Received: 4 August 2025 | Revised: 4 Novemeber 2025 | Accepted: 16 January 2026 | Published online: 11 Feburary 2026

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2026, Vol. 00(00) 1–11
DOI: 10.47852/bonviewJCCE62027069

BON VIEW PUBLISHING

FedProx-Enhanced Federated Transfer Learning for Heterogeneous 3D Medical Image Classification

Manjunath Naganna[1], **Guru Ramachandra Nayaka**[1,*], **Natesh Mahadev**[2] **and Mayura Tapkire**[3]

[1] *Department of Computer Science and Engineering, JSS Science and Technology University, India*

[2] *Department of Computer Science and Engineering, Vidyavardhaka College of Engineering, India*

[3] *Department of Computer Science and Engineering, The National Institute of Engineering, India*

Abstract: Artificial intelligence has revolutionized the analysis of medical images, but the creation of robust models is still problematic because of the strict regulation of data privacy and the very nature of medical imaging data heterogeneity. Federated learning (FL) is a privacy-preserving method with convergence instability and worse performance, particularly with non-Independent and Identically Distributed (non-IID) data. In addressing these disadvantages, this paper suggests a new FL framework based on the integration of transfer learning (TL) and FedProx to increase the resilience and stability of 3D medical image classification models in the distributed hospital context. TL offers a knowledgeable starting point to local models that enhances adaptation to different client data, whereas FedProx presents a proximal term to decrease the effects of data variability. The proposed framework was tested on three 3D medical imaging datasets, namely, OrganMNIST, FractureMNIST, and NoduleMNIST, with non-IID data distributions of 10 clients. It has been experimentally demonstrated that the performance is significantly improved, with an accuracy improvement of up to 18.2 and higher precision, recall, and F1 scores than current FL methods. All in all, the suggested solution offers a good and privacy-aware solution to collaborative learning in heterogeneous 3D medical imaging settings.

Keywords: federated learning, medical imaging, data heterogeneity, healthcare

1. Introduction

Analysis of medical images is the most important element of contemporary healthcare, which allows the improvement of diagnosis, treatment planning, and monitoring of the disease. Initial work in this area was mainly concentrated on 2D image information (e.g., X-rays), which, although useful, provided a poor depiction of anatomical structures. With the introduction of 3D imaging technologies like computed tomography (CT) and magnetic resonance imaging (MRI), medical image analysis has been revolutionized because it gives an opportunity to get a full picture of volumetric data, which results in more precise diagnoses and better comprehension of complicated pathologies [1].

The transition to 3D data, especially in the medical imaging field, presents immense challenges since it is a complex concept and more dimensional. Compared to 2D images, 3D medical images (e.g., CT, MRI, PET scans) are volumetric, so they are high-dimensional, more anatomically varied, complicated spatial relationships, and difficult to annotate [2]. Hundreds of slices in a single CT scan each have thousands of pixels, and the data points and computational costs increase exponentially. Moreover, there are vast differences in the anatomy of different people, and three-dimensional modeling makes it difficult to train the models. The analysis of 3D data needs to be done by comprehending intricate spatial association, which is a difficulty of conventional 2D methodologies, and manual annotation of 3D data is much more time-consuming and strenuous [3]. Deep learning (DL), in particular, 3D Convolutional Neural Networks, provides solutions with significant power due to the use of the 3D convolution kernel to uncover the spatial relationships in three dimensions and automatically identify the complex features [4]. Patches-based processing, special architectures, and weighted loss functions [5] are memory-efficient methods to use in tasks involving class imbalance and data augmentation, which are also effective in improving the capability of DL to process 3D data. Such developments open the possibilities of 3D medical data, which results in more precise diagnosis, individual treatment, and patient outcomes. One of the traditional approaches was the centralized data collection, where data across different sources were sent to one of the data centers, where a global model was trained [6–8]. This strategy, in its turn, is subjected to significant obstacles because of the strict data privacy policies (e.g., HIPAA) and increasing ethical issues regarding the dissemination of sensitive patient data.

*Corresponding author: Guru Ramachandra Nayaka, Department of Computer Science and Engineering, JSS Science and Technology University, India. Email: guruirg@sjce.ac.in

The weaknesses of centralized data collection have inspired efforts to introduce an alternative paradigm, one being federated learning (FL) [9]. The main differences between the centralized learning and the FL process are shown in Figure 1.

FL is a privacy-safe solution since it allows joint training of machine learning models, without actually sharing sensitive data. Models are trained locally in FL on the data of individual clients (e.g., hospitals), and model updates are only aggregated at a central server. This will reduce most of the privacy risks that are linked to centralized data collection. Nevertheless, there is a major challenge of the heterogeneity of medical data by nature. Medical images may differ significantly because of differences in imaging protocols (e.g., CT reconstruction parameters, MRI sequence parameters), type of scanner (e.g., different manufacturer, field strength), patient demographics (e.g., age, sex, ethnicity), and disease prevalence in different medical institutions. This heterogeneity in data has a major effect on the standard FL performance, like that of Federated Averaging (FedAvg), in which the data of the clients is assumed to be Independent and Identically Distributed (IID) [10]. FedAvg tends to cause model divergence, unstable training, and poor performance in non-IID applications, especially on the clients that are underrepresented in the training data.

The solution to this process is federated transfer learning (FTL), which utilizes the existing models to enhance learning in the intended federated environment. Knowledge transfer through a pretrained model can speed up convergence and enhance generalization, particularly in the small and or noisy local datasets. Nevertheless, it is a challenge to leverage transfer learning (TL) and FL to work together in a heterogeneous setting with 3D medical data. It is possible that standard FL algorithms are not always in the best position to deal with the nuanced interactions between TL and non-IID data. FedProx [11] is FedAvg with an additional proximal term to the local training objective, which criticizes local models that are too different than the global model.

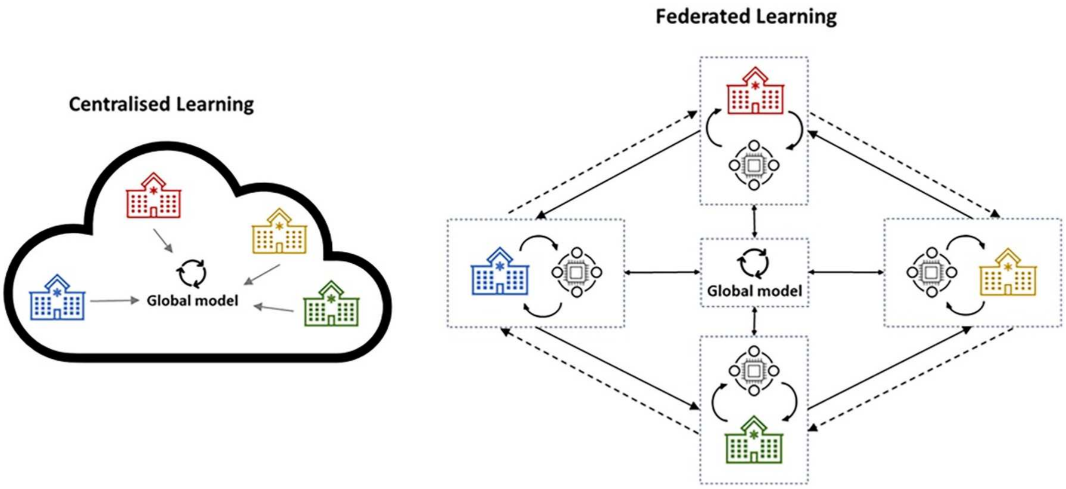
This methodology can be utilized to reduce the impact of data heterogeneity and enhance the stability and convergence of the training process. This paper advocates for the problem of data heterogeneity in FL for 3D medical image analysis by presenting a novel approach, referred to as FedProx TL. We claim that such a combination will enhance the performance, robustness, and fairness of collaboratively trained models, especially under realistic non-IID settings. The TL technique is adopted to overcome the raw heterogeneity of data, and the FedProx facilitates handling the disparate local training conditions in federated systems. The paper presents a thorough investigation of the effectiveness of FedProx with TL in 3D medical image analysis. This study fills a significant void in the literature related to how FL can be practically applied with challenging medical imaging data.

The key contributions of this study are as follows:

1) New architecture for non-IID 3D data: Released a new FL architecture that specifically uses TL and FedProx to solve the peculiarities of 3D medical imaging data.
2) Any data heterogeneity present in the data distribution was mitigated: Proximal regularization of FedProx and proximal regularization-based initialization enabled stable models to be well-trained on non-IID data distributions and enhanced inter-client consistency.
3) Scalability and practical feasibility: Illustrated the practical feasibility and scalability of the proposed strategy via a large scale of simulations with several federated clients with very little computational overhead that is worth being deployed in the real world.
4) Improved results and performance: Achieved higher performance metrics on all tested datasets, with accuracy improvements of up to 18.20% on OrganMNIST, 15.00% on NoduleMNIST, and 15.00% on FractureMNIST and significant improvements in precision, recall, F1-score, and training stability.

The rest of the paper will be organized in the following way. Section 2 is a thorough review of the relevant literature on FL, TL, and their use in medical imaging. Section 3 formulates the problem and outlines the intended methodology that incorporates TL and FedProx in the FL model. Section 4 describes the experimental setting, data description, and metrics of evaluation and makes a comprehensive commentary on the results obtained. Lastly, Section 5 is the conclusion of the paper as it summarizes the major findings of the paper and also talks about the possible future research directions.

Figure 1
Centralized learning vs federated learning



2. Related Works

In this section, the literature review on the existing work in the field of FL and its use in medicine is emphasized and the problems and gaps in research that are being resolved in this area.

2.1. Foundations of federated learning

FL has more recently been used in medical image analysis, where healthcare organizations seek to use distributed datasets without violating privacy laws, including HIPAA and GDPR. The initial background research by Nazir and Kaleem [12] proved that FL could be used to segment brain MRI data across 32 different institutions, with a 99% agreement with training centrality and data locality. The expansion of FL to other modalities followed, such as classifying chest X-rays, cardiac MRI, and low-dose PET denoising [13], making FL a framework that can be applied to develop collaborative models. Survey of the literature by Sharif et al. [14] is a systematic list of technical adaptations of the FL to medical imaging, which identified three continuing issues: (1) data distributions across institutions of varying types, (2) computational complexity of 3D medical volumes, and (3) lack of annotation in complex structures. Medical data is non-IID and, therefore, poses a major challenge to FL in 3D imaging and may significantly decrease the performance of standard algorithms such as FedAvg. Research has found that accuracy to deteriorate greater than 20% in activity with histopathology to PET denoising because of cross-site domain shifts. The proposed solutions include architectural modifications like model regularization (e.g., FedProx, a form of local update limitation) and TL architecture (e.g., FTL, which consolidates federated pretraining with local fine-tuning), which have demonstrated effectiveness in particular uses. There are also domain adaptation methods, such as client-specific decoders [15], which have been shown to be effective in preserving image quality in a variety of scanner conditions.

The existing constraints are fundamentally related to the fact that 2D-centric FL models have difficulties working with the specifics of 3D data. Patch-based processing, which is used by 68% of 3D FLs surveyed, breaks long-range spatial dependences required to analyze tumor invasion patterns [8]. The input patches smaller than 128 3 voxels used in 92% of methods are not sufficiently small in multi-organ interactions to image whole-body scans. Furthermore, most of the non-IID solutions are developed to work in single-modality settings, which cannot be generalized to multimodal federated settings where CT, MRI, and PET imaging data are available side by side [16]. The above gaps highlight a need to have native 3D FL architectures that can maintain the volumetric context but allow cross-modal knowledge transfer—which our FedProx–transfer integration explicitly targets by providing spatially sensitive regularization and hierarchical feature alignment.

2.2. Transfer learning in distributed medical imaging

TL has become an effective method of overcoming the issue of data scarcity through the utilization of knowledge that is already available in pretrained models. The models, including 3D ResNet-50, 3D UNet, and transformer-based models, are powerful feature extractors for different downstream tasks [17]. Nevertheless, TL integration in FL is a developing field of research with a lot of challenges, in terms of domain shifts, heterogeneity, and catastrophic forgetting. A number of studies have explored the workings of naive TL in FL. Yang et al. [18] showed that randomizing local models with ImageNet-trained weights resulted in a 14 percentage-point drop in segmentation Dice scores of federated brain MRI analysis as a result of domain discrepancy between natural and medical images. On the same note, naive model initialization was also found to provide nonuniform feature representation over clients, which restricts the ability of FL to generalize between institutions.

Researchers have tried adaptive TL methods to deal with problems of domain adaptation. Such federated pretraining methods as the ones suggested by Zhou et al. [19] have also demonstrated potential, aligning the distributions of features prior to fine-tuning on clients. Another effective way to reduce catastrophic forgetting is parameter freezing as Sharif et al. reported that accuracy in cross-site tumor classification was increased by 12% when federated feature extraction was coupled with centralized fine-tuning. In addition to the standard TL, new research focuses on modality-specific pretraining in order to improve feature representations in medical FL. A federated 3D hybrid network proposed by Liu et al. shared knowledge of 2D natural images to 3D CT volumes with anisotropic convolution, with a 40% reduced number of annotations. On the same note, the FedMRI framework used cross-site encoders to learn cross-site anatomical priors and kept institution-specific decoders to perform downstream reconstruction tasks. These works point to the opportunities of knowledge transfer on the hierarchical level but leave unanswered the questions about the best knowledge distillation methods to use in FL in 3D. In addition, few works have addressed cross-modal TL in FL. The multimodal FedMed-GAN [20] and FedFormer [21] have shown enhanced generalization with superior segmentation accuracy on multi-institutional MRI-CT datasets by a factor of up to 22% above single-modality models. Federated contrastive pretraining FedSSL [22] is also a contrastive self-supervised learning (SSL) method that has been studied, with significant enhancements of feature consistency across federated medical datasets achieved by federated contrastive pretraining. Such results indicate that self-supervised and contrastive pretraining may boost FTL by reducing the domain gaps.

There are also quite a number of research gaps that prevent a successful implementation of FL in 3D medical imaging. To begin with, the number of 3D-specific FL frameworks is extremely limited, and most of the literature is concerned with the analysis of 2D images even though volumetric modalities such as CT and MRI are widely used in clinical practice. Current 3D mappings are typically based on patch-based mappings that are computationally expensive to compute, and the mappings lose significant spatial information. Second, existing FL systems are likely to meet the heterogeneity in a vacuum, as they consider scanner-level variability or demographic characteristics of the population, instead of offering cohesive measures of simultaneous concurrent protocol, demographic, and pathological variations. Third, both TL and FL have proven to have their individual benefits, but their combined use is not yet developed. Initializing, updating, and stabilizing transferred features in non-IID federated environments requires a systematic methodology. Lastly, there are usually restricted evaluation protocols based on artificially split datasets of the population that do not reflect the realities of multi-institutional data, for example, shifts in the temporal distribution and inaccuracies in annotation. Such deficiencies underscore the importance of new architectures that are sensitive to 3D data properties and provide powerful knowledge exchange across heterogeneous federated networks. The ability to regularize models with volumetric complexity and non-IID distributions by combining principled model regularization with adaptive reuse of features is essential in taking the field of FL to the next level in medical imaging beyond incremental advances.

3. Problem Definition and Proposed Methodology

The use of artificial intelligence (AI) in 3D medical image classification, including the early identification of lung cancer and detection of brain aneurysms, is becoming an increasingly popular idea in hospitals around the world. Nevertheless, it is difficult to create powerful AI models because imaging data differ greatly. These differences are due to divergence in CT scan protocols (low-dose and contrast-enhanced scans), the scanner type (multi-detector and single-slice), the type of patients (smoking rates and nodules), and the software. This heterogeneity of data may lead to model generalization problems. Besides these issues, elaborate rules such as HIPAA and GDPR [23] restrict the dissemination of patient information, even in the anonymized state, complicating the process of centralized training of models. Due to that, effective methods that could be used in efficient data variability management and guarantee privacy are needed. This is crucial to supporting the joint AI production in the medical field. Consequently, FL has become a decentralized solution, which allows collaborative training among various hospitals without violating the privacy of data. Nonetheless, classical FL methods will assume IID data between clients, and this is not the case in a cross-silo medical scenario as datasets are highly non-IID. This heterogeneity of the data has a great influence on the convergence and generalizability of models, and the resultant models of the world are optimal. As a solution to this, our suggested method combines TL and FedProx with the FL framework to increase resilience and efficacy in cross-silo medical imaging conditions.

Let N represent the number of participating hospitals, each with a unique dataset D_i, where i = 1, 2,...,N. The set of all local datasets is denoted as:

$$D = \{D_1, \ D_2, \ \ldots, \ D_N\} \tag{1}$$

Each hospital trains a 3D image classification model with parameters ω_i in its local dataset D_i. The loss function of the local model on an input x with a ground truth label y is given by:

$$l(f_{\omega i}(x), y) \tag{2}$$

The objective is to minimize the global loss function in all hospitals.

$$\underset{\omega^g}{min} \sum\nolimits_{i=1}^{N} PiLi(\omega^g) \tag{3}$$

where $\mathrm{L}_i(\omega^g) = \mathrm{E}_{(x,y)\sim Di}[\ell(f_{\omega}g(x), y)]$ and p_i is the weight of the i-th hospital, typically proportional to the dataset size D_i.

Considering that the datasets at hospitals were non-IID, a straightforward aggregation of the local models into a global model ω^g may result in a lack of generalization. In order to solve this problem, every hospital starts by pretraining a local 3D medical imaging network in its facility. This method reduces the use of huge-labeled datasets and increases the speed of the convergence process. Also, FedProx is a variant of the FL aggregation method that adds a proximal term to the local optimization procedure. This balances the impact of the nonhomogeneous distribution of data and improves the stability of the model. Our proposed solution is developed by incorporating TL with FedProx, which improves the flexibility of the global model to different medical imaging scenarios, as well as enables efficient training under privacy-related limitations. The aim is to attain a strong and generalized 3D image classification model across various hospitals without necessarily sharing sensitive information about patients.

3.1. Federated Learning

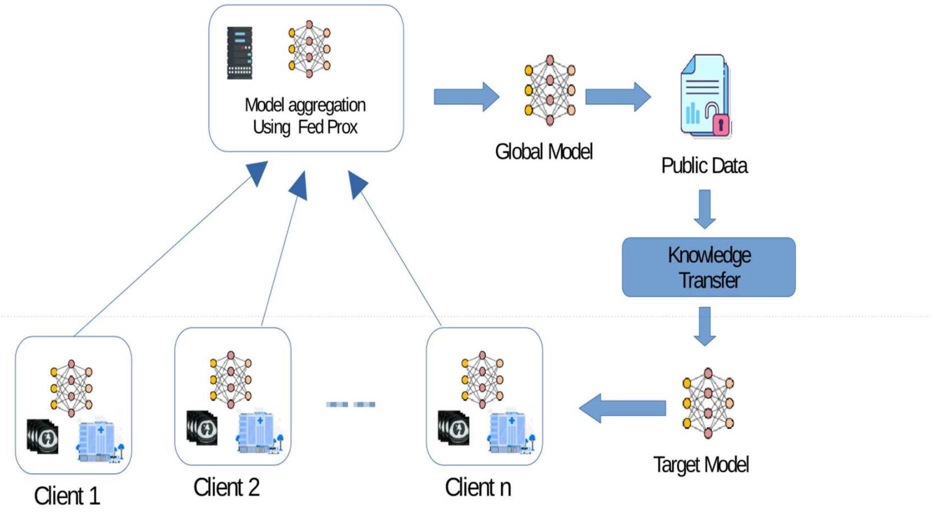
In FL, each hospital i trains a local model on its dataset D_i, and only model updates are shared with a central server for aggregation, ensuring data privacy. Let N represent the number of hospitals involved. Each hospital i has a local dataset.

$$D_i = \left\{(x_i^{(1)}, \ y_i^{(1)}\right), \ldots.., (x_i^{(mi)}, \ y_i^{(mi)})\} \tag{4}$$

where m_i is the number of samples in hospital i's dataset and x_i and y_i are the input image and label, respectively.

As shown in Figure 2, the proposed FL framework combines TL and FedProx to enhance model robustness and generalization across hospitals with heterogeneous 3D medical imaging data.

Figure 2
Proposed federated transfer learning framework with FedProx



The objective is to train a global model ω by aggregating the updates from all local models. The loss function for each local model is:

$$L_i(\omega_i) = \frac{1}{m_i}\sum\nolimits_{(x_i^{(j)},y_i^{(j)})\epsilon D_i} \ell\left(f_{\omega i}(x_i^{(j)}, y_i^{(j)} \right) \tag{5}$$

where $f_{\omega i}(x)$ is the local model's prediction and $\ell(\cdot)$ is the loss function (e.g., cross-entropy loss).

The global loss is defined as:

$$L_{global}(\omega) = \sum\nolimits_{i=1}^{N} p_i L_i(\omega) \tag{6}$$

where $p_i = \frac{m_i}{\sum_{i=1}^{N} m_i}$ is the weight assigned to each hospital based on the size of its local dataset.

3.2. Transfer Learning (TL)

TL allows leveraging a pretrained model ω_0 on a large dataset and fine-tuning it for the target task using a smaller dataset at each hospital. The pretrained model ω_0 is used to initialize each local model $\omega_i^{(0)}$:

$$\omega_i^{(0)} = \omega_0 \tag{7}$$

Each hospital then fine-tunes its local model using its own data. The fine-tuning loss function is:

$$L_i^{TL}(\omega_i) = \frac{1}{m_i}\sum\nolimits_{(x_i^{(j)},y_i^{(j)})\epsilon D_i} \ell\left(f_{\omega i}(x_i^{(j)}, y_i^{(j)} \right) \tag{8}$$

where ω_i is updated iteratively by minimizing this loss using gradient-based optimization. Typically, only the later layers of the network are fine-tuned, while the early layers remain frozen to preserve the knowledge learned from the pretrained model.

3.3. FedProx (Federated Proximal)

By adding a proximal term to the local loss function, FedProx improves on the conventional FedAvg algorithm. By penalizing differences between the local and global models, this addition lessens the impact of data distributions that are not IID.

The modified local loss function under FedProx is:

$$\begin{aligned} L_i^{FedProx}(\omega_i) &= \tfrac{1}{m_i}\sum\nolimits_{(x_i^{(j)},y_i^{(j)})\epsilon D_i} \ell\left(f_{\omega i}(x_i^{(j)}, y_i^{(j)} \right) \\ &+\mu\|\omega_i - \omega_0\|^2 \end{aligned} \tag{9}$$

where μ is a hyperparameter controlling the strength of the proximal term. This term ensures that the local model updates remain close to the global model, preventing large deviations that could arise due to the heterogeneity of local datasets. The global aggregation of local model updates is performed as follows:

$$\omega^g = \frac{1}{N}\sum\nolimits_{i=1}^{N} p_i \omega_i \tag{10}$$

The goal is to minimize this global loss while ensuring that local updates are stable and consistent with the global model.

3.4. Overall instructional strategy for federated learning using FedProx and transfer learning

As shown in Algorithm 1, the learning procedure in FTL, in collaboration with FedProx, seeks to leverage data from several hospitals while protecting data privacy and resolving the issues presented by non-IID data. Without requiring the exchange of raw data between various entities, this process is structured into iterative rounds to gradually improve a shared global model.

Pretraining, the initialization of a global model $\omega0$ using a pretrained model on a sizable, varied imaging dataset, is the first stage in the learning process. The global model ω_0 is used to initialize the local model ω_i for each hospital *i*. Since this initial model already encodes generalizable features learned from a wide range of data, it offers a solid foundation for local training. In the local training with TL phase, each hospital fine-tunes its local model ω_i on its private dataset D_i, which could consist of images of organs or diseases from its patients. The local loss function $\mathrm{L^{TL}}_i(\omega_i)$ is computed by comparing the model's predictions $f_\omega(x_i^{(j)})$ to the true labels $y_i^{(j)}$ of the local data. This fine-tuning step allows the model to specialize in downstream classification tasks, capturing local variations that may not be present in the global data.

After completing local training, each hospital applies FedProx regularization to its model to tackle issues such as model drift and the non-IID characteristics of data.

Algorithm 1:

FL with TL and FedProx

```
Input: Pretrained model $\omega_0$, $N$ hospitals, datasets $D_1$, $D_2$,...,$D_N$
Initialize global model $\omega_0$ using pretrained weights
for each communication round $t$ = 1, 2, ...,$T$ do
    for each hospital $i$ = 1, 2, ...,$N$ do
        Initialize local model $\omega_i^{(0)} = \omega_0$
        Fine-tune local model on dataset $D_i$
        Compute the local loss function:
            $L_i^{TL}(\omega_i) = \frac{1}{m_i}\sum_{(x_i^{(j)},y_i^{(j)})\epsilon D_i} \ell\left( f_{\omega i}(x_i^{(j)}, y_i^{(j)} \right)$
        Apply FedProx regularization:
            $L_i^{FedProx}(\omega_i) = \frac{1}{m_i}\sum_{(x_i^{(j)},y_i^{(j)})\epsilon D_i} \ell\left( f_{\omega i}(x_i^{(j)}, y_i^{(j)} \right) + \mu\|\omega_i - \omega_0\|^2$
        Update local model $\omega_i$ using gradient descent
    end for
    Aggregate local model updates:
        $\omega^g = \frac{1}{N}\sum_{i=1}^{N} p_i \omega_i$
end for
Output: Global model $\omega^g$
```

FedProx introduces a regularization term$\|\omega_i - \omega_2\|^2$ to the loss function, where ω_i is the updated local model and ω_0 is the global model from the previous round. This term discourages serious deviations around the global model so that the local models are near the global model but do still fit to their own local data. The global aggregation step is done in the central server and involves the aggregation of the updates received by all the hospitals into a new global model ω^g. The process of aggregation

Table 1
Summary of datasets used in the experiment

Dataset	Data modality	Num. of class	Num. of sample	Num. of clients
OrganMNIST (3D)	Abdominal CT	11	1742	10
NoduleMNIST (3D)	Chest CT	2	1633	10
FractureMNIST (3D)	Chest CT	3	1742	10

will provide a weighted average of the local models, with each hospital contribution being weighted based on its data volume or importance p_i. These aggregated updates are then used to build the global model by the central server so that the model makes use of the information provided by the local datasets and that the privacy of the users is not compromised.

The process is continued on and on. The local models are refined with every training round, and the global model is enhanced to make it better. This aims at reducing the global loss function Lglobal(0), which is a weighted average of local FedProx loss functions. This cyclic process continues to make the global model more correct and stronger to generalize more effectively with different medical data sources, that is, liver CT scans or brain MRIs, and deal with the non-IID data of other hospitals. FL combined with TL and FedProx can allow medical institutions to collaborate without involving raw patient data. It enables learning heterogeneous data, preserving privacy, and also has several benefits:

1) Data privacy: With FL, we will guarantee the safety of patient data in each hospital, but, at the same time, we will enable the global model to experience collaborative learning.
2) Non-IID data: FedProx is used to handle non-IID data by limiting local model updates to make the global model apply to the wider heterogeneous dataset more effectively.
3) Faster convergence: The TL process speeds up convergence, using pretrained models, which leads to less requirement of labeled data per hospital and a faster adaptation to the new environment.
4) Scalability and robustness: FL, TL, and FedProx can allow the model to be scaled to a variety of hospitals and endure the complications of heterogeneous data, which guarantees robust behavior on heterogeneous medical imaging data.

4. Experimental Setup

The experimental setup used to assess the efficacy of the suggested FedProx in conjunction with the TL framework is described in this section. We tested the flexibility and robustness of our method on a range of medical image modalities using several datasets from the MedMNIST 3D collection [24]. The following subsections provide detailed information on the experimental setup, dataset selection, computational resources used, and a detailed analysis of the observed results.

4.1. Dataset description

We used three distinct subsets from the MedMNIST3D collection to evaluate the efficacy of our proposed FedProx integrated with the TL approach. A thorough summary of the dataset is provided in Table 1. These datasets, which span a variety of medical imaging modalities and classification issues, offer a strong foundation for evaluating the adaptability and effectiveness of our FL methodology.

The first dataset, OrganMNIST3D, consists of abdominal CT scans used for an 11-class multi-class classification task. This subset's 1742 samples are split up into 610 testing, 161 validations, and 971 training samples. Because of the complexity of the abdominal organs and their intricate anatomical variability, OrganMNIST3D is a crucial benchmark for assessing our model's capacity to handle a range of anatomical structures. Using CT scans of the chest, the second dataset, NoduleMNIST3D, makes a binary classification between nodular and nonnodular structures. The 1633 samples in this dataset are divided into 1158 training samples, 165 validation samples, and 310 testing samples. We can evaluate our model's performance in the critical diagnostic task of identifying nodules in lung CT scans by testing it on NoduleMNIST3D. A 3-class multi-class classification task pertaining to bone fracture detection is addressed by the third dataset, FractureMNIST3D, which is also derived from chest CT scans. 1027 training samples, 103 validation samples, and 240 testing samples make up the 1370 samples in FractureMNIST3D. The robustness and generalization capabilities of our suggested method are severely tested by this dataset, given the difficulty of precisely identifying subtle fracture patterns.

We divided the OrganMNIST3D dataset into non-IID subsets across 10 fictitious hospitals in order to replicate a realistic FL scenario. By altering the class proportions across hospitals and modeling various hospital specializations and patient demographics, the non-IID distribution was accomplished. Together, these datasets cover a wide range of scenarios, including binary-class and multi-class problems, and successfully evaluate the flexibility and scalability of our FedProx-based TL methodology.

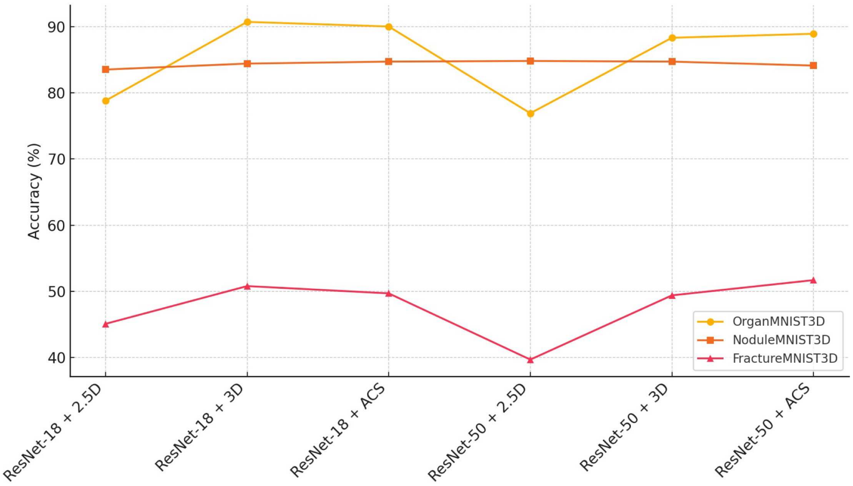
4.2. Implementation details

Python and the PyTorch DL library were used to implement our suggested FedProx with the TL framework. PyTorch was chosen because of its extensive support for intricate neural network architectures, strong flexibility, and simplicity of use. We used the official MedMNIST3D experimental setups and benchmarks[1] to efficiently assess different models.

Initially, we used the OrganMNIST3D, NoduleMNIST3D, and FractureMNIST3D datasets to evaluate several DL architectures, including various ResNet versions. To find the best architecture for our FL scenario, model performance comparisons were conducted using accuracy and precision–recall metrics. Comparative performance evaluations on reference datasets showed that ResNet-18 was the best choice due to its balanced performance and computational efficiency, as shown in Figure 3 [24]. To further enhance the performance of the pretrained MedicalNet model [25], which was specifically trained on a wide range of medical images, we applied TL techniques. By using MedicalNet's pretrained weights [26], we significantly accelerated training

[1] https://github.com/MedMNIST/experiments/tree/main/MedMNIST3D

Figure 3
Local training performance on all the datasets with different models



convergence and improved the model's robustness to variations in medical imaging data.

We distributed datasets among 10 clients, each of which represented a different medical institution, in order to simulate data decentralization for our FL environment [27]. In order to facilitate smooth integration within the FL process, each client's dataset was transformed into a federated dataset. Each communication round included five epochs for local training, following which each client model was updated and sent to a central server on a regular basis. An aggregation mechanism, created especially to efficiently manage the non-IID data distribution, was used at the server.

Eight Tesla 1-SXM2 GPUs, each with 32GB of RAM, 5.2TB of SSD storage, and a potent NVIDIA DGX Server Version 4.5.0 running GNU/Linux [28], were used for the experiments. With its extensive parallel processing capabilities, this computational infrastructure greatly reduced computation time while enabling the effective training of DL models on large-scale medical datasets. Over 100 communication rounds, the training was conducted [29]. To ensure robust model convergence across all client datasets, hyperparameters were adjusted for optimal convergence. These included a FedProx proximal term μ, a local batch size of 32, and a learning rate of $\eta = 0.01$.

4.3. Result analysis and discussion

Three different non-IID medical imaging datasets, namely, OrganMNIST, FractureMNIST, and NoduleMNIST, were tested using the FL methods, FedAvg, FedAvg+TL (with TL), FedProx, and the proposed method (FedProx with TL). Each experiment involved 10 federated clients, who were representing various hospitals, and exchanged more than 100 federated rounds. The non-IID distributions were simulated using the Dirichlet distribution, and therefore, realistic data heterogeneity among clients was modeled. Accuracy, precision, recall, F1-score, and convergence behavior metrics were used; they effectively operate on a wide range of data distributions, ensuring excellent model convergence on all client datasets.

The proposed method achieved excellent results with regard to the OrganMNIST dataset, with an accuracy of 83.17 (Figure 4). This outperformed FedProx (80.06%), FedAvg +TL (71.44%), and FedAvg (64.97%). Its accuracy (0.85), recall (0.84), and F1-score (0.84) also indicated that the model was reliable and consistent in all metrics of classification. These gains indicate that the heterogeneity of data can be significantly reduced when FedProx is used together with TL. The convergence curves showed that it attained higher accuracy levels in a faster and more dependable manner compared to other modes of doing the same, showing an increase in the stability of training and effectiveness during the rounds of federated communication.

Subsequently, the FractureMNIST dataset (Figure 5) was analyzed, where the complexity due to highly heterogeneous and challenging fracture patterns resulted in lower overall accuracies across all methods. Despite this, the proposed method showed unquestionable superiority over FedAvg (32.00%), FedAvg+TL (39.50%), and FedProx (42.0%), with an accuracy of 47.00%. Precision, recall, and F1-score metrics confirmed its resilience even in challenging data scenarios (0.49, 0.48, and 0.48, respectively). Importantly, the convergence behavior avoided the large performance swings observed with FedAvg and FedAvg+TL, demonstrating significantly improved stability and highlighting the method's versatility and resilience to a range of circumstances. Using the NoduleMNIST dataset (Figure 6), the proposed method was further validated and achieved an accuracy of 83.00%. This result significantly outperformed FedAvg (55.00%), FedAvg+TL (66.00%), and FedProx (73.00%). Metrics for precision, recall, and F1-score (0.85, 0.84, and 0.84, respectively) further demonstrated the method's overall capacity to handle non-IID data. The convergence trends [30], which demonstrated rapid improvement and stability over communication rounds, validated its strong inter-client consistency and robustness.

The comparative bar graph (Figure 7) and summary (Table 2) provided additional comparative insights, both of which demonstrated the suggested method's steady dominance across all datasets. The method's advantages in handling the heterogeneity common in federated medical imaging were amply demonstrated by the graphical representation. Due to disparities in the distribution of the data, FedAvg continuously performed the worst.

FedAvg+TL demonstrated a discernible improvement, confirming the gradual advantage of using TL alone. By specifically addressing client heterogeneity, FedProx added even more stability. Importantly, the suggested approach continuously produced better results by successfully utilizing the complementary advantages of both FedProx and TL. Scalability tests showed that

Figure 4
Non-IID OrganMNIST3D

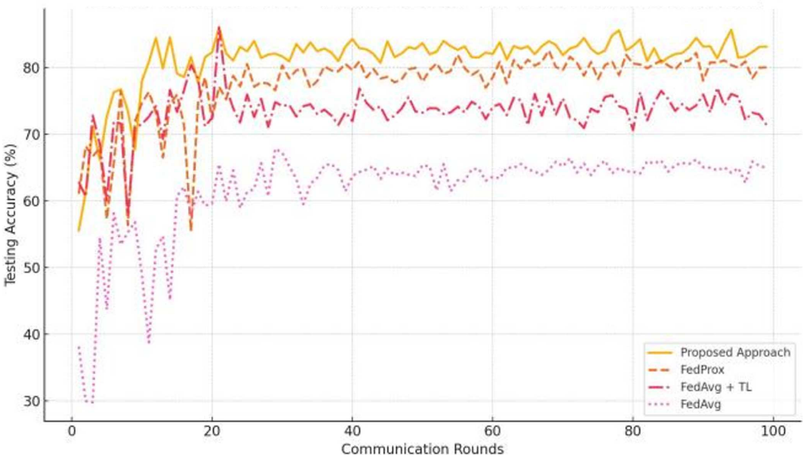


Figure 5
Non-IID FractureMNIST3D

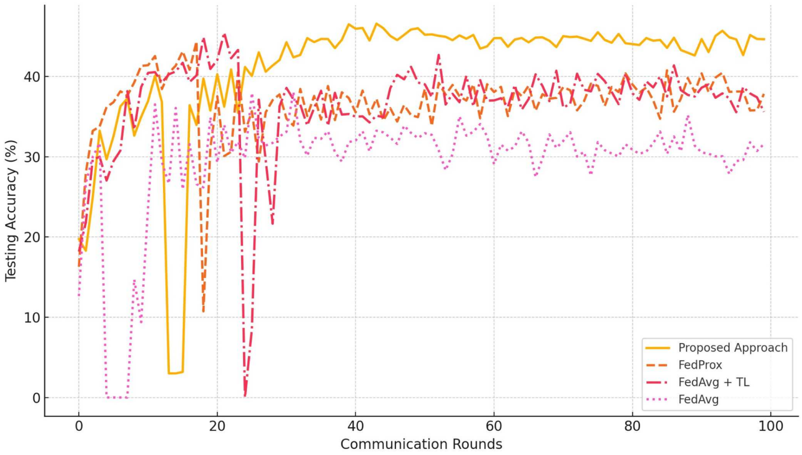


Figure 6
Non-IID NoduleMNIST3D

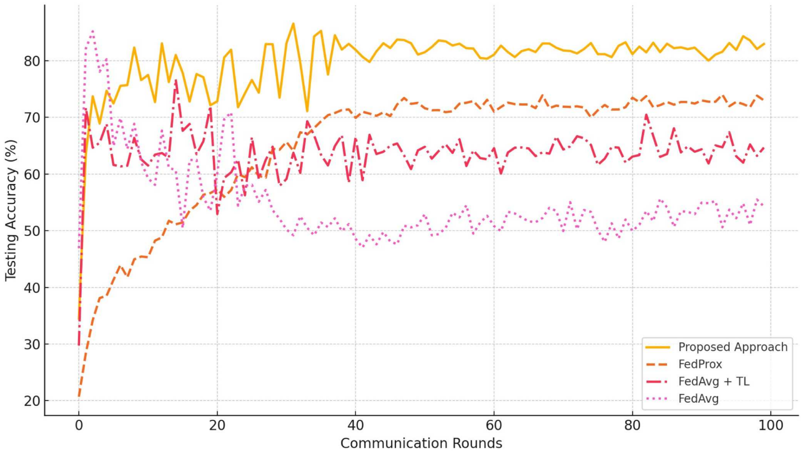


Figure 7
Overall comparison across different datasets and FL methods

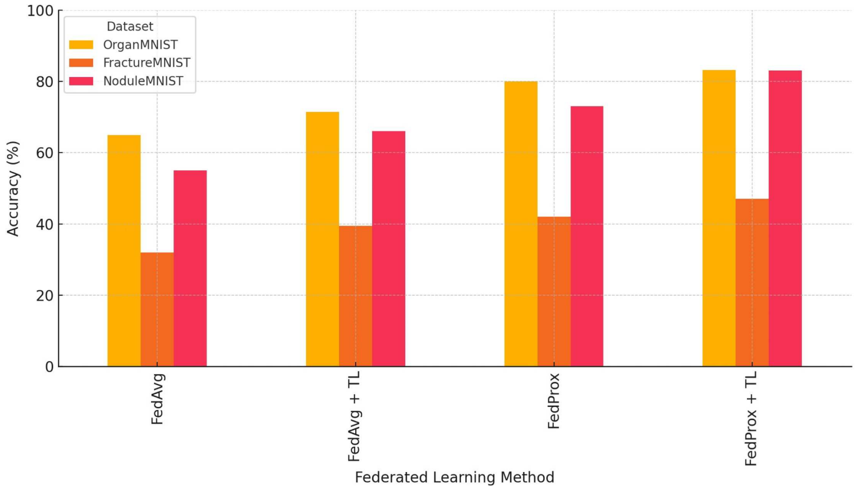


Table 2
Comparative performance evaluation across different FL methods on OrganMNIST, FractureMNIST, and NoduleMNIST (3D)

Dataset	Method	Accuracy (%)	Precision	Recall	F1-score
OrganMNIST	FedAvg	64.97	0.68	0.67	0.67
	FedAvg + TL	71.44	0.74	0.73	0.73
	FedProx	80.06	0.82	0.81	0.81
	Proposed	83.17	0.85	0.84	0.84
FractureMNIST	FedAvg	32.00	0.35	0.33	0.34
	FedAvg + TL	39.50	0.42	0.41	0.41
	FedProx	42.00	0.45	0.44	0.44
	Proposed	47.00	0.49	0.48	0.48
NoduleMNIST	FedAvg	55.00	0.59	0.58	0.58
	FedAvg + TL	66.00	0.68	0.67	0.67
	FedProx	73.00	0.76	0.75	0.75
	Proposed	83.00	0.85	0.84	0.84

the initialization of TL was largely responsible for the moderate communication costs and negligible additional computational overhead. Even as the number of participating hospitals increases, this finding demonstrates the method's practical viability for implementation in actual cross-silo federated healthcare environments. Overall, the integration of FedProx with TL offers significant promise for broader adoption in federated healthcare collaborations by effectively addressing the crucial issues of data heterogeneity, stability, and convergence efficiency in federated medical imaging applications.

5. Conclusion

Data heterogeneity and convergence instability were two significant issues in FL applied to non-IID 3D medical imaging that this study successfully addressed. Our suggested framework showed significant improvements by creatively combining FedProx for increased stability against statistical variance and TL for better local model adaptation. The method produced notable improvements in accuracy (up to 18.20%), precision, recall, and F1-score in comparison to baseline methods across a variety of 3D medical datasets (OrganMNIST, NoduleMNIST, and FractureMNIST), in addition to superior training stability. Most importantly, this was accomplished while maintaining the fundamental data privacy principles of FL. As a result, this study offers a workable, expandable, and clearly superior approach to cooperatively analyzing intricate medical images in diverse, real-world healthcare settings. To improve generalization and performance even more, future research could apply this strong methodology to multimodal data and look into sophisticated aggregation techniques. This framework will be expanded in our upcoming work to include multimodal medical data, which adds more levels of complexity and variability. The need for strong, privacy-preserving AI solutions in healthcare that can manage the growing complexity and diversity of medical data will be met by these extensions.

Acknowledgment

The authors are grateful to JSS Science and Technology University, Mysuru, Karnataka, for their support to carry out the research.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any of the authors.

Conflicts of Interest

The authors declare that there are no conflicts of interest to this work.

Data Availability Statement

The OrganMNIST, FractureMNIST, and NoduleMNIST datasets used in this study are part of the MedMNIST collection, a standardized benchmark suite of biomedical images for machine learning tasks. The data that support the findings of this study are openly available at https://medmnist.com/, https://github.com/MedMNIST/experiments/tree/main/MedMNIST3D, and https://github.com/TsingZ0/PFLlib/tree/master/dataset.

Author Contribution Statement

Manjunath Naganna: Conceptualization, Methodology, Software, Resources, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Guru Ramachandra Nayaka:** Formal analysis, Supervision. **Natesh Mahadev:** Validation, Investigation, Resources, Data curation. **Mayura Tapkire:** Conceptualization, Validation, Data curation, Writing – review & editing.

References

[1] Mienye, I. D., Swart, T. G., Obaido, G., Jordan, M., & Ilono, P. (2025). Deep convolutional neural networks in medical image analysis: A review. *Information*, *16*(3), 195. https://doi.org/10.3390/info16030195

[2] Sindhura, D. N., Pai, R. M., Bhat, S. N., & Pai, M. M. M. (2024). A review of deep learning and Generative Adversarial Networks applications in medical image analysis. *Multimedia Systems*, *30*(3), 161. https://doi.org/10.1007/s00530-024-01349-1

[3] Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, *18*(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

[4] Lepcha, D. C., Goyal, B., Dogra, A., Alkhayyat, A., Sahu, P. K., Ali, A., & Kukreja, V. (2025). Deep learning in medical image analysis: A comprehensive review of algorithms, trends, applications, and challenges. *Computer Modeling in Engineering & Sciences*, *145*(2), 1487–1573. https://doi.org/10.32604/cmes.2025.070964

[5] Yeung, M., Sala, E., Schönlieb, C.-B., & Rundo, L. (2022). Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Computerized Medical Imaging and Graphics*, *95*, 102026. https://doi.org/10.1016/j.compmedimag.2021.102026

[6] Chen, C., Mat Isa, N. A., & Liu, X. (2025). A review of convolutional neural network based methods for medical image classification. *Computers in Biology and Medicine*, *185*, 109507. https://doi.org/10.1016/j.compbiomed.2024.109507

[7] Ehab, W., Huang, L., & Li, Y. (2024). UNet and variants for medical image segmentation. *International Journal of Network Dynamics and Intelligence*, *3*(2), 100009. https://doi.org/10.53941/ijndi.2024.100009

[8] Teo, Z. L., Jin, L., Liu, N., Li, S., Miao, D., Zhang, X., . . . , & Ting, D. S. W. (2024). Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. *Cell Reports Medicine*, *5*(2), 101419. https://doi.org/10.1016/j.xcrm.2024.101419

[9] Papadopoulos, C., Kollias, K.-F., & Fragulis, G. F. (2024). Recent advancements in federated learning: State of the art, fundamentals, principles, IoT applications and future trends. *Future Internet*, *16*(11), 415. https://doi.org/10.3390/fi16110415

[10] Zhu, H., Xu, J., Liu, S., & Jin, Y. (2021). Federated learning on non-IID data: A survey. *Neurocomputing*, *465*, 371–390. https://doi.org/10.1016/j.neucom.2021.07.098

[11] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems*, *2*, 429–450.

[12] Nazir, S., & Kaleem, M. (2023). Federated learning for medical image analysis with deep neural networks. *Diagnostics*, *13*(9), 1532. https://doi.org/10.3390/diagnostics13091532

[13] Khan, R., Taj, S., Ma, X., Noor, A., Zhu, H., Khan, J., . . . , & Khan, S. U. (2024). Advanced federated ensemble internet of learning approach for cloud based medical healthcare monitoring system. *Scientific Reports*, *14*(1), 26068. https://doi.org/10.1038/s41598-024-77196-x

[14] Sharif, M. I., Mehmood, M., Uddin, M. P., Siddique, K., Akhtar, Z., & Waheed, S. (2024). Federated learning for analysis of medical images: A survey. *Journal of Computer Science*, *20*(12), 1610–1621. https://doi.org/10.3844/jcssp.2024.1610.1621

[15] Rehman, M. H. U., Hugo Lopez Pinaya, W., Nachev, P., Teo, J. T., Ourselin, S., & Cardoso, M. J. (2023). Federated learning for medical imaging radiology. *The British Journal of Radiology*, *96*(1150), 20220890. https://doi.org/10.1259/bjr.20220890

[16] Alekseenko, I., Karargyris, A., & Padoy, N. (2024). Distance-aware non-IID federated learning for generalization and personalization in medical imaging segmentation. In *Proceedings of the 7th International Conference on Medical Imaging with Deep Learning*, *250*, 33–47.

[17] Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., van Ginneken, B., Madabhushi, A., . . . , & Summers, R. M. (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, *109*(5), 820–838. https://doi.org/10.1109/JPROC.2021.3054390

[18] Yang, T., Yu, X., McKeown, M. J., & Wang, Z. J. (2024). When federated learning meets medical image analysis: A systematic review with challenges and solutions. *APSIPA Transactions on Signal and Information Processing*, *13*(1), 1–55. https://doi.org/10.1561/116.20240048

[19] Zhou, B., Miao, T., Mirian, N., Chen, X., Xie, H., Feng, Z., . . . , & Liu, C. (2023). Federated transfer learning for low-dose pet denoising: A pilot study with simulated heterogeneous data. *IEEE Transactions on Radiation and Plasma Medical Sciences*, *7*(3), 284–295. https://doi.org/10.1109/TRPMS.2022.3194408

[20] Wang, J., Xie, G., Huang, Y., Lyu, J., Zheng, F., Zheng, Y., & Jin, Y. (2023). FedMedGAN: Federated domain translation on unsupervised cross-modality brain image synthesis.

Neurocomputing, *546*, 126282. https://doi.org/10.1016/j.neucom.2023.126282

[21] Zhou, T., Ma, Z., Wen, Q., Wang, X., Sun, L., & Jin, R. (2022). FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *Proceedings of the 39th International Conference on Machine Learning*, *162*, 27268–27286.

[22] Tun, Y. L., Thwal, C. M., Le, H. Q., Nguyen, M. N. H., Huh, E.-N., & Hong, C. S. (2025). *Resource-efficient layer-wise federated self-supervised learning*. *arXiv Preprint: 2401.11647*

[23] Tovino, S. A. (2017). The HIPAA privacy rule and the EU GDPR: Illustrative comparisons. *Seton Hall Law Review*, *47*(4), 973–993.

[24] Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., . . . , & Pfister, H. (2023). MedMNIST v2—A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*, *10*(1), 41. https://doi.org/10.1038/s41597-022-01721-8

[25] Chen, S., Ma, K., & Zheng, Y. (2019). Med3D: Transfer learning for 3D medical image analysis. *arXiv Preprint: 1904.00625*

[26] Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. *Frontiers in Public Health*, *11*, 1273253. https://doi.org/10.3389/fpubh.2023.1273253

[27] Guan, H., Yap, P.-T., Bozoki, A., & Liu, M. (2024). Federated learning for medical image analysis: A survey. *Pattern Recognition*, *151*, 110424. https://doi.org/10.1016/j.patcog.2024.110424

[28] Kim, H. E., Cosa-Linan, A., Santhanam, N., Jannesari, M., Maros, M. E., & Ganslandt, T. (2022). Transfer learning for medical image classification: A literature review. *BMC Medical Imaging*, *22*(1), 69. https://doi.org/10.1186/s12880-022-00793-7

[29] Tapkire, M. D., & Arun, V. (2023). Application of artificial intelligence to correlate food formulations to disease risk prediction: A comprehensive review. *Journal of Food Science and Technology*, *60*(9), 2350–2357. https://doi.org/10.1007/s13197-022-05550-w

[30] Tapkire, M., Arun, V., Lavanya, M. S., & Shashidhar, R. (2025). Gluten identification from food images using advanced deep learning and transfer learning methods. *Journal of Food Science and Technology*, *62*(6), 1164–1172. https://doi.org/10.1007/s13197-024-06158-y

How to Cite: Naganna, M., Nayaka, G. R., Mahadev, N., & Tapkire, M. (2026). FedProx-Enhanced Federated Transfer Learning for Heterogeneous 3D Medical Image Classification. *Journal of Computational and Cognitive Engineering*. https://doi.org/10.47852/bonviewJCCE62027069