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

Peketi Divya, Krishna Mohan Chalavadi, Sobhan Babu, Sumanth Yenduri

भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

BMVC 2025

Introduction

- Federated medical segmentation operates on siloed MRIs; severe non-IID client distributions and unequal data scales induce local drift and aggregation bias in the global model.
- Prior remedies (FedProx, SCAFFOLD, MOON) largely target classification and are not scale-aware nor spatially structured, which limits effectiveness for dense prediction (lesion segmentation).
- FSF3A: CLSL (modified Dice + localized contrastive alignment to the global model, repelling the prior local state) + ASWA (adaptive scale-aware weighting) stabilizes optimization under heterogeneity and improves robustness on heterogeneous brain tumor data such as UPenn-GBM, UCSF-PDGM, MSD.

Motivation

- Federated brain-tumor segmentation (dense prediction) is challenged by non-IID client distributions (scanner/protocol/population shifts), data-scale imbalance, causing client drift, and unstable global convergence compared with classification settings.
- Prior FL remedies do not explicitly enforce spatial structure or voxel-wise alignment, leaving segmentation vulnerable to boundary errors, class imbalance, and representation mismatch across sites.
- We target local drift with client localized structural loss (CLSL), which couples a modified Dice term for spatial localization with a localized contrastive penalty that aligns the client's current features to the server (global) encoder while pushing away the client's previous state, improving cross-site representation consistency.
- The contrastive component operates spatially (region/voxel neighborhoods as positives) to preserve anatomical locality; negatives include historical local embeddings, stabilizing optimization without requiring raw data sharing.
- We address aggregation bias with Adaptive Scale-aware Weighted Aggregation (ASWA), assigning client weights proportional to effective data scale which improves fairness and round-to-round stability.
- Together, CLSL + ASWA reduce feature drift and aggregation variance, yielding consistent gains on UPenn-GBM, UCSF-PDGM, and MSD under strong heterogeneity.

FSF3A: Methodology

- Each round: Server broadcasts global model → Clients train locally with CLSL → Clients return updates → Server aggregates with ASWA → New global model redistributed.
- Backbone: 3D U-Net for segmentation. Improvements come from the loss (CLSL) and aggregation (ASWA), not architecture changes.

Client Localized Structural Loss (CLSL):

It balances precise mask boundaries with global feature alignment by combining a spatial structure term and a localized contrastive term.

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

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** enforce **volumetric overlap** and sharp, **anatomically aligned boundaries** by combining a soft Dice term with a 3D gradient-alignment penalty.

$$\mathcal{L}_{\text{SSC}} = L_{\text{Dice}} + \omega \cdot L_{\text{grad}}$$

- **Modified Dice Loss:** An ε-smoothed **3D soft-Dice** objective that computes voxel-wise overlap across ((i,j,k)) between (y_true) and probabilistic (y_(pred), mitigating class imbalance and stabilizing gradients when overlap is small: it replaces hard intersections with summed products and adds (\epsilon) to avoid division by zero.

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

- **Gradient Alignment Penalty:** To complement Dice overlap with sharper boundaries, we penalize mismatches between the spatial gradients of the prediction and ground truth. Let forward finite differences define voxel-wise gradients along each axis:

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

with symmetric/zero padding at borders. The per-axis gradient consistency losses use voxelwise MSE:

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

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

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

The overall penalty averages the axes:

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

- **Localized Contrastive Penalty Loss** is patch-level InfoNCE that **pulls the current local features toward the global model and pushes them away from the previous local state**, promoting alignment without losing locality:

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

- **sim_(pos)** (pres ↔ serv): Exponentiated cosine similarity between the **current local** and **global** patch; higher values **promote alignment** of client features with cross-client aggregates.

$$\text{sim}_{\text{pos}}^{(d,i,j)} = \exp\left(\frac{\text{patch}_{\text{pres}}(d,i,j) \cdot \text{patch}_{\text{serv}}(d,i,j)}{\tau}\right)$$

- **sim_(neg)** (prev ↔ pres): Exponentiated cosine similarity between **previous** and **current** local patches; higher values **penalize temporal inertia**, encouraging meaningful updates.

$$\text{sim}_{\text{neg}}^{(d,i,j)} = \exp\left(\frac{\text{patch}_{\text{prev}}(d,i,j) \cdot \text{patch}_{\text{pres}}(d,i,j)}{\tau}\right)$$

Adaptive Scale-aware Weighted Aggregation (ASWA):

A server-side aggregation rule that minimizes update variance under client size heterogeneity by weighting each client's contribution by its **effective data scale**, with optional tempering and smoothing to prevent dominance and improve stability.

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

Dataset Details

UCSF-PDGM (Client-1):

- 500 subjects; WHO grade 2–4 diffuse gliomas.
- Standardized 3T pre-op mpMRI (T1/T1ce/T2/FLAIR).
- Diffusion & perfusion MRI.
- Multi-compartment tumor segmentations.
- Tumor genetics, treatment, and survival metadata.

UPenn-GBM (Client-2):

- 630 patients with de novo GBM.
- mpMRI + demographics, outcomes, genomics.
- Longitudinal progression data.
- Computer-aided + expert-refined labels for multiple tumor sub-regions.

MSD (Client-3):

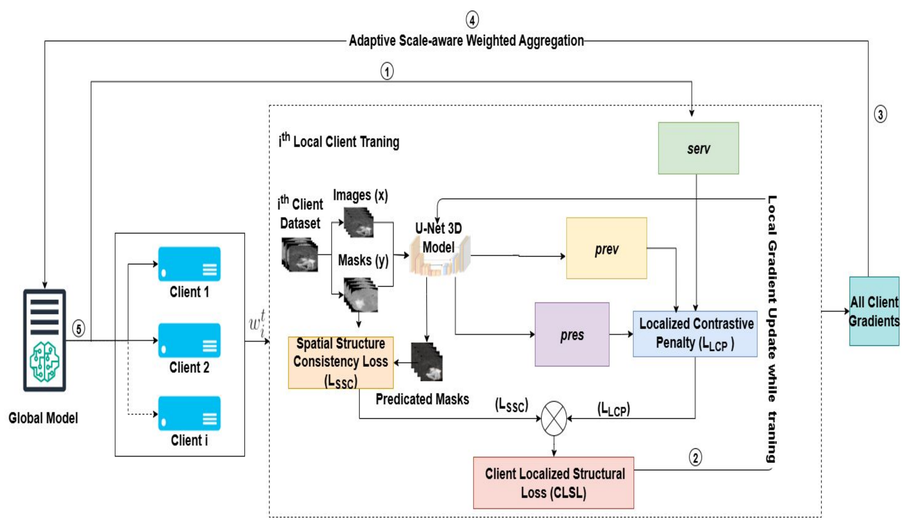
- **It consists of BraTS 2016–2017 subset,** 750 brain-tumor mpMRI volumes (GBM & LGG) .
- Expert-annotated tumor substructures.

Data Distribution:

We simulate a 3-client FL setting: **Client 1: UCSF-PDGM**, **Client 2: UPenn-GBM**, **Client 3: MSD (BraTS subset)**.

- Client 1 provides multi-parametric pre-op gliomas (grades II–IV).
- Client 2 focuses on de novo GBM with diverse demographics.
- Client 3 mixes GBM + lower-grade gliomas across scanners yielding realistic non-IID distributions.

Architecture



Results

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

Ablation Studies

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**

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

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**

More Empirical Analysis

Table A1: Scalability of FSF3A

Number of Clients	Dice ↑	IoU ↑	Prec. ↑	Sens. ↑
10	0.82	0.71	0.78	0.83
30	0.86	0.74	0.82	0.86
50	0.89	0.79	0.87	0.91

Table A2: Different segmentation models within FSF3A.

Data	Network	Dice ↑	IoU ↑	Prec. ↑	Sens. ↑
UCSF PDGM	VoxResNet	0.39	0.28	0.37	0.36
	SAM3D	0.43	0.38	0.48	0.58
UPenn-GBM	VoxResNet	0.37	0.24	0.32	0.39
	SAM3D	0.45	0.41	0.42	0.44
MSD	VoxResNet	0.37	0.24	0.32	0.39
	SAM3D	0.48	0.37	0.45	0.47

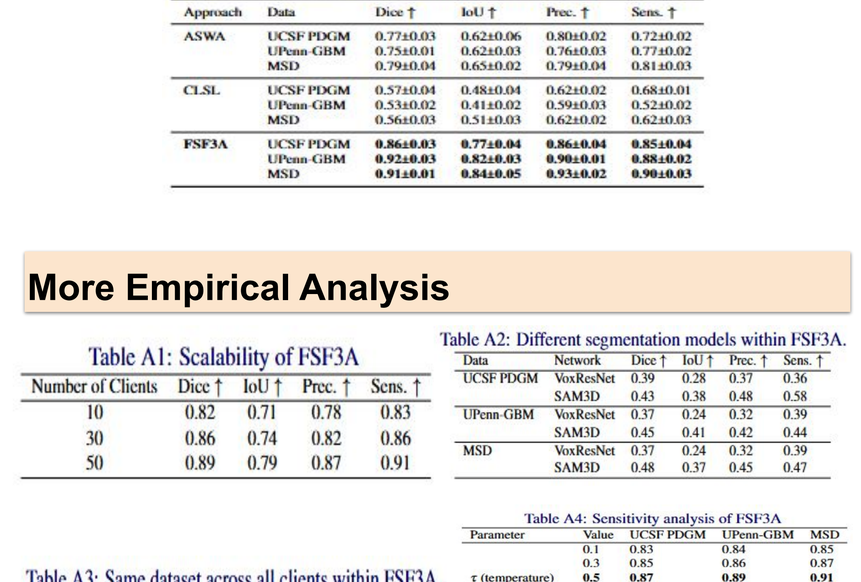
Table A3: Same dataset across all clients within FSF3A.

Data	Dice ↑	IoU ↑	Prec. ↑	Sens. ↑
UCSF PDGM	0.87	0.74	0.88	0.88
UPenn-GBM	0.94	0.83	0.92	0.90
MSD	0.93	0.86	0.94	0.92

Table A4: Sensitivity analysis of FSF3A

Parameter	Value	UCSF PDGM	UPenn-GBM	MSD
τ (temperature)	0.1	0.83	0.84	0.85
	0.3	0.85	0.86	0.87
	0.5	**0.87**	**0.89**	**0.91**
	0.7	0.85	0.87	0.88
	1.0	0.84	0.86	0.87
ω (grad weight)	0.1	0.84	0.85	0.86
	0.3	0.85	0.87	0.88
	0.5	**0.87**	**0.89**	**0.91**
	0.7	0.86	0.87	0.88
	1.0	0.84	0.86	0.87
γ (CLSL balance)	0.5	0.84	0.85	0.86
	0.75	0.86	0.87	0.88
	1.0	**0.87**	**0.89**	**0.91**
	1.25	0.86	0.87	0.88

Qualitative Analysis



More Information

- Email: ai21resch01001@iith.ac.in