Dynamic Weighting and Aggregation in Multimodal Federated Learning for Disease Prediction

Ali Anaissi[1,2], Lesley Manantan[1], Niushad Saeed[1], Jeevani Fernando[1], Manish Sharma [1], Puneykan Arora [1], Sai Dayakar Reddy Gollapudi [1], Weidong Huang[2], and Ali Braytee[2]

[1] University of Sydney, Australia
[2] University of Technology Sydney, Australia
ali.anaissi@sydney.edu.au, lman0453@uni.sydney.edu.au, nsae0696@uni.sydney.edu.au, sfer0786@uni.sydney.edu.au, msha94546@uni.sydney.edu.au, paro7137@uni.sydney.edu.au, sgol0977@uni.sydney.edu.au, weidong.huang@uts.edu.au, ali.braytee@uts.edu.au

Abstract. The advent of Natural Language Processing (NLP) and Large Language Models (LLMs) has transformed data-driven solutions in healthcare. However, significant challenges remain in managing the growing complexity and diversity of medical data under stringent privacy regulations. Traditional centralized models for disease prediction, which rely on data sharing, often face trade-offs between performance and patient confidentiality. These issues are further amplified in modern healthcare, where integrating multimodal data, such as medical images and clinical text, is critical for accurate and comprehensive diagnosis.

While federated learning frameworks address privacy concerns by keeping data decentralized, they often struggle with data heterogeneity and limited support for multimodal inputs. To address these limitations, this paper introduces a novel multimodal federated learning framework for disease prediction. The proposed framework employs ResNet-50 for image feature extraction and bidirectional LSTMs for clinical text analysis, enabling effective multimodal data fusion. Furthermore, an adaptive dynamic weighted aggregation algorithm is proposed to adjust client contributions based on data quality and model performance, thereby enhancing the stability and accuracy of the global model.

Evaluated on lung tumor datasets, the proposed approach outperformed state-of-the-art methods across multiple evaluation metrics, demonstrating its robustness, scalability, and strong potential for real-world healthcare applications.

Keywords: Federated Learning, Multimodel data, Dynamic aggregation algorithm.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

1 Introduction

In recent years, the application of deep learning in medical diagnosis, particularly for disease detection and prediction, has gained substantial traction [1]. However, the legitimate use of medical data remains a critical challenge. Traditional centralized training models rely on uploading raw data to a server, which often involves sensitive patient information. Strict data access restrictions within and across organizations lead to data silos, limiting the generalization of models and ultimately compromising their accuracy and performance [2].

Federated learning models offer a promising solution to address the data silo problem. This decentralized training approach eliminates the need for sharing raw data, instead enabling local model training while aggregating updates on a central server. Healthcare institutions train models locally on private data and share model parameters with a central server. The server aggregates these updates using methods such as weighted averaging, updates the global model, and redistributes the updated parameters. This iterative process improves the model's generalization capabilities while preserving data privacy, making federated learning a robust approach for healthcare data applications [3].

Despite its advantages, federated learning faces several challenges, particularly in aggregating multimodal data, such as medical images and clinical text, and designing adaptive aggregation algorithms. Integrating Natural Language Processing (NLP) and Large Language Models (LLMs) into federated learning frameworks can significantly advance these efforts. While visual data, such as medical images, captures anatomical features, clinical text provides crucial contextual information about patient history, diagnoses, and treatments. LLMs like BERT and GPT excel in processing unstructured text, extracting nuanced insights, and enabling the integration of multimodal data with visual features. However, federated learning systems often struggle to fuse such diverse data sources effectively, especially in Non-Independent and Identically Distributed (Non-IID) environments [4, 5].

This study aims to address the limitations of current federated learning models in healthcare by introducing a novel framework that incorporates NLP and LLMs for multimodal data fusion. Specifically, we propose a federated learning architecture that combines the ResNet-50 model for medical image analysis with a bidirectional LSTM and LLM-based embeddings for clinical text processing. The multimodal features are then fused into a unified framework, enabling comprehensive disease prediction and diagnosis.

Additionally, we introduce a dynamic weighted aggregation algorithm that adjusts global model aggregation weights based on the magnitude of updates from individual clients. This adaptive approach enhances model stability, particularly in heterogeneous and complex data environments. Furthermore, integrating LLMs for text feature extraction ensures the effective representation of clinical text, addressing challenges associated with noisy and incomplete datasets. By leveraging these innovations, our framework offers a scalable solution for collaborative healthcare applications while maintaining stringent privacy standards.

The key contributions of this work are:

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

– Integration of NLP and LLMs with deep learning models to process both medical images and clinical text, addressing the challenge of multimodal data fusion in federated learning.
– Development of an adaptive aggregation method that improves model stability and performance in complex, Non-IID environments by leveraging client-specific contributions.
– Utilization of advanced NLP techniques and LLMs, such as BERT and GPT, to extract robust features from clinical text, enabling comprehensive analysis when combined with visual features.
– Providing practical guidance for cross-institutional learning with stringent privacy protection, advancing the application of federated learning in healthcare.

The paper is organized as follows: Section II reviews related work, providing an overview of prior research in the field and highlighting gaps addressed by this study. Section III outlines the proposed methodology, detailing the framework, algorithms, and techniques employed. Section IV presents experimental results, including evaluations and comparative analyses. Finally, Section V concludes the paper, summarizing findings and discussing implications for future research.

2 Related Work

Federated Learning (FL) has gained significant traction in the medical domain as a transformative solution for addressing challenges related to data privacy, decentralized collaboration, and disease prediction. By enabling decentralized training, FL ensures sensitive patient data remains local, with only model updates shared to a central server. This approach mitigates privacy risks associated with traditional centralized models and fosters collaborative advancements across institutions. For instance, Google and the Mayo Clinic utilized FL for a cancer detection system, allowing model training on decentralized datasets without exposing sensitive information. Similarly, Owkin, a French biotech firm, deployed FL to analyze tumor data collaboratively, achieving enhanced diagnostic accuracy while maintaining patient confidentiality [1, 6]. Such applications illustrate the dual benefits of FL in advancing medical diagnostics and upholding stringent privacy standards [7]. Specific use cases include Google Health's FL-based diabetic retinopathy detection using retinal images from multiple hospitals to improve prediction accuracy across diverse patient groups [8], and Intel's brain tumor detection model developed with the University of Pennsylvania, pooling data from 14 hospitals to achieve unprecedented accuracy in predictive modeling [9]. These implementations highlight FL's capacity to leverage distributed data for creating accurate and generalizable models while addressing privacy concerns.

The integration of multimodal data, combining medical imaging, genomics, and clinical records, has expanded FL's potential. Multimodal FL synthesizes diverse data types, enabling personalized diagnostics and treatment recommendations. For example, during the COVID-19 pandemic, multimodal FL integrated

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

X-rays, ultrasound images, and clinical data to deliver tailored diagnoses and treatments [10]. These frameworks improve diagnostic generalization by incorporating extensive datasets from multiple institutions, as emphasized by Sheller [7], highlighting their critical role in advancing the robustness of medical models.

Despite its advantages, FL faces technical hurdles. Frequent parameter updates between clients and servers create communication bottlenecks, addressed through gradient compression [11], quantization [12], and asynchronous updates [13]. Uneven computing power among clients challenges model convergence, mitigated by adaptive task allocation and hierarchical aggregation [14, 15]. Aggregation algorithms play a pivotal role; FedAvg is effective in stable environments, while FedProx incorporates regularization to manage data and computational heterogeneity, making it suitable for complex medical applications [16].

Building on this foundation, our work introduces several innovations to enhance FL in heterogeneous environments. A weight difference method reduces communication overhead by transmitting only model weight changes since the last update. Dynamic weight adjustment allocates aggregation weights based on client contributions, enabling clients with greater computational power to have a more significant influence on the global model. This adaptive approach ensures robust performance in diverse and complex environments. Additionally, the weight differential mechanism enhances privacy protection, reduces data leakage risks, and improves generalization capabilities. Together, these advancements address key challenges in FL, offering scalable and efficient solutions for collaborative healthcare applications while maintaining stringent data privacy.

3 Methodology

Recent advancements in Natural Language Processing (NLP) and Large Language Models (LLMs) have demonstrated transformative potential across various domains, including healthcare. These technologies enable sophisticated analysis and understanding of textual data, complementing deep learning techniques for image processing. Leveraging the power of NLP and LLMs, we propose a novel federated learning framework that integrates a multimodal model with an innovative dynamic aggregation algorithm. This framework is specifically designed to tackle the challenges posed by heterogeneous data distributions in federated learning, including the need for effective fusion of diverse data modalities such as medical images and clinical text. The proposed framework utilizes advanced NLP techniques and LLMs for text analysis, coupled with deep learning methods like ResNet-50 for image processing. These components are integrated into a robust dynamic aggregation algorithm that adapts to variations in client contributions, ensuring improved global model stability and accuracy. By decentralizing training and maintaining strict data privacy, the framework addresses key challenges in healthcare applications, enabling secure and efficient learning in decentralized environments while enhancing performance and scalability.

First, we will present the multimodal federated learning framework, and then we will cover our novel dynamic aggregation algorithm. The core of the mul-

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

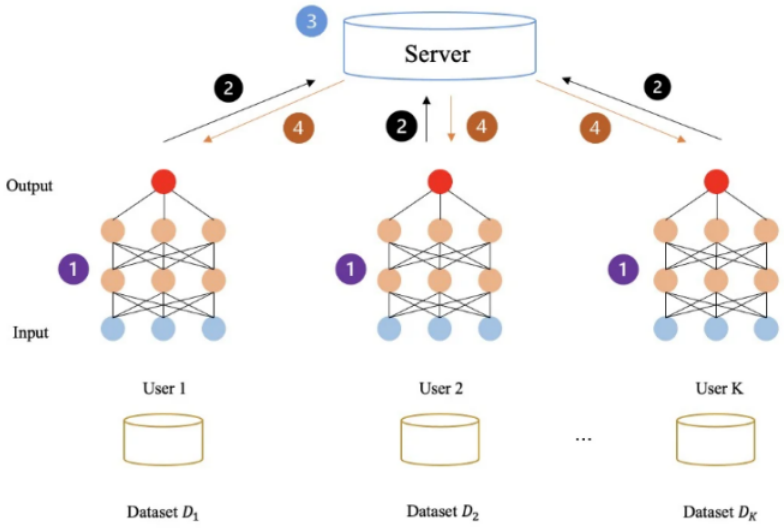


Fig. 1: Federated Learning Components and Processes

timodal model lies in its ability to process both image and text data, using distinct architectures tailored for each modality. Specifically, the framework employs a ResNet50 network for image processing and a bidirectional LSTM (Long Short-Term Memory) network for text processing. These components facilitate joint learning across modalities, enabling effective feature extraction and fusion to create a comprehensive representation of the multimodal data. ResNet50, a residual network architecture, is used for extracting deep visual features from images, making it particularly effective for tasks like medical image classification. On the other hand, the bidirectional LSTM captures context from both directions of the text sequence, enhancing the model's ability to understand complex semantic relationships in medical texts.

To address the challenges of heterogeneous data distributions due to the Non-IID (non-independent and identically distributed) problem and varying client capabilities in federated learning, we introduce an innovative, dynamic aggregation algorithm. This algorithm adjusts the contribution of each client to the global model based on the magnitude of changes in client model weights. The goal is to ensure stability by enhancing the robustness of the global model, especially in the presence of Non-IID data, and improve performance by mitigating degradation caused by uneven data distributions and computational resource disparities across clients. The server dynamically calculates client contributions during aggregation, enabling adaptive updates that enhance learning outcomes. To further improve system stability and ensure data privacy, the framework incorporates early stopping, which halts training when performance plateaus, and amplitude controls, which limit excessive weight changes during aggregation.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

Figure 1 illustrates the overall architecture of the federated learning framework and the interaction process between its components. The system consists of multiple clients and a central server, with training occurring across multiple steps.

Step 1: Local Training by the Client: Each client trains the model locally using its multimodal dataset (e.g., Dataset D_1, D_2, ..., D_K). This ensures data privacy, as the data remains local to each client.

Step 2: Uploading Model Weights to the Server: After local training, the client uploads the updated model weights to the server.

Step 3: Server Aggregates Weights: The server dynamically aggregates the model weights from all clients using the dynamic weighting algorithm. Each client's contribution is weighted according to the magnitude of changes in its model weights during training.

Step 4: Server Distributes Updated Global Model: The server distributes the aggregated global model weights back to each client for the next round of local training.

The server plays a central role in coordinating and aggregating the model updates from all clients. It starts by initializing a global model (e.g., a multimodal ResNet-LSTM model) and distributing it to the clients for initial local training. After each training round, the server collects the updated model weights from the clients, applies our proposed dynamic weighting algorithm to aggregate them, and sends the updated global model back to the clients. This process helps ensure the model converges quickly, even in environments with heterogeneous data.

The dynamic weighting algorithm offers a more precise and adaptable aggregation strategy than traditional federated learning methods such as FedAvg [17], which averages client weights without accounting for variations in client data distribution or computational contributions. Our approach ensures that clients with greater contributions to model training are given more influence in the aggregation process, enhancing the generalization ability and stability of the global model. The method is described in Algorithm 1 and operates as follows:

- *Local Weight Change Calculation:* After completing local training, each client calculates the changes in its model weights across all layers. These changes are aggregated to quantify the client's contribution during the training round.
- *Weight Coefficient Computation:* The server collects the weight change magnitudes from all clients and computes a weight coefficient for each. These coefficients are normalized based on the relative weight change ratios, ensuring that the influence of each client is proportional to its contribution.
- *Weighted Aggregation:* Using the computed coefficients, the server aggregates the weight changes for each layer, generating a weighted update that reflects the varying contributions of the clients.
- *Global Model Update:* The server applies the weighted update to the global model, resulting in an improved and updated global model.

By dynamically adjusting the aggregation process based on client contributions, this algorithm ensures balanced and robust model updates, particularly

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

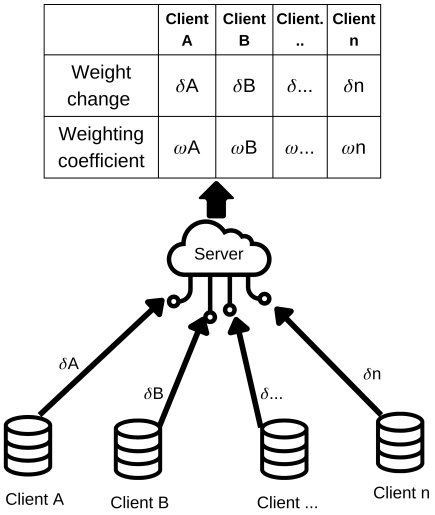


Fig. 2: Weighting coefficient calculation process

in heterogeneous data and computational environments. The detailed steps are described below.

In each round of local training, client k trains the model using its local dataset and updates the weights of each model layer. For the i-th layer, the weight change of client k is computed as the difference between the weight of the subsequent layer and the current layer, representing the weight update generated during the current round of training, without involving weight changes from previous rounds. The weight change magnitude δ_k for client k is defined as the sum of the L_2-norms of the weight changes across all layers and is calculated as follows:

$$\delta_k = \sum_{i=1}^{L} \left\| W_k^{i+1} - W_k^i \right\|_2 \tag{1}$$

Here, L represents the total number of layers in the model, i denotes the index of each layer, and k identifies the specific client.

On the server side, after collecting the weight change amplitudes δ_k uploaded by all clients, the server first computes the total sum of these weight change amplitudes ($\sum_{k=1}^{K} \delta_k$). This sum is then used to normalize the weight coefficient for each client. The weight coefficient ω_k for client k is determined based on its weight change amplitude, reflecting its relative contribution to the global model aggregation. Specifically, ω_k is calculated as the ratio of the client's weight change amplitude to the total sum of weight changes, as shown below

$$\omega_k = \frac{\delta_k}{\sum_{k=1}^{K} \delta_k + \epsilon} \tag{2}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

Here, ϵ is a small positive constant added to the denominator to prevent division by zero, ensuring numerical stability, and K is the total number of clients. The weight coefficient ω_k represents the relative influence of client k in the aggregation process. Clients with larger weight changes are assigned higher weight coefficients, significantly contributing to the global model update. These coefficients are subsequently used to perform a weighted average of the weight changes across all clients in the next step of the aggregation process. Figure 2 illustrates the process of calculating the weighting coefficients for each client during the aggregation phase.

The server then uses the weight coefficient ω_k of each client to perform hierarchical weighted aggregation on the weight updates of all clients, resulting in the final weight update ΔW_i for each layer of the global model. To further optimize communication efficiency and reduce the volume of transmitted data, the weight differential function is introduced in this step. The key idea behind the weight differential function is that clients upload only the weight changes for each layer from their local training rather than the complete weight values. This approach effectively minimizes the transmission bandwidth required for data exchange.

For each layer i of the model, the server computes the weighted aggregated weight change ΔW_i for the ith layer across all clients. The formula for this calculation is as follows:

$$\Delta W_i = \sum_{k=1}^{K} \omega_k \cdot \left(W_k^{i+1} - W_k^i\right) \tag{3}$$

Here, ω_k represents the weight coefficient of the client k, which is determined by the magnitude of the weight change amplitude for that client. The term within the parentheses denotes the weight update for the ith layer contributed by client k during the current round of local training, also referred to as the weight difference.

This formula shows that each client only needs to upload its weight updates (i.e., the differential values), enabling the server to compute the aggregated updates ΔW_i for each layer through a weighted summation of these updates. By introducing weight differentiation, clients transmit only the changes in weights rather than the complete weight information, significantly reducing communication overhead and improving efficiency. This also minimizes the risk of data leakage, as the transmitted updates do not directly expose sensitive information. Furthermore, by reducing transmission volume, the server's computational load during aggregation is decreased, allowing the process to complete more quickly.

The server then uses the weighted aggregate update ΔW_i for each layer to update the weights of the global model. This step integrates the training contributions of all clients to form an improved model. For each layer i of the model, the server updates the global model's weights using the following formula:

$$W_i^{t+1} = W_i^t + \Delta W_i \tag{4}$$

Here, t represents the training round of the global model, and i is the index of the layer. ΔW_i is the aggregated update for layer i, calculated in the previous

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

Table 1: Global Model Weight Updates for Each Layer

Layer Index	Weighted Update Weights
1	$\Delta W_1 = \omega_A \cdot \Delta W_{A1} + \omega_B \cdot \Delta W_{B1} + \cdots + \omega_n \cdot \Delta W_{n1}$
2	$\Delta W_2 = \omega_A \cdot \Delta W_{A2} + \omega_B \cdot \Delta W_{B2} + \cdots + \omega_n \cdot \Delta W_{n2}$
$\vdots$	$\vdots$
n	$\Delta W_n = \omega_A \cdot \Delta W_{An} + \omega_B \cdot \Delta W_{Bn} + \cdots + \omega_n \cdot \Delta W_{nn}$

Algorithm 1 Our proposed framework

Input:
  K clients, each with local dataset D_k, $k \in [1, K]$
  Global model with L layers
Output: Aggregated global model weights
for each client $k \in [1, K]$ **do**
  Client k trains the model locally using dataset D_k
  For each layer $i \in [1, L]$, compute the weight change δ_k for client using Eq. 1
end for
Server collects the weight change magnitudes δ_k from all clients
Compute the total sum of weight change magnitudes using $\sum_{k=1}^{K} \delta_k$
for each client $k \in [1, K]$ **do**
  Compute the weight coefficient ω_k for client k using Eq. 2
end for
Aggregation:
for each layer $i \in [1, L]$ **do**
  for each client $k \in [1, K]$ **do**
    Compute the weight update for client k for layer i using Eq. 3
  end for
end for
Update global model:
for each layer $i \in [1, L]$ **do**
  Update global model weight for layer i using Eq. 4
end for
Return updated global model

step, representing the combined contribution of all clients to the weight change for that layer. This operation is performed independently for each layer, ensuring that every layer of the global model can absorb the training updates from all clients and complete the weight update for that round.

As shown in Table 1, each layer's weight update is calculated using a weighted aggregation of the updates from all participating clients. The weight coefficients $\omega_A, \omega_B, \ldots, \omega_n$ correspond to the clients' contributions based on their weight change magnitudes, ensuring that more influential clients have a higher impact on the global model.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

4 Experiments

4.1 Dataset Preparation and Preprocessing

We utilized two public multi-modal chest X-ray datasets to enhance predictive modeling. The IU-CXR dataset [18] includes 7,470 images and radiology reports linked to unique patient IDs. Over 100 disease labels from the *Problems* column were grouped into 14 clinically relevant categories and numerically encoded. Reports were preprocessed by merging *Findings* and *Impression* sections into a *notes* column, tokenized with a 10,000-word vocabulary, converted to integer sequences, and padded to length 100. CXR images were converted to RGB, resized to 224×224, normalized to $[0, 1]$, and stored for model input.

The MIMIC-CXR dataset [19] contains 370,000 images and reports from 227,000 studies. Reports and images were preprocessed similarly to IU-CXR. Integrating textual and imaging data from both datasets enables robust multi-modal modeling for tasks such as disease classification, report generation, and representation learning.

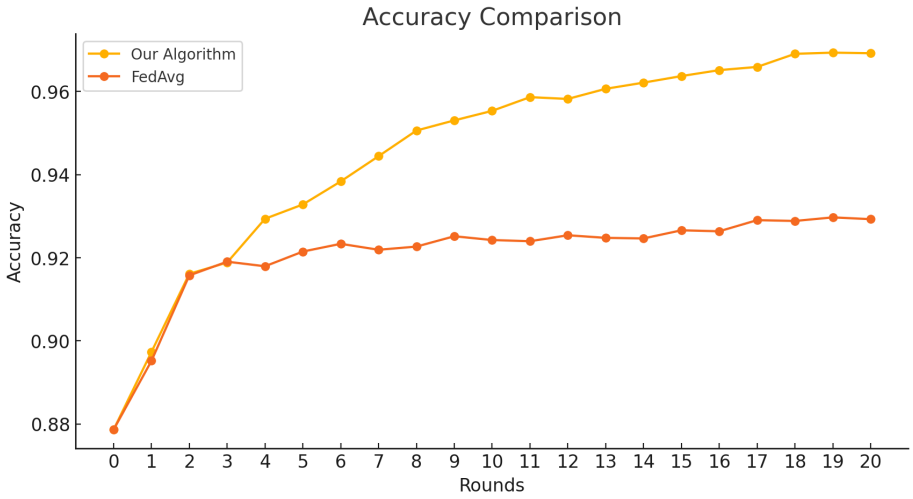
4.2 Results and Discussion

We evaluated the effectiveness of our proposed algorithm by comparing its performance with the FedAvg and FedProx algorithms, using identical model architectures and data allocations. Evaluation metrics, including accuracy, loss, and F1 score, were calculated and averaged across all clients to ensure fairness and minimize the impact of performance discrepancies. The final testing results were derived from the last validation round, providing a comprehensive comparison.

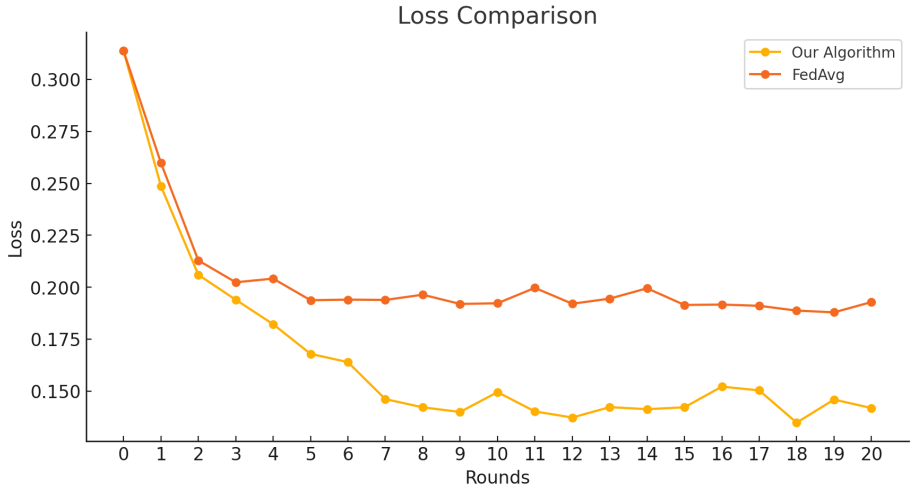
Figures 3a and 3b display the validation accuracy and loss trends for our algorithm compared to the traditional FedAvg algorithm using IU-CXR dataset. Both algorithms were trained using identical hyperparameters: 20 rounds, 10 epochs per round, and 3 clients. Results from the untrained global model at round 0 serve as a baseline. Our algorithm achieved a validation accuracy of 0.97, surpassing FedAvg's 0.93, and reduced the loss to 0.14, compared to FedAvg's 0.19, indicating faster convergence and superior overall performance. Using the MIMIC-CXR dataset, our algorithm demonstrated similarly outstanding performance. Figures 3a and 3b show the trends in validation accuracy and loss during training. Our algorithm achieved a validation accuracy of 0.97, outperforming FedAvg's 0.94, and reduced the loss to 0.15, compared to FedAvg's 0.22. These results reaffirm the robustness of our approach in achieving faster convergence and superior predictive capabilities across diverse datasets.

Table 2 shows that our algorithm outperformed FedAvg across all metrics. Notably, it achieved a significantly higher recall (0.85 vs. 0.56 for FedAvg), indicating its superior sensitivity in identifying positive class samples, a critical factor in medical applications where missing diagnoses could have serious consequences. Furthermore, the higher F1 score reflects a better balance between precision and recall, highlighting our algorithm's ability to maintain both accuracy and robustness. Similar trends were observed with the MIMIC-CXR dataset,

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11



(a) Validation accuracy comparison between our algorithm and FedAvg.



(b) Validation loss comparison between our algorithm and FedAvg.

Fig. 3: Comparison between our algorithm and FedAvg: (a) validation accuracy and (b) validation loss.

where our algorithm achieved a recall of 0.82, compared to FedAvg's 0.60, and an F1 score of 0.78, compared to FedAvg's 0.74.

We also compared our algorithm's performance with FedProx, a widely used baseline in federated learning. Our algorithm outperformed FedProx, achieving a validation accuracy of 0.97 compared to FedProx's 0.94 using IU-CXR dataset. Moreover, our method demonstrated a more balanced performance across key

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

Table 2: Performance comparison of FedAvg, FedProx, and proposed algorithm across datasets.

Metric	Dataset	FedAvg	FedProx	Proposed Algorithm
Recall	IU-CXR	0.56	0.60	0.85
	MIMIC-CXR	0.60	0.65	0.82
F1 Score	IU-CXR	0.72	0.75	0.88
	MIMIC-CXR	0.74	0.76	0.78
Loss	IU-CXR	0.19	0.18	0.14
	MIMIC-CXR	0.22	0.20	0.15
Accuracy	IU-CXR	0.93	0.94	0.97
	MIMIC-CXR	0.94	0.96	0.97

metrics such as precision, recall, and F1 score. Similarly, with the MIMIC-CXR dataset, our algorithm achieved a validation accuracy of 0.97, compared to FedProx's 0.95. While the accuracy improvement was modest, the recall of our algorithm was significantly higher (0.82 vs. 0.65 for FedProx), ensuring fewer false negatives. This aspect is particularly crucial in applications like healthcare, where missing high-risk cases can have severe consequences.

Our results show that assigning higher weights to clients with larger model weight updates plays a critical role in accelerating convergence and improving model stability. By leveraging this dynamic weighting strategy, our novel federated aggregation algorithm not only achieves higher validation accuracy but also ensures a more balanced performance across key metrics such as precision, recall, and F1 score. These advancements are particularly valuable for integrating multi-center medical data, enhancing both the robustness and predictive accuracy of the global model.

This work further demonstrates the potential of federated learning for multimodal medical data fusion, combining innovative aggregation techniques with robust data processing to improve predictive performance. The proposed algorithm offers an adaptable, scalable solution for collaborative medical prediction tasks, with applications extending to cross-institutional collaboration in areas where data privacy is a major concern.

5 conclusion

In this study, we presented a robust multimodal federated learning framework for disease prediction, integrating medical imaging and clinical text data using ResNet-50 and bi-directional LSTM models. The incorporation of a novel dynamic aggregation algorithm addressed challenges related to uneven data distribution and varying client contributions, significantly enhancing model convergence and stability. Our experimental results demonstrated that the proposed algorithm outperformed the standard FedAvg approach, achieving superior accuracy and recall rates, thereby highlighting its effectiveness in healthcare applications where accurate and reliable predictions are critical. The integration of

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

machine learning and deep learning into federated learning frameworks marks a transformative step for cross-domain collaborations, particularly in healthcare, where data privacy remains a paramount concern. Additionally, the proposed dynamic aggregation algorithm offers scalability and adaptability, extending its applicability beyond healthcare to a variety of federated learning tasks. Despite certain limitations, this work lays a solid foundation for future research, aiming to refine federated learning methodologies and foster more secure, collaborative, and effective systems for disease prediction and medical decision support.

References

1. Matteo Pennisi, Federica Proietto Salanitri, Giovanni Bellitto, Bruno Casella, Marco Aldinucci, Simone Palazzo, and Concetto Spampinato. Feder: Federated learning through experience replay and privacy-preserving data synthesis. *Computer Vision and Image Understanding*, 238:103882, 2024.
2. Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.
3. Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE signal processing magazine*, 37(3):50–60, 2020.
4. Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2):423–443, 2018.
5. S Hochreiter. Long short-term memory. *Neural Computation MIT-Press*, 1997.
6. Hao Guan, Pew-Thian Yap, Andrea Bozoki, and Mingxia Liu. Federated learning for medical image analysis: A survey. *Pattern Recognition*, page 110424, 2024.
7. Micah J Sheller, Brandon Edwards, G Anthony Reina, Jason Martin, Sarthak Pati, Aikaterini Kotrotsou, Mikhail Milchenko, Weilin Xu, Daniel Marcus, Rivka R Colen, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, 10(1):12598, 2020.
8. Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N Galtier, Bennett A Landman, Klaus Maier-Hein, et al. The future of digital health with federated learning. *NPJ digital medicine*, 3(1):1–7, 2020.
9. Ittai Dayan, Holger R Roth, Aoxiao Zhong, Ahmed Harouni, Amilcare Gentili, Anas Z Abidin, Andrew Liu, Anthony Beardsworth Costa, Bradford J Wood, Chien-Sung Tsai, et al. Federated learning for predicting clinical outcomes in patients with covid-19. *Nature medicine*, 27(10):1735–1743, 2021.
10. Jie Xu, Benjamin S Glicksberg, Chang Su, Peter Walker, Jiang Bian, and Fei Wang. Federated learning for healthcare informatics. *Journal of healthcare informatics research*, 5:1–19, 2021.
11. Hongyi Wang, Scott Sievert, Shengchao Liu, Zachary Charles, Dimitris Papailiopoulos, and Stephen Wright. Atomo: Communication-efficient learning via atomic sparsification. *Advances in neural information processing systems*, 31, 2018.
12. Jeremy Bernstein, Yu-Xiang Wang, Kamyar Azizzadenesheli, and Animashree Anandkumar. signsgd: Compressed optimisation for non-convex problems. In *International Conference on Machine Learning*, pages 560–569. PMLR, 2018.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11

13. Cong Xie, Sanmi Koyejo, and Indranil Gupta. Asynchronous federated optimization. *arXiv preprint arXiv:1903.03934*, 2019.
14. Lumin Liu, Jun Zhang, Shenghui Song, and Khaled B Letaief. Hierarchical federated learning with quantization: Convergence analysis and system design. *IEEE Transactions on Wireless Communications*, 22(1):2–18, 2022.
15. Siqi Luo, Xu Chen, Qiong Wu, Zhi Zhou, and Shuai Yu. Hfel: Joint edge association and resource allocation for cost-efficient hierarchical federated edge learning. *IEEE Transactions on Wireless Communications*, 19(10):6535–6548, 2020.
16. Li Li, Yuxi Fan, Mike Tse, and Kuo-Yi Lin. A review of applications in federated learning. *Computers & Industrial Engineering*, 149:106854, 2020.
17. H. Brendan McMahan, E. Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 1273–1282, 2017.
18. Dina Demner-Fushman, Marc D Kohli, Marc B Rosenman, Sonya E Shooshan, Laritza Rodriguez, Sameer Antani, George R Thoma, and Clement J McDonald. Preparing a collection of radiology examinations for distribution and retrieval. *Journal of the American Medical Informatics Association*, 23(2):304–310, 2016.
19. Alistair E. W. Johnson, Tom J. Pollard, Seth J. Berkowitz, Nathaniel R. Greenbaum, Matthew P. Lungren, Christopher Y. Deng, Roger G. Mark, Steven Horng, Jeffrey G. Rogers, Russell A. Taylor, and Adrian D. Haimovich. Mimic-cxr: A large publicly available database of labeled chest radiographs. *arXiv preprint arXiv:1901.07042*, 2019.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29924-6_11