

Enhancing Federated Learning for Imbalanced Medical Image Classification through Adaptive Tuning and Autoencoder-Based Reconstruction

Nadzurah Zainal Abidin¹, Amelia Ritahani Ismail², Cut Amalia Saffiera³, Nurul A. Emran⁴,
Zammarah Nuha Abdullah⁵

^{1,2} Department of Computer Science, International Islamic University Malaysia, Malaysia

³ Fakultas Komputer dan Multimedia, Universitas Islam Kebangsaan Indonesia, Indonesia

⁴ Fakulti Teknologi Maklumat dan Komunikasi, Universiti Teknikal Malaysia Melaka, Malaysia

⁵ Department of Computer and Information Science, Applied College, Taibahu University, Saudi Arabia

ARTICLE INFORMATION

Article History:

Received 10 November 2025

Revised 22 February 2026

Accepted 29 June 2026

Keywords:

Federated Learning;
Class Imbalance;
Medical Image Classification;
Adaptive Tuning;
Autoencoder

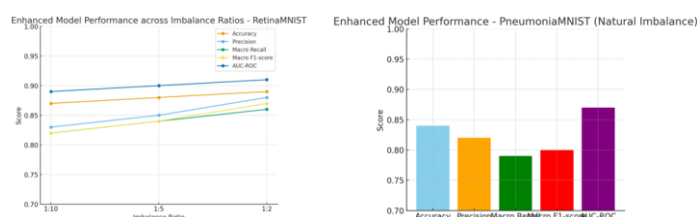
Corresponding Author:

Nadzurah Zainal Abidin,
Department of Computer Science,
International Islamic University
Malaysia, Malaysia.
nadzurah.zabidin@gmail.com

This work is open access under a
[Creative Commons Attribution-Share
Alike 4.0](https://creativecommons.org/licenses/by-sa/4.0/)



ABSTRACT



Medical image classification has advanced significantly through deep learning techniques, yet its performance remains limited by class imbalance and decentralized data silos commonly found in healthcare settings. These issues reduce model sensitivity to rare but clinically important cases, and standard Federated Learning (FL) further struggles under non-independent and identically distributed (non-IID) data. To address this, an enhanced federated model integrating unsupervised autoencoder-based reconstruction and adaptive tuning is proposed. The research contribution is an enhanced FL model that improves minority-class detection and overall classification performance under imbalanced medical image distributions, while remaining applicable across decentralized healthcare data sources. The method incorporates an autoencoder to compute reconstruction error, enabling emphasis on underrepresented samples, while adaptive tuning dynamically adjusts local hyperparameters and global aggregation weights based on sample difficulty. This integration strengthens minority-class learning without requiring additional labels or altering the decentralized structure. Experimental evaluations were conducted using two benchmark medical image datasets across three induced imbalance ratios (1:10, 1:5, 1:2) for RetinaMNIST and naturally induced imbalance for PneumoniaMNIST dataset. Results show that under severe imbalance (1:10), the enhanced model improves minority-class recall by 59.6%, F1-score by 33.9%, and AUC-ROC by 13.3% compared to standard FL. At 1:5 imbalance, recall increases by 41.3% and F1-score by 26.5%, with accuracy gains up to 6.0%. Even under mild imbalance (1:2), the model maintains consistent improvements, achieving a 26.4% recall gain and 18.9% increase in F1-score. The performance of the enhanced model was further evaluated against three baseline FL models such as standard federated learning (FedAvg), FL with GAN augmentation, and FL with standalone autoencoder-based reconstruction. The results consistently confirm that federated learning integrating adaptive tuning and autoencoder-based reconstruction outperform the three baselines FL-models for accuracy, recall, and F1-score. These findings also demonstrate that the enhanced model provides scalable, coordination-free improvement in imbalanced federated medical image classification, offering stronger performance and stability across real-world heterogeneous settings.

Document Citation:

N. Z. Abidin, A. R. Ismail, C. A. Saffiera, N. A. Emran, and Z. N. Abdullah, "Adaptive Federated Learning Model with Unsupervised Autoencoder-based Reconstruction for Imbalanced Medical Image Classification," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 8, no. 3, pp. 901-912, 2026, DOI: 10.12928/biste.v8i3.15266.

1. INTRODUCTION

Though AI-based medical imaging systems have proven successful, their performance often hinges on the availability of large and balanced datasets. In real-world clinical scenarios, however, medical image datasets are frequently imbalanced, where minority classes, typically representing rare but critical diseases, are significantly underrepresented [1][2]. For instance, severe diabetic retinopathy cases constitute a small proportion of retinal screening datasets, malignant melanoma cases are far fewer than benign lesions in dermatological datasets, and rare pulmonary conditions such as pneumothorax appear infrequently in chest X-ray collections. This imbalance biases the model predictions, where the classifier tends to favor the majority class, resulting in low sensitivity towards the clinically relevant class of interest, which is rare cases. Standard deep learning approaches, which assume uniform class distribution, therefore struggle to generalize well in such settings [3][4].

Federated learning (FL) has emerged as a decentralized paradigm that enables collaborative model training across multiple data sources or institutions without centralizing data or sharing their local datasets [5]-[10]. This approach leverages the diversity of distributed medical data while allowing each data source, such as hospitals, clinics, and medical institutions, to retain control of its own dataset and leads to improved model generalization across different populations. However, despite these advantages, FL faces two key significant challenges in medical image analysis: (1) non-IID (non-independent and identically distributed) data distributions across clients, where different data sources hold various class ratios or distinct feature distributions, and (2) class imbalance within local datasets, where uneven sample proportions across classes within each dataset [11][12], which both can negatively impact classification performance, especially on rare or underrepresented medical conditions [13]. This work primarily addresses the latter challenge by enhancing federated learning to improve minority-class recognition under imbalanced data conditions.

Standard FL struggles under these conditions due to its inability to adapt to data skewness and distributional differences, leading to poor convergence and reduced sensitivity on underrepresented classes [14][15]. Consequently, researchers have explored several enhancements to enhance FL performance under these challenges, broadly categorizing into contribution-based aggregation schemes and optimization-based methods. Contribution-based approaches adjust aggregation weights based on client reliability and performance [16]-[19], while optimization-based methods modify the training objective to improve convergence under heterogeneous data [19][20]. Another common approach is Generative Adversarial Networks (GANs)-based augmentation, which attempts to balance training samples by synthesizing realistic samples of minority classes under federated settings [21]-[23]. Similarly, autoencoder-based approaches, particularly autoencoders, have also been explored to identify rare or abnormal patterns in medical images, improving recognition of minority class instances [24][25]. Although GAN-based augmentation has shown effectiveness in centralized settings, its adversarial training process introduces instability and additional computational overhead in federated environments, particularly when minority samples are scarce across clients. In contrast, autoencoder-based reconstruction provides a stable, unsupervised mechanism for identifying rare samples through reconstruction error without generating synthetic data, making it more suitable for decentralized imbalance handling.

While these approaches show promise, they are typically applied separately and therefore do not fully address the combined challenges of class imbalance and decentralized learning [26]. Moreover, many existing approaches overlook the importance of dynamic training adjustments, training settings that adjust automatically during training rather than remain fixed, which have been shown to improve convergence and stability in non-IID settings. However, their integration with imbalance-handling techniques remains underexplored.

This research aims to address the challenge of class imbalance in federated medical image classification by enhancing the standard federated learning model that integrates two complementary approaches: (1) autoencoder-based reconstruction to improve recognition of minority class samples, and (2) adaptive tuning to adjust training behavior based on reconstruction error signals.

The research contribution of this work is the development of an integrated federated learning model that systematically evaluated against standard FedAvg, FL with GAN-based augmentation, and FL with standalone autoencoder-based reconstruction on RetinaMNIST and PneumoniaMNIST datasets. The experimental results demonstrate consistent improvements in minority-class detection across multiple imbalance ratios, confirming the effectiveness of the enhanced design without altering the core federated aggregation structure.

2. RESEARCH PROBLEM

Federated learning (FL) has emerged as a promising decentralized paradigm for training models collaboratively across multiple medical data sources while retaining local data. By combining knowledge or patterns from multiple medical data sources, FL has the potential to improve generalization across diverse

populations. However, FL presents unique challenges when applied to imbalanced medical image classification [27][28]. The decentralized and non-IID (non-independent and identically distributed) nature of FL data often amplifies class imbalance, as many medical data sources naturally contain insufficient samples from minority classes. As a result, the aggregated global model tends to underperform on rare conditions [29].

Existing solutions remain insufficient for addressing class imbalance in federated medical image classification. Data-level approaches, such as GAN-based augmentation, can generate synthetic samples but are often struggled to maintain stability and consistent quality in generating images in federated environments [14],[30]-[33]. Although several FL enhancement techniques have been proposed, many of these require substantial modifications to local training procedures [34][35]. This makes them impractical in real-world deployments, especially in smaller or resource-constrained medical data sources. Therefore, a method that can adaptively improve learning under imbalance and non-IID conditions without requiring major changes to local model at each data source is still needed.

This gap underscores the urgent need for an approach that improve FL's ability to recognize minority classes under imbalanced conditions, while preserving its advantages of decentralization and maintaining full control over local data.

3. METHODS

The research is designed as an experimental and comparative study that evaluates federated learning (FL) under class imbalance using two medical imaging datasets. Three baseline FL configurations are implemented for comparison against our enhanced federated learning model: standard Federated Averaging (FedAvg), FL with GAN-based augmentation, and FL with standalone autoencoder-based reconstruction.

3.1. Dataset and Preprocessing

This research employs two benchmarking datasets under a subset of the MedMNIST v2 collection, which are RetinaMNIST and PneumoniaMNIST datasets, representing induced (severe imbalance -1:10, moderate imbalance -1:5, mild imbalance -1:2) and natural imbalance settings respectively. These datasets are lightweight, standardized, and widely used for medical image classification tasks under constrained conditions.

3.1.1. RetinaMNIST

The dataset consists of 1600 retina fundus images, that are labeled for five levels of diabetic retinopathy severity (ranging from normal to severe), making the task an ordinal regression problem. Imbalance ratios of 1:10, 1:5, and 1:2, were artificially induced by downsampling minority-class samples to reflect realistic clinical scenarios [36][37]. These imbalance ratios represent severe, moderate, and mild imbalance conditions, respectively, enabling systematic evaluation of model robustness under different levels of class distribution.

The original fundus images have a resolution of 1736 x 1824 pixels with three color channels. For computational efficiency and consistency in training, all images are center-cropped and resized to 3 x 28 x 28. This preprocessing step ensures uniform input dimensions across all models while preserving essential diagnostic features.

3.1.2. PneumoniaMNIST

The dataset derived from chest X-ray images for the task of binary classification, distinguishing between normal cases and pneumonia-positive cases. It consists of 5856 grayscale images, reflecting a more natural imbalance in real-world clinical settings, where pneumonia cases are often less frequent compared to normal scans [36],[38]. Each image in PneumoniaMNIST dataset is standardized to a resolution of 1x28x28 (grayscale), aligning with the MedMNIST format. This ensures compatibility with lightweight neural networks and efficient training in federated settings. Table 1 summarizes the datasets used.

Table 1. Summary of Dataset

Dataset	Task Type	Imbalance Type	Training	Validation	Testing	Image Size
Retina	Multi-class (5 classes)	Induce	1080	120	400	3 x 28 x 28 (RGB)
Pneumonia	Binary classification	Natural	4708	524	624	1 x 28 x 28 (Grayscale)

To simulate federated learning, the datasets were distributed across 10 healthcare institutions, in a non-IID (non-independent and identically distributed) manner. Each data source receives training data with biased class distribution towards 2 or 3 dominant classes, simulating realistic scenarios in federated medical settings.

By enforcing this decentralized and non-IID partitioning, the research establishes a challenging yet practical testbed for assessing the robustness and generalization capability of federated learning under both data imbalance and client heterogeneity.

All experiments were conducted using Google Colab with GPU acceleration using PyTorch and Flower federated learning framework (FL library). Experiments were conducted by distributing both datasets to 10 different data sources over 50 global communication rounds, each performing 5 epochs with a batch size of 32 per round. These experiments used Adam optimizer, with a learning rate initially set to 0.01 and adaptively tuned in the enhanced model. Each experiment is repeated across the three imbalance settings to evaluate performance consistency and robustness.

3.2. Baseline Models

Three baseline FL configurations were implemented for comparison:

- Standard FL (FedAvg): Local data sources train independently and send model parameters to a central server for weighted averaging, with no imbalance mitigation.
- FL + GAN Augmentation: Minority-class samples are synthetically augmented at each client before local training
- FL + Autoencoder-based Reconstruction: Each client uses a pre-trained autoencoder to compute reconstruction error and emphasize samples that are harder to represent.

All baseline models were trained under identical settings across the three imbalance ratios to ensure fair comparison.

3.3. Enhanced Algorithm: Federated learning with Autoencoder-based Reconstruction and Adaptive Tuning

The enhanced model aims to address the limitations of standard federated learning under class imbalance by integrating two enhancement strategies: adaptive tuning and unsupervised autoencoder-based reconstruction.

3.3.1. Autoencoder-based Reconstruction

Autoencoder-based reconstruction has been widely applied for rare-event recognition [39][40]. In highly imbalanced datasets, the minority class can be treated as a form of anomalies since they are both underrepresented and statistically deviant from the majority [41][42]. Autoencoders learn compressed representations of dominant patterns and highlight deviations through reconstruction error, thereby enabling better detection of minority instances without requiring balanced training data [43][44].

In federated learning, the appeal of autoencoder-based reconstruction is amplified [45][46]. Since autoencoders can be trained locally on each institutional data, they do not rely on global label statistics or inter-client coordination [12],[47]. Each institution can independently emphasize rare patterns during local training, and reconstruction signals can be incorporated into the global aggregation process [48]-[51].

An unsupervised autoencoder is deployed at each data source to learn compact representations of local medical images. The autoencoder operates independently at each data source without sharing parameters across data sources [52][53]. For each input sample, a reconstruction error is computed locally and used as a signal to identify rare or difficult instances during training [12],[54]. This reconstruction-based signal is incorporated into the local optimization process to emphasize minority or underrepresented classes within each data source's dataset, while preserving the decentralized structure of federated learning. The reconstruction error is defined as the difference between the input image and its reconstructed output, computed independently at each data source during training [52][53].

3.3.2. Adaptive Tuning

Adaptive tuning in federated learning refers to strategies that dynamically adjust training process to handle client heterogeneity, non-IID data, and unstable convergence [55]-[57]. In standard federated learning, a fixed learning rate is used throughout all communication rounds and across all participating data sources [44]-[46]. However, this static approach often leads to suboptimal convergence, particularly when data distributions are non-identical (non-IID) and imbalanced across data sources [16],[58][59].

To mitigate this, the enhanced model introduces an adaptive learning rate tuning strategy, which dynamically adjusts each data source learning rate based on local validation loss feedback [16],[58][59]. The principle is straightforward:

- If validation loss decreases, the learning rate is retained or slightly increased to speed convergence.

- If validation loss stagnates or increases, the learning rate is reduced to prevent overshooting and improve stability.
By combining these two enhancements, the enhanced federated learning model aims to improve the robustness under real-world non-IID and imbalanced data conditions, ensuring the local updates are both reliable for global aggregation.

Algorithm: Federated Learning with Autoencoder-based Reconstruction and Adaptive Tuning
Input:

Set of client devices C
 Global model initialization w_0
 Number of communication rounds T
 Local mini-batch size B
 Number of local epochs E
 Initial learning rate a_0
 Pre-trained autoencoder AE
 Adaptive learning $Adapt()$

Initialize the global model: $w_t = w_0$

for $t = 1$ to T do:

 Select subset of client devices S_t from C

 for each client k in S_t do in parallel:

 Receive the current global model w_t from the server.

 Apply autoencoder-based reconstruction to the local dataset using the autoencoder AE

 Divide the reconstruction local data into mini batches of size B

 Compute adaptive learning rate $a_k = Adapt(a_0, validation_loss)$

 for each local epoch $e = 1$ to E :

 for each mini batch b do:

 Update the local model:

$$w_t^k = w_t^k - a_k \cdot \nabla \mathcal{L}(w_t^k, b)$$

 Send the updated global model w_t^k back to the server

 Aggregate the local models:

$$w_{t+1} = \left(\frac{1}{|S_t|} \right) \cdot \sum w_t^k,$$

 where $|S_t|$ is the number of clients in S_t

 and $\sum S_t$ represents the sum over all clients in S_t

 Send the updated global model w_{t+1} to all clients S_t

Output: Final global model w_t

By incorporating both an adaptive tuning mechanism and an unsupervised autoencoder-based reconstruction module as illustrated in [Figure 1](#), the model aims to achieve better classification performance, particularly on underrepresented classes, without compromising the federated learning structure.

3.4. Performance Metrics

All model performance is evaluated using five performance metrics: accuracy, precision, macro recall, macro F1-score, and AUC-ROC. While accuracy measures overall correctness, macro-averaged metrics assign equal importance to all classes and are therefore more appropriate under imbalanced conditions. In medical diagnosis, recall is particularly critical as it reflects the ability to correctly detect rare but clinically significant cases. AUC-ROC further evaluates the model's discrimination capability across decision thresholds.

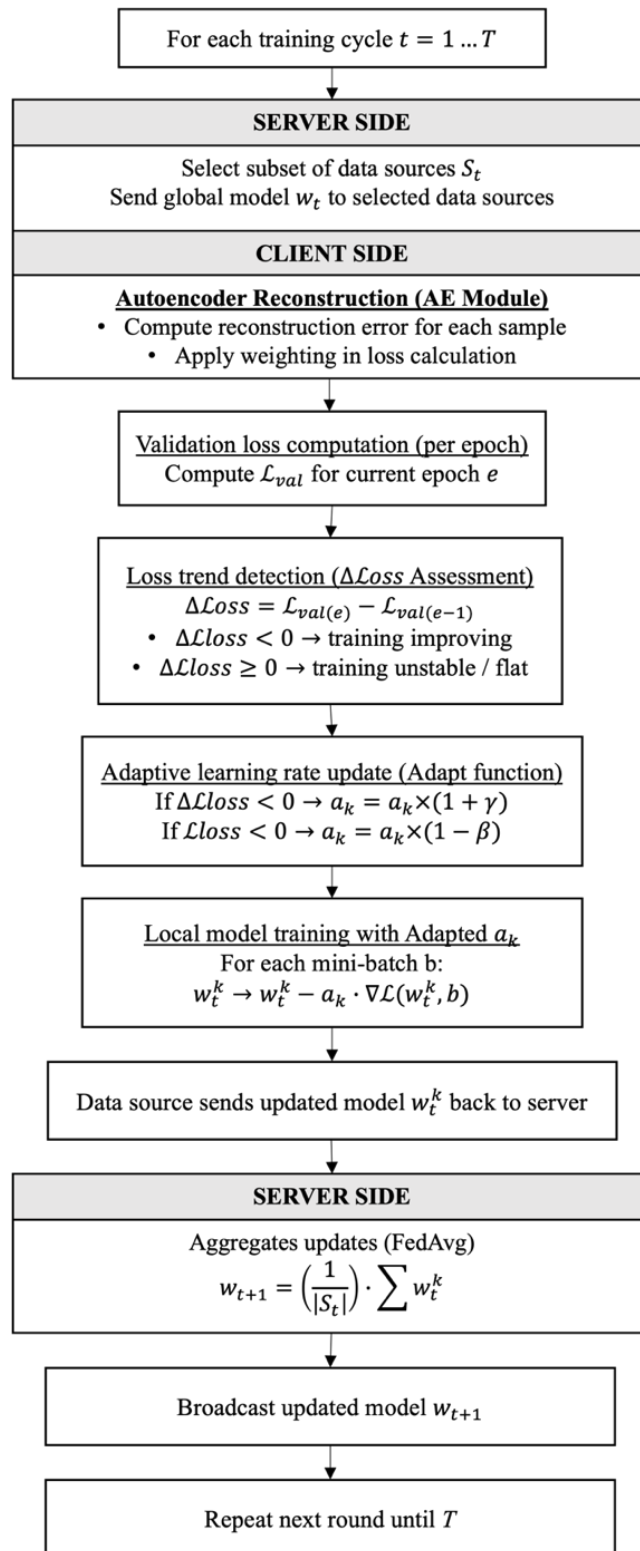


Figure 1. Workflow of the Enhanced Federated Learning

4. RESULT AND DISCUSSION

This subsection presents a general performance summary of the baseline models, (1) standard Federated Learning (FL), (2) FL with Generative Adversarial Networks (FL + GAN), and (3) FL with Unsupervised autoencoder-based reconstruction (FL + AE). The results are evaluated on two datasets: RetinaMNIST (Table

2) (with induced imbalance) and PneumoniaMNIST (Table 3) (natural imbalance) using common performance metrics including accuracy, precision, recall, F1-score and AUC-ROC.

Under the low imbalance condition (1:2), the enhanced model demonstrate strong performance across all evaluation metrics, including accuracy (0.90), macro recall (0.91), and macro F1-score (0.88). Notably, it maintains a high recall for the minority class, indicating that the model is still able to correctly identify a large portion of the underrepresented cases. This indicates that the effectiveness of autoencoder-based reconstruction and adaptive tuning remains robust even when the data distribution becomes moderately skewed. The model does not suffer a substantial drop in its ability to detect rare or less frequent class instances, which is critical in medical diagnosis scenarios where missing such cases could have serious implications.

As the imbalance increases to 1:5, the performance of the model remains stable, with only a slight drop in macro recall and F1-score. This suggests that the adaptive tuning mechanism helps maintain balance learning dynamics even when the minority class becomes less frequent. The autoencoder-based reconstruction further ensures that noisy or low-quality samples do not dominate the learning process, preserving the model's ability to generalize effectively.

Under the most severe imbalance (1:10), while a decline is observed in some metrics, particularly in recall, the enhanced model still outperforms all baseline models (as discussed in section 5.2). Macro recall remains around 0.80, indicating the model's continued attention to minority class detection, which is critical for medical diagnostic tasks, where missing rare conditions can be costly. The AUC-ROC also remains relatively high (0.85), indicating strong overall ability to distinguish between classes.

Overall, as illustrated in Figure 2, the most substantial improvements were observed under severe imbalance (1:10), where the enhanced federated learning model increased macro recall by 59.6 percent, F1-score by 33.9 percent, and AUC-ROC by 13.3 percent compared to standard FL. These results confirm that the combination of autoencoder-based reconstruction and adaptive tuning control enables the enhanced model to effectively manage data heterogeneity and class imbalance. The model achieves better robustness and fairness, and sensitivity to minority classes than baseline models.

Even though PneumoniaMNIST does not involve induced imbalance ratios, the enhanced model still achieves the strongest results across all metrics, including accuracy, recall, and AUC-ROC. The gains are particularly clear in recall and AUC-ROC as shown in Figure 3, which are crucial for identifying pneumonia-positive cases without excessive false alarms. Baselines show reasonable accuracy but lower sensitivity to positive cases, highlighting the enhanced model's advantage in real-world, naturally skewed datasets. Especially results on PneumoniaMNIST with natural imbalance, the enhanced model improved recall by 23.6 percent, F1-score by 18.7 percent, and accuracy by 8.3 percent.

These improvements are not just small numerical gains. They represent stronger generalization, better rare-case detection, and higher overall reliability across diverse data distributions. For practical deployment in medical settings, where errors in handling non-representative cases can lead to harmful consequences, the enhanced model provides consistent effectiveness for both artificially induced imbalance (RetinaMNIST) and natural imbalance (PneumoniaMNIST).

Consequently, performance metrics alone cannot determine whether observed improvements occur consistently across runs or simply by chance, particularly in federated learning, where variability arises from non-IID data distributions and data source-modification training dynamics. Therefore, statistical significance testing, Wilcoxon-signed rank test is performed to validate the reliability of the enhancement approaches.

Across all imbalance ratios in RetinaMNIST and on the naturally imbalanced PneumoniaMNIST dataset:

- Enhanced model against FL $\rightarrow p < 0.05$
- Enhanced model against FL + GAN $\rightarrow p < 0.05$
- Enhanced model against FL + AE $\rightarrow p < 0.05$

This demonstrates that the enhanced model improvements in macro recall, macro F1-score, and AUC-ROC are statistically significant, and not the result of random variation.

The strongest significance was observed under the 1:10 imbalance ratio, where baseline models show high variability, but the enhanced model produces consistently high minority-class performance.

In conclusion, compared to prior imbalance-handling strategies in federated learning that rely on synthetic data augmentation or static optimization adjustments, the enhanced integration of reconstruction-based emphasis and adaptive tuning demonstrates stronger minority-class recall under severe imbalance. This can be mainly because augmentation-based model only artificially increases minority classes, and may introduce noise or unrealistic medical features, while the enhanced model recalibrates the learning process to ensure minority classes are not ignored during decentralized training.

Table 2. Performance Results on RetinaMNIST

Model	Imbalance Ratio	Accuracy	Precision	Macro Recall	F1-Score	AUC-ROC
FL	1:10	0.81	0.70	0.52	0.59	0.75
	1:5	0.83	0.76	0.63	0.68	0.80
	1:2	0.85	0.80	0.72	0.74	0.83
FL + GAN	1:10	0.83	0.75	0.56	0.61	0.79
	1:5	0.85	0.80	0.68	0.71	0.84
	1:2	0.87	0.85	0.78	0.80	0.88
FL + AE	1:10	0.84	0.74	0.60	0.66	0.81
	1:5	0.86	0.79	0.72	0.75	0.86
	1:2	0.88	0.83	0.80	0.81	0.89
Enhanced Model	1:10	0.81	0.77	0.83	0.79	0.85
	1:5	0.88	0.84	0.89	0.86	0.91
	1:2	0.90	0.86	0.91	0.88	0.93

Table 3. Performance Results on PneumoniaMNIST

Model	Accuracy	Precision	Macro Recall	F1-Score	AUC-ROC
FL	0.84	0.79	0.72	0.75	0.82
FL + GAN	0.86	0.82	0.76	0.78	0.85
FL + AE	0.87	0.84	0.78	0.80	0.87
Enhanced Model	0.91	0.90	0.89	0.89	0.93

Enhanced Model Performance across Imbalance Ratios - RetinaMNIST

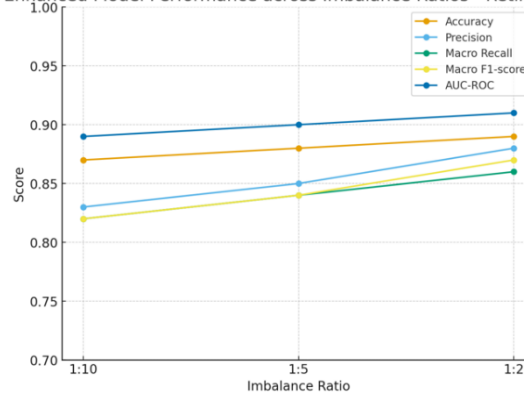


Figure 2. Enhanced Model Performance for RetinaMNIST

Enhanced Model Performance - PneumoniaMNIST (Natural Imbalance)

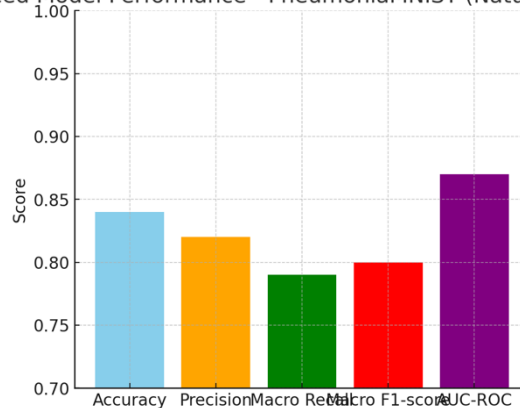


Figure 3. Enhanced Model Performance for PneumoniaMNIST

5. CONCLUSIONS

This research proposed an enhanced federated learning model integrating adaptive tuning and autoencoder-based reconstruction to address class imbalance in medical image classification. The performance of the enhanced model delivers clear advantages in classification performance and robustness under class imbalance and generalization compared to baseline FL based methods on both RetinaMNIST and PneumoniaMNIST. In RetinaMNIST, the enhancements mitigate the impact of increasing imbalance ratios. In

PneumoniaMNIST, the model handles naturally imbalanced binary data more effectively than the baselines. This dual validation across datasets strengthens the case for the enhanced model as a general solution for imbalanced medical image classification in federated environments. The enhanced model also demonstrates superior minority-class recognition compared to baseline FL-based imbalance mitigation strategies. Under severe imbalance (1:10), the enhanced federated learning model improved macro recall by 59.6 percent, F1-score by 33.9 percent, and AUC-ROC by 13.3 percent compared to standard federated learning, which also similarly observed under moderate and mild imbalance settings, as well as in naturally imbalanced binary classification tasks.

Nevertheless, the research was conducted using lightweight MedMNIST datasets within a simulated federated setting, which may not fully reflect large-scale, real-world multi-institutional deployments. Future work will extend validation to higher-resolution and clinically diverse datasets to further enhance robustness and scalability.

Overall, the enhanced federated learning demonstrates superior performance across varying imbalance medical representations against standard federated learning model, improve diagnostic sensitivity for rare medical conditions, and contributing to more reliable and equitable AI-driven healthcare systems.

DECLARATION

Author Contribution

All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding

This research received no external funding.

Acknowledgement

Write a thank you to those who have helped in this research who are not listed in the author's list.

Conflicts of Interest

The authors declare no conflict of interest.

ABBREVIATIONS

The following abbreviations are used in this manuscript.

FL	:	Federated Learning
FedAvg	:	Federated Averaging
Non-IID	:	Non-Independent and Identically Distributed
AT	:	Adaptive Tuning
AE	:	Auto-encoder
GAN	:	Generative Adversarial Network
AUC-ROC	:	Area under the Receiver Operating Characteristic Curve

REFERENCES

- [1] S. Kothawade, A. Savarkar, V. Iyer, L. Tamil, G. Ramakrishnan, and R. Iyer, "CLINICAL: Targeted Active Learning for Imbalanced Medical Image Classification," *In Workshop on Medical Image Learning with Limited and Noisy Data*, pp. 119-129, 2022, https://doi.org/10.1007/978-3-031-16760-7_12.
- [2] J. X. Zhuang, J. Cai, J. Zhang, W. S. Zheng, and R. Wang, "Class attention to regions of lesion for imbalanced medical image recognition," *Neurocomputing*, vol. 555, p. 126577, 2023, <https://doi.org/10.1016/j.neucom.2023.126577>.
- [3] A. A. Dharmasaputro, N. M. Fauzan, M. Kallista, Ig. P. D. Wibawa, and P. D. Kusuma, "Handling Missing and Imbalanced Data to Improve Generalization Performance of Machine Learning Classifier," in *2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE)*, pp. 140–145, 2022, <https://doi.org/10.1109/ISMODE53584.2022.9743022>.
- [4] B. Qi, W. Gong, and L. Li, "Probability Distribution Learning and Its Application in Deep Learning," *arXiv preprint arXiv:2406.05666*, 2024, <https://doi.org/10.48550/arXiv.2406.05666>.
- [5] M. Li, P. Xu, J. Hu, Z. Tang, and G. Yang, "From challenges and pitfalls to recommendations and opportunities: Implementing federated learning in healthcare," *Medical image analysis*, vol. 101, p. 103497, 2025, <https://doi.org/10.1016/j.media.2025.103497>.
- [6] S. Li *et al.*, "Federated Learning in Healthcare: A Benchmark Comparison of Engineering and Statistical Approaches for Structured Data Analysis," *Health Data Science*, vol. 4, p. 0196, 2024, <https://doi.org/10.34133/hds.0196>.

- [7] 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," In *Healthcare*, vol. 12, no. 24, p. 2587, 2024, <https://doi.org/10.3390/healthcare12242587>.
- [8] H. Chung and J. S. Lee, "Federated influencer learning for secure and efficient collaborative learning in realistic medical database environment," *Sci. Rep.*, vol. 14, no. 1, p. 22729, 2024, <https://doi.org/10.1038/s41598-024-73863-1>.
- [9] H. Byeon, A. Ullah, Z. Syed, A. Siddiqui, A. Aftab, and M. Sarfaraz, "Image Classification Using Federated Learning," In *International Conference on Recent Trends in Image Processing and Pattern Recognition*, pp. 66-75, 2023, https://doi.org/10.1007/978-3-031-53082-1_6.
- [10] S. Ren, Y. Hu, S. Chen, and G. Wang, "Federated Distillation for Medical Image Classification: Towards Trustworthy Computer-Aided Diagnosis," *arXiv preprint arXiv:2407.02261*, 2024, <https://doi.org/10.48550/arXiv.2407.02261>.
- [11] Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra, "Federated Learning with Non-IID Data," *arXiv preprint arXiv:1806.00582*, 2018, <https://doi.org/10.48550/arXiv.1806.00582>.
- [12] S. Laridi, G. Palmer, and K.-M. M. Tam, "Enhanced Federated Anomaly Detection Through Autoencoders Using Summary Statistics-Based Thresholding," *Scientific Reports*, vol. 14, no. 1, p. 26704, 2024, <https://doi.org/10.1038/s41598-024-76961-2>.
- [13] M. Babar, B. Qureshi, and A. Koubaa, "Investigating the impact of data heterogeneity on the performance of federated learning algorithm using medical imaging," *PLoS One*, vol. 19, no. 5, 2024, <https://doi.org/10.1371/journal.pone.0302539>.
- [14] B. Van Calster *et al.*, "Performance evaluation of predictive AI models to support medical decisions: Overview and guidance," *arXiv preprint arXiv:2412.10288*, 2024, <https://doi.org/10.48550/arXiv.2412.10288>.
- [15] M. Haller, C. Lenz, R. Nachtigall, F. M. Awayshehl, and S. Alawadi, "Handling Non-IID Data in Federated Learning: An Experimental Evaluation Towards Unified Metrics," in *2023 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCCom/CyberSciTech)*, pp. 0762-0770, 2023, <https://doi.org/10.1109/DASC/PiCom/CBDCCom/Cy59711.2023.10361408>.
- [16] F. Saile, J. Thomas, D. Kaaser, and S. Schulte, "Client-Side Adaptation to Concept Drift in Federated Learning," in *2024 2nd International Conference on Federated Learning Technologies and Applications (FLTA)*, pp. 71-78, 2024, <https://doi.org/10.1109/FLTA63145.2024.10840058>.
- [17] G. Long, M. Xie, T. Shen, T. Zhou, X. Wang, and J. Jiang, "Multi-center federated learning: clients clustering for better personalization," *World Wide Web*, vol. 26, no. 1, pp. 481-500, 2023, <https://doi.org/10.1007/s11280-022-01046-x>.
- [18] Y. -A. Chen and G. -L. Chen, "An Adaptive Clustering Scheme for Client Selections in Communication-Efficient Federated Learning," *2023 VTS Asia Pacific Wireless Communications Symposium (APWCS)*, pp. 1-3, 2023, <https://doi.org/10.1109/APWCS60142.2023.10234043>.
- [19] A. El-Niss, A. Alzu'Bi, A. Abuarqoub, M. Hammoudeh and A. Muthanna, "SimProx: A Similarity-Based Aggregation in Federated Learning With Client Weight Optimization," in *IEEE Open Journal of the Communications Society*, vol. 5, pp. 7806-7817, 2024, <https://doi.org/10.1109/OJCOMS.2024.3513816>.
- [20] X. Li, Z. Qu, B. Tang and Z. Lu, "FedLGA: Toward System-Heterogeneity of Federated Learning via Local Gradient Approximation," in *IEEE Transactions on Cybernetics*, vol. 54, no. 1, pp. 401-414, 2024, <https://doi.org/10.1109/TCYB.2023.3247365>.
- [21] C. Fan and P. Liu, "Federated Generative Adversarial Learning," In *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, pp. 3-15, 2020, https://doi.org/10.1007/978-3-030-60636-7_1.
- [22] J. Zhang, L. Zhao, K. Yu, G. Min, A. Y. Al-Dubai and A. Y. Zomaya, "A Novel Federated Learning Scheme for Generative Adversarial Networks," in *IEEE Transactions on Mobile Computing*, vol. 23, no. 5, pp. 3633-3649, 2024, <https://doi.org/10.1109/TMC.2023.3278668>.
- [23] A. Triastcyn and B. Faltings, "Federated Generative Privacy," in *IEEE Intelligent Systems*, vol. 35, no. 4, pp. 50-57, 2020, <https://doi.org/10.1109/MIS.2020.2993966>.
- [24] S. Kumari, C. Prabha, A. Karim, M. M. Hassan, and S. Azam, "A Comprehensive Investigation of Anomaly Detection Methods in Deep Learning and Machine Learning: 2019-2023," *IET Information Security*, vol. 2024, no. 1, p. 8821891, 2024, <https://doi.org/10.1049/2024/8821891>.
- [25] S. Schneider, D. Antensteiner, D. Soukup and M. Scheutz, "Autoencoders - A Comparative Analysis in the Realm of Anomaly Detection," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1985-1991, 2022, <https://doi.org/10.1109/CVPRW56347.2022.00216>.
- [26] C. Zhang, G. Shan, and B. Roh, "Communication-efficient federated multi-domain learning for network anomaly detection," *Digital Communications and Networks*, 2024, <https://doi.org/10.1016/j.dcan.2024.11.014>.
- [27] K. M. M. Dolaat, A. Erbad, and M. Ibrar, "Enhancing Global Model Accuracy: Federated Learning for Imbalanced Medical Image Datasets," in *2023 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1-4, 2023, <https://doi.org/10.1109/ISNCC58260.2023.10323682>.
- [28] H. Guan and M. Liu, "Federated Learning for Medical Image Analysis: A Survey," *Pattern recognition*, vol. 151, p. 110424, 2024, <https://doi.org/10.1016/j.patcog.2024.110424>.

- [29] L. I. Barona López and T. Borja Saltos, "Heterogeneity Challenges of Federated Learning for Future Wireless Communication Networks," *Journal of Sensor and Actuator Networks*, vol. 14, no. 2, p. 37, 2025, <https://doi.org/10.3390/jsan14020037>.
- [30] E. Diao, J. Ding, and V. Tarokh, "Heterofl: Computation and communication efficient federated learning for heterogeneous clients," *arXiv preprint arXiv:2010.01264*, 2020, <https://doi.org/10.48550/arXiv.2010.01264>.
- [31] L. U. Khan *et al.*, "Federated Learning for Edge Networks: Resource Optimization and Incentive Mechanism," *IEEE Communications Magazine*, vol. 58, no. 10, pp. 88–93, 2020, <https://doi.org/10.1109/MCOM.001.1900649>.
- [32] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated Learning: Strategies for Improving Communication Efficiency," *arXiv preprint arXiv:1610.05492*, 2016, <https://doi.org/10.48550/arXiv.1610.05492>.
- [33] F. Sattler, S. Wiedemann, K. R. Muller, and W. Samek, "Robust and Communication-Efficient Federated Learning from Non-i.i.d. Data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 9, pp. 3400–3413, 2020, <https://doi.org/10.1109/TNNLS.2019.2944481>.
- [34] M. Zhu, "Solving Class Imbalance in Medical Image Classification Based on Federated Learning," in *Proceedings of the 2025 2nd International Conference on Generative Artificial Intelligence and Information Security, GAIS 2025*, pp. 546–550, 2025, <https://doi.org/10.1145/3728725.3728811>.
- [35] X. Wei *et al.*, "Breaking Data Efficiency Dilemma: A Federated and Augmented Learning Framework for Alzheimer's Disease Detection Via Speech," *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 19147-19151, 2026, <https://doi.org/10.1109/ICASSP55912.2026.11463930>.
- [36] J. Yang *et al.*, "MedMNIST v2 -- A large-scale lightweight benchmark for 2D and 3D biomedical image classification," *Scientific data*, vol. 10, no. 1, p. 41, 2023, <https://doi.org/10.1038/s41597-022-01721-8>.
- [37] X. Xia *et al.*, "Benchmarking deep models on retinal fundus disease diagnosis and a large-scale dataset," *Signal Process. Image Commun.*, vol. 127, p. 117151, 2024, <https://doi.org/10.1016/j.image.2024.117151>.
- [38] A. Mabrouk, R. P. D. Redondo, A. Dahou, M. A. Elaziz, and M. Kayed, "Pneumonia Detection on chest X-ray images Using Ensemble of Deep Convolutional Neural Networks," *Applied Sciences*, vol. 12, no. 13, p. 6448, 2022, <https://doi.org/10.3390/app12136448>.
- [39] G. Dlamini and M. Fahim, "DGM: a data generative model to improve minority class presence in anomaly detection domain," *Neural Comput. Appl.*, vol. 33, no. 20, pp. 13635–13646, 2021, <https://doi.org/10.1007/s00521-021-05993-w>.
- [40] J. Kong, W. Kowalczyk, S. Menzel, and T. Bäck, "Improving imbalanced classification by anomaly detection," in *International conference on parallel problem solving from nature*, pp. 512–523, 2020, https://doi.org/10.1007/978-3-030-58112-1_35.
- [41] J. Stefanowski, "Overlapping, Rare Examples and Class Decomposition in Learning Classifiers from Imbalanced Data," in *Emerging paradigms in machine learning*, pp. 277–306, 2013, https://doi.org/10.1007/978-3-642-28699-5_11.
- [42] G. Pang, C. Shen, L. Cao, and A. Van Den Hengel, "Deep Learning for Anomaly Detection: A Review," *ACM computing surveys (CSUR)*, vol. 54, no. 2, pp. 1-38, 2021, <https://doi.org/10.1145/3439950>.
- [43] M. Lübbering, R. Ramamurthy, M. Gebauer, T. Bell, R. Sifa, and C. Bauckhage, "From Imbalanced Classification to Supervised Outlier Detection Problems: Adversarially Trained Auto Encoders," in *International Conference on Artificial Neural Networks*, pp. 27–38, 2020, https://doi.org/10.1007/978-3-030-61609-0_3.
- [44] M. C. Massi, F. Ieva, F. Gasperoni, and A. M. Paganoni, "Feature Selection for Imbalanced Data with Deep Sparse Autoencoders Ensemble," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 15, no. 3, pp. 376-395, 2022, <https://doi.org/10.1002/sam.11567>.
- [45] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriya, A. Dehghantanha and G. Srivastava, "Federated-Learning-Based Anomaly Detection for IoT Security Attacks," in *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2545-2554, 2022, <https://doi.org/10.1109/JIOT.2021.3077803>.
- [46] R. Alhammadi, A. Gawanmeh, S. Atalla, M. Q. Alkhatib and W. Mansoor, "Performance Evaluation of Federated Learning for Anomaly Network Detection," *2023 Congress in Computer Science, Computer Engineering, & Applied Computing (CSCE)*, pp. 116-122, 2023, <https://doi.org/10.1109/CSCE60160.2023.00024>.
- [47] L. Li, Y. Fan, M. Tse, and K. Y. Lin, "A review of applications in federated learning," *Computers & Industrial Engineering*, vol. 149, p. 106854, 2020, <https://doi.org/10.1016/j.cie.2020.106854>.
- [48] N. Yang, X. Chen, C. Z. Liu, D. Yuan, W. Bao, and L. Cui, "FedMAE: Federated Self-Supervised Learning with One-Block Masked Auto-Encoder," *arXiv preprint arXiv:2303.11339*, 2023, <https://doi.org/10.48550/arXiv.2303.11339>.
- [49] Y. Wu *et al.*, "Federated Self-Supervised Contrastive Learning and Masked Autoencoder for Dermatological Disease Diagnosis," *arXiv preprint arXiv:2208.11278*, 2022, <https://doi.org/10.48550/arXiv.2208.11278>.
- [50] M. Nardi, L. Valerio and A. Passarella, "Anomaly Detection Through Unsupervised Federated Learning," *2022 18th International Conference on Mobility, Sensing and Networking (MSN)*, pp. 495-501, 2022, <https://doi.org/10.1109/MSN57253.2022.00085>.
- [51] A. Anwar *et al.*, "FedAD-Bench: A Unified Benchmark for Federated Unsupervised Anomaly Detection in Tabular Data," *2024 2nd International Conference on Federated Learning Technologies and Applications (FLTA)*, pp. 115-122, 2024, <https://doi.org/10.1109/FLTA63145.2024.10839838>.

-
- [52] Y. Huh, B. Kim, and W. Choi, "Federated Learning Enhanced by Feature Reconstruction for Semantic Communication Module Updates of Agents," *arXiv preprint arXiv:2508.02048*, 2025, <https://doi.org/10.48550/arXiv.2508.02048>.
- [53] F. Angiulli, F. Fassetti, and L. Ferragina, "Reconstruction Error-based Anomaly Detection with Few Outlying Examples," *Neurocomputing*, vol. 675, p. 133002, 2026, <https://doi.org/10.1016/j.neucom.2026.133002>.
- [54] M. A. Rassam, "Autoencoder-Based Neural Network Model for Anomaly Detection in Wireless Body Area Networks," *Internet of Things*, vol. 5, no. 4, pp. 852–870, 2024, <https://doi.org/10.3390/iot5040039>.
- [55] Y. Liu, S. Rini and J. Chen, "GCFed: Exploiting Gradient Correlation for Client Selection and Rate Allocation in Federated Learning," in *IEEE Internet of Things Journal*, 2026, <https://doi.org/10.1109/JIOT.2026.3687127>.
- [56] Y. Zeng *et al.*, "Adaptive Federated Learning With Non-IID Data," *The Computer Journal*, vol. 66, no. 11, pp. 2758–2772, 2023, <https://doi.org/10.1093/comjnl/bxac118>.
- [57] M. Yashwanth, G. K. Nayak, A. Singh, Y. Simmhan, and A. Chakraborty, "Adaptive Self-Distillation for Minimizing Client Drift in Heterogeneous Federated Learning," *arXiv preprint arXiv:2305.19600*, 2023, <https://doi.org/10.48550/arXiv.2305.19600>.
- [58] X. Chen, Y. Liu, H. Du, M. Wang, and Y. Zhao, "Taming Gradient Variance in Federated Learning with Networked Control Variates," *arXiv preprint arXiv:2310.17200*, 2023, <https://doi.org/10.48550/arXiv.2310.17200>.
- [59] R. Niu, Y. Wang, C. Yang, and X. Cao, "CAAFL: A Client-Level Adaptive Accelerative Federated Learning Optimization Algorithm for Heterogeneous Data Scenarios," in *Proceedings of the 36th Chinese Control and Decision Conference, CCDC 2024*, pp. 3859–3864, 2024, <https://doi.org/10.1109/CCDC62350.2024.10587872>.