Sub-Region-Specific Deep Learning Model for Pediatric Brain Tumor Segmentation on Multi-Parametric MRI

Sanskriti Srivastava[1], Thulani Kehara[1], Harsh Yadav[1], and Anup Singh[1,2,3]

[1]Centre for Biomedical Engineering, Indian Institute of Technology Delhi, India.
[2]Department of Biomedical Engineering, All India Institute of Medical Sciences, Delhi, India
[3]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, India.
anupsm@iitd.ac.in

Abstract. Pediatric brain tumors (PBTs) such as glioblastoma, diffuse midline glioma, and diffuse intrinsic pontine glioma remain highly aggressive, with survival rates below 20%. Accurate segmentation of tumor subregions, including the enhancing region (ET), non-enhancing components (NET), cystic component (CC), and edema (ED), is essential for diagnosis and treatment but remains difficult due to heterogeneous pediatric neuroanatomy and imaging limitations. Therefore, we proposed a deep learning framework for automated segmentation of PBTs. Preprocessing includes skull stripping, intensity normalization, and data augmentation, followed by training 3D U-Net and Swin UNETR models using a suitable combination of loss functions for each tumor subregion (ET, NET, ED, and CC). Furthermore, the best-performing model is selected to predict the subregion. ET and NET are best predicted by Swin UNETR, and CC and ED are best predicted by 3D U-Net, which are combined into a single prediction mask in a fixed priority order of ET, NET, CC, and ED to resolve overlapping predictions within the subregion. Test-time augmentation is applied to all subregion models, followed by post-processing based on connected-component analysis, which reduces false positives and improves the overall dice score of the whole tumor. The proposed framework achieves a global dice similarity coefficient of 0.86 and 0.88 on the validation data for the whole tumor and tumor core, respectively.

Keywords: Pediatric brain tumor segmentation, Multi-Parametric MRI, Swin UNETR, 3D U-Net

1 Introduction

Central nervous system tumors and their uncontrolled growth are a major cause of morbidity in children. They are the second most common childhood cancer and the leading cause of cancer-related mortality in children [1]. Glioblastomas (GBMs), diffuse midline gliomas (DMGs), and diffuse intrinsic pontine gliomas (DIPGs) represent the major pediatric brain tumor subtypes [2]. The prognosis of these high-grade gliomas (HGGs) in children is as poor as in adults, largely because they arise near critical midline structures such as the pons, anterior cerebellar margins, thalamus, spinal cord, and occasionally the entire brainstem [3]. These tumors constitute a highly aggressive group, accounting for approximately 15% of all central nervous system (CNS)

neoplasms in children and adolescents. Despite advances in biological understanding, clinical outcomes remain dismal, with 5-year survival rates consistently below 20% [3].

For the structured and qualitative assessment of pediatric brain tumors, MRI (Magnetic Resonance Imaging) plays a pivotal role. MRI is the safest and most reliable technique, offering a radiation-free environment and enabling multiparametric imaging that provides complementary information for different tissue characterization. MRI sequences serve specific purposes: T1-weighted for anatomy, T2-weighted for disease visualization, FLAIR for suppressing cerebrospinal fluid, and contrast-enhanced T1 for enhancing tumors [4].

However, children's developing brains make tumor boundaries difficult to define, and tumors often contain mixed regions such as cystic, necrotic, or edematous tissue that appear similar on scans. This uncertainty complicates grading, delays treatment, and hinders clinical trials. In response, Artificial Intelligence (AI) has emerged as a promising tool. Deep learning (DL) models, particularly convolutional neural networks (U-NET) [5], transformer networks (Swin UNETR) [6], and diffusion networks [7], have achieved high accuracy in tumor segmentation. However, automatic segmentation with DL models often surpasses human consistency, delivering faster, more precise results. Their ability to adapt to the specific characteristics of brain lesions enables them to deliver segmentation results that are consistently more precise and reliable than those achieved with traditional techniques.

Recent work on pediatric brain tumor segmentation has focused on segmenting the whole tumor and its subregions. A study by Boyd et al. used T2-weighted MRI to develop a deep learning pediatric brain tumor segmentation model via stepwise transfer learning [8]. Another study by Mulvany et al. used cascading of a DL model for the segmentation of the whole tumor and its subregions, using BraTS-PEDs data [9]. Karayegen et al. used the BraTS-PEDs dataset to segment pediatric brain tumors with their developed 3D deep learning (DL) model [10]. Vossough et al. also used BraTS-PEDs data and external data to develop U-Net and DeepMedic segmentation models for pediatric brain tumor segmentation [11]. However, despite these advances, the automated segmentation of pediatric tumors remains challenging, and several limitations persist, such as the study by Boyd et al., which used a single sequence for segmentation, and other studies that require high computational resources and extensive preprocessing for tumor segmentation.

Therefore, in this study, we proposed a framework for automatic pediatric brain tumor segmentation using a subregion-specific model to predict each tumor subregion. The preprocessing involves skull stripping, intensity normalization, and data augmentation, followed by training 3D U-Net and Swin UNETR models using combination of multiple loss functions for each tumor subregion (ET, NET, ED, and CC). Further, post-processing of predicted masks is performed to reduce false predictions and improve the model's overall dice coefficient.

2 Methods

2.1 Dataset Description

In this study, we utilized the dataset from the BraTS 2026 Challenge [12] on Segmentation of Pediatric Brain Tumor Subregions in Pre- and Post-Treatment Scans. During the training phase, a multiparametric MRI dataset comprising 294 pediatric patients was provided, each accompanied by ground truth masks delineating tumor subregions. For the validation phase, an additional dataset of 91 pediatric patients was made available, though without corresponding masks. The multi-parametric sequences are T1-weighted (T1W), contrast-enhanced T1-weighted (T1CE), T2-weighted (T2W), and T2-weighted Fluid-Attenuated Inversion Recovery (T2-FLAIR). The data have 4 labels, as shown in Fig 1.

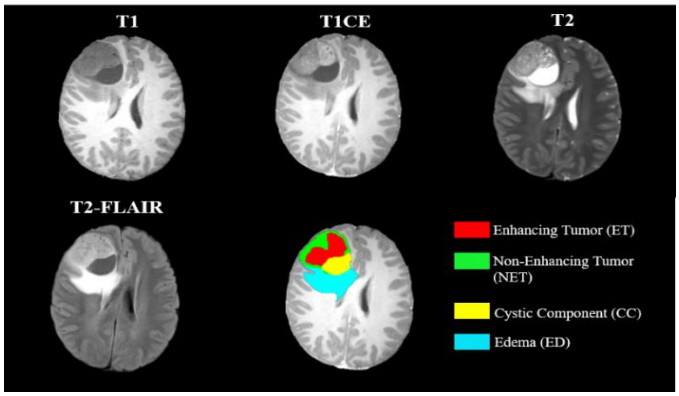


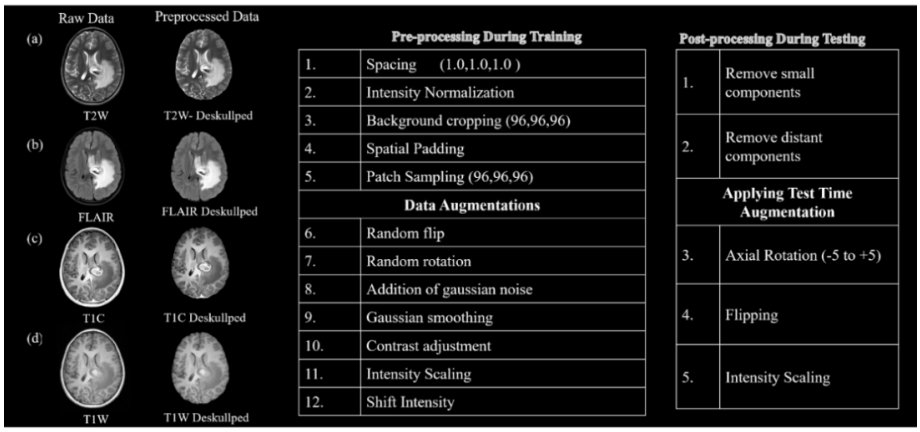
Fig 1. The four MRI sequences: here (a) T1-weighted (T1W); (b) contrast-enhanced T1-weighted (T1CE); (c) T2-weighted (T2W); (d) T2-weighted Fluid-Attenuated Inversion Recovery (T2-FLAIR); and (e) tumor subregion labels, where enhancing tumor (ET) is shown in red, the non-enhancing tumor (NET) in green, the cystic component (CC) in yellow, and the peritumoral edema (ED) in blue.

2.2 Preprocessing

Since the BraTS-PEDs 2026 dataset is only defaced and not skull-stripped, removing the skull and non-brain tissues is an important preprocessing step. For this, we used SynthStrip [13], a DL based tool from the FreeSurfer library that works across different MRI types, resolutions, and age groups. The T1CE sequence was chosen to generate the brain mask because it provides information on all anatomical features, including HGGs. This mask was then applied to all four MRI modalities: T1W, T1CE, T2W, and T2-FLAIR. By using the same mask across modalities, we ensured that the brain regions were perfectly aligned and that skull and non-brain tissues were consistently removed from each scan. Fig 2 shows the pipeline of the framework.

After skull stripping, we applied additional preprocessing steps to address the intensity variability in data, which arises from differences in imaging protocols, acquisition

parameters, and scanner manufacturers. First, intensity normalization was performed independently for each of the four MRI modalities to standardize the intensity distributions. Furthermore, foreground cropping was used to remove unnecessary background regions, maintaining a 10-voxel margin around the brain. During model training, we randomly sampled patches from the scans. To improve generalization, we added several data augmentation techniques, including random flipping, rotation, intensity shifting and scaling, as well as adding Gaussian noise and applying Gaussian smoothing.



Pre-processing During Training	
1.	Spacing (1.0,1.0,1.0)
2.	Intensity Normalization
3.	Background cropping (96,96,96)
4.	Spatial Padding
5.	Patch Sampling (96,96,96)
Data Augmentations	
6.	Random flip
7.	Random rotation
8.	Addition of gaussian noise
9.	Gaussian smoothing
10.	Contrast adjustment
11.	Intensity Scaling
12.	Shift Intensity

Post-processing During Testing	
1.	Remove small components
2.	Remove distant components
Applying Test Time Augmentation	
3.	Axial Rotation (-5 to +5)
4.	Flipping
5.	Intensity Scaling

Fig 2. Preprocessing and post-processing steps of the entire pipeline. Here, (a) represents the T2W MRI sequence; (b) represents the T2-FLAIR MRI sequence; (c) represents the T1CE MRI sequence; (d) represents the T1W MRI sequence.

2.3 Model Architectures

We implemented two DL architectures for whole-tumor and sub-region segmentation: 3D U-Net and Swin UNETR. For tumor segmentation, we adopted a one-model-per-subregion strategy. Instead of training a single multiclass model, four separate DL models were trained, each dedicated to one of the tumor subregions defined in the BraTS-PEDs 2026 challenge, enhancing tumor (ET), non-enhancing tumor (NET), cystic component (CC), and peritumoral edema (ED). For each model, the target subregion label was remapped to 1, while all other regions and backgrounds were set to 0. This design allowed each model to specialize in learning the unique intensity patterns and morphological features of its assigned subregion. By focusing on one subregion at a time, this approach improved segmentation accuracy compared to multiclass methods.

UNET: We developed four 3D U-Net [14] models for binary segmentation of individual brain tumor subregions. Each model processes four MRI modalities, T1W, T1CE, T2W, and T2-FLAIR, as input channels and produces a two-class output (background vs. subregion). The architecture follows an encoder-decoder design, built upon the original 3D U-Net framework. This structure leverages skip connections to preserve spatial details, enabling accurate localization of anatomical features such as tumor subregions.

Swin UNETR: We developed four Swin UNETR [15] models for each subregion, combining a hierarchical Swin Transformer encoder with shifted-window self-attention and

a U-Net-like decoder connected via multi-scale skip connections to produce volumetric segmentation outputs. The Swin transformer backbone (feature size=48) enables efficient long-range contextual modeling while maintaining computational efficiency.

2.4 Model Selection

The segmentation performance of each model was evaluated independently for each tumor component on the internal test dataset. The 3D U-Net achieved superior performance for CC and ED segmentation, whereas Swin UNETR yielded better results for ET and NET segmentation. Consequently, the final framework combined the best-performing architecture for each component: a 3D U-Net for CC and ED, and a Swin UNETR for ET and NET. Fig. 3 shows the proposed pipeline for brain tumor subregion segmentation.

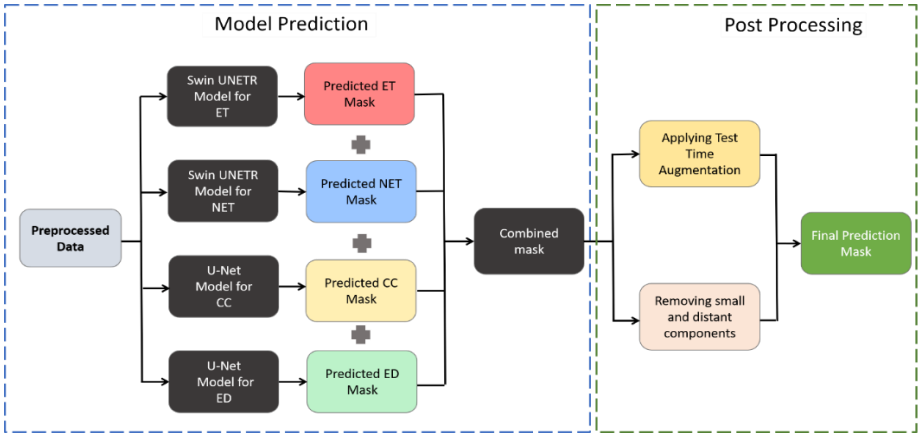


Fig 3. The proposed pipeline for segmentation of the brain tumor subregions.

2.5 Training Parameters

Both models were trained under a unified configuration using the AdamW [16] optimizer with an initial learning rate of 1×10−4, Batch Size of 4, and 200 epochs. Also, the models were trained separately using two patch sizes, (64 × 64 × 64) and (96 × 96 × 96) voxels, to investigate the effect of input spatial context on performance. For the segmentation of different subregions, different loss functions are employed. Three loss functions were utilized: Binary Cross-Entropy (BCE) [17], Focal Tversky Loss (FTL) [17], and Focal Dice Loss (FDL) [17]. For each region, a different combination of loss functions is employed to improve the overall DSC of the entire tumor, as shown in Table 1.

2.6 Post-processing

During inference, the four models generate independent binary prediction masks, which are combined into a single prediction mask in the fixed priority order of ET, NET, CC,

and ED to resolve overlapping predictions. The Test-Time Augmentation (TTA) [18] was applied to each subregion (ET, ED, NET, and CC) for mask prediction.

Table 1. Selection of a loss function for different subregions of the tumor.

	Brain tumor Subregion			
Model	**ED**	**CC**	**ET**	**NET**
3D U-Net	FTL + FDL	FTL + BCE	FTL + FDL + BCE	FTL +BCE
Swin UNETR	FTL + FDL	FTL+FDL+ BCE	FTL + FDL + BCE	FTL + FDL

Table description: FTL: Focal Tversky Loss, FDL: Focal Dice Loss, BCE: Binary Cross-Entropy, ED: Edema, CC: Cystic Component, ET: Enhancing Tumor, NET: Non-Enhancing Tumor.

Further, to reduce the false positives in isolated edema areas and improve overall dice, for each tumor subregion, any connected component smaller than the specified voxel count was removed during post processing by applying specified threshold such as for ET components smaller than 100 voxels were discarded, for NET components smaller than 200 voxels were discarded, for CC Components smaller than 500 voxels were discarded, similarly for ED region components smaller than 130 voxels were discarded. Also, the ED region located more than 15 mm away from the non-enhancing tumor was excluded. The application helps remove tiny, isolated regions that act as false positives by ensuring that only clinically meaningful structures remain in the segmentation. From the 294-case training set, 20 cases were randomly held out at the patient level to form an internal development set, used for architecture, loss, TTA, and post-processing threshold selection.

2.7 Performance Metrics

The segmentation performance was evaluated using the Dice Similarity Coefficient (DSC) [19] for the composite tumor regions. The whole tumor (WT) was defined as the combination of CC, ED, NET, and ET, while the tumor core (TC) comprised the NET, CC, and ET regions.

2.8 Training Hardware

This study was carried out on a high-performance computing system with dual NVIDIA A100 (40GB) and 2x Intel Xeon Platinum 8358 (32 cores, 2.6 GHz) connected to 200G HDR InfiniBand. The models were implemented in PyTorch 2.6.0 with MONAI 1.5.0, and cuDNN acceleration enabled. Multi-GPU parallelization was achieved using PyTorch's DataParallel module, allowing simultaneous use of multiple GPUs.

3 Results

Overall, the proposed framework, which combines 3D U-Net for the segmentation of CC and ED, and Swin UNETR for the segmentation of ET and NET, yields the best dice scores for the whole tumor and tumor core, at 0.86 and 0.88, respectively. The

model's performance on the test data (a randomly selected subset of the training data) and the validation data (provided by the BraTS challenge) are shown in Table 2. To evaluate the effect of patch size on model performance, models were trained with patch sizes of (64 × 64 × 64) and (96 × 96 × 96) voxels. The effect of patch size on the model performance is shown in Table 2. Fig 4 shows the best-performing and under-performing cases for the combined 3D U-Net and Swin UNETR model.

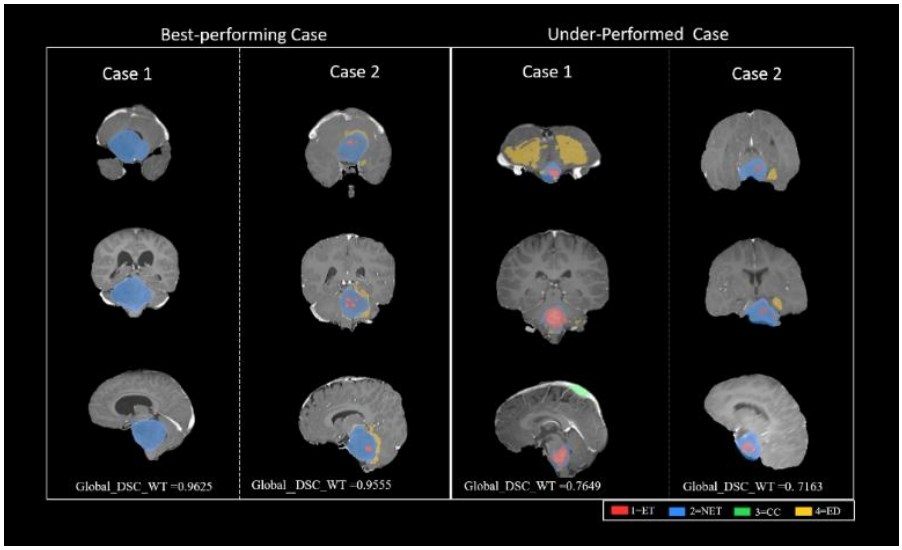


Fig 4. Comparison of brain tumor subregions segmentation between the best-performing and under-performing cases. Each case (Case 1 and Case 2) is shown across axial, coronal, and sagittal views.

Table 2. Segmentation performance comparison of 3D U-Net, Swin UNETR, optimized Swin UNETR, and the combination of 3D U-Net & Swin UNETR.

Evaluation on test data extracted from training dataset (20 cases)				
Models	**Global DSC NET**	**Global DSC ET**	**Global DSC Whole Tumor (WT)**	**Global DSC Tumor Core (TC)**
4×3D U-Net (64 × 64 × 64)	0.347	0.186	0.75	0.64
4×Swin UNETR (64 × 64 × 64)	0.565	0.528	0.76	0.68
4×Swin UNETR Optimized (96 × 96 × 96)	0.57	0.575	0.80	0.68
Combination of 3D U-Net & Swin UNETR (96 × 96 × 96)	0.57	0.68	0.82	0.70

Region-wise evaluation on separate validation data provided by BraTS-PED 2026 Challenge (91 cases)				
Models	**Global DSC NET**	**Global DSC ET**	**Global DSC Whole Tumor (WT)**	**Global DSC Tumor Core (TC)**
4×3D U-Net (64 × 64 × 64)	0.325	0.09	0.71	0.67
4×Swin UNETR (64 × 64 × 64)	0.551	0.103	0.83	0.84
4×Swin UNETR Optimized (96 × 96 × 96)	0.838	0.326	0.83	0.86
Combination of 3D U-Net & Swin UNETR (96 × 96 × 96)	0.848	0.426	0.86	0.88

4 Discussion

The proposed framework, which combines 3D U-Net for the segmentation of CC and ED, and Swin UNETR for the segmentation of ET and NET, yields the best dice scores for the whole tumor and tumor core, achieving a global DSC of 0.86 for WT and 0.88 for TC on the BraTS-PED 2026 challenge validation set, by leveraging the complementary strengths of two state-of-the-art deep learning architectures for different pediatric brain tumor subregions. Moreover, using different patches as inputs, applying different loss functions, and adding post-processing improve overall dice scores and make the pipeline more stable and generalized.

Increasing the training patch size from (64×64×64) to (96×96×96), the model gained improved dice scores. This led to a clear improvement in segmentation accuracy, with the NET dice score rising from 0.55 to 0.84. The larger patch size helped capture broader anatomical information and reduced false positives in the region of interest, resulting in more consistent labels. Additionally, using different loss functions for different subregions during training improved overall segmentation performance. Some studies have shown that using different loss functions for different subregions improves the overall DSC of the WT [20], [21]. FTL focuses on difficult-to-segment regions and reduces class imbalance. FDL enhances segmentation of small or irregular regions, while BCE improves voxel-wise prediction and boundary accuracy. So, FTL and FDL are common across all subregions. FTL, FDL, and BCE provide the most balanced approach for complex subregions, which is used in CC and ET while training Swin UNETR. The addition of TTA and post-processing steps further boosted performance; the dice scores improved to 0.88 for TC and 0.86 for WT, showing more stable predictions. The threshold values during post processing operation were optimized by testing multiple values on 20 test dataset and selecting those that maximized validation dice while minimizing the removal of true-positive voxels.

Studies have shown that using multiple ensembles and combining DL models improves overall segmentation accuracy and stabilizes the pipeline [22], [23]. Since pediatric brain tumor subregion segmentation is a challenging task, in this study, the component-specific strategy is employed to predict the mask. The WT consists of CC, ED,

NET, and ET, and the TC consists of ET, CC and NET. Consequently, selecting the best-performing architecture for each tumor subregion reduces segmentation errors in the final combined mask. Improved segmentation of individual subregions increases overall dice and makes the pipeline more clinically reliable. Fig 5 shows the improvement in mask prediction while combining 3D-UNET and Swin UNETR.

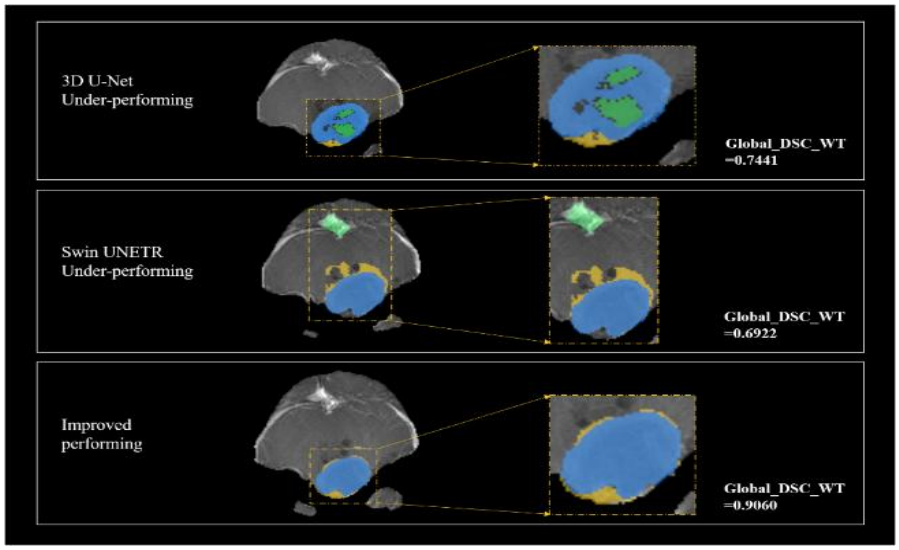


Fig 5. The improvement in WT dice when combining 3D U-Net for ED and CC prediction and Swin UNETR for NET and ET prediction

Also, the previous studies show that the diagnosis of brain tumors in infants and children is inherently challenging due to poorly defined anatomical structures and margins up to nine years of age[25]. Consequently, the identification and grading (Grade III and IV) of tumors in these regions becomes a challenging task. In contrast, the delineation of tumor subregions, such as cystic components, necrotic areas, and edematous localization, appears almost indistinct [2]. This complexity introduces significant variability in medical diagnosis and contributes to a cascade of clinical challenges, including diagnostic uncertainty, surgical limitations, and prognostic inaccuracy. The 3D U-Net has an encoder-decoder architecture with skip connections that effectively retain fine-grained local structures. This improves segmentation of CC and ED, which often appear as irregular, diffuse, or fragmented regions that require precise voxel-level detail. Whereas Swin UNETR uses a transformer architecture that considers the entire image, helping it identifies NET and ET by understanding the bigger picture and accurately detecting intensity changes across slices. Hence, by leveraging the strength of convolution and a transformer-based model, the overall dice score improved. Fig 6 shows the boxplots of the DSC distribution for the WT using the U-NET, Swin UNETR, and the combination of U-NET and Swin UNETR.

This study has some limitations; the combined 3D U-Net and Swin UNETR model performs well on NET and ET but struggles to predict CC and ED. In these regions, the dice score was often below 0.1. The 3D U-Net often produced false positives, while the Swin UNETR generalized better but still struggled to detect edema. The prevalence of CC in the dataset was around 38%, and the ED boundary is diffuse and has intensity overlapping with surrounding tissue, which creates data imbalance and makes it hard for these models to perform well across CC and ED. Hence, to improve the stability, robustness, and generalizability of the model, it should be trained on larger and more diverse datasets and validated across independent multicentric cohorts. Federated learning can facilitate the collaborative use of data from multiple institutions while preserving data privacy, thereby enabling access to larger and more diverse datasets without requiring direct data sharing [24].

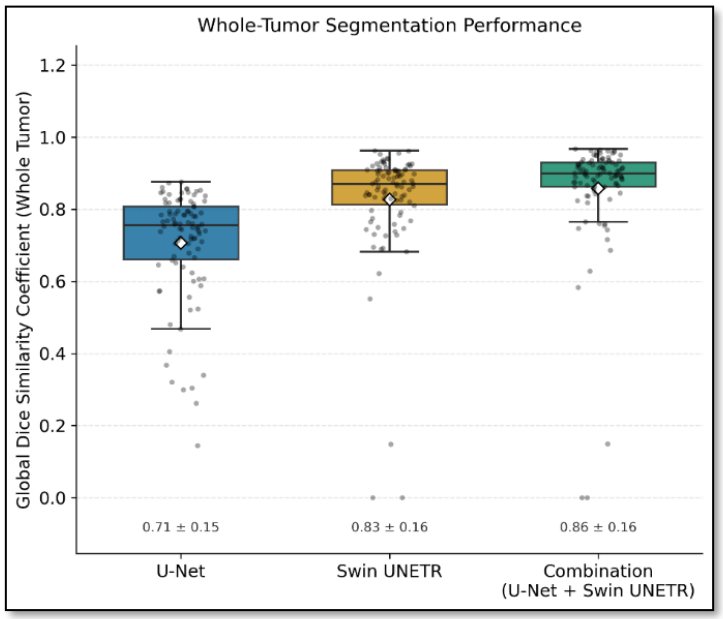


Fig 6. The box plot of the distribution of the Global Dice Similarity Coefficient (DSC) for whole-tumor (WT) segmentation across U-NET, Swin UNETR, and a combination of 3D U-Net and Swin UNETR models, evaluated on the 91-case challenge validation data.

5 Conclusion

The proposed framework provides a fast and effective approach for predicting pediatric brain tumors by leveraging the complementary strengths of convolutional and transformer-based networks. Furthermore, using larger input patches, along with a combination of optimal loss functions and post-processing, improves the overall performance

of the framework. Future work will focus on improving the dice for the CC and ED to make the framework more robust and generalizable for clinical use.

Code Availability. https://github.com/Kehara2002/BraTS-PED-2026.

Acknowledgments. The authors acknowledge funding support from ICMR (project number: CAR-2024-01-000187) and WIN-COE (project number: IA-0000000037), the IIT Delhi HPC facility for computational resources, and thank all MedImg lab members for their guidance and support.

Disclosure of Interests. The authors declare no conflicts of interest.

References

1. N. Raja, L. Hayes, N. Basta, and R. J. Q. McNally, "International trends in the incidence of brain tumours in children and young-adults and their association with indicators of economic development," Cancer Epidemiol., vol. 74, Oct. 2021, doi: 10.1016/j.canep.2021.102006.
2. T. J. MacDonald, D. Aguilera, and C. M. Kramm, "Treatment of high-grade glioma in children and adolescents," Neuro. Oncol., vol. 13, no. 10, pp. 1049–1058, Oct. 2011, doi: 10.1093/neuonc/nor092.
3. E. Szychot et al., "European standard clinical practice recommendations for paediatric high-grade gliomas," EJC Paediatric Oncology, vol. 5, Jun. 2025, doi: 10.1016/j.ejcped.2024.100210.
4. F. G. Gonçalves, A. N. Viaene, and A. Vossough, "Advanced Magnetic Resonance Imaging in Pediatric Glioblastomas," Nov. 10, 2021, Frontiers Media S.A. doi: 10.3389/fneur.2021.733323.
5. S. Maurya, V. Kumar Yadav, S. Agarwal, and A. Singh, "Brain Tumor Segmentation in mpMRI Scans (BraTS-2021) Using Models Based on U-Net Architecture," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2022, pp. 312–323. doi: 10.1007/978-3-031-09002-8_28.
6. H. Pang, W. Guo, and C. Ye, "Multi-modal brain MRI synthesis based on SwinUNETR," Jun. 2025, [Online]. Available: http://arxiv.org/abs/2506.02467
7. Z. Xing et al., "Diff-UNet: A diffusion embedded network for robust 3D medical image segmentation," Med. Image Anal., vol. 105, Oct. 2025, doi: 10.1016/j.media.2025.103654.
8. A. Boyd et al., "Stepwise Transfer Learning for Expert-level Pediatric Brain Tumor MRI Segmentation in a Limited Data Scenario," Radiol. Artif. Intell., vol. 6, no. 4, Jul. 2024, doi: 10.1148/ryai.230254.
9. T. Mulvany and D. Griffiths-King, "Segmentation of Pediatric Brain Tumors using a Radiologically informed, Deep Learning Cascade." Doi: https://www.synapse.org/Synapse:syn53708249/wiki/626323
10. G. Karayegen and M. F. Aksahin, "Brain tumor prediction on MR images with semantic segmentation by using deep learning network and 3D imaging of tumor region," Biomed. Signal Process. Control, vol. 66, Apr. 2021, doi: 10.1016/j.bspc.2021.102458.
11. A. Vossough et al., "Training and Comparison of nnU-Net and DeepMedic Methods for Autosegmentation of Pediatric Brain Tumors," American Journal of Neuroradiology, vol. 45, no. 8, pp. 1081–1089, Aug. 2024, doi: 10.3174/ajnr.A8293.

12. A. F. Kazerooni et al., “The Brain Tumor Segmentation in Pediatrics (BraTS-PEDs) Challenge: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs),” *arXiv preprint arXiv:2404.15009*, 2024, doi: 10.48550/arXiv.2404.15009.
13. A. Hoopes, J. S. Mora, A. V. Dalca, B. Fischl, and M. Hoffmann, “SynthStrip: skull-stripping for any brain image,” Neuroimage, vol. 260, Oct. 2022, doi: 10.1016/j.neuroimage.2022.119474.
14. O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Verlag, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.
15. A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, “Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images.” Doi: https://monai.io/research/swin-unetr
16. I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” in *Proceedings of the 7th International Conference on Learning Representations (ICLR)*, New Orleans, LA, USA, 2019.
17. M. Yeung, E. Sala, C.-B. Schönlieb, and L. Rundo, “Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation,” *Computerized Medical Imaging and Graphics*, vol. 95, p. 102026, Jan. 2022. doi: 10.1016/j.compmedimag.2021.102026.
18. N. Moshkov, B. Mathe, A. Kertesz-Farkas, R. Hollandi, and P. Horvath, “Test-time augmentation for deep learning-based cell segmentation on microscopy images,” Sci. Rep., vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-61808-3.
19. Zou, K. H., Warfield, S. K., Bharatha, A., et al. Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index. *Academic Radiology*. 2004;11(2):178–189.
20. G. Visalakshi and L. Mohan, “A novel sub-differentiable hausdorff loss combined with BCE for MRI brain tumor segmentation using UNet variants,” Sci. Rep., vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-33136-x.
21. Montazerolghaem M, Sun Y, Sasso G, Haworth A. U-Net Architecture for Prostate Segmentation: The Impact of Loss Function on System Performance. Bioengineering (Basel). 2023 Mar 26;10(4):412. doi: 10.3390/bioengineering10040412.
22. S. Srivastava, K. Raghuwanshi, and A. Singh, “Condition-Based Ensemble Modelling of Swin UNETR and 3D U-Net for Meningioma Segmentation in Radiotherapy Planning,” in Lecture Notes in Computer Science, Springer Science and Business Media Deutschland GmbH, 2026, pp. 148–158. doi: 10.1007/978-3-032-16365-3_14.
23. S. Pandey, K.-F. Chen, and E. B. Dam, “Comprehensive Multimodal Segmentation in Medical Imaging: Combining YOLOv8 with SAM and HQ-SAM Models.”24
24. A. Karargyris et al., “Federated benchmarking of medical artificial intelligence with MedPerf,” *Nature Machine Intelligence*, vol. 5, pp. 799–810, 2023, doi: 10.1038/s42256-023-00652-2.
25. B. B. Paraxatovna, ’Anatomical and Physiological Characteristics of Early and Preschool-Age Children," American Journal of Medicine and Medical Sciences, vol. 15, no. 6, pp. 1646 to 1649, 2025, doi: 10.5923/j.ajmms.20251506.02.