FedAMM: Federated Learning for Brain Tumor Segmentation with Arbitrary Missing Modalities

Yukun Shi[1], Meiting Xue[1], Yan Zeng[2(✉)], Jilin Zhang[1], Jian Wan[2,3], and Ye Zhou[4]

[1] School of Cyberspace, Hangzhou Dianzi University, Hangzhou, 310018, China
[2] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou, 310018, China
yz@hdu.edu.cn
[3] Zhejiang University of Science and Technology, Hangzhou, 310023, China
[4] Affiliated Hangzhou First People's Hospital, School of Medicine, Westlake University, Hangzhou, 310006, China

Abstract. Brain tumor segmentation and detection have advanced significantly with the introduction of multimodal magnetic resonance imaging. However, data privacy concerns restrict most studies to centralized environments, limiting their real-world applicability. While federated learning (FL) offers a privacy-preserving solution for cross-institutional brain tumor research, existing multimodal FL approaches primarily address scenarios wherein clients possess either a single modality or complete missing modality data. These methods fail to account for the modality heterogeneity caused by arbitrary missing modalities, a frequent challenge in clinical practice. To address this issue, we propose FedAMM, a novel FL framework designed for brain tumor segmentation under arbitrary missing modalities. FedAMM incorporates multiple strategies to mitigate discrepancies arising from varying modality combinations across clients. First, FedAMM introduces a unimodal prototype distillation technique during local training to balance the contributions of different modalities. Additionally, the server aggregates multimodal prototypes uploaded by clients to generate cluster centers that represent the global modality distribution, thereby guiding local training toward global optimality. Furthermore, we implement a weighted aggregation strategy based on modality proportions. Experimental results on the BraTS2020 dataset demonstrate that FedAMM outperforms existing methods in handling arbitrary missing modalities, highlighting its strong adaptability to imbalanced and heterogeneous federated systems. The code is available at `https://github.com/13sky/FedAMM.git`.

Keywords: Federated learning · Modality missing · Medical image segmentation · Brain tumor segmentation.

1 Introduction

Multimodal learning plays a key role in disease diagnosis and treatment [1, 2], integrating information from multiple medical imaging modalities significantly

enhances diagnostic accuracy and comprehensiveness. Among these imaging modalities, multimodal magnetic resonance imaging (MRI) has demonstrated remarkable progress in brain tumor segmentation [3]. By leveraging the complementary information from T1-weighted (T1), post-contrast T1 (T1c), T2-weighted (T2), and fluid-attenuated inversion recovery (FLAIR) modalities, segmentation models can more precisely capture tumor characteristics [4]. However, data privacy constraints restrict most current studies [5, 6] to centralized learning scenarios. For instance, in clinical practice, medical data are typically distributed across multiple hospitals and institutions, and patient privacy concerns preclude direct data aggregation for model training [7]. Thus, privacy challenges pose a significant barrier to the adoption of multimodal segmentation technologies in cross-institutional research.

Federated learning (FL), a privacy-preserving distributed learning framework, presents a potential solution to the above issues as it enables collaborative model training without requiring raw data exchange [8]. This approach has demonstrated promise across various domains, including medical imaging [9]. However, in practical applications, the phenomenon of modality missingness [10] is widespread owing to differences in imaging protocols, equipment, and clinical constraints, posing a significant challenge for multimodal FL approaches [11]. Existing centralized methods [12–14] typically address missing modalities through zero-filling or data generation, disregarding the inherent correlations among different modalities. Consequently, these methods are not directly applicable to federated scenarios. In FL, missing modalities [15] are classified into two groups: **complete missingness**, where each client consistently lacks a specific modality across all samples, and **arbitrary missingness**, where different samples for the same client have varying combinations of available modalities. Although approaches such as FedNorm [16] and FedMM [17] mitigate complete missing-modality instances using regularization and representation learning, they are ineffective at handling the more complex challenge of arbitrary missing modalities, where modality combinations vary across samples for the same client. This inconsistency leads to severe distribution shifts, which is a key limitation that the proposed FedAMM approach is designed to overcome. Compared to complete missingness, arbitrary missingness introduces greater heterogeneity across samples and clients owing to the uneven distribution of modalities. Furthermore, the global model must perform well under all potential missing-modality combinations during inference.

To address this critical challenge, we propose FedAMM, a federated brain tumor segmentation method specifically designed to handle arbitrary missing modalities. Previous research [18, 19] indicates that different modalities contribute unequally to model performance, resulting in significant modality imbalances during training when samples contain varying modality combinations. To counter modality imbalances, especially dominance in model updates, we propose prototype [20, 21] distillation for balanced information sharing. This approach ensures that each modality, regardless of its combination, meaningfully contributes to the model's predictive performance. Furthermore, to address the

inter-client heterogeneity introduced by missing modalities, each client's learned prototypes are uploaded to the server and categorized based on modality combinations. Representative prototypes are then selected to regularize local models, aligning them and minimizing divergence across clients. Finally, instead of relying on conventional data-volume-based aggregation, we aggregate model encoders based on the number of available modalities. This strategy mitigates the heterogeneity introduced by arbitrary missing modalities and enhances inference performance across all possible modality combinations. To the best of our knowledge, this is the first study to examine arbitrary missing modalities in the context of FL for brain tumor segmentation.

The primary contributions of this study include the following: 1. We present the first study on arbitrary missing modalities in multimodal FL, addressing a more realistic and practical challenge. 2. We introduce a prototype-based strategy to balance cross-modality discrepancies and mitigate client heterogeneity, further alleviating distribution mismatches through modality-specific encoder aggregation. 3. The proposed method outperforms previous approaches on public brain tumor segmentation datasets under diverse missing-modality conditions.

2 Method

This study focuses on the collaborative task of brain tumor segmentation with missing modalities across K clients. First, each client's dataset is defined as $D_k = \{(x_n, y_n)\}_{n=1}^{N_k}$, with each sample $x_n = \{x^1, x^2, x^3, ..., x^m\}$ containing up to M modalities, all sharing the same ground truth y_n. Accounting for potential missing modalities, the number of possible non-empty modality subsets is $\bar{M} = 2^M - 1$. The modality combination for each sample is denoted as $\bar{m}$ redefining X_n as $X_n^{\bar{m}}$, where $\bar{m} \in \{1, 2, ..., \bar{M}\}$. We consider a scenario wherein the modality combinations for each sample X_n are arbitrary. To handle this scenario, we train a global model using all available modality information from the datasets, allowing it to generate predictions for samples with any modality combination during inference. The objective is to optimize the following loss function:

$$\underset{\theta}{\operatorname{argmin}} \frac{1}{\sum_{k=1}^{K} N_k} \sum_{k=1}^{K} \sum_{n=1}^{N_k} \mathcal{L}\left(f\left(\theta, x_n^{\bar{m}}, y_n\right)\right), where, \bar{m} \in \{1, 2, ..., \bar{M}\} \quad (1)$$

To this end, we propose FedAMM. Fig.1 illustrates the overall framework of this approach. We employ a classical multimodal segmentation model comprising four modality-specific encoders and a shared decoder. For incomplete modality inputs, only the available modality encoders are activated, and their fused representations are decoded to produce final predictions. To address the imbalance introduced by modality heterogeneity across different samples, in addition to the widely used cross-entropy loss L_{ce} and Dice loss L_{dice} [22], we introduce an intra-client modality balance loss L_{mb} and an inter-client modality combination balance loss L_{mc}. Furthermore, we optimize the model aggregation stage to enhance performance.

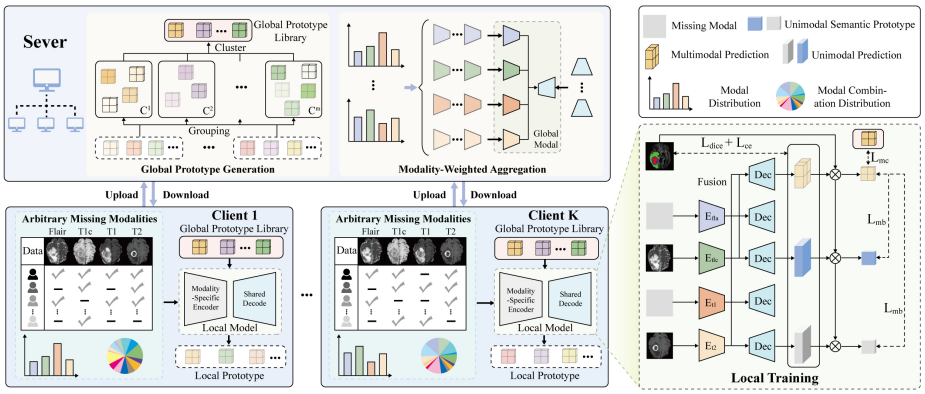

Fig. 1: Overview of the proposed FedAMM.

Intra-client Modal Balance. Given the varying distribution patterns of pixel classifications across modalities, the training process frequently favors the modality containing richer information, referred to as the "fast" modality, while the other is termed the "slow" modality [23]. Under these circumstances, the model learns incompletely from the slow modality and biases itself toward the fast modality. To mitigate this imbalance, we propose enriching unimodal representations by transferring knowledge from multimodal representations to individual modalities, thereby negating the impact of information differences. Drawing inspiration from prior studies [24, 25], we introduce prototype knowledge distillation. Here, the prototype for each modality is computed based on the class features of the pixels in the modal representation. In this case, the multimodal representation serves as the teacher prototype $p^t_{n,c}$, while the unimodal representation acts as the student prototype $p^m_{n,c}$, enabling knowledge transfer between modalities. The prototypes are defined as follows:

$$p^t_{n,c} = \frac{\sum_{i=1}^{I} \hat{y}_{n,i} \mathbb{I}\left[y_{n,i} = c\right]}{\sum_{i=1}^{I} \mathbb{I}\left[y_{n,i} = c\right]} \qquad p^m_{n,c} = \frac{\sum_{i=1}^{I} \hat{y}^m_{n,i} \mathbb{I}\left[y_{n,i} = c\right]}{\sum_{i=1}^{I} \mathbb{I}\left[y_{n,i} = c\right]} \tag{2}$$

where I indicates the number of pixels in the modality, C denotes the classification, $\hat{y}$ indicates pseudo labels, and $\mathbb{I}\left[y_{n,i} = c\right]$ represents an indicator function that equals one if $y_{n,i} = c$ and zero otherwise. We use cosine similarity to quantify the similarity between the pixel class prototype and the pseudo label, capturing semantic information both within and between classes:

$$P^t_{n,c} = \sum_{i=1}^{I} \frac{\hat{y}_i \cdot p^t_{n,c}}{\|y_i\| \left\|p^t_{n,c}\right\|} \qquad P^m_{n,c} = \sum_{i=1}^{I} \frac{\hat{y}^m_i \cdot p^m_{n,c}}{\|y^m_i\| \left\|p^m_{n,c}\right\|} \tag{3}$$

where $P^t_{n,c}$ and $P^m_{n,c}$ denote the semantic prototypes of the teacher and student modalities, respectively. To balance the knowledge disparity between modalities,

we minimize the distance between unimodal and multimodal semantic prototypes using L2 loss, defined as the modal distillation loss:

$$\mathcal{L}_{md}^{m} = \sum_{i=1}^{I} \sum_{c=1}^{C} \left\| P_{n,c}^{m}(i) - P_{n,c}^{t}(i) \right\|_2^2 \tag{4}$$

The semantic difference between unimodal and multimodal is defined as:

$$\mathcal{D}_{n}^{m} = \sum_{i=1}^{I} \sum_{c=1}^{C} \left\| P_{n,c}^{m}(i) - P_{n,c}^{t}(i) \right\|_2 \tag{5}$$

A larger value of D_n^m indicates that modality m is a slow modality, implying that its learning proportion should be increased. The imbalance rate between any two single modalities is defined as follows:

$$\rho_j^i = \frac{D_n^i}{D_n^j} \;\; i, j \in M, i \neq j \tag{6}$$

We regulate the loss weights using coefficients α and β, determined as follows:

$$\begin{cases} \alpha = \text{clip}\left(0, \frac{1}{\rho_j^i} - 1, 1\right), \beta = 0 \;\; \rho_j^i < 1 \\ \alpha = 0, \beta = \text{clip}\left(0, \rho_j^i - 1, 1\right) \;\; \rho_j^i \geqslant 1 \end{cases} \tag{7}$$

where i and j represent any two modalities, and the *clip* function bounds the loss weight between 0 and 1. The modality-balanced loss is formulated as:

$$\mathcal{L}_{mb} = \alpha \mathcal{L}_{md}^{i} + \beta \mathcal{L}_{md}^{j} \;\; i, j \in M \;\; and \;\; i \neq j \tag{8}$$

For all available modalities, we compute the modality balancing loss over each pair of modalities to ensure that no single modality dominates the optimization process.

Inter-client Modal Combination Balance. While the previous section focused on intra-client knowledge transfer from multimodal representations to balance cross-modal information disparities within clients, global modality distribution differences persist across clients owing to variations in available modal combinations. To address this, we introduce a global mixed-modal semantic prototype library to harmonize client model training. In particular, we aggregate semantic prototypes from all client samples on the server and categorize them into modal combination groups, defined as: $P_{n,c}^{t,\bar{M}} = \{P_{n,c}^{t,1}, P_{n,c}^{t,2}, ..., P_{n,c}^{t,\bar{m}}\}$. We then apply k-means clustering to derive global cluster centroids, denoted as:

$$P^{g,\bar{M}} = Cluster\{P_{n,c}^{t,1}, P_{n,c}^{t,2}, ..., P_{n,c}^{t,\bar{m}}\} = \{P^{g,1}, P^{g,2}, ..., P^{g,\bar{m}}\} \tag{9}$$

These centroids $P^{g,\bar{M}}$ represent unified semantic references for each modal combination, effectively mitigating interclient model discrepancies. To align client

sample prototypes with the global centroid prototypes of their modality combination, we use L2 loss, defined as:

$$\mathcal{L}_{mc}^{\bar{m}} = \sum_{i=1}^{I} \sum_{c=1}^{C} \left\| P_{n,c}^{t,\bar{m}}(i) - P^{g,\bar{m}}(i) \right\|_2^2 \tag{10}$$

Given that each client has a different number of modal combinations, we design weights for the modal combination loss based on the modal combination ratio. We determine the number of samples $d_k^{\bar{m}} = \{d^1, d^2, ..., d^{\bar{M}}\}$ for each modal combination in a client's training dataset, with d denoting the global number of samples. The weight $\gamma^{\bar{m}}$ is then computed as:

$$\gamma^{\bar{m}} = \frac{d_k^{\bar{m}}}{d}, \bar{m} \in \{1, 2, ..., \bar{M}\} \tag{11}$$

Thus, the local modal combination loss is expressed as follows: $\mathcal{L}_{mc} = \gamma^{\bar{m}} \mathcal{L}_{mc}^{\bar{m}}$.

Modality-weighted Aggregation. Traditional FL methods typically use data volume weights to aggregate client models during the model aggregation phase. However, in multimodal models, the actual knowledge learned by each modality-specific encoder does not include sample size but the number of modalities. To account for this, we aggregate the M modality-specific encoders based on modality ratio and combine shared decoders based on data volume. Specifically, we calculate the number of samples for each modality in each client, denoted as s_k^m. The total number of samples for each modality across all clients is then given as $s^m = \sum_{k=1}^{K} s_k^m$. The weight for each modality-specific encoder is then computed as $\omega_k^m = \frac{s_k^m}{s^m}, m \in \{1, 2, ..., M\}$. Thus, the aggregation of the global model proceeds as follows:

$$F_g = \sum_{k=1}^{K} \{\omega_k^1 * Enc_k^1, \omega_k^2 * Enc_k^2, ..., \omega_k^m * Enc_k^m, \frac{D_k}{D} * Dec_k\} \tag{12}$$

where F_g represents the global model, Enc_k^m denotes the m-modality-specific encoder for client k, and Dec_k represents the shared decoder for client k.

3 Experiments

Dataset and Settings. We evaluated our method on widely used brain tumor segmentation benchmarks from the BraTS2020 challenge [26], which includes MRI scans from 369 subjects. Here, each subject has four MRI modalities: FLAIR, T1, T1c, and T2. Following previous studies, we preprocessed the data and split them into 219, 50, and 100 subjects for training, validation, and testing, respectively. The ground truth annotations defined three nested tumor subregions: whole tumor (WT), tumor core (TC), and enhancing tumor (ET).

Table 1: Performance comparison of FedAMM with State-of-the-art Methods for tumor region segmentation (WT, TC, ET) under varying levels of modality heterogeneity on BraTS2020.

Method	$\alpha = 0.001$				$\alpha = 0.1$				$\alpha = 1$			
	WT	TC	ET	Avg	WT	TC	ET	Avg	WT	TC	ET	Avg
FedAvg	73.42	57.75	43.09	58.09	79.42	69.24	53.40	67.35	83.39	72.35	56.36	70.70
FedProx	69.08	55.68	43.61	56.12	79.24	69.02	54.25	67.50	82.97	72.43	54.47	69.96
FedNorm	70.29	54.72	42.25	55.75	78.86	68.06	52.04	66.32	82.40	66.90	53.65	67.65
FedMM	75.46	60.28	47.00	60.91	80.00	70.33	54.74	68.36	82.72	72.54	55.75	70.34
FedMEMA	76.19	60.42	48.59	61.73	82.01	70.21	55.07	69.10	84.04	73.82	57.83	71.90
FedAMM(ours)	**82.19**	**72.34**	**54.62**	**69.71**	**83.86**	**74.32**	**56.67**	**71.62**	**85.04**	**76.85**	**58.20**	**73.36**

To simulate a realistic scenario with missing modalities, the training set is evenly partitioned into four clients, with modality distribution controlled by the Dirichlet parameter α. When $\alpha = 1$, samples tend to have completely missing modalities. As α decreases, the setting progressively shifts toward an arbitrarily missing modality scenario, leading to higher modality heterogeneity among clients. We use the Dice similarity coefficient [27] as the performance metric.

Implementation Details. We implemented FedAMM using PyTorch 1.11.0 and trained it on four RTX 4090 GPUs. Each client model was trained on a separate GPU, with one GPU reserved for server-side aggregation. For fairness, all methods use RFNet [28] as the backbone network. Local training was performed with a batch size of one using an Adam optimizer with a learning rate of 0.0002 and a weight decay of 0.0001. The model was trained for 1,000 rounds, with each round comprising one epoch of local training.

Comparison with State-of-the-art Methods. We compared FedAMM with several baseline methods, including FedAvg [29], FedProx [30], FedNorm [16], FedMM [17], and FedMAME [31]. Among these, FedAvg and FedProx address unimodal heterogeneity, while FedNorm, FedMM, and FedMAME focus on multimodal scenarios.

Table 2: Performance comparison of FedAMM and State-of-the-art Methods across 15 modality combination scenarios.

Flair	●	○	●	●	●	●	●	○	●	○	●	●	○	○	○	Avg
T1c	●	●	●	○	●	●	○	○	○	●	●	○	○	●	○	
T1	●	●	○	●	●	○	○	●	●	●	○	○	●	○	○	
T2	●	●	●	●	○	○	●	●	○	○	●	○	○	○	●	
FedAvg	70.83	69.79	69.91	53.91	70.90	69.48	53.01	50.29	50.95	61.70	67.90	46.97	33.15	55.66	46.82	58.09
FedProx	69.64	66.26	69.54	52.20	68.56	69.26	53.91	46.60	49.65	56.83	65.92	49.33	20.63	54.97	48.48	56.12
FedNorm	69.17	66.52	68.20	52.63	68.65	66.91	51.97	48.79	48.76	58.38	65.20	47.46	21.01	56.37	46.25	55.75
FedMM	74.24	71.59	75.20	56.78	72.19	75.32	57.53	53.03	49.46	58.41	72.39	51.46	28.14	65.12	52.89	60.91
FedMEMA	77.77	75.09	77.56	56.83	77.01	76.95	57.55	52.08	51.65	53.80	74.52	51.76	28.54	65.40	49.51	61.73
FedAMM(our)	**80.39**	**78.75**	**79.84**	**64.05**	**79.81**	**78.78**	**63.55**	**62.54**	**60.52**	**77.28**	**79.44**	**57.14**	**50.45**	**75.04**	**58.12**	**69.71**

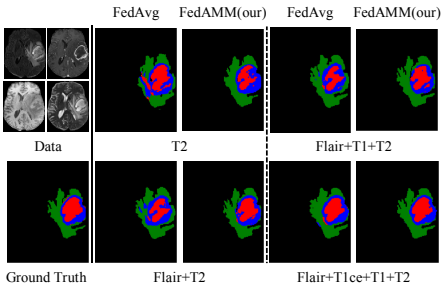

Fig. 2: Visualization results of FedAvg and FedAMM predictions under four modality combinations.

Table 3: Ablation study results showing the impact of three key components on the performance of the FedAMM.

L_{mb}	L_{mc}	Aggregation	WT	TC	ET	Avg
			73.42	57.75	43.09	58.09
✓			79.29	68.37	51.92	66.53
	✓		76.88	65.55	49.33	63.92
✓	✓		81.16	70.37	53.09	68.21
✓	✓	✓	82.19	72.34	54.62	69.71

We evaluated the performance of all methods on 15 modality combinations under three different levels of missing modality scenarios. Table 1 shows that as modality heterogeneity increases, all methods suffer performance degradation. However, FedAMM remains more stable than others. Specifically, FedProx, designed to address data heterogeneity in unimodal settings, proves ineffective in multimodal scenarios and performs comparably to FedAvg. Notably, FedNorm performs even more poorly than FedAvg, likely owing to its simplistic modality normalization approach, which prevents the model from capturing the rich semantic information across modalities. While FedMM and FedMAME demonstrate improvements over FedAvg, their performance gains remain limited compared to FedAMM. Specifically, when $\alpha = 0.001$, FedAMM outperforms FedAvg by 20% and the second-best method, FedMEMA, by 12%.

Table 2 presents a comparison of inference performance across different modality combinations. The results indicate that FedAMM performs comparably to other methods when inferring from samples with a greater number of modalities. However, as the number of modalities decreases, FedAMM demonstrates notably superior inference performance. Overall, the visualization results in Fig.2 indicate that FedAMM produces better segmentation outcomes than the baseline methods across various modality combinations.

We attribute the observed performance improvement to FedAMM's ability to transfer knowledge from multimodal settings to unimodal samples while maintaining consistency across client models. This dual capability enables FedAMM to achieve competitive performance in multimodal combinations while significantly enhancing performance in low-modality scenarios. Consequently, FedAMM consistently outperforms all baseline methods in terms of overall performance.

Ablation Study. To evaluate the effectiveness of the three key components in FedAMM, we conducted an ablation study, as shown in Table 3. When none of these components are used, FedAMM degenerates into FedAvg. We found that the modal balance loss has a greater impact on the performance of FedAMM compared to the modal combination loss when they are used separately. We

believe this is because, when there are significant modality differences within a client, merely balancing sample combinations across clients provides limited improvement. By combining both losses, the model's performance is significantly enhanced, which further validates our hypothesis. Finally, a modality-based model aggregation strategy further reduces inter-client differences and enhances overall performance. In summary, these three key components play a crucial role in enhancing the performance of FedAMM.

4 Conclusion

This study presents an FL framework for multimodal brain tumor segmentation that effectively addresses heterogeneity caused by arbitrary missing modalities. Our method balances intra-client modality variations while ensuring model consistency across clients. Additionally, a modality-weighted aggregation strategy enhances global performance. Experiments on the BraTS2020 dataset demonstrate the framework's superior and robust segmentation performance across various modality combinations, highlighting its potential for real-world multimodal medical imaging applications. Future research will focus on expanding its applicability, improving adaptability, and enhancing privacy preservation.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (NSFC) under Grant No.62302133; the Key Research and Development Program of Zhejiang Province under Grant (2024C01026, 2023C03194); the Yangtze River Delta Project under Grant No.2023ZY1068; the Key Research and Development Program of Hangzhou City under Grant 2024SZD1A02.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Behrad, F., Abadeh, M.S.: An overview of deep learning methods for multimodal medical data mining. Expert Systems with Applications **200**, 117006 (2022)
2. Cao, B., Bi, Z., Hu, Q., Zhang, H., Wang, N., Gao, X., Shen, D.: Autoencoder-driven multimodal collaborative learning for medical image synthesis. International Journal of Computer Vision **131**(8), 1995–2014 (2023)
3. Liu, Z., Tong, L., Chen, L., Jiang, Z., Zhou, F., Zhang, Q., Zhang, X., Jin, Y., Zhou, H.: Deep learning based brain tumor segmentation: a survey. Complex & intelligent systems **9**(1), 1001–1026 (2023)
4. Zhang, Y., He, N., Yang, J., Li, Y., Wei, D., Huang, Y., Zhang, Y., He, Z., Zheng, Y.: mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 107–117. Springer (2022)
5. Qiu, Y., Chen, D., Yao, H., Xu, Y., Wang, Z.: Scratch each other's back: Incomplete multi-modal brain tumor segmentation via category aware group self-support learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 21317–21326 (2023)

6. Wang, Y., Zhang, Y., Liu, Y., Lin, Z., Tian, J., Zhong, C., Shi, Z., Fan, J., He, Z.: Acn: adversarial co-training network for brain tumor segmentation with missing modalities. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VII 24. pp. 410–420. Springer (2021)
7. Pan, H., Zhao, X., He, L., Shi, Y., Lin, X.: A survey of multimodal federated learning: background, applications, and perspectives. Multimedia Systems **30**(4), 222 (2024)
8. Wen, J., Zhang, Z., Lan, Y., Cui, Z., Cai, J., Zhang, W.: A survey on federated learning: challenges and applications. International Journal of Machine Learning and Cybernetics **14**(2), 513–535 (2023)
9. Guan, H., Yap, P.T., Bozoki, A., Liu, M.: Federated learning for medical image analysis: A survey. Pattern Recognition p. 110424 (2024)
10. Zhou, T., Ruan, S., Hu, H.: A literature survey of mr-based brain tumor segmentation with missing modalities. Computerized Medical Imaging and Graphics **104**, 102167 (2023)
11. Che, L., Wang, J., Zhou, Y., Ma, F.: Multimodal federated learning: A survey. Sensors **23**(15), 6986 (2023)
12. Azad, R., Khosravi, N., Merhof, D.: Smu-net: Style matching u-net for brain tumor segmentation with missing modalities. In: International Conference on Medical Imaging with Deep Learning. pp. 48–62. PMLR (2022)
13. Liu, H., Wei, D., Lu, D., Sun, J., Wang, L., Zheng, Y.: M3ae: multimodal representation learning for brain tumor segmentation with missing modalities. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 37, pp. 1657–1665 (2023)
14. Wang, H., Chen, Y., Ma, C., Avery, J., Hull, L., Carneiro, G.: Multi-modal learning with missing modality via shared-specific feature modelling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15878–15887 (2023)
15. Lin, Y.M., Gao, Y., Gong, M.G., Zhang, S.J., Zhang, Y.Q., Li, Z.Y.: Federated learning on multimodal data: A comprehensive survey. Machine Intelligence Research **20**(4), 539–553 (2023)
16. Bernecker, T., Peters, A., Schlett, C.L., Bamberg, F., Theis, F., Rueckert, D., Weiß, J., Albarqouni, S.: Fednorm: Modality-based normalization in federated learning for multi-modal liver segmentation. arXiv preprint arXiv:2205.11096 (2022)
17. Peng, Y., Bian, J., Xu, J.: Fedmm: Federated multi-modal learning with modality heterogeneity in computational pathology. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1696–1700. IEEE (2024)
18. Zhang, Y., Li, Z., Li, H., Tao, D.: Prototype-driven and multi-expert integrated multi-modal mr brain tumor image segmentation. IEEE Transactions on Instrumentation and Measurement (2024)
19. Shi, J., Shang, C., Sun, Z., Yu, L., Yang, X., Yan, Z.: Passion: Towards effective incomplete multi-modal medical image segmentation with imbalanced missing rates. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 456–465 (2024)
20. Hinton, G.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
21. Pahde, F., Puscas, M., Klein, T., Nabi, M.: Multimodal prototypical networks for few-shot learning. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 2644–2653 (2021)

22. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV). pp. 565–571. Ieee (2016)
23. Zhou, Y., Wang, X., Chen, H., Duan, X., Zhu, W.: Intra-and inter-modal curriculum for multimodal learning. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 3724–3735 (2023)
24. Jiang, Z., Li, Y., Yang, C., Gao, P., Wang, Y., Tai, Y., Wang, C.: Prototypical contrast adaptation for domain adaptive semantic segmentation. In: European conference on computer vision. pp. 36–54. Springer (2022)
25. Wang, S., Yan, Z., Zhang, D., Wei, H., Li, Z., Li, R.: Prototype knowledge distillation for medical segmentation with missing modality. In: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 1–5. IEEE (2023)
26. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE transactions on medical imaging **34**(10), 1993–2024 (2014)
27. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
28. Ding, Y., Yu, X., Yang, Y.: Rfnet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3975–3984 (2021)
29. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)
30. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems **2**, 429–450 (2020)
31. Dai, Q., Wei, D., Liu, H., Sun, J., Wang, L., Zheng, Y.: Federated modality-specific encoders and multimodal anchors for personalized brain tumor segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 38, pp. 1445–1453 (2024)