

# Coarse-to-Fine Meta-Reweighting with Dynamic Retrieval for Adult-to-Pediatric Domain Adaptation in Tumor Segmentation

Abdulkhalek Al-Fakih<sup>1</sup>, Kyu Sung Choi<sup>2</sup>, Dong-Hyun Kim<sup>1</sup>, Kanghyun Ryu<sup>3,4,\*</sup>, and Mohammed A. Al-masni<sup>5,\*</sup>

<sup>1</sup> Department of Electrical and Electronic Engineering, Yonsei University, Seoul, Republic of Korea

<sup>2</sup> Department of Radiology, Seoul National University Hospital, Seoul, Republic of Korea

<sup>3</sup> Intelligence and Interaction Research Center, Korea Institute of Science and Technology, Seoul, Republic of Korea

<sup>4</sup> KIST School, Division of AI and Robotics, University of Science and Technology, Seoul, Republic of Korea

<sup>5</sup> Department of Bioengineering, King Fahd University of Petroleum & Minerals (KFUPM), Dhahran, Saudi Arabia

alfakih@sju.ac.kr, ent1127@snu.ac.kr, donghyunkim@yonsei.ac.kr, khryu@kist.re.kr, m.almasani@sejong.ac.kr

**Abstract.** Deep learning models face challenges in medical image segmentation under limited annotations, particularly in pediatric brain tumor cases. Simply augmenting pediatric datasets with adult cases is suboptimal due to anatomical and developmental differences. We propose a two-stage meta-reweighting framework: a coarse stage that performs cross-domain binary segmentation for stable localization, and a fine stage that refines pediatric-specific subregions with gradient-alignment-based reweighting. To further improve pediatric relevance, we introduce a gradient-similarity-driven retrieval mechanism that selects representative pediatric meta-cases as a stable reference set for weight computation, avoiding reliance on noisy online samples. This design enables effective loss reweighting while maintaining computational efficiency. Experiments on BraTS-PEDs demonstrate that the proposed method significantly outperforms state-of-the-art reweighting approaches, achieving a 6.1% improvement in Dice score and a 20.84 mm reduction in HD95, highlighting its robustness for pediatric brain tumor segmentation in low-data settings. Code is available at <https://github.com/abod-alfakih/MDR2>.

**Keywords:** Pediatric Brain Segmentation · Data Scarcity · Meta reweighting · Dynamic Retrieval

---

\* Corresponding author

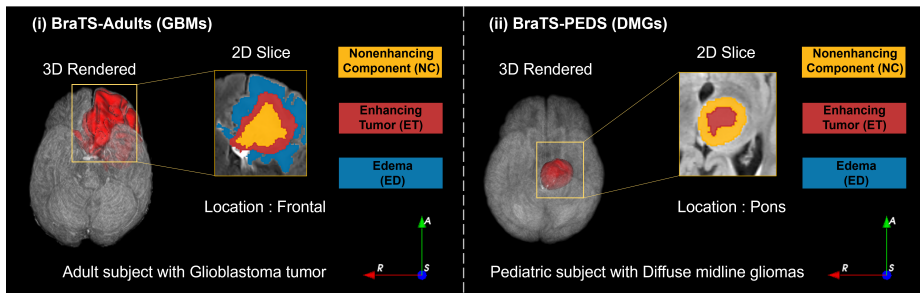
## 1 Introduction

Deep learning has advanced medical image segmentation, yet pediatric brain tumor segmentation remains critically challenging due to severe data scarcity. Unlike adult gliomas with thousands of annotated cases, pediatric datasets comprise only 99 labeled cases, reflecting tumor rarity (incidence 1 per 100,000) [1], high annotation costs, and ethical constraints [2, 3]. To address this scarcity, augmentation with adult glioma cases has become necessary [4]. However, this introduces domain shifts from fundamental differences in patient populations and tumor biology—including variations in brain anatomy, tissue composition [5], and tumor types [6]. As shown in Fig. 1, adult glioblastomas and pediatric diffuse midline gliomas differ substantially in their internal subregion composition and spatial characteristics.

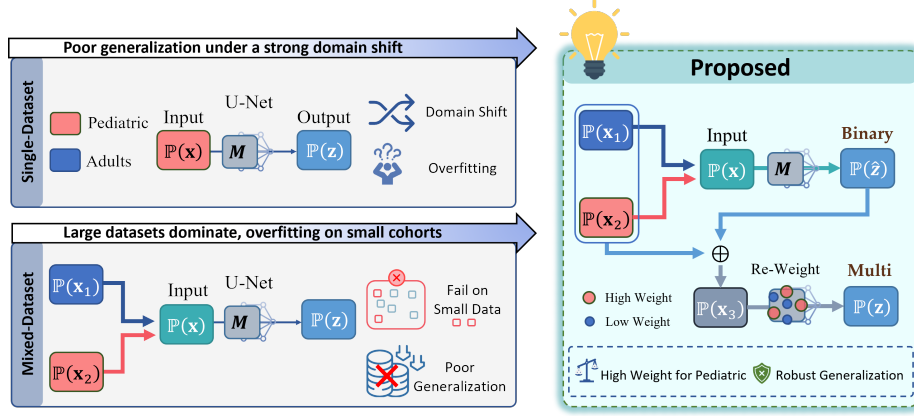
Models trained on mixed datasets naturally overfit adult-dominant features due to population imbalance, limiting pediatric generalization [17, 6]. While meta-learning offers a principled framework for handling such imbalance through adaptive sample reweighting, it remains largely unexplored for this problem. We address this gap with a two-stage meta-reweighting framework designed to mitigate pediatric-adult domain discrepancy.

We propose a two-stage meta-reweighting framework that strategically decomposes the problem (Fig. 2). Recognizing that tumor localization shows cross-domain consistency while subregion characteristics are population-specific, the coarse stage learns robust binary segmentation from mixed data, while the fine stage focuses on pediatric-specific refinement through adaptive sample reweighting.

One of the key innovation is our dynamic retrieval mechanism: instead of reweighting against random pediatric samples as in prior meta-learning [7–10], we identify and maintain a stable set of representative cases based on gradient similarity. This ensures reweighting is guided by informative references that truly



**Fig. 1.** Comparison of adult and pediatric brain tumors. Adult glioblastomas (i) involve the frontal/temporal lobes with typical peripheral enhancement, central necrosis, and pronounced peritumoral edema, whereas pediatric diffuse midline gliomas (DMGs) (ii) involve midline structures such as the brainstem with relatively reduced peritumoral edema.



**Fig. 2.** Comparison of two training strategies for pediatric brain tumor segmentation: (a) naïve joint training on combined pediatric and adult datasets, and (b) the proposed adaptive meta-reweighting strategy.

reflect pediatric characteristics, not noisy individual samples. Combined with efficient gradient alignment, this achieves substantial gains (10.7% Dice, 24.6mm HD95) without bi-level optimization overhead. Our main contributions are: (1) A coarse-to-fine framework that exploits cross-domain consistency while addressing population-specific differences; (2) A gradient-similarity retrieval mechanism for stable meta-case selection; (3) Efficient gradient-based reweighting achieving state-of-the-art pediatric segmentation performance.

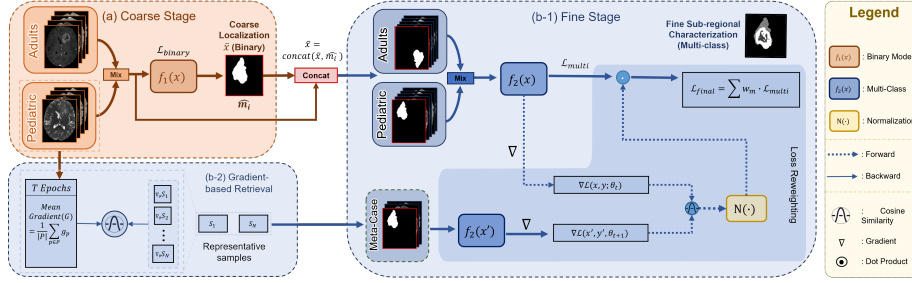
## 2 Materials and Methods

### 2.1 Datasets

We utilize the BraTS 2021 (1,251 adult cases) [11] and BraTS-PEDs 2024 (99 pediatric cases) [12] labeled training cohort. Annotations comprise Edema (ED), Enhancing Tumor (ET), and Non-Enhancing Component (NC). From these labels, Tumor Core ( $TC = NC \cup ET$ ) and Whole Tumor ( $WT = ED \cup NC \cup ET$ ) are derived. BraTS-PEDs is randomly divided into 70 training and 29 testing cases. To mitigate variance caused by the small sample size, we repeat experiments under a 4-fold cross-validation scheme and report each results. At each training epoch, 10 pediatric subjects are sampled as a meta dataset using the proposed dynamic retrieval strategy.

### 2.2 Proposed Framework

The proposed framework consists of two stages (Fig. 3). The coarse stage performs cross-domain binary segmentation to obtain robust tumor localization, while the fine stage refines these predictions into pediatric-specific subregions.



**Fig. 3.** Overview of the proposed two-stage meta-reweighting framework: (a) binary coarse localization; (b-1) pediatric multi-class refinement; (b-2) gradient-based dynamic retrieval.

In the fine stage, a meta-reweighting strategy with dynamic retrieval adaptively reweights training samples, prioritizing pediatric cases and adult samples that align with pediatric gradients.

**Coarse Stage (Cross-Domain Binary Segmentation):** As shown in Fig. 3(a), the coarse stage trains a binary segmentation model on both adult and pediatric data by collapsing all tumor subregions into a single class. The rationale is that overall tumor appearance and whole-tumor features are broadly consistent across populations, making binary segmentation a stable cross-domain task. Let  $\mathcal{X} = \{x_i\}_{i=1}^N$  denote the MR volumes and  $\mathcal{Y}^{(b)} = \{y_i^{(b)}\}_{i=1}^N$  the binary masks ( $y_i^{(b)} \in \{0, 1\}$ ). The model  $f_\theta$  is trained with the sum of Dice and BCE losses:

$$\mathcal{L}_{\text{binary}} = \mathcal{L}_{\text{Dice}}(f_\theta(x_i), y_i^{(b)}) + \mathcal{L}_{\text{BCE}}(f_\theta(x_i), y_i^{(b)}), \quad (1)$$

The resulting coarse masks  $\hat{m}_i$  provide auxiliary input to the fine stage, which then focuses on refining population-specific subregion differences through meta-reweighting.

**Fine Stage (Multi-Class Segmentation):** The second stage (Fig 3(b-1)) refines binary tumor segmentation into subregions specific to pediatric cases (i.e., ET, TC, and ED). The binary tumor mask  $\hat{m}_i$  predicted in Stage 1 is concatenated with the original MR modalities  $x_i$  to form the augmented input.

$$\tilde{x}_i = \text{concat}(x_i, \hat{m}_i), \quad \tilde{x}_i \in \mathbb{R}^{H \times W \times D \times (K+1)}, \quad (2)$$

where  $K$  is the number of MR modalities and  $(K+1)$  accounts for the additional binary mask channel. The multi-class segmentation model  $f_\theta$  maps  $\tilde{x}_i$  to  $C$  output channels, with  $C = 3$  representing the tumor subregions. The objective is a weighted combination of Dice and CE losses:

$$\mathcal{L}_{\text{multi}}(\theta) = \frac{1}{|B|} \sum_{i \in B} \left[ \mathcal{L}_{\text{Dice}}^{\text{multi}}(f_\theta(\tilde{x}_i), y_i) + \mathcal{L}_{\text{CE}}(f_\theta(\tilde{x}_i), y_i) \right], \quad (3)$$

where  $y_i \in \{0, 1, \dots, C\}^{H \times W \times D}$  are voxel-wise labels, and  $B$  is the batch. This stage emphasizes pediatric-specific subregion characteristics, guided by meta-reweighting.

**Dynamic Meta-Cases Retrieval:** We use a gradient-based meta-reweighting strategy with two components: (i) selecting representative pediatric meta-cases through gradient similarity, and (ii) reweighting training samples based on their alignment with these meta-cases.

(i) *Retrieval-Based Meta-Case Selection.* Every 5 epochs, we construct a dynamic set of pediatric reference samples, termed meta-cases. Let  $\mathcal{P}$  denote the set of pediatric training samples. For each  $p \in \mathcal{P}$ , we compute its gradient:

$$g_p = \nabla_{\theta} \mathcal{L}_{\text{multi}}(x_p, y_p; \theta). \quad (4)$$

The reference pediatric gradient is the mean over  $\mathcal{P}$ :

$$G_{\text{ref}} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} g_p. \quad (5)$$

Each pediatric sample is then scored by cosine similarity with  $G_{\text{ref}}$ :

$$S(g_p, G_{\text{ref}}) = \frac{g_p \cdot G_{\text{ref}}}{\|g_p\| \|G_{\text{ref}}\|}. \quad (6)$$

The top 10 samples form the meta-case set  $M$ . This design ensures that reference cases are not chosen arbitrarily but are representative of the overall pediatric distribution, making gradient alignment more stable. Because  $M$  is updated as parameters evolve, it reflects the current training state as well (Fig. 3(b-2)).

(ii) *Meta-Gradient Reweighting.* The meta-cases  $M$  yield an aggregated pediatric meta-gradient:

$$G_m = \frac{1}{|M|} \sum_{m \in M} \nabla_{\theta} \mathcal{L}(x_m, y_m; \theta). \quad (7)$$

For each training sample  $(x_i, y_i)$  in batch  $B$ , we compute its sample-level gradient  $g_i$  (obtained via the same patch-based aggregation used during training) and measure its alignment with  $G_m$ :

$$S(g_i, G_m) = \frac{g_i \cdot G_m}{\|g_i\| \|G_m\|}. \quad (8)$$

Negatively aligned samples are suppressed using ReLU, followed by normalization:

$$\tilde{w}_i = \frac{\max(0, S(g_i, G_m))}{\sum_{j \in B} \max(0, S(g_j, G_m))}, \quad \forall i \in B. \quad (9)$$

The final reweighted objective becomes:

$$\mathcal{L}_{\text{final}} = \sum_{i \in B} \tilde{w}_i \left[ \mathcal{L}_{\text{Dice}}^{\text{multi}}(f_{\theta}(x_i), y_i) + \mathcal{L}_{\text{CE}}(f_{\theta}(x_i), y_i) \right]. \quad (10)$$

This strategy avoids the costly bi-level optimization commonly used in meta-learning-based reweighting methods [8]. Instead of repeatedly solving inner-outer loops or computing higher-order gradients, our approach relies on a single gradient alignment operation within standard training, with lower computational and memory overhead.

**Table 1.** Quantitative comparison of DA, FCL, MWN, L2RW, and MGR-DAS against the proposed method. <sup>†</sup>  $p < 0.05$  (paired t-test vs Proposed).

| Method          | DSC (%) $\uparrow$ |                    |              |                    | HD95 (mm) $\downarrow$ |                    |              |                    | Time $\downarrow$<br>(min/epoch) |
|-----------------|--------------------|--------------------|--------------|--------------------|------------------------|--------------------|--------------|--------------------|----------------------------------|
|                 | WT                 | TC                 | ET           | Avg                | WT                     | TC                 | ET           | Avg                |                                  |
| DA [6]          | 0.828 <sup>†</sup> | 0.808 <sup>†</sup> | 0.486        | 0.707 <sup>†</sup> | 23.65 <sup>†</sup>     | 20.52 <sup>†</sup> | 51.81        | 31.99 <sup>†</sup> | 1.83                             |
| FCL [14]        | 0.800 <sup>†</sup> | 0.728 <sup>†</sup> | 0.483        | 0.670 <sup>†</sup> | 31.74 <sup>†</sup>     | 22.26 <sup>†</sup> | 52.49        | 35.50 <sup>†</sup> | <b>1.32</b>                      |
| MWN [8]         | 0.789 <sup>†</sup> | 0.731 <sup>†</sup> | 0.481        | 0.667 <sup>†</sup> | 12.50                  | 15.62              | 92.83        | 40.32              | 26.42                            |
| L2RW [7]        | 0.797 <sup>†</sup> | 0.751 <sup>†</sup> | 0.482        | 0.677 <sup>†</sup> | 27.97 <sup>†</sup>     | 20.97 <sup>†</sup> | 52.28        | 33.74 <sup>†</sup> | 28.32                            |
| MGR-DAS [9, 10] | 0.805 <sup>†</sup> | 0.761 <sup>†</sup> | 0.483        | 0.683 <sup>†</sup> | 23.81 <sup>†</sup>     | 21.32 <sup>†</sup> | 53.83        | 32.98 <sup>†</sup> | 6.82                             |
| <b>Proposed</b> | <b>0.891</b>       | <b>0.865</b>       | <b>0.549</b> | <b>0.768</b>       | <b>10.04</b>           | <b>9.01</b>        | <b>14.39</b> | <b>11.15</b>       | 3.26                             |

### 2.3 Experiment Details

We conducted three experiments to evaluate the proposed framework: (i) a step-wise ablation study to quantify the contribution of two-stage training, meta-reweighting, and retrieval-based meta-case selection; (ii) comparison with existing mixed-training strategies; and (iii) robustness evaluation using four-fold cross-validation to verify consistency across different data splits, followed by evaluation across multiple backbone architectures to assess generalizability across different network designs. For fair comparison with prior work, all primary results are reported using a fixed pediatric train/test split (70/29 cases), while the four-fold cross-validation experiments were conducted separately as an additional robustness analysis (Table 3(A)).

Models were trained for 90 epochs using AdamW with an initial learning rate of  $1 \times 10^{-4}$  and a learning rate scheduler to ensure stable convergence. Training was performed with a batch size of 16 and a patch size of  $64 \times 64 \times 64$ . Performance was evaluated using the Dice Similarity Coefficient (DSC) and the 95th percentile Hausdorff Distance (HD95).

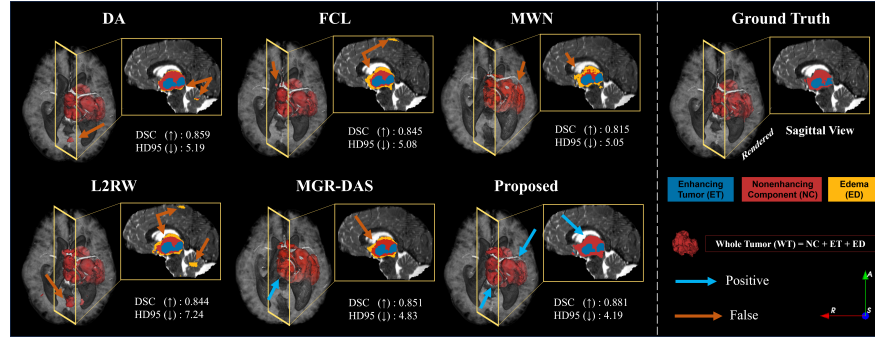
### 2.4 Baseline Study

We compare the proposed method with representative reweighting and domain generalization approaches, including supervised Domain Adaptation (DA) [6], which transfers knowledge between two models by first training on mixed-domain data and then adapting to pediatric data; Fixed Curriculum Learning (FCL) [14], which prioritizes pediatric cases in an easy-to-hard schedule; Meta-Weight-Net (MWN) [8], which learns adaptive sample weights via a lightweight MLP; Learning-to-Reweight (L2RW) [7], which optimizes per-sample weights through bilevel meta-learning; and MGR-DAS [9, 10], which weights samples based on gradient similarity. All methods are trained on the same mixed adult–pediatric dataset with identical backbone and training settings for fair comparison.

## 3 Results and Discussion

### 3.1 Comparison with Alternative Training Strategies

As shown in Table 1, DA achieved average DSC values of 0.707, while FCL, MWN, L2RW, and MGR-DAS provided modest gains with 0.670, 0.667, 0.677,



**Fig. 4.** 3D visualization of binary and multi-class tumor segmentation across methods. Most methods overestimated perilesional edema due to adult-type priors, whereas the proposed method demonstrated more accurate edema delineation

and 0.683 respectively. In contrast, the proposed method achieved the highest average DSC (0.768) and the lowest HD95 (11.15 mm). The visual results in Fig. 4 further confirm that our method reduces false positives and yields more precise tumor boundaries, demonstrating robust improvements in segmentation accuracy and boundary precision. These findings highlight that both the retrieval strategy and the progressive refinement introduced by the two-stage architecture are critical for achieving robust pediatric tumor segmentation.

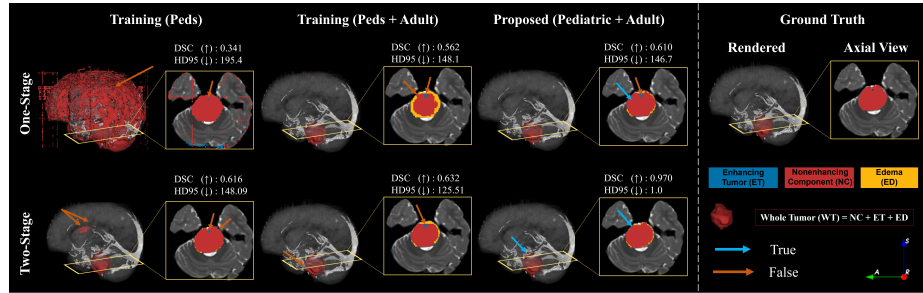
### 3.2 Ablation Study

Baseline mixed-domain training achieved a DSC of 0.661 (Table 2). Adding meta-reweighting improved performance to 0.677, demonstrating that our method can mitigate domain imbalance. Incorporating dynamic retrieval further increased DSC to 0.690, confirming the benefit of stable reference selection. The dual-stage design alone (without reweighting) achieved 0.725, while the full framework yielded the best performance (DSC 0.768, HD95 11.15 mm).

Overall, the two-stage framework improves DSC by 7.8% over single-stage mixed training (0.690→0.768), and surpasses pediatric-only (+5.4%) and naive

**Table 2.** Ablation study showing the effect of Mix Data, reweighting, retrieval, and dual-stage design on segmentation performance. The first row (baseline) corresponds to training using only pediatric data. ✓ indicates the component is used.

| Mix | Data | Reweight | Retrieval | Dual Stage | DSC (%) ↑    |              |              |              | HD95 (mm) ↓  |             |              |              |
|-----|------|----------|-----------|------------|--------------|--------------|--------------|--------------|--------------|-------------|--------------|--------------|
|     |      |          |           |            | WT           | TC           | ET           | Avg          | WT           | TC          | ET           | Avg          |
| ×   | ×    | ×        | ×         | ×          | 0.759        | 0.261        | 0.050        | 0.357        | 42.76        | 153.5       | 174.7        | 123.6        |
| ✓   | ×    | ×        | ×         | ×          | 0.792        | 0.710        | 0.482        | 0.661        | 36.57        | 17.48       | 53.20        | 35.75        |
| ✓   | ✓    | ×        | ×         | ×          | 0.790        | 0.754        | 0.488        | 0.677        | 23.24        | 19.28       | 53.87        | 32.11        |
| ✓   | ✓    | ✓        | ✓         | ×          | 0.803        | 0.786        | 0.480        | 0.690        | 23.10        | 23.01       | 52.84        | 32.98        |
| ✓   | ×    | ×        | ×         | ✓          | 0.883        | 0.842        | 0.450        | 0.725        | 14.43        | <b>8.85</b> | 52.03        | 25.10        |
| ✓   | ✓    | ✓        | ✓         | ✓          | <b>0.891</b> | <b>0.865</b> | <b>0.549</b> | <b>0.768</b> | <b>10.04</b> | 9.01        | <b>14.39</b> | <b>11.15</b> |



**Fig. 5.** 3D visualization results for binary segmentation and 2D slice results for multi-class segmentation on pediatric datasets.

mixed training (+4.3%). The proposed method also correctly predicts the absence of ET when ET is not present in the ground truth, indicating improved class-specific reliability and reduced over-segmentation Fig. 5. This confirms that decomposing into binary localization and multi-class refinement effectively addresses both data scarcity and domain imbalance.

### 3.3 Reproducibility Across Folds and Backbones

Across all folds, the proposed method consistently outperformed conventional mixed-domain training, yielding DSC improvements of +0.118, +0.178, +0.114, and +0.093, along with notable reductions in HD95 (up to −16.39 mm) see Table. 3. Consistent performance gains were also observed across U-Net [16], Swin UNETR [15], UNETR [13], and CNMCPMI2023 [18] backbones, with DSC increases of up to +0.249 and substantial HD95 reductions exceeding 100 mm for U-Net and Swin UNETR. These results demonstrate the stability, reproducibility, and backbone-agnostic effectiveness of the proposed strategy.

**Table 3.** Performance comparison between the proposed framework and conventional training across (A) four-fold cross-validation and (B) different backbone architectures. For (A), the four folds contain 25, 25, 25, and 24 testing cases, respectively. For (B), results are reported using the primary 70/29 pediatric train/test split.

| Mode                 | Setting     | DSC ↑        |              | HD95 ↓       |              |
|----------------------|-------------|--------------|--------------|--------------|--------------|
|                      |             | Proposed     | Conventional | Proposed     | Conventional |
| (A) Cross-Validation | Fold 1      | <b>0.789</b> | 0.671        | <b>29.41</b> | 37.58        |
|                      | Fold 2      | <b>0.789</b> | 0.611        | <b>16.88</b> | 32.80        |
|                      | Fold 3      | <b>0.733</b> | 0.619        | <b>29.13</b> | 45.52        |
|                      | Fold 4      | <b>0.789</b> | 0.696        | <b>23.49</b> | 26.02        |
| (B) Backbones        | U-Net       | <b>0.686</b> | 0.441        | <b>25.45</b> | 145.20       |
|                      | Swin UNETR  | <b>0.757</b> | 0.508        | <b>16.80</b> | 155.60       |
|                      | UNETR       | <b>0.768</b> | 0.651        | <b>11.15</b> | 29.01        |
|                      | CNMCPMI2023 | <b>0.822</b> | 0.715        | <b>8.60</b>  | 30.30        |

## 4 Conclusion

Our two-stage framework with meta-reweighting effectively mitigates data imbalance by enhancing pediatric-specific learning while leveraging adult data as complementary dataset. By adaptively assigning higher weights to pediatric cases—or to adult cases most similar to pediatric ones—the method achieves improved generalization across datasets dominated by adult samples. Beyond pediatric imaging, our method might be applied to other cohorts or tumor types.

**Acknowledgments.** This work was supported by the Korea Institute of Science and Technology (KIST) Institutional Program under Grant No. 26E0212, 26E0214. This work was also supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00561616).

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

## References

1. Kirscher, T., Faisan, S., Coubez, X., Barrier, L., Meyer, P.: PSAT: Pediatric segmentation approaches via adult augmentations and transfer learning. In: Proceedings of the 28th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) (2025).
2. Li, J., Zhu, G., Hua, C., Feng, M., Bennamoun, B., Li, P., Lu, X., Song, J., Shen, P., Xu, X., Mei, L., Zhang, L., Shah, S.A.A., Bennamoun, M.: A systematic collection of medical image datasets for deep learning. *ACM Computing Surveys* **56**(5), Article 116 (2023).
3. Wang, H., Jin, Q., Li, S., Liu, S., Wang, M., Song, Z.: A comprehensive survey on deep active learning in medical image analysis. *Medical Image Analysis* **95**, 103201 (2024).
4. Rajaraman, S., Yang, F., Zamzmi, G., Xue, Z., Antani, S.: Can deep adult lung segmentation models generalize to the pediatric population? *Expert Systems with Applications* **229**, 120531 (2023).
5. Chatterjee, D., Kanhere, A., Doo, F.X., Zhao, J., Chan, A., Welsh, A., Kulkarni, P., Trang, A., Parekh, V.S., Yi, P.H.: Children are not small adults: Addressing limited generalizability of an adult deep learning CT organ segmentation model to the pediatric population. *Journal of Imaging Informatics in Medicine* **38**, 1628–1641 (2025).
6. Fu, J., Bendazzoli, S., Smedby, Ö., Moreno, R.: Unsupervised domain adaptation for pediatric brain tumor segmentation. *arXiv preprint arXiv:2406.16848* (2024).
7. Ren, M., Zeng, W., Yang, B., Urtasun, R.: Learning to reweight examples for robust deep learning. In: *International Conference on Machine Learning (ICML)*, pp. 4334–4343 (2018).
8. Shu, J., Xie, Q., Yi, L., Zhao, Q., Zhou, S., Xu, Z., Meng, D.: Meta-weight-net: Learning an explicit mapping for sample weighting. In: *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 32, pp. 1917–1928 (2019).

9. Al-Fakih, A., Rezk, A.M., Shazly, A., Ryu, K., Al-masni, M.A.: Overcoming data scarcity: Brain tumor segmentation in pediatric and African populations. In: MIC-CAI Workshop on Data Engineering in Medical Imaging, pp. 11–20. Springer Nature Switzerland (2025).
10. Al-Fakih, A., Rezk, A.M., Shazly, A., Ryu, K., Al-masni, M.A.: Learning robust brain tumor segmentation under label corruption and data scarcity. *Eng. Appl. Artif. Intell.* 162(A), 112322 (2025)
11. de Verdier, M.C., Saluja, R., Gagnon, L., LaBella, D., Baid, U., Tahon, N.H., Foltyn-Dumitru, M., Zhang, J., Alafif, M., Baig, S.: The 2024 brain tumor segmentation (BraTS) challenge: Glioma segmentation on post-treatment MRI. *arXiv preprint arXiv:2405.18368* (2024).
12. Kazerooni, A.F., Khalili, N., Liu, X., Haldar, D., Jiang, Z., Zapaishchykova, A., Pavaine, J., Shah, L.M., Jones, B.V., Sheth, N.: BraTS-PEDs: Results of the multi-consortium international pediatric brain tumor segmentation challenge 2023. *arXiv preprint arXiv:2407.08855* (2024).
13. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: UNETR: Transformers for 3D medical image segmentation. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 574–584 (2022).
14. Hacothen, G., Weinshall, D.: On the power of curriculum learning in training deep networks. In: *Proceedings of the 36th International Conference on Machine Learning (ICML)*, vol. 97, pp. 2535–2544 (2019).
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241. Springer (2015).
16. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, BrainLes 2021, LNCS*, vol. 12962, pp. 272–284. Springer (2022).
17. Zeineldin, R.A., Karar, M.E., Burgert, O., Mathis-Ullrich, F.: Multimodal CNN networks for brain tumor segmentation in MRI: a BraTS 2022 challenge solution. *arXiv preprint arXiv:2212.09310* (2022). <https://doi.org/10.48550/arXiv.2212.09310>
18. Capellán-Martín, D., Jiang, Z., Parida, A., Liu, X., Lam, V., Nisar, H., Tapp, A., Elsharkawi, S., Ledesma-Carbayo, M.J., Anwar, S.M., Linguraru, M.G.: Model ensemble for brain tumor segmentation in magnetic resonance imaging. In: *Brain Tumor Segmentation and Cross-Modality Domain Adaptation for Medical Image Segmentation, Lecture Notes in Computer Science (LNCS)*, vol. 14669, pp. 221–232. Springer, Cham (2024).