Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

A Robust Federated Learning Framework for Healthcare Across IID and Non-IID Data Distributions

Mohammed Kamel BENKADDOUR [1], Mohammed El Aymene BEYAT [1], Mohammed El Amine BEYAT [1], Abdessalam MOUALDI [2].

[1] *Department of Computer Science and Information Technology, Laboratory of Artificial Intelligence and Information Technologies, University of Kasdi Merbah, 30000, Ouargla, Algeria.*

[2] *Department of Electronics and Telecommunications, Laboratory of Electrical Engineering, University of Kasdi Merbah, 3000, Ouargla, Algeria.*

*Corresponding author: Mohammed Kamel Benkaddour. Email: Benkaddour.kamel@univ-ouargla.dz

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

Privacy-preserving machine learning has become essential in healthcare, where sensitive patient data cannot be centralized without risking confidentiality and regulatory non-compliance. Federated learning (FL) offers a viable alternative by enabling collaborative model training while retaining data on local medical institutions. This study presents a robust federated learning framework designed to maintain strong predictive performance across both independent and identically distributed (IID) and non-IID data scenarios, reflecting realistic variability in healthcare environments. Using EfficientNet-B0 as the core architecture and the PathMNIST dataset as the benchmark, we evaluate the framework across federations of 5 and 10 clients, systematically comparing centralized and federated setups. Experimental results demonstrate that the proposed framework achieves 95.29% accuracy under IID conditions with 5 clients and 94.99% with 10 clients, closely matching centralized performance. Under non-IID distributions generated via a Dirichlet partitioning, the framework maintains competitive performance with 94.26% accuracy for 5 clients and 93.20% for 10 clients. Additional metrics highlight the system's robustness: precision reaches up to 95.38%, recall up to 95.35%, and F1-score up to 95.23% in centralized benchmarking, with only marginal degradation under federated settings. Convergence curves show stable optimization in IID scenarios and controlled fluctuations under non-IID heterogeneity, confirming the resilience of the federated averaging strategy. These findings demonstrate that the proposed federated learning framework delivers high model utility while ensuring decentralized data governance, making it suitable for scalable, privacy-conscious medical image analysis.

Keywords: Federated Learning, Privacy Preservation, Non-IID Distributuion, Secure Aggregation, PathMNIST.

1. INTRODUCTION

Data security and privacy preservation are critical challenges in domains handling sensitive information, particularly in healthcare [1]. Traditional centralized learning methods require aggregating all data in a single location, increasing the risk of privacy breaches and unauthorized access [2]. Federated learning (FL) has emerged as a promising solution, enabling multiple institutions to collaboratively train a global model without exchanging raw data [3]. In this framework, each client trains a local model on its private dataset, and only model updates are shared with a central server for aggregation. This decentralized approach mitigates privacy risks while allowing effective distributed learning across multiple organizations.

Despite its advantages, federated learning faces challenges when client data are heterogeneous or non-identically distributed (non-IID). Such statistical heterogeneity can affect model convergence and performance, particularly in

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

practical healthcare scenarios with diverse data sources. Addressing these challenges requires frameworks that maintain high model accuracy and stable convergence across IID and non-IID data distributions. In this work, we present a federated learning framework tailored for privacy-sensitive healthcare applications. The system orchestrates local model training across multiple medical centers and securely aggregates updates to construct a high-quality global model. Experiments on the PathMNIST dataset [4], using configurations of 5 and 10 clients under

both IID and non-IID settings, demonstrate that the proposed framework closely matches centralized performance, achieving 94.79 % accuracy with stable convergence. These results highlight the framework's resilience to statistical heterogeneity, computational efficiency, and practical applicability in decentralized medical environments. Privacy-preserving federated learning has emerged as a key paradigm for collaborative model training in healthcare, allowing institutions to jointly benefit from distributed medical data without compromising patient confidentiality. Building upon foundational mechanisms such as differential privacy and secure aggregation, recent research has focused on improving the robustness of federated learning under heterogeneous and non-IID data distributions that commonly arise in real-world medical environments.

Adaptive aggregation strategies have been introduced to enhance convergence and model stability by dynamically adjusting the contribution of local updates according to data divergence, achieving improved performance on medical datasets such as tuberculosis chest X-rays and brain tumor MRIs [5]. Frameworks such as Health-FedNet provide scalable architectures for privacy-preserving medical analytics across institutions [6], while MultiProg addresses feature heterogeneity in multi-source Electronic Health Records through multi-channel architectures and feature calibration techniques [7]. These advances highlight the growing maturity of federated learning frameworks designed for healthcare applications, emphasizing both privacy preservation and learning efficiency under diverse data conditions. At the same time, recent studies continue to stress the importance of ethical and legal compliance, including informed patient consent and adherence to healthcare data protection regulations during deployment of federated learning systems [8].

2. PROPOSED METHODOLOGY

A. System Overview

The proposed framework implements standard federated learning (FL) for privacy-sensitive healthcare applications [1]. The system enables multiple medical centers to collaboratively train a global model without sharing raw patient data. Each client trains a local model on its private dataset and sends model updates to a central server for aggregation [9]. The server constructs a high-quality global model by combining the updates, which is then redistributed to clients for the next training round. This approach allows effective distributed learning while mitigating privacy risks associated with centralized data aggregation.

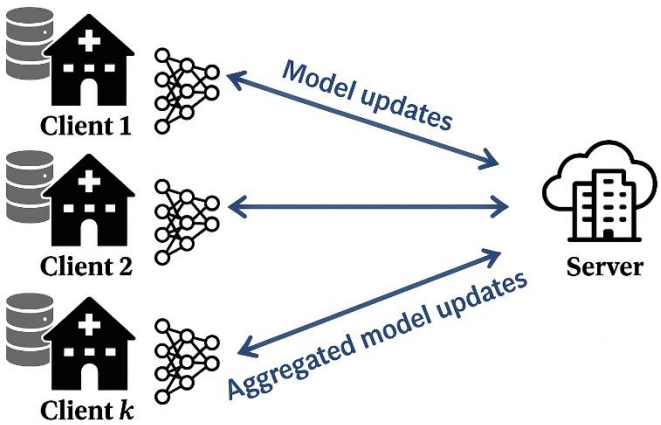

Fig 1: Overview of the proposed federated learning framework.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

We describe the workflow using simplified algorithmic representations of the client update and global aggregation procedures. These algorithms reflect the standard federated learning pipeline used in our experiments on the PathMNIST dataset [4].

Algorithm 1 describes the local training process performed independently on each healthcare client, where the model is updated using private data and the resulting weight differences are computed without exposing any raw information.

Algorithm 1 — ClientUpdate	
Require:	**Local model W_t, local dataset D_c, learning rate η**
Ensure:	**Local model update ΔW_t^(c)**
1	**Initialize local model W_t^(c) ← W_t**
2	**Train W_t^(c) on D_c using gradient descent or optimizer**
3	**Compute update : ΔW_t^(c) = W_t^(c) − W_t**
4	**return ΔW_t^(c)**

Algorithm 2 outlines the global aggregation step using Federated Averaging, combining all client updates proportionally to their dataset sizes to produce a new global model that reflects collaborative learning across heterogeneous environments.

Algorithm 2 — FederatedAveraging	
Require :	**Global model W_t, client updates {ΔW_t^(c)}_{c=1..K}, client dataset sizes {n_c}**
Ensure:	**Updated global model W_{t+1}**
1	**Compute total samples N = sum_{c=1..K} n_c**
2	**Initialize aggregated update ΔW_t = 0**
3	**for c = 1 to K do**
4	**ΔW_t ← ΔW_t + (n_c / N) * ΔW_t^(c)**
5	**end for**
6	**Update global model: W_{t+1} ← W_t + ΔW_t**
7	**return W_{t+1}**

B. Experimental Environment

1) Hardware and Execution Platform

All experiments were performed using Google Colab [10], equipped with an NVIDIA L4 GPU (22.5 GB VRAM), an Intel Xeon CPU, and approximately 53 GB of RAM. This environment provides a scalable and reliable platform for evaluating federated learning experiments.

2) Software Stack

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

The framework was implemented in Python using PyTorch for deep learning. Federated orchestration was managed with Flower [11], allowing efficient client-server coordination and global model aggregation.

3) Dataset and Configurations

Experiments were conducted on the PathMNIST dataset [4]. We evaluated configurations with 5 and 10 clients under both IID and non-IID data distributions to examine the framework's ability to maintain stable convergence and high accuracy in heterogeneous environments.

3. EXPERIMENTAL SETUP

A. Dataset Description

Given the privacy-sensitive nature of healthcare data, our evaluation focuses on medical image classification. We employed the PathMNIST subset from the MedMNIST v2 benchmark [4], which is a standardized collection designed for lightweight, reproducible experimentation in biomedical imaging.

PathMNIST is derived from colorectal histopathology datasets (NCT-CRC-HE-100K and CRC-VAL-HE-7K) and contains 107,180 RGB image patches resized to 28×28 pixels. The classification task involves nine tissue types, making it a 9-class multi-class problem. The dataset has 89,996 training, 10,004 validations, and 7,180 test samples. An overview of this process is provided in Table I.

Table I: PathMNIST Dataset Overview.

Aspect	Details
Modality	Colon Histopathology
Total samples	107,180
Training / Validation / Test	89,996 / 10,004 / 7,180
Image format	RGB, 28×28 pixels
Classes	9 tissue types
Task	Multi-class classification

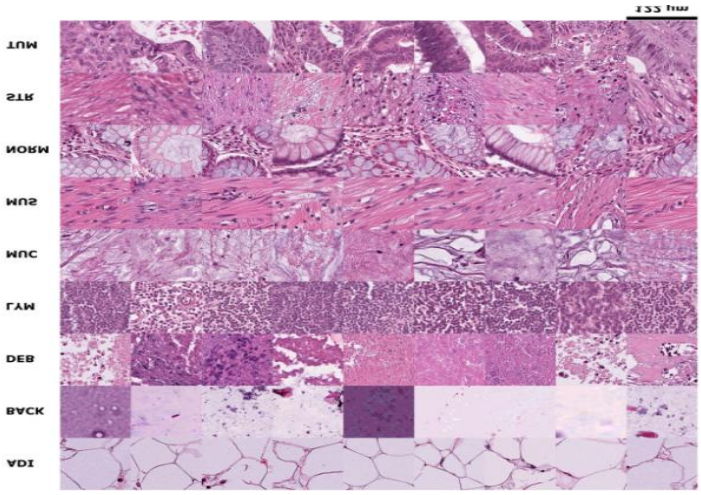

Fig 2: Example images from each of the nine tissue classes in PathMNIST.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Data preprocessing included:

• Tensor Conversion: Converting images to tensors for GPU efficiency.

• Normalization: Scaling pixel intensities to [0, 1].

• Data Augmentation: Random flips, rotations, and crops to improve generalization.

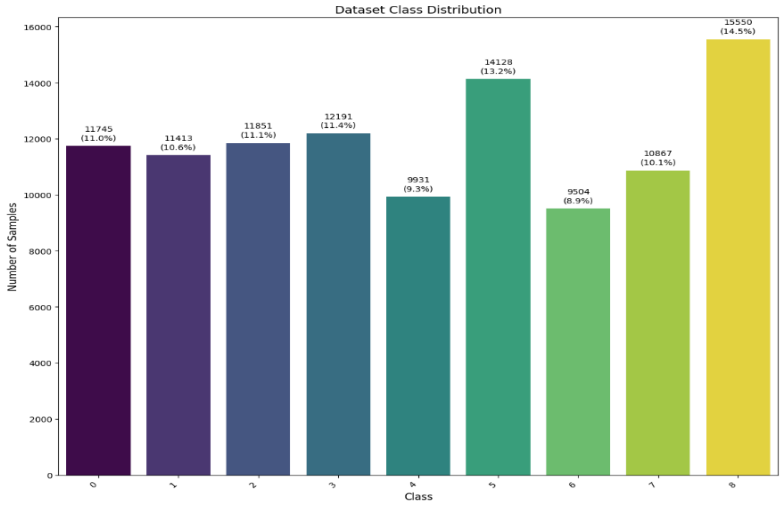

Fig 3: Class distribution in PathMNIST.

B. Performance Evaluation Metrics

We evaluated the framework under centralized and federated settings with 5 and 10 clients, using the following standard classification metrics:

1) Accuracy:

$$Accuracy = (Number\ of\ Correct\ Predictions) / (Total\ Number\ of\ Predictions)$$

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \quad \text{.....(1)}$$

2) Precision:

$$Precision = TP / (TP + FP) \quad \text{.....(2)}$$

3) Recall:

$$Recall = TP / (TP + FN) \quad \text{.....(3)}$$

4) 1-Score:

$$F1 = 2 \times (Precision \times Recall) / (Precision + Recall) \quad \text{.....(4)}$$

5) Confusion Matrix: For multi-class evaluation, we define the confusion matrix C ∈ RC×C , where C is the number of classes. Each element Ci, j represents the number of samples whose true label is i and predicted label is j.

Where T P , T N , F P , and F N are True Positives, True Negatives, False Positives, and False Negatives, respectively.

C. Computational Performance Metrics

To assess computational efficiency of federated learning compared to centralized training, we measured:

• Training Time: Duration of local model updates per client.

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

• Communication Time: Time for model updates to be transmitted and aggregated across clients.

• Total Round Time: End-to-end time per communication round.

4. RESULTS

Experiments were conducted under IID and non-IID data distributions with 5 and 10 clients. Each client locally updated the model for a fixed number of epochs, and the global model was obtained by averaging client weights after each communication round. The protocol ensured stable convergence across heterogeneous client data.

A. Centralized Benchmarking of Deep Learning Architectures

Before deploying our federated learning framework, we performed extensive benchmarking with centralized deep learning models to determine the most performant architecture for downstream evaluation. This preliminary stage was essential for establishing a high-quality baseline and ensuring a fair comparison between federated and distributed setups. We evaluated both custom convolutional neural networks and state-of-the-art pretrained models using the PathMNIST dataset. The models considered included CNNMed (with and without self-attention enhancement), ResNet-18 [14], MobileNetV3 Small [15], and EfficientNet-B0 [16].

1) Custom CNN Architectures:

CNNMed is a lightweight convolutional neural network designed for histopathological image classification [17]. It consists of three convolutional layers interspersed with batch normalization, ReLU activations, and maximum pooling operations [18]. Dropout and adaptive average pooling were used to ensure regularization and feature dimension consistency. The classification head consists of two fully connected layers, followed by a 9-class output layer. An improved variant, CNNMed with self-attention, incorporates a spatial attention mechanism to improve feature extraction, selectively emphasizing informative regions in the image to improve classification accuracy while incurring minimal computational overhead.

2) Pre-trained Architectures:

To leverage prior knowledge from large-scale image models, we fine-tuned three well established architectures:

•ResNet-18: A residual network mitigating vanishing gradients via identity shortcuts [14].

•MobileNetV3 Small: A resource-efficient model designed for low-power devices [15]

•EfficientNet-B0: A highly optimized model employing compound scaling [16].

These models were selected for their generalization capabilities and varying complexity levels, making them suitable for comparative analysis.

3) Results of Centralized Architectures:

All models were trained with a batch size of 64, an initial learning rate of 0.001, and up to 30 epochs (10 for pretrained networks to avoid overfitting). Evaluation employed standard classification metrics: precision, recall, F1-score, and accuracy.

Table II: Comparison of Centralized Deep Learning Models

Model	Epochs	Precision	Recall	F1 -score	Accuracy
CNNMed	30	93.57	93.55	93.33	93.55
CNNMed + Self-Attn.	30	94.00	94.11	93.94	94.11
MobileNetV3	10	94.83	94.85	94.82	94.85
ResNet-18	10	95.15	95.04	94.91	95.04
EfficientNet-B0	10	95.38	95.35	95.23	95.35

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

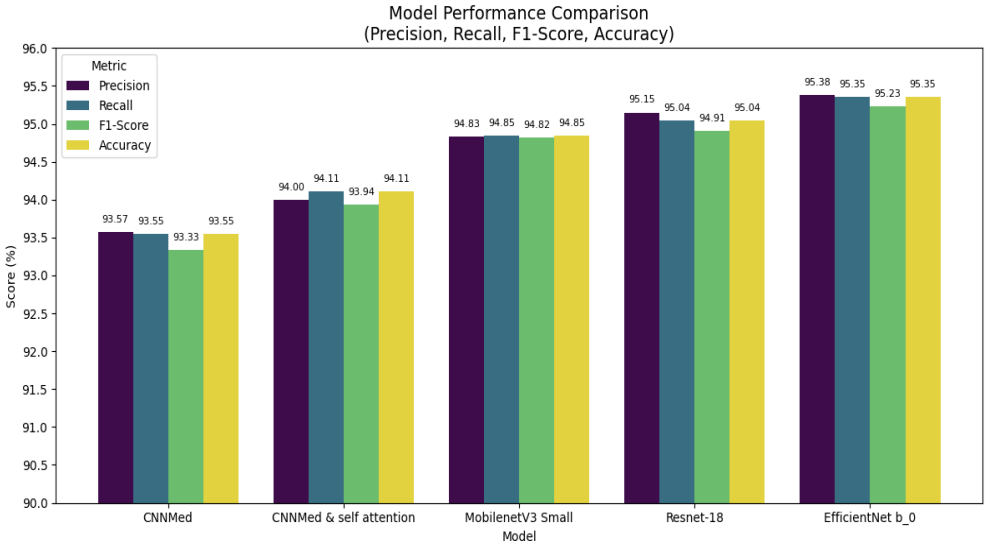

Fig 4: Model performance comparison (accuracy, precision, recall, F1).

EfficientNet-B0 achieved the highest accuracy and was selected as the baseline for subsequent federated experiments.

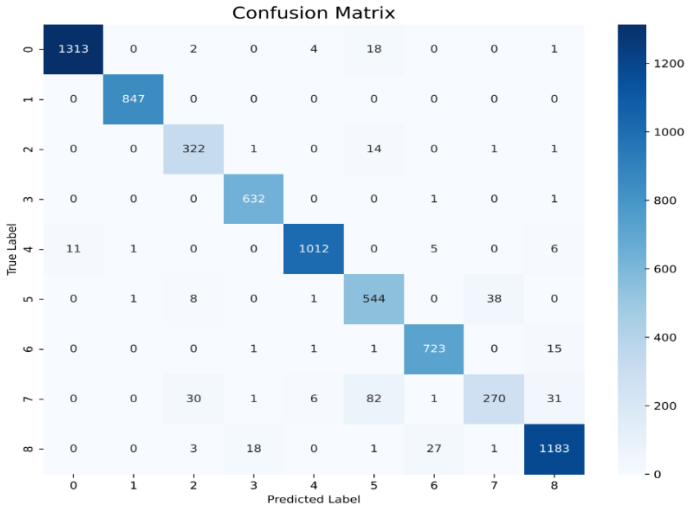

Fig 5: Confusion matrix for EfficientNet-B0.

B. Federated Learning under IID and Non-IID Distributions

We evaluated the federated learning paradigm using EfficientNet-B0 across two data distribution regimes: IID and non-IID. Experiments were conducted with 5 and 10 clients, trained with the Federated Averaging (FedAvg) algorithm. Hyperparameters are summarized in Table III.

Table III: Federated Learning Configuration

Hyperparameter	Settings
Client Model Architecture	EfficientNet-B0
Number of clients	5 and 10
Data distribution types	IID, Non-IID (Dirichlet α = 0.5)
Communication rounds	10

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Local batch size	64
Initial LR	0.001
Aggregation strategy	FedAvg

1) IID Distribution: Uniform Statistical Balance: In the IID scenario, each client received a balanced subset of data. Figure 6 shows accuracy and loss curves for 5 and 10 clients. Training was stable and performance degradation with more clients was marginal.

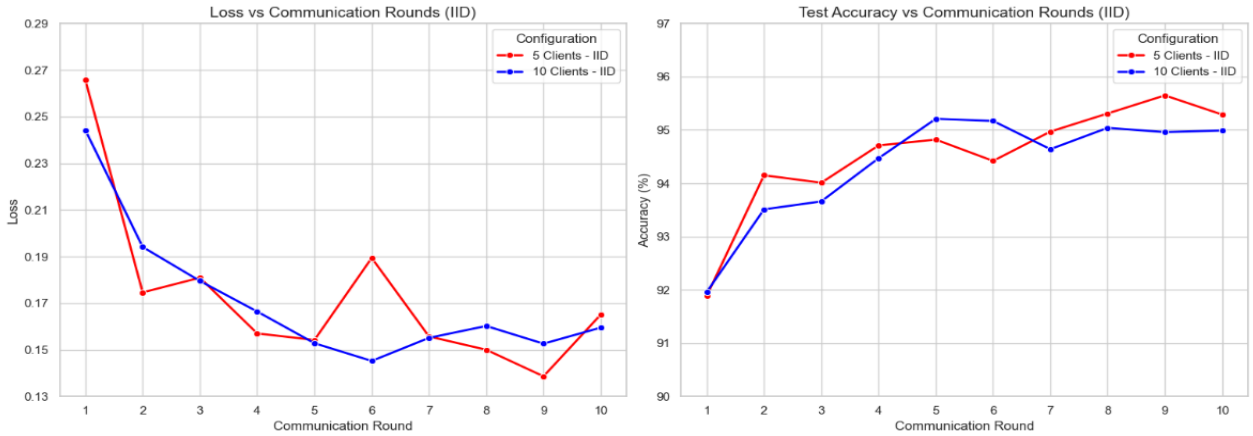

Fig 6: Federated Learning Accuracy and Loss Curves (IID).

2) Non-IID Distribution: Real-World Statistical Heterogeneity: For non-IID data, we used a Dirichlet distribution (α = 0.5) to induce class imbalance. Learning curves for 5 and 10 clients under non-IID conditions are shown in Figure 7. Training is less stable and fluctuations are more pronounced with ten clients.

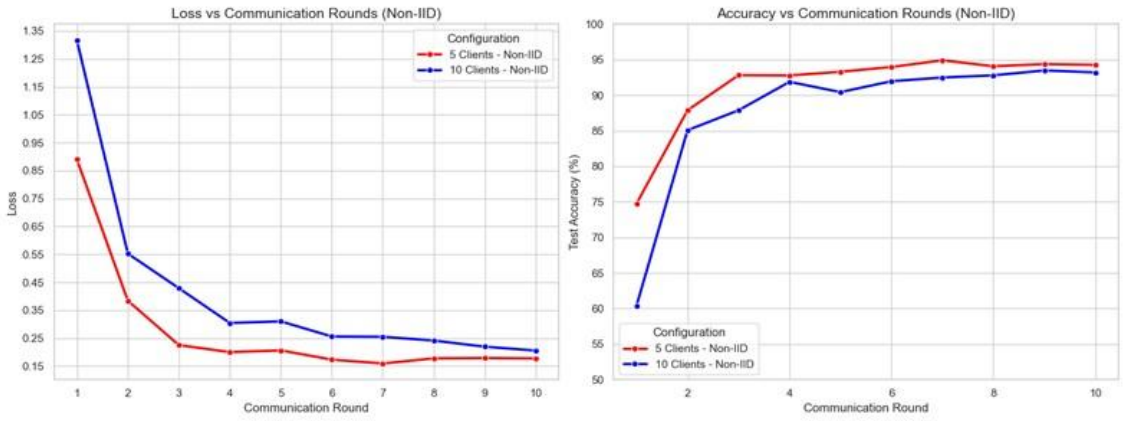

Fig 7: Federated Learning Accuracy and Loss Curves (NonIID).

Table IV: Federated Performance: IID vs Non-IID.

Data Type	Clients	Accuracy (%)	Loss
IID	**5**	**95.29**	**0.1651**
IID	**10**	**94.99**	**0.1595**
Non-IID	**5**	**94.26**	**0.1774**
Non-IID	**10**	**93.20**	**0.2055**

Journal of Information Systems Engineering and Management
2025, 10(63s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

These results indicate that federated learning closely matches centralized performance under IID conditions and remains competitive under non-IID, albeit with slightly reduced accuracy and increased loss due to statistical heterogeneity.

5. DISCUSSION

The experiments conducted in this study provide strong evidence that the proposed federated learning framework can maintain high predictive performance while preserving data privacy in decentralized medical imaging applications. Centralized benchmarking identified EfficientNet-B0 as the most suitable model architecture, achieving 95.35% accuracy on the PathMNIST dataset with balanced computational efficiency. This choice ensured that subsequent federated experiments would rely on a robust baseline for fair comparison. Federated learning experiments under both IID and nonIID data distributions demonstrated the framework's ability to closely match centralized performance. For IID distributions, the federated model reached 95.29% and 94.99% accuracy across 5 and 10 clients, respectively, with stable convergence observed throughout training rounds. These results highlight that, when client data are statistically homogeneous, federated averaging effectively aggregates model updates without significant loss in accuracy or increase in loss.

Under non-IID conditions, simulating realistic heterogeneity in client data, the model achieved 94.26% and 93.20% accuracy for 5 and 10 clients, respectively. Although slight reductions in performance and slower convergence were observed due to statistical skew, the framework remained robust, showing consistent improvement over training rounds. These findings indicate that the proposed framework is resilient to client-level variability and can generalize well even under heterogeneous data distributions. The overall results demonstrate that the framework combines high model utility (up to 94.26% accuracy) with computational efficiency, as evidenced by rapid convergence across different federation scales. The minor performance degradation observed in larger and heterogeneous federations is within acceptable bounds, confirming the practical applicability of the method in real-world, privacy-sensitive domains such as healthcare.In summary, these outcomes suggest that carefully designed federated learning systems, leveraging well-performing architectures like EfficientNet-B0 and accommodating client heterogeneity, can deliver near-centralized performance while keeping data decentralized. This positions the proposed framework as a viable solution for scalable, privacy-conscious medical AI applications.

6. CONCLUSION

This study presented a federated learning framework tailored for privacy-sensitive healthcare applications, capable of handling heterogeneous (non-IID) client data while maintaining high predictive performance. Through experiments on the PathMNIST dataset under both IID and non-IID distributions, across configurations of 5 and 10 clients, the framework achieved up to 95.29% accuracy, closely matching the performance of centralized training. The results demonstrate stable convergence and robust generalization, even in the presence of statistical heterogeneity among clients. The framework's design ensures computational efficiency, with rapid convergence observed across different federation scales, and only minor performance degradation in larger or more heterogeneous federations. These outcomes highlight the practical applicability of federated learning in decentralized medical imaging scenarios, where data privacy and security are critical concerns.

In summary, the proposed approach confirms that carefully orchestrated federated learning systems, leveraging well performing model architectures and accommodating client heterogeneity, can deliver near-centralized accuracy while keeping sensitive data decentralized. This positions the framework as a viable, scalable, and privacy-conscious solution for real world healthcare AI deployments.

REFRENCES

[1] Z. L. Teo, L. Jin, N. Liu, S. Li, D. Miao, X. Zhang, W. Y. Ng, T. F. Tan, D. M. Lee, K. J. Chua et al., "Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture," Cell Reports Medicine, vol. 5, no. 2, 2024.

[2] N. Yadav, S. Pandey, A. Gupta, P. Dudani, S. Gupta, and K. Rangarajan, "Data privacy in healthcare: in the era of artificial intelligence," Indian Dermatology Online Journal, vol. 14, no. 6, pp. 788–792, 2023.

[3] F. Zhang, D. Kreuter, Y. Chen, S. Dittmer, S. Tull, T. Shadbahr, M. Schut, F. Asselbergs, S. Kar, S. Sivapalaratnam et al., "Recent methodological advances in federated learning for healthcare," Patterns, vol. 5, no. 6, 2024.
[4] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni, "Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification," Scientific Data, vol. 10, no. 1, p. 41, 2023.
[5] T. JLoftus, M. Ruppert, B. Shickel, T. Ozrazgat-Baslanti et al. Federated learning for preserving data privacy in collaborative healthcare research. Digital Health, 2022.
[6] S. R. Abbas, Z. Abbas, A. Zahir, and S. W. Lee, "Federated learning in smart healthcare: a comprehensive review on privacy, security, and predictive analytics with iot integration," Healthcare, vol. 12, no. 24. MDPI, 2024.
[7] A. Thakur, S. Molaei, P. C. Nganjimi, F. Liu, A. Soltan, P. Schwab, K. Branson, and D. A. Clifton, "Knowledge abstraction and filtering based federated learning over heterogeneous data views in healthcare," npj Digital Medicine, vol. 7, no. 1, p. 283, 2024.
[8] S. Rajendran, Z. Xu, W. Pan, A. Ghosh, and F. Wang, "Data heterogeneity in federated learning with electronic health records: Case studies of risk prediction for acute kidney injury and sepsis diseases in critical care," PLOS Digital Health, vol. 2, no. 3, p. e0000117, 2023.
[9] P. Qi, D. Chiaro, A. Guzzo, M. Ianni, G. Fortino, and F. Piccialli, "Model aggregation techniques in federated learning: A comprehensive survey," Future Generation Computer Systems, vol. 150, pp. 272–293, 2024.
[10] Google, "Google colaboratory," https://colab.research.google.com/, 2024, accessed: 20-06-2024.
[11] D. Beutel, T. Topal, A. Mathur, X. Qiu, T. Parcollet, and N. D. Lane, "Flower: A friendly federated learning research framework," in Proceedings of the 2022 ACM Conference on Distributed Machine Learning (DistML '22). New York, NY, USA: Association for Computing Machinery, 2022.
[12] S. Chen, J. Liu, P. Wang, C. Xu, S. Cai, and J. Chu, "Accelerated optimization in deep learning with a proportional-integral-derivative controller," Nature Communications, vol. 15, no. 1, p. 10263, 2024.
[13] X. Yue, M. Nouiehed, and R. Al Kontar, "Salr: Sharpness-aware learning rate scheduler for improved generalization," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, 2023.
[14] E. Jing, H. Zhang, Z. Li, Y. Liu, Z. Ji, I. Ganchev. ECG heartbeat classification based on an improved ResNet-18 model. Computational and Mathematical Methods in Medicine, 2021.
[15] C. Guo, Q. Zhou, J. Jiao, Q. Li, and L. Zhu, "A modified mobilenetv3 model using an attention mechanism for eight-class classification of breast cancer pathological images," Applied Sciences, vol. 14, no. 17, p. 7564, 2024.
[16] L. Arora, S. K. Singh, S. Kumar, H. Gupta, W. Alhalabi, V. Arya, S. Bansal, K. T. Chui, and B. B. Gupta, "Ensemble deep learning and efficientnet for accurate diagnosis of diabetic retinopathy," Scientific Reports, vol. 14, no. 1, p. 30554, 2024.
[17] G. U. Nneji, H. N. Monday, G. T. Mgbejime, V. S. R. Pathapati, S. Nahar, and C. C. Ukwuoma, "Lightweight separable convolution network for breast cancer histopathological identification," Diagnostics, vol. 13, no. 2, p. 299, 2023.
[18] Z. Rasheed, Y.-K. Ma, I. Ullah, M. Al-Khasawneh, S. S. Almutairi, and M. Abohashrh, "Integrating convolutional neural networks with attention mechanisms for magnetic resonance imaging-based classification of brain tumors," Bioengineering, vol. 11, no. 7, p. 701, 2024.