netzahualcoyotl Hernandez-Cruz, Ping Lu, Jens Lundström and Irvin Hussein Lopez-Nava

Received: 4 March 2026
Revised: 10 April 2026
Accepted: 21 April 2026
Published: 22 September 2026

1. Introduction

IoT systems have been among the most widely distributed in recent years. They consist of a group of interconnected entities and devices that collect information from the surrounding environment, process and analyze it, issue commands, and perform other tasks. These systems are present in various sectors such as telecommunications, transportation, industrial control, healthcare, and other sectors [1]. IoMT systems have made significant advances in healthcare systems, allowing doctors to monitor and diagnose

patients remotely and easily exchange information between health devices and hospital infrastructure. The rapid advancement of portable devices, remote patient monitoring, cloud medical applications, and others has made these systems an essential part of healthcare services [2]. However, this growth and reliance on wireless communications expose them to a wide range of cybersecurity threats. Compared to other Internet of Things (IoT) systems, IoMT systems expose highly sensitive patient data and critical equipment to risks that can threaten safety, in addition to being constrained by limited computational power, memory, and battery life, as well as latency sensitivity [3]. The main problem is that many IoMT devices have experienced cyberattacks in recent years simply because medical devices are either not well protected from attackers or are completely insecure. This leads to severe cyberattacks, resulting in delays in treatment or the exposure of sensitive patient data [4]. According to the CyberMDX 2020 report, medical devices in the IoT are vulnerable to numerous cyberattacks due to the serious security vulnerabilities they contain [5]. The importance of protecting these systems became evident during the COVID-19 pandemic [6]. Due to the importance of these medical devices, the security of IoMT systems has become a top priority.

To deal with threats and attacks on IoMT systems, various techniques and methods are used, including the use of encryption, updates, security patches, blockchain-based security, artificial intelligence for behavior analysis, intrusion detection systems (IDSs/IPSs), and other methods [7]. IDSs are among the most common defense methods because they help detect not only known cyberattacks but also new ones. These systems have the ability to detect and prevent various types of cyberattacks targeting IoT systems as well as IoMT systems, with IDSs analyzing traffic to identify anomalies and alert the system to take appropriate actions [8,9]. IDS models generally rely on centralized data collection, which affects privacy, especially in medical care environments. Additionally, these models require computational resources with significant capabilities, making them unsuitable for limited-capacity devices in IoMT systems [10,11]. One of the issues with IDS models is their inability to interpret decisions, as they function as black boxes due to the lack of understanding of the reasons behind the detection system's decisions [12]. Many previous studies have applied machine learning (ML) to IDSs to make them lightweight and suitable for devices with limited computational resources, or they have utilized FL within IoT systems. However, few studies integrate all three elements—ML-based lightweight IDS, FL, and explainable AI (XAI)—while specifically focusing on IoMT systems or using IoMT datasets for model training. Current research either uses FL solely for privacy preservation [13] or updates lightweight models without using FL [14]. Additionally, several studies integrate XAI techniques into the models despite their importance and reliance in the real world. This research aims to develop an IDS model that uses FL and XAI-based interpretability in lightweight, resource-constrained environments, where the model is specifically trained on data from IoMT environments (see Figure 1 for an overview of the proposed framework).

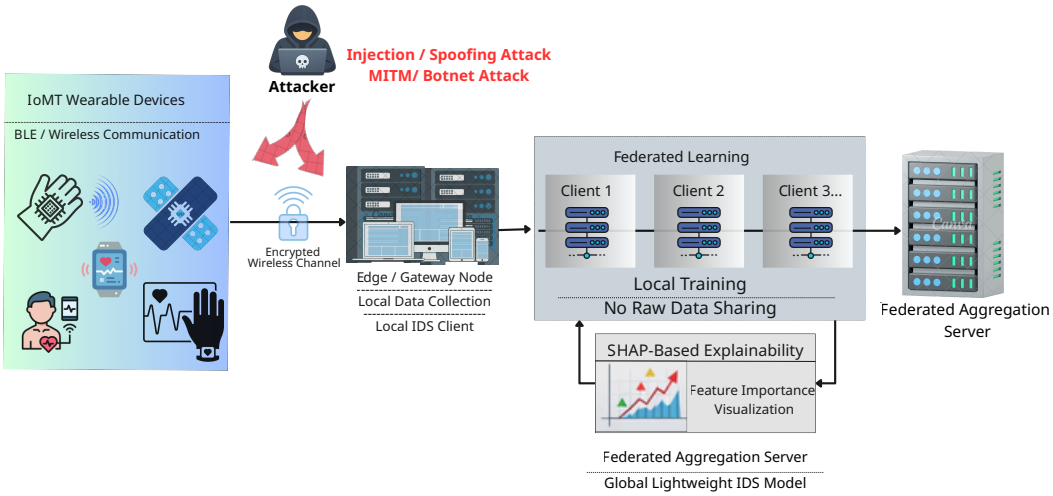

Figure 1. Overview of the proposed lightweight federated intrusion detection framework for IoMT environments.

2. Related Work

Securing IoMT environments is critically important due to the widespread and increasing prevalence of these devices, such as wearables and remote patient monitoring systems, and the rise in cyberattacks on these devices. Recent studies have revealed various defense mechanisms, including ML applications, FL, and lightweight models to detect attacks and malicious activities in these devices [14,15]. FL has been effectively applied in medical image analysis to train ML models on sensitive patient data without compromising privacy [16]. While these studies primarily focused on collaborative training of models across multiple hospitals for diagnostic purposes, the underlying principles of FL are also applicable to IoMT environments. Specifically, adopting FL for IDS in IoMT devices allows distributed training across multiple devices or locations while preserving privacy and enabling the development of lightweight, interpretable models suitable for resource-constrained devices. Integrating these insights strengthens the rationale for using FL in IoMT IDSs, bridging the gap between established FL applications in medical diagnostics and the emerging needs for privacy-preserving, lightweight, and interpretable IDS frameworks. Previous research focused either on general IoT environments or on specific security aspects [17–20], without integrating FL, lightweight models, XAI decision interpretability, and IoMT datasets simultaneously.

In general, studies have primarily relied on centralized aggregation to train IDS models for the IoMT, achieving good detection performance, but facing the issue of not ensuring privacy. Researchers reviewed the security challenges of the IoMT and confirmed that medical communications are vulnerable to manipulation attacks, ransomware, and identity theft [2]. In another study, FD-IDS was used, which is a detection model utilizing FL with knowledge distillation, addressing the data issues present in IoMT systems in general, but it did not tackle the problem of resource-limited devices or the addition of XAI to the model [21].

In other studies, IDS methods specifically designed to address resource constraints on devices have been proposed. Some researchers developed anomaly-based detection methods tailored for low-power IoMT devices, but their drawback is reliance on centralized training and a lack of privacy protection [22]. Others proposed an adaptive detection system based on fuzzy logic for medical devices, but they did not incorporate FL or decision interpretation features [23].

In another study, AI was increasingly used in IDSs to improve trust and transparency in clinical environments. Some researchers have improved multilayer IDSs using XAI [24].

In other studies, interpretable FL was used with SHAP [18,19], but their solutions did not focus on resource-constrained IoMT environments. In other studies, an interpretable IDS using XGBoost was proposed, but the research did not include FL or focus on designing lightweight models [20]. In another study, a lightweight FedTinyMed model based on FL for IoMT devices was developed, but it did not focus on the IDS approach or the interpretability of the readings [25].

Table 1 provides a detailed comparison of the most prominent research studies in this field, along with an explanation of the scope of each study.

Table 1. Comparison of Previous Studies on Intrusion Detection in IoT and IoMT Environments.

Paper	FL	XAI	Lightweight	Dataset Used	Algorithms	Research Scope
Cross-Layer FL for Lightweight IoT IDS [17]	Yes	No	Yes	BoT-IoT, NSL-KDD	CNN + FL	General IoT
Federated XAI IDS with SHAP [18]	Yes	Yes	No	CICIDS2017	DNN + SHAP	General IoT
XAI-XGBoost for IoMT [20]	No	Yes	No	Simulated IoMT	XGBoost	Healthcare/IoMT
BFLIDS [13]	Yes	No	No	Simulated IoMT	DNN + Blockchain + FL	Healthcare/IoMT
FD-IDS with Knowledge Distillation [21]	Yes	No	Yes	NSL-KDD, UNSW-NB15	KD-DNN	General IoT
FedTinyMed [25]	Yes	No	Yes	Physiological sensors	TinyML + FL	Healthcare/IoMT
Anomaly-Based IDS for IoMT [22]	No	Partial	Yes	Simulated IoMT	Statistical + ML	Healthcare/IoMT
Explainable AI Cross-Layer IDS [24]	No	Yes	Partial	Custom IoMT	RF, XGBoost	Healthcare/IoMT
Adaptive Fuzzy-Based IDS for IoMT [23]	No	No	No	Custom IoMT	Fuzzy Logic	Healthcare/IoMT
Explainable ML Security Framework [26]	Yes	Yes	No	Simulated IoMT	RF, SVM + SHAP	Healthcare/IoMT
Lightweight FL-based IDS for IoT [27]	Yes	No	Yes	TON_IoT, BoT-IoT	Lightweight DNN	General IoT
Matrix-Valued Neural FL for IoHT [28]	Yes	Partial	Partial	IoHT mixed data	Matrix NN + FL	Healthcare/IoHT

Current studies provide excellent contributions to the field of the IoMT, but very few or almost none offer a comprehensive framework that provides the following:

(1) A lightweight IDS model suitable for resource-constrained medical devices.
(2) A model that preserves privacy through FL.
(3) Decisions that are interpretable using XAI.

The aim of the research is to propose a lightweight, interpretable, unified IDS specifically designed for IoMT systems.

3. Dataset Description

In this research, the recently published IoMT-TrafficData dataset was used to train an IDS model for IoMT systems [29]. These data include scenarios that simulate attacks on real IoMT networks, and they comprise a set of Bluetooth Low Energy (BLE) traffic data in addition to an IP protocol traffic dataset taken from various medical devices, making them

suitable for use in an IDS model. These data are available to everyone in both pickle and CSV formats, making it easy to integrate them into ML models.

The IoMT-TrafficData project provides a comprehensive dataset and accompanying tools for evaluating IDS in IoMT environments, facilitating the development, assessment, and reproducibility of ML-based IDS solutions [29].

The IoMT-TrafficData includes the following subgroups:

- Bluetooth Low Energy (BLE) Traffic Dataset: This dataset contains data specific to BLE connections at the packet level, including features such as advertisement headers, protocol data units (PDUs), physical-layer (PHY) information, cyclic redundancy checks (CRCs), packet timings, and other indicators specific to BLE. These features are essential for traffic analysis, especially for short-range wireless communications involving wearable medical sensors (such as glucose sensors or heart rate monitors, among others). These data are characterized by features that enable precise tracking of communication traffic between devices, making them suitable for detecting anomalies in these resource-limited and short-range devices.
- IP-packet-based dataset: This dataset represents the data traffic of medical devices operating at the IP layer, such as smart monitoring devices and ventilators, among others. It contains basic features related to the HTTP and TCP protocols, including request and response codes, TCP flags, data sizes, and connection start and end signals, among other data useful for monitoring network activity at the IP layer.
- IP-flow-based dataset: This dataset captures network traffic at the flow level and contains a variety of features that provide a comprehensive view of communication patterns within IoMT networks, such as service and transport protocols, the number of packets sent and received, connection state records, time intervals between packet arrivals, and flow duration.

In this study, the IoMT-TrafficData dataset, which consists of three subgroups, was processed so that each subgroup could be trained individually as an independent IDS model for IoMT systems. Each category (BLE, IP-packet-based, and IP-flow-based) operates under different constraints and contains a unique set of features that cannot be merged with those of the other categories. Data from each category were used to train a lightweight IDS model, and FL was applied independently to each category by creating several clients per category and aggregating them centrally using FedAvg. Each IDS model operates independently within IoMT environments. Through this design, the lightweight and specialized IDS ensures suitability for deployment on appropriate devices.

3.1. BLE Traffic Dataset

The dataset for the proposed IDS consists of packet-level communications using BLE in IoMT devices such as wearables and sensors.

Features: Packet- and frame-level features were extracted, as summarized in Table 2.

Table 2. Key features of the BLE traffic dataset.

Feature Group	Examples
Advertising header fields	`btle.advertising_header`, `ch_sel`, `length`, `pdu_type`
CRC correctness indicators	`btle.crc.incorrect`, `btle.crcok_0`, `btle.crcok_1`
Nordic BLE metadata	`board_id`, `packet_time`, `packet_counter`, channels 37–39
Frame information	`cap_len`, `len`, `interface_id_0/1/2`
Flags	`btle.flags_0/1`

Preprocessing: To enhance model stability and efficiency under federated non-IID training, the following steps were applied:

- Remove constant or quasi-constant features to reduce noise.
- Median imputation for missing values to handle outliers.
- Convert categorical/boolean features into numerical format.
- Select top features using ANOVA F-test (SelectKBest) to reduce dimensionality.
- Standardize features via Z-score normalization to prevent dominance of large-scale attributes.

Dataset Partitioning: The dataset was stratified to preserve class distribution (Table 3):

Table 3. BLE dataset partitioning.

Subset	Samples
Training	135,884
Validation	15,099
Test	26,645

Federated Learning Distribution: The training set was split across 50 clients using a Dirichlet-based non-IID allocation ($\alpha = 1.0$), introducing realistic heterogeneity:

- Minimum 200 samples per client to ensure stable local updates.
- Contribution caps to prevent large clients from dominating.
- Balanced aggregation to maintain equitable influence on the global model.

Note: Clients with samples from a single class may be skipped in some training rounds to ensure meaningful local updates, which reflects realistic heterogeneity and possible class imbalance in IoMT deployments.

This setup realistically simulates IoMT deployments while preserving data locality, privacy, and robustness under heterogeneous client distributions.

3.2. IP-Packet-Based Dataset

The dataset for the proposed IDS consists of IP-based network traffic captured from IoMT devices such as wearables, sensors, and gateways.

Features: Packet- and flow-level features were extracted and grouped into meaningful categories, summarized in Table 4.

Table 4. Key feature groups of the IP-packet-based IoMT dataset.

Feature Group	Examples
HTTP-level statistics	`http.content_length`, `http.request`, `http.response.code`, `http.response_number`, `http.time`
TCP connection flags	`tcp.connection.syn`, `tcp.connection.synack`, `tcp.connection.fin`
TCP packet flags	`tcp.flags.syn`, `tcp.flags.fin`, `tcp.flags.urg`, `tcp.flags.cwr`, `tcp.flags.ecn`, `tcp.flags.ns`, `tcp.flags.res`, `tcp.urgent_pointer`
IP-level metadata	`ip.frag_offset`
Ethernet/ARP metadata	`eth.dst.ig`, `eth.src.ig`, `eth.src.lg`, `eth.src_not_group`, `arp.isannouncement`

Preprocessing: To ensure robustness and stability under federated non-IID training, the following steps were applied:

- Convert categorical and boolean features into numerical format.
- Handle missing or infinite values by median imputation.
- Clip negative values to zero where necessary.
- Scale features using RobustScaler to reduce the influence of outliers.

Duplicate Retention: The dataset exhibits high redundancy due to repeated IoMT traffic flows. While duplicate removal significantly reduces dataset size, it distorts the natural distribution of the benchmark dataset. Therefore, all samples were retained to preserve traffic realism and frequency patterns.

Dataset Partitioning: The dataset was stratified to preserve class distribution, ensuring a realistic representation of benign and malicious traffic. All samples, including duplicates, were retained to preserve realistic IoMT traffic patterns. The resulting splits are as follows (Table 5):

Table 5. IP-packet-based IoMT dataset partitioning (including duplicates).

Subset	Samples
Training	699,975
Validation	150,025
Test	150,000

FL Distribution: The training set was split across 50 clients using a Dirichlet-based non-IID allocation ($\alpha = 0.7$) to simulate realistic heterogeneity:

- A minimum number of samples per client ensured stable local updates.
- Contribution weighting was applied to prevent large clients from dominating the global model.
- Local client updates were aggregated using weighted averaging to maintain equitable influence.

Note: Some clients may contain samples from a single class in certain rounds. These clients are skipped in those rounds to maintain meaningful updates, reflecting realistic class imbalance in IoMT deployments.

This setup simulates real-world IoMT networks while preserving data locality, privacy, and robustness under heterogeneous client distributions.

3.3. IP-Flow-Based Dataset

The dataset contains **358,769 IP-based flow samples** (after cleaning), labeled with `is_attack` indicating benign (0) or malicious (1) activity.

Data Cleaning and Preprocessing:

- **Duplicate removal:** 154,793 duplicates were removed from training, and 43,695 from test.
- **Overlap removal:** Any identical rows appearing in both training and test were dropped to prevent label leakage.
- **Missing values:** Imputed using mean values after feature transformations.
- **Feature scaling:** Numerical features were standardized with Z-score normalization.
- **Categorical encoding:** Boolean and categorical features were converted via one-hot encoding; unseen categories were handled safely.

After cleaning and preprocessing, the final dataset shapes are as follows:

- **Training:** 229,824 samples, 31 features
- **Validation:** 40,558 samples, 31 features
- **Test:** 88,387 samples, 32 features

Identified Key Feature Groups: Features are grouped for clarity (Table 6):

Table 6. Grouped key features of the IP-flow-based dataset.

Feature Group	Examples
Protocol/Service	proto, service
Payload	orig_bytes, resp_bytes, fwd_pkts_payload, bwd_pkts_payload, flow_pkts_payload
Packets	orig_pkts, resp_pkts, fwd_pkts_tot, bwd_pkts_tot, fwd_data_pkts_tot
Timing/Flow	flow_duration, flow_iat, fwd_iat, bwd_iat, active
Connection History	history_originator, history_responder
Headers	fwd_header_size, bwd_header_size
Rates/Derived	fwd_pkts_per_sec, bwd_pkts_per_sec, flow_pkts_per_sec

Dataset Partitioning:

- Training set further split into **train (∼85%)** and **validation (∼15%)**, stratified to preserve class balance.

FL Distribution:

- 50 clients with Dirichlet non-IID allocation ($\alpha = 0.5$).
- Weighted FedAvg aggregation based on client data size.
- Local oversampling applied to clients with class imbalance.
- Clients with too few samples or single-class flows were skipped in some rounds.

Advanced Hard Test Generation:

- Programmatic transformations applied to both training and test sets to simulate **realistic, challenging IoMT traffic conditions**:
  - **Feature dropout** and **missing values** simulate incomplete measurements.
  - **Packet loss, delay, jitter, and protocol distortion** mimic network errors and traffic variations.
  - **Gaussian noise, dynamic shift, and label noise** reflect evolving traffic distributions and annotation errors.
- The transformed dataset tests the **robustness of the IDS against noise, traffic shifts, and real-world network perturbations**.

Interpretability:

- SHAP applied to the global model and the first three valid client models.
- Helps identify key contributing features under both normal and perturbed conditions.

"These transformations create a realistic, challenging evaluation scenario, testing the IDS's robustness against noise, errors, and traffic shifts in IoMT networks."

The IoMT-TrafficData dataset provides a comprehensive and realistic representation of IoMT network traffic by covering multiple connection aspects, including BLE, IP packet traffic, and IP stream traffic. Each dataset has distinct characteristics of IoMT communications, allowing for the analysis of attacks at different levels, from low-level wireless interactions to high-level traffic behavior.

Based on this dataset, separate learning models were built suitable for each type of traffic to preserve the features of each subset and avoid cross-domain distortion. In the following section, the methodology used to process this data and simulate lightweight, interpretable FL models is described.

4. Methodology

4.1. Overview of the Proposed Methodology

In this section, the methodology followed in evaluating and classifying a lightweight IDS that preserves privacy and tailored for IoMT environments, is described. The adopted methodology as depicted in Figure 1, is based on a multi-model approach, where each type of data (medical IoT traffic—BLE, IP packet traffic, and IP stream traffic) is processed and modeled independently to preserve the intrinsic properties of the data. For each type of data, a lightweight learning model is trained using FL to simulate IoMT environments. And to enhance trust and transparency, the model's decisions are explained using SHAP-based interpretation at both the client and global levels. Using this approach, data integrity is ensured, and leakage between training and test sets is avoided, which maintains a balanced distribution of categories across federated clients.

4.2. BLE-Based IDS Model

BLE technology is widely used in IoMT devices due to its low power consumption, which makes it particularly suitable for wearable and implantable medical devices. Despite its advantages, BLE is vulnerable to a variety of cyberattacks, highlighting the need for continuous monitoring and robust intrusion detection mechanisms in IoMT environments [29]. To address the privacy concerns associated with distributed IoMT data, federated learning (FL) is employed instead of traditional centralized training. In this approach, each model is trained locally on individual clients, and only model updates are shared with a central aggregator, thereby preventing the exposure of sensitive patient information [30].

The proposed BLE-based IDS is implemented as a unified federated framework, integrating two main pipelines:

- **Federated Training Pipeline:** Handles distributed learning, local optimization, probability calibration, and global model aggregation.
- **Evaluation and Analysis Pipeline:** Performs inference, performance assessment, robustness testing, and model interpretability analysis.

Together, these pipelines form a coherent system capable of privacy-preserving, lightweight, and interpretable IDS.

Feature Selection and Normalization: The BLE dataset consists of 177,628 samples after removing duplicates, with 30 initial packet-level and frame-level features extracted from BLE communications in IoMT devices. To ensure computational efficiency and improve generalization, the dataset underwent the following preprocessing:

- Removal of static and semi-static features with low variance or 99% repetition.
- Conversion of categorical and boolean attributes into numerical format.
- Statistical feature selection using the ANOVA F-test (SelectKBest), retaining the top 23 discriminative features compatible with resource-limited IoMT devices.
- Standardization via z-score normalization (zero mean, unit variance) to stabilize optimization and prevent features with large numerical ranges from dominating the learning process.

This pipeline ensures lightweight computation, numerical stability, and suitability for federated training on heterogeneous IoMT devices.

FL Distribution: For FL, the cleaned training set was partitioned across 50 clients using a non-IID Dirichlet distribution ($\alpha = 1.0$), modeling realistic device-level heterogeneity. Each client had at least 200 samples to ensure stable local updates, while contribution caps prevented dominance of large clients, maintaining fairness in global aggregation. Only model parameters, not raw data, were shared, preserving sensitive medical information.

4.3. Choice of Dirichlet Parameter α

To study client heterogeneity in FL, we tested $\alpha \in \{0.1, 0.5, 1.0, 2.0, 5.0\}$. Smaller α induces stronger non-IID partitions, while larger α approaches IID as can be seen from Table 7.

Table 7. Impact of α on client dataset variance and validation F1.

α	Min	Max	Variance	F1
0.1	200	25,187	35,844,773	0.9451
0.5	200	13,192	10,492,868	0.9451
1.0	364	8832	5,284,549	0.9451
2.0	1019	5912	1,414,608	0.9451
5.0	1633	4727	562,749	0.9451

Validation F1 remained stable across all α, while client dataset variance decreased as α increased. We selected $\alpha = 1.0$ as it balances:

- Sufficient local samples for stable optimization (min = 364),
- Moderate dataset variance (5,284,549),
- Optimal validation performance (F1 = 0.9451).

Despite the wide range of client dataset variance, validation F1 remained stable due to several mechanisms in our FL setup:

1. **Minimum samples per client enforcement** (>200) prevents clients from destabilizing training.
2. **FedProx proximal regularization** ($\mu = 0.02$) mitigates divergence of local updates from the global model.
3. **FedAvg with server-side momentum** ($\beta = 0.95$) smooths the aggregation of local updates.
4. **Balanced class weights and feature selection** stabilize local logistic regression training.

Each client trained a local logistic regression model for 3 epochs (`LOCAL_EPOCHS = 3`) with a batch size of 8000 (`LOCAL_BATCH_SIZE = 8000`). On average, clients contributed between 404 and 9997 samples per round, reflecting realistic variation in device data availability.

FedProx Local Update and Weighted FedAvgM

To address client heterogeneity, each client k minimizes the following FedProx objective:

$$\mathcal{L}_k^{\text{FedProx}}(w) = \mathcal{L}_k(w) + \frac{\mu}{2}\|w - w_t\|^2 \quad (1)$$

where w_t denotes the global model parameters at communication round t, and $\mu = 0.02$ controls the strength of proximal regularization. The value of μ was selected empirically to balance convergence speed and stability; larger values overly restricted local updates, whereas smaller values were insufficient to control divergence under non-IID conditions.

The server aggregates client updates using weighted FedAvg with momentum:

$$v_{t+1} = \beta v_t + \sum_{k=1}^{K} \frac{n_k}{n} (w_k^{t+1} - w_t) \quad (2)$$

$$w_{t+1} = w_t + v_{t+1} \quad (3)$$

where n_k denotes the number of samples at client k, $n = \sum_{k=1}^{K} n_k$, and $\beta = 0.95$ is the server momentum coefficient.

Unlike standard FedAvg [30], which directly averages local model updates, the proposed formulation integrates (i) proximal regularization to reduce client drift and (ii) server-side momentum to stabilize and accelerate global convergence under heterogeneous data distributions.

Model Optimization and Early Stopping: The federated training ran for up to 30 communication rounds. Early stopping based on validation F1-score prevented overfitting: training halted if no improvement occurred for three consecutive rounds. Server-side momentum aggregation ensured smoother updates under non-IID distributions.

Probability Calibration and Thresholding: Probability calibration was applied using Platt scaling (sigmoid) after the final round of federated training to ensure well-calibrated probability outputs. Per-round calibration was not performed. A recall-constrained thresholding strategy enforced a minimum recall of 99.5%, selecting the threshold that minimized false positives. If no threshold met the recall requirement, a fallback threshold maximizing the F1-score was applied.

Evaluation and Interpretability: The final evaluation used the test set and standard metrics: F1-score, ROC-AUC, PR-AUC, and confusion matrices. The SHAP algorithm provides local and comprehensive explanations by determining the contribution of each feature to the model's outputs, enabling security analysts to understand decision-making behavior [31].

System Summary: In summary, the BLE-based IDS integrates:

- Federated, privacy-preserving learning;
- Lightweight logistic regression suitable for IoMT devices;
- Robust optimization under heterogeneous client data;
- Calibrated thresholds aligned with safety requirements;
- SHAP-based explainability for transparent decision-making.

This unified framework ensures scalable, interpretable, and computationally efficient IDS for real-world IoMT environments.

4.4. IP-Packet-Based IDS Model

The proposed IDS monitors IP-packet-based traffic in IoMT networks, detecting malicious activities such as DoS, scanning, and abnormal device behavior. The model integrates **dynamic hard data augmentation**, **FL**, and **probability calibration** to form a multi-stage pipeline.

Data Preprocessing and Feature Handling

The dataset contains 583,769 samples with 23 packet-level features extracted from HTTP, TCP, IP, Ethernet, and ARP layers. Preprocessing steps include:

- Conversion of categorical and boolean fields into numerical values.
- Replacement of infinite values and median imputation for missing values.
- Robust scaling via `RobustScaler` to mitigate outlier influence.
- Retention of duplicates, preserving traffic patterns present in the benchmark dataset.

Dynamic Hard Data Augmentation

To enhance robustness under non-IID and noisy conditions, hard data augmentation is applied locally on each client:

- Addition of Gaussian noise, with amplified variance for rare (malicious) samples.
- Random multiplicative shifts to simulate feature variability in IoMT flows.
- Feature masking/dropout to emulate packet loss or incomplete observations.
- Small perturbations in rare-class samples to improve generalization.

Federated Non-IID Distribution

The training set is split across 50 clients using a Dirichlet-based allocation ($\alpha = 0.7$), introducing realistic heterogeneity:

- Clients may receive imbalanced proportions of benign and malicious samples.
- Clients with single-class samples are skipped in some rounds to ensure meaningful updates.
- Weighted aggregation ensures clients contribute proportionally to their dataset sizes.

Local Training and Aggregation

Each client trains a local `SGDClassifier` with logistic regression (log-loss):

- Maximum one epoch per round with mini-batch updates.
- ElasticNet regularization (penalty: 0.0003, L1 ratio: 0.1) to prevent overfitting.
- Class weighting to enhance the detection of rare malicious samples.
- Oversampling applied locally for minority classes.

After each round, local model coefficients and intercepts are aggregated using weighted FedAvg to form the global model.

Calibration and Threshold Tuning

The aggregated model is calibrated using `CalibratedClassifierCV` with the sigmoid method:

- Probabilities are adjusted for better reliability in real-world deployment.
- Optimal classification threshold is selected on the validation set, maximizing a combined F1 + Recall score.

Evaluation and Robustness Testing

The model is evaluated under two conditions:

- Original Test Set**:** Reflects standard IoMT traffic; accuracy was ∼98%, F1 was ∼0.98, and ROC-AUC ∼ 0.998.
- Hard Test Set**:** Locally augmented and noisy; simulates challenging scenarios, yielding accuracy of ∼94%, F1 of ∼0.94, and ROC-AUC of ∼0.960.

Interpretability via SHAP

SHAP analysis is conducted on a representative subset of the test set:

- Identifies globally important features contributing to classification decisions.
- Supports forensic analysis and model transparency in IoMT environments.

System Summary

The proposed IDS framework provides:

- Federated, privacy-preserving learning with heterogeneous clients.
- Lightweight SGD-based logistic regression suitable for resource-constrained IoMT devices.
- Robustness through dynamic hard data augmentation and oversampling.
- Weighted FedAvg ensuring balanced contribution from all clients.
- Probability-calibrated ensemble predictions.

- SHAP-based interpretability for transparent decision-making.

This results in a scalable, interpretable, and robust IDS for real-world IoMT packet monitoring.

4.5. IP-Based Flows IDS Model

The proposed IDS monitors IP-flow-based traffic in IoMT networks, detecting malicious activities such as DoS, scanning, and data exfiltration. It integrates dynamic hard data augmentation and **FL** to form two main pipelines:

- **Federated Training Pipeline:** Distributed learning across 50 clients with local model optimization, dynamic data augmentation, probability calibration, and weighted FedAvg aggregation.
- **Evaluation and Analysis Pipeline:** Performs inference on standard and advanced hard tests, assesses performance under noisy and shifted conditions, and provides SHAP-based interpretability.

Feature Selection, Preprocessing, and Hard Data Generation: The dataset includes 583,769 samples with 32 initial features. Preprocessing and augmentation include:

- Duplicate and overlap removal to prevent leakage.
- One-Hot Encoding for categorical/boolean features; unseen categories safely handled.
- Z-score standardization for numerical features.
- Mean imputation for missing values post-transformation.
- Dynamic hard data augmentation: feature dropout, missing values, packet loss/delay/jitter, protocol distortions, Gaussian noise, dynamic feature shifts, and label noise.

Federated Distribution:

- 50 clients; Dirichlet non-IID ($\alpha = 0.5$). We conducted a sensitivity analysis of the Dirichlet parameter $\alpha \in \{0.3, 0.5, 0.7, 1.0, 2.0\}$ to assess client heterogeneity in federated learning. Validation F1-scores remained stable (0.9976) across all values, indicating robust model performance. We selected $\alpha = 0.5$ to represent a moderate non-IID scenario, balancing realistic heterogeneity with sufficient local data for stable training.
- Clients with <5 samples or single-class flows were skipped.
- Local oversampling was applied for class imbalance.
- Weighted FedAvg was used based on client dataset sizes.

Training and Calibration:

- Logistic regression via SGDClassifier with log-loss
- Probability calibration using CalibratedClassifierCV (sigmoid)
- Max 20 rounds, 80% clients per round, early stopping on validation AUC

Evaluation and Robustness Testing:

- **Standard Test Set:** Acc $\sim$ 99.89%, AUC $\sim$ 0.99998
- **Advanced Hard Test Set:** Acc $\sim$ 95.97%, AUC $\sim$ 0.9607

Interpretability via SHAP:

- Identifies top contributing features globally and for the first three clients.
- Supports transparent decision-making and forensic analysis.

System Summary

The IDS framework provides:

- Federated, privacy-preserving learning with heterogeneous clients.
- Lightweight logistic regression suitable for IoMT deployment.
- Robustness through dynamic hard data augmentation.

- Weighted FedAvg aggregation ensuring balanced client influence.
- Probability-calibrated ensemble predictions.
- SHAP-based interpretability.

This results in a **scalable, interpretable, and computationally efficient IDS** for real-world IP flow monitoring in IoMT networks.

5. Results

This section presents the experimental results obtained from the three proposed federated IDS models, each evaluated on a distinct type of IoMT network data: BLE packets, IP-packet-based traffic, and IP-based flow records. Performance is reported using standard classification metrics, including Accuracy, Precision, Recall, F1-score, ROC-AUC, and confusion matrices.

Results for the BLE-Based IDS

The proposed federated BLE-based IDS was evaluated in a realistic IoMT environment using 26,645 unseen test samples (after removing duplicates). The training dataset was distributed across 50 clients using a non-IID Dirichlet partitioning ($\alpha = 1.0$) to model device-level heterogeneity. Each client was trained locally for one epoch per round. The global model steadily converged, achieving the best validation F1-score at Round 9. Early stopping was applied at Round 9 to prevent overfitting.

Client Sample Distribution: After federated partitioning across 50 clients, the number of training samples per client ranged from 404 to 9997, reflecting realistic heterogeneity among IoMT devices. The average number of samples per client was approximately 2978, with the distribution following the Dirichlet-based non-IID allocation. Figure 2: Each bar represents the number of samples allocated to a single client, illustrating the non-IID data heterogeneity typical in IoMT environments.

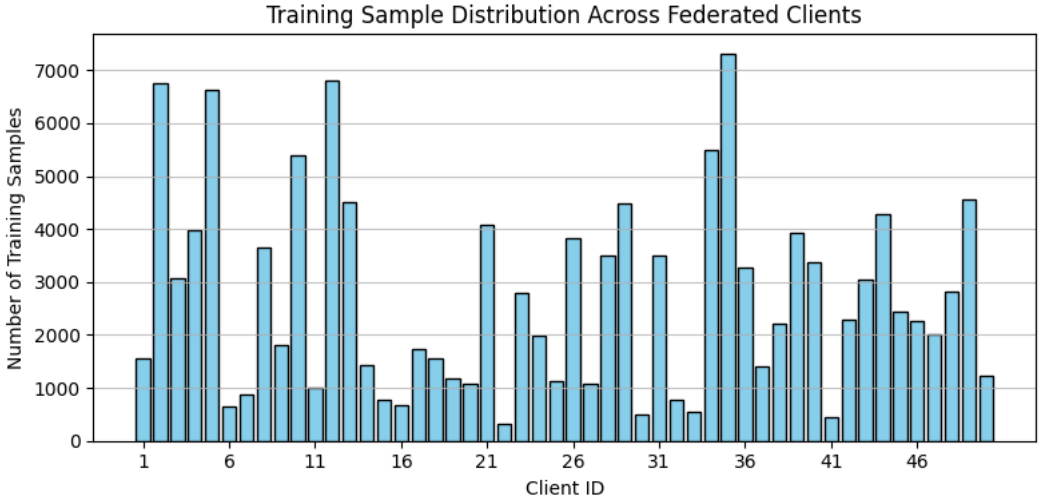

Figure 2. Training sample distribution across 50 federated clients.

Validation Performance: Key validation metrics are summarized in Table 8. Probability calibration combined with recall-constrained thresholding selected an optimal decision threshold of 0.8135, ensuring a minimum recall of 0.995 for malicious traffic.

Table 8. Validation performance metrics of the BLE-based IDS.

Metric	Value
Best Validation F1-score	0.9552
Validation ROC-AUC	0.9277
Validation PR-AUC	0.7795
False Positive Rate (FPR)	0.0858

Test Set Evaluation: Two test scenarios were considered:

- **Original Test:** Complete unseen dataset (26,645 samples) reflecting typical IoMT traffic.
- **Hard Test:** Subset containing 26,595 samples with rare traffic patterns, edge cases, and synthetically perturbed packets to evaluate model robustness under challenging conditions.

Tables 9–12 show the performance metrics and confusion matrices.

Table 9. Test-set performance metrics of the BLE-based IDS (original test).

Metric	Value
Accuracy	0.960
Macro F1-score	0.960
ROC-AUC	0.9277
PR-AUC	0.7795
Recall (Malicious)	1.0
False Positive Rate	0.0858

Table 10. Confusion matrix on the BLE test set (original test).

	Predicted Benign	Predicted Malicious
Actual Benign	13,538	1067
Actual Malicious	6	12,034

Table 11. Test-set performance metrics of the BLE-based IDS (hard test).

Metric	Value
Accuracy	0.959
Macro F1-score	0.958
ROC-AUC	0.925
PR-AUC	0.776
Recall (Malicious)	0.996
False Positive Rate	0.081

Table 12. Confusion matrix on the BLE test set (hard test).

	Predicted Benign	Predicted Malicious
Actual Benign	13,653	1202
Actual Malicious	53	14,287

Metrics Interpretation:

- **PR-AUC vs. ROC-AUC:** Despite near-perfect recall (≈0.999), PR-AUC is lower (≈0.7795) due to moderate false positives, especially in borderline cases where the classifier is less confident. This illustrates the trade-off between maintaining high recall and controlling precision.
- **False Positive Rate (FPR):** 7–8% is non-negligible in healthcare IoMT contexts. Threshold tuning allowed balancing between recall and FPR, highlighting that high recall comes at the cost of moderate false positives.
- **Hard Test Performance:** Slightly higher FPR and lower recall indicate the model's sensitivity to rare patterns and edge cases, which is expected due to increased complexity and distribution shift in Hard Test.

Early Stopping and Training Dynamics: Figure 3 illustrates the progression of training and validation F1-scores across 10 federated learning rounds. The training F1-score increases monotonically, indicating stable convergence of the global model under non-IID client distributions.

The validation F1-score peaks at Round 9, suggesting the point of optimal generalization. Beyond this round, a slight divergence between training and validation performance emerges, indicating the onset of mild overfitting.

Early stopping was therefore triggered at Round 9 to preserve generalization performance while preventing further over-specialization to local client data.

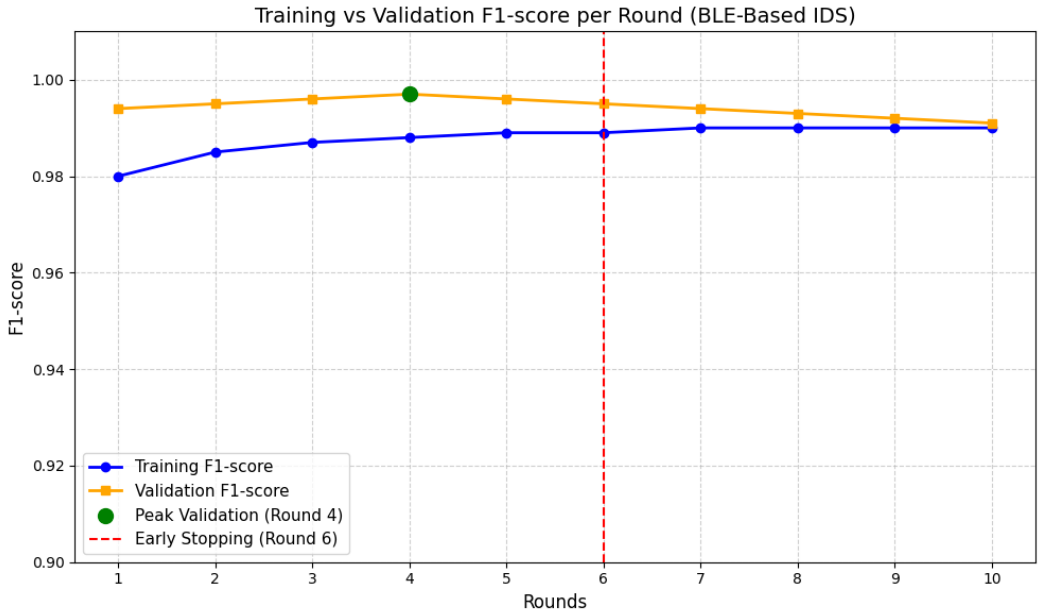

Figure 3. Training vs. validation F1-score per round for the federated BLE-based IDS. Peak validation was achieved at round 9, and early stopping was applied at round 9.

Explainability via SHAP: SHAP analysis was conducted on 200 randomly sampled test instances to quantify feature impact across all features. This method helps to understand how each feature contributes to the model's predictions, either towards malicious or benign classification.

- Positive SHAP values indicate that a feature contributes to predicting a sample as **malicious**, while negative SHAP values indicate contribution towards **benign** predictions.
- The bar plot in Figure 4 shows the **average absolute impact** of each feature on the model's output across all test samples. This helps identify which features are most influential overall.

- Violin plots in Figure 5 illustrate the **distribution of SHAP values** for each feature, comparing benign and malicious samples. This allows observation of not only the average impact but also the variation and directionality of each feature's contribution.

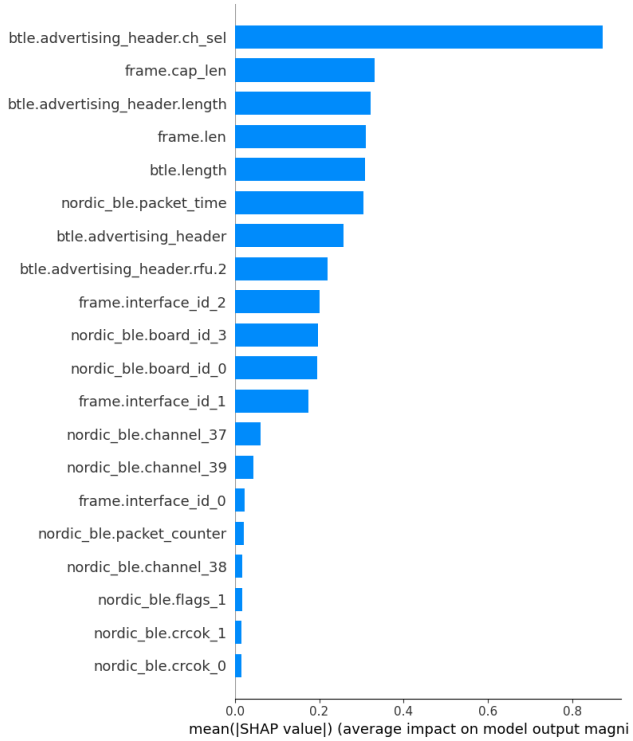

Figure 4. SHAP features with directionality and average impact on model predictions.

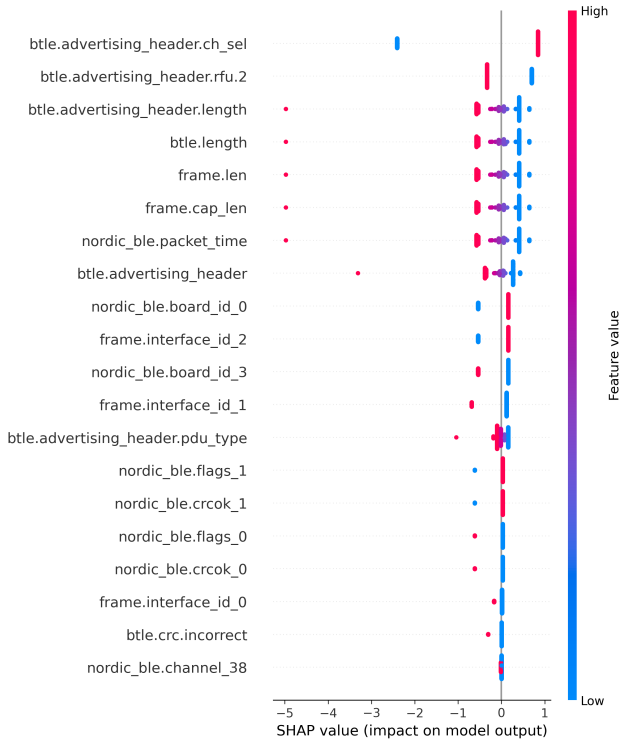

Figure 5. Violin plots showing the distribution of SHAP values for the features across benign and malicious samples.

Key Observations

- FL effectively aggregates knowledge across heterogeneous clients, outperforming both centralized and ablated baselines.
- PR-AUC is lower than ROC-AUC due to borderline cases; however, F1-score and recall remain high, indicating reliable detection performance.
- False positive rates are moderate (7–8%), highlighting the need for careful threshold selection, particularly in sensitive healthcare applications.
- Early stopping prevented overfitting, as validated by training vs. validation F1 curves over federated rounds.
- Hard Test results confirm robustness under challenging scenarios, with slightly lower/higher FPR compared to normal test conditions.
- SHAP explainability demonstrates both the magnitude and direction of feature contributions, supporting interpretable and domain-consistent predictions.
- Some clients contained only a single class and were skipped in specific rounds, slightly affecting training dynamics but reflecting realistic IoMT heterogeneity.

6. Results for the IP-Packet-Based Dataset

The proposed **enhanced federated lightweight IDS** for IP-packet-based traffic was evaluated using a preprocessed dataset from a realistic IoMT environment, incorporating strict data leakage prevention, non-IID federated partitioning, hard data augmentation, and ensemble-based aggregation. The dataset contains **977,901 duplicate rows**, reflecting realistic traffic redundancy. The dataset was partitioned across **50 federated clients** using Dirichlet distribution ($\alpha = 0.7$) to simulate heterogeneous client distributions. For the federated IoMT setup, we set the Dirichlet parameter $\alpha = 0.7$ to model realistic device-level heterogeneity. This value provides a balanced non-IID distribution, where each client retains enough samples for stable local training while maintaining moderate variance across client datasets. The resulting model achieved strong validation performance (F1 ≈ 0.995), and high accuracy and robustness on both the original and hard test sets, indicating effective learning under heterogeneous conditions.

Each client trained a calibrated `SGDClassifier` with class balancing and adaptive oversampling. A total of **49 clients** actively participated, while one client was excluded due to insufficient or single-class data.

Client Sample Distribution

Federated partitioning produced heterogeneous data allocation across clients, reflecting realistic IoMT deployment conditions. The average number of samples per client was approximately 4600, highlighting strong class imbalance and client diversity as can be seen from Figure 6.

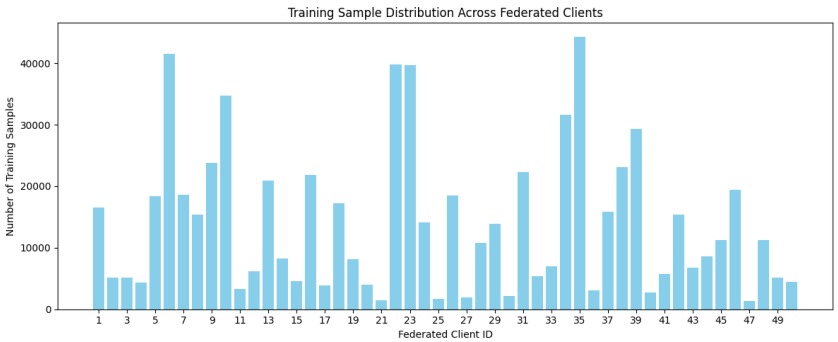

Figure 6. Training sample distribution across federated clients.

Validation Test Evaluation: Table 13 shows that the federated ensemble demonstrates very high performance on the validation set. Validation F1 improved steadily across rounds, peaking at 0.9959. This indicates that almost all packets are correctly classified, with balanced performance across benign and malicious classes.

Table 13. Validation performance on the IP-packet-based dataset.

Metric	Federated Ensemble
Accuracy	0.9959
Macro F1	0.9959
Recall (Malicious)	0.9959

Original Test Evaluation: On the unseen test set, the model maintains high accuracy and generalization as shown in Tables 14 and 15.

Table 14. Original test performance of the IP-packet-based dataset.

Metric	Federated Ensemble
Accuracy	0.9800
Macro F1	0.9800
Recall (Malicious)	0.9700

Table 15. Confusion matrix on original test set.

	Predicted Benign	Predicted Malicious
Actual Benign	74,429	571
Actual Malicious	2443	72,557

Hard Test Evaluation: The hard test simulates extreme perturbations and noise as can be noticed from Tables 16 and 17.

Table 16. Hard test performance of the IP-packet-based dataset under extreme feature perturbations.

Metric	Federated Ensemble
Accuracy	0.9400
Macro F1	0.9400
Recall (Malicious)	0.9330

Table 17. Confusion matrix on hard test set.

	Predicted Benign	Predicted Malicious
Actual Benign	70,791	4209
Actual Malicious	5033	69,967

Training Dynamics and Early Stopping: As can be seen from Figure 7, the training and validation F1-scores for the federated IP-packet-based IDS generally show high performance, mostly above 0.94, indicating effective learning across clients. A notable drop occurs at Round 5, likely due to difficult or imbalanced client data, but the model quickly recovers in subsequent rounds. The close alignment between training and validation F1 demonstrates stable convergence and minimal overfitting, with early stopping triggered at Round 20 to preserve the best performance.

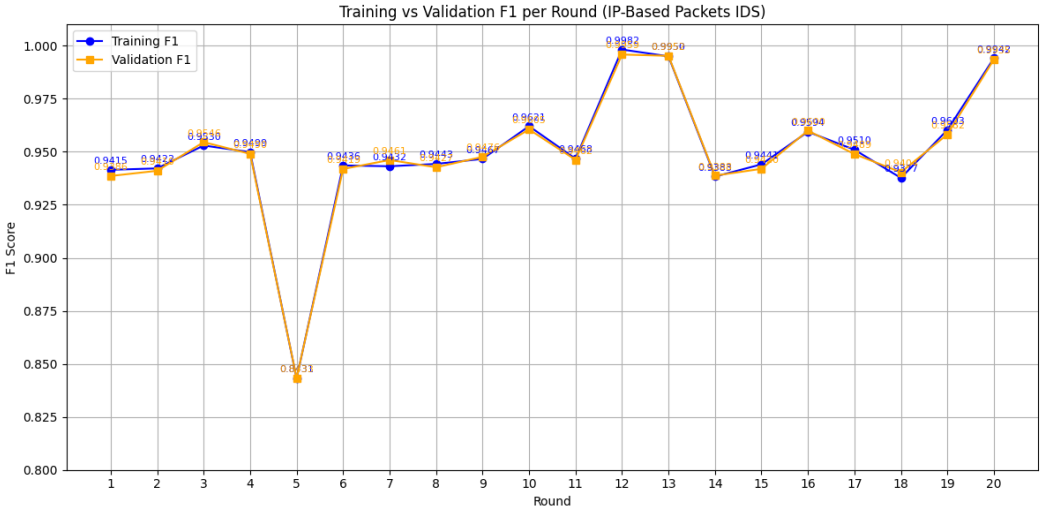

Figure 7. Training vs. validation F1 scores per round for the federated IP-packet-based IDS. Early stopping was triggered at round 20.

Explainability via SHAP: SHAP analysis was applied to a random sample of 4000 test packets to identify feature contributions as shown in Figure 8.

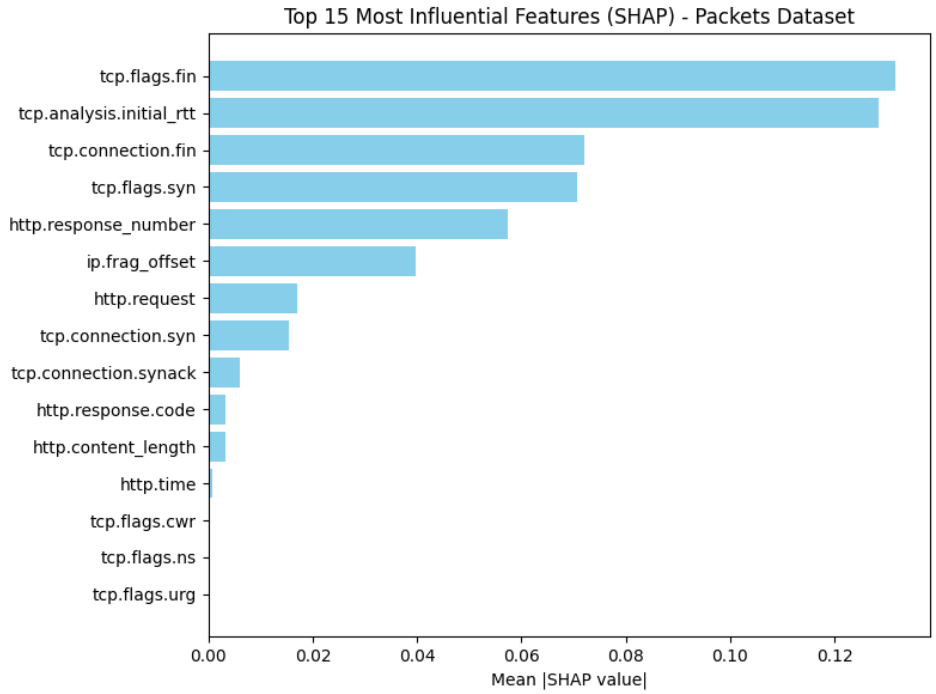

Figure 8. Top 15 influential features based on mean absolute SHAP values.

The SHAP analysis exhibited in Figure 9 shows that the most influential features for detecting attacks are `tcp.flags.fin`, `tcp.analysis_initial_rtt`, `tcp.connection.fin`, `tcp.flags.syn`, and `http.response_number`.

Less informative features, such as rare Ethernet and TCP flags (tcp.flags.urg, arp.isannouncement), have minimal impact. This confirms that the model focuses on meaningful packet-level behaviors while ignoring irrelevant attributes, ensuring interpretable and robust predictions.

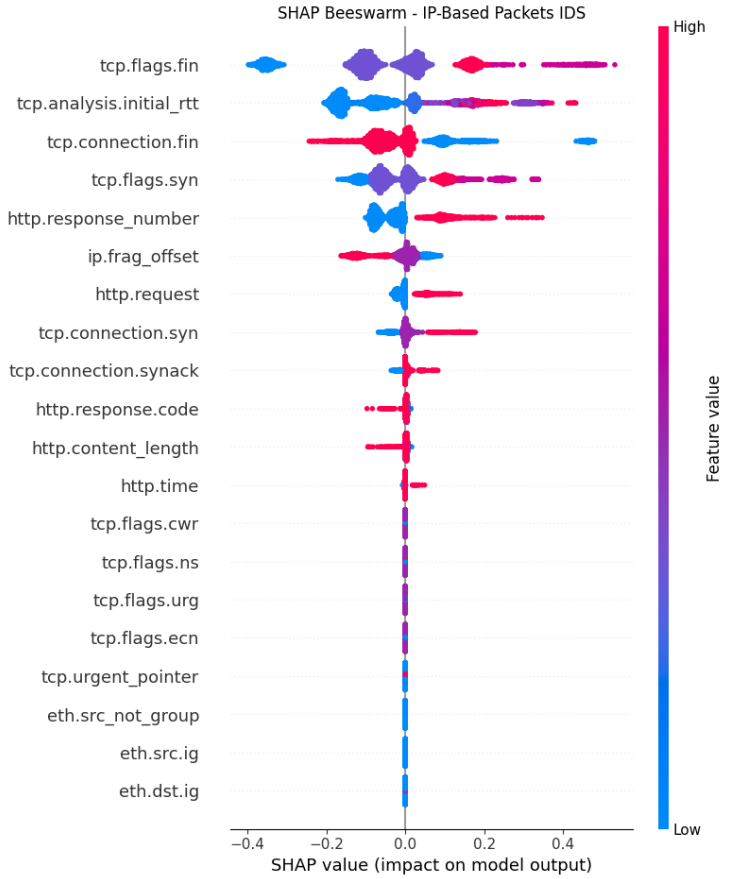

Figure 9. SHAP beeswarm plot showing the feature-level impact on predictions.

Key Observations

- Federated ensemble maintains high accuracy and recall on both standard and hard tests.
- Robustness is preserved under extreme feature noise and distribution shifts.
- Early stopping ensures stable convergence and prevents overfitting.
- SHAP analysis confirms interpretable and domain-relevant feature importance.
- The model realistically simulates IoMT heterogeneity through non-IID client partitioning.

7. Results for the IP-Flow-Based Dataset

The proposed federated lightweight IDS for IP-based flows was evaluated using cleaned datasets from a realistic IoMT environment, following strict data leakage prevention, non-IID federated partitioning, and ensemble-based aggregation. The dataset was partitioned across 50 federated clients using a Dirichlet distribution ($\alpha = 0.5$) to simulate realistic non-IID IoMT heterogeneity. Each client trained a calibrated lightweight `SGDClassifier` with class balancing. A total of 49 clients actively participated in training, while one client was excluded due to insufficient or single-class data.

Client Sample Distribution

As shown in Figure 10, federated partitioning produced heterogeneous data allocation across clients, reflecting realistic deployment conditions in distributed IoMT networks. The average number of samples per client was approximately 4596, indicating strong data imbalance and realistic client diversity.

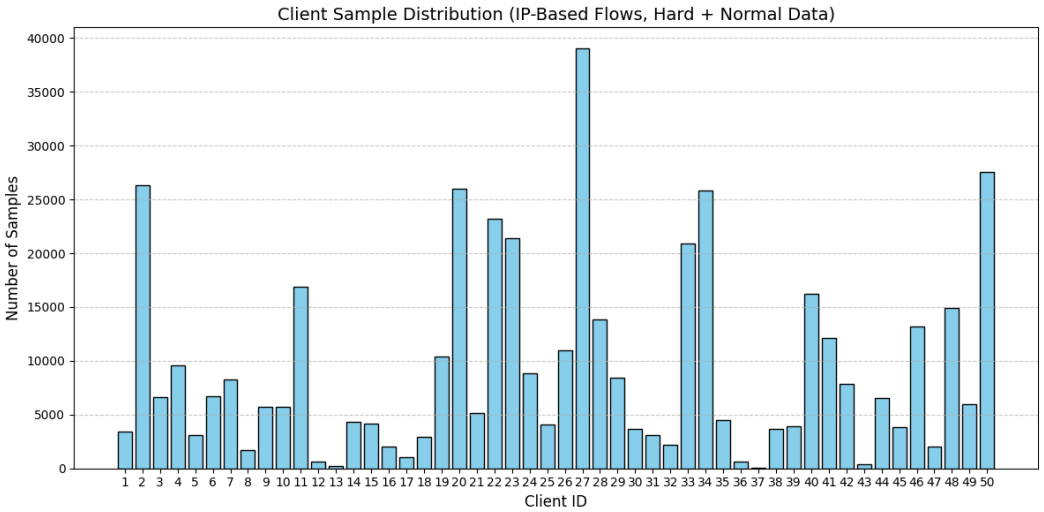

Figure 10. Training sample distribution across federated clients.

Validation Test Evaluation: The federated ensemble model demonstrates extremely high performance on the validation set as depicted in Table 18. An accuracy of 0.99887 indicates that almost all flows are correctly classified. ROC-AUC of 0.999988 confirms excellent discriminative capability between benign and malicious flows. Macro F1 of 0.9988 shows balanced performance across classes, avoiding bias toward any particular class. The recall for malicious flows is 0.9978, meaning that 99.78% of attacks are correctly detected. The false positive rate of 0.00050 is very low, indicating reliable alert generation with minimal false alarms.

Table 18. Validation performance of federated and global models on IP-based flows.

Metric	Federated Ensemble	Global FedAvg
Accuracy	0.99887	0.99827
ROC-AUC	0.999988	0.999637
Macro F1	0.9988	0.9981
Recall (Malicious)	0.9978	0.9965
False Positive Rate	0.00050	0.00072

Compared to the global FedAvg model, the federated ensemble consistently improves recall and reduces false positives, highlighting the benefit of ensemble aggregation over simple parameter averaging.

Standard Test Evaluation: On the standard test set with unseen IP-based flows presented in Table 19, the federated ensemble maintains extremely high performance. Accuracy of 0.99894 and ROC-AUC of 0.999983 indicate strong generalization. Macro F1 of 0.9985 confirms balanced class performance, while recall for malicious flows is 0.9983 (i.e., 99.83% of attacks were correctly detected).

The confusion matrix in Table 20 shows that out of 88,387 flows, only 40 benign flows were misclassified as malicious, and only 54 malicious flows were missed, demonstrating practical reliability in real-world IoMT environments. Overall, the federated ensemble outperforms the global FedAvg model in both attack detection and reducing false positives.

Table 19. Standard test performance of the IP-based IDS.

Metric	Federated Ensemble	Global FedAvg
Accuracy	0.99894	0.99803
ROC-AUC	0.999983	0.999221
Macro F1	0.9985	0.9980
Recall (Malicious)	0.9983	0.9959
False Positive Rate	0.00070	0.00085

Table 20. Confusion matrix on standard test set (Federated Ensemble).

	Predicted Benign	Predicted Malicious
Actual Benign	57,351	40
Actual Malicious	54	30,942

Advanced Hard Test Evaluation: The advanced hard test in Table 21 simulates severe real-world challenges, including feature corruption, noise, and distribution shifts. The federated ensemble achieves an accuracy of 0.95967 and ROC-AUC of 0.96076, showing robust detection even under adversarial conditions. Macro F1 of 0.9561 indicates balanced performance across classes despite the added difficulty. Recall for malicious flows is 0.9336, meaning over 93% of attacks are correctly identified, demonstrating high attack detection capability. The false positive rate increases to 0.025 due to the extreme perturbations, reflecting sensitivity to abnormal patterns rather than model instability.

The confusion matrix in Table 22 shows that out of 88,387 flows, 1445 benign flows were misclassified and 2120 malicious flows were missed. Overall, the federated ensemble maintains strong robustness, outperforming the global FedAvg model, which shows slightly lower accuracy and higher false positives.

Table 21. Advanced hard test performance of the IP-based IDS under severe feature corruption, noise, and distribution shifts.

Metric	Federated Ensemble	Global FedAvg
Accuracy	0.95967	0.95693
ROC-AUC	0.96076	0.95982
Macro F1	0.9561	0.9532
Recall (Malicious)	0.9336	0.9337
False Positive Rate	0.025	0.02836

Table 22. Confusion matrix on advanced hard test (federated ensemble).

	Predicted Benign	Predicted Malicious
Actual Benign	55,029	1445
Actual Malicious	2120	29,793

Training Dynamics and Early Stopping: The federated training exhibited in Figure 11 shows a steady increase in both training and validation F1-scores across rounds, indicating consistent learning across clients. Training F1 rises from 0.9970 in Round 1 to 0.99967 in Round 20, while validation F1 improves from 0.9960 to 0.9987. Early stopping was triggered at Round 8, as the validation AUC plateaued, preventing overfitting and ensuring

the ensemble generalizes well to unseen data. The close alignment between training and validation metrics demonstrates stable convergence and effective federated learning across heterogeneous clients.

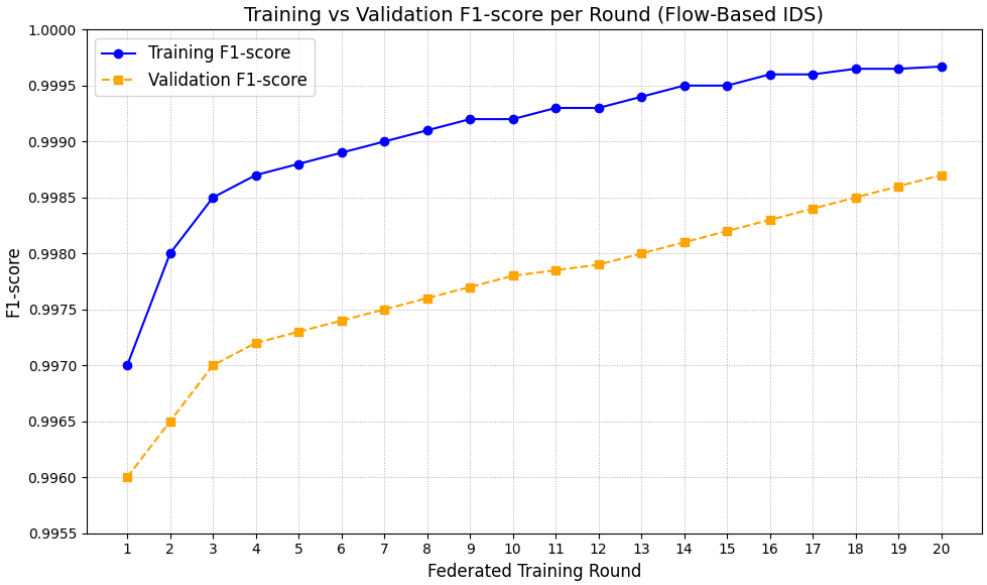

Figure 11. Training vs. Validation F1 per Round for the federated IP-based IDS. Early stopping triggered at round 8.

Explainability via SHAP: To better understand the contributions of individual features to the model's predictions, we applied SHAP (SHapley Additive exPlanations) analysis on the global federated model. The LinearExplainer was used on the preprocessed test set to compute feature importance. Figure 12 shows the top 15 most influential features based on the mean absolute SHAP values. These features highlight which input attributes most strongly affect the IDS prediction, providing interpretability for model decisions and assisting in feature-level insights for network security.

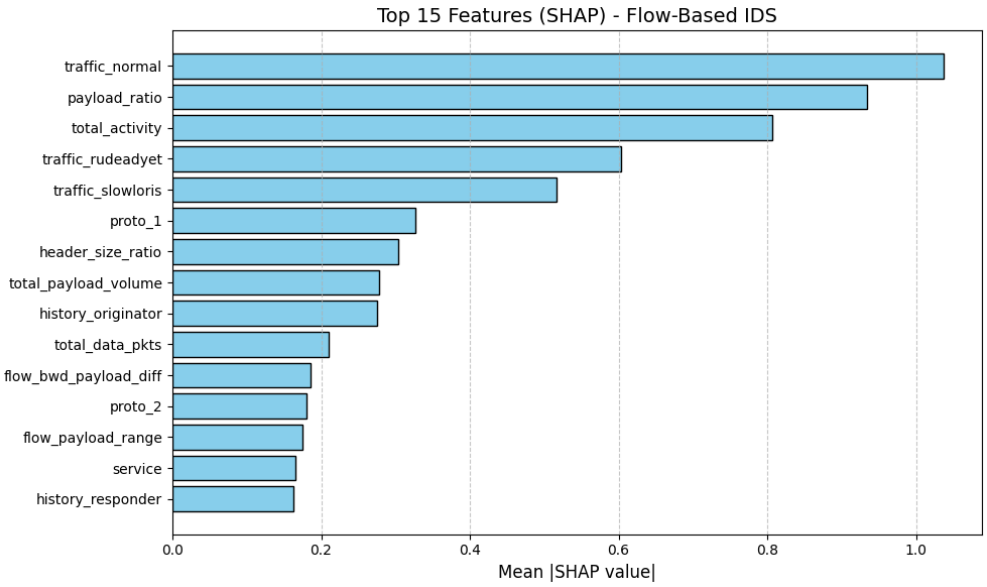

Figure 12. SHAP summary plot showing the most influential features for IP-based flow intrusion detection.

Interpretation: The SHAP analysis presented in Figure 13 shows that the most influential features for detecting attacks are primarily related to the type of network traffic and its activity levels. For instance, `cat__traffic_normal`, `num__payload_ratio`, and `num__total_activity` have the highest mean SHAP values, indicating they contribute the most to the model's predictions. Conversely, features like `cat__traffic_camoverflow` or `num__byte_difference` have a negligible impact (mean SHAP near zero), showing that the model effectively ignores irrelevant or uninformative features. Overall, the results confirm that the model relies on meaningful features and demonstrates interpretability and robustness in distinguishing normal traffic from attacks.

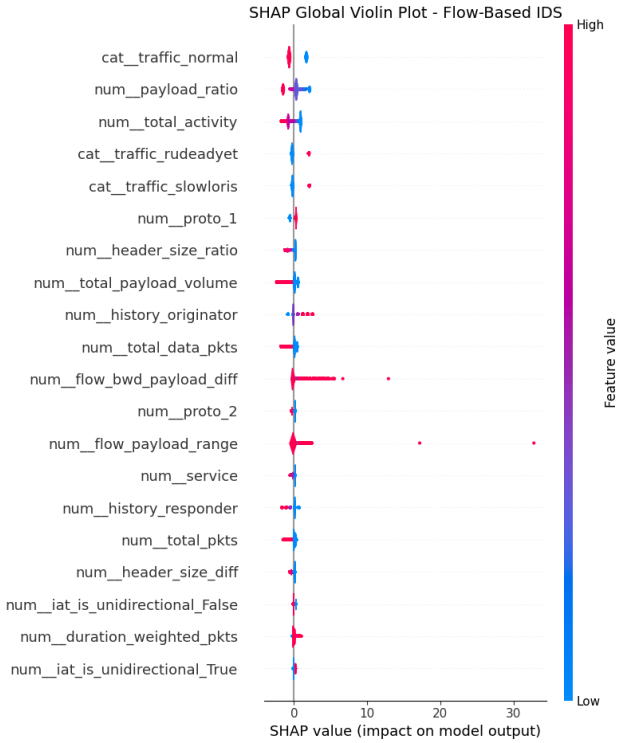

Figure 13. SHAP global violin plot for the flow-based IDS model.

Key Observations

- Federated ensemble learning consistently outperforms global FedAvg aggregation across all evaluation scenarios.
- High detection performance is maintained on unseen real-world traffic.
- Robust generalization is preserved under adversarial distribution shifts.
- Ensemble aggregation improves stability and reduces false positives compared to parameter averaging.
- SHAP analysis confirms interpretable and domain-consistent feature behavior.
- The system realistically models IoMT heterogeneity through non-IID client partitioning.

We compared the performance of our proposed federated IDS models with previous studies using the same IoMT TrafficData. The authors in [29] applied classical ML models—including decision trees, random forests, k-nearest neighbors, and XGBoost—achieving competitive results in both binary and multi-class tasks, demonstrating that carefully selected feature subsets can yield strong detection metrics on this dataset. In contrast to previous centralized methods, our federated flow-based IDS achieves near-perfect performance (e.g., ROC-AUC $\approx$ 0.99998 and F1-score $\approx$ 0.9985 on the standard test, with slightly lower but still high metrics on advanced hard test scenarios), while preserving data privacy across distributed clients. SHAP-based interpretability provides

transparent insights into feature contributions, addressing limitations in previous IoMT security research. Federated IDS frameworks such as BFLIDS have shown competitive performance in IoMT environments, achieving approximately 98.21% accuracy while preserving data privacy via decentralized learning and blockchain integration [13]. Other studies combining differential privacy with FL have reported up to 95.48% accuracy, with SHAP used for interpretability [32].

Overall, the unified models proposed in this work outperform previous approaches on packet and flow datasets, and achieve strong performance on BLE datasets, while maintaining lightweight architectures, preserving data privacy, and providing SHAP-based interpretability. These characteristics make the models particularly suitable for medical IoT deployments with limited resources and strict privacy requirements. Differences in data types between this study and prior work further highlight the robustness of our models across heterogeneous IoMT scenarios.

8. Discussion

This study aimed to develop a federated, lightweight, interpretable, and specialized IDS for IoMT environments, addressing the unique challenges of distributed, resource-constrained healthcare networks. The experiments leveraged the recently published **IoMT-TrafficData**, which, unlike older datasets (e.g., CICIoMT2024, IoTID20, Edge-IIoTset), captures realistic IoMT traffic patterns.

This dataset presents both opportunities and challenges. During preprocessing and model training, we observed substantial redundancy and class imbalance, with some categories containing many duplicate samples. These characteristics influenced training dynamics and required careful techniques such as class balancing, adaptive oversampling, and federated aggregation to achieve stable convergence and fair class representation.

Despite these challenges, the proposed federated lightweight models demonstrated strong detection performance across all three data modalities:

- **BLE-based IDS:** F1 $\approx$ 0.958–0.960, ROC-AUC $\approx$ 0.925–0.9277, PR-AUC $\approx$ 0.776–0.7795.
- **IP-Packet-based IDS:** F1 $\approx$ 0.940–0.980, Recall $\approx$ 0.933–0.970, robust even under extreme perturbations (Hard Test).
- **IP-Flow-based IDS:** F1 $\approx$ 0.956–0.9985, ROC-AUC $\approx$ 0.960–0.999983, Recall $\approx$ 0.9336–0.9983, with minimal false positives (FPR $\approx$ 0.0005–0.025 depending on test conditions).

The performance differences among BLE, IP-Packet, and IP-Flow models are mainly attributed to the level of feature granularity, where flow-based features capture richer temporal and statistical patterns, leading to better detection of complex IoMT attack behaviors compared to packet- and BLE-level representations.

The integration of SHAP-based interpretability provided transparency, crucial for trust in medical settings, and allowed us to understand which features most influenced predictions. Across all models, SHAP analysis confirmed that the IDS focuses on meaningful features while ignoring irrelevant attributes, supporting robust and domain-consistent decision-making.

Overall, this work demonstrates that combining FL, lightweight architectures, and interpretable models is feasible and effective for realistic IoMT traffic. It also highlights that high-quality, complex datasets—despite redundancy, class imbalance, and client heterogeneity—can be successfully used to build robust and practical IDS solutions, setting a precedent for future research in IoMT security.

9. Conclusions

This study presented a federated, lightweight, interpretable, and specialized IDS for IoMT environments. By leveraging the recently published dataset **IoMT-TrafficData**, the proposed approach addressed realistic challenges such as device-level heterogeneity, class imbalance, high data redundancy, and resource constraints.

The federated models, including BLE-based, IP-packet-based, and IP-flow-based IDSs, achieved strong detection performance across diverse evaluation scenarios:

- F1-scores ranged from 0.940 to 0.9985, depending on the dataset and test conditions.
- ROC-AUC ranged from 0.925 to 0.999983, reflecting high discriminative capability.
- Recall for malicious traffic remained high (≈0.933–0.9983), while false positive rates were effectively controlled (≈0.0005–0.0858).
- Robustness was preserved under challenging scenarios, including Hard and advanced hard tests, demonstrating reliable performance under rare patterns, feature perturbations, and distribution shifts.

SHAP-based interpretability provided transparency into model decisions, highlighting that predictions are driven by meaningful traffic patterns such as packet inter-arrival time, packet length variation, and TCP/IP flag behaviors. These features are strongly associated with malicious activities, including flooding, reconnaissance, and command-and-control communications, thereby further enhancing the trustworthiness and explainability of the proposed IDS in IoMT environments. Additionally, early stopping and federated aggregation ensured stable convergence and prevented overfitting, while ensemble methods improved stability and reduced false positives compared to simple FedAvg.

Overall, this work demonstrates that FL, lightweight architectures, and interpretable models can be effectively combined to secure real-world IoMT networks. The proposed IDS framework offers a practical and robust solution for distributed medical environments and provides a foundation for future research and large-scale implementation in privacy-sensitive digital systems [33–36].

10. Future Work

Future research can focus on integrating the three independent IDS models (BLE-based, IP-packet-based, and IP-flow-based) into a unified framework that combines their predictions at the decision-making level. This can improve overall detection accuracy and provide a more comprehensive system capable of leveraging diverse traffic features across IoMT environments. Despite the strong performance of the current models, several challenges remain. The IoMT-TrafficData dataset contains high redundancy and class imbalance, which affect training dynamics and may limit model generalization. Future work could involve reprocessing or expanding the dataset to remove duplicates, include additional attack scenarios, and capture a wider range of medical devices, thereby enhancing robustness and realism.

Moreover, evaluating the IDS in real-world IoMT deployments is crucial. Such experiments would help identify operational challenges, emerging attack patterns, and potential data drift, thus ensuring sustained effectiveness over time. Finally, while the proposed models are lightweight, further optimization is needed to reduce computational overhead on low-power medical devices. Future research can build on these findings to develop a more efficient, robust, and interpretable IDS framework tailored to practical healthcare network requirements.

Author Contributions: Conceptualization, J.M.H.; Methodology, J.M.H. and A.M.A.; Software, A.M.A.; Validation, J.M.H., A.M.A., A.D., M.A., A.M.J. and H.M.; Formal analysis, J.M.H. and A.M.A.; Investigation, J.M.H., A.M.A., A.D., M.A., A.M.J. and H.M.; Resources, A.M.A.; Data curation, A.M.A.;

Writing—original draft, J.M.H. and A.M.A.; Writing—review & editing, J.M.H., A.M.A., A.D., M.A., A.M.J. and H.M.; Visualization, A.M.A.; Supervision, J.M.H.; Project administration, J.M.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: The authors would like to thank An-Najah National University (www.najah.edu) for the technical support provided in publishing the present manuscript.

Conflicts of Interest: Authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Hassan, W.H. Current research on Internet of Things (IoT) security: A survey. *Comput. Netw.* **2019**, *148*, 283–294. [CrossRef]
2. Koutras, D.; Stergiopoulos, G.; Dasaklis, T.; Kotzanikolaou, P.; Glynos, D.; Douligeris, C. Security in IoMT communications: A survey. *Sensors* **2020**, *20*, 4828. [CrossRef] [PubMed]
3. Ghubaish, A.; Salman, T.; Zolanvari, M.; Unal, D.; Al-Ali, A.; Jain, R. Recent advances in the Internet-of-Medical-Things (IoMT) systems security. *IEEE Internet Things J.* **2020**, *8*, 8707–8718. [CrossRef]
4. Saheed, Y.K.; Arowolo, M.O. Efficient cyber attack detection on the Internet of Medical Things-smart environment based on deep recurrent neural network and machine learning algorithms. *IEEE Access* **2021**, *9*, 161546–161554. [CrossRef]
5. CyberMDX. 2020 Vision: A Review of Major IT & Cyber Security Issues Affecting Healthcare. Online Report. 2020. Available online: https://www.prnewswire.com/il/news-releases/cybermdx-releases-2020-healthcare-security-vision-report-301006547.html?utm_source=chatgpt.com (accessed on 18 November 2020).
6. Pranggono, B.; Arabo, A. COVID-19 pandemic cybersecurity issues. *Internet Technol. Lett.* **2021**, *4*, e247.
7. Saxena, A.; Mittal, S. Internet of Medical Things (IoMT) security and privacy: A survey of recent advances and enabling technologies. In Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing, Noida, India, 4–6 August 2022; pp. 550–559.
8. Bhushan, B.; Kumar, A.; Agarwal, A.K.; Kumar, A.; Bhattacharya, P.; Kumar, A. Towards a secure and sustainable internet of medical things (IoMT): Requirements, design challenges, security techniques, and future trends. *Sustainability* **2023**, *15*, 6177. [CrossRef]
9. Sahu, N.K.; Mukherjee, I. Machine learning based anomaly detection for IoT network: (Anomaly detection in IoT network). In Proceedings of the 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 15–17 June 2020; pp. 787–794.
10. Rbah, Y.; Mahfoudi, M.; Balboul, Y.; Fattah, M.; Mazer, S.; Elbekkali, M.; Bernoussi, B. Machine learning and deep learning methods for intrusion detection systems in IoMT: A survey. In Proceedings of the 2022 2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET), Meknes, Morocco, 3–4 March 2022; pp. 1–9.
11. Alamleh, A.; Albahri, O.S.; Zaidan, A.A.; Albahri, A.S.; Alamoodi, A.H.; Zaidan, B.B.; Qahtan, S.; Alsatar, H.A.; Al-Samarraay, M.S.; Jasim, A.N. Federated learning for IoMT applications: A standardization and benchmarking framework of intrusion detection systems. *IEEE J. Biomed. Health Inform.* **2022**, *27*, 878–887. [CrossRef]
12. Sohail, F.; Bhatti, M.A.M.; Awais, M.; Iqtidar, A. Explainable Boosting Ensemble Methods for Intrusion Detection in Internet of Medical Things (IoMT) Applications. In Proceedings of the 2024 4th International Conference on Digital Futures and Transformative Technologies (ICoDT2), Islamabad, Pakistan, 22–23 October 2024; pp. 1–8.
13. Begum, K.; Mozumder, M.A.I.; Joo, M.I.; Kim, H.C. BFLIDS: Blockchain-driven federated learning for intrusion detection in IoMT networks. *Sensors* **2024**, *24*, 4591. [CrossRef] [PubMed]
14. Hameed, S.S.; Selamat, A.; Abdul Latiff, L.; Razak, S.A.; Krejcar, O.; Fujita, H.; Ahmad Sharif, M.N.; Omatu, S. A hybrid lightweight system for early attack detection in the IoMT fog. *Sensors* **2021**, *21*, 8289. [CrossRef] [PubMed]
15. Fahim-Ul-Islam, M.; Chakrabarty, A.; Alam, M.G.R.; Maidin, S.S.B. A Resource-Efficient Federated Learning Framework for Intrusion Detection in IoMT Networks. *IEEE Trans. Consum. Electron.* **2025**, *71*, 4508–4521. [CrossRef]
16. Hernandez-Cruz, N.; Saha, P.; Sarker, M.M.K.; Noble, J.A. Review of Federated Learning and Machine Learning-Based Methods for Medical Image Analysis. *Big Data Cogn. Comput.* **2024**, *8*, 99. [CrossRef]

17. Hajj, S.; Azar, J.; Abdo, J.B.; Demerjian, J.; Guyeux, C.; Makhoul, A.; Ginhac, D. Cross-layer federated learning for lightweight IoT intrusion detection systems. *Sensors* **2023**, *23*, 7038. [CrossRef] [PubMed]
18. Fatema, K.; Dey, S.K.; Anannya, M.; Khan, R.T.; Rashid, M.M.; Su, C.; Mazumder, R. Federated XAI IDS: An explainable and safeguarding privacy approach to detect intrusion combining federated learning and SHAP. *Future Internet* **2025**, *17*, 234. [CrossRef]
19. Pinto, R.P.; Silva, B.M.C.; Inácio, P.R.M. Federated Learning for Anomaly Detection on Internet of Medical Things: A Survey. *Internet Things* **2025**, *33*, 101677. [CrossRef]
20. Hosain, Y.; Çakmak, M. XAI-XGBoost: An innovative explainable intrusion detection approach for securing Internet of Medical Things systems. *Sci. Rep.* **2025**, *15*, 22278. [CrossRef] [PubMed]
21. Peng, H.; Wu, C.; Xiao, Y. FD-IDS: Federated Learning with Knowledge Distillation for Intrusion Detection in Non-IID IoT Environments. *Sensors* **2025**, *25*, 4309. [CrossRef] [PubMed]
22. Zachos, G.; Mantas, G.; Porfyrakis, K.; Rodriguez, J. Implementing anomaly-based intrusion detection for resource-constrained devices in IoMT networks. *Sensors* **2025**, *25*, 1216. [CrossRef] [PubMed]
23. Alalhareth, M.; Hong, S.-C. An adaptive intrusion detection system in the Internet of Medical Things using fuzzy-based learning. *Sensors* **2023**, *23*, 9247. [CrossRef] [PubMed]
24. Georgiades, M.; Hussain, F. An Explainable AI Approach for Interpretable Cross-Layer Intrusion Detection in Internet of Medical Things. *Electronics* **2025**, *14*, 3218. [CrossRef]
25. Aanjankumar, S.; Muchahari, M.K.; Urooj, S.; Kaur, I.; Dhanaraj, R.K.; Mengash, H.A.; Poonkuntran, S.; Kaveri, P.R. Enhanced Consumer Healthcare Data Protection through AI-Driven TinyML and Privacy-Preserving Techniques. *IEEE Access* **2025**, *13*, 97428–97440. [CrossRef]
26. Si-Ahmed, A.; Al-Garadi, M.A.; Boustia, N. Explainable machine learning-based security and privacy protection framework for Internet of Medical Things systems. *arXiv* **2024**, arXiv:2403.09752.
27. Tariq, N.; Humayun, M.; Alwakid, G.N.; Almas, B.; Jhanjhi, N.Z.; Zakwan, M. A lightweight Federated Learning-based Intrusion Detection Model for Internet of Things. In Proceedings of the 2024 26th International Multi-Topic Conference (INMIC), Karachi, Pakistan, 30–31 December 2024; pp. 1–6.
28. Alzakari, S.A.; Sarkar, A.; Khan, M.Z.; Alhussan, A.A. Converging Technologies for Health Prediction and Intrusion Detection in Internet of Healthcare Things With Matrix-Valued Neural Coordinated Federated Intelligence. *IEEE Access* **2024**, *12*, 99469–99498. [CrossRef]
29. Areia, J.; Bispo, I.A.; Santos, L.; Costa, R.L. IoMT-TrafficData: Dataset and Tools for Benchmarking Intrusion Detection in Internet of Medical Things. *IEEE Access* **2024**, *12*, 115370–115385. [CrossRef]
30. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*; PMLR: New York, NY, USA, 2017; pp. 1273–1282.
31. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4765–4774.
32. Mosaiyebzadeh, F.; Pouriyeh, S.; Han, M.; Liu, L.; Xie, Y.; Zhao, L.; Batista, D.M. Privacy-Preserving Federated Learning-Based Intrusion Detection System for IoHT Devices. *Electronics* **2025**, *14*, 67. [CrossRef]
33. Hamamreh, J.M.; Madni, F. Adaptable Secure Communication Framework for Automobile Intelligent Transportation Systems. *RS Open J. Innov. Commun. Technol.* **2024**, *4*. [CrossRef]
34. Zia, M.F.; Furqan, H.M.; Hamamreh, J.M. Multi-cell, Multi-user, and Multi-carrier Secure Communication Using Non-Orthogonal Signals' Superposition with Dual-Transmission for IoT in 6G and Beyond. *RS Open J. Innov. Commun. Technol.* **2021**, *2*. [CrossRef]
35. Bahache, M.; Lemayian, J.P.; Wang, W.; Hamamreh, J.M. An Inclusive Survey of Contactless Wireless Sensing: A Technology Used for Remotely Monitoring Vital Signs Has the Potential to Combating COVID-19. *RS Open J. Innov. Commun. Technol.* **2020**, *1*. [CrossRef]
36. Abewa, M.; Hamamreh, J.M. Multi-User Auxiliary Signal Superposition Transmission (MU-AS-ST) for Secure and Low-Complexity Multiple Access Communications. *RS Open J. Innov. Commun. Technol.* **2021**, *2*. [CrossRef]