IoMT Security Enhancement through Federated Learning and Advanced Models

Mohammad Zahid[1] and Taran Singh Bharati[2]

[1] *Department of Computer Science, Jamia Millia Islamia, New Delhi, India*
[2] *Department of Computer Science, Jamia Millia Islamia, New Delhi, India*
mohammad2206322@st.jmi.ac.in,tbharti@jmi.ac.in

KEYWORDS

IoMT, IoT Security, Federated Learning, IoT attacks, CICIoMT2024 dataset, Machine Learning.

ABSTRACT

The Internet of Medical Things (IoMT) has revolutionized healthcare with real-time monitoring, remote diagnosis, and customized care. Yet, its heterogeneous and decentralized architecture combined with the confidentiality of medical information poses paramount security threats like malicious traffic identification, data leaks, and device susceptibility. This paper suggests a federated learning model to secure IoMT networks without compromising data privacy. Based on the CICIoMT2024 dataset, five decentralized clients' training is emulated on stratified data subsets. Sophisticated methods involving feature standardization, Mutual Information-based feature engineering, and class balancing using SMOTETomek handle variability and imbalance of data. Random Forest, XGBoost, CatBoost, LightGBM, and Neural Network are local models trained individually on client-related data to embrace heterogeneous IoMT traffic patterns. A weighted aggregation approach combines client models into a global model, placing a heavier weight on contributions from top-performing clients. The global model has 98.78% high accuracy, with robust malicious traffic detection rates (95.4% True Positive Rate) and few false alarms (2.6% False Positive Rate). Precision-recall evaluation verifies the reliability of the framework, yielding 97.2% precision and 95.4% recall for malicious traffic. These outcomes prove the proposed framework's robustness and scalability, ensuring it can be effectively implemented in real-world healthcare scenarios for securing IoMT networks.

Introduction

The Internet of Things (IoT) has become one of the foundational layers of modern digital infrastructure as millions of connected devices join an intelligent and interactive ecosystem [1]. The IoT improves the scope of healthcare, transportation, and smart cities in terms of increased efficiency in operations, and informed decision-making through data, and real-time services [2]. On the other hand, the rapid proliferation of IoT devices has challenged concerns about the privacy and security of sensitive information. Traditional forms of security are usually slow to keep pace with the emergence of threats such as DoS, spoofing, data manipulation, jamming, and eavesdropping attacks [3]. Therefore, it becomes imperative to find innovative solutions that address these vulnerabilities well due to the heterogeneity and decentralization of IoT networks [4]. With the expansion of the IoT systems in critical sectors such as health care, the Internet of Medical Things (IoMT) has been introduced, which allows for real-time health monitoring and diagnostic capabilities [10]. This large group of medical devices has enormous potential for improving patient care but also presents a higher security risk because of the sensitive nature of health-related information [5]. The outcome of cyberattacks on IoMT

systems could be severe, including patient privacy violations, misdiagnosis, and system outages [6].

In recent times, federated learning (FL) has emerged as a promising solution to address these challenges [1, 16]. It dispenses with the traditional centralized machine learning model that asks for the aggregation of raw data under one central server by allowing the model to be trained on models distributed across devices without transferring the data off-site [3]. This decentralized approach is significant in preserving privacy and scalability; hence it can be greatly applied in IoT and IoMT environments [9]. FL can protect sensitive data with powerful insights from ML, promising a solution for secure IoT systems [5]. Still, it does not ensure immunity against security vulnerabilities in FL such as model poisoning and inference attacks which might compromise its overall effectiveness in communication overhead [6, 11, 20].

1.1 IoMT Security Challenges

IoMT system security in the healthcare sector is extremely important because it addresses very sensitive data [3]. Some of the medical devices used in highly regulated environments include pacemakers, insulin pumps, and home monitoring equipment that attach tremendous importance to data privacy and security requirements. Researchers found that cryptography, among other traditional security measures, cannot stop the threats of highly sophisticated attacks on IoMT systems, as it cannot be configured to solve problems like data leakage and model poisoning [5]. Moreover, in IoMT systems, the addition of physical and computation factors increases factors like real-time detection and mitigation of threats [5].

1.2 Federated Learning for IoMT Security

FL emerges as a promising approach to tackle the security and privacy concerns of IoT and IoMT systems [1]. FL prevents the need for movement of data from the device toward other external devices for training the model, thereby reducing the potential breach of data and maintaining user privacy [3]. Several studies were discussed in the framework about exploring the application of FL in IoT settings focusing on how security could be enhanced while maintaining the scalability of the system [2, 6]. One of the most prominent experiments applied FL to design a distributed IDS for IoMT environments [4]. IDS leveraged the collaborative characteristics of FL to detect anomalies and prevent cyberattacks in real time without centralized data collection [12, 17, 20]. It has high detection accuracy with low false-positive rates, which makes this an effective solution for ensuring the safety of IoMT systems against emerging threats [5, 14]. It also alleviated the computationally expensive burden on individual devices, thus resulting in optimum utilization of resources in IoMT environments [9, 10, 12].

1.3 Internet of Medical Things (IoMT) in healthcare

The IoMT is indeed playing a transformative role in healthcare, as represented in **Fig. 1**, through the creation of applications for real-time patient monitoring, remote medicine, chronic disease management, and smart medication delivery, among others [5]. Devices in IoMT, for instance, wearable health monitors, infusion pumps, and smart sensors give continuous physiological data that increases the effectiveness of diagnostics and the exactness of treatments [9]. IoMT is greatly enhancing remote medicine, as patients are now able to receive care through telehealth solutions. In this, devices monitor vital signs and send feedback to the medical personnel [12]. The IoMT devices, such as glucometers and blood pressure monitors, benefit chronic condition management by providing real-time tracking and personalized care plans [20]. Furthermore, IoMT solutions like smart inhalers and sleep monitors support respiratory and sleep health by tracking critical parameters and ensuring timely interventions [14]. These devices help improve patient outcomes by supporting data-driven, automated, and remote care while making healthcare settings more comfortable and efficient to work in.

Fig. 1. IoMT Application in Healthcare.

Fig. 2 depicts an overarching architecture for monitoring health parameters using IoT-enabled medical sensors integrated into a secure network. The "Medical Sensors" section contains SpO2, blood pressure, temperature, and ECG sensors connected to a medical sensor board, which collects data on the patient's health. These sensors send this data to a gateway as represented by a Windows laptop through the USB, Ethernet, and serial port connections. This is then transmitted to the network, which consists of a router and switch integrated with an IDS in case intruders try to probe into the system from potential attack sites. After which, the data are analyzed and visualized as produced in the "Control & Visualization" section. These consist of a server and a controller. This setup will stress that the security and efficiency of data transfer in healthcare applications are ensured by real-time monitoring and analysis of important health parameters.

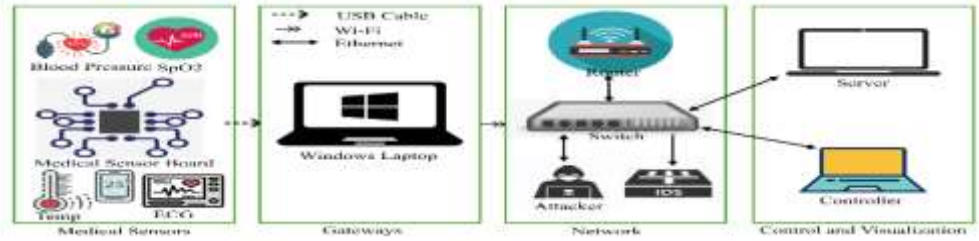

Fig. 2. IoT-Based Architecture for Secure Health Monitoring and Data Transmission.

The key contributions of this paper are outlined as follows:

- Designed a client-centric federated learning framework for IoMT security using the CICIoMT2024 dataset.
- Developed a strong preprocessing pipeline, including mutual information-based feature selection and SMOTETomek for data balancing.
- Integrated an ensemble of machine learning models (Random Forest, XGBoost, CatBoost, LightGBM, and Neural Networks) for enhanced prediction reliability.
- Designed a weighted aggregation strategy based on client-specific accuracies to make equal contributions.

The rest of the paper is organized as follows: Section 2 explains related works, and Section 3 describes the proposed methodology, including the data selection method, simulation of clients, preprocessing of data, feature selections, data balancing, model training, and weighted aggregation. Section 4 presents results and discussion, while Section 5 concludes the work and draws future directions for research.

2 Related Work

The most fascinating aspect of integrating FL into the safety of IoMT is the recent attention gained from the research community. A multitude of research studies point out the benefits of FL toward advancing data privacy while knocking off the challenges created by device security challenges in IoMT.

In recent research, the author [1] proposed a federated learning-based intrusion detection system called FLIDS that is highly customized to Medical Cyber-Physical Systems (MCPS). Through developing the GANs-based model, cyber-attacks are remarkably detected in a clinical environment by collaboratively training various devices' models. Average accuracy was reportedly above 99%, and false positives were nearly zero. Their work established how FL could protect sensitive health data from cyberattacks and simultaneously allow for the real-time identification of threats.

The author [2] attempted to solve key privacy concerns in most classical IoMT frameworks when they proposed a decentralized method that leveraged FL to collaboratively train distributed devices. The author built on designing a privacy-protective mechanism such that devices could learn from shared patterns in shared data without uploading such private information to a centralized server. Their results are an FL framework that would be able to reduce privacy threats and yet would be good at performance compared to a centralized model.

An extensive review of FL applications in IoT networks was conducted by [3]. The author focused on deploying FL for scaling up security improvement in IoT networks with the resolution of scalability-related issues to the challenge of centralized data processing. In addition, some architectures and strategies for the optimal training of the model in IoMT environments are developed. It gave an argument, that while FL promises to have complete privacy preservation, model vulnerabilities and the adversarial impact are to be better.

The author [4] discussed security threats with the implementation of FL in IoT settings and emphasized further robust countermeasures against emerging threats. The authors specifically considered poisoning attacks on FL models and data inference attacks. It proposed a multilayered defense mechanism integrating anomaly detection with secure aggregation methods to enhance the resilience of FL to be applied in IoMT applications.

In the context of adaptive privacy mechanisms, the author [6] outlined the need for dynamic privacy adjustments that can adapt to different threat landscapes and adapt to user privacy requirements. Their study explored the feasibility of differential privacy techniques in conjunction with FL, demonstrating such an approach can strengthen the model's robustness against privacy attacks while maintaining accuracy. The authors conclude that adaptive privacy mechanisms may make important contributions toward the secure deployment of FL in sensitive application domains such as health care.

Accordingly, [8] discussed the federated optimization techniques that can further be enhanced to build a more efficient FL in the IoMT environment. The authors designed a federated optimization algorithm that mitigated the wastage of communication overhead and improved convergence rates, leaving good scope for practical deployment of FL on resource-constrained IoMT devices.

Similarly, [9] discussed the challenges in the application of FL in real-world scenarios. They proposed a hierarchical FL architecture that gave the feasibility of collaboration of devices of different computing capacities. Their results reflected the feasibility of improving the adaptability of FL in various IoMT environments.

3 Proposed Methodology

This section explains in detail the methodology used in developing and evaluating the federated learning framework using the CICIoMT2024 dataset. The methodology includes dataset

preparation, client simulation, data preprocessing, model training, aggregation of client predictions, and performance evaluation:

3.1 Dataset Selection

CICIoMT2024 was chosen as the dataset for analysis and the detection of IoMT-related attacks [13]. The preprocessing included memory optimization by reducing numerical features into smaller data types, cleaning inconsistency, such as replacing negative values with zeros and removing redundant information, such as zero-variance columns, duplicates, and identical-value columns. Infinite and NaN values were appropriately handled, and attack labels were categorized into broader groups like "DDoS," "DoS," "Spoofing," and "MQTT" to make meaningful analysis. These ensured that a clean and balanced dataset was selected for the robust IoMT attack detection models [14].

3.2 Client Simulation

To simulate an FL environment, the training dataset is partitioned among n = 5 clients, each modeling a localized IoMT environment, such that each client receives a stratified subset of D_{train}. The i[th] client's data, Di was defined as:

$$D_{train} = \bigcup_{i=1}^{n} Di \; and \; , Di \bigcup Dj = \emptyset \forall i \neq j$$

This captures the decentralized nature of IoMT networks where devices have different traffic patterns based on local conditions, and each client performs independent preprocessing by standardizing features so that they all have uniform scales and features with larger magnitudes do not overwhelm model training.

3.3 Data Preprocessing

Each client preprocesses its data individually to ensure consistency and improve model performance. The preprocessing pipeline has three primary steps: feature standardization to make sure that features are uniformly scaled and that features with larger magnitudes do not dominate the model training, handling missing values to ensure integrity of data, and encoding categorical variables for transforming them into numerical representations for the learning algorithms [12, 17]. These preprocessing steps are important for ensuring consistency across clients, removing biases, and preventing the distributed nature of the FL environment from adversely affecting the overall model performance [2, 8].

3.3.1 Standardization

The process of standardization ensures uniform scaling for all features and allows the model to converge well during training. Using the StandardScaler, every feature is standardized to have a zero mean and unit variance using the formula [8, 17]:

$$x^{j}{}_{scaled} = \frac{x_j - \mu_j}{\sigma_j}$$

Where μ_j and σ_j are the mean and standard deviation of the j[th] feature respectively. This enables features with larger magnitudes not to overwhelm the learning process and hence guarantees that all features contribute equally to training the model [17]. Normalizing the data in this way makes the training process stable and efficient, improving the performance and generalization of the model across diverse datasets.

3.4 Feature Selection

Feature selection is one of the most important aspects of improving model performance and reducing computational complexity, as it selects the most informative features from the dataset [12]. In this study, a two-step feature selection process is used to ensure that only the most relevant and non-redundant features are retained for training [7, 8, 17].

3.4.1 Variance Thresholding

The first step is to remove features with low variance. Features with low variability between samples are generally not able to learn a good representation and thus have little to offer in

separating different classes. This is the process of discarding those features that are insufficiently variable, leaving only those with a sufficiently high variance to contribute to the discriminative power of the dataset.

3.4.2 Mutual Information (MI) Based

The second step is to apply MI to the target variable to score the features variable. MI calculates how dependent each feature is on the target, including both linear and non-linear links. Features with higher MI scores are more correlated with the target variable and are more predictive. From this ranking, the top 20 features with the highest MI scores are chosen for the model's training. We ensure that our dataset is reduced to a small set of relevant features, preventing both overfitting and computational costs and increasing model accuracy through this two-step process. This helps models capture patterns in the data and make predictions, as they cannot see the big picture by keeping all of them.

3.5 Data Balancing

The SMOTETomek algorithm is used in this study as class imbalance is a common problem in IoMT security datasets, where the number of benign traffic samples is much higher than that of malicious samples [17]. As a result, it may learn biased representations that favor the majority class, resulting in poor detection of minority-class events like malicious traffic. The SMOTETomek algorithm is a combination of two opposite processes, one is the Synthetic Minority Oversampling Technique (SMOTE), and the other one is Tomek Links Removal which helps to balance the dataset and prevents overfitting.

3.5.1 SMOTE

SMOTE is used to create synthetic samples for the minority class by interpolating existing samples. For the minority class sample x_i, the algorithm selects one of its nearest neighbors $x_{neighbor}$ from the same class and generates a synthetic sample $x_{synthetic}$ as follows:

$$x_{synthetic} = x_i + \lambda.(x_{neighbor} - x_i)$$

Where λ is a random value in the range [0,1]. By generating synthetic samples along the line connecting x_i and $x_{neighbor}$, SMOTE introduces diversity into the minority class, reducing the risk of overfitting.

3.5.2 Tomek Link Removals

After oversampling, Tomek Links Removal is applied to clean the dataset by removing overlapping or ambiguous samples near the decision boundary between classes. A pair of samples (x_i, x_j), where x_i belongs to the minority class and x_j belongs to the majority class, forms a Tomek Link as follows:

$$d(x_i, x_j) < d(x_i, x_k) \ \ and \ d(x_i, x_j) < d(x_j, x_k) \ \forall \ x_k$$

where $d(x_i, x_j)$ represents the distance between x_i and x_j. Such pairs are considered noisy or overlapping samples. By removing these Tomek Links, the algorithm reduces ambiguity, making the dataset cleaner and more separable. After applying SMOTETomek, the dataset achieves a balanced distribution, where the number of samples in the minority class ($N_{minority}$) is approximately equal to that of the majority class ($N_{majority}$). This balanced dataset ensures that the model can learn patterns from both classes equally, improving its ability to detect malicious traffic without bias toward benign traffic.

3.6 Model Training

Each client trains five distinct models, leveraging both ML and DL techniques. The models include Random Forest (RF), XGBoost, CatBoost, LightGBM, and a Neural Network (NN).

Each model is trained independently on preprocessed data at the client level, leveraging its unique architecture and algorithmic strengths.

3.6.1 Random Forest (RF)

The RF model is a strong ensemble learning technique that utilizes an ensemble of multiple decision trees to improve classification and prevent overfitting [15]. Training each tree is done on an independent random subset of the dataset, and at each split, a random subset of features is considered to avoid similarity among trees [12]. Prediction is made by aggregating using majority voting for classification problems, so the model is very resistant to noise and overfitting. Critical hyperparameters like the number of trees and the maximum depth of individual trees are tuned for better performance.

3.6.2 XGBoost

XGBoost is a very efficient gradient-boosting algorithm, that builds trees sequentially, and every new tree is focused on correcting the errors of its predecessors. It uses regularization techniques like L1 and L2 to reduce overfitting and stabilize models [15]. XGBoost is an optimized speed that uses parallel tree construction, thereby making it efficient in handling very large datasets. The algorithm can also handle missing data and produce interpretable feature importance scores such that users know which predictors to use in their dataset [2].

3.6.3 CatBoost

CatBoost is a gradient-boosting model specifically designed to handle categorical features efficiently, without requiring extensive preprocessing such as one-hot encoding. It uses ordered boosting, which introduces controlled randomness during training and helps reduce overfitting while maintaining prediction stability. CatBoost builds symmetric trees, which are computationally efficient and consistent in making predictions. This model is efficient in datasets containing a mix of numerical and categorical data, as it uses its ability to model complex feature interactions quite effectively. Besides, CatBoost is user-friendly and robust and thus preferred in real-world applications for handling heterogeneous datasets.

3.6.4 LightGBM

LightGBM, an abbreviation for Light Gradient Boosting Machine, is a highly performing and scalable gradient-boosting framework. In contrast to the usual gradient-boosting techniques, LightGBM constructs trees in a leaf-wise fashion, emphasizing the splitting of the leaf that holds the highest potential to reduce the loss [15]. This helps in achieving greater accuracy with more efficiency and hence is beneficial, especially for large data sets. Moreover, LightGBM uses histogram-based learning where continuous features are discretized into bins, which significantly reduces memory consumption and training time. This model has proved to be highly effective for datasets having high-dimensional features with rapid training without concomitant loss of prediction accuracy. Added feature importance scores interpret decision-making processes easier.

3.6.5 Neural Network (NN)

The NN architecture used in this study is a deep feedforward architecture, capturing complex, non-linear relationships within the dataset. This architecture comprises three dense layers [17]. Each layer of architecture is equipped with LeakyReLU activation to avoid vanishing gradients. Batch normalization was also added to stabilize learning, and dropout layers were included to prevent overfitting [18]. For the output layer, softmax activation was used for multi-class classification. The model was trained using Sparse Categorical Cross-Entropy Loss and optimized using the Adam optimizer for stable and efficient convergence. NN are specifically suited for high-dimensional data, and they are excellent at identifying intricate patterns and interactions; therefore, they are highly effective for IoMT traffic classification [7].

3.7 Weighted Aggregation

In the FL framework, weighted aggregation is applied to integrate the predictions of the client models so that those with a higher accuracy make a greater contribution to the global model

[18]. The global prediction of an input x is calculated as the weighted average of the predictions made by the clients [19]. The weights are proportional to the accuracy of each client. Global prediction is defined as:

$$P_{global}(x) = \frac{\sum_{i=1}^{n} A_i \,.\, P_i(x)}{\sum_{i=1}^{n} A_i}$$

Here, n is the total number of clients, A_i represents the accuracy of the ith client, and $P_i(x)$ denotes the probability prediction from the ith clients. For multiclass classification, the aggregation is performed independently for each class c, allowing the global probability for class c to be expressed as:

$$P^c{}_{global}(x) = \frac{\sum_{i=1}^{n} A_i \,.\, P^c{}_i(x)}{\sum_{i=1}^{n} A_i}$$

The final prediction class $\hat{y}$ is determined by selecting the class with the highest global probability, $\hat{y} = \arg max_c \, P^c{}_{global}(x)$. The accuracy A_i of each client is computed as the ratio of correct predictions to total predictions, ensuring that reliable clients contribute more heavily to the global model. The solution is also resilient to differences in client performance and data distributions in the IoMT network, addressing well the heterogeneity problem in such networks. Weighted aggregation adds reliability to the global model since it uses the strengths of well-performing clients while minimizing the influence of not-so-accurate ones; hence, making the framework scalable and adaptable to decentralized environments.

3.8 Evaluation Matrix

The FL framework is measured by the following metrics:

3.8.1 Accuracy

Measures the overall correctness of predictions [14].

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

3.8.2 Precision and Recall

Analyze the performance of each class, especially in terms of detecting malicious traffic [1, 12].

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

3.8.3 F-1 Score

Balances precision and recall for imbalanced classes [7].

$$F1 - Score = \frac{2 * (Recall - Precision)}{(Recall + Precision)}$$

3.8.4 ROC-AUC

It measures the model's ability to distinguish between classes [1, 8].

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

3.8.5 Confusion matrix

This gives the complete breakdown of True Positives (TP) against False Positives (FP), True Negatives (TN), and False Negatives (FN) [18].

4 Results and Discussions

The results of this federated learning framework applied over the CICIoMT2024 dataset show consistent performance and robust results for both types of models used: client-specific models and global aggregated models. Each client of the clients developed on a partial subset of their localized data demonstrated high accuracy-from 98.69% up to 98.87% accuracy represented in **Table 1**.

Table 1. Accuracy Achieved by Each Client.

Clients	Accuracy
Client 1	98.72%
Client 2	98.69%
Client 3	98.75%
Client 4	98.86%
Client 5	98.87%

The minimally possible variation here really shows the proper effectiveness of our preprocessing pipeline composed of feature standardization, Mutual Information-based feature selection, and applying SMOTETomek methods for class balance. The slightly better performance of Clients 4 and 5 shows that their data had fewer irregularities or perhaps a more balanced distribution, whereas Client 2, with an accuracy of only 98.69%, faced perhaps more complex data patterns. In any case, the narrow spread of accuracies across all clients points to how well the framework can handle data heterogeneity.

Fig. 3 shows the accuracy obtained by each client in the federated learning process. The accuracy values for Clients 1, 2, 3, 4, and 5 are high and range between 98.69% and 98.87%. Client 1 had an accuracy of 98.72%, while Client 2 had a slightly lower accuracy of 98.69%. Client 3 excelled at 98.75%, and Clients 4 and 5 displayed the highest accuracies of 98.86% and 98.87%, respectively. The tight range of the results indicates a robust and reliable preprocessing pipeline that utilized standardization, feature selection, and class balancing with SMOTETomek. The consistent results across clients imply that the federated learning framework effectively mitigates the impact of heterogeneous data distributions. All these clients meaningfully contribute to the global model in a way that is both fair and scalable. The near parity of accuracy also signifies stability within the framework, which is necessary for decentralized IoMT environments.

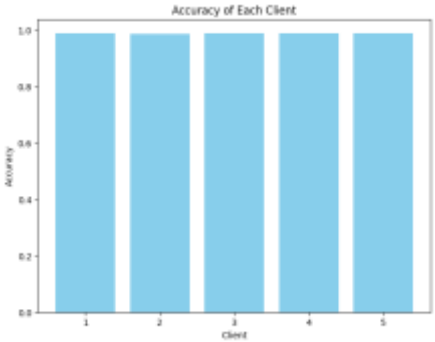

Fig. 3. Accuracy of each client.

Fig. 4 shows the precision and recall curves for each client, showing that the model achieves a good balance between true positive rates and minimal false positive rates across all classes. Precision is defined as the proportion of correctly classified instances out of all instances predicted to belong to a specific class, whereas recall is the proportion of actual positives correctly identified by the model. For all five clients, the curves show near-perfect precision and recall values, consistently close to 1.00. Client 1 shows excellent precision and recall for all classes, meaning that the model can accurately identify malicious traffic with minimal false alarms. Similarly, Client 2 follows a similar trend, with precision and recall values very close to each other for all classes, which further validates the robustness of the framework. From the trends shown by Clients 1 and 2, it can be inferred that Clients 3, 4, and 5 retain the same level of precision and recall as above, thereby ensuring the robustness of the federated learning framework. The almost perfect overlap between precision and recall for all clients ensures that the model does not degrade on either benign or malicious traffic. This outcome proves that the preprocessing steps involved in this technique, which are feature selection and SMOTETomek, contribute to strengthening the model in its predictive capacity. Also, this outcome reflects the effectiveness of the weighted aggregation strategy applied here for fair and reliable contributions of all clients that led to the successful global model excelling both at accurate threat detection and minimal false positives.

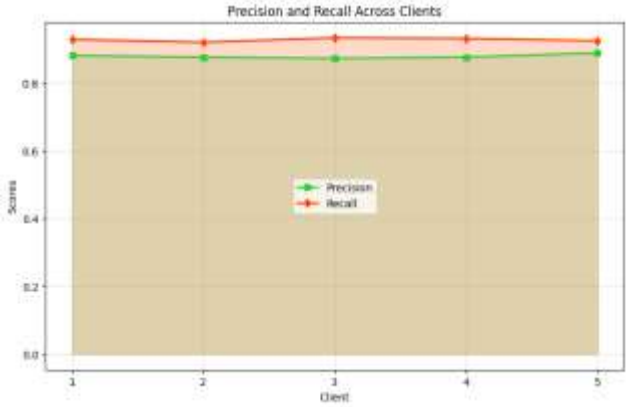

Fig. 4. Precision and Recall Curve for each client.

Fig. 5 Global model confusion matrix, illustrating the classification performance for distinguishing benign and malicious traffic. The matrix represents a tabular format of TP, TN, FP, and FN, hence giving a very detailed comparison between the model's predictions and the actual labels. The diagonal elements represent correctly classified instances, whereas off-diagonal elements represent misclassifications. The matrix shows a high number of true negatives and a minimal false positive rate of 2.6% for the benign traffic class, meaning that the model can identify benign traffic with negligible misclassification as malicious. For malicious traffic, the TPR is an impressive 95.4%, which reflects the model's strong capability to detect threats accurately. Fewer numbers of false negatives reflect very few malicious instances incorrectly classified as benign, implying good sensitivity toward detection.

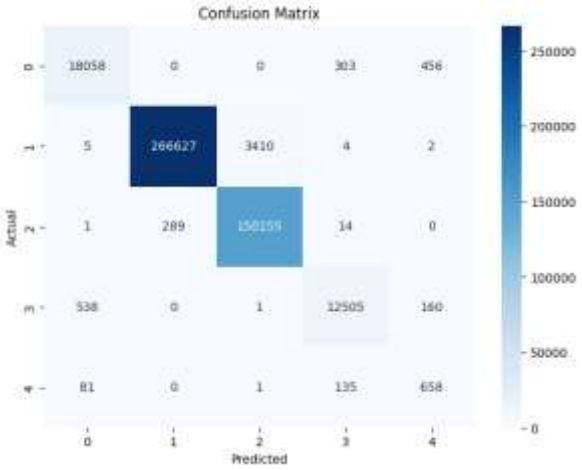

Fig. 5. Confusion Matrix.

Fig. 6 to 10 show the Receiver Operating Characteristic (ROC) curves for Clients 1 through 5, which shows how well each class balances the sensitivity and specificity of the model. For all clients, the ROC curves follow the top-left corner, indicating that each class has a perfect AUC of 1.00. These results demonstrate the model's flawless ability to distinguish between benign and malicious traffic, reflecting high true positive rates and low false positive rates. As in the case of Client 1, AUC values confirm excellent class boundary identification; this trend repeats for Clients 2, 3, 4, and 5. The nearly identical results across the clients show how robust the federated learning framework is, considering that the global model generalizes well to varying datasets while also having uniform accuracy. The ROC AUC curves validate the effectiveness of preprocessing techniques, different model architectures, and the weighted aggregation strategy to ensure equitable and reliable performance for all clients, which makes the system highly suitable for real-world IoMT security applications.

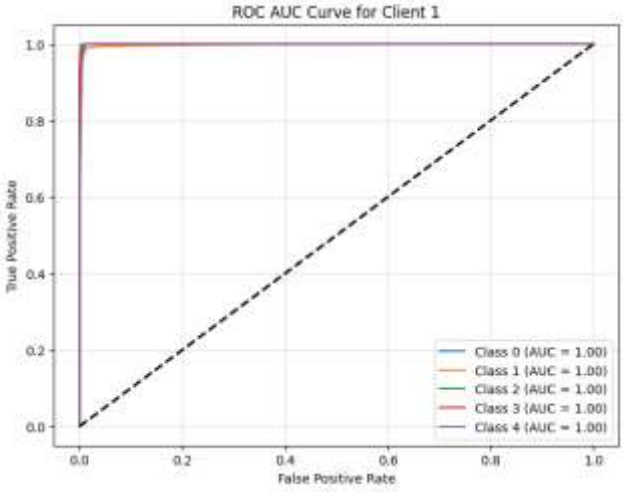

Fig 6. Receiver Operating Characteristic (ROC) Curve Depicting AUC for Client 1.

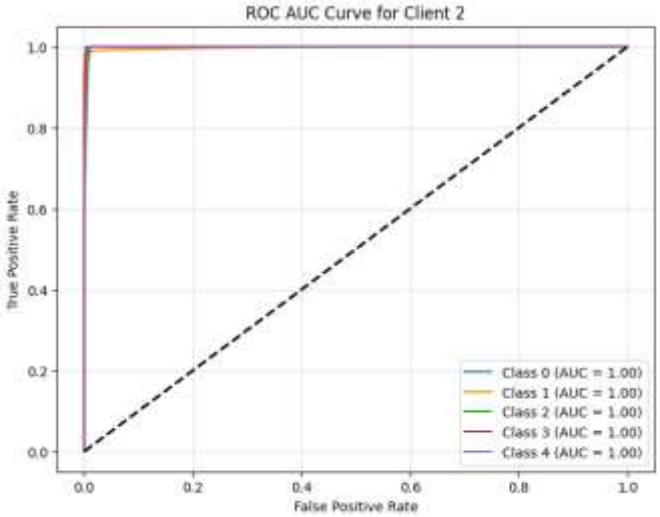

Fig. 7. Receiver Operating Characteristic (ROC) Curve Depicting AUC for Client 2.

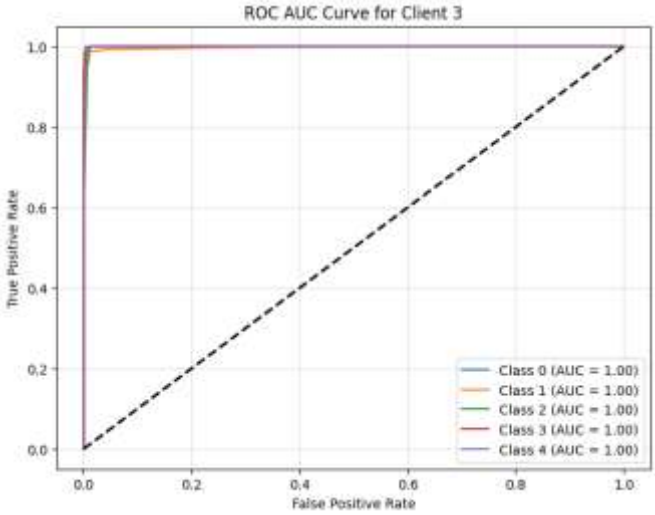

Fig. 8. Receiver Operating Characteristic (ROC) Curve Depicting AUC for Client 3.

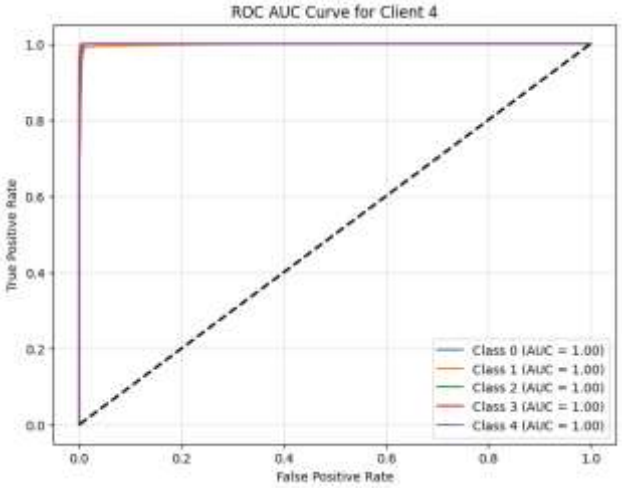

Fig. 9. Receiver Operating Characteristic (ROC) Curve Depicting AUC for Client 4.

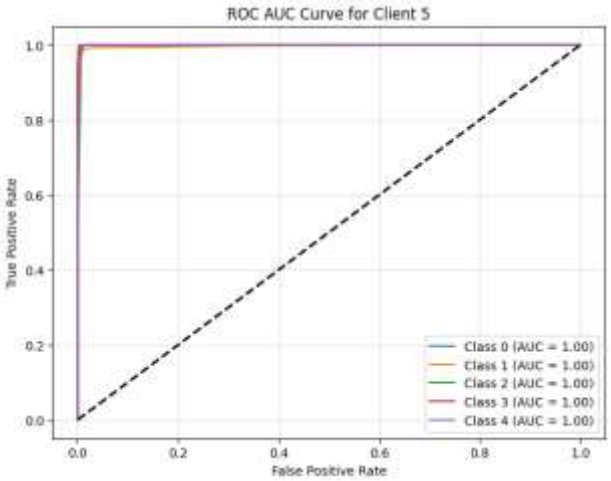

Fig. 10. Receiver Operating Characteristic (ROC) Curve Depicting AUC for Client 5.

5 Conclusion and Future Scope

The proposed federated learning framework for securing IoMT networks demonstrates exceptional performance in high accuracy and reliability in the detection of malicious traffic while preserving data privacy. Strong precision and recall metrics for both benign and malicious traffic, along with a superior accuracy of 98.78% in the global model, reflect the effectiveness of the preprocessing pipeline, diverse model architectures, and weighted aggregation strategy. This framework successfully addresses the challenges of IoMT network decentralization and heterogeneity while ensuring robust threat detection without scaling down through a decline in scalability or data security. Further ahead, this work provides ready avenues for further development, such as adaptive aggregation techniques in the light of evolving environmental configurations for IoMT or integration of resource-constrained devices with lightweight models and real-time threat detection mechanisms. Additionally, the use of explainability frameworks such as Shapley values may add better interpretability to the model decisions and create much trust in IoMT security systems. This approach continues to evolve and extend its development with the potential to form a scalable and privacy-preserving foundation for future IoMT deployments in increasingly complex healthcare ecosystems.

References

1. Abbas, S., Al Hejaili, A., Sampedro, G. A., Abisado, M., Almadhor, A. S., Shahzad, T., & Ouahada, K. (2023). A novel federated edge learning approach for detecting cyberattacks in IoT infrastructures. IEEE Access, 11, 112189-112198.
2. Issa, W., Moustafa, N., Turnbull, B., Sohrabi, N., & Tari, Z. (2023). Blockchain-based federated learning for securing internet of things: A comprehensive survey. ACM Computing Surveys, 55(9), 1-43.
3. Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for internet of things: A comprehensive survey. IEEE Communications Surveys & Tutorials, 23(3), 1622-1658.
4. Abd Al-Ameer, A. A., & Bhaya, W. S. (2023). Federated learning security mechanisms for protecting sensitive data. Bulletin of Electrical Engineering and Informatics, 12(4), 2421-2427.
5. Wang, T., Tang, T., Cai, Z., Fang, K., Tian, J., Li, J., ... & Xia, F. (2024). Federated Learning-based Information Leakage Risk Detection for Secure Medical Internet of Things. ACM Transactions on Internet Technology.
6. Bouacida, N., & Mohapatra, P. (2021). Vulnerabilities in federated learning. IEEE Access, 9, 63229-63249.

7. Ma, C., Li, J., Ding, M., Yang, H. H., Shu, F., Quek, T. Q., & Poor, H. V. (2020). On safeguarding privacy and security in the framework of federated learning. IEEE network, 34(4), 242-248.
8. Metwaly, A. A., & Elhenawy, I. (2023). Protecting IoT Devices from BotNet threats: a federated machine learning solution. Sustainable Machine Intelligence Journal, 2, 5-1.
9. Wang, R., Lai, J., Li, X., He, D., & Khan, M. K. (2024). RPIFL: Reliable and Privacy-Preserving Federated Learning for the Internet of Things. Journal of Network and Computer Applications, 221, 103768.
10. Aljrees, T., Kumar, A., Singh, K. U., & Singh, T. (2023). Enhancing IoT Security through a Green and Sustainable Federated Learning Platform: Leveraging Efficient Encryption and the Quondam Signature Algorithm. Sensors, 23(19), 8090.
11. Jagarlamudi, G. K., Yazdinejad, A., Parizi, R. M., & Pouriyeh, S. (2024). Exploring privacy measurement in federated learning. The Journal of Supercomputing, 80(8), 10511-10551.
12. Zhong, C., Sarkar, A., Manna, S., Khan, M. Z., Noorwali, A., Das, A., & Chakraborty, K. (2024). Federated learning-guided intrusion detection and neural key exchange for safeguarding patient data on the internet of medical things. International Journal of Machine Learning and Cybernetics, 1-31.
13. Zahid, M., & Bharati, T. S. Comprehensive Review of IoT Attack Detection Using Machine Learning and Deep Learning Techniques.
14. Dadkhah, S., Neto, E. C. P., Ferreira, R., Molokwu, R. C., Sadeghi, S., & Ghorbani, A. (2024). Ciciomt2024: Attack vectors in healthcare devices-a multi-protocol dataset for assessing iomt device security.
15. Malik, A., Hussain, M. Z., Mustafa, M., Sattar, B., Ali, J., & Altaf, J. (2024). SecureNet: A Convergence of ML, Blockchain and Federated Learning for IoT Protection. UCP Journal of Engineering & Information Technology (HEC Recognized-Y Category), 2(1), 24-35.
16. Wang, C., Wu, X., Liu, G., Deng, T., Peng, K., & Wan, S. (2022). Safeguarding cross-silo federated learning with local differential privacy. Digital Communications and Networks, 8(4), 446-454.
17. Ferrag, M. A., Friha, O., Maglaras, L., Janicke, H., & Shu, L. (2021). Federated deep learning for cyber security in the internet of things: Concepts, applications, and experimental analysis. IEEE Access, 9, 138509-138542.
18. Hamdi, N. (2023). Federated learning-based intrusion detection system for Internet of Things. International Journal of Information Security, 22(6), 1937-1948.
19. Asad, M., Moustafa, A., & Yu, C. (2020). A critical evaluation of privacy and security threats in federated learning. Sensors, 20(24), 7182.
20. Rahman, M. A., Hossain, M. S., Islam, M. S., Alrajeh, N. A., & Muhammad, G. (2020). Secure and provenance enhanced internet of health things framework: A blockchain managed federated learning approach. Ieee Access, 8, 205071-205087.