Asymmetric Mutual Learning for Decentralized Federated Medical Imaging

Jiaqi Wang
jqwang@psu.edu
The Pennsylvania State University

Houping Xiao
hxiao@gsu.edu
Georgia State University

Fenglong Ma
fenglong@psu.edu
The Pennsylvania State University

Abstract

Rigorous medical data privacy laws significantly obstruct data collection necessary for centralized machine learning, creating a substantial barrier to medical artificial intelligence (AI) application advancements. Centralized federated learning offers a potential solution by enabling collaborative model training without direct data exchange. However, it encounters notable challenges, including reliance on a trusted central server and the requirement for participants to disclose model information explicitly. These issues can hinder widespread adoption, especially in sensitive domains such as healthcare, where data privacy and security are paramount. To address these challenges, this study introduces a novel decentralized federated learning framework, named FedDAL, designed to enhance privacy and efficiency in collaborative model training. Our framework innovatively maintains local private models and proxy models through asymmetrical mutual learning, tailored to consider the varying sizes and knowledge bases of different models involved. This asymmetrical approach ensures that local models can remain private, while only compact proxy models are shared during the iterative update process, significantly reducing the risk of sensitive data leakage and communication overhead. We validate the efficacy and practicality of our proposed framework through comprehensive experiments on medical image classification and segmentation tasks. Our results demonstrate that FedDAL not only achieves superior performance compared to traditional centralized federated learning methods but also offers a more secure and efficient approach. This framework represents a significant advancement in federated learning, providing a viable pathway for integrating AI into medical practice while upholding stringent privacy standards[1].

CCS Concepts

• **Applied computing** → **Health informatics**; • **Computing methodologies** → **Distributed algorithms**; *Machine learning algorithms*; • **Information systems** → *Data mining*.

Keywords

Medical Image Analysis, Distributed Learning, Federated Learning

[1]Source code is available at https://github.com/JackqqWang/ACMBCB24

ACM-BCB '24, November 22–25, 2024, Shenzhen, China

ACM ISBN 979-8-4007-1302-6/24/11
https://doi.org/3698587.3701344

ACM Reference Format:
Jiaqi Wang, Houping Xiao, and Fenglong Ma. 2024. Asymmetric Mutual Learning for Decentralized Federated Medical Imaging. In *Proceedings of ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB '24).* ACM, New York, NY, USA, 9 pages. https://doi.org/3698587.3701344

1 Introduction

The stringent privacy regulations governing the storage and utilization of medical data pose a significant challenge for training machine learning models through conventional centralized methods. Addressing this challenge, federated learning (FL) [4, 22, 37, 38, 40] has emerged as a solution, enabling collaborative model training across multiple entities without the necessity of direct data sharing. FL models predominantly fall into two categories: centralized federated learning and decentralized federated learning. Centralized FL federated learning relies on a shared central server for exchanging or aggregating client model parameters, a concept extensively explored in medical tasks [6, 8, 15, 25, 25, 39, 43, 44, 49, 52]. However, establishing a universally trusted server for model aggregation poses challenges.

To address this issue, decentralized federated learning has been proposed as a viable solution, eliminating the need for a central, trustworthy server [3, 16, 28, 33]. In decentralized federated learning, each client communicates directly with others within the network, effectively bypassing the challenge of finding a universally trusted aggregation point. In the medical domain, recent research [13] introduces a method where each client maintains a proxy model for parameter exchange, safeguarding the integrity of their private models from complete exposure. However, this approach overlooks the intrinsic differences between private and proxy models. The typically larger private model is confined to processing local data from its respective clients, while the smaller proxy model benefits from a richer diversity of information obtained through communication with others. Neglecting these distinctions in simple mutual learning processes can have a detrimental impact on the performance of local private models.

To address these challenges simultaneously, we propose FedDAL, an innovative cross-silo **d**ecentralized federated learning framework with **a**symmetrical mutua**l** learning, as shown in Figure 1. Each client will maintain a private model and a proxy model. To address the disparities in size and informational capacity between private and proxy models, we design an asymmetrical mutual learning strategy for model updates, a method detailed in Section 2.2. Considering the continuous learning process of the two models, we further introduce a dynamic strategy to manage the interaction between these two models, leveraging a consensus-based approach detailed in Section 2.3. Finally, to improve model synergy during

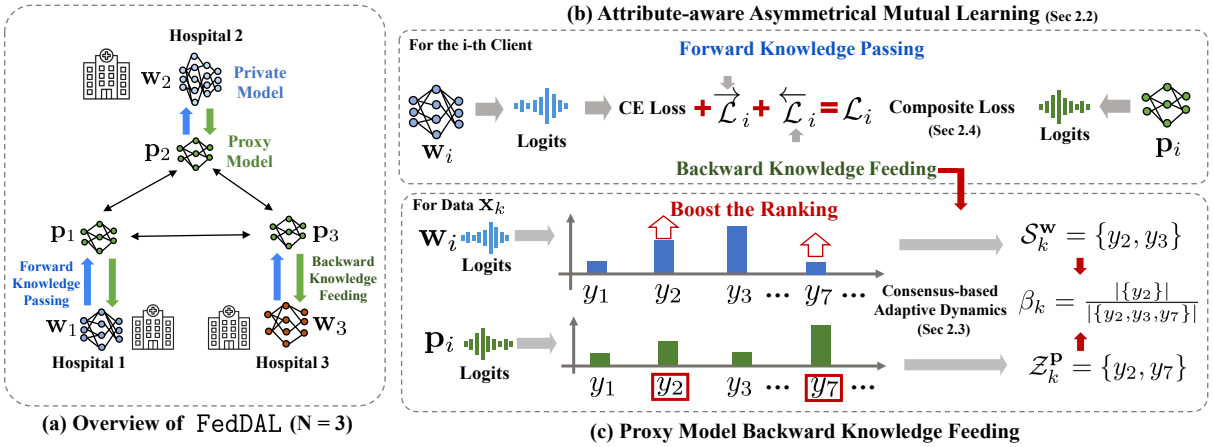

Figure 1: The framework of the proposed FedDAL. (*a*) shows the overview of the proposed FedDAL, which is a decentralized federated learning framework. Each client contains two models – a private model $\mathbf{w}_i$ and a proxy model $\mathbf{p}_i$. (*b*) demonstrates the designed attribute-aware asymmetrical mutual learning, which contains three losses – a traditional cross-entropy (CE) loss, a forward knowledge passing loss, and a backward knowledge feeding loss. (*c*) illustrates the details of the backward knowledge feeding loss. In this example, we select the top 2 labels. The private model $\mathbf{w}_i$ predicts the label of the input data $\mathbf{x}_k$ to be y_3 or y_2 according to the logit value, i.e., $\mathcal{S}_k^{\mathbf{w}} = \{y_2, y_3\}$. However, the proxy model $\mathbf{p}_i$ predicts the label set to be $\mathcal{Z}_k^{\mathbf{p}} = \{y_2, y_7\}$. The proposed FedDAL will enforce the private model to enhance the logit values of y_2 and y_7. Simultaneously, it can calculate the consensus score β_k, a key factor in the backward knowledge feeding loss.

local training, we adopt an alternating optimization strategy to ensure both models are continuously refined by minimizing a composite loss, further discussed in Section 2.4. After the local update, only the proxy models will be exchanged with other clients, enhancing privacy and reducing communication costs. We perform comprehensive experiments on the medical image classification and segmentation tasks with four datasets, evaluating the proposed FedDAL framework in both heterogeneous and homogeneous scenarios. The experiment results on four datasets with both classification and segmentation tasks consistently affirm the effectiveness of FedDAL.

The proposed FedDAL framework stands out with its distinct benefits over conventional methodologies. Primarily, its decentralized nature is particularly well-suited for the medical sector, addressing the issue of securing a reliable server, a common challenge in traditional FL setups. Furthermore, our unique approach to asymmetrical mutual learning considers the distinct characteristics of models, effectively managing potential disturbances from the proxy model while simultaneously bolstering the performance of the local private model. Additionally, by incorporating consensus-based adaptive dynamics, we refine the interplay between models and employ an alternative optimization technique to ensure both models remain current. Moreover, this comprehensive strategy safeguards the privacy of the local private model and notably diminishes communication costs across both homogeneous and heterogeneous environments.

2 Methodology

2.1 Model Overview

The proposed FedDAL framework is shown in Figure 1. This framework involves N hospitals, and they communicate with each other through the decentralized federated learning paradigm. Each hospital i will update its proxy $\mathbf{w}_i$ and its private model $\mathbf{p}_i$ with its own data $\mathcal{D}_i$ through the forward knowledge passing and backward knowledge passing algorithms, respectively. The proxy models are significantly smaller than the respective private models and standardized across hospitals. Subsequently, the hospitals transmit their proxy models to other hospitals for the purpose of aggregation. Then the aggregated proxy model parameters will serve to initialize the proxy model for the next communication round. It is important to note that exchanging only proxy models between the hospitals is instrumental in safeguarding model privacy and optimizing communication efficiency. Notably, our framework permits hospitals to employ diverse private models, further accommodating the heterogeneous nature of medical data and institutional capabilities. The full algorithm can be found in Algorithm 1.

2.2 Asymmetrical Mutual Learning

In each local client's setting, there exists a private model $\mathbf{w}i$ and a proxy model $\mathbf{p}i$. The most straightforward method is to engage in mutual learning, akin to the approach described in [13]. Yet, this method encounters various shortcomings, as outlined in a previous section. To address these issues, we introduce an asymmetrical mutual learning strategy aimed at updating both the private and proxy models while acknowledging their distinct characteristics. This strategy encompasses a bidirectional knowledge transfer between the private and proxy models, facilitating both forward and backward learning processes shown in Figure 1 (b). We use the classification task to introduce the algorithm in detail and apply the same techniques at the pixel level for the segmentation task.

2.2.1 Private Model Forward Knowledge Passing. First, we introduce how to conduct forward knowledge passing from the private

Algorithm 1: Algorithm Flow of FedDAL.

Input: Client training data $\mathcal{D}_1, \cdots, \mathcal{D}_N$, private models $\{\mathbf{w}_1, \cdots, \mathbf{w}_N\}$, proxy models $\{\mathbf{p}_1, \cdots, \mathbf{p}_N\}$, communication round T, the number of active clients at each communication round B, local training epoch V, and hyperparameters.

for each communication round $t = 1, 2, \cdots, T$ **do**
  Client Update
  for each epoch $v = 1, \cdots, V$ **do**
    for active client $i \in [1, \cdots, B]$ **do**
      Conduct asymmetrical mutual learning in Sec.2.2;
      Obtain β_k by calculating the consensus-based adaptive dynamics in Sec.2.3;
    end
  end
  Obtain private models $\{\mathbf{w}_1^t, \cdots, \mathbf{w}_N^t\}$ and proxy models $\{\mathbf{p}_1^t, \cdots, \mathbf{p}_N^t\}$;
  Upload proxy models $\{\mathbf{p}_1^t, \cdots, \mathbf{p}_N^t\}$ to the server;
  Server Update
  Conduct FedAvg or other existing methods to obtain aggregated model $\mathbf{p}_g^t$;
  Distribute the aggregated model $\mathbf{p}_t^g$ back to clients.
end

model $\mathbf{w}_i$ to the proxy model $\mathbf{p}_i$. This forward loss is as follows:

$$\overrightarrow{\mathcal{L}}_i = \sum_{k=1}^{|\mathcal{D}_i|} \mathrm{KL}(\mathbf{w}_i(\mathbf{x}_k), \mathbf{p}_i(\mathbf{x}_k)), \tag{1}$$

where $\mathbf{x}_k$ is the input data feature, $|\mathcal{D}_i|$ denotes the number of data in $\mathcal{D}_i$, and KL is the Kullback–Leibler divergence.

2.2.2 Proxy Model Backward Knowledge Feeding. Compared with the private model, the proxy model has two main unique properties: (1) Its capability is weaker than the private model due to its significant difference in the model size. (2) It carries a broader spectrum of information, given its role in facilitating exchanges among clients for aggregation. The conventional approach of knowledge distillation fails to account for these unique attributes, potentially leading to a decline in the performance of the private model.

To mitigate this, we have conceptualized a strategy for soft knowledge feeding tailored to enable proxy models to effectively impart knowledge to private models, as illustrated in Figure 1 (c). Specifically, for a given input $\mathbf{x}_k$, we organize the class probability output of $\mathbf{p}_i(\mathbf{x}_k)$ in descending order. For the top V classes, we aim to elevate their ranking in the output of $\mathbf{w}_i(\mathbf{x}_k)$ through the training of the private model $\mathbf{w}_i$. To facilitate this, we introduce a specialized soft knowledge feeding loss as follows:

$$\begin{aligned} \overleftarrow{\mathcal{L}}_i &= \sum_{k=1}^{|\mathcal{D}_i|} \beta_k \sum_{j \in \mathcal{Z}_k^{\mathbf{p}}} \log\left(\frac{\exp(\mathbf{w}_i(\mathbf{x}_k)[j])}{\Phi_k}\right), \\ \Phi_k &= \sum_{z \in \mathcal{Z}_k^{\mathbf{p}}} \exp(\mathbf{w}_i^t(\mathbf{x}_k)[z]) + \sum_{z' \in \mathcal{Z}_k^{\mathbf{p}'}} \exp(\mathbf{w}_i(\mathbf{x}_k)[z']), \end{aligned} \tag{2}$$

where $\mathcal{Z}_k^{\mathbf{p}}$ is the top V class set from the output of the proxy model $\mathbf{p}_i$, $\mathcal{Z}_k^{\mathbf{p}'}$ is the complement set with lower ranks, and β_v is the hyperparameter to control how much soft knowledge will be fed to the private model. In the next subsection, we will introduce how we adaptively decide the value of the hyperparameter β_k.

2.3 Consensus-based Adaptive Dynamics

The parameter β_k determines the soft knowledge transfer from proxy to private models. A static β_k overlooks the proxy model's dynamic updates via client aggregations and the mutual influence between proxy and private models, making it insufficient for managing knowledge flow. Thus, we introduce a consensus-based dynamic method to adjust β_k, adapting to the models' evolving interactions.

As shown in Figure 1 (c), for a given input $\mathbf{x}_k$, alongside the private model $\mathbf{w}_i$ and the proxy model $\mathbf{p}_i$, we define the top-V class sets as $\mathcal{S}_k^{\mathbf{w}}$ and $\mathcal{Z}_k^{\mathbf{p}}$, respectively. The interaction during local training sessions and communications with other clients can lead to variations in these sets. To quantify the consensus between the outputs of the two models, we use the following formula:

$$\beta_k = \frac{|\mathcal{S}_k^{\mathbf{w}} \cap \mathcal{Z}_k^{\mathbf{p}}|}{|\mathcal{S}_k^{\mathbf{w}} \cup \mathcal{Z}_k^{\mathbf{p}}|}, \tag{3}$$

where $|\cdot|$ represents the cardinality of the set. This method offers multiple advantages: (1) It allows the weight to adapt dynamically in tandem with the models' updates, eliminating the need for a fixed, manually-set value, (2) The adaptation is grounded in the consensus between the outputs of the two models, enhancing when both models concur on the high-confidence classes for the given input and diminishing otherwise. This approach safeguards the private model's learning process from potential disruptions caused by the smaller proxy model, thereby enhancing local model performance.

2.4 Composite Loss and Alternative Optimization

At each local client, we have two models and they affect each other during the training process. To enhance their interaction at every step of training, we propose an alternative optimization strategy to minimize the composite loss, as detailed below:

$$\mathcal{L}_i = \sum_{k=1}^{|\mathcal{D}_i|} \mathrm{CE}(\mathbf{w}_i(\mathbf{x}_k), \mathbf{y}_k) + \overrightarrow{\mathcal{L}}_i + \overleftarrow{\mathcal{L}}_i,$$

where we freeze $\mathbf{w}_i$ to optimize $\mathbf{p}_i$ and freeze $\mathbf{p}_i$ to optimize $\mathbf{w}_i$ alternatively for each training batch at each training epoch. *Note that we use classification tasks as an example to demonstrate the proposed* FedDAL*, which can be further applied to the segmentation tasks at the pixel level.*

3 Experiments

3.1 Experimental Setups

3.1.1 Datasets. For the classification task, we utilize two datasets: Fed-ISIC19 [36] and the Pneumonia dataset from Chest X-ray [42]. The Fed-ISIC19 dataset, which encompasses 23,247 dermoscopy images, is employed for the classification of eight distinct melanoma types. We also perform a binary classification evaluation for pneumonia using 5,863 chest x-ray images. For the segmentation task,

we select the 2D lung segmentation dataset[2] and 3D brain segmentation dataset [36] in our experiments.

Classification Datasets. Following the data distribution scheme of FLamby [23], we segregate six clients into two groups: three with larger data volumes and three with smaller ones. The allocation of training/validation/testing data among these clients is as follows: 8,689/1,241/2,483, 2,768/395/791, 2,654/336/673 for the larger clients, and 655/164, 351/88, 180/45 for the smaller ones. The chest X-ray dataset allocation involves two major clients and four minor clients, with the training, validation, and testing data distributed as follows: 2,743/391/374, 917/131/124, 370/52/49, 278/39/37, 187/26/24, and 196/13/12, respectively.

Segmentation Datasets. The 2D lung segmentation dataset comprises 704 images distributed among 4 clients, with respective training, validation, and test data allocations of 254/31/31, 254/31/31, and 58/7/7. The 3D brain T1 magnetic resonance imaging (MRI) dataset, sourced from the Information Extraction from Images (IXI) database, adheres to the data distribution format of Fed-IXI, as referenced by FLamby [23]. This setup involves training, validation, and testing images, 218/31/62, 127/18/36, and 51/8/15, respectively.

3.1.2 Comparison Scenarios and Baselines. In our experiments, we aim to thoroughly evaluate the proposed FedDAL in two scenarios: **homogeneous** and **heterogeneous** evaluation. In the homogeneous evaluation, all the clients use the same client model, while in the heterogeneous evaluation, clients use different models.

The **homogeneous** setting employs FedAvg [22] and FedProx [18] as baselines:

- **FedAvg** [22]: It is the vanilla version of FL. The active local clients train their models and send the parameters to a central server. This server aggregates the model parameters by calculating their average, forming a global model that is then distributed back to the clients for further rounds of local training.
- **FedProx** [18]: This work introduces a proximal term in the local training of each client, quantifying the divergence between the local and global models. As it is an approach based on FedAvg, the model structures of local clients need to be identical.

There is no study focusing on **heterogeneous**, decentralized, federated learning. To evaluate the proposed FedDAL in our setting, we slightly modify the setting of ProxyFL [13], which can be used as a baseline in this setting:

- **ProxyFL** [13]: This approach is a decentralized FL framework. In the original work, the clients share identical local private models and proxy models with each other. They simply conduct the knowledge distillation with each other. For purposes of a fair assessment, our implementation adapts this approach by allowing clients to possess distinct private models while maintaining uniform proxy models, mirroring our configuration.

In addition, we also take the single client training as a baseline, denoted as "**Central**," where the client trains the private model only using its private data. This is not a federated learning baseline.

[2]https://www.kaggle.com/datasets/nikhilpandey360/chest-xray-masks-and-labels/

3.1.3 Implementation Details. For the classification tasks, we use ResNet-34 as the private model and ResNet-18 as the proxy model for all the clients in the homogeneous setting for the proposed FedDAL, while for baselines FedAvg and FedProx, they use ResNet-34 as the private models. In the heterogeneous setting, we randomly deploy either ResNet-101 or ResNet-34 on each client as the private model, and the proxy model is still ResNet-18. For the segmentation tasks, we use U-net [27] with different depths for different clients. Following previous work [12, 20, 45], we use accuracy and dice score as the metric.

We set communication round $T = 100$, the local training epoch is equal to 5, $V = 2$ for the Fed-ISIC19 dataset, $V = 1$ for other datasets, and the local learning rate is equal to 0.0001. We conduct our experiment under the cross-silo setting, where every client will participate in the update process at each communication round following [13, 36]. All the models, baselines, and the proposed FedDAL are implemented using PyTorch 2.0.1. Our experiments are conducted on an NVIDIA A100 with CUDA version 12.0, running on a Ubuntu 20.04.6 LTS server.

3.2 Experiment Results

3.2.1 Classification Results. We use Fed-ISIC19 and the Pneumonia dataset to demonstrate the performance of the classification task in Table 1. We provide several observations and discussions based on the results.

Our proposed method FedDAL consistently outperforms all other baseline methods in both homogeneous and heterogeneous settings. While clients with more extensive training data, typically show better performance, an interesting observation from our study is that our approach notably enhances the performance of clients with less data. This improvement is significant even when taking into account the local computational limitations of these clients with less data.

In certain scenarios, we noticed that training models on a single client can surpass the performance of some FL baselines. This suggests that small clients could potentially detract from the overall learning process, affecting other clients negatively. However, our approach not only improves the performance of individual clients but also elevates the average performance across clients compared with learning. This effectively mitigates the adverse impact small clients might have in the FL setup, ensuring a more beneficial learning environment for all participants.

3.2.2 Segmentation Results. For the segmentation task, we employ the 2D lung segmentation dataset and the 3D brain segmentation dataset to demonstrate the effectiveness of our proposed FedDAL in Table 2. Consistent with the findings from the classification task, our method FedDAL surpasses the baseline approaches across both the 2D and 3D segmentation datasets in terms of accuracy and Dice coefficient, under both homogeneous and heterogeneous settings. Besides, we randomly select two data from clients and visualize their segmentation results, as shown in Figures 2 and 3, compared with ProxyFL. The visualization experiments further show the effectiveness of FedDAL across various datasets and tasks.

Task	Client		Client 1	Client 2	Client 3	Client 4	Client 5	Client 6	Average
Fed-ISIC19	Training Data		8,689	2,763	2,654	655	351	180	-
	Homogeneous	FedAvg	0.4883	0.4790	0.4642	0.3877	0.3726	0.2401	0.4053
		FedProx	0.4932	0.4821	0.4693	0.3989	0.3803	0.2466	0.4117
		FedDAL	**0.5375**	**0.5061**	**0.4816**	**0.4126**	**0.4055**	**0.2857**	**0.4382**
		(% imp.)	8.89%↑	4.98%↑	2.62%↑	3.43% ↑	6.63%↑	15.86% ↑	6.44%↑
	Heterogeneous	Central	0.5505	0.5093	0.4827	0.3264	0.3103	0.2048	0.3973
		ProxyFL	0.5600	0.5086	0.5005	0.4131	0.4084	0.3041	0.4491
		FedDAL	**0.5841**	**0.5318**	**0.5147**	**0.4706**	**0.4519**	**0.3682**	**0.4869**
		(% imp.)	4.30%↑	4.42%↑	2.84%↑	13.92%↑	10.65%↑	21.08%↑	8.42%↑
Pneumonia	Training Data		2,743	917	370	278	187	196	-
	Homogeneous	FedAvg	0.7715	0.7608	0.7432	0.7144	0.6706	0.6515	0.7187
		FedProx	0.7762	0.7812	0.7560	0.7106	0.6845	0.6523	0.7268
		FedDAL	**0.8336**	**0.8204**	**0.7764**	**0.7481**	**0.7219**	**0.6608**	**0.7602**
		(% imp.)	7.40%↑	5.02%↑	2.70%↑	4.72%↑	5.46%↑	1.30%↑	4.60%↑
	Heterogeneous	Central	0.7889	0.7967	0.7412	0.7186	0.6696	0.6447	0.7266
		ProxyFL	0.8424	0.8206	0.7633	0.7338	0.7220	0.6654	0.7579
		FedDAL	**0.8813**	**0.8625**	**0.8141**	**0.7809**	**0.7711**	**0.7327**	**0.8071**
		(% imp.)	4.62%↑	5.12%↑	5.10%↑	6.42%↑	6.80%↑	10.11%↑	6.73%↑

Table 1: Accuracy comparison on the Fed-ISIC19 and Pneumonia datasets. "Underline" indicates the best baseline performance, and "bold" denotes the best performance. "% imp." denotes the improvement percentage compared to the best baseline.

Task	Client		Client 1		Client 2		Client 3		Client Avg	
2D Lung	Training Data		254		254		58		-	
	Metric		Acc	DC	Acc	DC	Acc	DC	Acc	DC
	Homogeneous	FedAvg	0.8525	0.8304	0.8662	0.8429	0.8130	0.7745	0.8439	0.8159
		FedProx	0.9074	0.8811	0.8917	0.8610	0.8621	0.8419	0.8871	0.8613
		FedDAL	**0.9249**	**0.8893**	**0.9145**	**0.8853**	**0.8870**	**0.8676**	**0.9088**	**0.8807**
		(% imp.)	1.93%↑	0.93%↑	2.56%↑	2.82%↑	2.89%↑	3.05%↑	2.45%↑	2.25%↑
	Heterogeneous	Central	0.9048	0.8875	0.8907	0.8768	0.7955	0.7680	0.8637	0.8441
		ProxyFL	0.9258	0.9076	0.9236	0.8944	0.8890	0.8703	0.9128	0.8908
		FedDAL	**0.9414**	**0.9211**	**0.9377**	**0.9205**	**0.9164**	**0.8868**	**0.9318**	**0.9095**
		(% imp.)	1.69%↑	1.49%↑	1.53%↑	2.92%↑	3.08%↑	1.90%↑	2.09%↑	2.10%↑
3D Brain	Training Data		218		127		51		-	
	Metric		Acc	DC	Acc	DC	Acc	DC	Acc	DC
	Homogeneous	FedAvg	0.7651	0.7408	0.7511	0.7267	0.6922	0.6432	0.7361	0.7149
		FedProx	0.7688	0.7586	0.7647	0.7279	0.6989	0.6502	0.7441	0.7160
		FedDAL	**0.7884**	**0.7677**	**0.7709**	**0.7418**	**0.7100**	**0.6778**	**0.7564**	**0.7291**
		(% imp.)	2.55%↑	1.20%↑	0.81%↑	1.91%↑	1.59%↑	4.24%↑	1.65%↑	1.83%↑
	Heterogeneous	Central	0.7692	0.7505	0.7598	0.7246	0.7001	0.6723	0.7430	0.7158
		ProxyFL	0.7850	0.7617	0.7784	0.7413	0.7055	0.6704	0.7563	0.7245
		FedDAL	**0.8217**	**0.8046**	**0.7932**	**0.7758**	**0.7206**	**0.6885**	**0.7785**	**0.7563**
		(% imp.)	4.68%↑	5.63%↑	1.90%↑	4.65%↑	2.14%↑	2.40%↑	2.94%↑	4.39%↑

Table 2: Results of the two segmentation tasks. "Acc" means the accuracy score, and "DC" denotes the dice score. "Underline" indicates the best baseline performance, and "bold" denotes the best performance. "% imp." denotes the improvement percentage compared to the best baseline.

3.3 Ablation Study

We evaluate several variations of our proposed model, FedDAL, to assess the impact of different components on its performance:

- $\texttt{FedDAL}_{\text{v}}$: In this variant, we enable the proxy model to transfer knowledge of the top-**v** classes to the private model without prioritizing an increase in their ranking during the back-and-forth knowledge exchange process.

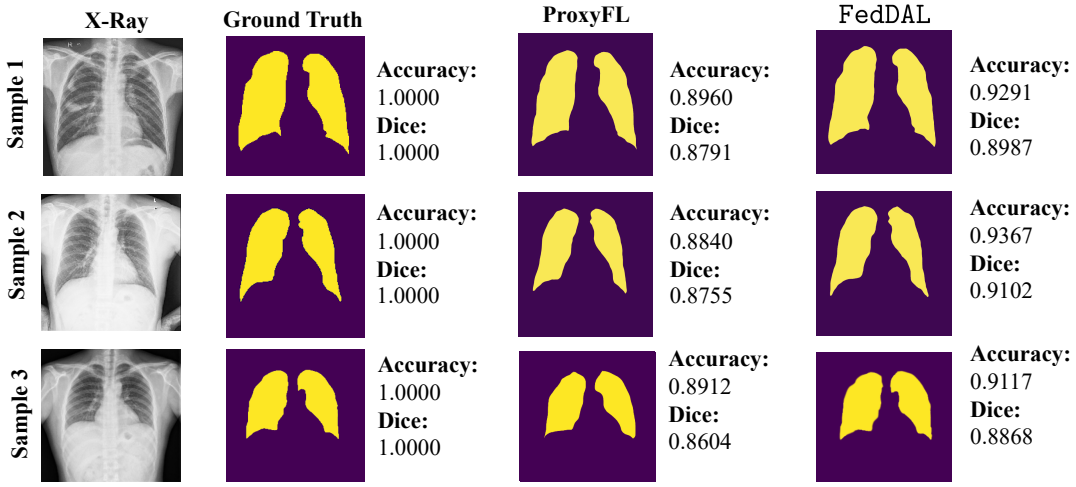

Figure 2: Visualization of the 2D lung segmentation task.

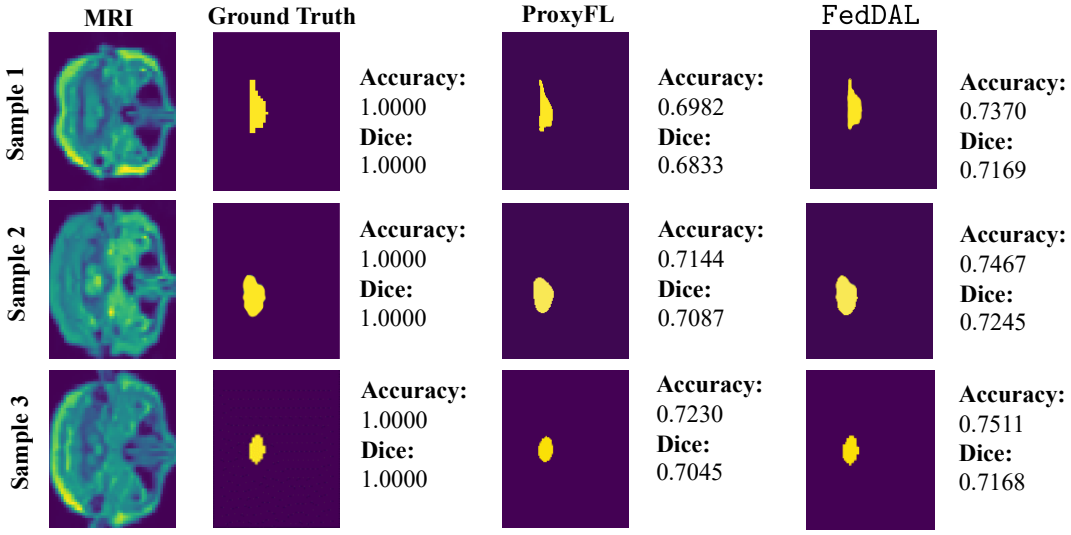

Figure 3: Visualization of the 3D brain MRI segmentation task.

- FedDAL$_{\beta}$: Here, we set the value of β_k to a constant 1, as opposed to employing a dynamic value determined by consensus.
- FedDAL$_{\text{alt}}$: This variant excludes the alternative optimization training strategy. Instead of simultaneous training, we train the proxy model for E epochs, followed by separate training of the private model for E epochs.

The results are reported in Figure 4. The outcomes of this study are presented in Figure 4. The findings indicate that omitting any of these design elements results in a performance decrease, with the most significant impact observed in the FedDAL$_{\text{v}}$ setting. This underscores the critical role of the ranking loss mechanism detailed in section 2.2 of our work. Additionally, the performance decline in the FedDAL$_{\beta}$ configuration highlights the importance of dynamically adjusting the value of β, as introduced in section 2.3. Lastly, the results from FedDAL$_{\text{alt}}$ demonstrate the effectiveness of the alternative training strategy outlined in section 2.4, further boosting the performance of our proposal. This ablation study methodically examines each design component, illustrating how their collection in FedDAL significantly enhances the overall results.

3.4 Model Insight Analysis

In this section, we delve into an analysis of model insights, focusing on the dynamics of β as described in Eq. 3 and the convergence behavior of the framework. In addition, we quantitatively compare the reduction of communication costs between the proposed FedDAL and baselines.

3.4.1 Analysis of Consensus Dynamics. With the Fed-ISIC19 dataset, we visualize the average value of β at each communication round

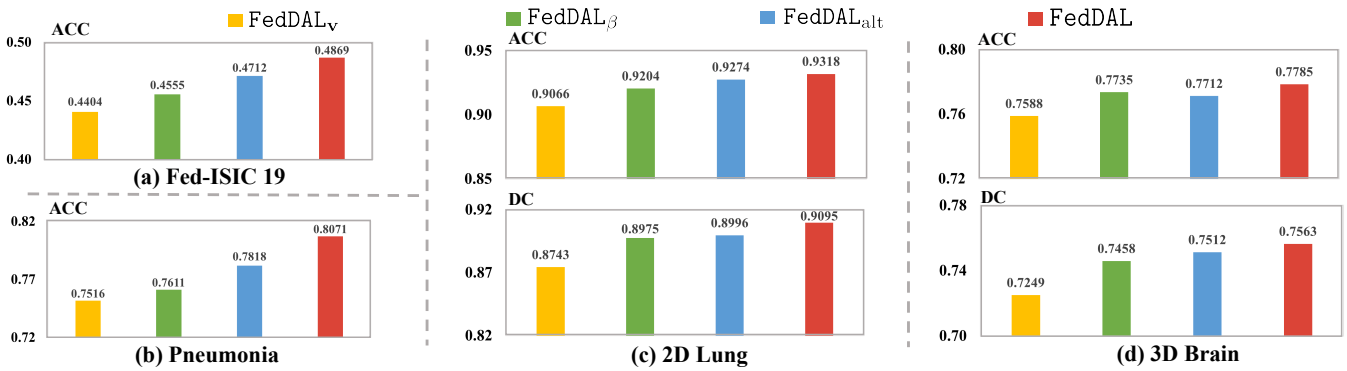

Figure 4: Ablation study of the classification and segmentation tasks on 4 datasets. (a) and (b) show the accuracy results on two classification tasks. (c) and (d) show the accuracy and dice scores on two segmentation tasks.

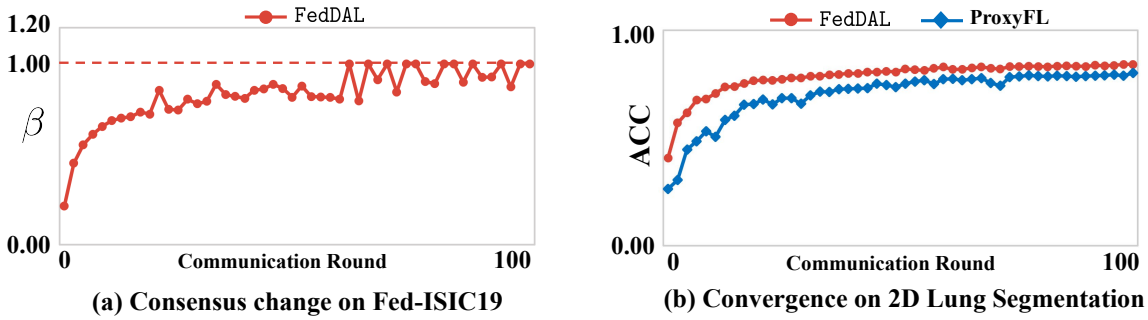

Figure 5: Model insight analysis.

in Figure 5 (a). Observations reveal a steady escalation in β values, signifying an enhanced level of agreement between the private and proxy models with the training process. These results underscore the evolving nature of β within the learning cycle and affirm our approach of progressively amplifying the influence of knowledge transfer from the proxy to the private model with increasing consensus.

3.4.2 Convergence Visualization. In Figure 5 (b), we show the average testing accuracy of FedDAL and ProxyFL over communication rounds on the 2D lung segmentation dataset. FedDAL not only surpasses ProxyFL in final accuracy but also reaches convergence faster and more stable, avoiding the late-stage fluctuations seen in ProxyFL. This enhanced performance and stability of FedDAL may be attributed to the effective integration of our asymmetrical mutual learning strategy (Sec.2.2) and consensus-based dynamics(Sec. 2.3), which tightly regulate the feedback of knowledge to the private model, minimizing potential disturbances from the proxy model. Additionally, our alternating optimization of the composite loss guarantees regular, detailed updates between the models, enhancing the learning process's effectiveness and stability.

3.4.3 Communication Cost Reduction. To quantitatively assess the benefits of this method, we conducted an experiment using the Fed-ISIC 19 dataset for a classification task. We evaluated the communication efficiency by measuring the sum of model parameters exchanged during each round of communication. The traditional approach of exchanging complete models between clients would necessitate the transfer of approximately 0.9569 billion parameters. However, with FedDAL, only about 0.1676 billion parameters are required to be exchanged, resulting in a substantial 82.49% reduction in communication costs. Furthermore, we compared this approach to a hypothetical extreme scenario within a typical centralized federated learning framework (such as FedAvg), where clients only communicate with a central server. Under identical conditions, the FedAvg-based method would require the exchange of 0.1914 billion parameters, which is still 14.20% more than what is needed with FedDAL. This comparison underscores the significant advantages of FedDAL in saving communication costs.

4 Related Work

In this section, we briefly summarize the studies in both centralized federated learning and decentralized federated learning approaches. We mainly discuss the work in a heterogeneous setting. Besides, our proposed FedDAL introduces a novel mutual learning. Thus, we also discuss the work in dual knowledge distillation.

4.1 Centralized Federated Learning

Most existing studies on centralized federated learning [2, 7, 21, 22, 34, 50, 51] have predominantly focused on the homogeneous setting, which necessitates uniform model structures across all clients. Most

existing work in the medical domain [6, 8, 15, 25, 25, 39, 44, 49, 52] also falls into this category.

Recent advancements have started delving into heterogeneous settings, where clients are allowed to employ varying model architectures. Within this realm, methods such as sub-model training [1], sparse model adaptation [5], and hypernetworks [29] are being investigated. These techniques, however, enforce specific constraints on the relationship between individual client models and the overarching global model, thereby limiting the clients' flexibility in model selection.

To promote more adaptable collaboration among diverse models, several studies have explored alternatives to parameter averaging, opting instead to aggregate supplementary information like logits [10], class scores [17], and label-wise representations [35, 46]. Yet, these approaches can potentially compromise privacy. Consequently, recent research has pivoted towards strategies such as ensemble learning [19], mutual learning [31, 47], or model reassembly [41] to exchange model parameters.

These methodologies face two significant limitations: (1) Dependence on public datasets. In practical scenarios, accessing public data might be impractical, and identifying appropriate public datasets without prior knowledge of the client's data is a complex issue. (2) The necessity of sharing sensitive information, including model architecture, model parameters, or data-derived insights from the client's private models with the server, poses substantial privacy risks.

4.2 Decentralized Federated Learning

Decentralized federated learning [3, 16, 28, 33] is a variation of federated learning that operates without a central server, allowing clients to collaborate directly with each other to train a shared global model. This peer-to-peer approach enhances privacy by eliminating a central entity that could access sensitive data, improves fault tolerance by avoiding single points of failure, and potentially scales better by distributing communication and computation loads across clients. However, decentralized Federated Learning also introduces new security challenges, requiring advanced cryptographic techniques or consensus mechanisms to ensure the integrity and authenticity of model updates.

In the medical field, recent research [13] has proposed a method where each client uses a proxy model to exchange parameters, thereby protecting the integrity of their private models from full exposure. However, this method fails to account for the intrinsic differences between private and proxy models. The larger private model is restricted to processing local data from its respective client, whereas the smaller proxy model gains from a more diverse range of information through communication with others. Ignoring these distinctions in straightforward mutual learning processes can adversely affect the performance of local private models.

4.3 Dual Knowledge Distillation

Knowledge distillation [9] treats the large model as a teacher, which passes knowledge to a small student model to enhance its performance. The most relevant work is bidirectional or dual knowledge distillation [14], enabling the teacher and student to learn knowledge from each other. In [14, 26], the bidirectional distillation technique is utilized to solve the top-k ranking research problem and machine translation [11, 48, 53]. Although a few studies apply bidirectional knowledge distillation in federated learning to conduct the tasks of distracted driving detection [30], medical relation extraction [32], and the IoT system [24], they all treat both teacher and student models equally yet ignore the importance of distinguishing the differences between the private and proxy models.

5 Conclusion

In our work, we present FedDAL, a novel decentralized federated learning framework designed to optimize for communication efficiency and model privacy within both heterogeneous and homogeneous settings, tailored specifically for hospital collaboration. This framework incorporates an attribute-aware asymmetrical mutual learning strategy, enabling the training of models of varying sizes, capabilities, and knowledge bases. Additionally, we employ a consensus-based dynamics mechanism to regulate the exchange of knowledge among models. The efficacy of FedDAL is validated across multiple medical datasets encompassing both classification and segmentation tasks, showcasing its potential to significantly enhance hospital collaboration and related healthcare applications in real-world scenarios. In our future work, we will investigate more effective approaches for the proxy model backward knowledge feeding and deploy the proposed solutions with our collaborators of multiple health institutions.

Acknowledgments

This work is partially supported by the National Science Foundation under Grant No. 2238275 and 2348541 and the National Institutes of Health under Grant No. R01AG077016.

References

[1] Samiul Alam, Luyang Liu, Ming Yan, and Mi Zhang. 2022. Fedrolex: Model-heterogeneous federated learning with rolling sub-model extraction. *Advances in Neural Information Processing Systems* 35 (2022), 29677–29690.

[2] Wenxuan Bao, Haohan Wang, Jun Wu, and Jingrui He. 2023. Optimizing the Collaboration Structure in Cross-Silo Federated Learning. *arXiv preprint arXiv:2306.06508* (2023).

[3] Enrique Tomás Martínez Beltrán, Mario Quiles Pérez, Pedro Miguel Sánchez Sánchez, Sergio López Bernal, Gérôme Bovet, Manuel Gil Pérez, Gregorio Martínez Pérez, and Alberto Huertas Celdrán. 2023. Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges. *IEEE Communications Surveys & Tutorials* (2023).

[4] Liwei Che, Jiaqi Wang, Yao Zhou, and Fenglong Ma. 2023. Multimodal federated learning: A survey. *Sensors* 23, 15 (2023), 6986.

[5] Daoyuan Chen, Liuyi Yao, Dawei Gao, Bolin Ding, and Yaliang Li. 2023. Efficient Personalized Federated Learning via Sparse Model-Adaptation. *arXiv preprint arXiv:2305.02776* (2023).

[6] Zhifang Deng, Dandan Li, Shi Tan, Ying Fu, Xueguang Yuan, Xiaohong Huang, Yong Zhang, and Guangwei Zhou. 2023. FedGrav: An Adaptive Federated Aggregation Algorithm for Multi-institutional Medical Image Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 170–180.

[7] Don Kurian Dennis, Tian Li, and Virginia Smith. 2021. Heterogeneity for the win: One-shot federated clustering. In *International Conference on Machine Learning*. PMLR, 2611–2620.

[8] Nicola K Dinsdale, Mark Jenkinson, and Ana IL Namburete. 2022. FedHarmony: unlearning scanner bias with distributed data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 695–704.

[9] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* (2015).

[10] Wenke Huang, Mang Ye, and Bo Du. 2022. Learn from others and be yourself in heterogeneous federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10143–10153.

[11] Hirofumi Inaguma, Tatsuya Kawahara, and Shinji Watanabe. 2021. Source and Target Bidirectional Knowledge Distillation for End-to-end Speech Translation. In *2021 Conference of the North American Chapter of the Association for Computational*

Linguistics: Human Language Technologies, NAACL-HLT 2021. Association for Computational Linguistics (ACL), 1872–1881.

[12] Meirui Jiang, Holger R Roth, Wenqi Li, Dong Yang, Can Zhao, Vishwesh Nath, Daguang Xu, Qi Dou, and Ziyue Xu. 2023. Fair Federated Medical Image Segmentation via Client Contribution Estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16302–16311.

[13] Shivam Kalra, Junfeng Wen, Jesse C Cresswell, Maksims Volkovs, and HR Tizhoosh. 2023. Decentralized federated learning through proxy model sharing. *Nature communications* 14, 1 (2023), 2899.

[14] Wonbin Kweon, SeongKu Kang, and Hwanjo Yu. 2021. Bidirectional distillation for top-K recommender system. In *Proceedings of the Web Conference 2021*. 3861–3871.

[15] Kyungsu Lee, Haeyun Lee, Thiago Coutinho Cavalcanti, Sewoong Kim, Georges El Fakhri, Dong Hun Lee, Jonghye Woo, and Jae Youn Hwang. 2023. Fine-Tuning Network in Federated Learning for Personalized Skin Diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 378–388.

[16] Chengxi Li, Gang Li, and Pramod K Varshney. 2021. Decentralized federated learning via mutual knowledge transfer. *IEEE Internet of Things Journal* 9, 2 (2021), 1136–1147.

[17] Daliang Li and Junpu Wang. 2019. Fedmd: Heterogenous federated learning via model distillation. *arXiv preprint arXiv:1910.03581* (2019).

[18] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. 2020. Federated optimization in heterogeneous networks. *Proceedings of Machine learning and systems* 2 (2020), 429–450.

[19] Tao Lin, Lingjing Kong, Sebastian U Stich, and Martin Jaggi. 2020. Ensemble distillation for robust model fusion in federated learning. *Advances in Neural Information Processing Systems* 33 (2020), 2351–2363.

[20] Quande Liu, Cheng Chen, Jing Qin, Qi Dou, and Pheng-Ann Heng. 2021. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1013–1023.

[21] Othmane Marfoq, Giovanni Neglia, Richard Vidal, and Laetitia Kameni. 2022. Personalized federated learning through local memorization. In *International Conference on Machine Learning*. PMLR, 15070–15092.

[22] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*. PMLR, 1273–1282.

[23] Jean Ogier du Terrail, Samy-Safwan Ayed, Edwige Cyffers, Felix Grimberg, Chaoyang He, Regis Loeb, Paul Mangold, Tanguy Marchand, Othmane Marfoq, Erum Mushtaq, et al. 2022. FLamby: Datasets and Benchmarks for Cross-Silo Federated Learning in Realistic Healthcare Settings. *Advances in Neural Information Processing Systems* 35 (2022), 5315–5334.

[24] Peihan Qi, Xiaoyu Zhou, Yuanlei Ding, Zhengyu Zhang, Shilian Zheng, and Zan Li. 2022. Fedbkd: Heterogenous federated learning via bidirectional knowledge distillation for modulation classification in iot-edge system. *IEEE Journal of Selected Topics in Signal Processing* 17, 1 (2022), 189–204.

[25] Xiaoming Qi, Guanyu Yang, Yuting He, Wangyan Liu, Ali Islam, and Shuo Li. 2022. Contrastive re-localization and history distillation in federated CMR segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 256–265.

[26] Sashank Reddi, Rama Kumar Pasumarthi, Aditya Menon, Ankit Singh Rawat, Felix Yu, Seungyeon Kim, Andreas Veit, and Sanjiv Kumar. 2021. Rankdistil: Knowledge distillation for ranking. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 2368–2376.

[27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 234–241.

[28] Abhijit Guha Roy, Shayan Siddiqui, Sebastian Pölsterl, Nassir Navab, and Christian Wachinger. 2019. Braintorrent: A peer-to-peer environment for decentralized federated learning. *arXiv preprint arXiv:1905.06731* (2019).

[29] Aviv Shamsian, Aviv Navon, Ethan Fetaya, and Gal Chechik. 2021. Personalized federated learning using hypernetworks. In *International Conference on Machine Learning*. PMLR, 9489–9502.

[30] Ertong Shang, Hui Liu, Zhuo Yang, Junzhao Du, and Yiming Ge. 2023. FedBiKD: Federated Bidirectional Knowledge Distillation for Distracted Driving Detection. *IEEE Internet of Things Journal* (2023).

[31] Tao Shen, Jie Zhang, Xinkang Jia, Fengda Zhang, Zheqi Lv, Kun Kuang, Chao Wu, and Fei Wu. 2023. Federated mutual learning: a collaborative machine learning method for heterogeneous data, models, and objectives. *Frontiers of Information Technology & Electronic Engineering* 24, 10 (2023), 1390–1402.

[32] Dianbo Sui, Yubo Chen, Jun Zhao, Yantao Jia, Yuantao Xie, and Weijian Sun. 2020. Feded: Federated learning via ensemble distillation for medical relation extraction. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)*. 2118–2128.

[33] Tao Sun, Dongsheng Li, and Bao Wang. 2022. Decentralized federated averaging. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 4 (2022), 4289–4301.

[34] Canh T Dinh, Nguyen Tran, and Josh Nguyen. 2020. Personalized federated learning with moreau envelopes. *Advances in Neural Information Processing Systems* 33 (2020), 21394–21405.

[35] Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. 2022. Fedproto: Federated prototype learning across heterogeneous clients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 8432–8440.

[36] Jean Ogier du Terrail, Samy-Safwan Ayed, Edwige Cyffers, Felix Grimberg, Chaoyang He, Regis Loeb, Paul Mangold, Tanguy Marchand, Othmane Marfoq, Erum Mushtaq, et al. 2022. Flamby: Datasets and benchmarks for cross-silo federated learning in realistic healthcare settings. *arXiv preprint arXiv:2210.04620* (2022).

[37] Jiaqi Wang, Yuzhong Chen, Yuhang Wu, Mahashweta Das, Hao Yang, and Fenglong Ma. 2024. Rethinking Personalized Federated Learning with Clustering-Based Dynamic Graph Propagation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 155–167.

[38] Jiaqi Wang and Fenglong Ma. 2023. Federated learning for rare disease detection: a survey. *Rare Disease and Orphan Drugs Journal* (2023).

[39] Jiaqi Wang, Cheng Qian, Suhan Cui, Lucas Glass, and Fenglong Ma. 2022. Towards federated covid-19 vaccine side effect prediction. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 437–452.

[40] Jiaqi Wang, Xiaochen Wang, Lingjuan Lyu, Jinghui Chen, and Fenglong Ma. 2024. FEDMEKI: A Benchmark for Scaling Medical Foundation Models via Federated Knowledge Injection. *Advances in Neural Information Processing Systems* (2024).

[41] Jiaqi Wang, Xingyi Yang, Suhan Cui, Liwei Che, Lingjuan Lyu, Dongkuan Xu, and Fenglong Ma. 2023. Towards Personalized Federated Learning via Heterogeneous Model Reassembly. In *Thirty-seventh Conference on Neural Information Processing Systems*.

[42] Xiaosong Wang, Yifan Peng, Le Lu, Zhiyong Lu, Mohammadhadi Bagheri, and Ronald M Summers. 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2097–2106.

[43] Xiaochen Wang, Jiaqi Wang, Houping Xiao, Jinghui Chen, and Fenglong Ma. 2024. FEDKIM: Adaptive Federated Knowledge Injection into Medical Foundation Models. *Proceedings of the 2024 conference on empirical methods in natural language processing (EMNLP)* (2024).

[44] Nannan Wu, Li Yu, Xin Yang, Kwang-Ting Cheng, and Zengqiang Yan. 2023. FedIIC: Towards robust federated learning for class-imbalanced medical image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 692–702.

[45] An Xu, Wenqi Li, Pengfei Guo, Dong Yang, Holger R Roth, Ali Hatamizadeh, Can Zhao, Daguang Xu, Heng Huang, and Ziyue Xu. 2022. Closing the generalization gap of cross-silo federated medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20866–20875.

[46] Liping Yi, Gang Wang, Xiaoguang Liu, Zhuan Shi, and Han Yu. 2023. FedGH: Heterogeneous Federated Learning with Generalized Global Header. *arXiv preprint arXiv:2303.13137* (2023).

[47] Sixing Yu, Wei Qian, and Ali Jannesari. 2022. Resource-aware Federated Learning using Knowledge Extraction and Multi-model Fusion. *arXiv preprint arXiv:2208.07978* (2022).

[48] Huaao Zhang, Shigui Qiu, and Shilong Wu. 2021. Dual knowledge distillation for bidirectional neural machine translation. In *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–7.

[49] Ruipeng Zhang, Ziqing Fan, Qinwei Xu, Jiangchao Yao, Ya Zhang, and Yanfeng Wang. 2023. Grace: A generalized and personalized federated learning method for medical imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 14–24.

[50] Xu Zhang, Yinchuan Li, Wenpeng Li, Kaiyang Guo, and Yunfeng Shao. 2022. Personalized federated learning via variational bayesian inference. In *International Conference on Machine Learning*. PMLR, 26293–26310.

[51] Yao Zhou, Jun Wu, Haixun Wang, and Jingrui He. 2022. Adversarial Robustness through Bias Variance Decomposition: A New Perspective for Federated Learning. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management, Atlanta, GA, USA, October 17-21, 2022*. ACM, 2753–2762.

[52] Wei Zhu and Jiebo Luo. 2022. Federated medical image analysis with virtual sample synthesis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 728–738.

[53] Yimeng Zhuang and Mei Tu. 2023. Pretrained Bidirectional Distillation for Machine Translation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Toronto, Canada, 1132–1145. https://doi.org/10.18653/v1/2023.acl-long.63