FSF3A: Federated Spatial Feature Alignment and Adaptive Aggregation for Heterogeneous Brain Tumor Segmentation

Peketi Divya[1]
ai21resch01001@iith.ac.in
Krishna Mohan Chalavadi[2]
ckm@cse.iith.ac.in
Sobhan Babu[2]
sobhan@cse.iith.ac.in
Sumanth Yenduri[3]
sumanth@shsu.edu

[1] Department of AI, Indian Institute of Technology, Hyderabad
[2] Department of CSE, Indian Institute of Technology, Hyderabad
[3] Sam Houston State University, Texas, United States

Abstract

Deep learning has the potential to identify complex patterns in brain tumors, but privacy concerns hinder centralized approaches. Federated learning (FL) addresses privacy concerns by enabling collaborative model training without exchanging local client data. However, managing data heterogeneity in FL is challenging due to the varying distributions and data scales at the local clients. To address these problems, we propose FSF3A, a novel federated framework with two approaches: **(i)** *client localized structural loss (CLSL)*, a local training loss function that can handle varying data distributions by effectively finding a trade-off between spatial localization of local data, using modified dice loss, and the similarity between model representations to correct the local training of individual clients through our localized contrastive penalty and **(ii)** *adaptive scale-aware weighted aggregation (ASWA)* method which facilitates weighted aggregation at the global model, prioritizing clients based on the data scales to ensure fairness in representation. The efficacy of FSF3A is demonstrated through comprehensive experiments on UCSF PDGM, UPenn-GBM, and MSD datasets for brain tumor segmentation.

1 Introduction

Deep learning (DL) achieves cutting-edge performance in medical image segmentation by leveraging extensive labeled datasets [41], [37], [52], [13]. The absence of such datasets leads to notable degradation in DL performance [23]. However, medical data are typically isolated within individual medical centers and hospitals [54]. The aggregation of extensive data containing highly sensitive and private medical information in a single location is impractical and may even breach legal regulations. These challenges can be addressed with federated learning (FL), a practical machine learning approach in which distributed clients, such as individual medical institutions, collaboratively learn a shared model while maintaining the privacy of their raw data locally [14, 40, 43, 44]. Federated learning has become a significant area in machine learning, capturing substantial research attention [6, 11, 16, 19, 25, 28, 32, 36, 46]. Furthermore, federated learning has many applications in various domains, including medical imaging [24, 26], object detection [35], and landmark classification [18]. A significant challenge in federated learning arises from the heterogeneity of the data distribution between different clients [23]. In most real-world applications, the data among clients can be non-identically distributed, posing a challenge that can compromise the performance of federated learning [25, 30, 33]. Hence, we address this non-IID issue through a novel and intuitive observation in federated learning: the global model, functioning as a central federated aggregator, adeptly captures diverse patterns from varied data distributions and feature representations. To transfer the knowledge of the global model to local client models at the representation level, we introduce a localized contrastive penalty. This maximizes similarity between the current local and global models while enforcing dissimilarity with the previous local model, thereby promoting effective generalization. Building upon this, we integrate the localized contrastive penalty with the local client's modified dice loss and introduce a novel loss function named spatial-feature alignment loss (CLSL) for each client training. CLSL effectively balances the spatial localization of local data, using modified dice loss, with feature similarity using localized contrastive penalty.

As each client updates its local model, the local objective may diverge significantly from the global objective. When intermediate model updates are shared and aggregated at a central server, the averaged global model may drift from the global optimum. Several weighted aggregation studies have aimed to address this issue, one such work is FedProx [32], which places direct constraints on local updates using the L2-norm distance, and the other approach is SCAFFOLD [25] that addresses local updates through variance reduction [22]. For instance, Model-contrastive federated learning (MOON) incorporates model-level contrastive learning between the server model and client models to align the global representation learned by clients [29]. However, existing methods predominantly target image-level classification tasks, rather than dense prediction tasks like lesion segmentation, which are critical in medical image processing. Moreover, these methods neither consider data-scaled aggregation—crucial for prioritizing clients by dataset size—nor address model drift reduction through uncertainty handling, which prevents overfitting to local data and enhances global generalization. Thus, these methods lack weighted data-scaled aggregation, which is crucial for prioritizing individual client data weights based on their respective data sizes. This becomes particularly important in real-world scenarios where dealing with varying sizes of individual clients can impact the aggregation process. To address this, we propose ASWA, a server-side aggregation strategy that prioritizes clients by data scale while reducing global model drift from the optimum. This practical approach prioritizes clients with larger datasets, ensuring greater consideration in aggregation, thereby influencing the global model more by their distribution and minimizing the impact of fine-tuning with smaller datasets on its overall distribution. Our framework demonstrates improved performance over the other state-of-the-art's across different brain image segmentation datasets provided at three clients, namely UCSF-PDGM [4], UPenn-GBM [3], and MSD [2].

The following are the main contributions of this work:(i) We propose two novel approaches to address data heterogeneity in federated learning: (a) feature enhancements in the local training phase and (b) improvements to the aggregation phase, thereby tackling challenges arising from diverse brain tumor data sources across individual clients. (ii) Local client training uses the proposed client localized structural loss (CLSL) to manage data variability, balancing spatial localization with modified dice loss and aligning model representations through a localized contrastive penalty for improved local training, (iii) We propose an adaptive scale-aware weighted aggregation (ASWA) approach at the server in the federated framework that ensures weighted aggregation at the global model, prioritizing clients based on data scale to ensure fair representation, and (iv) We empirically demonstrate the effectiveness of the proposed methodology on three standard datasets, namely, UPenn-GBM (client-1), UCSF-PDGM (client-2), and MSD (client-3).

2 Related Work

To address data heterogeneity in federated learning, extensive research is being carried out, particularly in federated contrastive learning.

Federated Learning. FedAvg [36] is a notable aggregation approach where the server consolidates model weights through averaging to generate a global model for the subsequent training rounds. Several studies have aimed to enhance FedAvg's performance on non-IID data. FedProx [32] utilized a proximal term into the objective function during the local training process, constraining model updates by evaluating the L2-norm distance between local and global models. Scaffold [25] improves local updates by incorporating control variates and adjusting gradients based on differences between local and global control variates. However, both FedProx and Scaffold have mainly been evaluated on classification datasets, with limited exploration in segmentation tasks. FedMA [46] employs Bayesian non-parametric methods to match and average the weights layer-wise, while FedNova [49] normalizes the local updates before performing the averaging process. In addressing the challenge of model heterogeneity among clients, Li et al. [27] propose a method using knowledge distillation to enable communication between models. In this approach, the server gathers class scores from a public dataset across client models and averages them to form the updated consensus. Diao et al. [12] introduce an adaptive parameter allocation strategy, where a subset of global model parameters is dynamically allocated as local model parameters based on each client's capabilities. Finally, Liang et al. [34] present an algorithm that concurrently trains both the compact local representation and the global model of the client. FedMix [55] applies a mixup operation using average data from other clients and local data to approximate a global mixup. Meanwhile, FedFA [56] permits clients to extract new training samples from a universal statistic derived from all participants, aiming to alleviate feature drift specific to each client.

Federated Learning for Medical Image Segmentation. Decentralized medical datasets benefit from the potential of federated medical image segmentation while preserving privacy. Current research primarily focuses on addressing challenges related to data heterogeneity, including distribution shift [21, 31, 47, 48, 51], inconsistent ROIs [53], and inter-client unfairness [20]. The existing research focused on enhancing client training to address local drift by incorporating diverse strategies. These strategies involve employing data pre-processing to minimize data heterogeneity [44], a domain loss for classifying heterogeneous electroencephalography [15], and contrastive learning for volumetric medical image segmentation [50].

Contrastive Learning. Contrastive learning (CL) represents a self-supervised method for acquiring valuable visual representations by utilizing unlabeled data. CL approaches [8, 9, 17, 38] have demonstrated state-of-the-art results in learning visual representations. The core principle of CL involves minimizing the gap between representations of diverse augmented views of the same image (referred to as positive pairs) and simultaneously maximizing the gap between representations of augmented views belonging to different images (referred to as negative pairs). An exemplary contrastive learning frameworks are Local CL [7], SimCLR [8], infoNCE [39], and SwAV [6]. Li et al. [29] introduced MOON as a method to address the challenge of heterogeneous local data distributions among participants. Recent work, relaxed contrastive learning (RCL), enhances federated learning by mitigating gradient inconsistency across clients and preventing representation collapse [42]. However, these works primarily target image-level classification tasks, making them less suited for segmentation tasks where pixel-level spatial and semantic details are essential. Thus, client drift in FL segmentation remains an open challenge, and existing approaches continue to underperform in this setting (see Sec. 4). To address this, we designed a localized contrastive penalty that captures 3D feature patterns, unlike traditional methods. In this work, we simultaneously address local (using CLSL) and global (using ASWA) strategies in federated learning to effectively handle heterogeneous data sourced from entirely distinct brain tumor data sources on individual clients.

3 Methodology

Our federated framework (Figure 1) addresses non-IID data using the global model as a central aggregator of diverse patterns. We introduce a localized contrastive penalty (L_{LCP}) for knowledge transfer to local models, enforcing spatial structure consistency (L_{SSC}) and forming a client-localized structural loss (L_{CLSL}). To balance client data scales, we propose adaptive scale-aware weighted aggregation (ASWA). The FSF3A framework employs 3D U-Net [10] for segmentation, with detailed objectives and aggregation described in subsequent sections.

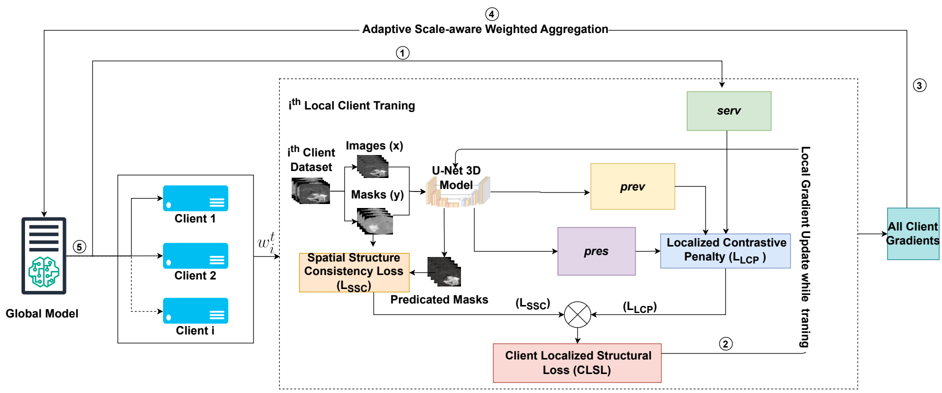

Figure 1: Overview of FSF3A. ① Global model broadcast, ② local training with CLSL, ③ updates sent to the server, ④ ASWA aggregates by dataset size, ⑤ global model redistributed.

Client Localized Structural Loss (CLSL). We propose a client localized structural loss (CLSL) function (Eq. 1) that integrates spatial structure consistency with a localized contrastive penalty. This formulation is designed to simultaneously capture fine-grained spatial features and feature-based representations from client data, enabling a robust segmentation process. The CLSL function is expressed as:

$$\mathcal{L}_{\text{CLSL}} = \mathcal{L}_{\text{SSC}} + \gamma \cdot \mathcal{L}_{\text{LCP,}} \tag{1}$$

where γ are coefficients that determine the relative contributions of contrastive terms. This balance ensures that the local model learns spatially consistent segmentations while preserving alignment with the global model.

Spatial Structure Consistency Loss (L_{SSC}). To ensure accurate spatial boundaries, crucial for tasks like medical imaging, we employ a spatial structure consistency loss. This loss function combines modified dice loss and gradient alignment penalty to preserve shape fidelity and align spatial features in the predicted segmentation with the ground truth.

Modified Dice Loss. The Dice loss quantifies the overlap between the predicted segmentation (y_{pred}) and the ground truth mask (y_{true}), making it well-suited for tasks with class imbalances by emphasizing accurate segmentation of regions of interest. Added upon this, our modified dice loss measures pixel-wise intersections of the original mask and predicted mask in i,j, and k spatial dimensions as shown below:

$$L_{\text{Dice}} = 1 - \frac{2\sum_{i,j,k}\left(y_{\text{true}}^{(i,j,k)} \cdot y_{\text{pred}}^{(i,j,k)}\right)}{\sum_{i,j,k}\left(y_{\text{true}}^{(i,j,k)} + y_{\text{pred}}^{(i,j,k)}\right) + \varepsilon}. \tag{2}$$

The term $y_{\text{true}}^{(i,j,k)}$ denotes the ground truth binary mask, while $y_{\text{pred}}^{(i,j,k)}$ refers to the predicted binary mask. Also, ε (typically set to a small value such as 10^{-5}) is included to prevent division by zero. This adjustment helps stabilize the training process and ensures that the loss remains non-zero, even in cases where the intersection between predicted and ground truth masks is minimal.

Gradient Alignment Penalty. Dice loss measures the overlap between the predicted mask and the ground truth mask. To improve spatial consistency, we incorporate a gradient alignment term that enforces precise alignment of the predicted segmentation boundaries (y_{pred}) with those of the ground truth mask (y_{true}). Gradients are calculated along each spatial axis (e.g., x, y, and z), reinforcing boundary accuracy in the segmentation. To ensure alignment between the spatial gradients of y_{true} and y_{pred}, we incorporate a gradient consistency component evaluates gradients along each axis in 3D volumes, promoting accurate boundary delineation in the segmentation. This approach not only enhances the spatial precision of the predicted mask but also reinforces the alignment of edges and boundaries with the ground truth. For instance, gradients along the x-axis are computed as:

$$\nabla_x y_{\text{true}}^{(i,j,k)} = y_{\text{true}}^{(i+1,j,k)} - y_{\text{true}}^{(i,j,k)}, \nabla_x y_{\text{pred}}^{(i,j,k)} = y_{\text{pred}}^{(i+1,j,k)} - y_{\text{pred}}^{(i,j,k)} \tag{3}$$

Similar calculations are performed for the y and z axes to ensure comprehensive gradient alignment across all spatial dimensions.

The mean squared error (MSE) between the gradients of y_{true} and y_{pred} is used to calculate the gradient consistency loss for each spatial axis as shown below :

$$L_{\text{grad},x} = \frac{1}{N}\sum_{i,j,k}\left(\nabla_x y_{\text{true}}^{(i,j,k)} - \nabla_x y_{\text{pred}}^{(i,j,k)}\right)^2 \tag{4}$$

$$L_{\text{grad},y} = \frac{1}{N}\sum_{i,j,k}\left(\nabla_y y_{\text{true}}^{(i,j,k)} - \nabla_y y_{\text{pred}}^{(i,j,k)}\right)^2 \tag{5}$$

$$L_{\text{grad},z} = \frac{1}{N} \sum_{i,j,k} \left(\nabla_z y_{\text{true}}^{(i,j,k)} - \nabla_z y_{\text{pred}}^{(i,j,k)} \right)^2 \tag{6}$$

This ensures precise alignment of gradients and strengthens boundary delineation. The overall gradient consistency loss is derived by taking the average of the individual losses computed along the x, y, and z, ensuring a balanced alignment of spatial gradients across all dimensions:

$$L_{\text{grad}} = \frac{L_{\text{grad},x} + L_{\text{grad},y} + L_{\text{grad},z}}{3} \tag{7}$$

By minimizing L_{grad} loss during training, the model is encouraged to produce segmentation with highly accurate spatial boundaries and consistent edge alignments, improving both the quality and reliability of the predictions. The spatial structure consistency loss (L_{ssc}) is a combined loss that ptimize both region-wise overlap and boundary alignment. It combines the Dice loss, which measures volumetric similarity, with the gradient consistency loss, which enforces spatial boundary coherence. The formulation is expressed as:

$$L_{\text{spatial}} = L_{\text{Dice}} + \omega \cdot L_{\text{grad}} \tag{8}$$

In this equation, the edge weight ω serves as a scaling factor to balance the influence of the gradient consistency loss relative to the Dice loss, enabling precise control over boundary refinement during optimization.

Localized Contrastive Penalty (L_{LCP}). Federated learning faces the challenge of balancing global model alignment with preserving local model diversity. Clients need to update their models consistently across epochs while adapting to their unique data, ensuring both global convergence and local specificity. To address this, contrastive learning is employed at the patch level, allowing the model to emphasize localized features while achieving alignment. This involves comparing the current local model representation (*pres*) and the previous local model representation (*prev*) with the global model representation (*serv*). This comparison forms positive and negative patch pairs. Minimizing contrastive loss trains the model to align similar features (positive pairs) while preserving distinct client-specific characteristics by distinguishing them from global features (negative pairs). Patch-based contrastive learning facilitates localized feature capture, enabling fine-grained feature learning for segmentation tasks. By minimizing interference between global and local features, this approach effectively preserves client-specific characteristics.

Patch Extraction: At each depth slice d, patches are extracted from the feature maps to encapsulate fine-grained spatial information. These patches are derived from the representations of the previous model state (prev), the current model state (pres), and the global model (serv), facilitating the alignment of the temporal and global features while preserving spatial locality. To achieve this, a slicing mechanism is employed to partition the feature maps at each depth into smaller, localized regions, enabling the extraction of detailed spatial information. For a patch size of $s \times s$, each representation is partitioned as $\text{patch}(d,i,j) = \text{prev}[d, i : i+s, j : j+s]$ where prev represents the local model's feature representation from the preceding iteration, d denotes the depth slice of the feature map, and (i, j) identifies the coordinates of the top-left corner of the extracted patch. The same slicing mechanism is applied to the representations of the current model state (pres) and the global model (serv), ensuring effective alignment and differentiation during the contrastive learning process.

Similarity computations. In this work, patches extracted from the local client model's representation at two different temporal states, specifically, the previous epoch's representation (*prev*) and the current epoch's representation (*pres*), are treated as positive pairs. The rationale behind this pairing is rooted in the temporal continuity of learned features: *prev* encapsulates the knowledge acquired in the preceding epoch, while *pres* embodies the updated representation based on the same local client data. By establishing these as positive pairs, the model is incentivized to maintain temporal consistency, which fosters the preservation of critical features over successive updates while allowing for incremental refinements.

Conversely, patches derived from the previous epoch's representation (*prev*) and the global server model's representation (*serv*) are designated as negative pairs. This approach emphasizes the contrast between local and global feature distributions. The *serv* representation encapsulates the aggregated knowledge across all clients and, therefore, serves as a proxy for global patterns. By treating *prev* and *serv* as negatives, the model is encouraged to differentiate between client-specific nuances and the overarching global trends. This differentiation mechanism is pivotal in ensuring that the model retains the unique characteristics of local client data without succumbing to excessive generalization. The combined use of temporal consistency through positive pairings and feature differentiation via negative pairings yields a dual benefit. It preserves the granularity of client-specific data, ensuring adaptability to diverse and heterogeneous data distributions, while simultaneously leveraging the global model to guide alignment without compromising local specificity. The positive similarity between patches is defined as:

$$\mathrm{sim}_{\mathrm{pos}}^{(d,i,j)} = \exp\left(\frac{\mathrm{patch}_{\mathrm{prev}}(d,i,j)\cdot\mathrm{patch}_{\mathrm{pres}}(d,i,j)}{\tau}\right) \tag{9}$$

where τ is the temperature parameter regulating the sharpness of the similarity distribution. Similarly, the negative similarity is computed as:

$$\mathrm{sim}_{\mathrm{neg}}^{(d,i,j)} = \exp\left(\frac{\mathrm{patch}_{\mathrm{prev}}(d,i,j)\cdot\mathrm{patch}_{\mathrm{serv}}(d,i,j)}{\tau}\right) \tag{10}$$

These similarity metrics facilitate the contrastive learning objective, encouraging the alignment of positive pairs while promoting separation between negative pairs. Finally, the patch-based localized contrastive penalty for a given patch, combined with the aggregation over all spatial regions and depth slices, is computed as:

$$\mathcal{L}_{\mathrm{LCP}} = -\frac{1}{D\times M\times N}\sum_{d,i,j}\log\left(\frac{\mathrm{sim}_{\mathrm{pos}}^{(d,i,j)}}{\mathrm{sim}_{\mathrm{pos}}^{(d,i,j)}+\mathrm{sim}_{\mathrm{neg}}^{(d,i,j)}}\right), \tag{11}$$

where D is the total number of depth slices, and M and N represent the number of patches in each spatial direction. The term $\mathrm{sim}_{\mathrm{pos}}^{(d,i,j)}$ and $\mathrm{sim}_{\mathrm{neg}}^{(d,i,j)}$ denote the positive and negative similarities for the patch at depth d and coordinates (i,j). This ensures that the contrastive loss enforces consistency and alignment across all depth slices and spatial regions, fostering robust feature learning while maintaining the model's ability to adapt to client-specific data distributions in the federated setting. By aggregating the loss over all spatial regions and depth slices, the model learns to preserve both global coherence and localized specificity.

Adaptive Scale-aware Weighted Aggregation (ASWA). ASWA weighted data-scaled aggregation method prioritizes clients based on their data scales, ensuring that those with larger datasets receive greater consideration in the aggregation process. By prioritizing larger datasets, the global model is more influenced by their distribution, minimizing the impact of fine-tuning with smaller datasets on its overall distribution. This aggregation is performed on the server side by considering the ratio of varying sizes of client datasets (n_i) to the total dataset size across all clients (N_{total}), along with their corresponding weights (w_i^t) as shown below:

$$w^T = \sum_{i=1}^{N} \frac{n_i}{N_{total}} w_i^t. \tag{12}$$

Workflow of FSF3A. At each round, the global model is sent to clients, which update it locally using CLSL to enforce spatial consistency and feature alignment. The updates are then returned to the server, where ASWA combines them in a scale-aware manner. The new global model is redistributed, coupling CLSL at the client side with ASWA at the server side for robust federated optimization.

4 Experiments

Implementation Details and Baselines: We conduct experiments on three public datasets: UCSF-PDGM [4], UPenn-GBM [3], and MSD [2] (see Appendix A2 for dataset details). Each is randomly split 75/15/10 for train/validation/test, and all volumes are resampled to $128 \times 128 \times 128$ for computational feasibility. While higher resolutions (e.g., 256^3) preserve more fine-grained details, pilot experiments showed only marginal improvements (+0.01 Dice) at the cost of > 2× GPU memory. Thus, 128^3 represents a practical trade-off between accuracy and efficiency in federated settings. We adopt this multi-site setup to capture both scale and distribution heterogeneity. Each client uses a 3D U-Net backbone and we extract feature vectors from the final encoder convolution for all methods. All models (FSF3A + six baselines) are implemented in TensorFlow on an NVIDIA V100 GPU. We train for 120 communication rounds, with 20 local epochs per round, using the Adam optimizer (learning rate 1×10^{-5}, $\beta_1 = 0.9, \beta_2 = 0.999$) and batch size 8. The contrastive temperature and gradient-weight coefficient (ω) are set to 0.5. Segmentation performance is measured by three-fold cross-validation at each client. In each fold, 75% of samples are used for training, 15% for validation, and 10% for testing. We report Dice coefficient, intersection-over-union (IoU), precision, and sensitivity.
Baselines: To evaluate our client Localized structural loss (CLSL), we embed six state-of-the-art loss functions into the same federated loop with our Adaptive Scale-Aware Weighted Aggregation (ASWA) and compare their segmentation performance in Table 1. These include SimCLR [8] and SwAV [6] for contrastive pre-training (FedSimCLR, FedSwAV), Local CL [7] for client-level contrastive learning (FedLocal-CL), InfoNCE [39] for mutual-information maximization (FedInfoNCE), Dice Loss [45] for overlap-based segmentation (FedDice), and Focal Tversky Loss [1] for class-imbalance handling (FedFoc-Tvers). To isolate the benefit of ASWA itself, we further compare it against five standard server aggregators—FedAvg [36], FedProx [32], SCAFFOLD [25], FedNova [46], and FedMA [49] in Table 3.
Results. Table 1 presents the 120-round results for each client dataset. On UPenn-GBM, FSF3A reaches 0.92 versus 0.77 for SimCLR (15 percentage points higher). On MSD, FSF3A scores 0.91 compared with 0.77 for InfoNCE (14 percentage points higher). FSF3A also records the highest intersection-over-union, precision, and sensitivity in all cases, with absolute gains of 0.03 to 0.15 over the strongest competing methods. These consistent improvements confirm the effectiveness of combining CLSL and ASWA for federated segmentation on heterogeneous data. These gains arise because CLSL produces more consistent segmentations at each client by enforcing both spatial boundary alignment and feature-level

Table 1: Comparison of FSF3A with baselines using ASWA as the server aggregator, evaluated at 40, 80, and 120 communication rounds (CR) with 3-fold cross-validation on UCSF PDGM, UPenn-GBM, and MSD datasets. Standard deviations across folds range from 0.01 to 0.06.

Data	Baselines	CR = 40				CR = 80				CR = 120			
		Dice ↑	IoU ↑	Prec. ↑	Sens. ↑	Dice ↑	IoU ↑	Prec. ↑	Sens. ↑	Dice ↑	IoU ↑	Prec. ↑	Sens. ↑
UCSF PDGM	FedDice	0.22	0.12	0.15	0.38	0.64	0.52	0.57	0.61	0.77	0.62	0.80	0.72
	FedInfoNCE	0.56	0.39	**0.63**	0.51	0.60	0.42	0.62	0.63	0.70	0.53	0.64	0.74
	FedSimCLR	**0.59**	0.43	0.58	0.62	0.65	0.44	0.58	0.61	0.75	0.63	0.68	0.68
	FedFoc-Tvers	0.25	0.38	0.15	**0.73**	0.64	0.47	0.57	0.74	0.75	0.58	0.61	**0.88**
	FedLocal-CL	0.32	0.28	0.40	0.22	0.65	0.52	**0.62**	0.58	0.73	0.69	0.77	0.69
	FedSwAV	0.48	**0.51**	0.54	0.61	0.62	0.60	0.58	0.64	0.73	0.68	0.64	0.71
	FSF3A	0.57	0.43	0.60	0.59	**0.75**	**0.62**	0.78	**0.75**	**0.86**	**0.77**	**0.86**	0.85
UPenn-GBM	FedDice	0.51	0.35	0.52	0.50	0.67	0.55	0,64	0.63	0.75	0.62	0.76	0.77
	FedInfoNCE	0.66	**0.53**	0.67	0.70	0.67	0.51	0.73	0.69	0.63	0.55	0.76	0.66
	FedSimCLR	**0.68**	0.51	**0.69**	0.70	0.69	0.54	0.74	0.66	0.73	0.58	0.79	0.73
	FedFoc-Tvers	0.36	0.22	0.25	0.67	0.70	0.55	0.61	**0.85**	0.77	0.62	0.67	**0.89**
	FedLocal-CL	0.51	0.42	0.42	0.20	0.59	0.55	0.61	0.49	0.75	0.63	0.74	0.67
	FedSwAV	0.53	0.39	0.58	0.55	0.64	0.56	0.64	0.68	0.70	0.59	0.68	0.73
	FSF3A	0.69	0.56	0.73	**0.70**	**0.81**	**0.70**	**0.83**	0.82	**0.92**	**0.82**	**0.90**	0.88
MSD	FedDice	0.38	0.22	0.28	0.62	0.66	0.49	0.62	0.71	0.79	0.65	0.79	0.81
	FedInfoNCE	0.65	0.59	0.74	0.73	0.70	0.66	0.79	0.78	0.77	0.73	0.82	0.87
	FedSimCLR	0.66	**0.69**	**0.78**	**0.80**	0.69	0.65	0.77	**0.83**	0.77	0.72	0.86	0.85
	FedFoc-Tvers	0.34	0.21	0.22	0.66	0.39	0.24	0.33	0.77	0.67	0.51	0.53	**0.97**
	FedLocal-CL	0.38	0.42	0.36	0.51	0.63	0.53	0.59	0.66	0.79	0.68	0.72	0.83
	FedSwAV	0.50	0.43	0.52	0.57	0.61	0.51	0.62	0.67	0.65	0.58	0.66	0.71
	FSF3A	**0.71**	0.57	0.76	0.69	**0.82**	**0.73**	**0.84**	0.80	**0.91**	**0.84**	**0.93**	0.90

consistency, while ASWA ensures that larger, more reliable datasets exert appropriate influence during global aggregation. Together, these components reduce client drift and improve overall model convergence under heterogeneous data.

Table 2 presents a performance comparison between the proposed FSF3A framework and state-of-the-art (SOTA) methods such as MOON, FedMix, and FedRCL on UCSF-PDGM, UPenn-GBM, and MSD datasets at each client. FSF3A outperforms these methods with higher robustness and accuracy in handling heterogeneous medical image segmentation. MOON's model-level contrastive learning benefits classification but lacks pixel-level precision for segmentation. FedMix's mixup blurs spatial detail and structural accuracy, while FedRCL improves feature diversity but omits scale-adaptive aggregation and precise pixel alignment. FSF3A addresses these challenges by uniting scale-adaptive aggregation with a contrastive segmentation loss, yielding superior performance across data scales while enhancing client-localized structural learning for medical image segmentation.

Table 2: FSF3A vs SOTA methods at each client.

Method	UCSF-PDGM	UPenn-GBM	MSD
MOON	0.74 ± 0.01	0.76 ± 0.02	0.77 ± 0.03
FedMix	0.79 ± 0.02	0.80 ± 0.04	0.81 ± 0.02
FedRCL	0.71 ± 0.02	0.72 ± 0.04	0.78 ± 0.02
FSF3A	**0.86 ± 0.03**	**0.92 ± 0.03**	**0.91 ± 0.01**

4.1 More Empirical Analysis and Ablation Studies

We compare several server aggregation methods on three client's merged test data and observed that our adaptive scale-aware weighted aggregation (ASWA) method outperforms other baselines with a mean dice coefficient of 0.78 as shown in Table

Table 3: Aggregation methods on server performance.

Method	Dice ↑	IoU ↑	Prec. ↑	Sens. ↑
Fedavg	0.50±0.04	0.36±0.07	0.65±0.01	0.58±0.02
Fedprox	0.69±0.01	0.61±0.02	0.70±0.02	**0.86±0.04**
Scaffold	0.73±0.02	0.57±0.02	0.74±0.03	0.73±0.01
FedNova	0.77±0.02	0.63±0.01	0.73±0.02	0.71±0.02
FedMA	0.76±0.03	0.64±0.05	0.78±0.01	0.81±0.02
ASWA	**0.78±0.03**	**0.65±0.03**	**0.77±0.02**	0.81±0.02

3.

Compared to Fedavg, ASWA demonstrates an impressive 28% and 13% improvement over Fedprox, highlighting its efficacy in enhancing segmentation outcomes. Compared to other methods such as Scaffold, FedNova, and FedMA, ASWA maintains consistent improvements of 5%, 1%, and 2%, respectively. During experiments, we noticed that Fedavg and Fedprox may not be effective in dealing with the heterogeneity of data across different clients as they treat all clients equally during aggregation, which leads to sub-optimal performance in the FSF3A setup, which has varied data distribution across clients. On the other hand, Scaffold, FedNova, and FedMA incorporate more sophisticated mechanisms to handle data heterogeneity, such as using control variates or momentum terms. However, these methods fall short under highly skewed data distributions, as they fail to fully leverage larger client datasets, yielding a less accurate global model. Unlike other aggregation methods, ASWA prioritizes larger datasets, reducing bias from smaller clients and enabling the global model to capture the overall distribution for more precise segmentation.

To assess the contributions of FSF3A's core components, ASWA and CLSL, we performed an ablation study (Table 4). First, ASWA was used with a basic Dice loss for local training. The resulting dice coefficients are 77% for UCSF PDGM, 75% for UPenn-GBM, and 79% for MSD. Subsequently, we employed only CLSL as the client training loss and replaced ASWA with FedAvg. This produced dice coefficients of 57% for UCSF PDGM, 53% for UPenn-GBM, and 56% for MSD. Finally, within the FSF3A framework, we employed both CLSL for local client training and ASWA for server aggregation. This approach achieved high dice coefficients of 86% for UCSF PDGM, 92% for UPenn-GBM, and 91% for MSD. These results underscore the importance of CLSL and ASWA, which together enable client-localized structural learning and scale-aware aggregation across heterogeneous data for efficient brain tumor segmentation.

Table 4: Ablation results of different components of FSF3A.

Approach	Data	Dice ↑	IoU ↑	Prec. ↑	Sens. ↑
ASWA	UCSF PDGM	0.77±0.03	0.62±0.06	0.80±0.02	0.72±0.02
	UPenn-GBM	0.75±0.01	0.62±0.03	0.76±0.03	0.77±0.02
	MSD	0.79±0.04	0.65±0.02	0.79±0.04	0.81±0.03
CLSL	UCSF PDGM	0.57±0.04	0.48±0.04	0.62±0.02	0.68±0.01
	UPenn-GBM	0.53±0.02	0.41±0.02	0.59±0.03	0.52±0.02
	MSD	0.56±0.03	0.51±0.03	0.62±0.02	0.62±0.03
FSF3A	UCSF PDGM	**0.86±0.03**	**0.77±0.04**	**0.86±0.04**	**0.85±0.04**
	UPenn-GBM	**0.92±0.03**	**0.82±0.03**	**0.90±0.01**	**0.88±0.02**
	MSD	**0.91±0.01**	**0.84±0.05**	**0.93±0.02**	**0.90±0.03**

5 Conclusion

This work aims to propose a novel federated framework, incorporating the spatial-feature alignment loss (CLSL) and the adaptive scale-aware weighted aggregation (ASWA) method, which demonstrates better efficacy in brain tumor segmentation. CLSL effectively handles varying data distributions by balancing spatial localization and model representation similarity. ASWA ensures fairness by prioritizing clients based on data scale. Our approach achieves improved segmentation performance through comprehensive experiments on UPenn-GBM, UCSF PDGM, and MSD datasets while preserving privacy and addressing data heterogeneity. Federated learning emerges as a promising avenue for advancing brain tumor segmentation while overcoming challenges associated with decentralized data. FSF3A may be less effective for extremely small client datasets, where ASWA down-weights their contribution, and for highly imbalanced classes beyond Dice supervision. Future work will investigate robustness under these conditions and extend FSF3A to other modalities beyond brain MRI.

References

[1] Nabila Abraham and Naimul Mefraz Khan. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 683–687. IEEE, 2019.

[2] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022.

[3] Spyridon Bakas, Chiharu Sako, Hamed Akbari, Michel Bilello, Aristeidis Sotiras, Gaurav Shukla, Jeffrey D Rudie, Natali Flores Santamaría, Anahita Fathi Kazerooni, Sarthak Pati, et al. The university of pennsylvania glioblastoma (upenn-gbm) cohort: Advanced mri, clinical, genomics, & radiomics. *Scientific data*, 9(1):453, 2022.

[4] Evan Calabrese, Javier E Villanueva-Meyer, Jeffrey D Rudie, Andreas M Rauschecker, Ujjwal Baid, Spyridon Bakas, Soonmee Cha, John T Mongan, and Christopher P Hess. The university of california san francisco preoperative diffuse glioma mri dataset. *Radiology: Artificial Intelligence*, 4(6):e220058, 2022.

[5] Sebastian Caldas, Sai Meher Karthik Duddu, Peter Wu, Tian Li, Jakub Konečnỳ, H Brendan McMahan, Virginia Smith, and Ameet Talwalkar. Leaf: A benchmark for federated settings. *arXiv preprint arXiv:1812.01097*, 2018.

[6] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33:9912–9924, 2020.

[7] Krishna Chaitanya, Ertunc Erdil, Neerav Karani, and Ender Konukoglu. Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in neural information processing systems*, 33:12546–12558, 2020.

[8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[9] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems*, 33:22243–22255, 2020.

[10] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[11] Zhongxiang Dai, Bryan Kian Hsiang Low, and Patrick Jaillet. Federated bayesian optimization via thompson sampling. *Advances in Neural Information Processing Systems*, 33:9687–9699, 2020.

[12] Enmao Diao, Jie Ding, and Vahid Tarokh. Heterofl: Computation and communication efficient federated learning for heterogeneous clients. *arXiv preprint arXiv:2010.01264*, 2020.

[13] Hao Dong, Guang Yang, Fangde Liu, Yuanhan Mo, and Yike Guo. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In *Medical Image Understanding and Analysis: 21st Annual Conference, MIUA 2017, Edinburgh, UK, July 11–13, 2017, Proceedings 21*, pages 506–517. Springer, 2017.

[14] Qi Dou, Tiffany Y So, Meirui Jiang, Quande Liu, Varut Vardhanabhuti, Georgios Kaissis, Zeju Li, Weixin Si, Heather HC Lee, Kevin Yu, et al. Federated deep learning for detecting covid-19 lung abnormalities in ct: a privacy-preserving multinational validation study. *NPJ digital medicine*, 4(1):60, 2021.

[15] Dashan Gao, Ce Ju, Xiguang Wei, Yang Liu, Tianjian Chen, and Qiang Yang. Hhhfl: Hierarchical heterogeneous horizontal federated learning for electroencephalography. *arXiv preprint arXiv:1909.05784*, 2019.

[16] Chaoyang He, Songze Li, Jinhyun So, Xiao Zeng, Mi Zhang, Hongyi Wang, Xiaoyang Wang, Praneeth Vepakomma, Abhishek Singh, Hang Qiu, et al. Fedml: A research library and benchmark for federated machine learning. *arXiv preprint arXiv:2007.13518*, 2020.

[17] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[18] Tzu-Ming Harry Hsu, Hang Qi, and Matthew Brown. Federated visual classification with real-world data distribution. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part X 16*, pages 76–92. Springer, 2020.

[19] Sixu Hu, Yuan Li, Xu Liu, Qinbin Li, Zhaomin Wu, and Bingsheng He. The oarf benchmark suite: Characterization and implications for federated learning systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13(4):1–32, 2022.

[20] 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*, pages 16302–16311, 2023.

[21] Meirui Jiang, Hongzheng Yang, Chen Cheng, and Qi Dou. Iop-fl: Inside-outside personalization for federated medical image segmentation. *IEEE Transactions on Medical Imaging*, 2023.

[22] Rie Johnson and Tong Zhang. Accelerating stochastic gradient descent using predictive variance reduction. *Advances in neural information processing systems*, 26, 2013.

[23] Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210, 2021.

[24] Georgios A Kaissis, Marcus R Makowski, Daniel Rückert, and Rickmer F Braren. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence*, 2(6):305–311, 2020.

[25] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, pages 5132–5143. PMLR, 2020.

[26] Rajesh Kumar, Abdullah Aman Khan, Jay Kumar, Noorbakhsh Amiri Golilarz, Simin Zhang, Yang Ting, Chengyu Zheng, Wenyong Wang, et al. Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging. *IEEE Sensors Journal*, 21(14):16301–16314, 2021.

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

[28] Qinbin Li, Zeyi Wen, and Bingsheng He. Practical federated gradient boosting decision trees. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 4642–4649, 2020.

[29] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10713–10722, 2021.

[30] Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. Federated learning on non-iid data silos: An experimental study. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 965–978. IEEE, 2022.

[31] Shuailin Li, Zhitong Gao, and Xuming He. Superpixel-guided iterative learning from noisy labels for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 525–535. Springer, 2021.

[32] 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*, 2:429–450, 2020.

[33] Xiang Li, Kaixuan Huang, Wenhao Yang, Shusen Wang, and Zhihua Zhang. On the convergence of fedavg on non-iid data. *arXiv preprint arXiv:1907.02189*, 2019.

[34] Paul Pu Liang, Terrance Liu, Liu Ziyin, Nicholas B Allen, Randy P Auerbach, David Brent, Ruslan Salakhutdinov, and Louis-Philippe Morency. Think locally, act globally: Federated learning with local and global representations. *arXiv preprint arXiv:2001.01523*, 2020.

[35] Yang Liu, Anbu Huang, Yun Luo, He Huang, Youzhi Liu, Yuanyuan Chen, Lican Feng, Tianjian Chen, Han Yu, and Qiang Yang. Fedvision: An online visual object detection platform powered by federated learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 13172–13179, 2020.

[36] 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*, pages 1273–1282. PMLR, 2017.

[37] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[38] Ishan Misra and Laurens van der Maaten. Self-supervised learning of pretext-invariant representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6707–6717, 2020.

[39] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[40] Nicola Rieke, Jonny Hancox, Wenqi Li, Fausto Milletari, Holger R Roth, Shadi Albarqouni, Spyridon Bakas, Mathieu N Galtier, Bennett A Landman, Klaus Maier-Hein, et al. The future of digital health with federated learning. *NPJ digital medicine*, 3(1): 119, 2020.

[41] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 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*, pages 234–241. Springer, 2015.

[42] Seonguk Seo, Jinkyu Kim, Geeho Kim, and Bohyung Han. Relaxed contrastive learning for federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12279–12288, 2024.

[43] Micah J Sheller, G Anthony Reina, Brandon Edwards, Jason Martin, and Spyridon Bakas. Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, pages 92–104. Springer, 2019.

[44] Micah J Sheller, Brandon Edwards, G Anthony Reina, Jason Martin, Sarthak Pati, Aikaterini Kotrotsou, Mikhail Milchenko, Weilin Xu, Daniel Marcus, Rivka R Colen, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, 10(1):12598, 2020.

[45] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017.

[46] Hongyi Wang, Mikhail Yurochkin, Yuekai Sun, Dimitris Papailiopoulos, and Yasaman Khazaeni. Federated learning with matched averaging. *arXiv preprint arXiv:2002.06440*, 2020.

[47] Jiacheng Wang, Yueming Jin, and Liansheng Wang. Personalizing federated medical image segmentation via local calibration. In *European Conference on Computer Vision*, pages 456–472. Springer, 2022.

[48] Jiacheng Wang, Yueming Jin, Danail Stoyanov, and Liansheng Wang. Feddp: Dual personalization in federated medical image segmentation. *IEEE Transactions on Medical Imaging*, 2023.

[49] Jianyu Wang, Qinghua Liu, Hao Liang, Gauri Joshi, and H Vincent Poor. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in neural information processing systems*, 33:7611–7623, 2020.

[50] Yawen Wu, Dewen Zeng, Zhepeng Wang, Yiyu Shi, and Jingtong Hu. Federated contrastive learning for volumetric medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*, pages 367–377. Springer, 2021.

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

[52] Xiaowei Xu, Tianchen Wang, Yiyu Shi, Haiyun Yuan, Qianjun Jia, Meiping Huang, and Jian Zhuang. Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pages 477–485. Springer, 2019.

[53] Xuanang Xu, Hannah H Deng, Jamie Gateno, and Pingkun Yan. Federated multi-organ segmentation with inconsistent labels. *IEEE Transactions on Medical Imaging*, 2023.

[54] Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2):1–19, 2019.

[55] Tehrim Yoon, Sumin Shin, Sung Ju Hwang, and Eunho Yang. Fedmix: Approximation of mixup under mean augmented federated learning. *arXiv preprint arXiv:2107.00233*, 2021.

[56] Tianfei Zhou and Ender Konukoglu. Fedfa: Federated feature augmentation. *arXiv preprint arXiv:2301.12995*, 2023.