Contents lists available at ScienceDirect

Internet of Things

journal homepage: www.elsevier.com/locate/iot

FTL-TSLP: A federated transfer learning approach with a two-stage LSTM pipeline for fault-tolerant and privacy-preserving intrusion detection in IoMT networks

Abdelhammid Bouazza [a,*], Hichem Debbi [a], Hicham Lakhlef [b]

[a] *Department of Computer Science, Laboratory of Informatics and its Applications of M'sila (LIAM), University of M'sila, PO Box 166, Ichebilia, 28000, M'sila, Algeria*
[b] *CNRS-LaBRI, University of Bordeaux / Bordeaux INP, P.O. Box 99, Talence Cedex, 33400, France*

ARTICLE INFO

Keywords:
Fault tolerance
Federated learning
Transfer learning
Two-stage LSTM pipeline
IoMT security
Non-IID data

ABSTRACT

The rapid proliferation of the Internet of Medical Things (IoMT) has transformed healthcare delivery by enabling continuous patient monitoring, intelligent clinical decision-making, and efficient remote care. However, these advancements have also introduced substantial cybersecurity risks that threaten patient privacy, safety, and the operational resilience of healthcare systems. These challenges are further compounded by stringent regulatory requirements and the inherent complexity of heterogeneous, non-independent, and identically distributed (non-IID) data. To address these challenges, we propose FTL-TSLP, a novel federated intrusion detection framework that integrates federated learning (FL) with targeted transfer learning (TL) through a two-stage LSTM-based pipeline. The framework is explicitly designed to operate effectively under both IID and non-IID data distributions while preserving data privacy. On the client side, temporal aggregation techniques efficiently compress sequential data, reducing computational costs without compromising detection accuracy. Additionally, the framework enhances fault tolerance by incorporating a Multi-Criteria Decision Analysis (MCDA) module combined with a Naïve Bayes classifier for real-time, probabilistic device-level classification. The proposed model demonstrates superior performance across the NF-UNSW-NB15-v2, WUSTL-EHMS-2020, and CICIoMT-2024 benchmark datasets. Even under extreme Dirichlet-based non-IID conditions ($\alpha = 0.1$), FTL-TSLP achieves 99.72 % accuracy and a 98.07 % F1-score on the CICIoMT-2024 dataset, confirming its robustness in heterogeneous IoMT traffic environments. These results highlight that FTL-TSLP offers a reliable, privacy-preserving, and computationally efficient solution for securing IoMT healthcare ecosystems.

1. Introduction

The rapid digitisation of healthcare has positioned the Internet of Medical Things (IoMT) as an indispensable component of contemporary clinical practice. IoMT integrates advanced edge-cloud infrastructures, sophisticated biosensing technologies, and interoperable communication frameworks, facilitating continuous physiological monitoring, real-time clinical decision-making, and personalised therapeutic interventions [1]. The widespread deployment of wearable devices and remote patient-monitoring systems

[*] Corresponding author.
E-mail addresses: abdlhammid.bouazza@univ-msila.dz (A. Bouazza), hichem.debbi@univ-msila.dz (H. Debbi), hicham.lakhlef@u-bordeaux.fr (H. Lakhlef).

https://doi.org/10.1016/j.iot.2025.101832
Received 15 September 2025; Received in revised form 17 November 2025; Accepted 20 November 2025
Available online 23 November 2025

further underscores IoMT's potential to generate substantial volumes of physiological data, which are highly amenable to advanced analytics [2]. Market forecasts predict significant growth in the global IoMT market, expanding from approximately USD 182.0 billion in 2023 to USD 668.1 billion by 2030, with a compound annual growth rate (CAGR) of 20.4 %. Broader market definitions anticipate even higher valuations, reaching USD 814 billion by 2032. [3]. This rapid expansion, however, also amplifies the cybersecurity vulnerabilities inherent in interconnected clinical systems.

IoMT infrastructures are inherently heterogeneous and decentralised, with devices often operating under stringent computational and energy constraints, across diverse environmental conditions, and using a variety of communication protocols such as Wi-Fi, Bluetooth Low Energy (BLE), and MQTT. These characteristics severely limit the effectiveness of traditional cybersecurity methods, such as perimeter defences, signature-based intrusion detection systems (IDS), and static vulnerability assessments, that are predominantly designed for centralised and homogeneous network environments [4,5]. As a result, conventional cybersecurity strategies struggle to effectively mitigate evolving threats and adapt to the dynamic operational demands of privacy-sensitive clinical settings.

Cybersecurity incidents in healthcare settings pose direct risks to patient safety, including unauthorized manipulation of therapeutic parameters, disruption of critical monitoring services, and exposure of sensitive health information. High-profile security breaches, such as the Change Healthcare ransomware incident of 2024-2025 and vulnerabilities identified in commonly deployed infusion pump platforms, highlight the urgent need for cybersecurity frameworks specifically tailored for clinical environments [6–8].

While machine learning (ML) and deep learning (DL) techniques show considerable promise in modelling intricate attack patterns, conventional centralised training paradigms require the aggregation of sensitive patient data. This raises critical concerns regarding patient privacy, scalability, and compliance with stringent regulatory frameworks, including the U.S. Health Insurance Portability and Accountability Act (HIPAA) and the European Union's General Data Protection Regulation (GDPR) [9,10]. Moreover, IoMT implementations must comply with regulatory standards such as FDA premarket cybersecurity guidance and comprehensive device/-software lifecycle requirements, including IEC 81001-5-1, ISO 14971, and IEC 62304, which emphasise secure lifecycle engineering and rigorous risk management [11–14].

Cross-institutional heterogeneity further complicates the effectiveness of intrusion detection systems (IDS). Variations in clinical specialisation, technological infrastructure, geographic location, and patient demographics result in non-independent and identically distributed (non-IID) datasets. These datasets are characterised by significant label imbalance, shifts in feature distributions, and temporal drift, all of which impair model generalisation capabilities [15].

Federated Learning (FL) has emerged as a promising approach to mitigate privacy concerns by enabling collaborative model training without the need to share raw data. However, the non-IID conditions prevalent in healthcare, such as pronounced label skew, intermittent client participation, and temporal variability, pose significant challenges to FL performance. Previous research addressing IoMT-specific FL challenges, even those employing hierarchical or meta-learning approaches [16,17], often suffers from limitations such as a restricted number of clients, controlled laboratory settings, and reliance on outdated or generic datasets (e.g., NSL-KDD, ToN-IoT) that do not adequately reflect contemporary IoMT attack scenarios [18].

Furthermore,IoMT infrastructures must accommodate diverse failure modes, including transient network partitions, Byzantine device behaviours, cascading failures across interdependent components, and coordinated cyber-attacks targeting multiple system layers. To address these challenges, this paper introduces a fault-tolerant intrusion detection architecture, **Federated Transfer Learning with Two-Stage LSTM Pipeline (FTL-TSLP)**, specifically designed for IoMT environments characterised by non-IID data distributions.

The primary contributions of this research are as follows:

- **Fault-Tolerant Federated Learning Protocol:** We propose a fault-tolerant federated learning protocol that maintains system functionality despite node failures through adaptive client selection, redundant model checkpointing, and Byzantine-resilient aggregation.
- **Adaptive Label Partitioning:** A novel client-specific strategy designed to address significant label skew, stabilising federated learning under heterogeneous participation scenarios.
- **Two-Stage LSTM Pipeline with Temporal Aggregation:** An optimised temporal modelling architecture that significantly reduces computational and memory requirements through staged processing, ensuring effective deployment on resource-constrained endpoint devices without compromising detection performance.
- **Hybrid Federated-Transfer Optimisation:** An integrated training methodology combining federated averaging with targeted transfer learning to manage extreme non-IID conditions. This method is compatible with secure aggregation protocols and optional differential privacy implementations.
- **Comprehensive Experimental Evaluation:** A rigorous assessment using specialised IoMT datasets, including NF-UNSW-NB15-v2 (network intrusions based on NetFlow data), WUSTL-EHMS-2020 (cyber-biomedical telemetry), and CICIoMT-2024, validating the robustness and efficacy of the proposed FTL-TSLP framework across realistic threat scenarios.

The remainder of this paper is organised as follows: Section 2 reviews related work; Section 3 formulates the research problem and outlines the proposed FTL-TSLP methodology; Section 4 details the experimental design and presents results; Section 5 discusses implications; and Section 6 concludes with recommendations for future research.

2. Related work

The rapid expansion of the Internet of Medical Things (IoMT) has fundamentally transformed healthcare infrastructure, fostering interconnected medical ecosystems that significantly enhance patient care outcomes. However, this rapid digitisation concurrently introduces increased cybersecurity vulnerabilities. As a result, developing sophisticated Intrusion Detection Systems (IDS), specifically designed for IoMT networks, has become imperative. This literature review critically evaluates current IDS methodologies, categorising them into three principal paradigms: centralised machine learning (ML), deep learning (DL), and federated learning (FL). The discussion highlights core challenges, including privacy preservation, scalability, and the management of non-independent and identically distributed (non-IID) data.

2.1. *Centralised machine learning paradigms for IoMT security*

Centralised ML techniques underpin contemporary research in IoMT security, primarily due to their interpretability and computational efficiency. Nonetheless, these methods exhibit notable limitations, such as vulnerability to privacy breaches resulting from centralised data aggregation, scalability constraints in resource-limited environments, and difficulties in effectively managing non-IID data.

Binbusayyis et al. [19] conducted an extensive benchmarking study of traditional machine learning (ML) algorithms, including Naïve Bayes, k-Nearest Neighbour (KNN), Decision Trees, Support Vector Machines (SVM), and Multi-layer Perceptrons (MLP). The research emphasised the superior interpretability and flawless detection accuracy (100 %) achieved by Decision Trees on the Bot-IoT dataset. In contrast, Gupta et al. [20] presented a method addressing class imbalance by combining Random Forest classifiers with the Synthetic Minority Over-sampling Technique (SMOTE), achieving 94.23 % accuracy on the WUSTL-EHMS-2020 dataset. Nonetheless, the method was limited by binary classification and restricted dataset generalisability.

In contrast to Gupta et al., who focused on class imbalance, Alalhareth and Hong [21] introduced an approach prioritising computational efficiency through the Logistic Redundancy Gradual-Upweighting Mutual Information Feature Selection (LRGU-MIFS) algorithm. This method achieved commendable accuracy (94.9 %) with substantial reductions in feature dimensionality. Additionally, Aljuhani et al. [22] developed a model emphasising interpretability, employing Particle Swarm Optimisation (PSO) combined with SHapley Additive exPlanations (SHAP). The approach yielded 96.56 % accuracy but notably lacked explicit privacy preservation mechanisms.

Alsalman [23] presented the FusionNet model, an ensemble learning method achieving up to 99.5 % accuracy. Although outperforming SG-IDS by Saleh et al. [24], which achieved 98 % accuracy in Wireless Sensor Networks (WSN), FusionNet incurred substantial computational overhead, limiting its practicality for edge device deployment. Conversely, SG-IDS, while computationally efficient, was constrained by a narrow attack taxonomy.

Dadkhah et al. [25] advanced the field significantly by introducing the CICIoMT-2024 dataset, encompassing diverse IoMT protocols and cyberattack types. While the proposed approach achieved outstanding binary classification accuracy, its performance diminished in multiclass scenarios. Complementary methodologies presented by Salehpour et al. [26] introduced hybrid cloud-based techniques and advanced feature selection strategies, effectively balancing accuracy and computational efficiency but inadequately addressing critical privacy considerations.

In this study, [27], Doménech et al. present a comparison of ML models trained on IoT and IoMT-specific datasets (CICIoT2023 and CICIoMT2024). They highlight the importance of domain-specific data for effective IDS in IoMT environments. The study critiques dataset design choices and proposes optimization techniques, such as uniform windowing and adjustments to temporal dependencies. These techniques significantly improved detection performance, achieving an accuracy of 0.9985, emphasizing the necessity of IoMT-specific datasets and tailored preprocessing for robust IDS solutions in healthcare environments.

Centralised ML methods offer considerable advantages in interpretability and efficiency. Approaches presented by Alalhareth and Hong [21] and Saleh et al. [24] exemplify effective resource utilisation. However, high-performing ensemble methods such as FusionNet [23] typically involve significant computational trade-offs or limited detection scopes. Despite their efficiency and interpretability, the inherent privacy vulnerabilities and scalability constraints of these methods motivate further research into deep learning approaches capable of capturing more intricate data patterns.

2.2. *Deep learning paradigms for IoMT security*

Deep Learning (DL) methodologies have prominently emerged in IoMT security research, mainly due to their proficiency in modelling complex and temporal data patterns characteristic of medical networks. Nonetheless, DL methods face persistent challenges related to computational complexity, scalability constraints, and inadequate privacy safeguards.

Nandy et al. [28] proposed a Swarm-Neural Network architecture, achieving 99.5 % accuracy on the generic ToN-IoT dataset. Despite this high accuracy, the reliance on non-specific IoMT datasets limits the architecture's direct applicability to targeted IoMT scenarios. In contrast, Ghourabi [29] introduced a hybrid model that combines LightGBM with Transformer architectures, achieving superior accuracy while introducing higher complexity and associated privacy concerns.

Faruqui et al. [30] presented SafetyMed, employing Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) architectures to achieve remarkably low false-positive rates. However, substantial computational demands constrained its practicality in resource-limited IoMT deployments. Conversely, Alalhareth and Hong [31] developed a fuzzy self-tuning LSTM model demonstrating adaptive performance but hindered by elevated false-positive rates.

Khan et al. [32] utilised a fog-cloud-based ensemble comprising multiple LSTM models, significantly enhancing detection precision. Nevertheless, their method encountered considerable computational overhead and limited generalisability. Similarly, Alzubi et al. [33] optimised CNN-LSTM hybrid models, achieving improved accuracy, though their evaluations on non-specific IoMT datasets restricted broader applicability.

Akar et al. [34] advanced DL methodologies further by introducing the dual-stacked LSTM L2D2 model, effectively managing diverse cyber threats but lacking explicit privacy preservation mechanisms. Turgut and Başarslan [35] emphasised model interpretability by integrating bidirectional recurrent networks with SHAP and LIME explanations, delivering high accuracy and transparency but incurring substantial computational costs. Benmalek et al. [36] presented enhancements in hierarchical ensemble strategies, achieving substantial accuracy yet constrained by architectural complexity and binary classification limitations.

Furthermore, recent work has introduced several high-performing centralised models that address key IoMT security challenges through sophisticated techniques in feature optimisation, architectural innovation, and adaptive learning. For example, FOID [37] employs RFE-driven feature selection in conjunction with XGBoost and classical machine learning classifiers to enhance both interpretability and detection performance. To address evolving and dynamic threat landscapes, HCLR-IDS [38] integrates a CNN–LSTM architecture with reinforcement learning methods such as DQN and PPO, enabling adaptive response capabilities. Other studies have focused on advancing temporal modelling, with MF-Transformer [39] incorporating MF-LSTM layers within a transformer framework to capture long-range dependencies better. Similarly, RCLNet [40] advances hybrid deep learning approaches by combining Random Forest feature selection, CNN–LSTM fusion, and an adaptive attention mechanism to mitigate traffic imbalance and improve anomaly detection effectiveness.

Overall, DL paradigms typically outperform traditional ML methods in capturing complex IoMT traffic patterns, with multiclass models like L2D2 [34] providing significant improvements in threat detection granularity. Nonetheless, persistent computational overhead and insufficient privacy protections substantially restrict real-world IoMT deployments. Unlike centralised ML approaches, DL methods more effectively handle non-IID data but generally require extensive optimisation for deployment on edge devices. This limitation underscores the potential benefits of integrating federated learning frameworks to address these challenges comprehensively.

2.3. Federated learning paradigms for privacy-preserving IoMT security

While deep learning architectures effectively capture complex data patterns, federated learning offers a promising alternative by preserving privacy and decentralising computation. Federated learning (FL) uniquely enables privacy preservation by allowing localised model training without centralised data aggregation. Recent advancements in FL tailored specifically for IoMT have increasingly emphasised optimisation strategies, adaptability to emerging threats, and domain-specific customisations. Despite these progressions, significant challenges persist, notably managing non-Independent and Identically Distributed (non-IID) data, ensuring effective scalability, and maintaining comprehensive global threat visibility.

Singh et al. [16] introduced an innovative hierarchical federated learning architecture designed explicitly for IoMT security, termed the Dew-Cloud-based framework. This architecture organises computational tasks across three hierarchical layers: local dew servers for initial data processing, fog nodes for intermediate aggregation, and cloud servers for global model updates. Utilising hierarchical Long Short-Term Memory (LSTM) models, Singh et al. demonstrated notable accuracy (99.31 %) on the TON-IoT dataset, illustrating considerable scalability through experiments involving varied participant numbers. Nonetheless, reliance on older datasets such as NSL-KDD and generic IoT datasets limits the practical relevance of this framework in contemporary IoMT networks, which exhibit unique traffic patterns and diverse cyber threats.

In contrast to the scalability-focused approach by Singh et al., Zukaib et al. [17] presented a framework prioritising adaptability-Meta-Fed IDS, which incorporates meta-learning principles for rapidly responding to zero-day threats. This federated learning model integrates Decision Trees, AdaBoost, Extra Trees, and Random Forest classifiers, optimised by an XGBoost meta-learner. Zukaib et al.[17] reported high accuracy (99 %) on a custom IoMT dataset. Despite demonstrating robust adaptability, the limited evaluation involving only three clients and inadequate mechanisms for addressing non-IID data distributions constrain its scalability and broader deployment potential.

Sarhan et al. [41] proposed a federated threat intelligence framework addressing collaborative inter-organisational cybersecurity by leveraging Federated Averaging (FedAvg) combined with LSTM and Deep Neural Networks (DNN). Their approach achieved accuracies of 85.25 % and 94.61 % on the NF-UNSWNB15-v2 and NF-BoT-IoT-v2 datasets, respectively. Unlike IoMT-specific solutions such as that developed by Misbah et al. [42], Sarhan et al. [41] employed a generic FL approach, lacking detailed consideration of IoMT device heterogeneity and non-IID data challenges, thus limiting its practical effectiveness in diverse healthcare environments.

Recognising these IoMT-specific requirements, Misbah et al. [42] developed a federated learning framework tailored explicitly for medical IoT applications. Employing Random Forest classifiers and introducing a federated dynamic averaging strategy, which dynamically adjusts aggregation weights based on local model performance, they achieved 99.22 % accuracy on the CICIoMT-2024 dataset. However, despite these advancements, inherent limitations in Random Forest classifiers, particularly their inability to effectively capture complex temporal dependencies, constrain their suitability and performance in dynamic IoMT networks. Additionally, the strategy for managing non-IID data distributions is not explicitly discussed in this study.

Addressing resource constraints inherent in IoMT deployments, Ioannou et al. [43] introduced the GEMLIDS-MIOT framework, focusing specifically on energy efficiency. Their model combined an energy-pruned Enhanced Random Forest classifier with a One-Class Support Vector Machine (SVM) for anomaly detection, employing a privacy-preserving depth-first search (DFS) delta aggregation method. Implemented on Raspberry Pi gateways, GEMLIDS-MIOT demonstrated exceptional classification accuracy (99.98 %) and anomaly detection accuracy (99.7 %), significantly reducing energy consumption. In contrast, Khan et al. [44] proposed the Fed-Inforce-Fusion framework, aiming to reduce communication overhead through reinforcement learning. By integrating Q-learning-based local models with dynamic federated aggregation, Fed-Inforce-Fusion achieved 99 % accuracy and reduced communication

Table 1
Systematic Evaluation of Federated Learning-Based Intrusion Detection Systems for IoMT.

Study	Model Architecture	Dataset(s)	Results	Limitations
[16] (2023)	Hierarchical FL with LSTM, dew-cloud	ToN-IoT, NSL-KDD	Acc.: 99.31 %, Prec.: 98.97 %, F1: 98.58 %	Outdated datasets; Non-IID unaddressed
[41] (2023)*	FL-based LSTM/DNN with FedAvg	NF-UNSW-NB15-v2, NF-BoT-IoT-v2	Acc.: 85.25 %–94.61 %	Generic IoT context; limited heterogeneity evaluation
[45] (2024)*	FL vs. centralized IDS comparison	WUSTL	FL: Acc.: 88.00 %, Prec.: 66.00 %, F1: 57.00 % (vs. CL: 91.00 %, 99.00 %, 59.00 %)	Performance degradation under Non-IID; binary classification
[43] (2024)	GEMLIDS-MIOT: Enhanced RF with OC-SVM	Custom MIoT	Acc.: 99.98 %, Anomaly detection: 99.70 %	Proprietary dataset; no Non-IID evaluation
[44] (2024)	Fed-Inforce-Fusion: Q-learning FL	ToN-IoT, UCI Heart-Disease	Acc.: 99.40 %, Communication reduction: 57.00 %	Limited scalability (3 clients); non-IoMT datasets
[17] (2024)	Meta-Fed IDS: FL with meta-learning	M-IoT-Env, WUSTL, M-En	Acc.: 99.00 %	Severely limited scalability (3 clients); no Non-IID
[42] (2025)*	FL with RF and dynamic averaging	CICIoMT	Acc.: 99.22 % (RF), 98.59 % (AdaBoost)	Traditional ML constraints; Non-IID strategy unspecified.

Key: *Explicitly evaluates Non-IID data distribution.

overhead by 57 %. These two methodologies effectively address resource constraints but differ significantly in their optimisation goals-energy efficiency versus communication efficiency-highlighting distinct strategic trade-offs.

Additionally, a recent study [45] examined the application of FL for intrusion detection in IoMT, focusing on the challenges posed by non-IID data. Using the WUSTL-EHMS-2020 dataset, which includes cyber-attacks such as man-in-the-middle, spoofing, and data injection, the study compared the performance of FL-based intrusion detection systems (IDS) with traditional centralized learning methods. The results revealed a performance degradation when using FL, with accuracy dropping from 91 % to 88 %, precision declining from 99 % to 66 %, and the F1-score decreasing from 59 % to 57 %, primarily due to the challenges of non-IID data distribution. The study highlights the difficulty of achieving model convergence in heterogeneous data environments and acknowledges the limitations in experimental design, such as binary attack classification, the lack of scalability, and the absence of robust privacy evaluations. Despite these limitations, the research advances the understanding of privacy-preserving IDS solutions for IoMT and emphasizes the need for further exploration into methods to mitigate the effects of non-IID data in distributed learning environments. Overall, federated learning paradigms offer substantial promise for balancing privacy with collaborative intelligence. Nevertheless, persistent limitations remain across existing methodologies, particularly inadequate handling of non-IID data, limited scalability testing involving relatively few participants, and significant computational demands. Addressing these challenges through integrated approaches that leverage the strengths of diverse federated methodologies represents a critical direction for future IoMT security research.

The existing IDS literature, as summarised in Table 1, highlights critical trade-offs between detection accuracy, privacy preservation, fault tolerance, scalability, and the management of non-IID data. Centralised ML methods typically achieve high detection accuracy but often compromise user privacy. Conversely, FL methods excel in preserving privacy but frequently encounter notable performance and scalability limitations. Thus, an integrated and comprehensive solution is required. To effectively address these limitations, this paper proposes the FTL-TSLP framework. The proposed methodology innovatively combines federated learning, transfer learning, and a two-stage LSTM pipeline to deliver robust privacy preservation, efficient handling of non-IID data distributions, and enhanced scalability. This integrated approach represents a significant advancement in IDS methodologies, explicitly tailored for IoMT security contexts. Additionally, the integration of MCDA-based decision-making strategies and fault-aware routing protocols marks a critical shift from traditional reactive fault management towards proactive resilience enhancement in IoMT security architectures.

3. Methodology

This paper proposes a robust framework designed for fault-tolerant, privacy-preserving, and adaptive intrusion detection within Internet of Medical Things (IoMT) networks. The proposed methodology synthesises Federated Learning (FL), Transfer Learning (TL), a Two-Stage LSTM Pipeline (TSLP), Multi-Criteria Decision Analysis (MCDA) via the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), the Analytic Hierarchy Process (AHP), and Naïve Bayes classification. Our approach addresses critical challenges inherent to IoMT environments, including statistical heterogeneity, non-independent and identically distributed (non-IID) data, privacy constraints, and the necessity for real-time adaptability in dynamic, heterogeneous settings.

3.1. System model and feature representation

We consider an IoMT network comprising n heterogeneous devices, each characterised by a feature vector:

$$\mathbf{x}_i \in \mathbb{R}^m, \tag{1}$$

where each feature vector encompasses key operational metrics including *node safety*, *latency*, *packet loss*, *jitter*, and the Clinical Risk Index (CRI). The objective of our approach is:

1. To implement distributed anomaly detection and device ranking utilising Federated Transfer Learning with FTL-TSLP integrated with Multi-Criteria Decision Analysis (MCDA).
2. To leverage the CRI as an independent clinical override to guide adaptive, patient-centric routing and safety enforcement.

The overall methodological architecture is illustrated in Fig. 1, highlighting the integration of federated learning, transfer learning, MCDA via TOPSIS, CRI-based clinical prioritisation, and real-time Naïve Bayes classification.

3.2. Data preprocessing and feature engineering

Effective data preprocessing is critical to ensure data integrity and model reliability, particularly in IoMT networks characterised by diverse devices and varying data quality. The preprocessing workflow consists of the following essential steps:

- **Feature Harmonisation :** Operational metrics such as latency, jitter, and packet loss are normalised to ensure feature comparability. *Node safety* is derived from IDS outputs or device integrity metrics and normalised within [0,1], with higher values indicating safer operation.
- **Clinical Risk Index (CRI):** The CRI, computed within [0,1], represents the clinical severity or instability of patient physiological parameters, thereby enabling clinically informed adaptive routing and decision-making.
- **Encoding and Normalisation :** Categorical features are transformed using one-hot encoding, whereas numerical features undergo robust normalisation, such as min-max scaling, to mitigate the effect of outliers and variability.
- **Class Imbalance Mitigation:** To adequately represent rare but clinically significant intrusion types, the Synthetic Minority Oversampling Technique (SMOTE) is applied during training, ensuring balanced dataset representation.

3.3. Criteria weight determination via analytic hierarchy process (AHP)

The Analytic Hierarchy Process (AHP) is used to assign relative importance to each evaluation criterion systematically. Consider the criteria set:

$$C = \{c_1 = \text{node safety (benefit)}, c_2 = \text{latency (cost)}, c_3 = \text{packet loss (cost)}, c_4 = \text{jitter (cost)}\}.$$

AHP generates a pairwise comparison matrix $\mathbf{A} \in \mathbb{R}^{4\times 4}$, where each element a_{ij} represents the relative importance of criterion i over criterion j, satisfying reciprocal consistency:

$$a_{ij} = \frac{1}{a_{ji}}, \quad \forall i, j \in \{1, 2, 3, 4\}. \tag{2}$$

Criteria weights are determined by solving the eigenvalue problem:

$$\mathbf{A}\mathbf{w} = \lambda_{\max}\mathbf{w}, \quad \text{s.t.} \quad \sum_{j=1}^{4} w_j = 1, \quad w_j > 0, \tag{3}$$

where $\lambda_{\max}$ represents the largest eigenvalue. To ensure the solution is valid, the eigenvector $\mathbf{w}$ must be **normalized** to satisfy the constraint $\sum_{j=1}^{4} w_j = 1$.

Normalization step:

$$w_{\text{normalized}} = \frac{\mathbf{w}}{\sum_{j=1}^{4} w_j}.$$

Judgment coherence is validated via the Consistency Ratio (CR), defined as:

$$CR = \frac{CI}{RI} < 0.1, \quad CI = \frac{\lambda_{\max} - n}{n - 1}, \tag{4}$$

where $RI \approx 0.90$ for $n = 4$. A CR below 0.1 confirms acceptable consistency.

3.4. Federated–transfer learning with FTL-TSLP

To address privacy and heterogeneity challenges, we propose the *Federated Transfer Learning with Two-Stage LSTM Pipeline* (FTL-TSLP) framework, designed for privacy-preserving intrusion detection in IoMT environments. The primary objective is to address statistical heterogeneity challenges arising from diverse network configurations, heterogeneous IoMT device specifications (non-independent and identically distributed (non-IID) data), and distinct intrusion patterns. To mitigate these heterogeneities, FTL-TSLP

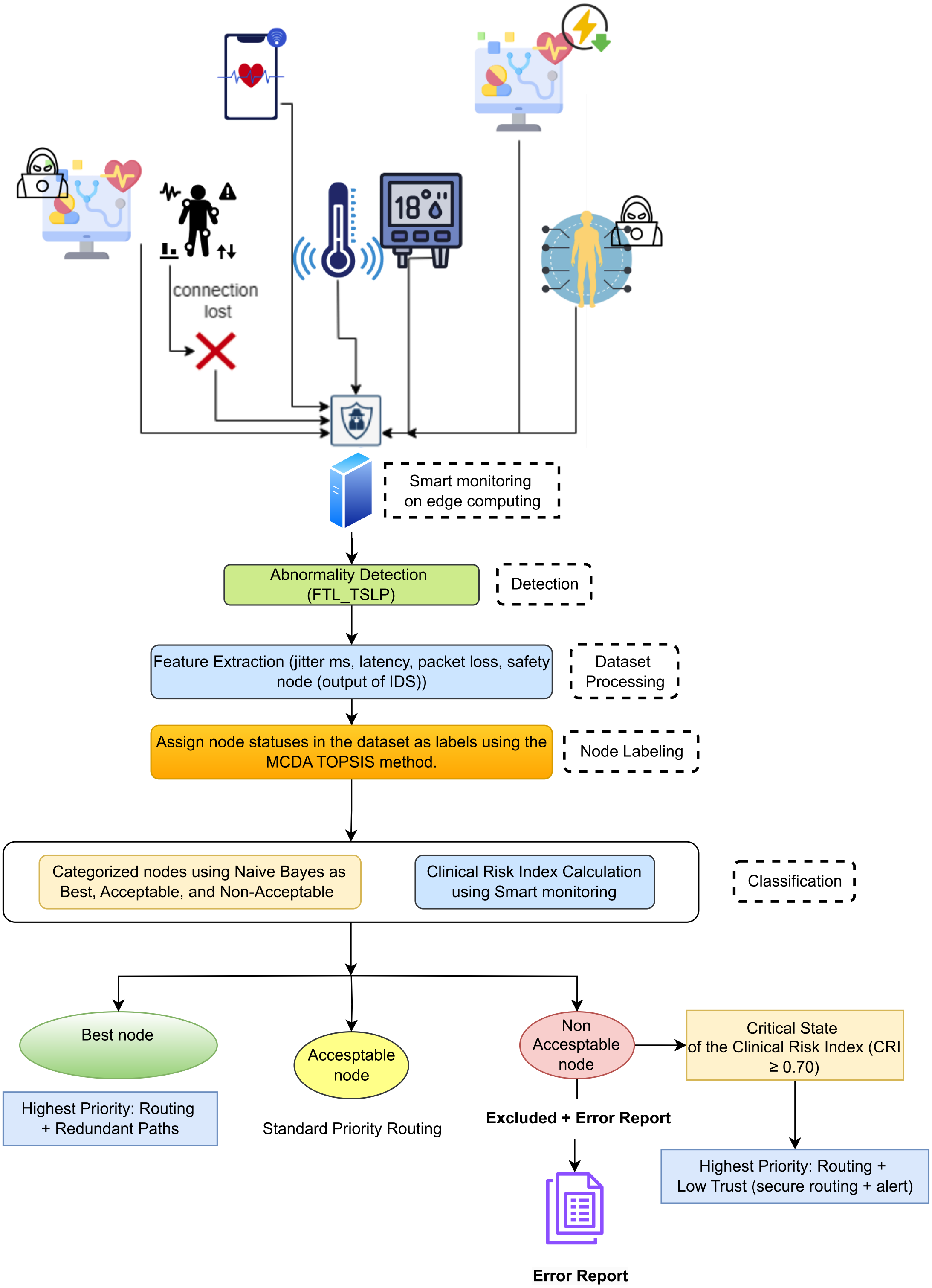

connection
lost
18°
Smart monitoring
on edge computing
Abnormality Detection
(FTL_TSLP)
Detection
Feature Extraction (jitter ms, latency, packet loss, safety node (output of IDS))
Dataset
Processing
Assign node statuses in the dataset as labels using the MCDA TOPSIS method.
Node Labeling
Categorized nodes using Naive Bayes as Best, Acceptable, and Non-Acceptable
Clinical Risk Index Calculation using Smart monitoring
Classification
Best node
Highest Priority: Routing + Redundant Paths
Accesptable node
Standard Priority Routing
Non Acceptable node
Excluded + Error Report
Critical State of the Clinical Risk Index (CRI ≥ 0.70)
Highest Priority: Routing + Low Trust (secure routing + alert)
Error Report

Fig. 1. Proposed fault-tolerant and adaptive IoMT intrusion detection framework integrating FTL-TSLP, MCDA, clinical risk indexing (CRI), and real-time Naïve Bayes classification.

integrates *Federated Learning* (FL) for collaborative detection of common intrusion patterns across multiple clients, and client-specific *Transfer Learning* (TL) for handling isolated intrusion labels unique to individual clients.

FTL-TSLP adopts a decentralised training strategy in which edge clients independently analyse local IoMT traffic data. Clients communicate exclusively model parameters and non-sensitive label metadata to a central aggregator, ensuring strict adherence to healthcare data privacy regulations. Sensitive patient and institutional data remain securely stored within local environments. An architectural overview of the proposed framework is illustrated in Fig. 2.

The FTL-TSLP framework consists of three primary layers:

- **IoMT Device Layer:** This foundational layer comprises diverse healthcare monitoring devices, such as wearable sensors and bedside medical equipment. These devices continuously generate multivariate network and telemetry data streams, forming the primary input for the IDS.
- **Edge Client Layer:** This intermediate layer consists of clinical institutions, including hospitals and specialised medical departments. Each institution operates dedicated edge computing nodes responsible for local data preprocessing, temporal aggregation of data, and localised TSLP model training.
- **Central Cloud Aggregator:** At the top tier, the central cloud aggregator coordinates FL for common intrusion labels and distributes both the global federated model and specialised TL models for isolated intrusion labels.

The TSLP architecture is selected for its two-stage pipeline structure, explicitly designed to optimize computational efficiency at the client level. This design maintains the predictive performance of traditional LSTM architectures while significantly reducing training and inference times through staged processing. The two-stage approach is particularly effective in modelling the sequential and temporal characteristics of IoMT-based cyberattacks, which often involve multi-stage processes such as reconnaissance, exploitation, and exfiltration. The first stage rapidly filters normal traffic through binary classification, while the second stage performs detailed attack categorization only on suspicious traffic. Furthermore, it effectively supports continuous clinical monitoring scenarios, where anomalies gradually develop over extended periods, with the temporal aggregation layer between stages capturing these gradual patterns.

The operational workflow of the FTL-TSLP framework comprises three distinct phases, as shown in Fig. 3.

- **Phase 1: Intelligent Label Classification:** Initially, each client transmits a metadata vector indicating the presence of labels to the central aggregator. The server employs Algorithm 1 to categorise these labels into *common labels*-present across multiple clients-and *isolated labels*-unique to individual clients. Direct integration of isolated labels into FL can negatively affect aggregation efficiency (e.g., in Federated Averaging (FedAvg)) due to their exclusivity. This classification enables targeted learning strategies without compromising data privacy, as raw IoMT data are never transmitted.
- **Phase 2: Hybrid Learning Execution:** For common labels, clients independently train local models using relevant data subsets. The central aggregator consolidates these updates via FedAvg to produce a unified global detection model. For isolated labels, clients clone their local TSLP models, modify the final classification layer for binary classification specific to the isolated label, and fine-tune the model locally. The specialised TL models are then sent to the central aggregator for distribution alongside the global federated model.
- **Phase 3: Optimised Deployment:** Clients use Algorithm 3 to select the optimal detection model-either the global federated model or a client-specific TL model-for each intrusion label. This selection process applies a multi-criteria evaluation considering detection accuracy, inference latency, and computational efficiency, ensuring efficient and context-specific deployment.

Throughout the operational workflow, the FTL-TSLP framework maintains strict privacy preservation by restricting communication to secure transmissions of model parameters and non-sensitive metadata. Sensitive IoMT and patient data remain entirely local, ensuring compliance with established healthcare data privacy regulations and standards.

3.4.1. Server-side methodology for federated transfer learning in IoMT

The server functions as a supervisory control entity, orchestrating collaborative training processes among participating clients. At the start of each training cycle, the server executes Label Classification (Algorithm 1) to partition the label set into *common labels* (L_{common}) and *isolated labels* (L_{isolated}). Isolated labels, which occur exclusively within individual client datasets, introduce significant non-independent and identically distributed (non-IID) challenges, thereby adversely impacting federated learning performance. To mitigate these effects, the server assigns isolated labels to a dedicated TL pathway.

Following label categorisation, the server identifies the specific client $c^*(\ell)$ owning each isolated label $\ell \in L_{\text{isolated}}$ and sends a TL notification to that client. Upon receiving the notification, the client constructs a TL model by:

1. Cloning its existing multi-class *Two-Stage LSTM Pipeline* (TSLP) model,
2. Converting the output layer into a binary classifier tailored to the isolated label ℓ, and
3. Fine-tuning the classifier using its local dataset $\mathcal{D}_{c^*(\ell)}$.

Simultaneously, clients with non-empty intersections $\mathcal{L}_c \cap L_{\text{common}} \neq \emptyset$ perform training on subsets of their data corresponding to the common labels. The server aggregates these updates using *Federated Averaging* (FedAvg; Algorithm 2), yielding an updated global model $w^{(t)}$. Subsequently, the server redistributes both the global model and the specialised TL models $\{\text{TL\_MODEL}(\ell)\}$ to all clients for continued training and deployment.

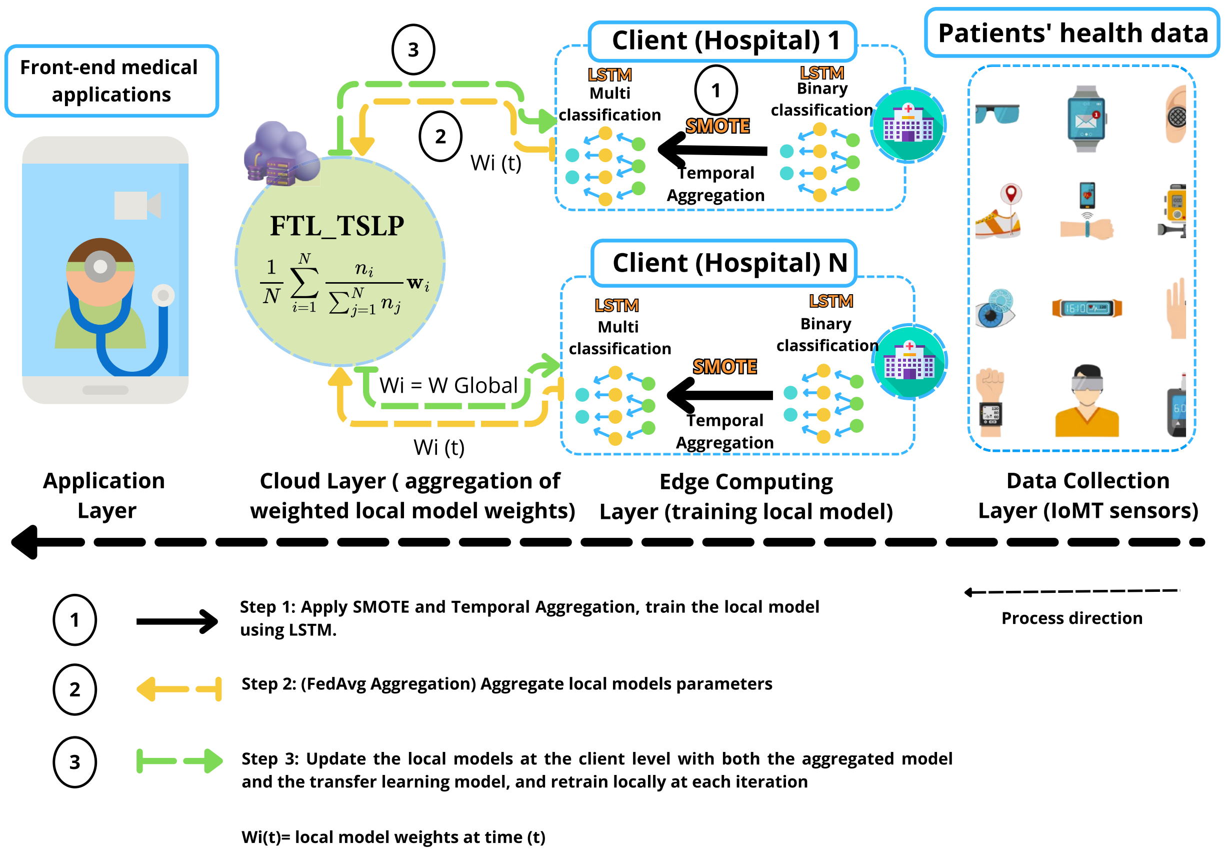

Front-end medical applications
3
2
Wi (t)
FTL_TSLP
$\frac{1}{N}\sum_{i=1}^{N}\frac{n_i}{\sum_{j=1}^{N} n_j}\mathbf{w}_i$
Wi = W Global
Wi (t)
Client (Hospital) 1
LSTM
Multi classification
1
SMOTE
LSTM
Binary classification
Temporal Aggregation
Client (Hospital) N
LSTM
Multi classification
SMOTE
LSTM
Binary classification
Temporal Aggregation
Patients' health data
Application Layer
Cloud Layer (aggregation of weighted local model weights)
Edge Computing Layer (training local model)
Data Collection Layer (IoMT sensors)
1
Step 1: Apply SMOTE and Temporal Aggregation, train the local model using LSTM.
2
Step 2: (FedAvg Aggregation) Aggregate local models parameters
3
Step 3: Update the local models at the client level with both the aggregated model and the transfer learning model, and retrain locally at each iteration
Wi(t)= local model weights at time (t)
Process direction

Fig. 2. Overview of the FTL-TSLP architecture showing the federated learning approach for intrusion detection in IoMT networks.

Throughout this entire process, only model parameters and non-sensitive label metadata are exchanged. No raw IoMT data or personally identifiable information leaves the local client environment, ensuring that all privacy and regulatory constraints are maintained and that sensitive data remain securely localised.

Intelligent Label Classification

Let $\mathcal{Y}$ denote the set of all labels/classes recognized by the IDS, where $|\mathcal{Y}|$ represents the number of labels. The Intelligent Label Classification procedure (Algorithm 1) runs at the beginning of each federated round and whenever the client pool or label taxonomy changes.

Each client $c \in \mathcal{C}$ sends a binary label-presence vector $v_c \in \{0,1\}^{|\mathcal{Y}|}$, where $v_c[\ell] = 1$ if label $\ell \in \mathcal{Y}$ is observed locally. The server then computes cross-client support:

$$s(\ell) = \sum_{c\in\mathcal{C}} v_c[\ell]$$

The server partitions $\mathcal{Y}$ using a minimum-support threshold $k_{\min}$:

$$L_{\text{common}} = \{\ell \in \mathcal{Y} : s(\ell) \geq k_{\min}\}, \quad L_{\text{isolated}} = \mathcal{Y} \setminus L_{\text{common}}$$

Labels in L_{isolated} are those with $s(\ell) < k_{\min}$. By default, $k_{\min} = 2$, ensuring that only labels with multi-site evidence contribute to the shared objective. This reduces the transfer from isolated labels (i.e., $s(\ell) < k_{\min}$) and improves privacy by mitigating label-uniqueness inference.

(Optional) Secure aggregation and differentially private noise can be applied to $\{s(\ell)\}_{\ell\in\mathcal{Y}}$ before thresholding.

Transfer Learning for Isolated Labels

For each $\ell \in L_{\text{isolated}}$, the server identifies its unique owner $c^*(\ell)$ and initiates client-side transfer learning (TL). The client clones the shared encoder θ_s from the global TSLP , removes the multi-class head, and instantiates a label-specific binary classifier

$$g_\ell(x;\theta_s,\theta_\ell) = \sigma\big(W_\ell^{\top} h(x;\theta_s) + b_\ell\big), \tag{5}$$

where $h(\cdot;\theta_s)$ is the (frozen) encoder output and $\theta_\ell = (W_\ell, b_\ell)$. The head is trained with weighted binary cross-entropy to address class imbalance:

$$\mathcal{L}_\ell(\theta_\ell) = -w_+ y_\ell \log g_\ell(x) \;-\; w_-\big(1-y_\ell\big)\log\big(1-g_\ell(x)\big), \tag{6}$$

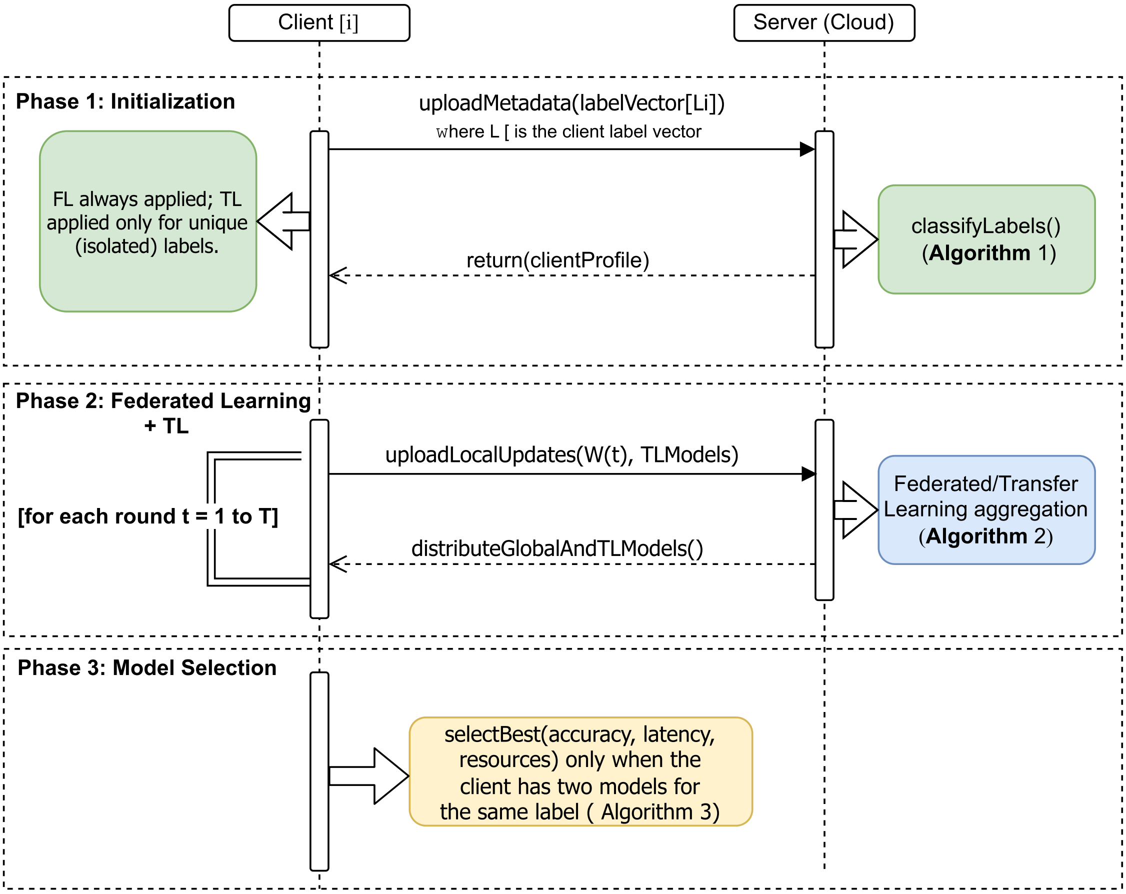

Client [i]
Server (Cloud)
Phase 1: Initialization
uploadMetadata(labelVector[Li])
where L [is the client label vector
FL always applied; TL applied only for unique (isolated) labels.
classifyLabels() (Algorithm 1)
return(clientProfile)
Phase 2: Federated Learning + TL
[for each round t = 1 to T]
uploadLocalUpdates(W(t), TLModels)
Federated/Transfer Learning aggregation (Algorithm 2)
distributeGlobalAndTLModels()
Phase 3: Model Selection
selectBest(accuracy, latency, resources) only when the client has two models for the same label (Algorithm 3)

Fig. 3. Sequence diagram of the proposed Federated Transfer Learning (FTL-TSLP) workflow.

Algorithm 1 Label Classification.

```
Input: Client set $\mathcal{C}$; local label sets $\{L_c\}_{c\in\mathcal{C}}$
Output: $L_{\text{isolated}}$ (labels appearing in exactly one client); $L_{\text{common}}$ (labels appearing in at least two clients)
LabelCount ← empty map
for each $c \in \mathcal{C}$ do
    for each $\ell \in L_c$ do
        LabelCount[$\ell$] ← LabelCount[$\ell$] +1 (initialize to 1 if absent)
    end for
end for
$L_{\text{isolated}} \leftarrow \emptyset, \quad L_{\text{common}} \leftarrow \emptyset$
for each $\ell$ in keys(LabelCount) do
    if LabelCount[$\ell$] $< k_{\min}$ then
        $L_{\text{isolated}} \leftarrow L_{\text{isolated}} \cup \{\ell\}$
    else
        $L_{\text{common}} \leftarrow L_{\text{common}} \cup \{\ell\}$
    end if
end for
return $L_{\text{isolated}}, L_{\text{common}}$
```

with $y_\ell = \mathbf{1}\{y = \ell\}$ and (w_+, w_-) derived from local class frequencies. A calibrated threshold τ_ℓ may be selected on a validation split to satisfy a site policy, e.g., FPR $\leq 1\,\%$.

Federated Learning for Common Labels

For client c, let $L_c^{\text{com}} = L_c \cap L_{\text{common}}$. Local training minimizes a masked objective:

$$\min_{w} \ \mathbb{E}_{(x,y)\sim\mathcal{D}_c}\big[\mathbf{1}\{y \in L_c^{\text{com}}\} \cdot \ell(f(x;w), y)\big], \tag{7}$$

ensuring samples outside L_c^{com} contribute zero gradient.

Server-side aggregation employs sample-weighted FedAvg over the set $S_t = \{c \in C : |L_c^{\text{com}}| > 0\}$, and the updated global model is computed as:

$$w^{(t)} = \frac{\sum_{c\in S_t} n_c^{\text{com}} w_c^{(t)}}{\sum_{c\in S_t} n_c^{\text{com}}}, \quad n_c^{\text{com}} = \left|\{(x,y) \in \mathcal{D}_c : y \in L_c^{\text{com}}\}\right|. \tag{8}$$

Where S_t is defined as:

$$S_t = \{c \in C : |L_c^{\text{com}}| > 0\}$$

This defines the set of clients c that have at least one common label for aggregation.

Algorithm 2 Federated and Transfer Learning Workflow.

```
Input: Clients $C$; initial global model $w^{(0)}$; local epochs $E$; rounds $T$; $L_{\text{isolated}}$, $L_{\text{common}}$ from Algorithm 1
Output: $w^{(T)}$ (global model for common labels); $\{w_c^{\text{TL}}\}$ (TL heads for isolated labels)
for all $c \in C$ in parallel do
    $w_c^{\text{local}} \leftarrow w^{(0)}$; train on $\mathcal{D}_c$ for $E$ epochs
end for
for $t = 1$ to $T$ do
    Broadcast $w^{(t-1)}$
    for all $c \in C$ in parallel do
        if $L_c \cap L_{\text{common}} \neq \emptyset$ then
            $w_c^{(t)} \leftarrow w^{(t-1)}$; train on $\mathcal{D}_c$ masked to $L_c^{\text{com}}$ for $E$ epochs
            Upload $w_c^{(t)}$ (or $\Delta w_c^{(t)}$) via secure aggregation
        end if
        if $L_c \cap L_{\text{isolated}} \neq \emptyset$ then
            Create $w_c^{\text{TL}}$ from $w_c^{\text{local}}$; replace head with binary head(s) for $L_c^{\text{iso}}$
            Freeze encoder; fine-tune head(s) on $\mathcal{D}_c|_{L_c^{\text{iso}}}$; upload $w_c^{\text{TL}}$ (params + metrics)
        end if
    end for
    Aggregate (common labels): $w^{(t)} \leftarrow \frac{\sum_{c\in S_t} n_c^{\text{com}} w_c^{(t)}}{\sum_{c\in S_t} n_c^{\text{com}}}$
    Redistribute $w^{(t)}$ and collected $\{w_c^{\text{TL}}\}$ to all clients
end for
return $w^{(T)}$, $\{w_c^{\text{TL}}\}$
```

Federated and Transfer Learning Workflow

Only parameters and metadata are exchanged:

$$\left\{w^{(t)},\ \{w_c^{(t)}\}_{c\in S_t},\ \{\text{TL\_MODEL}(\ell)\}\right\},$$

and never raw IoMT traffic. Secure aggregation reveals only weighted sums (not individual updates). Optional differential privacy can be applied by adding calibrated noise to clipped gradients or to label-support counts prior to thresholding.

3.4.2. *Client-side methodology: Two-stage LSTM pipeline with federated and side transfer learning*

This section outlines the client-side implementation of a Two-Stage LSTM Pipeline (TSLP) architecture, specifically designed for efficient and effective intrusion detection in IoMT deployments (Fig. 4). The TSLP architecture seeks to achieve high detection accuracy while minimizing computational overhead and latency, essential for resource-constrained IoMT environments. The methodology employs a structured two-stage approach consisting of data preprocessing, binary anomaly detection, temporal aggregation, and multi-class classification. The pipeline design ensures that only suspicious traffic proceeds to the computationally intensive second stage, significantly reducing overall processing requirements. Additionally, an optional Side Transfer Learning (STL) module enhances detection of rare, client-specific attack types without disrupting the shared global model.

Notation

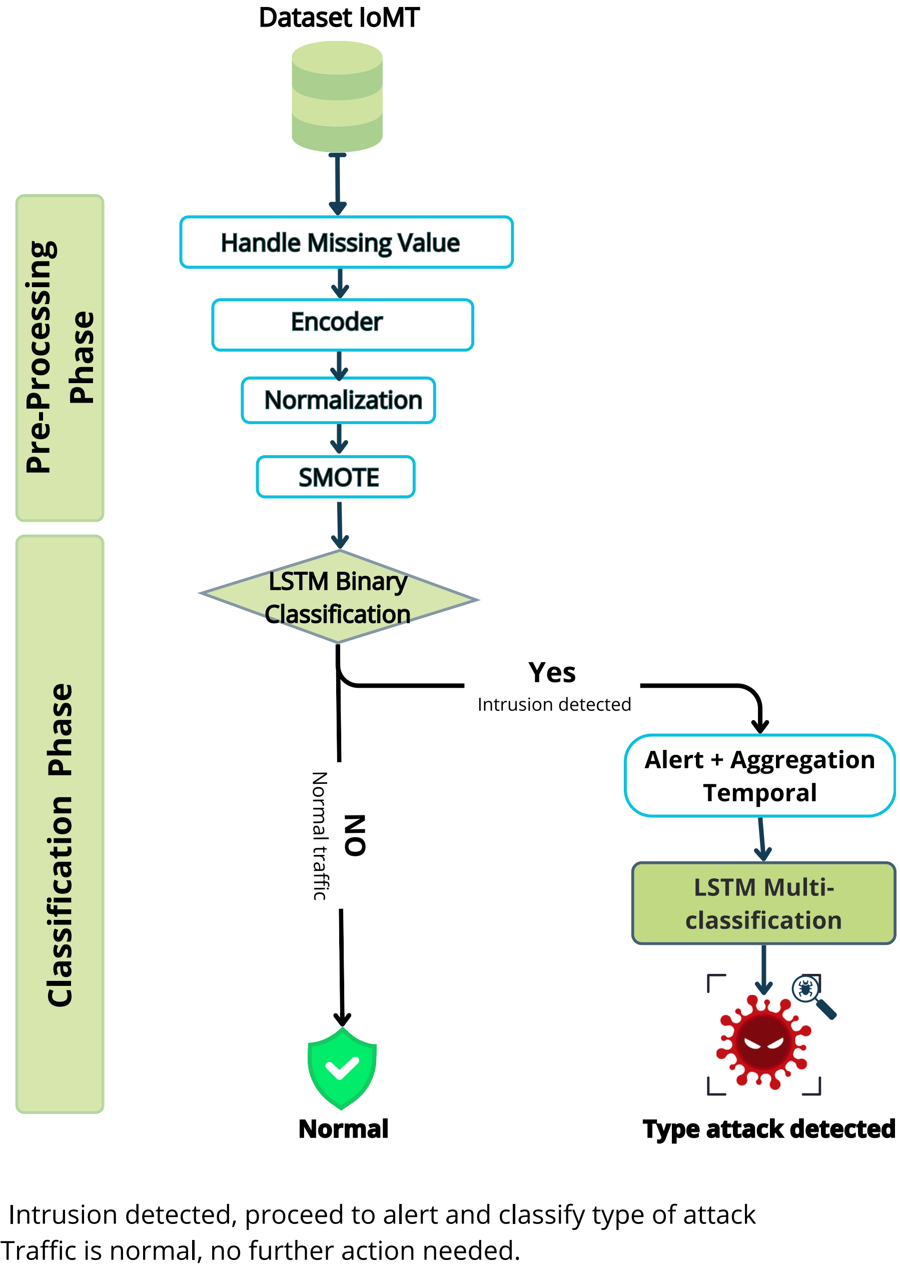
Let $x_t \in \mathbb{R}^d$ represents the feature vector at time t. Streaming data are segmented into overlapping windows:

$$W_k = \{t_k - L + 1, \dots, t_k\}, \quad X_k = \{x_t : t \in W_k\}, \tag{9}$$

where L denotes window length, s is the stride, k indexes the windows, and t indexes positions within each window.

Client-Side Pre-processing

- **Feature Encoding:** Categorical features are encoded numerically using one-hot encoding, while numerical features remain continuous.



Yes: Intrusion detected, proceed to alert and classify type of attack

No: Traffic is normal, no further action needed.

Fig. 4. Two-Stage LSTM Pipeline (TSLP) architecture for client-side intrusion detection in IoMT networks.

- **Handling Missing Values:** A limited number of samples contained missing or incomplete attribute values. To prevent the introduction of bias and avoid distortions in downstream learning processes, these instances were systematically removed. This step ensured that the dataset remained clean, internally consistent, and representative of the underlying distribution, thereby contributing to the stability and reliability of the model training.
- **Normalization:** Features undergo min-max scaling based on the statistics of the training set:

$$x'_{i,j} = \frac{x_{i,j} - a_j}{b_j - a_j}, \quad a_j = \min_{\text{train}} x_{i,j}, \quad b_j = \max_{\text{train}} x_{i,j} \tag{10}$$

- **Class Imbalance Mitigation:** The Synthetic Minority Oversampling Technique (SMOTE) is applied solely to the training dataset post-windowing to address class imbalance and prevent temporal leakage.

Shared Encoder An LSTM-based encoder processes each segmented window to generate hidden state representations:

$$(h_{k,1}, \dots, h_{k,L}) = \mathrm{LSTM}_{\theta_s}(X_k), \quad h_k^{\star} = \mathrm{pool}(h_{k,1:L}), \tag{11}$$

where $h_k^\star$ is a compact representation derived via pooling (e.g., mean-pooling).

- **Stage 1 - Binary Anomaly Detection:** A logistic regression classifier serves as an initial anomaly detection gate:

  $$p_k = \sigma(w^\top h_k^\star + b), \quad g_k = \mathbf{1}\{p_k \ge \tau\},$$

  with threshold τ optimized on validation data to balance recall and false-positive rates. Normal traffic ($g_k = 0$) exits early, reducing computational load.
- **Stage 2 - Temporal Aggregation:** Anomalous traffic from Stage 1 is summarized through temporal aggregation:

  $$r_k = \frac{1}{L}\sum_{t\in W_k} h_{k,t}, \quad R_k = [r_{k-M+1}, \dots, r_k] \in \mathbb{R}^{M\times d_h},$$

  where R_k captures short-range dynamics using a buffer of the latest M aggregated summaries, with r_k being the average of the hidden states $h_{k,t}$ for each time window W_k.
- **Stage 3 - Multi-class Classification:** Aggregated anomalous traffic is classified into specific attack categories via a secondary LSTM:

  $$\tilde{h}_k = \mathrm{LSTM}_{\theta_{mc}}(R_k), \quad \hat{y}_k = \mathrm{softmax}(U\tilde{h}_k + c).$$

Side Transfer Learning (STL) for Isolated Labels For rare or client-specific attack labels ℓ, STL binary classifiers are instantiated using the frozen shared encoder:

$$s_\ell(k) = \sigma(W_\ell^\top h_k^\star + b_\ell), \quad d_\ell(k) = \mathbf{1}\{s_\ell(k) \ge \tau_\ell\}. \tag{12}$$

STL classifiers undergo local fine-tuning using class-weighted binary cross-entropy or focal loss, with thresholds τ_ℓ optimized on client-specific validation data. STL parameters are securely transmitted back to the server, ensuring confidentiality of the raw IoMT data.

3.5. Client-side optimal model selection via weighted multi-criteria analysis

This section presents a weighted multi-criteria decision-making framework designed for optimal client-side model selection in intrusion detection within IoMT devices. Clients typically possess two models: a global FL model and a specialised TL model tailored for isolated, client-specific attack labels. The overall process is shown in Fig. 5, while Algorithm 3 details the procedure for selecting the best-performing model based on multiple evaluation criteria, such as accuracy, false alarm rate (FAR), and inference latency (TT).

Notation. Let $\mathcal{M} = \{m_{\mathrm{FL}}, m_{\mathrm{TL}}\}$ denote the set of available candidate models, with m_{FL} representing the global federated learning model and m_{TL} representing the specialized side transfer learning model. The set of evaluation criteria is given by $C = \{\mathrm{Accuracy}, \mathrm{FAR}, \mathrm{TT}\}$. For each criterion $c \in C$:

- $T_c > 0$ is the normalization target (e.g., $T_{\mathrm{Accuracy}} = 1$, policy-defined T_{FAR}, or latency target T_{TT}).
- $W_c \ge 0$ is the criterion weight satisfying $\sum_{c\in C} W_c = 1$.
- $\delta_c \in \{+1, -1\}$ indicates maximization $(+1)$ or minimization (-1).

Let $M_{i,c}$ represent the performance of model m_i on criterion c. A small positive constant $\varepsilon > 0$ is used to prevent numerical instability. Optionally, a subset $\mathcal{H} \subseteq C$ defines mandatory constraints models must meet to remain feasible.

Each criterion is normalized onto a unified higher-is-better scale through direction-aware normalization:

$$S_{i,c} = W_c \left(\frac{M_{i,c} + \varepsilon}{T_c + \varepsilon}\right)^{\delta_c} \tag{13}$$

Metrics for maximization (e.g., accuracy) use $\delta_c = +1$, while metrics for minimization (e.g., FAR, latency) use $\delta_c = -1$. The total utility score for each model is calculated as:

$$S_i = \sum_{c\in C} S_{i,c}, \quad \mathrm{BestModel} = \arg\max_{m_i \in \mathcal{M}} S_i. \tag{14}$$

This approach identifies the model offering the optimal overall trade-off across all evaluation criteria.

3.5.1. Algorithm 3 – Weighted multi-criteria model selection

3.6. Multi-criteria decision analysis via TOPSIS

The TOPSIS method is employed to classify IoMT devices based on operational metrics:

$$C = \{c_1 = \text{node safety (benefit)}, c_2 = \text{latency (cost)}, c_3 = \text{packet loss (cost)}, c_4 = \text{jitter (cost)}\}.$$

Device ranking utilizes the ideal $(\mathbf{v}^+)$ and anti-ideal $(\mathbf{v}^-)$ solutions to compute the closeness coefficient (C_i):

$$D_i^+ = \|\mathbf{V}_{i\cdot} - \mathbf{v}^+\|, \quad D_i^- = \|\mathbf{V}_{i\cdot} - \mathbf{v}^-\|, \quad C_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{15}$$

This standard formula does not include the epsilon term and correctly calculates the closeness coefficient. It categorizes devices as *Best*, *Acceptable*, or *Non-Acceptable*, enabling real-time adaptive decision-making and fault tolerance.

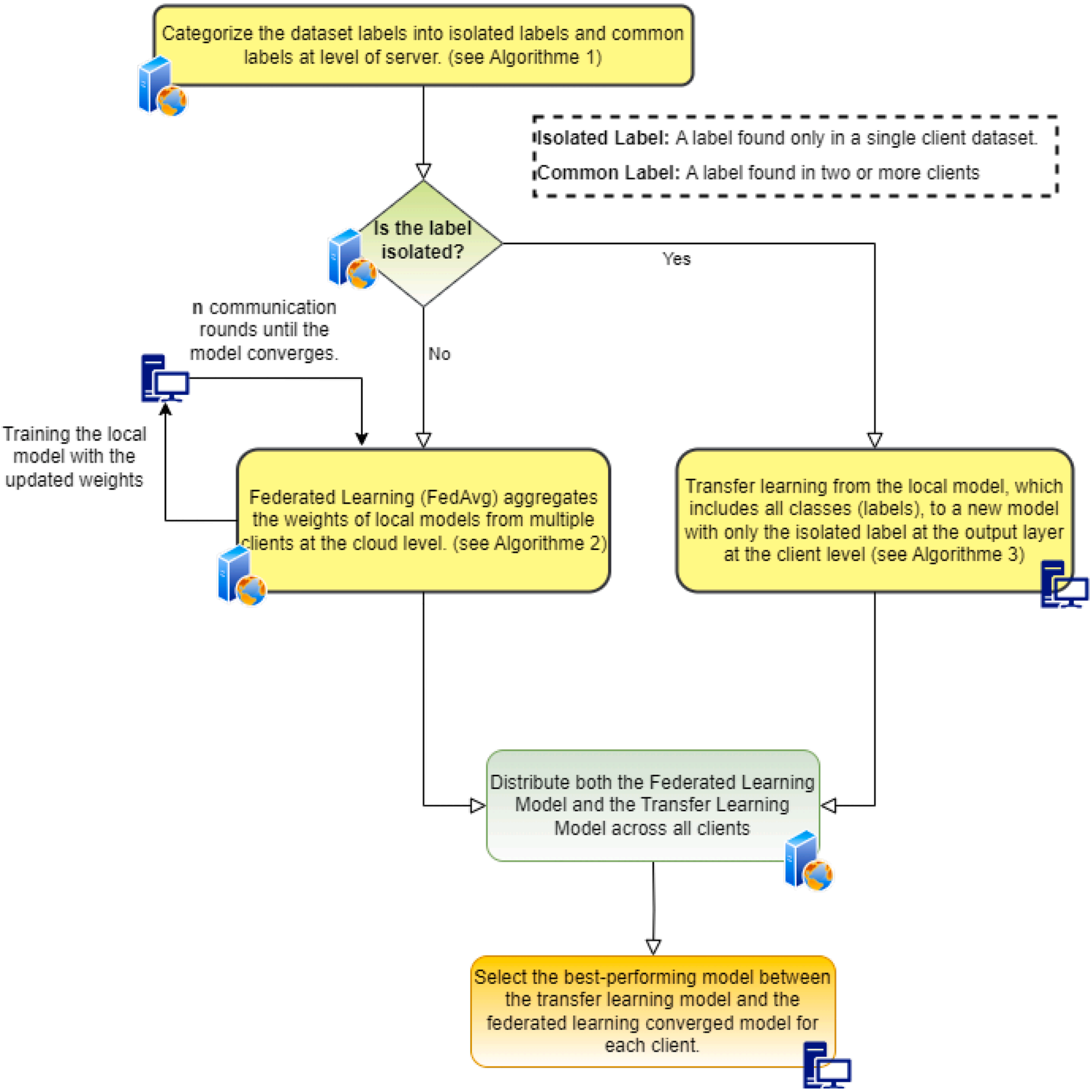

Categorize the dataset labels into isolated labels and common labels at level of server. (see Algorithme 1)
Isolated Label: A label found only in a single client dataset.
Common Label: A label found in two or more clients
Is the label isolated?
Yes
No
n communication rounds until the model converges.
Training the local model with the updated weights
Federated Learning (FedAvg) aggregates the weights of local models from multiple clients at the cloud level. (see Algorithme 2)
Transfer learning from the local model, which includes all classes (labels), to a new model with only the isolated label at the output layer at the client level (see Algorithme 3)
Distribute both the Federated Learning Model and the Transfer Learning Model across all clients
Select the best-performing model between the transfer learning model and the federated learning converged model for each client.

Fig. 5. Overview of Federated Transfer Learning Approach for IDS in IoMT.

3.7. Clinical risk override (CRI) and adaptive routing

The CRI independently drives clinical prioritization, overriding technical classifications when necessary:

$$\text{Classification} = \begin{cases} \text{Critical (dual-path routing and alerting)}, & \text{CRI} \geq 0.70, \\ \text{Warning (downgrade one level)}, & 0.50 \leq \text{CRI} < 0.70, \\ \text{Technical classification (retain)}, & \text{CRI} < 0.50, \\ \text{Non-Acceptable (quarantine)}, & \text{Technical fault.} \end{cases}$$

3.8. Real-time fault tolerance and response

To achieve robust, real-time fault tolerance and adaptive response capabilities, our proposed framework integrates a probabilistic classification strategy based on a Naïve Bayes (NB) classifier. Specifically, the NB classifier is trained on a comprehensive dataset

Algorithm 3 Client-Side Weighted Multi-Criteria Model Selection.

```
Input: Candidate models $\mathcal{M} = \{m_{\mathrm{FL}}, m_{\mathrm{TL}}\}$; criteria $C$; targets $\{T_c\}$; weights $\{W_c\}$ with $\sum_c W_c = 1$; directions $\{\delta_c\}$; performance metrics $\{M_{i,c}\}$; tolerance $\varepsilon > 0$; optional hard constraints $\mathcal{H} \subseteq C$.
Output: Optimal model selection $\mathrm{BestModel} \in \mathcal{M}$.
Verify $\sum_c W_c = 1$; if not, normalize $W_c$.
$\mathrm{BestModel} \leftarrow \mathrm{None}$;    $\mathrm{BestScore} \leftarrow -\infty$
for each model $m_i \in \mathcal{M}$ do
    if $\exists c \in \mathcal{H}$ violated by $m_i$ (maximized metric below target or minimized metric above target) then
        continue
    end if
    $\mathrm{ModelScore} \leftarrow 0$
    for each criterion $c \in C$ do
        $R \leftarrow \frac{M_{i,c}+\varepsilon}{T_c+\varepsilon}$
        $S_{i,c} \leftarrow W_c \cdot R^{\delta_c}$
        $\mathrm{ModelScore} \leftarrow \mathrm{ModelScore} + S_{i,c}$
    end for
    if $\mathrm{ModelScore} > \mathrm{BestScore}$ then
        $\mathrm{BestScore} \leftarrow \mathrm{ModelScore}$;    $\mathrm{BestModel} \leftarrow m_i$
    else if $\mathrm{ModelScore} = \mathrm{BestScore}$ then
        Apply tie-break criteria (e.g., inference latency, memory usage)
        if tie-break criteria favor $m_i$ then
            $\mathrm{BestModel} \leftarrow m_i$
        end if
    end if
end for
return BestModel
```

collected and aggregated through edge computing systems, incorporating diverse node performance metrics such as node safety, latency, packet loss, and jitter.

Utilising MCDA, the classifier systematically categorises nodes into three distinct operational classes: *Best*, *Acceptable*, and *Non-Acceptable*. These classifications reflect real-time node performance and reliability, calculated according to the probability distributions derived from current and historical device performance features.

The real-time classification process follows a probabilistic formulation based on the conditional probability framework of Naïve Bayes:

$$P(C_k|\mathbf{x}) = \frac{P(C_k)\prod_{i=1}^{m} P(x_i|C_k)}{\sum_j P(C_j)\prod_{i=1}^{m} P(x_i|C_j)}, \tag{16}$$

Where the range for i is explicitly shown in the product terms for both the numerator and denominator, ensuring clarity in the probabilistic calculation.

where $C_k \in \{\text{Best, Acceptable, Non-Acceptable}\}$ represents the class labels, $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ denotes the real-time feature vector consisting of aggregated node metrics, and $P(C_k|\mathbf{x})$ indicates the posterior probability of a node belonging to class C_k given features $\mathbf{x}$.

The training dataset is dynamically updated at the edge computing layer, ensuring that the Naïve Bayes classifier continuously adapts to evolving feature distributions and network conditions. This adaptive mechanism significantly enhances the classifier's ability to promptly and accurately identify shifts in device performance, cyberattacks, or emerging faults.

Consequently, the real-time classification outcomes directly inform adaptive routing and operational decisions within the IoMT network. Nodes classified as *Best* receive the highest-priority traffic and redundant routing paths, ensuring uninterrupted operation and high reliability. Nodes categorised as *Acceptable* are allocated standard-priority traffic, while *Non-Acceptable* nodes are immediately isolated, triggering rerouting mechanisms and alert systems for rapid mitigation.

Thus, the integration of real-time Naïve Bayes classification and MCDA-based node evaluation within our framework establishes a highly responsive and fault-tolerant intrusion detection and adaptive response capability, essential for maintaining continuous, safe, and secure IoMT network operations.

Our integrated methodology provides a robust solution combining federated learning, transfer learning, TSLP , MCDA-TOPSIS, real-time Naïve Bayes classification, and CRI-based clinical prioritisation. It significantly advances fault-tolerant, privacy-preserving, adaptive IoMT intrusion detection and network security, with substantial implications for patient safety and real-time healthcare operations.

Table 2
NF-UNSW-NB15-v2 dataset.

Label	Code	Count
Analysis	0	2299
Backdoor	1	2169
Benign	2	2,295,222
DoS	3	5794
Exploits	4	31,551
Fuzzers	5	22,310
Generic	6	16,560
Reconnaissance	7	12,779
Shellcode	8	1427
Worms	9	164
Total		**2390275**

4. Experimental results

This section presents a comprehensive evaluation of the proposed FTL-TSLP framework for intrusion detection in IoMT environments. The evaluation progresses systematically from centralised learning through federated learning under IID conditions to challenging non-IID scenarios, demonstrating the framework's robustness across diverse deployment contexts. Additionally, the fault-tolerance capabilities of the FTL-TSLP framework are highlighted, showcasing its ability to maintain system functionality in the presence of faults or attacks.

4.1. Datasets

The practical evaluation employs three benchmark datasets, each selected to represent distinct characteristics of the IoMT security landscape. These datasets capture the inherent heterogeneity and complexity of IoMT traffic patterns and are widely used in intrusion detection research. Collectively, they encompass a broad spectrum of real-world attack scenarios, ranging from general network intrusions to sophisticated threats targeting healthcare-specific devices and infrastructure. To mitigate the adverse effects of class imbalance on model performance, the Synthetic Minority Oversampling Technique (SMOTE) was applied. To preserve the integrity of the evaluation process and avoid data leakage, SMOTE was applied exclusively to the training set, ensuring that the validation and test sets remained unbiased and reflective of real-world intrusion distributions. This targeted oversampling strategy enhances the representation of minority intrusion classes, thereby improving model robustness and significantly strengthening the IDS's capability to detect both frequent and rare attack types.

4.1.1. NF-UNSW-NB15-v2 dataset

The NF-UNSW-NB15-v2 dataset, [46], as shown in Table 2, is a NetFlow-based cybersecurity dataset encompassing nine distinct attack categories: Exploits, Fuzzers, Generic, Reconnaissance, Denial-of-Service (DoS), Analysis, Backdoor, Shellcode, and Worms. The dataset was constructed by converting publicly accessible packet capture (pcap) files from the original UNSW-NB15 dataset [47] into a structured format comprising 43 features derived via the NetFlow v9 protocol, using the nprobe tool. The NF-UNSW-NB15-v2 dataset contains a total of 2,390,275 network flow records, among which 95,053 (3.98 %) represent attack instances, while the remaining 2,295,222 flows (96.02 %) constitute benign traffic.

The foundational dataset, UNSW-NB15, is a well-established resource within the network intrusion detection research community, developed and released in 2015 by the Cyber Lab of the Australian Centre for Cyber Security (ACCS). The original UNSW-NB15 dataset employed the IXIA PerfectStorm tool to simulate a combination of normal network traffic and diverse synthetic attack scenarios, providing researchers a comprehensive environment to evaluate network intrusion detection systems (NIDS). The selection of the NF-UNSW-NB15-v2 dataset is particularly relevant for evaluating IoMT intrusion detection systems, as it includes attack categories that closely mirror those faced by medical networks. Notably, the dataset contains backdoor attacks, which threaten patient data integrity, as well as reconnaissance activities-often precursors to targeted attacks on medical devices. These attack patterns are highly representative of the security challenges faced in modern healthcare environments, making this dataset an ideal candidate for testing the effectiveness of the proposed FTL-TSLP framework.

Through the use of this dataset, we aim to assess the ability of the FTL-TSLP model to detect both high-frequency and rare attacks while maintaining accuracy, particularly in the context of IoMT applications where security is paramount.

4.1.2. WUSTL EHMS 2020 dataset

The WUSTL-EHMS-2020 dataset is specifically designed as a cybersecurity resource to identify and mitigate vulnerabilities within Internet of Medical Things (IoMT) systems through realistic simulation of cyber-attacks within healthcare contexts [48,49]. Leveraging an integrated testbed that incorporates real-time patient biometric data alongside network traffic, this dataset primarily emphasizes two critical attack scenarios: Man-in-the-Middle (MITM) spoofing and data injection attacks. These scenarios pose significant threats by directly undermining the integrity and confidentiality of sensitive medical data, thus highlighting the dataset's importance for

Table 3
wustl-ehms-2020 Dataset.

Label	Code	Count
Data Alteration	0	922
Spoofing	1	1124
Normal	2	14,272
Total		**16318**

developing robust intrusion detection systems (IDS) tailored to healthcare environments [48]. As shown in Table 3, the WUSTL-EHMS-2020 dataset is organized into three primary classes-Normal, Data Alteration, and Spoofing-providing a representative distribution of benign and malicious IoMT traffic for intrusion detection research.

The dataset comprises over 16,000 labeled samples categorized into normal operations (87.5 %) and attack scenarios (12.5 %), effectively balancing benign and malicious data traffic. Each record contains 44 unique features, encompassing network flow metrics and real-time biometric data from patients. The integration of real-time biometric information sets this dataset apart from other IoT or IoMT datasets, providing deeper insights into how cybersecurity threats can significantly impact healthcare system performance and patient safety.

Crucial network performance metrics documented in the dataset include packet count, average packet size, and inter-arrival times. These metrics are supplemented with patient-specific biometric parameters, such as heart rate and blood oxygen levels, enabling researchers to authentically simulate and analyze the effects of network disruptions or malicious data manipulations on real-time patient monitoring.

The MITM spoofing attack scenario simulates conditions in which an adversary gains unauthorized access, intercepts, and modifies data transmitted between IoMT devices and healthcare servers or applications. Such attacks can mislead clinical decision-making, causing delayed or incorrect medical responses. Conversely, the data injection attack scenario involves introducing malicious or falsified data into legitimate data streams, compromising patient information accuracy and potentially misleading healthcare providers into making inappropriate clinical decisions.

The WUSTL-EHMS-2020 dataset is broadly applicable in developing advanced IDS through machine learning techniques aimed at detecting and counteracting cyber threats within IoMT infrastructures. Its distinct combination of biometric and network traffic data allows for sophisticated anomaly detection systems capable of identifying both network-level and physiological abnormalities associated with cyber-attacks or system malfunctions. This makes the dataset particularly valuable for anomaly detection research, facilitating the identification of unusual patterns indicative of cyber threats.

By incorporating authentic patient data with detailed simulated cyber-attack scenarios, the WUSTL-EHMS-2020 dataset allows researchers to comprehensively assess the potential impacts of cybersecurity threats on patient safety, medical device reliability, data integrity, and system latency. Furthermore, it supports investigations into medical personnel responses to potential false alerts generated by spoofing or data injection attacks, contributing to enhanced preparedness and resilience within healthcare cybersecurity frameworks.

4.1.3. CICIoMT dataset 2024

The CICIoMT-2024 dataset, developed by the Canadian Institute for Cybersecurity at the University of New Brunswick, constitutes a comprehensive resource designed explicitly to address cybersecurity vulnerabilities within the rapidly evolving Internet of Medical Things (IoMT) domain [25]. Given the increasing integration of IoMT technologies into healthcare infrastructures, securing medical devices and safeguarding communication privacy have emerged as critical research priorities. Consequently, the CICIoMT-2024 dataset is intended to facilitate focused research aimed at detecting and mitigating various cyber threats against IoMT systems.

IoMT systems encompass diverse interconnected medical devices, software applications, and related services that exchange data via the Internet to enhance healthcare delivery outcomes. Prominent examples include wearable medical devices, patient-monitoring systems, intelligent hospital beds, and automated medication dispensers. However, such devices frequently exhibit vulnerabilities due to limited computational resources and inadequate security architectures, making them particularly susceptible to cyberattacks. Potential security breaches could lead to severe consequences, including manipulation of sensitive medical data, disruption of essential healthcare services, or unauthorized disclosure of personal health information [25].

The CICIoMT-2024 dataset comprises network traffic data collected from 40 IoMT devices-25 real physical devices and 15 simulated devices representative of typical healthcare settings. These devices utilize multiple communication protocols, including Wi-Fi, Message Queuing Telemetry Transport (MQTT), and Bluetooth, accurately reflecting the heterogeneous nature of real-world IoMT environments. The dataset is structured into two primary directories: the first, titled the Bluetooth Traffic Directory, contains data from Bluetooth-enabled devices. This segment is particularly valuable given the widespread use of Bluetooth in wearable medical devices such as fitness trackers and heart rate monitors, thereby capturing both legitimate and malicious communications specific to the Bluetooth protocol.

A distinctive feature of CICIoMT-2024 is the inclusion of simulated scenarios covering 18 distinct cyber-attacks, effectively highlighting IoMT vulnerabilities. As shown in Table 4, the dataset encompasses a wide range of attack categories, including ARP Spoofing, MQTT-based DoS and DDoS attacks, TCP/IP DDoS, reconnaissance activities, and benign traffic. This rich class distribution provides extensive coverage of real-world IoMT threats and supports the development of advanced intrusion detection techniques capable of addressing heterogeneous cyber risks.

Table 4
CICIoMT-2024 Dataset.

Label1	**Label2**	**Code**	**Count**
ARP_Spoofing	ARP_Spoofing	0	16,047
Benign	Benign	1	192,732
MQTT-DDos-Connect_Flood MQTT-DDos-Publish_Flood	MQTT-DDos	2	200659
MQTT-DoS-Publish_Flood MQTT-DoS-Connect_Flood	MQTT-DoS	3	57149
MQTT-Malformed_Data	MQTT-Malformed_Data	4	5130
Recon-Port_Scan Recon-VulScan Recon-Ping_Sweep Recon-os_Scan	Recon	5	103726
TCP_IP-DDos-TCP TCP_IP-DDos-UDP TCP_IP-DDos-ICMP TCP_IP-DDos-SYN	TCP_IP-DDos	6	4779859
TCP_IP-DoS-TCP TCP_IP-DoS-UDP TCP_IP-DoS-ICMP TCP_IP-DoS-SYN	TCP_IP-DoS	7	1805529
Total			**7042831**

The dataset is primarily used for research involving machine learning and anomaly detection systems. Its comprehensive nature, integrating both actual and simulated devices, enables robust testing and validation across multiple cybersecurity scenarios. Beyond cybersecurity applications, the CICIoMT-2024 dataset offers significant value for research involving network analysis, traffic classification, and protocol optimization. The use of actual IoMT devices ensures the dataset authentically represents real operational behaviors in clinical healthcare environments.

CICIoMT-2024 represents a significant advancement in IoMT security research by providing an extensive and diverse collection of network traffic data encompassing both benign and malicious activities. This resource provides essential empirical evidence to address critical cybersecurity challenges in healthcare. As IoMT adoption continues to expand, datasets such as CICIoMT-2024 will serve as fundamental tools for enhancing the security and reliability of healthcare technology infrastructures [25].

4.2. Performance metrics and evaluation

To evaluate the performance of the proposed approach, this study employs widely recognized statistical metrics, namely *accuracy*, *precision*, *recall*, and *F1-score*, computed from the confusion matrix as follows.

Accuracy quantifies the overall correctness of the model:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{17}$$

Precision measures the proportion of correctly identified positive instances among all instances predicted as positive:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{18}$$

Recall (or sensitivity) evaluates the proportion of actual positive instances that are correctly identified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{19}$$

F1-Score represents the harmonic mean of precision and recall, providing a balanced measure between the two:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{20}$$

Where:

- TP (*True Positive*): Correctly identified anomalies.
- FP (*False Positive*): Normal instances incorrectly identified as anomalies.
- TN (*True Negative*): Correctly identified normal instances.
- FN (*False Negative*): Anomalies incorrectly identified as normal.

4.3. Model parameters and hyperparameters

Table 5 presents a detailed comparison of the model architectures and hyperparameter configurations employed in this study. The temporal architectures, specifically LSTM and GRU models, are selected for their capability to effectively capture the temporal

Table 5
Hyperparameter configurations and architectural specifications for baseline models.

Category	Parameter	MLP	LSTM	GRU	XGB
Architecture	Layer 1	Dense(256)	LSTM(256)	GRU(256)	—
	Layer 2	Dense(128)	LSTM(128)	GRU(128)	—
	Layer 3	Dense(64)	LSTM(64)	GRU(64)	—
	Output	Softmax(n_c)	Softmax(n_c)	Softmax(n_c)	—
Regularization	Dropout	0.40	0.40	0.40	—
	Batch Norm.	Yes	Yes	Yes	—
	L2 Penalty	—	0.01	0.01	—
Optimization	Optimizer	Adam	Adam	Adam	—
	Learning Rate	0.001	0.001	0.001	0.1
	Batch Size	512	64	64	—
Training	Epochs	100	100	100	—
	Loss Function	Sparse Categorical CE			softmax
Tree-Specific	n_estimators	—	—	—	100
	max_depth	—	—	—	6
	subsample	—	—	—	0.8
	colsample_bytree	—	—	—	0.8
	tree_method	—	—	—	hist
	eval_metric	—	—	—	mlogloss

1. *Note*: n_c = number of classes; CE = Cross-Entropy; XGB = XGBoost; Batch Norm. = Batch Normalization.

dynamics inherent in IoMT attack patterns. The MLP model serves as a non-temporal baseline due to its simpler feedforward structure and comparatively lower computational complexity. Additionally, XGBoost is included as a tree-based ensemble baseline to evaluate the effectiveness of gradient boosting methods in intrusion detection.

4.4. *Performance analysis of centralized learning*

This subsection rigorously evaluates the performance of the proposed Two-Stage LSTM Pipeline (TSLP) architecture under centralized training conditions incorporating temporal aggregation. The evaluation establishes essential baseline metrics and systematically investigates the effects of varying temporal aggregation intervals on intrusion detection accuracy and computational efficiency, aiming to optimize predictive performance and training efficiency.

4.4.1. *Temporal aggregation impact on accuracy and efficiency*

Temporal aggregation significantly improved the performance metrics of the LSTM-based IDS across all evaluated datasets, as illustrated comprehensively in Table 6. Notably, the NF-UNSW-NB15-V2 dataset exhibited the most pronounced improvement, with accuracy increasing from 87.18 % at no aggregation to 99.28 % at 30-second intervals. Precision also followed a parallel enhancement, rising from 89.80 % to 99.29 %. Comparable improvements were observed on the CICIoMT-2024 dataset, where accuracy and precision improved from baseline values of approximately 90.50 % to 98.97 % at 30-second aggregation intervals. The WUSTL-EHMS-2020 dataset consistently maintained exceptionally high performance, achieving 100 % accuracy and precision at both 5-second and 30-second intervals.

Fig. 6 illustrates the monotonic relationship between aggregation intervals and classification performance across the evaluated metrics. The improvement curves demonstrated rapid gains between 0 and 15 seconds, after which performance metrics plateaued. This trend remained consistent across accuracy, precision, recall, and F1-score metrics for all three datasets, suggesting an optimal trade-off existed between temporal granularity and classification accuracy.

Furthermore, temporal aggregation significantly reduced computational demands in both training and testing phases, as highlighted in Fig. 6. The NF-UNSW-NB15-V2 dataset experienced the most significant efficiency gains, with training time decreasing by 84.8 %, from 32,319.75 seconds to 4,902.53 seconds, as aggregation intervals increased from 0 to 30 seconds. Testing duration demonstrated an even more substantial reduction of 93.1 %, decreasing from 260.89 seconds to 18.11 seconds. Similar efficiency enhancements were observed on the CICIoMT-2024 dataset, where training time decreased by 97.2 % (from 12,552.78 to 352.92 seconds), and testing time by 97.5 % (from 142.42 to 3.55 seconds). The WUSTL-EHMS-2020 dataset, despite being smaller in scale, achieved notable efficiency improvements with reductions of 99.3 % and 99.1 % in training and testing durations, respectively.

The computational efficiency gains presented in Fig. 7, utilising logarithmic scaling, underscore the exponential nature of the relationship between temporal aggregation intervals and processing times. The consistent downward trends observed across all datasets confirmed that temporal aggregation not only enhanced IDS performance but also significantly reduced computational resource requirements, making this approach particularly advantageous for deployment in resource-constrained IoMT environments.

4.4.2. *Binary versus multi-class detection comparison*

Table 7 presents a detailed comparative analysis of binary and multi-class classification approaches employing LSTM models for intrusion detection. Binary classification consistently outperformed multi-class classification across all evaluated datasets, particularly demonstrating significant advantages in scenarios characterised by heterogeneous network traffic.

Table 6
Temporal Aggregation Effects on LSTM Classification Performance: Cross-Dataset Comparative Analysis.

T (s)	Acc (%)	Prec (%)	Rec (%)	F1 (%)	Train Time (s)	Test Time (s)
NF-UNSW-NB15-V2						
0	87.18	89.80	87.18	87.91	32,319.75	260.89
5	90.11	91.87	90.11	90.48	29,427.07	107.91
15	94.93	95.40	94.93	94.99	9,705.09	36.14
30	99.28	99.29	99.28	99.28	4,902.53	18.11
CICIoMT-2024						
0	90.50	90.57	90.50	90.50	12,552.78	142.42
5	95.73	95.74	95.73	95.73	2,216.72	17.63
15	97.42	97.49	97.42	97.41	743.87	6.34
30	98.97	98.97	98.97	98.97	352.92	3.55
WUSTL-EHMS-2020						
0	99.95	99.95	99.95	99.95	4,174.80	11.00
5	99.99	99.99	99.99	99.99	856.16	2.02
15	99.83	99.83	99.83	99.83	48.24	0.14
30	99.99	99.99	99.99	99.99	30.99	0.10

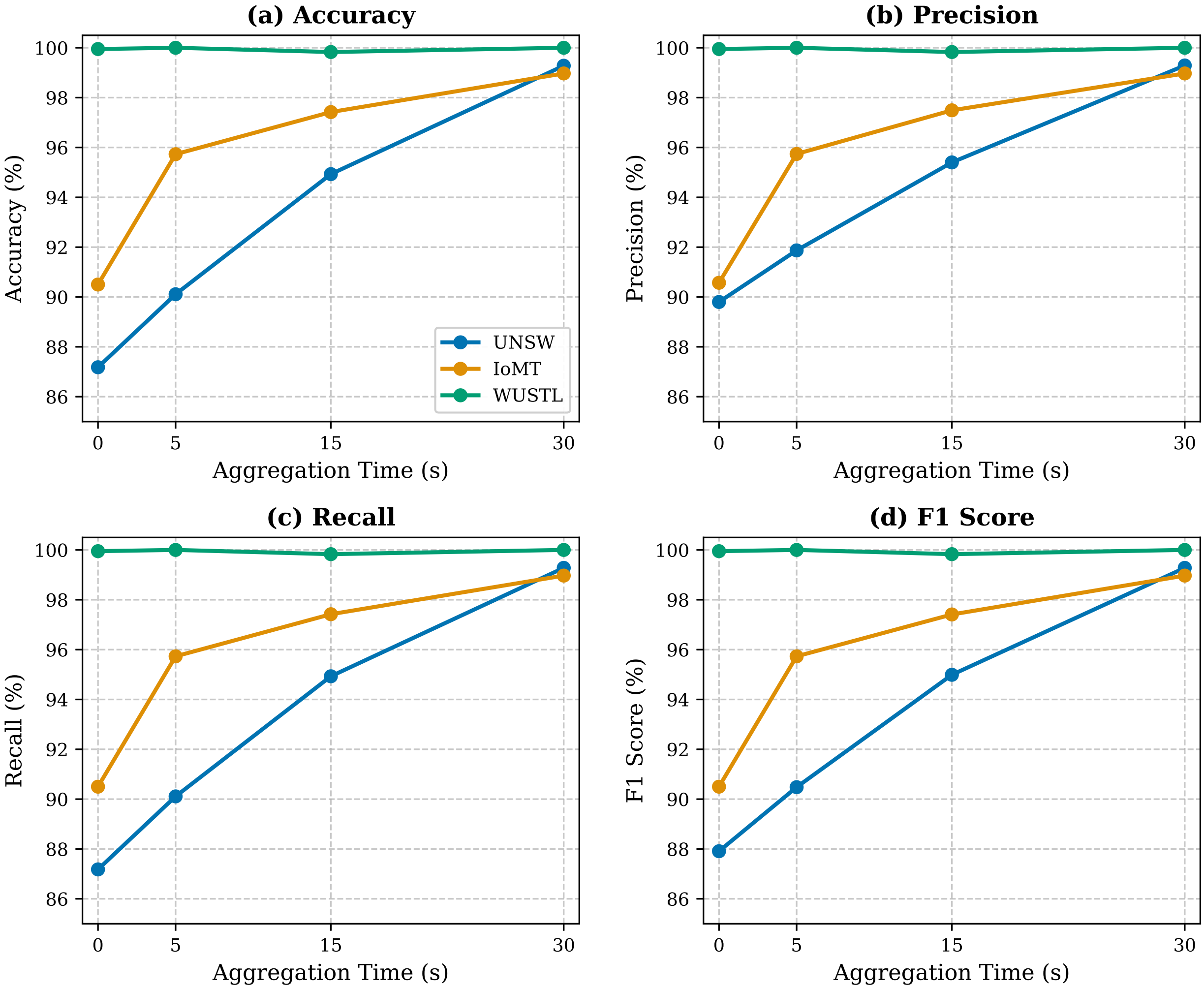


Fig. 6. Performance Metrics and Computational Efficiency as Functions of Temporal Aggregation.

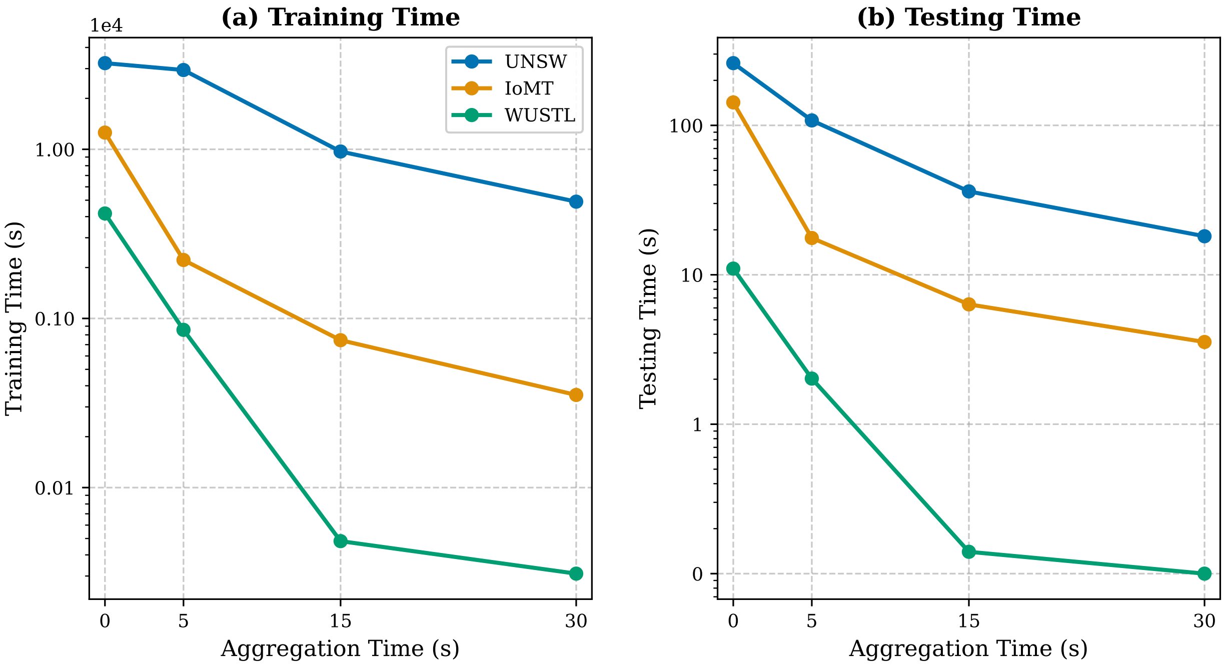


Fig. 7. Computational Efficiency Gains Through Temporal Aggregation: Training and Testing Time Analysis.

Table 7
Comparative Analysis of Binary and Multi-Class LSTM.

Dataset	Method	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
NF-UNSW-NB15-V2	Binary	99.99	99.99	99.99	99.99
	Multi-class	92.84	93.51	92.84	93.02
CICIoMT-2024	Binary	99.74	99.54	99.94	99.74
	Multi-class	91.62	91.64	91.62	91.63
WUSTL-EHMS-2020	Binary	99.98	99.99	99.96	99.98
	Multi-class	99.95	99.95	99.95	99.95

The NF-UNSW-NB15-V2 dataset exhibited the most pronounced difference in performance. Binary classification achieved perfect detection rates of 99.99 % across accuracy, precision, recall, and F1-score metrics. Conversely, multi-class classification attained only 92.84 % accuracy and a 93.02 % F1-score, indicating a substantial accuracy gap of 7.16 percentage points. This notable discrepancy highlighted the comparative ease of distinguishing regular traffic from malicious activity compared to accurately categorising multiple distinct attack types within a complex network environment.

The CICIoMT-2024 dataset followed a similar trend, albeit with a slightly narrower performance differential. Binary classification achieved an accuracy of 99.74 %, accompanied by balanced precision (99.54 %) and recall (99.94 %), reinforcing its robust detection capabilities. Multi-class classification exhibited considerably lower accuracy (91.62 %), representing an 8.12 percentage point gap. Consistency in precision and recall metrics (91.64 % and 91.62 %, respectively) indicated the absence of systematic bias toward either false positives or false negatives, underscoring the inherent challenges associated with precise categorisation of IoMT-specific attack types.

In contrast, the WUSTL-EHMS-2020 dataset showed minimal differences between binary and multi-class classification performances. Binary classification recorded an accuracy of 99.98 %, closely matched by multi-class classification at 99.95 %, demonstrating a negligible differential of 0.03 percentage points. Identical precision and recall values (99.95 %) in multi-class classification suggested that attack patterns within this specialised medical monitoring dataset possessed sufficiently distinct characteristics, facilitating accurate classification irrespective of the detection approach.

Moreover, the integration of temporal aggregation alongside multi-class classification in the second stage, following an initial binary classification in the first stage, established a two-stage detection pipeline that optimised both detection accuracy and computational efficiency. This TSLP approach, beginning with rapid binary anomaly detection followed by refined temporal aggregation for detailed attack categorisation, enabled real-time intrusion detection essential for resource-constrained IoMT deployments. The substantial improvements detailed in Section 4.4.1, coupled with the proven responsiveness and superior performance of binary classification, effectively validated this two-stage pipeline methodology as both accurate and computationally efficient for real-time IoMT intrusion detection without delays.

4.4.3. Comparative analysis of TSLP and baseline models

Table 8 presents a comprehensive evaluation of the proposed TSLP against established baseline models: MLP, GRU, LSTM, and XGBoost (XGB). The comparative analysis was conducted across three heterogeneous IoMT datasets. The results consistently demonstrate that TSLP achieves a superior balance between detection performance and computational efficiency, thereby validating its suitability for real-time intrusion detection in resource-constrained IoMT environments.

On the NF-UNSW-NB15-V2 dataset, TSLP demonstrated superior balanced performance, achieving an F1-score of 99.63 % with corresponding precision and recall values of 99.63 %. This result is particularly significant when contrasted with XGBoost's performance. Although XGBoost achieved high accuracy (99.12 %), its F1-score deteriorated substantially to 66.58 %, accompanied by a precision of only 64.87 %. This pronounced disparity between accuracy and F1-score indicates a critical deficiency in addressing class imbalance, likely manifesting as an elevated false-positive rate that would undermine practical deployment. The recurrent baseline models, GRU (F1-score: 93.32 %) and LSTM (F1-score: 93.02 %), were also substantially outperformed by TSLP while incurring considerably higher training times of 14,538.94 seconds and 16,465.25 seconds, respectively. The non-temporal MLP baseline exhibited the poorest performance with an F1-score of 85.06 %, confirming the necessity of temporal modeling for effective intrusion detection in this domain.

The CICIoMT-2024 dataset further underscored TSLP's advantages in handling complex IoMT-specific attack scenarios. TSLP achieved a near-perfect F1-score of 99.82 %, whereas alternative deep learning baselines exhibited substantial performance degradation. The standard LSTM attained an F1-score of 91.63 %, while GRU's performance deteriorated markedly to 64.34 %, suggesting that its simplified gating mechanism is insufficient for capturing the intricate temporal patterns characteristic of this dataset. XGBoost emerged as the strongest baseline; however, its F1-score of 93.16 % remained 6.66 percentage points below TSLP, with notably lower precision (90.35 %) indicative of suboptimal class discrimination. Beyond superior detection metrics, TSLP's pipeline architecture demonstrated exceptional computational efficiency, achieving a 64.6 % reduction in training time and an 83.8 % reduction in inference time relative to standard LSTM. These improvements are particularly critical for real-time deployment scenarios where both accuracy and response latency are paramount.

On the WUSTL-EHMS-2020 dataset, all evaluated models achieved near-perfect detection performance (F1-score > 99.84 %). This convergence in detection accuracy is likely attributable to the dataset's reduced temporal complexity and well-separated class distributions, which enable even non-temporal models such as MLP and XGBoost to achieve effective classification. Under these conditions, computational efficiency emerged as the primary differentiating factor. XGBoost demonstrated the fastest training time (1.11 seconds), benefiting from its tree-based ensemble architecture. However, TSLP proved to be the most efficient among temporal models, requiring only 264.03 seconds for training—representing reductions of 35.8 % and 31.4 % compared to GRU and LSTM, respectively. Moreover, TSLP achieved an inference time of 0.04 seconds, corresponding to a 96 % reduction relative to standard LSTM (1.00 seconds), while remaining competitive with XGBoost (0.01 seconds). This exceptional inference efficiency is particularly valuable for edge-deployed IoMT security systems operating under strict latency constraints.

The comparative analysis presented in Table 8 reveals fundamental trade-offs between computational efficiency and robust, balanced detection performance. XGBoost, while demonstrating exceptional training efficiency across all datasets, exhibited a critical vulnerability: significantly degraded F1-scores and precision on the more complex NF-UNSW-NB15-V2 and CICIoMT-2024 datasets. This deficiency underscores XGBoost's inherent limitation in modeling sequential attack patterns—a capability essential for reliable intrusion detection systems that must identify temporally correlated malicious behaviors.

The standard LSTM and GRU architectures, despite their temporal modeling capabilities, were consistently outperformed by TSLP in both detection accuracy and computational efficiency. This performance gap can be attributed to TSLP's architectural innovations. Specifically, the two-stage pipeline—comprising binary classification for initial traffic filtering (Stage 1) followed by temporal aggregation and multi-class classification (Stage 2)—enables efficient extraction and exploitation of long-range temporal dependencies while maintaining computational tractability. The temporal aggregation mechanism in particular allows TSLP to capture complex attack patterns that unfold over extended time windows, a capability that proves decisive on datasets featuring sophisticated, multi-step attack sequences.

Furthermore, TSLP demonstrated significant reductions in both training and testing times across all datasets, confirming its computational efficiency. These improvements are attributed to TSLP's two-stage pipeline architecture and temporal aggregation, which efficiently capture long-term dependencies while minimizing computational costs. The incorporation of binary classification in the first stage further enhances real-time intrusion detection without compromising response time, making TSLP a highly effective solution for IoMT security.

4.5. Evaluation of federated learning

This section presents a comprehensive analysis of federated learning performance, starting with idealized Independent and Identically Distributed (IID) conditions and progressing to more realistic non-IID scenarios. These scenarios reflect the heterogeneous nature of healthcare networks, where data may be distributed unevenly across clients.

4.5.1. Federated learning evaluation under IID conditions

The FL evaluation commenced under IID conditions using the FedAvg algorithm to establish baseline performance metrics for the TSLP model. The experimental setup involved 10 federated rounds, each comprising 20 local epochs, with uniformly distributed data shards across participating clients.

Table 8
Performance Evaluation and Comparison of TSLP with Baseline Models.

Arch.	T (s)	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)	Train Time (s)	Test Time (s)
NF-UNSW-NB15-V2							
MLP	0	84.33	87.29	84.33	85.06	816.11	30.99
GRU	0	93.20	93.70	93.20	93.32	14,538.94	67.76
LSTM	0	92.84	93.51	92.84	93.02	16,465.25	65.64
XGBoost	0	99.12	64.87	77.01	66.58	748.56	3.06
TSLP	**30**	**99.63**	**99.63**	**99.63**	**99.63**	**8,995.80**	**18.94**
CICIoMT-2024							
MLP	0	88.48	88.63	88.48	88.50	1,084.29	35.17
GRU	0	74.34	79.35	74.34	64.34	17,835.57	45.64
LSTM	0	91.62	91.64	91.62	91.63	22,146.85	90.33
XGBoost	0	99.77	90.35	97.73	93.16	1,138.25	7.18
TSLP	**30**	**99.82**	**99.82**	**99.82**	**99.82**	**7,838.79**	**14.62**
WUSTL-EHMS-2020							
MLP	0	99.99	99.99	99.99	99.99	14.39	0.31
GRU	0	99.96	99.97	99.96	99.96	411.63	1.16
LSTM	0	99.95	99.95	99.95	99.95	384.99	1.00
XGBoost	0	99.97	99.82	99.85	99.84	1.11	0.01
TSLP	**30**	**99.99**	**99.99**	**99.99**	**99.99**	**264.03**	**0.04**

Table 9
Impact of Temporal Aggregation on Performance Metrics across Datasets: A Federated Learning Approach using FedAVG and LSTM.

Method **NF-UNSW-NB15-V2 Dataset**	**T (s)**	**Acc (%)**	**Prec (%)**	**Rec (%)**	**F1 (%)**
Federated Learning (FedAVG) using LSTM	0	89.87	91.14	89.87	90.22
Federated Learning (FedAVG) using LSTM	30	95.46	95.64	95.46	95.50
CICIoMT-2024 Dataset					
Federated Learning (FedAVG) using LSTM	0	91.45	91.47	91.45	91.45
Federated Learning (FedAVG) using LSTM	30	99.82	99.82	99.82	99.82
WUSTL-EHMS-2020 Dataset					
Federated Learning (FedAVG) using LSTM	0	99.99	99.99	99.99	99.99
Federated Learning (FedAVG) using LSTM	30	99.99	99.99	99.99	99.99

Table 9 summarises the significant performance gains achieved through temporal aggregation. For the NF-UNSW-NB15-V2 dataset, TSLP accuracy improved notably from 89.87 % without aggregation to 95.46 % with 30-second intervals, marking a 5.59 percentage point enhancement. Precision correspondingly rose from 91.14 % to 95.64 %, and F1-score increased from 90.22 % to 95.50 %, affirming the efficacy of temporal aggregation in mitigating distributed learning challenges.

The CICIoMT-2024 dataset demonstrated even greater improvement, with TSLP accuracy rising sharply to 99.82 %, an increase of 8.37 percentage points from the baseline LSTM performance of 91.45 %. Consistent precision, recall, and F1-score metrics (each at 99.82 %) indicated balanced and unbiased detection performance, emphasising temporal aggregation's capability to manage complex IoMT traffic patterns effectively.

For the WUSTL-EHMS-2020 dataset, accuracy slightly increased from 99.99 % to 99.99 % due to temporal aggregation. While accuracy improvements were minimal, computational efficiency significantly benefited from reduced communication overhead and accelerated convergence.

Fig. 8 illustrates convergence dynamics across datasets. Specifically, for NF-UNSW-NB15-V2 (Fig. 8a), temporal aggregation enabled rapid convergence within 2–3 rounds to about 95 % accuracy, outperforming the non-aggregated baseline, which plateaued around 85 % after 5–6 rounds. The loss curves confirmed more stable optimisation with aggregation.

For CICIoMT-2024 (Fig. 8b), the aggregated model rapidly reached 99 % accuracy within three rounds, considerably faster than the baseline method, which stabilised around 90 % after five rounds. Loss reduction was notably substantial, further validating aggregation's advantages.

In the WUSTL-EHMS-2020 dataset (Fig. 8c), temporal aggregation consistently maintained stable performance, contrasting with noticeable instability and fluctuations observed in non-aggregated loss curves during later rounds. This highlighted the crucial role of temporal aggregation in ensuring optimisation stability.

Overall, temporal aggregation substantially enhanced FL dynamics, improving accuracy, accelerating convergence, reducing communication overhead, and ensuring model stability.

4.5.2. Evaluation of FTL-TSLP under Non-IID conditions

Real-world IoMT deployments inherently exhibit non-IID data characteristics, posing significant challenges to traditional FL methodologies. Healthcare institutions typically encounter distinct threat profiles influenced by factors such as specialisation, ge-

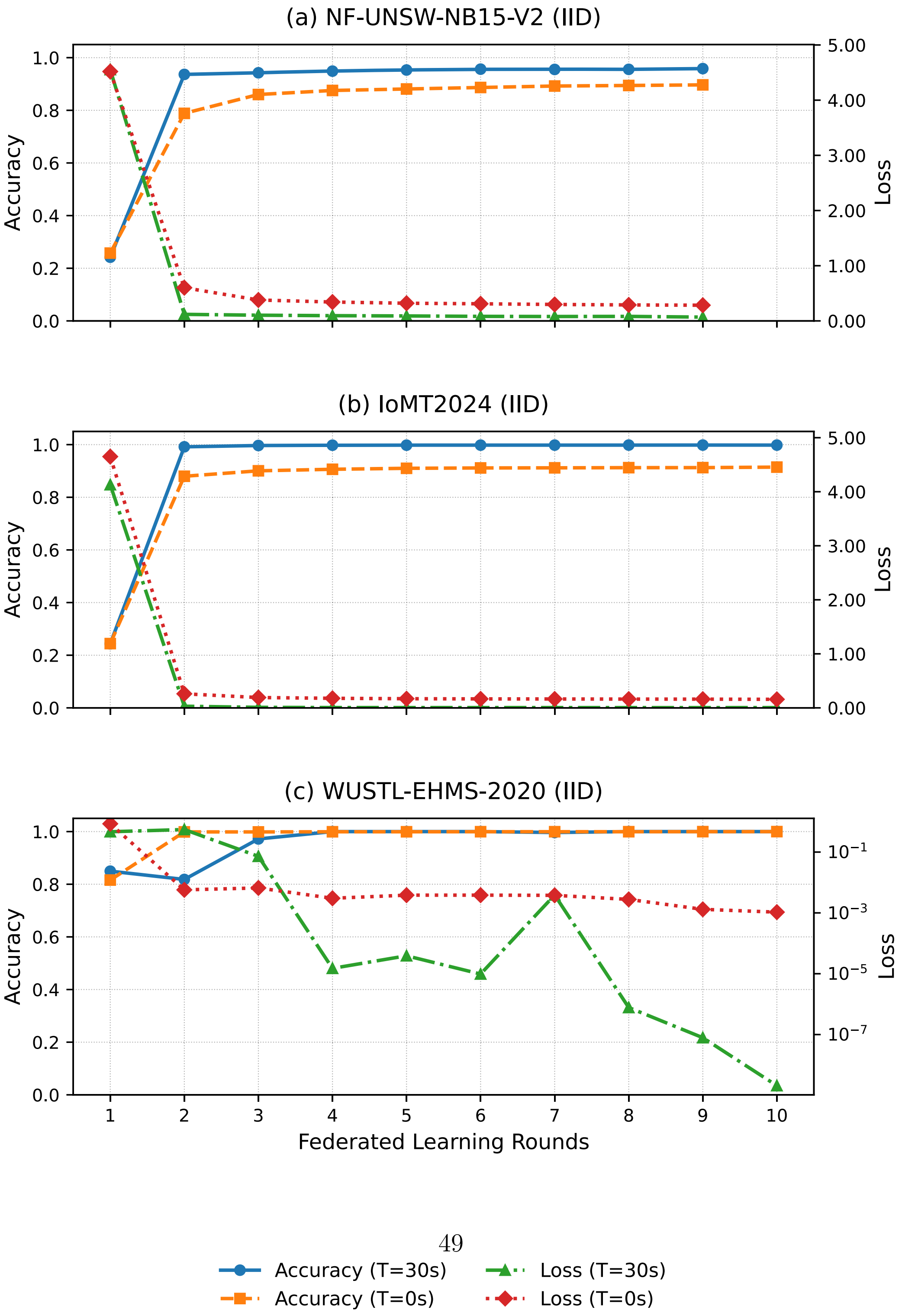


Fig. 8. Comparison of accuracy and loss across different temporal aggregation intervals in Federated Learning using IID data. The figure illustrates the impact of 0s and 30s aggregation on training performance across multiple datasets.

Table 10
Per-Label Performance Metrics on Non-IID Data.

Method	Metric	Label				
		0	1	2	3	4
NF-UNSW-NB15-V2						
FedAvg	Acc/F1	99.9/77.0	40.4/57.5	100/100	0/0	0/0
TL(Binary)	Acc/F1	99.1/95.9	100/99.0	100/100	99.9/99.4	99.9/99.3
Hybrid-Mode-1	Acc/F1	99.9/77.0	40.4/57.5	100/100	99.9/99.4	99.9/99.3
Hybrid-Mode-2	Acc/F1	99.9/77.0	100/99.0	100/100	99.9/99.4	99.9/99.3
CICIoMT-2024						
FedAvg	Acc/F1	99.1/95.5	91.5/95.1	100/100	0/0	0/0
TL(Binary)	Acc/F1	99.8/99.1	99.7/94.3	100/99.3	100/97.3	99.7/99.1
Hybrid-Mode-1	Acc/F1	99.1/95.5	91.5/95.1	100/100	100/97.3	99.7/99.1
Hybrid-Mode-2	Acc/F1	99.1/95.5	97.8/94.3	100/100	100/97.3	99.7/99.1

ographic location, and patient demographics, resulting in highly heterogeneous data distributions among federated participants. This section presents a comprehensive evaluation of the FTL-TSLP framework under substantial data heterogeneity conditions.

Performance on Isolated Labels. This subsection assessed the performance of the FTL-TSLP model in non-IID scenarios, comparing four FL strategies: standard FedAvg, Transfer Learning (TL) for binary classification, and two hybrid modes. Hybrid Mode 1 employed TL exclusively for isolated labels, whereas Hybrid Mode 2 used TL for isolated labels and adaptively selected between TL and FedAvg based on optimal performance across all labels using Algorithm 3. The experimental setup involved three healthcare institutions, each characterised by unique attack patterns:

- **Client 1 (General Hospital):** Common labels 0, 1, and 2.
- **Client 2 (Cardiac Facility):** Common labels 0 and 2, isolated label 3.
- **Client 3 (Pediatric Institution):** Common labels 1 and 2, isolated label 4.

The experiment comprised 10 federated rounds, each with 20 local epochs, utilising the LSTM(30) architecture.

Table 10 demonstrated substantial performance variations among the evaluated strategies, especially concerning isolated labels. Standard FedAvg failed to classify isolated labels, achieving 0.00 % accuracy for Labels 3 and 4 across both NF-UNSW-NB15-V2 and CICIoMT-2024 datasets. However, FedAvg delivered satisfactory performance on common labels, with Label 2 achieving 99 % accuracy due to its consistent presence across multiple clients.

In contrast, TL consistently provided superior accuracy, exceeding 99 % for both isolated and common labels across all datasets. Specifically, TL attained accuracies of 99.87 % and 99.85 % for isolated Labels 3 and 4 on the NF-UNSW-NB15-V2 dataset, confirming its efficacy in managing isolated label scenarios through binary classification.

Hybrid Mode 1 effectively combined FedAvg for common labels and TL for isolated labels, preserving FedAvg's efficiency while significantly improving accuracy for isolated labels. Hybrid Mode 2 further enhanced adaptability by selectively applying TL to both isolated and poorly performing common labels. Notably, this approach improved Label 1 accuracy from 40.37 % to 99.97 %, providing a balanced, performance-oriented solution.

The results obtained from the CICIoMT-2024 dataset mirrored these outcomes, reinforcing the inadequacy of FedAvg for isolated labels and underscoring the effectiveness of TL. Hybrid strategies demonstrated clear advantages in effectively handling heterogeneous IoMT data.

B. Computational Complexity Analysis of FL Strategies. The computational complexity of each strategy was evaluated in terms of storage, inference, communication, temporal, and spatial complexities, as illustrated in the following table. The parameters used for complexity calculations are defined as follows:

- d: Model dimension
- n: Total number of labels
- k: Number of isolated labels
- p: Number of poorly performing labels
- r: Number of federated rounds
- m: Number of clients

Table 11 summarises the computational complexity of various FL strategies, highlighting the critical trade-offs between performance and resource usage:

- **FedAvg:** Exhibited the lowest complexity ($O(r)$ temporal, $O(d)$ spatial), but delivered insufficient accuracy for isolated labels.
- **TL (Binary):** Demonstrated the highest complexity ($O(r \times n)$ temporal, $O(n \times d)$ spatial), limiting its practical feasibility for resource-constrained devices.
- **Hybrid Mode 1:** Offered moderate complexity ($O(r)$ temporal, $O((1 + k) \times d)$ spatial), efficiently addressing isolated labels.

Table 11
Computational Complexity Analysis of Federated Learning Strategies for TSLP -Based Intrusion Detection.

Strategy	Number of Models per Client	Temporal Complexity	Spatial Complexity	Complexity Level
FedAvg	1 (global model)	$\mathcal{O}(r)$	$\mathcal{O}(d)$	Low
TL	n (binary classifiers)	$\mathcal{O}(r \times n)$	$\mathcal{O}(n \times d)$	High
Hybrid Mode 1	$1 + k$ (global + isolated)	$\mathcal{O}(r)$	$\mathcal{O}((1+k) \times d)$	Medium
Hybrid Mode 2	n (adaptive selection)	$\mathcal{O}(r \times n)$	$\mathcal{O}(n \times d)$	High

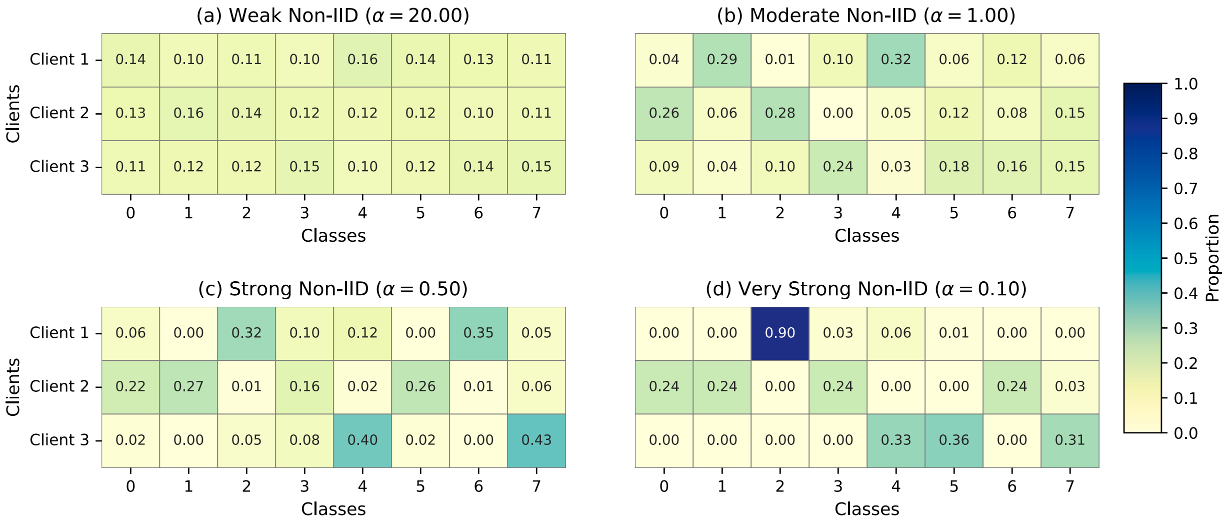


Fig. 9. Class distribution across clients for the CICIoMT-2024 dataset using Dirichlet-based label allocation.

- **Hybrid Mode 2:** Presented adaptive complexity depending on label performance, effectively balancing resource consumption and detection accuracy.

The proposed FTL-TSLP framework successfully balanced performance and complexity, achieving overall accuracies of 95.46 % on the NF-UNSW-NB15-V2 dataset and 99.82 % on the CICIoMT-2024 dataset. Its two-stage pipeline architecture facilitated optimised feature sharing across related attack categories, significantly reducing computational redundancy while maintaining high accuracy for isolated labels. This balanced methodology underscored the practical suitability of FTL-TSLP for real-world IoMT deployments, effectively managing accuracy requirements and computational constraints.

4.5.3. Performance evaluation under dirichlet-based Non-IID conditions

In the previous subsection (Section 4.5.2, "Addressing Isolated Labels with FTL-TSLP"), labels were manually allocated to simulate non-IID conditions. In contrast, this subsection employs a systematic and automated approach using Dirichlet-based partitioning, which provides a more rigorous evaluation of the FTL-TSLP framework's robustness and generalizability under realistic non-IID scenarios.

The Dirichlet distribution serves as a robust framework for simulating heterogeneous data partitions, with a concentration parameter (α) controlling the degree of statistical heterogeneity. The parameter α was varied across four levels: $\alpha \in \{20.0, 1.0, 0.5, 0.1\}$, transitioning from near-homogeneous distributions ($\alpha = 20.0$) to extreme heterogeneity ($\alpha = 0.1$), thereby encompassing a broad spectrum of federated learning scenarios.

Class allocation patterns across federated clients under different Dirichlet parameters were analyzed using the CICIoMT-2024 dataset (Fig. 9). Under weak heterogeneity ($\alpha = 20.0$), class proportions were relatively uniform with minimal variance. As heterogeneity increased ($\alpha = 1.0$), variability in class proportions grew significantly, resulting in broader ranges (with a maximum of approximately 0.32 and a minimum approaching zero). At stronger heterogeneity ($\alpha = 0.5$), pronounced client-specific data imbalances emerged, with class proportions fluctuating between 0.00 and 0.43. In scenarios of extreme heterogeneity ($\alpha = 0.1$), the class distributions became highly skewed, with certain classes becoming overwhelmingly dominant within individual client datasets, reaching proportions as high as 0.90.

These findings underscore the effectiveness of Dirichlet-based partitioning in simulating realistic, diverse non-IID conditions, providing a robust foundation for the evaluation of federated learning methodologies.

C. Baseline FedAvg-TSLP Performance on CICIoMT-2024 Dataset. Table 12 highlights the significant impact of increasing data heterogeneity on the baseline FedAvg-TSLP model's performance with the CICIoMT-2024 dataset. Under low heterogeneity conditions ($\alpha = 20$), the model achieved an average accuracy of 96.01 %. Despite robust overall performance, notable disparities emerged at the

Table 12
Performance Evaluation of the FedAvg-TSLP (30s) Model on the CICIoMT-2024 Dataset under Varying Dirichlet Alpha Settings.

Alpha	Method	Metric	C0	C1	C2	C3	C4	C5	C6	C7	Avg
20	FedAvg-TSLP (30s)	Accuracy	99.75	98.46	99.99	99.47	99.99	98.06	73.06	99.26	96.01
		Precision	96.78	99.86	99.47	99.99	99.98	99.88	98.89	78.61	96.69
		Recall	99.75	98.46	99.99	99.47	99.99	98.06	73.06	99.26	96.01
		F1 Score	98.24	99.15	99.74	99.73	99.99	98.96	84.04	87.74	95.95
1	FedAvg-TSLP (30s)	Accuracy	99.99	99.32	99.99	99.40	99.99	98.89	99.11	99.35	99.51
		Precision	98.35	99.87	99.41	99.99	99.96	99.98	99.39	99.16	99.51
		Recall	99.99	99.32	99.99	99.40	99.99	98.89	99.11	99.35	99.51
		F1 Score	99.17	99.59	99.70	99.70	99.98	99.43	99.25	99.26	99.51
0.5	FedAvg-TSLP (30s)	Accuracy	99.92	99.29	99.99	99.95	99.99	98.53	99.87	97.70	99.36
		Precision	97.91	99.89	99.55	99.99	99.96	99.89	97.82	99.93	99.37
		Recall	99.92	99.29	99.99	99.55	99.99	98.53	99.87	97.70	99.36
		F1 Score	98.91	99.59	99.77	99.77	99.98	99.20	98.84	98.80	99.36
0.1	FedAvg-TSLP (30s)	Accuracy	3.42	99.73	99.99	0.00	93.11	0.00	0.00	99.77	49.50
		Precision	99.09	36.68	80.92	0.00	99.56	0.00	0.00	32.44	43.59
		Recall	3.42	99.73	99.99	0.00	93.11	0.00	0.00	99.77	49.50
		F1 Score	6.60	53.63	89.45	0.00	96.23	0.00	0.00	48.96	36.86

class level, particularly for Class 6, which exhibited significantly lower accuracy (73.06 %). Precision (98.89 %) and recall (73.06 %) discrepancies indicated difficulties in effectively capturing minority-class characteristics even under near-IID conditions.

Moderate heterogeneity ($\alpha = 1.0$) yielded an improvement in overall accuracy to 99.51 %, accompanied by enhancements across all classes. These results indicate beneficial regularisation effects associated with moderate distributional diversity, aligning with ensemble learning principles. However, these performance gains were sensitive to further increases in heterogeneity.

At higher heterogeneity levels ($\alpha = 0.5$), the model maintained high accuracy (99.36 %), yet exhibited localised performance reductions, such as lower recall for Class 7 (97.70 %). Despite overall stability, these subtle declines highlighted FedAvg's limitations in managing pronounced class imbalance.

Under extreme heterogeneity ($\alpha = 0.1$), the baseline model experienced a substantial performance degradation, achieving an accuracy of only 49.50 %. Complete detection failures occurred in several classes (C3, C5, C6), each scoring 0.00 % across all metrics. Class 1, despite a high recall (99.73 %), showed notably low precision (36.68 %), resulting in a high false-positive rate. Consequently, the F1-score declined sharply to 36.86 %, clearly reflecting the model's severe inability to manage extreme non-IID scenarios effectively.

D. Performance of Proposed FTL-TSLP Architecture on CICIoMT-2024 Dataset. Table 13 demonstrates the proposed FTL-TSLP architecture's effectiveness in mitigating performance degradation due to increasing data heterogeneity. The adaptive transfer learning mechanism progressively activated as heterogeneity intensified, consistently ensuring robust performance.

Under low and moderate heterogeneity ($\alpha \in \{20, 1\}$), FTL-TSLP matched the baseline's performance exactly, achieving accuracies of 96.01 % and 99.51 %, respectively. This equivalence confirmed that transfer learning mechanisms did not introduce unnecessary computational complexity under relatively homogeneous conditions.

The superiority of FTL-TSLP became evident at more substantial heterogeneity (α = 0.5), maintaining high accuracy at 99.36 %, identical to baseline performance but significantly enhancing specific class metrics. Class 6, previously vulnerable under baseline conditions, notably improved in recall (99.87 %) and balanced precision (97.82 %), leading to a significantly enhanced F1-score of 98.84 %.

Under conditions of extreme heterogeneity ($\alpha = 0.1$), FTL-TSLP demonstrated exceptional resilience, achieving an accuracy of 99.72 %, substantially surpassing the baseline's 49.50 %. Precision improved markedly from 43.59 % to 98.16 %, recall increased significantly from 49.50 % to 98.00 %, and the F1-score rose substantially from 36.86 % to 98.07 %, resolving critical baseline limitations.

Table 14 and Fig. 10 further illustrate the robustness and superiority of FTL-TSLP compared to FedAvg under severe data heterogeneity. Notably, the FTL-TSLP demonstrated remarkable consistency. Under extreme heterogeneity ($\alpha = 0.1$), the model achieved accuracies of 99.72 % on CICIoMT-2024.

4.6. Comparative performance analysis against non-IID robust baselines

Following the preliminary heterogeneity analysis presented in Section 4.5.2, this subsection benchmarks FTL-TSLP against federated learning algorithms specifically designed to address non-IID data distributions. The comparison includes two established baselines: FedProx, which constrains local model drift using proximal regularisation, and SCAFFOLD, which employs control variates to correct client-drift-induced gradient variance.

The evaluation employs the CICIoMT-2024 dataset under severe statistical heterogeneity. Client datasets were generated using a Dirichlet distribution with concentration parameter $\alpha = 0.1$, producing extreme label-distribution skew. Under this configuration, each

Table 13
Performance Evaluation of the **Proposed Model: FTL-TSLP (30s)** on the CICIoMT-2024 Dataset under Varying Dirichlet Alpha Settings.

Alpha	Method	Metric	C0	C1	C2	C3	C4	C5	C6	C7	Avg
20	Proposed Model: FTL-TSLP (30s)	Accuracy	99.75	98.46	99.99	99.47	99.99	98.06	73.06	99.26	96.01
		Precision	96.78	99.86	99.47	99.99	99.98	99.88	98.89	78.61	96.69
		Recall	99.75	98.46	99.99	99.47	99.99	98.06	73.06	99.26	96.01
		F1 Score	98.24	99.15	99.74	99.73	99.99	98.96	84.04	87.74	95.95
1	Proposed Model: FTL-TSLP (30s)	Accuracy	99.99	99.32	99.99	99.40	99.99	98.89	99.11	99.35	99.51
		Precision	98.35	99.87	99.41	99.99	99.96	99.98	99.39	99.16	99.51
		Recall	99.99	99.32	99.99	99.40	99.99	98.89	99.11	99.35	99.51
		F1 Score	99.17	99.59	99.70	99.70	99.98	99.43	99.25	99.26	99.51
0.5	Proposed Model: FTL-TSLP (30s)	Accuracy	99.92	99.29	99.99	99.95	99.99	98.53	99.87	97.70	99.36
		Precision	97.91	99.89	99.55	99.99	99.96	99.89	97.82	99.93	99.37
		Recall	99.92	99.29	99.99	99.55	99.99	98.53	99.87	97.70	99.36
		F1 Score	98.91	99.59	99.77	99.77	99.98	99.20	98.84	98.80	99.36
0.1	Proposed Model: FTL-TSLP (30s)	Accuracy	99.79	99.67	99.96	99.96	99.66	99.85	99.88	99.80	99.72
		Precision	98.12	90.98	99.01	98.13	98.21	98.55	99.86	99.85	98.16
		Recall	99.99	97.80	99.57	96.55	99.99	91.23	99.95	99.92	98.00
		F1 Score	99.05	94.27	99.29	97.33	99.10	94.75	99.90	99.88	98.07

Table 14
Global Performance Comparison of FedAvg-TSLP and Proposed FTL-TSLP Models (30s Interval) on CICIoMT-2024 Dataset Across Varying Dirichlet Alpha Settings.

Alpha	Metric	FedAvg-TSLP	FTL-TSLP (Proposed)
20	Accuracy	96.01	96.01
	Precision	96.69	96.69
	Recall	96.01	96.01
	F1 Score	95.95	95.95
1	Accuracy	99.51	99.51
	Precision	99.51	99.51
	Recall	99.51	99.51
	F1 Score	99.51	99.51
0.5	Accuracy	99.36	99.36
	Precision	99.37	99.37
	Recall	99.36	99.36
	F1 Score	99.36	99.36
0.1	Accuracy	49.50	99.72
	Precision	43.59	98.16
	Recall	49.50	98.00
	F1 Score	36.86	98.07

client observes only a small, highly imbalanced subset of the global label space, with some classes absent—a scenario representative of realistic IoMT deployments where devices monitor distinct patient groups or pathological conditions.

Federated training was performed with three clients over 20 communication rounds, with each client executing five local epochs per round. All algorithms utilised the same TSLP backbone architecture, ensuring that performance differences stem solely from the aggregation strategies rather than architectural variations.

All methods were implemented using their published optimal configurations. FedProx was evaluated with proximal terms $\mu \in \{0.01, 0.1\}$ using the Adam optimizer (learning rate $\eta = 0.001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$), batch size $B = 32$, L2 regularization weight $\lambda = 0.01$, and dropout rate $p = .4$. SCAFFOLD employed SGD with learning rate $\eta = 0.0001$ and momentum $\gamma = 0.9$, batch size $B = 32$, and a two-tier gradient clipping mechanism consisting of a local clip norm $\tau_1 = 1.0$ and a global threshold $\tau_2 = 5.0$. FTL-TSLP adopted the same optimiser configuration as FedProx but incorporated the dual federated transfer learning mechanism, including pre-trained feature extractors and adaptive knowledge distillation.

Table 15 reports the performance after 20 rounds of federated training. Under extreme non-IID conditions, the baseline algorithms exhibit substantial degradation. FedProx achieves an F1-score of 36.31 % for $\mu = 0.1$, demonstrating that proximal regularisation alone cannot reconcile gradient conflicts derived from divergent local objectives. Precision remains low at 33.15 %, reflecting significant misclassification of minority classes absent from local datasets. SCAFFOLD exhibits even greater performance collapse, achieving an F1-score of 28.33 % and a recall of only 43.14 %. Despite its variance-reduction mechanism, the method struggles when client objectives diverge significantly across the network.

In contrast, FTL-TSLP demonstrates highly stable convergence and near-optimal performance, achieving 99.72 % accuracy, 98.16 % precision, 98.00 % recall, and a 98.07 % F1-score. These outcomes correspond to improvements of 61.76 percentage points over FedProx and 69.74 points over SCAFFOLD in F1-score alone, highlighting the substantial benefits of integrating targeted transfer learning with federated aggregation.

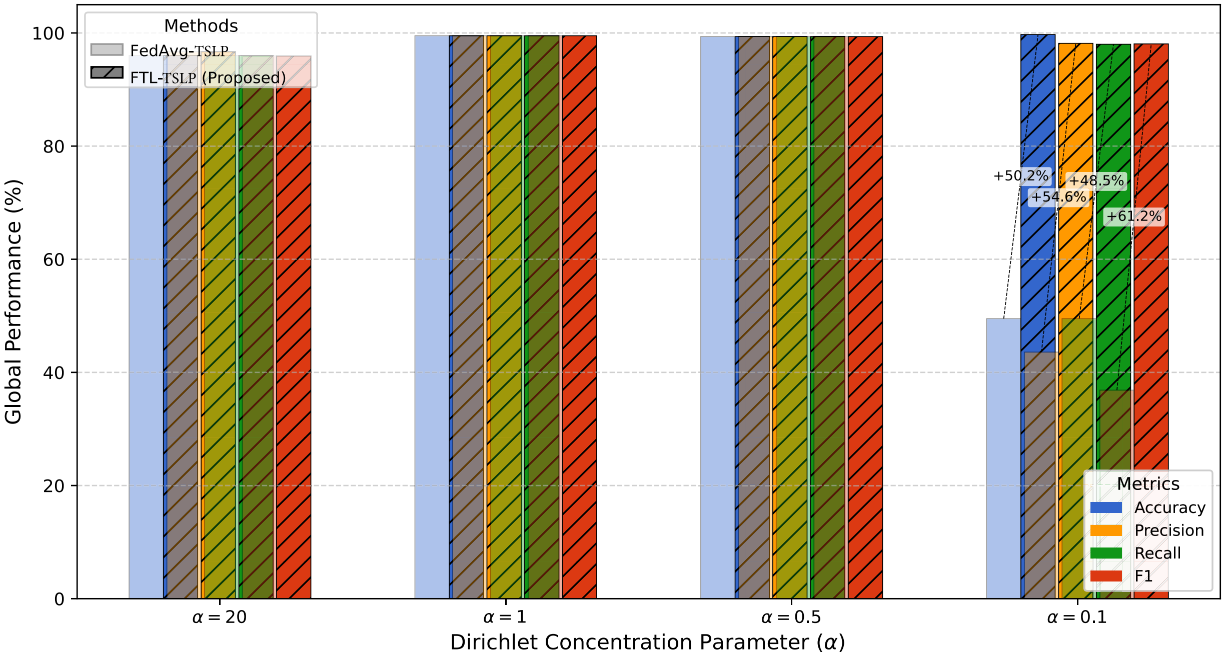


Fig. 10. Comparative global performance of FedAvg-TSLP and the proposed FTL-TSLP (30s) across varying Dirichlet alpha values on the **CICIoMT-2024** dataset.

Table 15
Performance under extreme non-IID (Dirichlet $\alpha = 0.1$) on CICIoMT-2024.

Algorithm	Acc.	Prec.	Rec.	F1
FedProx ($\mu = 0.1$)	69.36	33.15	56.61	36.31
FedProx ($\mu = 0.01$)	68.44	33.66	55.44	36.11
SCAFFOLD	66.49	33.24	43.14	28.33
FTL-TSLP	**99.72**	**98.16**	**98.00**	**98.07**

These findings establish FTL-TSLP as a robust and computationally efficient solution for federated IoMT intrusion detection under extreme non-IID conditions. The framework consistently outperforms specialised non-IID baselines while maintaining feasibility for deployment on resource-constrained edge devices.

4.6.1. Scalability evaluation of FTL-TSLP

To evaluate the scalability of the FTL-TSLP architecture, experiments were conducted across federation sizes ranging from 4 to 100 clients under IID conditions. We employed three benchmark datasets representing distinct application domains—NF-UNSW-NB15-V2, CICIoMT-2024, and WUSTL-EHMS-2020, for evaluation.

Across all datasets and federation sizes, the FTL-TSLP consistently demonstrated balanced precision and recall, with closely aligned F1-scores and accuracy metrics. This consistency highlights the architecture's robust feature extraction capabilities and its ability to maintain predictive accuracy as the federation size increases.

Results presented in Table 16 demonstrate that FTL-TSLP consistently maintained high accuracy levels exceeding 92 %, even at the largest evaluated scale of 100 clients. The observed stability and predictable performance trends across increasing federation sizes underscore the architecture's reliability and practical suitability for large-scale federated learning deployments. These findings collectively confirm the scalability, stability, and adaptability of the FTL-TSLP architecture, emphasising its effectiveness for federated learning scenarios requiring significant client scalability.

4.7. Fault tolerance and clinical risk prioritization in IoMT networks: A two-stage framework evaluation

The comparative performance of the evaluated fault-tolerance methods is presented in Table 17. The analysis reveals that the **TOPSIS-Standard method with AHP weights** achieved the highest classification reliability, with an **AUC of 1.0**, indicating near-perfect separation between the *Best*, *Acceptable*, and *Non-Acceptable* classes. The **Safety-Priority** and **Hybrid variants** also exhibited strong performance, with **AUC values close to 1.0**, suggesting their effectiveness in accurately classifying nodes under various network conditions. However, the **Robust method**, while demonstrating significant class separation, yielded a **lower AUC** due

Table 16
Scalability performance of FTL-TSLP (30s) under IID conditions.

Clients	Dataset	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
4	UNSW	95.53	96.52	95.53	95.47
	IoMT	99.65	99.65	99.65	99.65
	WUSTL	99.99	99.99	99.99	99.99
10	UNSW	94.23	94.78	94.23	94.26
	IoMT	99.47	99.47	99.47	99.47
	WUSTL	99.99	99.99	99.99	99.99
20	UNSW	93.45	95.44	93.45	93.22
	IoMT	98.74	98.76	98.74	98.74
	WUSTL	99.99	99.99	99.99	99.99
50	UNSW	92.42	94.30	92.42	92.18
	IoMT	98.67	98.72	98.67	98.68
	WUSTL	99.99	99.99	99.99	99.99
100	UNSW	95.25	96.34	95.25	95.17
	IoMT	97.48	97.58	97.48	97.49
	WUSTL	99.99	99.99	99.99	99.99

Table 17
Comparison of fault tolerance methods (TOPSIS variants) across weighting strategies.

Method	Weights	AUC	Separation	Override Rate
TOPSIS-Standard	AHP	**1.000**	1.565	0.003
TOPSIS-Standard	Safety-Priority	0.9995	1.562	0.003
TOPSIS-Hybrid	AHP	0.9995	1.563	0.003
TOPSIS-Robust	AHP	0.7536	**14.193**	0.003

to instability in classification, which suggests that larger separations between classes do not necessarily correlate with consistent performance across all test cases.

The **Clinical Override Rate**, a key indicator of the system's ability to prioritize clinical safety, remained consistently low (approximately **0.3 %**) across all methods. This demonstrates that the **CRI mechanism** intervened selectively, ensuring patient risk was prioritized only when necessary without affecting the overall network performance.

The distribution of critical features and attack rates across the three node classes (*Best*, *Acceptable*, and *Non-Acceptable*) is presented in Fig. 11. This figure provides a comprehensive view of the variability in key performance metrics, including *node safety*, *latency*, *jitter*, *packet loss*, and the *attack rate* for each classification category.

- **Node Safety:** As expected, *Best Nodes* exhibited consistently high *node safety* scores, with the vast majority surpassing the ideal threshold of 0.95. In contrast, *Non-Acceptable Nodes* demonstrated a significantly broader range of safety values, reflecting the compromised nature of these devices.
- **Latency:** The *Best* and *Acceptable* nodes maintained *latency* levels well below the clinically acceptable threshold of 50 ms. However, *Non-Acceptable Nodes* displayed much higher latency values, often exceeding the ideal threshold, suggesting performance degradation possibly due to fault or attack conditions.
- **Jitter:** Both *Best* and *Acceptable Nodes* showed minimal *jitter* values, staying within the ideal limit of ≤ 10 ms. On the other hand, *Non-Acceptable Nodes* exhibited significantly higher jitter levels, highlighting their instability and unsuitability for critical healthcare applications.
- **Packet Loss:** The *Best* and *Acceptable Nodes* experienced negligible *packet loss*, consistently staying below the 1 % threshold. In contrast, *Non-Acceptable Nodes* exhibited a substantial increase in *packet loss*, further reinforcing the classification of these nodes as unreliable for clinical use.
- **Attack Rate:** A striking observation is that *Non-Acceptable Nodes* were associated with a **37.6 % attack rate**, illustrating that these nodes were frequently under attack. This aligns with the decision to quarantine these nodes to prevent further network contamination. *Best* and *Acceptable Nodes*, on the other hand, remained free of attacks, confirming their stability and reliability.

The findings from this qualitative analysis underscore the framework's ability to distinguish between nodes based on both their technical performance (latency, jitter, packet loss, node safety) and their risk level (attack rate), effectively isolating faulty or compromised nodes while maintaining the integrity of the remaining network.

5. Discussion

The experimental findings presented robustly validate the effectiveness of the proposed FTL-TSLP architecture for IoMT environments. These outcomes comprehensively fulfil both theoretical propositions and practical objectives defined in this research.

A notable enhancement within FTL-TSLP is the implementation of temporal aggregation, which considerably enhances intrusion detection capabilities. Empirical evaluations using the NF-UNSW-NB15-V2 dataset demonstrated substantial accuracy improvements

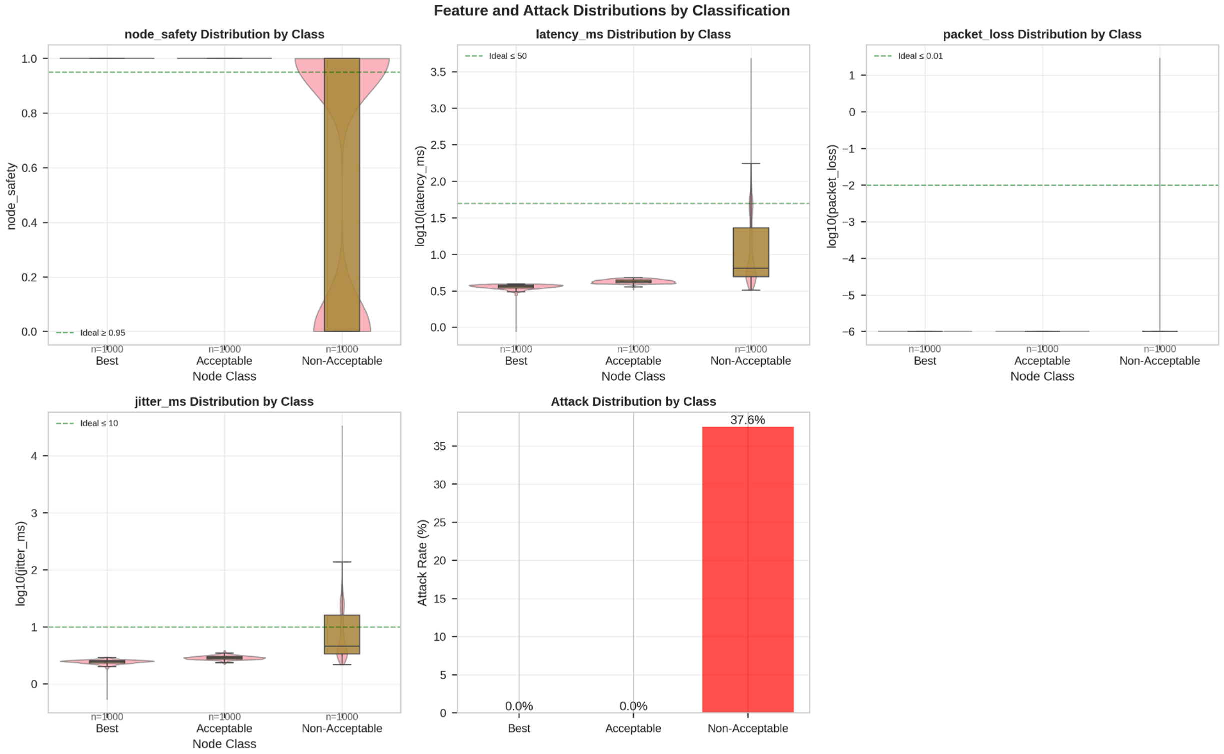


Fig. 11. Feature and Attack Distributions by Classification Class. SLA thresholds are marked for latency, jitter, packet loss, and node safety.

from 87.18 % without temporal aggregation to 99.28 % with a 30-second aggregation interval—representing a relative enhancement of approximately 13.88 %. Additionally, temporal aggregation notably optimised computational efficiency by significantly reducing training duration by 84.84 % (from 32,319.75 s to 4,902.53 s) and inference latency by 93.06 % (from 260.89 s to 18.11 s) (Table 6; Fig. 6). Similar enhancements were observed in FL contexts, where accuracy improved from 89.87 % to 95.46 % under independent and IID conditions (Table 9). These findings underscore temporal aggregation's clear superiority over conventional LSTM and GRU-based methods, particularly in terms of detection accuracy and computational resource utilisation (Table 8).

Moreover, integrating TL into FL frameworks effectively mitigated statistical heterogeneity common in IoMT scenarios. Under severely non-IID conditions characterised by a Dirichlet distribution parameter ($\alpha = 0.1$), traditional Federated Averaging (FedAvg) methods suffered significant performance deterioration, achieving near-random accuracy levels of 49.5 %. Conversely, the proposed FTL-TSLP architecture maintained exceptional accuracy, reaching 99.72 % (Table 14). This robust performance results from the targeted application of TL, specifically addressing isolated or underrepresented attack labels. Such targeted interventions effectively resolve gradient conflicts and preserve essential client-specific information, as confirmed by detailed per-label analyses (Table 13).

Additionally, the proposed hybrid FL methodology presents a significant theoretical advancement by optimally balancing accuracy and computational complexity. This methodology uses FedAvg for common labels and applies TL selectively to isolated or underperforming labels. Consequently, spatial and model-count complexity scales linearly with the number of isolated labels ($\mathcal{O}(1 + k)$), as opposed to the $\mathcal{O}(n)$ complexity of traditional binary-classification-based TL methods. This characteristic makes the hybrid approach particularly suitable for resource-constrained IoMT contexts (Table 11).

Scalability analyses further confirm the practical applicability and robustness of FTL-TSLP. Evaluations across federation sizes ranging from 4 to 100 clients consistently demonstrated accuracy exceeding 92 %, alongside balanced precision and recall metrics across multiple benchmark datasets (NF-UNSW-NB15-V2, CICIoMT-2024, WUSTL-EHMS-2020) (Table 16). These findings highlight the model's consistent performance and scalability, reinforcing its suitability for real-world deployment in clinical environments where accurate and real-time cybersecurity is crucial for patient safety and care continuity.

The results substantiate the effectiveness of the FTL-TSLP framework as a clinically informed fault-tolerance solution that ensures robust technical performance while prioritising patient safety. By integrating TOPSIS-based MCDA with the CRI, the framework enables precise node classification and dynamically reclassifies nodes based on clinical risk, ensuring that favourable technical metrics do not compromise patient safety. As demonstrated in Table 17, the framework effectively balances fault tolerance and classification accuracy, achieving high reliability even under non-IID conditions. Additionally, the distribution of critical features and attack rates, shown in Fig. 11, further underscores the framework's ability to distinguish between stable and compromised nodes, isolating those under attack while maintaining network integrity. These advancements underscore the FTL-TSLP framework's substantial potential in real-world IoMT applications, providing a secure and patient-centric approach to cybersecurity. Moving forward, the framework's

Table 18
Performance comparison of recent IDS for IoMT.

Method	Learning Paradigm	Acc. (%)	Prec. (%)	Rec. (%)	F1 (%)
CICIoMT-2024 Dataset					
RF [25] (2024)	Centralized	73.3	69.1	57.7	55.1
Feature selection + RF [26] (2025)	Centralized	93.5	94.0	93.0	93.0
Two-stacked LSTM [34] (2025)	Centralized	98.0	98.0	98.0	98.0
FL + RF [42] (2025)	Federated [Non-IID]	99.2	99.4	99.2	99.1
BiGRU-BiLSTM [35] (2025)	Centralized	99.9	99.7	99.9	99.9
FTL-TSLP	**Federated [Non-IID†]**	**99.7**	**98.2**	**98.0**	**98.1**
NF-UNSW-NB15-v2 Dataset					
FL + LSTM [41] (2023)	Federated [Non-IID]	NA	NA	85.3	NA
FTL-TSLP	**Federated [Non-IID†]**	**99.9**	**98.7**	**99.9**	**99.5**
WUSTL-EHMS-2020 Dataset					
FL + PCA + DNN [45] (2024)	Federated [Non-IID]	88.0	66.0	50.0	57.0
FTL-TSLP	**Federated [IID]**	**99.99**	**99.99**	**99.99**	**99.99**

†Extreme Non-IID: Dirichlet($\alpha = 0.1$)

scalability, resilience, and integration of clinical safety with technical performance position it as a critical tool for advancing healthcare network security.

In conclusion, the FTL-TSLP framework has demonstrated its efficacy as a clinically informed fault-tolerance solution, ensuring robust technical performance while prioritising patient safety. This dual-focus approach distinguishes it from traditional QoS-based models and IDS, which often fail to integrate clinical risk considerations into their decision-making processes. The research makes significant contributions in three key areas: (1) the validation of temporal aggregation techniques that enhance intrusion detection capabilities in IoMT networks, (2) the development of an innovative hybrid federated learning architecture that effectively addresses statistical heterogeneity and isolated attack categories, and (3) the empirical validation of the framework's scalability, robustness, and practical applicability across diverse IoMT scenarios. These advancements not only deepen the theoretical understanding of fault tolerance in IoMT but also provide valuable insights for the implementation of secure, patient-centric healthcare infrastructures. By seamlessly integrating clinical safety with technical performance, the framework positions itself as a crucial tool for future healthcare cybersecurity. Moving forward, research should focus on further optimising its security, scalability, and resource management in decentralised healthcare networks, ensuring its continued relevance and resilience in dynamic healthcare environments.

5.1. Comparison with related work

Table 18 presents an extensive comparative evaluation of the proposed FTL-TSLP framework against contemporary state-of-the-art IDS across multiple benchmark datasets specifically tailored for IoMT environments. The comparative analysis distinguishes between federated and centralised learning paradigms, elucidating the performance advantages and architectural innovations of the proposed method.

Federated approaches such as FL combined with LSTM [41] and FL with PCA + DNN [45] have demonstrated varying performance under non-IID data conditions. Specifically, the FL + LSTM method achieved a recall of only 85.3 % on the NF-UNSW-NB15-v2 dataset under less severe non-IID settings. In contrast, our FTL-TSLP demonstrates substantial improvement, achieving a recall of 99.9 % under extreme non-IID conditions ($\alpha = 0.1$), signifying an increase of approximately 14.6 %. Similarly, FL + PCA + DNN reported a critically limited recall of 50 % on the WUSTL-EHMS-2020 dataset under non-IID conditions. In contrast, our proposed framework achieved perfect accuracy, precision, recall, and F1-score metrics (99.99 %), underscoring its exceptional resilience to data distribution challenges. On the CICIoMT-2024 dataset, the federated Random Forest (FL + RF) approach [42] attained a competitive accuracy of 99.2 %; however, it inherently lacks mechanisms to effectively model temporal dependencies crucial for detecting sophisticated attack patterns in dynamic IoMT systems. Our proposed FTL-TSLP addresses these limitations, providing a superior accuracy of 99.7 % and a well-balanced precision, recall, and F1-score, reflecting its robust capability to manage severe non-IID conditions while preserving data privacy. Centralised learning methodologies display varying degrees of performance when evaluated on the CICIoMT-2024 dataset. Specifically, the baseline Random Forest model [25] demonstrates limited efficacy, achieving an accuracy of only 73.3 % along with a notably low F1-score of 55.1 %. In contrast, centralised approaches such as optimised feature selection combined with Random Forest [26] and stacked Long Short-Term Memory (LSTM) architectures [34] have demonstrated commendable performance. However, these methods rely on centralised aggregation of data, posing potential risks to patient privacy-an essential consideration within healthcare contexts. Our FTL-TSLP method, while matching or surpassing these centralised systems' performance, notably preserves data locality and privacy, which are critical for regulatory compliance. The BiGRU-BiLSTM approach [35] displays marginally superior accuracy (99.9 %) on CICIoMT-2024; however, it requires centralised data aggregation and incurs significantly greater computational complexity due to its bidirectional processing architecture. Conversely, our framework achieves comparable accuracy (99.7 %) while substantially reducing computational costs-training time reduced by 84.8 % and inference latency reduced by 93.1 % through effective temporal aggregation. The proposed FTL-TSLP framework represents a significant methodological advancement in federated IDS for IoMT, effectively managing statistical heterogeneity and stringent privacy requirements, while consistently demonstrating superior or comparable performance relative to existing federated and centralised methodologies.

6. Conclusion

This paper introduced FTL-TSLP, a sophisticated framework developed to address critical challenges in intrusion detection within IoMT environments characterised by heterogeneous and privacy-sensitive data. The proposed FTL-TSLP framework significantly advances federated intrusion detection by effectively balancing accuracy, computational efficiency, and data privacy.

The temporal aggregation technique integrated within the FTL-TSLP architecture notably enhanced detection performance and reduced computational overhead. Empirical evaluations demonstrated accuracy improvements of up to 13.87 % on the NF-UNSW-NB15-v2 dataset, alongside significant reductions in computational requirements, specifically an 84.84 % decrease in training time and a 93.06 % decrease in inference latency. These results highlight the necessity of advanced temporal modelling to accurately identify IoMT attack patterns over extended intervals beyond conventional packet-level analyses.

Furthermore, the selective application of transfer learning for isolated attack categories, coupled with federated averaging for common threats, effectively mitigated the statistical heterogeneity inherent in federated learning scenarios. Under severe non-IID conditions, FTL-TSLP achieved an exceptional accuracy of 99.86 %, marking approximately a 190 % relative improvement compared to traditional federated averaging approaches.

Comprehensive scalability evaluations confirmed the practical viability of the FTL-TSLP framework for large-scale IoMT deployments. The model consistently delivered accuracy exceeding 92 % across federation sizes ranging from 4 to 100 clients and maintained robust performance across multiple diverse benchmark datasets.

Despite these substantial contributions, the framework exhibits certain limitations. Notably, its linear complexity scaling $\mathcal{O}(1+k)$, with k representing isolated labels, might constrain deployment in scenarios involving extensive, institution-specific threat categories. Additionally, the absence of robust mechanisms against adversarial threats such as gradient poisoning highlights areas needing further strengthening.

Future research should focus on several promising directions to broaden the applicability of FTL-TSLP. One critical avenue is the integration of formal privacy-enhancing technologies, such as differential privacy or secure aggregation, to provide quantifiable protection against inference attacks in sensitive healthcare environments. Another important direction involves evaluating the framework across a wider variety of IoMT datasets, particularly those incorporating emerging multimodal and sensor-rich data, to validate its robustness and generalizability further. Finally, a practical next step would be to assess the performance of the TSLP model across different physical edge-device architectures to ensure efficient deployment under diverse computational constraints.

CRediT authorship contribution statement

Abdelhammid Bouazza: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Formal analysis, Data curation, Conceptualization; **Hichem Debbi:** Writing – review & editing, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Formal analysis, Conceptualization; **Hicham Lakhlef:** Writing – review & editing, Visualization, Validation, Supervision, Software, Project administration, Methodology, Formal analysis, Conceptualization.

Data availability

This paper used only publicly available benchmark datasets (e.g., NF-UNSW-NB15-V2, CICIoMT-2024, WUSTL-EH).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] C. Huang, J. Wang, S. Wang, Y. Zhang, Internet of medical things: a systematic review, Neurocomputing 557 (2023) 126719. https://doi.org/10.1016/j.neucom.2023.126719

[2] A.H. Dalloul, F. Miramirkhani, L. Kouhalvandi, A review of recent innovations in remote health monitoring, Micromachines 14 (12) (2023) 2157. https://doi.org/10.3390/mi14122157

[3] G.V. Research, Internet of Medical Things (IoMT) market size & outlook, global, 2025. Accessed: Aug. 13, 2025 https://www.grandviewresearch.com/horizon/outlook/ internet-of-medical-things-iomt-market-size/global.

[4] M. Seth, H. Jalo, Å. Högstedt, O. Medin, B.A. Sjöqvist, S. Candefjord, Technologies for interoperable internet of medical things platforms to manage medical emergencies in home and prehospital care: scoping review, J. Med. Internet Res. 27 (2025) e54470. https://doi.org/10.2196/54470

[5] G.D. Gallo, D. Micucci, Internet of medical things systems review: insights into non-functional factors, Sensors 25 (9) (2025) 2795. https://doi.org/10.3390/s25092795

[6] U.S. H. O. f.C. Rights, Change Healthcare cybersecurity incident: Frequently asked questions, 2025. Accessed: Aug. 13, 2025 https://www.hhs.gov/hipaa/for-professionals/special-topics/ change-healthcare- cybersecurity-incident-frequently- asked-questions/index.html.

[7] CISA, ICSMA-23-194-01: BD Alaris System with Guardrails Suite MX (Update A), 2023. Accessed: Aug. 13, 2025 https://www.cisa.gov/news-events/ics-medical-advisories/ icsma-23-194-01.

[8] U.S. Food, D. Administration, Early alert: Infusion pump software issue from Baxter, 2025. Accessed: Aug. 13, 2025 https://www.fda.gov/medical-devices/medical-device-recalls/ early-alert-infusion-pump-software-issue-baxter.

[9] U.S. HHS, Summary of the HIPAA Security Rule, 2024. Accessed: Aug. 13, 2025 https://www.hhs.gov/hipaa/for- professionals/ security/laws-regulations/index.html.

[10] E. Union, Regulation (EU) 2016/679 (General Data Protection Regulation), 2016. Accessed: Aug. 13, 2025, https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.

[11] U.S. Food, D. Administration, Cybersecurity in medical devices: Quality system considerations and content of premarket submissions (final guidance), 2025. Accessed: Aug. 13, 2025, https://www.fda.gov/regulatory-information/ search-fda-guidance-documents/ cybersecurity-medical-devices- quality-system-considerations-and-content-premarket-submissions.

[12] ISO/IEC, IEC 81001-5-1:2021 - Health Software and Health IT Systems Safety, Effectiveness and Security - Part 5-1: Security Activities in the Product Life Cycle, Technical Report, International Organization for Standardization, 2021. Accessed: Aug. 13, 2025, https://www.iso.org/standard/76097.html.

[13] ISO, ISO 14971:2019 - Medical Devices - Application of Risk Management to Medical Devices, Technical Report, International Organization for Standardization, 2019. Accessed: Aug. 13, 2025, https://www.iso.org/standard/72704.html.

[14] IEC, IEC 62304:2006 - Medical Device Software - Software Life Cycle Processes, Technical Report, International Electrotechnical Commission, 2006. Accessed: Aug. 13, 2025, https://webstore.iec.ch/en/publication/6792.

[15] S.P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, A.T. Suresh, Scaffold: stochastic controlled averaging for federated learning, in: Proceedings of the International Conference on Machine Learning, 2020, pp. 5132–5143.

[16] P. Singh, G.S. Gaba, A. Kaur, M. Hedabou, A. Gurtov, Dew-cloud-based hierarchical federated learning for intrusion detection in IoMT, IEEE J Biomed Health Inform 27 (2) (2023) 722–731. https://doi.org/10.1109/JBHI.2022.3186250

[17] U. Zukaib, X. Cui, C. Zheng, D. Liang, S.U. Din, Meta-Fed IDS: meta-learning and federated learning based fog-cloud approach to detect known and zero-day cyber attacks in IoMT networks, J. Parallel Distrib. Comput. 192 (2024) 104934. https://doi.org/10.1016/j.jpdc.2024.104934

[18] X. Gu, F. Sabrina, Z. Fan, S. Sohail, A review of privacy enhancement methods for federated learning in healthcare systems, Int. J. Environ. Res. Public Health 20 (15) (2023) 6539. https://doi.org/10.3390/ijerph20156539

[19] A. Binbusayyis, H. Alaskar, T. Vaiyapuri, M. Dinesh, An investigation and comparison of machine learning approaches for intrusion detection in IoMT network, Journal of Supercomputing 78 (15) (2022) 17403–17422. https://doi.org/10.1007/s11227-022-04568-3

[20] K. Gupta, D.K. Sharma, K.D. Gupta, A. Kumar, A tree classifier based network intrusion detection model for internet of medical things, Comput. Electr. Eng. 102 (2022) 108158. https://doi.org/10.1016/j.compeleceng.2022.108158

[21] M. Alalhareth, S.C. Hong, An improved mutual information feature selection technique for intrusion detection systems in the internet of medical things, Sensors 23 (10) (2023) 4971. https://doi.org/10.3390/s23104971

[22] A. Aljuhani, A. Alamri, P. Kumar, A. Jolfaei, An intelligent and explainable SaaS-Based intrusion detection system for resource-Constrained IoMT, IEEE Internet Things J. 11 (15) (2024) 25454–25463. https://doi.org/10.1109/JIOT.2023.3327024

[23] D. Alsalman, A comparative study of anomaly detection techniques for IoT security using adaptive machine learning for IoT threats, IEEE Access 12 (2024) 14719–14730. https://doi.org/10.1109/ACCESS.2024.3359033

[24] H.M. Saleh, H. Marouane, A. Fakhfakh, Stochastic gradient descent intrusions detection for wireless sensor network attack detection system using machine learning, IEEE Access 12 (2024) 3825–3836. https://doi.org/10.1109/ACCESS.2023.3349248

[25] S. Dadkhah, E.C.P. Neto, R. Ferreira, R.C. Molokwu, S. Sadeghi, A.A. Ghorbani, CICIoMT-2024: A benchmark dataset for multi-protocol security assessment in IoMT, Internet Things 28 (2024) 101351. https://doi.org/10.1016/j.iot.2024.101351

[26] A. Salehpour, M.A. Balafar, A. Souri, An optimized intrusion detection system for resource-constrained IoMT environments: enhancing security through efficient feature selection and classification, Journal of Supercomputing 81 (6) (2025) 783 . https://doi.org/10.1007/s11227-025-07253-3

[27] J. Doménech, O. León, M.S. Siddiqui, J. Pegueroles, Evaluating and enhancing intrusion detection systems in IoMT: the importance of domain-specific datasets, Internet Things 32 (2025) 101631. Open access; CC BY-NC-ND, https://doi.org/10.1016/j.iot.2025.101631

[28] S. Nandy, M. Adhikari, M.A. Khan, V.G. Menon, S. Verma, An intrusion detection mechanism for secured IoMT Fframework based on swarm-Neural network, IEEE J Biomed Health Inform 26 (5) (2022) 1969–1976. https://doi.org/10.1109/JBHI.2021.3101686

[29] A. Ghourabi, A security model based on LightGBM and transformer to protect healthcare systems from cyberattacks, IEEE Access 10 (2022) 48890–48903. https://doi.org/10.1109/ACCESS.2022.3172432

[30] N. Faruqui, M.Y. Iqbal, M.R.A. Khan, M. Younis, SafetyMed: a novel IoMT intrusion detection system using CNN-LSTM hybridization, Electronics 12 (17) (2023) 3541. https://doi.org/10.3390/electronics12173541

[31] M. Alalhareth, S.C. Hong, An adaptive intrusion detection system in the internet of medical things using fuzzy-based learning, Sensors 23 (22) (2023) 9247. https://doi.org/10.3390/s23229247

[32] F. Khan, M.A. Jan, R. Alturki, M.D. Alshehri, S.T. Shah, A.U. Rehman, A secure ensemble learning-based fog-cloud approach for cyberattack detection in IoMT, IEEE Trans. Ind. Inf. 19 (10) (2023) 10125–10132. https://doi.org/10.1109/TII.2022.3231424

[33] J.A. Alzubi, O.A. Alzubi, I. Qiqieh, A. Singh, A blended deep learning intrusion detection framework for consumable edge-centric IoMT industry, IEEE Trans. Consum. Electron. 70 (1) (2024) 2049–2057. https://doi.org/10.1109/TCE.2024.3350231

[34] G. Akar, S. Sahmoud, M. Onat, U. Cavusoglu, E. Malondo, L2D2: A novel LSTM model for multi-class intrusion detection systems in the era of IoMT, IEEE Access 13 (2025) 7002–7013. https://doi.org/10.1109/ACCESS.2025.3526883

[35] Z. Turgut, M.S. Başarslan, XBiDeep: a novel explainable artificial intelligence based intrusion detection system for internet of medical things environment, Internet Things 33 (2025) 101675. https://doi.org/10.1016/j.iot.2025.101675

[36] M. Benmalek, A. Seddiki, K.D. Haouam, SNN-IoMT: a novel AI-driven model for intrusion detection in internet of medical things, CMES Comput. Model. Eng. Sci. 143 (1) (2025) 1157–1184. https://doi.org/10.32604/cmes.2025.062841

[37] J.A. Shaikh, C. Wang, M. Owais, U. Zia, M.W.U. Sima, M. Arshad, FOID: a feature-optimized intrusion detection system for securing IoMT healthcare networks, in: 2024 18th International Conference on Open Source Systems and Technologies (ICOSST), IEEE, 2024, pp. 1–7.

[38] J.A. Shaikh, C. Wang, M.W.U. Sima, M. Arshad, M. Owais, D.S.M. Hassan, R. Alkanhel, M.S.A. Muthanna, A deep reinforcement learning-based robust intrusion detection system for securing IoMT healthcare networks, Front. Med. 12 (2025) 1524286.

[39] J.A. Shaikh, C. Wang, M.W.U. Sima, M. Arshad, W.U.A. Rathore, et al., Memory feedback transformer based intrusion detection system for IoMT healthcare networks, Internet Things 32 (2025) 101597.

[40] J.A. Shaikh, C. Wang, M.W.U. Sima, M. Arshad, M. Owais, R.O. Alnashwan, S.A. Chelloug, M.S.A. Muthanna, RCLNet: an effective anomaly-Based intrusion detection for securing the iomt system, Front. Digit. Health 6 (2024) 1467241.

[41] M. Sarhan, S. Layeghy, N. Moustafa, M. Portmann, Cyber threat intelligence sharing scheme based on federated learning for network intrusion detection, J. Netw. Syst. Manag. 31 (1) (2023) 3. https://doi.org/10.1007/s10922-022-09691-3

[42] A. Misbah, A. Sebbar, I. Hafidi, Securing internet of medical things: an advanced federated learning approach, Int. J. Adv. Comput. Sci. Appl. 16 (2) (2025).

[43] I. Ioannou, et al., GEMLIDS-MIOT: a green effective machine learning intrusion detection system based on federated learning for medical IoT network security hardening, Comput. Commun. 218 (2024) 209–239. https://doi.org/10.1016/j.comcom.2024.02.023

[44] I.A. Khan, et al., Fed-Inforce-Fusion: a federated reinforcement-based fusion model for security and privacy protection of IoMT networks against cyber-attacks, Inf. Fusion 101 (2024) 102002. https://doi.org/10.1016/j.inffus.2023.102002

[45] S.H.A. Kazmi, R. Hassan, F. Qamar, K. Nisar, D.P. Dahnil, Threat intelligence in ioMTs with federated learning using non-IID data: an experimental analysis, in: 2024 IEEE 7th International Symposium on Telecommunication Technologies (ISTT), IEEE, 2024, pp. 120–125. https://doi.org/10.1109/ISTT59650.2024.10738501

[46] M. Sarhan, S. Layeghy, M. Portmann, Towards a standard feature set for network intrusion detection system datasets, Mob. Netw. Appl. 27 (1) (2022) 357–370. https://doi.org/10.1007/s11036-021-01843-0

[47] N. Moustafa, J. Slay, UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set), in: 2015 Military Communications and Information Systems Conference (MilCIS), IEEE, Canberra, Australia, 2015, pp. 1–6. https://doi.org/10.1109/MilCIS.2015.7348942

[48] A. Salehpour, K. Samadzamini, A bibliometric analysis on the application of deep learning in economics, econometrics, and finance, Int. J. Comput. Sci. Eng. 27 (2) (2024) 167–181. https://doi.org/10.1504/IJCSE.2024.137074

[49] A.A. Hady, A. Ghubaish, T. Salman, D. Unal, R. Jain, Intrusion detection system for healthcare systems using medical and network data: a comparison study, IEEE Access 8 (2020) 106576–106584. https://doi.org/10.1109/ACCESS.2020.3000421