Adaptive Aggregation Weights for Federated Segmentation of Pancreas MRI

Published in:
ISBI 2025 - 2025 IEEE 22nd International Symposium on Biomedical Imaging, Proceedings

Publication status and date:
Published: 01/01/2025

DOI (link to publisher):
10.1109/ISBI60581.2025.10981148

Document Version
Publisher's PDF, also known as Version of record

ADAPTIVE AGGREGATION WEIGHTS FOR FEDERATED SEGMENTATION OF PANCREAS MRI

Hongyi Pan[1], *Gorkem Durak*[1], *Zheyuan Zhang*[1], *Yavuz Taktak*[2], *Elif Keles*[1], *Halil Ertugrul Aktas*[1], *Alpay Medetalibeyoglu*[3], *Yury Velichko*[1], *Concetto Spampinato*[4], *Ivo Schoots*[5], *Marco J. Bruno*[6], *Rajesh N. Keswani*[7], *Pallavi Tiwari*[8], *Candice Bolan*[9], *Tamas Gonda*[10], *Michael G. Goggins*[11], *Michael B. Wallace*[9], *Ziyue Xu*[12], *Ulas Bagci*[1]

[1]Department of Radiology, Northwestern University, Chicago, IL, USA
[2]Department of Radiology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey
[3]Department of Internal Medicine, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey
[4]Department of Electrical, Electronic and Computer Engineering, University of Catania, Catania, Italy
[5]Department of Radiology and Nuclear Medicine, Erasmus Medical Center, Rotterdam, Netherlands
[6]Department of Gastroenterology and Hepatology, Erasmus Medical Center, Rotterdam, Netherlands
[7]Department of Gastroenterology and Hepatology, Northwestern University, Chicago, IL, USA
[8]Department of Biomedical Engineering and Radiology, University of Wisconsin, Madison, WI, USA
[9]Division of Gastroenterology and Hepatology, Mayo Clinic, Jacksonville, FL, USA
[10]Division of Gastroenterology and Hepatology, New York University, New York, NY, USA
[11]Division of Gastroenterology and Hepatology, Johns Hopkins University, Baltimore, MD, USA
[12]NVIDIA, Bethesda, MD, USA

2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI) | 979-8-3315-2052-6/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISBI60581.2025.10981148

ABSTRACT

Federated learning (FL) enables collaborative model training across institutions without sharing sensitive data, making it an attractive solution for medical imaging tasks. However, traditional FL methods, such as Federated Averaging (FedAvg), face difficulties in generalizing across domains due to variations in imaging protocols and patient demographics across institutions. This challenge is particularly evident in pancreas MRI segmentation, where anatomical variability and imaging artifacts significantly impact performance. In this paper, we conduct a comprehensive evaluation of FL algorithms for pancreas MRI segmentation and introduce a novel approach that incorporates adaptive aggregation weights. By dynamically adjusting the contribution of each client during model aggregation, our method accounts for domain-specific differences and improves generalization across heterogeneous datasets. Experimental results demonstrate that our approach enhances segmentation accuracy and reduces the impact of domain shift compared to conventional FL methods while maintaining privacy-preserving capabilities. Significant performance improvements are observed across multiple hospitals (centers).

Index Terms— Federated learning, adaptive aggregation weights, pancreas segmentation, magnetic resonance imaging

1. INTRODUCTION

Magnetic Resonance Imaging (MRI) has become an essential tool in medical imaging, particularly for pancreas volumetry (requiring segmentation), pancreatic cancer risk estimation, and treatment planning [1, 2, 3, 4, 5]. The accurate segmentation of the pancreas from MRI scans is fundamental for quantitative analyses that inform clinical decision-making and personalized treatment strategies. However, developing robust deep-learning models for pancreas segmentation presents significant challenges. The pancreas exhibits a highly complex and variable anatomical structure, which complicates creating models that can consistently delineate its boundaries across diverse patient populations. Additionally, variability in MRI scanner hardware, imaging protocols, and patient demographics introduces a phenomenon known as domain shift, where models trained on data from one institution or imaging setup perform poorly when applied to data from different sources [1, 2].

To address these challenges, alternative strategies such as federated learning (FL) have gained traction, allowing institutions to collaboratively train models without sharing raw data [6, 7, 8, 9, 10, 11, 12]. This decentralized approach is particularly advantageous in medical imaging, where privacy regulations, restrict the transfer of sensitive patient data [13]. However, despite its promise, FL faces significant challenges in medical applications due to the inherent heterogeneity of data across institutions. Variability in imaging protocols, scanner resolutions, and patient populations introduces domain shifts that can impair model generalization.

These challenges are especially evident in pancreas MRI segmentation, which has never been performed under the FL setting. Besides, the pancreas is a small and anatomically variable organ, making it difficult to segment across datasets consistently. Available algorithms are focused on CT scans, and MRI based algorithms are absent. Regarding conventional FL methods, Federated Averaging (FedAvg) [14, 15] and its variant FedProx [16], aggregate models uniformly by assuming homogeneity among clients. This assumption can degrade model performance, as some clients may contribute less reliable updates due to differences in local data distributions. To address this issue, adaptive aggregation strategies [17, 18, 19, 20] have been introduced to weigh client contributions based on their local models' performance and relevance to the global model.

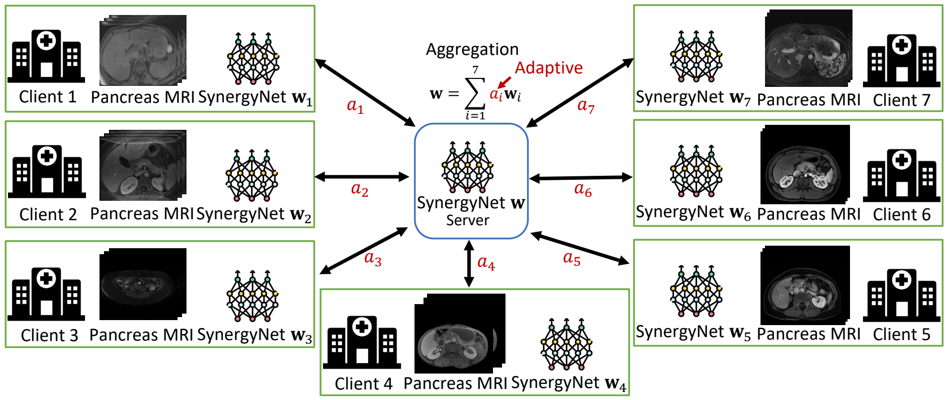

Fig. 1. Adaptive aggregation weights for federated pancreas MRI segmentation. Unlike the traditional FedAvg algorithm, where aggregation weights remain fixed, our method updates weights dynamically based on the validation loss gap before and after aggregation, ensuring that clients contributing positively to model performance are given more weight in subsequent rounds.

In this paper, we comprehensively evaluate FL algorithms for pancreas MRI segmentation. Building upon this evaluation, we propose a novel approach that leverages adaptive aggregation weights to improve FL performance. By dynamically adjusting the contribution of each client during model aggregation, our method enhances the model's ability to generalize across diverse institutions as Fig. 1. Unlike traditional FL methods, which assume equal contributions from all clients, our approach customizes the aggregation process based on local domain shifts. This leads to more robust and accurate segmentation performance, addressing the inherent heterogeneity in real-world medical imaging datasets.

2. METHODOLOGY

In this section, we will first formulate the FL problem. Then, we will review the Federated Averaging (FedAvg) algorithm, which is widely used in federated settings. Finally, we will introduce our approach for updating the aggregation weights dynamically to enhance the FedAvg algorithm, as illustrated in Fig. 1.

2.1. Federated Learning Problem Formulation

FL facilitates collaborative model training across multiple institutions without sharing raw data. Consider a central server with K clients, where each client has a training dataset denoted as $\Psi_i = \{\psi_{i,0}, \ldots, \psi_{i,n_i-1}\}$ containing n_i samples. The loss of model weights $\mathbf{w}$ on the i-th client is defined as the average loss over all samples in Ψ_i:

$$\mathcal{L}(\mathbf{w}, \Psi_i) = \frac{1}{n_i} \sum_{j=0}^{n_i-1} \mathcal{L}(\mathbf{w}, \psi_{i,j}), \tag{1}$$

where $\mathcal{L}(\mathbf{w}, \psi_{i,j})$ computes the loss on $\psi_{i,j}$ using model weights of $\mathbf{w}$. The server's objective is to leverage the data from all clients to identify an optimal model that minimizes the average loss across clients. Consequently, we aim to find a single model that minimizes the weighted average of client losses based on the number of samples:

$$\underset{\mathbf{w}}{\text{minimize}} \sum_{i=0}^{K-1} n_i \mathcal{L}(\mathbf{w}, \Psi_i), \tag{2}$$

Due to privacy constraints, clients are prohibited from sharing their local data with other clients in FL. Instead, each client uploads only its locally optimized model weights to the central server, where these weights are aggregated to update the global model. This process ensures that sensitive data remains decentralized, addressing privacy concerns while still enabling collaborative model training.

2.2. Background: Federated Averaging

FedAvg [14] is the most widely used FL algorithm, which works by distributing a global model to local clients, who then update it using their private data. After several local updates, the clients send their model updates (gradients or weights) back to a central server. The server then aggregates the updates by computing a weighted average, which is typically based on the number of data samples each client has, forming new global model weights $\mathbf{w}^{t+1}$:

$$\mathbf{w}^{t+1} = \sum_{i=0}^{K-1} \frac{n_i}{\sum_{j=0}^{K-1} n_j} \mathbf{w}_i^t, \tag{3}$$

where, $\mathbf{w}_i^t$ are the model weights updated from client i at round t.

However, FedAvg assumes that client data is independently and identically distributed (IID), which is rare in medical applications like pancreas MRI segmentation. Domain shifts between institutions due to different imaging protocols, patient populations, or scanner types can result in suboptimal generalization of the global model. To address these challenges, we propose a modified aggregation strategy that dynamically adapts to client-specific data distributions in the following section.

2.3. Adaptive Aggregation Weights for FedAvg

Inspired by the FedDG-GA algorithm [18], to account for data heterogeneity across clients, we introduce adaptive aggregation weights into the FedAvg framework. Our approach assigns a unique weight to each client based on their local data characteristics, including performance and domain shifts. This adaptive weighting mechanism ensures that clients with more relevant or higher-quality data contribute more to the global model, while clients with outlier data have less influence:

$$\mathbf{w}^{t+1} = \sum_{i=0}^{K-1} a_i^t \mathbf{w}_i^t, \tag{4}$$

where, a_i^t is the adaptive aggregation weight at round t. The aggregation weights are initialized the same as the one in the vanilla FedAvg algorithm:

$$a_i^0 = \frac{n_i}{\sum_{j=0}^{K-1} n_j}, \tag{5}$$

and they are updated based on the validation loss gap observed before and after aggregation: Let $P_i = \mathcal{L}(\mathbf{w}_i^t, \Phi_i)$ and $Q_i = \mathcal{L}(\mathbf{w}^t, \Phi_i)$ stand for the validation loss on validation dataset Φ_i before and after model aggregation, respectively, where $\mathcal{L}$ denotes the loss function. We first compute the loss gap:

$$G_i = Q_i - P_i, i = 0, \dots K-1. \tag{6}$$

Then, we update the aggregation weights based on the normalized loss gap:

$$a_i^{t+1} = a_i^t + \frac{G_i \cdot s}{\max(|G_0|, \dots, |G_{K-1}|)}, i = 0, \dots K-1. \tag{7}$$

Here, s is the step size for aggregation weights updating. In this work, we dynamically set s at epoch t as $0.1(1 - \frac{t}{T})$, training the model for a total of T epochs. This strategy allows the aggregation weights to change significantly at the beginning of the training process, gradually stabilizing as training progresses To ensure the aggregation weights remain within the range $[0, 1]$ and that their sum equals to 1, we clip the aggregation weights to the interval $[0, 1]$ and normalize them by dividing by their sum. In this approach, if a client contributes to better convergence of model weights–specifically if its corresponding validation loss before aggregation is smaller than that after aggregation–we increase its associated aggregation weights. Conversely, if a client detracts from model weight convergence–indicated by its validation loss before aggregation being worse than that after aggregation–we decrease its corresponding aggregation weights. Algorithm 1 outlines the process of incorporating adaptive aggregation weights into the FedAvg algorithm.

Algorithm 1 Federated averaging with adaptive aggregation weights across K centers.

Input: Global model initial wights $\mathbf{w}^0$, training datasets Ψ_i for $i = 0, 1, \dots, K-1$, where the i-th dataset contains n_i images, validation datasets Φ_i for $i = 0, 1, \dots, K-1$.
Hyperparameters: number of epochs T, adaptive aggregation weights step size s.
Output: Well-trained global model wights $\mathbf{w}^T$.

- $a_i^0 = \frac{n_i}{\sum_{j=0}^{K-1} n_j}$.
- **for** $t = 0, 1, \dots, T-1$ **do**
  - **for** $i = 0, 1, \dots, K-1$ **do**
    - Assign weights to local model: $\mathbf{w}_i^t = \mathbf{w}^t$;
    - Update $\mathbf{w}_i^t$ using Ψ_i;
  - **end for**
  - Aggregate models: $\mathbf{w}^{t+1} = \sum_{i=0}^{K-1} a_i^t \mathbf{w}_i^t$;
  - **for** $i = 0, 1, \dots, K-1$ **do**
    - Compute the validation loss before model aggregation: $P_i = \mathcal{L}(\mathbf{w}_i^t, \Phi_i)$;
    - Compute the validation loss after model aggregation: $Q_i = \mathcal{L}(\mathbf{w}^t, \Phi_i)$;
    - Compute the validation loss gap: $G_i = Q_i - P_i$;
  - **end for**
  - **for** $i = 0, 1, \dots, K-1$ **do**
    - Update the weight: $\tilde{a}_i^{t+1} = a_i^t + \frac{G_i \cdot s}{\max(|G_0|, \dots, |G_{K-1}|)}$;
    - Clip the weight: $\tilde{a}_i^{t+1} = \min(\max(\tilde{a}_i^{t+1}, 1), 0)$;
  - **end for**
  - **for** $i = 0, 1, \dots, K-1$ **do**
    - Normalize the weight: $a_i^{t+1} = \frac{\tilde{a}_i^{t+1}}{\sum_{j=0}^{k-1} \tilde{a}_j^{t+1}}$;
  - **end for**
- **end for**

3. EXPERIMENTAL RESULTS

This work is implemented using the extended version of the pancreas MRI dataset introduced in [21]. The dataset includes 723 T1-weighted and 738 T2-weighted MRI images, along with corresponding segmentation masks, collected from 7 centers: New York University Langone Health (NYU), Mayo Clinic Florida (MCF), Northwestern University (NU), Allegheny Health Network (AHN), Mayo Clinic Arizona (MCA), Istanbul University Faculty of Medicine (IU), and Erasmus Medical Center (EMC). The number of images available from each center is listed in Table 1. Samples of the MRI slices are shown in Fig. 1. We split the dataset into 80% for training and 20% for testing. Data is available at OSF[1]. The experiments are implemented using PyTorch [22].

SynergyNet [23] has achieved state-of-the-art performance in various 2D medical image segmentation tasks. In this study, we extended SynergyNet-8s2h to a 3D version by replacing its 2D convolutional blocks with 3D convolutional blocks, making it suitable for pancreas MRI segmentation. Other approaches can be adapted to our framework once they are shown suitable for pancreas segmentation. The training process employed the AdamW optimizer [24] with a learning rate of 1e-4, a batch size of 8, and a weight decay of 0.01. We trained the model for 300 epochs to minimize the Dice-BCE loss $\mathcal{L}$:

$$\mathcal{L}_{Dice} = 1 - \frac{2 \cdot |\mathbf{Y} \cap \hat{\mathbf{Y}}|}{|\mathbf{Y}| + |\hat{\mathbf{Y}}|} = 1 - \frac{2 \cdot \sum_{i=0}^{N-1} y_i \cdot \hat{y}_i}{\sum_{i=0}^{N-1} y_i + \sum_{i=0}^{N-1} \hat{y}_i}, \tag{8}$$

$$\mathcal{L}_{BCE} = -\frac{1}{N} \sum_{i=0}^{N-1} \left[y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i) \right], \tag{9}$$

$$\mathcal{L} = \mathcal{L}_{Dice} + \mathcal{L}_{BCE}, \tag{10}$$

where, $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ stand for the true and the predicted segmentation masks with N pixels, respectively. y_i and $\hat{y}_i$ denote the i-th pixels of $\mathbf{Y}$ and $\hat{\mathbf{Y}}$. All images were resized to $80 \times 256 \times 256$ for training and testing. To compare the performance of different methods, we evaluated the models based on the dice coefficient, Jaccard index (intersection over union), precision, recall, Hausdorff distance (HD95), and average symmetric surface distance (ASSD). The lower HD95 and ASSD and the higher other metrics indicate better performance.

Table 2 presents a comprehensive analysis of pancreas MRI seg-

Table 1. Pancreases Data Distribution.

Data Centers	Modalities	
	T1	T2
New York University Langone Health (NYU)	162	162
Mayo Clinic Florida (MCF)	148	143
Northwestern University (NU)	206	207
Allegheny Health Network (AHN)	17	27
Mayo Clinic Arizona (MCA)	25	23
Istanbul University Faculty of Medicine (IU)	74	73
Erasmus Medical Center (EMC)	91	103
Total	**723**	**738**

[1] https://osf.io/74vfs/

Table 2. Pancreas MRI segmentation results. "AAW" denotes our proposed Adaptive Aggregation Weights method. Improved results are highlighted in pink, while degraded results are highlighted in cyan.

	T1-Weighted Modality						T2-Weighted Modality					
Method	**Dice↑**	**Jaccard↑**	**Precision↑**	**Recall↑**	**HD95↓**	**ASSD↓**	**Dice↑**	**Jaccard↑**	**Precision↑**	**Recall↑**	**HD95↓**	**ASSD↓**
Center 1: New York University Langone Health (NYU)												
No FL	0.8194	0.7065	0.8502	0.7979	2.3686	0.4106	0.7863	0.6763	0.8200	0.7690	3.9018	0.7015
FedAvg [14]	0.6839	0.5310	0.6071	0.8096	7.1474	1.4332	0.6515	0.5064	0.6071	0.7498	9.5187	2.1262
FedAvg+AAW	0.7133	0.5709	0.6269	0.8447	8.1295	1.9601	0.5812	0.4342	0.5534	0.6657	12.9598	3.5244
Center 2: Mayo Clinic Florida (MCF)												
No FL	0.8279	0.7226	0.8367	0.8287	3.4076	0.6147	0.7993	0.7089	0.8735	0.7762	3.9265	1.0606
FedAvg [14]	0.6657	0.5183	0.6649	0.7393	10.6307	2.2033	0.5888	0.4811	0.7825	0.5628	17.9339	4.9035
FedAvg+AAW	0.7510	0.6222	0.6917	0.8613	7.2299	1.6896	0.6090	0.4829	0.7356	0.6031	15.9095	3.6724
Center 3: Northwestern University (NU)												
No FL	0.8184	0.7205	0.8443	0.8056	4.4057	1.5844	0.8417	0.7595	0.8717	0.8306	2.2739	0.5431
FedAvg [14]	0.6578	0.5045	0.5181	0.9145	4.6909	1.2883	0.6785	0.5416	0.5917	0.8185	4.4509	1.2742
FedAvg+AAW	0.6736	0.5241	0.5536	0.8752	5.2923	1.4514	0.6150	0.4688	0.5354	0.7512	5.2034	1.5686
Center 4: Allegheny Health Network (AHN)												
No FL	0.8321	0.7179	0.8549	0.8129	1.7661	0.3060	0.6492	0.5306	0.7420	0.5996	8.1092	1.8968
FedAvg [14]	0.4117	0.2727	0.6065	0.4587	24.3632	7.2514	0.3664	0.2376	0.5362	0.3190	25.4220	6.8224
FedAvg+AAW	0.6953	0.5568	0.7053	0.7997	5.1412	1.0850	0.4916	0.3380	0.5650	0.4672	12.1286	2.8360
Center 5: Mayo Clinic Arizona (MCA)												
No FL	0.7674	0.6471	0.7686	0.7682	2.3627	0.5067	0.6059	0.5003	0.6336	0.5969	10.2899	2.3163
FedAvg [14]	0.5975	0.4453	0.4937	0.7779	8.4508	1.9823	0.4077	0.3112	0.4278	0.3943	13.8540	4.0907
FedAvg+AAW	0.6680	0.5121	0.5597	0.8386	3.9422	0.9641	0.4282	0.3167	0.3717	0.5103	8.9490	3.1265
Center 6: Istanbul University Faculty of Medicine (IU)												
No FL	0.8394	0.7454	0.8455	0.8344	2.7811	0.4969	0.8360	0.7578	0.8292	0.8654	3.7258	0.8394
FedAvg [14]	0.6277	0.4716	0.4844	0.9122	8.2974	1.7792	0.7179	0.5859	0.6186	0.8681	7.1870	1.3420
FedAvg+AAW	0.6564	0.5038	0.5418	0.8499	6.2866	1.3795	0.5506	0.3994	0.5100	0.6396	9.7447	2.5334
Center 7: Erasmus Medical Center (EMC)												
No FL	0.8263	0.7402	0.9124	0.8021	4.0791	1.5077	0.8107	0.7350	0.8626	0.7987	3.9553	1.1469
FedAvg [14]	0.6952	0.5398	0.5762	0.9010	8.2871	1.2091	0.6387	0.5155	0.5821	0.7362	11.0488	3.1895
FedAvg+AAW	0.7121	0.5743	0.6593	0.8411	4.9892	1.2401	0.6012	0.4679	0.5259	0.7441	8.1523	2.2427
Average												
No FL	0.8223	0.7203	0.8507	0.8100	3.4022	0.9351	0.8011	0.7103	0.8422	0.7888	3.8224	0.9030
FedAvg [14]	0.6583	0.5061	0.5733	0.8367	7.9286	1.7303	0.6324	0.5034	0.6254	0.7140	10.4819	2.7442
FedAvg+AAW	0.7017	0.5593	0.6146	0.8554	6.3298	1.5523	0.5870	0.4467	0.5697	0.6727	10.2151	2.6916

mentation performance across multiple client centers using the proposed adaptive aggregation weights strategy within an FL framework. The reported average results are weighted based on the number of test samples from each center, ensuring that larger datasets have a proportionate influence on the overall performance metrics. For benchmarking purposes, we also include segmentation results obtained without the use of FL, which serves as an upper-bound reference. We observed that on the T1 modality, the adaptive aggregation weights significantly improve the FedAvg results with region-based metrics such as Dice and Jaccard. On the T2 modality, we observed reduced HD95 and ASSD metrics while did not observe much improvement from region-based metrics. We also observed, naturally, that some centers underperformed in the FL setting. This was expected too because of the number of images they contributed was small and quality varies significantly. Most of the results were better than conventional FedAvg, and approaching no-FL results.

4. CONCLUSION

This study presents an innovative adaptive aggregation weights strategy tailored for FL frameworks, specifically addressing the complexities of pancreas MRI segmentation. Traditional FL approaches aggregate model updates uniformly across clients, which can be suboptimal in heterogeneous clinical environments where local data distributions vary significantly. Our proposed method dynamically adjusts the contribution of each client during the model aggregation phase based on intrinsic local data characteristics, thereby mitigating the adverse effects of domain shift and enhancing model generalization across diverse clinical centers. This targeted aggregation approach not only improves the robustness of the segmentation model but also maintains the integrity and privacy of patient data, as raw data remains localized and secure within each institution.

Experimental validation was conducted using multi-center MRI datasets (of the pancreas) encompassing various imaging protocols and patient populations. Our results demonstrate a significant enhancement in segmentation performance metrics, notably a reduction in the HD95 and the ASSD. These improvements indicate a finer anatomical boundary delineation and more consistent segmentation accuracy across varying datasets. Moreover, the adaptive aggregation strategy effectively addresses the variability inherent in FL environments, ensuring that the global model remains resilient and generalizable without compromising patient privacy.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by Northwestern University (No. STU00214545).

6. ACKNOWLEDGMENTS

This work was supported by the following grants: NIH NCI R01-CA246704, R01-CA240639, U01-CA268808, NIH/NHLBI R01-HL171376, and NIH/NIDDK #U01 DK127384-02S1.

7. REFERENCES

[1] Jinzheng Cai, Le Lu, Zizhao Zhang, Fuyong Xing, Lin Yang, and Qian Yin, "Pancreas segmentation in mri using graph-based decision fusion on convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 442–450.

[2] Haribalan Kumar, Steve V DeSouza, and Maxim S Petrov, "Automated pancreas segmentation from computed tomography and magnetic resonance images: A systematic review," *Computer methods and programs in biomedicine*, vol. 178, pp. 319–328, 2019.

[3] Jinzheng Cai, Le Lu, Fuyong Xing, and Lin Yang, "Pancreas segmentation in ct and mri via task-specific network design and recurrent neural contextual learning," *Deep Learning and Convolutional Neural Networks for Medical Imaging and Clinical Informatics*, pp. 3–21, 2019.

[4] Xu Yao, Yuqing Song, and Zhe Liu, "Advances on pancreas segmentation: a review," *Multimedia Tools and Applications*, vol. 79, no. 9, pp. 6799–6821, 2020.

[5] Zheyuan Zhang, Ulas Bagci, Elif KELES, Camila Lopes Vendrami, Sachin Jambawalikar, Derk C Klatte, Megan M Engels, Candice W Bolan, Emil Agarunov, Tamas A Gonda, et al., "Tu1986 a fully automatic ai system for pancreas segmentation from multicenter mri scans," *Gastroenterology*, vol. 164, no. 6, pp. S–1175, 2023.

[6] Li Li, Yuxi Fan, Mike Tse, and Kuo-Yi Lin, "A review of applications in federated learning," *Computers & Industrial Engineering*, vol. 149, pp. 106854, 2020.

[7] Chen Zhang, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao, "A survey on federated learning," *Knowledge-Based Systems*, vol. 216, pp. 106775, 2021.

[8] Pengfei Guo, Dong Yang, Ali Hatamizadeh, An Xu, Ziyue Xu, Wenqi Li, Can Zhao, Daguang Xu, Stephanie Harmon, Evrim Turkbey, et al., "Auto-fedrl: Federated hyperparameter optimization for multi-institutional medical image segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 437–455.

[9] Meirui Jiang, Holger R Roth, Wenqi Li, Dong Yang, Can Zhao, Vishwesh Nath, Daguang Xu, Qi Dou, and Ziyue Xu, "Fair federated medical image segmentation via client contribution estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 16302–16311.

[10] Peyman Gholami and Hulya Seferoglu, "Digest: Fast and communication efficient decentralized learning with local updates," *IEEE Transactions on Machine Learning in Communications and Networking*, 2024.

[11] Runxuan Miao and Erdem Koyuncu, "Contrastive and non-contrastive strategies for federated self-supervised representation learning and deep clustering," *IEEE Journal of Selected Topics in Signal Processing*, 2024.

[12] Hongyi Pan, Debesh Jha, Koushik Biswas, and Ulas Bagci, "Frequency-based federated domain generalization for polyp segmentation," *arXiv preprint arXiv:2410.02044*, 2024.

[13] Mustafa Y Topaloglu, Elisabeth M Morrell, Suraj Rajendran, and Umit Topaloglu, "In the pursuit of privacy: the promises and predicaments of federated learning in healthcare," *Frontiers in Artificial Intelligence*, vol. 4, pp. 746497, 2021.

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

[15] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai, "Fedavg with fine tuning: Local updates lead to representation learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 10572–10586, 2022.

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

[17] Lei Tan, Xiaoxi Zhang, Yipeng Zhou, Xinkai Che, Miao Hu, Xu Chen, and Di Wu, "Adafed: Optimizing participation-aware federated learning with adaptive aggregation weights," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 4, pp. 2708–2720, 2022.

[18] Ruipeng Zhang, Qinwei Xu, Jiangchao Yao, Ya Zhang, Qi Tian, and Yanfeng Wang, "Federated domain generalization with generalization adjustment," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 3954–3963.

[19] Jianqing Zhang, Yang Hua, Hao Wang, Tao Song, Zhengui Xue, Ruhui Ma, and Haibing Guan, "Fedala: Adaptive local aggregation for personalized federated learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2023, vol. 37, pp. 11237–11244.

[20] Yiming Chen, Nan He, and Lifeng Sun, "Fedawa: Aggregation weight adjustment in federated domain generalization," in *2024 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2024, pp. 451–457.

[21] Zheyuan Zhang, Elif Keles, Gorkem Durak, Yavuz Taktak, Onkar Susladkar, Vandan Gorade, Debesh Jha, Asli C Ormeci, Alpay Medetalibeyoglu, Lanhong Yao, et al., "Large-scale multi-center ct and mri segmentation of pancreas with deep learning," *Medical Image Analysis*, vol. 99, pp. 103382, 2025.

[22] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al., "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[23] Vandan Gorade, Sparsh Mittal, Debesh Jha, and Ulas Bagci, "Synergynet: Bridging the gap between discrete and continuous representations for precise medical image segmentation," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 7768–7777.

[24] I Loshchilov, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.