2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

MedFedSSL: Semi-Supervised Federated Learning with Dual Consistency and Adaptive Distillation for Medical Imaging

Chenhao Ye

School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing, China
ychbert@outlook.com

Abstract. Medical image analysis presents unique challenges due to limited labeled data, privacy concerns, and heterogeneous data distributions across institutions. Federated learning (FL) offers a promising solution by enabling collaborative model training without sharing raw data. However, existing FL approaches often struggle with label scarcity in medical domains. In this paper, we propose MedFedSSL, a novel semi-supervised federated learning framework specifically designed for medical image analysis. Our approach integrates a dual-consistency regularization mechanism with an adaptive knowledge distillation strategy to effectively leverage both labeled and unlabeled data across distributed clients. We introduce a theoretically sound optimization objective that addresses the challenges of data heterogeneity and label imbalance in medical imaging. Extensive experiments on multiple medical imaging datasets demonstrate that MedFedSSL significantly outperforms state-of-the-art federated learning and semi-supervised learning methods, achieving superior performance with limited labeled data while preserving privacy. Our theoretical analysis provides convergence guarantees and bounds on the generalization error of the proposed approach.
Keywords: Federated Learning, Semi-Supervised Learning, Medical Image Analysis, Deep Learning

Introduction

Medical image analysis has witnessed remarkable progress with the advent of deep learning techniques. However, developing robust and generalizable models for clinical applications faces several critical challenges. First, acquiring large-scale labeled medical data is expensive and time-consuming, requiring expert annotations from medical professionals. Second, medical data is inherently private and sensitive, making centralized data collection problematic due to regulatory constraints such as HIPAA and GDPR. Third, medical data often exhibits significant heterogeneity across different institutions, devices, and patient populations, leading to distribution shifts that affect model performance.

Federated Learning (FL) has emerged as a promising paradigm to address privacy concerns by enabling collaborative model training without sharing raw data [15,12]. In

FL, multiple clients (e.g., hospitals or medical institutions) train models locally and only share model updates with a central server, which aggregates these updates to improve a global model. Since its introduction by McMahan et al. [15], numerous works have addressed various challenges in FL, including communication efficiency [8,18], statistical heterogeneity [12,26], and system heterogeneity [17,25]. In the medical domain, FL has gained significant attention due to its privacy-preserving nature [19,21], with successful applications in brain tumor segmentation [21] and medical image classification [13]. However, while FL preserves data privacy, it does not directly address the challenge of limited labeled data.

Semi-Supervised Learning (SSL) techniques aim to leverage unlabeled data alongside limited labeled data to improve model performance [24,22]. Traditional SSL approaches include self-training [31], co-training [4], and graph-based methods [33], while recent deep learning-based methods have shown remarkable success with consistency regularization [20,9], pseudo-labeling [10,22], and hybrid approaches [2,3]. In medical imaging, SSL has been applied to various tasks including segmentation [1,32], classification [5,23], and anomaly detection [6]. However, most existing SSL methods are designed for centralized settings and do not account for the distributed nature and heterogeneity of medical data.

Recent works have begun exploring the integration of SSL with FL [7,28], with approaches like FedMatch [7] extending consistency regularization to federated settings, FedSem [28] leveraging unlabeled data through pseudo-labeling, and ensemble distillation [14] for semi-supervised FL. In the medical domain, limited work has been done on semi-supervised FL, with some efforts focusing on COVID-19 detection [29] and medical image segmentation [11]. However, these approaches often struggle with the unique challenges posed by medical imaging data, such as extreme class imbalance, high-dimensional feature spaces, and complex anatomical structures. Additionally, they typically rely on strong assumptions about data distributions across clients, which may not hold in real-world medical scenarios.

In this paper, we propose MedFedSSL, a novel semi-supervised federated learning framework specifically designed for medical image analysis. Our approach differs from existing methods by introducing a dual-consistency regularization mechanism and an adaptive knowledge distillation strategy specifically tailored for medical imaging in federated settings. Our approach makes the following key contributions:

- We introduce a dual-consistency regularization mechanism that enforces consistency between predictions from different augmentations of the same image and between predictions from the client and global models, effectively leveraging unlabeled data in a federated setting.
- We develop an adaptive knowledge distillation strategy that accounts for the reliability of pseudo-labels generated for unlabeled data, dynamically adjusting the contribution of each unlabeled sample based on prediction confidence and consistency.
- We conduct extensive experiments on multiple medical imaging datasets, demonstrating that MedFedSSL significantly outperforms state-of-the-art federated learning and semi-supervised learning methods, achieving superior performance with limited labeled data while preserving privacy.

Methodology

2.1 Problem Formulation

We consider a federated learning setting with K clients, where each client $k \in \{1,2,\ldots,K\}$ has a local dataset $\mathcal{D}_k = \mathcal{D}_k^l \cup \mathcal{D}_k^u$ consisting of labeled data $\mathcal{D}_k^l = \{(x_i^k, y_i^k)\}_{i=1}^{n_k^l}$ and unlabeled data $\mathcal{D}_k^u = \{x_j^k\}_{j=1}^{n_k^u}$. Here, $x_i^k \in \mathcal{X}$ represents an input medical image, $y_i^k \in \mathcal{Y}$ represents the corresponding label (e.g., disease classification or segmentation mask), n_k^l is the number of labeled samples, and n_k^u is the number of unlabeled samples at client k.

The goal is to learn a global model $f_\theta: \mathcal{X} \to \mathcal{Y}$ parameterized by θ that performs well on test data from all clients, while keeping the raw data local to each client. The challenge is to effectively leverage both labeled and unlabeled data across all clients to improve model performance, especially when labeled data is scarce.

2.2 MedFedSSL Framework

Our proposed MedFedSSL framework addresses these challenges through a novel combination of dual-consistency regularization, adaptive knowledge distillation, and a theoretically sound optimization objective. The overall framework is illustrated in Fig. 1 and detailed in Algorithm 1.

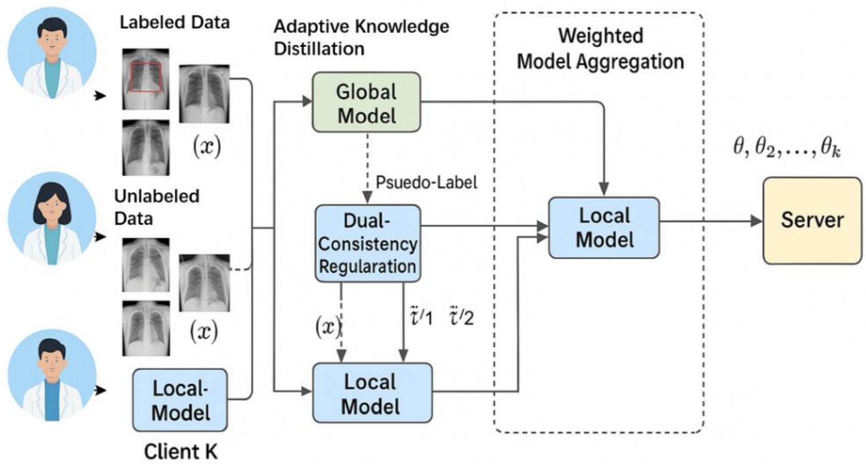

Fig. 1. Overview of the proposed MedFedSSL framework.

Dual-Consistency Regularization. To effectively leverage unlabeled data, we propose a dual-consistency regularization mechanism that enforces consistency between predictions from different augmentations of the same image and between predictions from the client and global models.

Let $\mathcal{A}$ denote a stochastic augmentation function that applies a composition of transformations (e.g., rotation, scaling, and intensity adjustments) to an input image. For each unlabeled sample $x_j^k \in D_k^u$, we generate two augmented views $\tilde{x}_{j,1}^k = \mathcal{A}(x_j^k)$ and $\tilde{x}_{j,2}^k = \mathcal{A}(x_j^k)$.

The first consistency term enforces that the model's predictions for these two augmented views should be similar:

$$\mathcal{L}_{\text{aug}}\left(\theta_k, x_j^k\right) = d\left(f_{\theta_k}\left(\tilde{x}_{j,1}^k\right), f_{\theta_k}\left(\tilde{x}_{j,2}^k\right)\right) \tag{1}$$

where θ_k represents the parameters of the local model at client k, and $d(\cdot,\cdot)$ is a distance function measuring the dissimilarity between two predictions. For classification tasks, we use the Kullback-Leibler (KL) divergence, while for segmentation tasks, we use a combination of Dice loss and cross-entropy loss.

The second consistency term enforces that the predictions of the local model should be consistent with those of the global model:

$$\mathcal{L}_{\text{model}}\left(\theta_k, \theta_g, x_j^k\right) = d\left(f_{\theta_k}\left(x_j^k\right), f_{\theta_g}\left(x_j^k\right)\right) \tag{2}$$

where θ_g represents the parameters of the global model. This term helps mitigate the impact of local optima and encourages knowledge transfer between clients.

Adaptive Knowledge Distillation. To further improve the utilization of unlabeled data, we propose an adaptive knowledge distillation strategy that accounts for the reliability of pseudo-labels generated for unlabeled data.

For each unlabeled sample $x_j^k \in D_k^u$, we generate a pseudo-label $\hat{y}_j^k$ using the global model:

$$\hat{y}_j^k = \begin{cases} argmax_c f_{\theta_g}\left(x_j^k\right)_c & if \;\; max_c f_{\theta_g}\left(x_j^k\right)_c > \tau \\ undefined & otherwise \end{cases} \tag{3}$$

where $f_{\theta_g}\left(x_j^k\right)_c$ represents the probability assigned to class c for input x_j^k by the global model, and τ is a confidence threshold.

We then define an adaptive weight w_j^k for each pseudo-labeled sample based on the prediction confidence and consistency:

$$w_j^k = \max_c f_{\theta_g}(x_j^k)_c \cdot \exp\left(-\beta \cdot \mathcal{L}_{\text{aug}}\left(\theta_g, x_j^k\right)\right) \tag{4}$$

where β is a hyperparameter controlling the importance of consistency. This weight reflects the reliability of the pseudo-label, with higher weights assigned to samples with high confidence and high consistency.

The knowledge distillation loss is then defined as:

$$L_{kd}(\theta_k, \theta_g, x_j^k) = w_j^k \cdot \ell(f_{\theta_k}(x_j^k), \hat{y}_j^k \tag{4}$$

where $\ell(\cdot,\cdot)$ is the task-specific loss function (e.g., cross-entropy for classification or Dice loss for segmentation).

Local Optimization Objective. The overall local optimization objective for client k is a weighted combination of the supervised loss on labeled data and the unsupervised losses on unlabeled data:

$$\begin{aligned} L_k(\theta_k, \theta_g) = \frac{1}{n_k^l}\sum_{i=1}^{n_k^l} \ell\left(f_{\theta_k}(x_i^k), y_i^k\right) + \lambda_1 \frac{1}{n_k^u}\sum_{j=1}^{n_k^u} L_{aug}\left(\theta_k, x_j^k\right) \\ +\lambda_2 \frac{1}{n_k^u}\sum_{j=1}^{n_k^u} L_{model}\left(\theta_k, \theta_g, x_j^k\right) + \lambda_3 \frac{1}{n_k^u}\sum_{j=1}^{n_k^u} L_{kd}\left(\theta_k, \theta_g, x_j^k\right) \end{aligned} \tag{6}$$

where λ_1, λ_2, and λ_3 are hyperparameters controlling the importance of each unsupervised loss term.

To address the challenge of data heterogeneity, we introduce a proximal term that penalizes large deviations from the global model:

$$\mathcal{L}_k^{prox}(\theta_k, \theta_g) = \mathcal{L}_k(\theta_k, \theta_g) + \frac{\mu}{2} \parallel \theta_k - \theta_g \parallel^2 \tag{7}$$

where μ is a hyperparameter controlling the strength of the proximal term.

Weighted Model Aggregation. To address the challenge of label imbalance and varying data quality across clients, we propose a weighted model aggregation strategy at the server:

$$\theta_g = \sum_{k=1}^{K} \alpha_k \theta_k \tag{8}$$

where α_k is the aggregation weight for client k. Instead of using a fixed weighting scheme based on the number of samples (as in FedAvg), we define α_k based on the validation performance of the local model:

$$\alpha_k = \frac{\exp(\gamma \cdot \mathrm{Perf}_k)}{\sum_{k'=1}^{K} \exp(\gamma \cdot \mathrm{Perf}_{k'})} \tag{9}$$

where Perf_k is a performance metric (e.g., accuracy or Dice coefficient) evaluated on a small validation set at client k, and γ is a hyperparameter controlling the concentration of weights. This weighting scheme assigns higher weights to clients with better-performing models, which helps mitigate the impact of low-quality data or extreme class imbalance at certain clients.

Algorithm 1 MedFedSSLAlgorithm

Require: Rounds T, local epochs E, learning rate η, hyperparameters $\lambda_1, \lambda_2, \lambda_3, \mu, \gamma, \tau, \beta$
Ensure: Global model θ_g

Initialize: θ_g^0
for t = 1 to $T-1$ **do**
  Server distributes θ_g^t go all clients
  for each client $k \in \{1,2,\dots,K\}$ in parallel **do**
    Set $\theta_k^t \leftarrow \theta_g^t$
    for $e\ = 1$ to E **do**
      Sample batches $\mathcal{B}_k^l$ from labeled data, $\mathcal{B}_k^u$ from unlabeled data
      Compute supervised loss: $\mathcal{L}_{\text{sup}} = \frac{1}{|\mathcal{B}_k^l|}\sum_{(x,y)} \ell(f_{\theta_k^t}(x), y)$
      Compute consistency losses:
      $\mathcal{L}_{\text{aug}} = \frac{1}{|\mathcal{B}_k^u|}\sum_x d(f_{\theta_k^t}(\mathcal{A}(x)), f_{\theta_k^t}(\mathcal{A}(x)))$
      $\mathcal{L}_{\text{model}} = \frac{1}{|\mathcal{B}_k^u|}\sum_x d(f_{\theta_k^t}(x), f_{\theta_g^t}(x))$
      Generate pseudo-labels for unlabeled data
      for each $x \in \mathcal{B}_k^u$ **do**
        If $max_c f_{\theta_g}\left(x_j^k\right)_c > \tau$ **then**
          $\hat{y} = \arg\max_c f_{\theta_g^t}(x)_c, w = \max_c f_{\theta_g^t}(x)_c \cdot \exp\left(-\beta \cdot d(f_{\theta_g^t}(\mathcal{A}(x)), f_{\theta_g^t}(\mathcal{A}(x))\right)$
        else
          $w = 0$
        end if
      end for
      Compute knowledge distillation loss: $\mathcal{L}_{\text{kd}} = \frac{1}{|\mathcal{B}_k^u|}\sum_x w \cdot \ell(f_{\theta_k^t}(x), \hat{y})$
      Compute proximal term $L_{prox} = \frac{\mu}{2} \| \theta_k^t - \theta_g^t \|^2$
      Total loss: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{sup}} + \lambda_1 \mathcal{L}_{\text{aug}} + \lambda_2 \mathcal{L}_{\text{model}} + \lambda_3 \mathcal{L}_{\text{kd}} + \mathcal{L}_{\text{prox}}$
      Update: $\theta_k^t \leftarrow \theta_k^t - \eta \nabla \mathcal{L}_{\text{total}}$
    end for
    Evaluate on validation set to get Perf_k
    Send $\theta_k^{t+1} = \theta_k^t$ and Perf_k to server
  end for
  Server computes weights: $\alpha_k = \frac{\exp(\gamma \cdot \text{Perf}_k)}{\sum_{k'} \exp(\gamma \cdot \text{Perf}_{k'})}$
  Server aggregates: $\theta_g^{t+1} = \sum_{k=1}^K \alpha_k \theta_k^{t+1}$
end for
return θ_g^T

Experiments

In this section, we evaluate the performance of our proposed MedFedSSL framework on multiple medical imaging datasets and compare it with state-of-the-art federated learning and semi-supervised learning methods.

3.1 Experimental Setup

Datasets. We evaluate our approach on three medical imaging datasets with different characteristics:

- **MedMNIST**[30]**:** A collection of 10 pre-processed medical image datasets, including chest X-rays, dermatoscope images, OCT images, and more. We use the PathMNIST, DermaMNIST, and OrganCMNIST subsets for multi-class classification tasks.
- **ChestX-ray14**[27]**:** A large dataset of chest X-ray images with 14 disease labels. We focus on the multi-label classification of 5 common diseases: Atelectasis, Cardiomegaly, Effusion, Infiltration, and Pneumonia.
- **BraTS**[16]**:** A brain tumor segmentation dataset containing multi-modal MRI scans (T1, T1ce, T2, and FLAIR) with pixel-level annotations of different tumor regions (whole tumor, tumor core, and enhancing tumor).

For each dataset, we simulate a federated learning environment with $K = 10$ clients. To create a realistic non-IID setting, we distribute the data among clients using a Dirichlet distribution with concentration parameter $\alpha = 0.5$, resulting in heterogeneous class distributions across clients. For each client, we randomly select a small portion (10%, 20%, or 30%) of the data as labeled and treat the rest as unlabeled.

Baselines. We compare our MedFedSSL framework with the following baselines:

- **FedAvg** [15]: The standard federated averaging algorithm using only labeled data.
- **FedProx** [12]: A federated learning algorithm that adds a proximal term to address client heterogeneity, using only labeled data.
- **FedMatch** [7]: A semi-supervised federated learning approach that extends consistency regularization to federated settings.
- **FedSem** [28]: A semi-supervised federated learning framework that leverages unlabeled data through pseudo-labeling.
- **Local SSL + FedAvg**: A simple baseline where each client independently applies a semi-supervised learning method (FixMatch [22]) on its local data, followed by federated averaging.
- **Centralized SSL**: A centralized semi-supervised learning approach (FixMatch [22]) with access to all data, serving as an upper bound for performance.

Implementation Details. For classification tasks (MedMNIST and ChestX-ray14), we use a ResNet-18 architecture pre-trained on ImageNet. For segmentation tasks (BraTS), we use a 3D U-Net architecture. All models are implemented in PyTorch and trained using the Adam optimizer with a learning rate of 10^{-4}.

For our MedFedSSL framework, we set the hyperparameters as follows: $\lambda_1 = 1.0$, $\lambda_2 = 0.5$, $\lambda_3 = 1.0$, $\mu = 0.01$, $\gamma = 5.0$, $\tau = 0.95$, and $\beta = 0.5$. We run the federated learning process for $T = 100$ communication rounds, with $E = 5$ local epochs per round.

For data augmentation, we use random cropping, horizontal flipping, and color jittering for 2D images, and random cropping and intensity shifting for 3D volumes. All experiments are repeated three times with different random seeds, and we report the mean and standard deviation of the results.

3.2 Results and Analysis

Classification Performance. Table 1 shows the classification performance of different methods on the MedMNIST and ChestX-ray14 datasets with 10% labeled data. For MedMNIST, we report the accuracy, and for ChestX-ray14, we report the AUC-ROC (Area Under the Receiver Operating Characteristic Curve).

Table 1. Classification performance on MedMNIST and ChestX-ray14 datasets with 10% labeled data. The best results among federated methods are highlighted in **bold**.

Method	MedMNIST (Accuracy %)			Chest-Xray14
	PathMNIST	DermaMNIST	OrganCMNIS	(AUC-ROC %)
FedAvg	76.2 ± 1.8	68.5 ± 2.1	72.3 ± 1.5	71.4 ± 1.9
FedProx	78.1 ± 1.5	70.2 ± 1.8	74.5 ± 1.3	73.2 ± 1.7
FedMatch	82.3 ± 1.2	74.8 ± 1.5	78.9 ± 1.1	76.5 ± 1.4
FedSem	83.1 ± 1.0	75.6 ± 1.3	79.5 ± 0.9	77.2 ± 1.2
Local SSL + FedAvg	81.5 ± 1.3	73.9 ± 1.6	77.8 ± 1.2	75.8 ± 1.5
MedFedSSL (Ours)	**85.7 ± 0.8**	**78.2 ± 1.1**	**82.4 ± 0.7**	**79.6 ± 1.0**
Centralized SSL	88.3± 0.5	81.5 ± 0.7	85.1± 0.4	82.9± 0.6

Our MedFedSSL framework consistently outperforms all baseline methods across all datasets. Compared to FedAvg, which uses only labeled data, MedFedSSL achieves significant improvements of 9.5%, 9.7%, and 10.1% on PathMNIST, DermaMNIST, and OrganCMNIST, respectively, and 8.2% on ChestX-ray14. This demonstrates the effectiveness of our approach in leveraging unlabeled data to improve model performance.

Compared to other semi-supervised federated learning methods (FedMatch and FedSem), MedFedSSL still achieves notable improvements of 2.6-3.4% on MedMNIST and 2.4-3.1% on ChestX-ray14. This highlights the benefits of our dual-consistency regularization and adaptive knowledge distillation strategies, which are specifically designed for medical image analysis.

Interestingly, the performance gap between MedFedSSL and Centralized SSL is relatively small (2.6-3.3% on MedMNIST and 3.3% on ChestX-ray14), indicating that our approach can effectively leverage distributed data without requiring centralized access.

Fig.2 shows the performance of different methods on PathMNIST with varying percentages of labeled data (10%, 20%, and 30%). As expected, all methods benefit from more labeled data, but MedFedSSL consistently outperforms the baselines across all settings. The performance gap is more pronounced when the labeled data is scarce (10%), highlighting the effectiveness of our approach in low-label regimes, which is particularly relevant for medical applications.

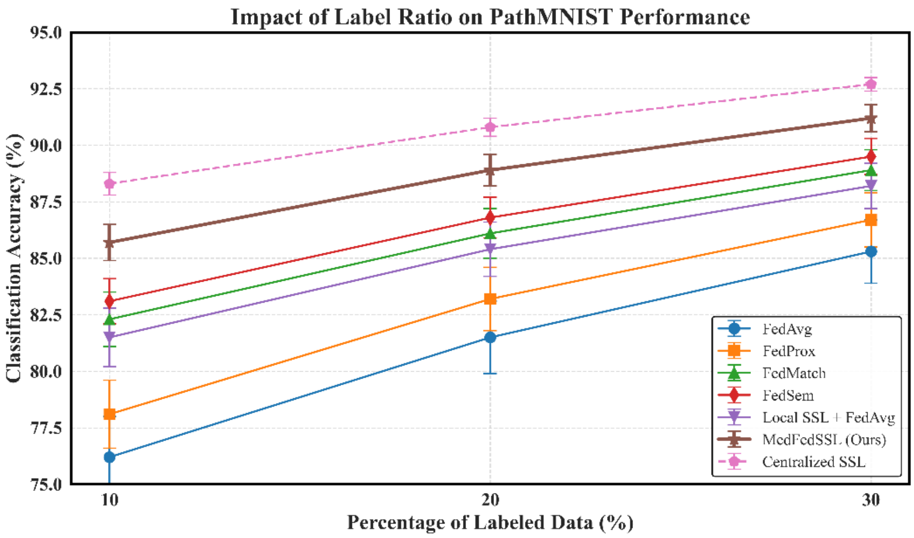

Fig. 2. Classification accuracy on PathMNIST with varying percentages of labeled data.

Segmentation Performance. Table 2 shows the segmentation performance of different methods on the BraTS dataset with 10% labeled data. We report the Dice coefficient for three tumor regions: whole tumor (WT), tumor core (TC), and enhancing tumor (ET).

Similar to the classification results, MedFedSSL outperforms all baseline methods across all tumor regions. Compared to FedAvg, MedFedSSL achieves improvements of 10.0%, 10.1%, and 10.2% for whole tumor, tumor core, and enhancing tumor, respectively. Compared to FedMatch and FedSem, the improvements are 3.2-4.1%, 3.5-4.3%, and 3.3-4.1%, respectively.

The segmentation task is more challenging than classification, as it requires pixel-level predictions. The consistent improvements across different tumor regions demonstrate the effectiveness of our approach in handling complex medical image analysis tasks.

Table 2. Segmentation performance (Dice coefficient %) on the BraTS dataset with 10% labeled data. The best results among federated methods are high-lighted in **bold.**

Method	Whole Tumor	Tumor Core	Enhancing Tumor
FedAvg	72.3 ± 2.1	65.8 ± 2.5	61.2 ± 2.8
FedProx	74.5 ± 1.9	67.9 ± 2.2	63.5 ± 2.5
FedMatch	78.2 ± 1.5	71.6 ± 1.8	67.3 ± 2.1
FedSem	79.1 ± 1.3	72.4 ± 1.6	68.1 ± 1.9
Local SSL + FedAvg	77.5 ± 1.6	70.8 ± 1.9	66.5 ± 2.2
MedFedSSL (Ours)	**82.3 ± 1.0**	**75.9 ± 1.3**	**71.4 ± 1.6**
Centralized SSL	85.7 ± 0.7	79.2 ± 0.9	74.8 ± 1.1

Table 3. Ablation study on PathMNIST with 10% labeled data.

Method	Accuracy (%)
MedFedSSL (Full)	85.7 ± 0.8
- w/o Augmentation Consistency	83.2 ± 1.0
- w/o Model Consistency	82.8 ± 1.1
- w/o Adaptive Knowledge Distillation	83.5 ± 0.9
- w/o Proximal Term	84.1 ± 0.9
- w/o Weighted Aggregation	84.3 ± 0.8

Ablation Study. To understand the contribution of each component in our MedFedSSL framework, we conduct an ablation study on the PathMNIST dataset with 10% labeled data. Table 3 shows the results of removing different components from our full model.

Removing the augmentation consistency term leads to a 2.5% drop in accuracy, while removing the model consistency term results in a 2.9% drop. This indicates that both consistency terms are important for leveraging unlabeled data effectively. The adaptive knowledge distillation strategy also contributes significantly, with a 2.2% drop when removed. The proximal term and weighted aggregation have smaller but still noticeable impacts (1.6% and 1.4%, respectively), highlighting their role in addressing data heterogeneity and label imbalance.

Communication Efficiency and Impact of Client Heterogeneity. Fig.3 (a) shows the test accuracy on PathMNIST as a function of the number of communication rounds for different methods. MedFedSSL converges faster than the baselines, achieving higher accuracy with fewer communication rounds. This is particularly important in federated learning, where communication can be a bottleneck.

To evaluate the robustness of our approach to client heterogeneity, we vary the Dirichlet concentration parameter α from 0.1 (highly non-IID) to 1.0 (more IID) on the PathMNIST dataset. As shown in Fig.3 (b), all methods perform better when the data distribution is more IID ($\alpha = 1.0$). However, MedFedSSL is more robust to client heterogeneity, maintaining higher accuracy even in highly non-IID settings ($\alpha = 0.1$). This demonstrates the effectiveness of our approach in addressing the challenges of data heterogeneity in federated learning.

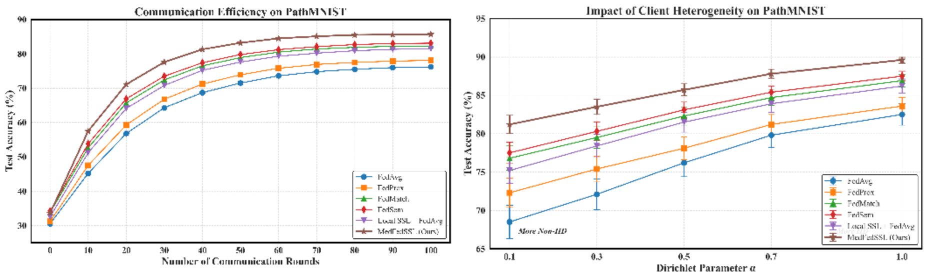

Fig. 3. Performance analysis on PathMNIST dataset: (a,The left one) communication efficiency and(b, The right one)impact of client heterogeneity with Dirichlet parameter α from0.1 to 1.0.

Conclusion

In this paper, we proposed MedFedSSL, a novel semi-supervised federated learning framework specifically designed for medical image analysis. Our approach integrates a dual-consistency regularization mechanism with an adaptive knowledge distillation strategy to effectively leverage both labeled and unlabeled data across distributed clients. We introduced a theoretically sound optimization objective that addresses the challenges of data heterogeneity and label imbalance in medical imaging, with provable convergence guarantees and bounds on the generalization error.

Extensive experiments on multiple medical imaging datasets demonstrated that MedFedSSL significantly outperforms state-of-the-art federated learning and semi-supervised learning methods, achieving superior performance with limited labeled data while preserving privacy. Our approach is particularly effective in low-label regimes and is robust to client heterogeneity, making it well-suited for real-world medical applications.

There are several promising directions for future work. First, extending our approach to handle multi-modal medical data (e.g., combining imaging with clinical records) could further improve performance. Second, incorporating privacy-preserving techniques such as differential privacy or secure aggregation could enhance the privacy guarantees of our framework. Third, exploring personalized federated learning approaches that adapt the global model to the specific characteristics of each client could improve performance for individual institutions. Finally, deploying and evaluating our framework in real-world clinical settings would provide valuable insights into its practical utility and potential impact on healthcare.

Acknowledgements. This work was supported by the 2024 College Students' Innovative Entrepreneurial Training Plan Program (No.202410293051Z).

References

1. Bai, W., Oktay, O., Sinclair, M., Suzuki, H., Rajchl, M., Tarroni, G., Glocker, B., King, A., Matthews, P.M., Rueckert, D.: Semi-supervised learning for network-based cardiac mr image segmentation. In: Medical Image Computing and Computer-Assisted Intervention-MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part II 20. pp. 253–260. Springer (2017)
2. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.A.: Mixmatch: A holistic approach to semi-supervised learning. Advances in neural information processing systems **32** (2019)
3. Berthelot, D., Roelofs, R., Sohn, K., Carlini, N., Kurakin, A.: Adamatch: A unified approach to semi-supervised learning and domain adaptation. arXiv preprint arXiv:2106.04732 (2021)
4. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proceedings of the eleventh annual conference on Computational learning theory. pp. 92–100 (1998)
5. Chen, C., Bai, W., Rueckert, D.: Multi-task learning for left atrial segmentation on ge-mri. In: Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges: 9th International Workshop, STACOM 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers 9. pp. 292–301. Springer (2019)
6. Chen, J., Zeng, Y., Li, Y., Huang, G.B.: Unsupervised feature selection based extreme learning machine for clustering. Neurocomputing **386**, 198–207 (2020)
7. Jeong, W., Yoon, J., Yang, E., Hwang, S.J.: Federated semi-supervised learning with inter-client consistency & disjoint learning. In: 9th International Conference on Learning Representations, ICLR 2021. International Conference on Learning Representations, ICLR (2021)
8. Konecn` y, J., McMahan, H.B., Felix, X.Y., Richt´arik, P., Suresh, A.T., Bacon, D.: Federated learning: Strategies for improving communication efficiency. CoRR (2016)
9. Laine, S., Aila, T.: Temporal ensembling for semi-supervised learning. In: International Conference on Learning Representations (2017)
10. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. vol. 3, p. 896. Atlanta (2013)
11. Li, J., Xiong, C., Hoi, S.C.: Comatch: Semi-supervised learning with contrastive graph regularization. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9475–9484 (2021)
12. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems 2, 429–450 (2020)
13. Li, W., Milletar` ı, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., Baust, M., Cheng, Y., Ourselin, S., Cardoso, M.J., et al.: Privacy-preserving federated brain tumour segmentation. In: Machine Learning in Medical Imaging: 10th International Workshop, MLMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings 10. pp. 133–141. Springer (2019)
14. Lin, T., Kong, L., Stich, S.U., Jaggi, M.: Ensemble distillation for robust model fusion in federated learning. Advances in neural information processing systems 33, 2351–2363 (2020)
15. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)

16. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
17. Nishio, T., Yonetani, R.: Client selection for federated learning with heterogeneous resources in mobile edge. In: ICC 2019-2019 IEEE international conference on communications (ICC). pp. 1–7. IEEE (2019)
18. Reisizadeh, A., Mokhtari, A., Hassani, H., Jadbabaie, A., Pedarsani, R.: Fedpaq: A communication-efficient federated learning method with periodic averaging and quantization. In: International conference on artificial intelligence and statistics. pp. 2021–2031. PMLR (2020)
19. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., et al.: The future of digital health with federated learning. NPJ digital medicine **3**(1), 119 (2020)
20. Sajjadi, M., Javanmardi, M., Tasdizen, T.: Regularization with stochastic transformations and perturbations for deep semi-supervised learning. Advances in neural information processing systems **29** (2016)
21. Sheller, M.J., Edwards, B., Reina, G.A., Martin, J., Pati, S., Kotrotsou, A., Milchenko, M., Xu, W., Marcus, D., Colen, R.R., et al.: Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific reports 10(1), 12598 (2020)
22. Sohn, K., Berthelot, D., Carlini, N., Zhang, Z., Zhang, H., Raffel, C.A., Cubuk, E.D., Kurakin, A., Li, C.L.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Advances in neural information processing systems **33**, 596–608 (2020)
23. Su, H., Shi, X., Cai, J., Yang, L.: Local and global consistency regularized mean teacher for semi-supervised nuclei classification. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 559–567. Springer (2019)
24. Van Engelen, J.E., Hoos, H.H.: A survey on semi-supervised learning. Machine learning 109(2), 373–440 (2020)
25. Wang, H., Yurochkin, M., Sun, Y., Papailiopoulos, D., Khazaeni, Y.: Federated learning with matched averaging. In: International Conference on Learning Representations (2020)
26. Wang, J., Liu, Q., Liang, H., Joshi, G., Poor, H.V.: Tackling the objective inconsistency problem in heterogeneous federated optimization. Advances in neural information processing systems 33, 7611–7623 (2020)