



Improving Pre-trained Adult Glioma Segmentation Models Using only Post-processing Techniques

Abhijeet Parida^{1,2}, Daniel Capellán-Martín^{1,2}, Zhifan Jiang¹,
Nishad Kulkarni¹, Krithika Iyer¹, Austin Tapp¹, Syed Muhammad Anwar^{1,3},
María J. Ledesma-Carbayo², and Marius George Linguraru^{1,3}(✉)

¹ Sheikh Zayed Institute for Pediatric Surgical Innovation, Children's National Hospital, Washington, DC, USA

² Universidad Politécnica de Madrid and CIBER-BBN, ISCIII, Madrid, Spain

³ School of Medicine and Health Sciences, George Washington University, Washington, DC, USA

milingura@childrensnational.org

Abstract. Gliomas are the most common malignant brain tumors in adults and are among the most lethal. Despite aggressive treatment, the median survival rate is less than 15 months. Accurate multiparametric MRI (mpMRI) tumor segmentation is critical for surgical planning, radiotherapy, and disease monitoring. While deep learning models have improved the accuracy of automated segmentation, large-scale pre-trained models generalize poorly and often underperform, producing systematic errors such as false positives, label swaps, and slice discontinuities in slices. These limitations are further compounded by unequal access to GPU resources and the growing environmental cost of large-scale model training. In this work, we propose adaptive post-processing techniques to refine the quality of glioma segmentations produced by large-scale pretrained models developed for various types of tumors. We demonstrated the techniques in multiple BraTS 2025 segmentation challenge tasks, with the ranking metric improving by 14.9 % for the sub-Saharan Africa challenge and 0.9% for the adult glioma challenge. This approach promotes a shift in brain tumor segmentation research from increasingly complex model architectures to efficient, clinically aligned post-processing strategies that are precise, computationally fair, and sustainable.

Keywords: Brain MRI · BraTS Challenge · Glioma segmentation · Medical image analysis · Resource-aware AI

1 Introduction

Gliomas are the most common malignant brain tumors in adults and remain among the deadliest among cancer types. Despite aggressive treatment strategies (maximal surgical resection, chemotherapy, radiotherapy), median overall

A. Parida, D. Capellán-Martín and Z. Jiang—These authors contributed equally.

survival duration is around 15 months. Malignant gliomas collectively account for approximately 2.5% of all cancer-related deaths [26]. Considering these alarming facts, there is a pressing need for tools that enable earlier detection, accurate characterization, and effective treatment planning to improve patient outcomes. Central to the clinical management of gliomas is multi-parametric magnetic resonance imaging (mpMRI), which provides non-invasive, high-resolution visualization of tumor anatomy and physiology. mpMRI-based tumor segmentation of tumors is routinely used to: (i) guide surgical resection margins [27], (ii) paint radiotherapy dose [8], and (iii) monitor volumetric disease progression [10]. Despite its critical importance, tumor segmentation is still performed manually: an iterative and labor intensive process that can consume substantial expert time per case [4]. In addition, the manual segmentation process is highly subjective, reducing the reproducibility between observers. Several deep learning models have recently been proposed to segment various pathologies [1, 30]. Automating the tumor segmentation step with deep learning models promises significant time savings and greater standardization across centers.

Since 2012, Brain Tumor Segmentation (BraTS) challenges, held in conjunction with the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), have become a leading contributor towards the development and benchmarking of automated tumor segmentation models by providing large, expertly annotated, and standardized datasets. The 2025 edition introduced the largest expert-annotated glioma data set and standardized lesion-wise normalized surface distance (NSD) metrics [5, 22]. During this period, the best performing methods have primarily relied on the nnU-Net framework [14], which extends the original U-Net architecture [24] with modality- and dataset-specific adaptations to deliver state-of-the-art performance for 3D biomedical image tasks. In contrast, large-scale pre-trained models underperform in glioma segmentation and may perform inconsistently and generalize poorly even after light fine-tuning [12]. Their residual errors (tiny false positive islands), segmentation label swaps, or slice-wise discontinuities are systematic and easily correctable. Previous BraTS winners have shown that post-processing steps such as ensemble voting [21], size-sensitive connected component filtering [9], ET-to-NCR relabeling heuristics [9], and adaptive refinement [16, 23] can significantly improve segmentation accuracy.

Beyond the choice of model and post-processing algorithms, socioeconomic conditions also strongly affect the challenge outcomes. Access to GPUs remains highly unequal, with recent surveys showing that GPU resources are highly concentrated in high-income settings, leaving researchers, especially in low and middle-income countries, without adequate resources [19]. Moreover, training complex deep learning models on GPUs consumes considerable energy; therefore, the environmental impact of using GPUs is becoming impossible to ignore. However, post hoc refinement using simple image-processing techniques typically requires only a few CPU hours. Thus, adaptive post-processing offers a compute-democratic path, allowing all steps to run on commodity CPUs.

This can broaden meaningful participation in BraTS while simultaneously shrinking the carbon footprint of medical-AI research.

Taken together, the saturation of performance gains in large-scale pre-trained segmentation model architectures, the unequal access to computational resources, and the environmental costs of large-scale training underscore the need for more efficient alternatives. Therefore, this article proposes samplewise adaptive post-processing techniques on foundation models for the segmentation of adult gliomas (**GLI**) and adult gliomas in sub-Saharan African patients (**SSA**). We show performance gains over the base foundation model without additional GPU training. We advocate a shift in brain MRI segmentation research from ever deeper backbones toward smarter, greener, and clinically aligned post-processing – promoting accuracy, computational equity, and sustainability.

2 Previous Work

2.1 Glioma Segmentation Models

The *BraTS orchestrator* [18] provides dockerized access to the best performing solutions from previous BraTS challenges, enabling fair, side-by-side evaluation of winning models. Since these models generalize well and have been trained on large datasets using the best practices to get the best performance, they are considered our large-scale pre-trained segmentation models. For pretreatment adult glioma (**GLI-pre**), the BraTS 2023 winner combined SwinUNETR [13] and nnU-Net [14] in an ensemble trained on real and GAN-augmented data [11]. A closely related ensemble from the same group secured the first place in the post-treatment GLI (**GLI-post**) task at BraTS 2024.

For adult glioma in sub-Saharan African patients (**SSA**), the leading model of BraTS 2024 first pre-trained on the large GLI-pre cohort and then fine-tuned on the modest SSA dataset, using a MedNeXT [25]nnU-Net ensemble [23]. Although these architectures differ in backbone details, their training recipes converge on heavy GPU usage, large-scale augmentation, and multifold ensembling to squeeze out the last increments in Dice.

2.2 Adaptive Post-processing

To mitigate systematic residual errors that survive end-to-end training, several groups have proposed *adaptive post-processing*, i.e., selecting sample-specific refinements rather than applying a fixed post-processing pipeline to every case. Jiang *et al.* [15] formalized the concept and demonstrated that radiomic features of the predicted mask can guide which operations (connected-component filtering, ET→NCR relabeling, etc.) maximize Dice gain for a given sample. Follow-up work clustered cases in radiomic space to learn cluster-wise thresholds for component removal [16], and Parida *et al.* [23] also applied similar heuristic rules to win the BraTS 2024 SSA challenge. These refinements run entirely on CPUs and boost generalization.

3 Challenge and Data Description

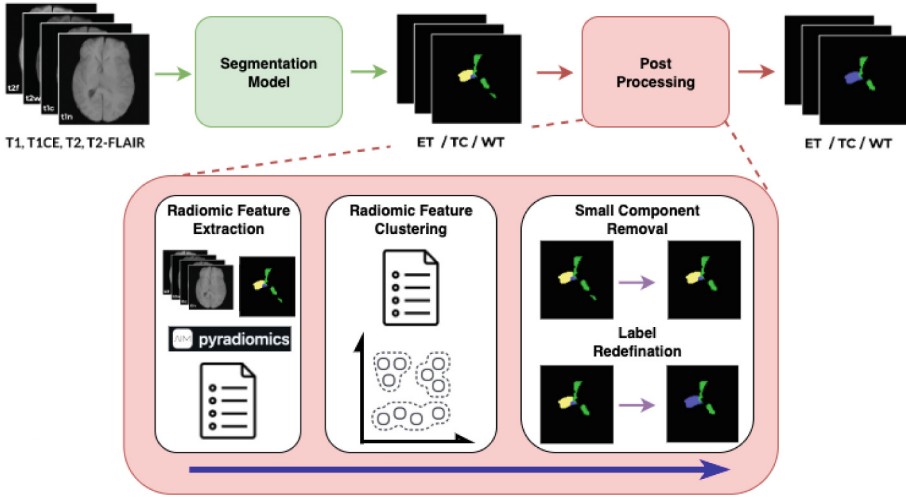


Fig. 1. The process pipeline shows the high-level overview of our proposed post-processing pipeline. It shows the major steps of PyRadiomics-based feature extraction, clustering of samples based on the radiomic signature, removal of small components, and small label redefinition to get the final segmentation.

BraTS 2024 GLI [6, 7, 29] aims at automatic segmentation of diffuse gliomas on multi-institutional, clinically acquired pre and post-treatment mpMRI scans. Each case provides four sequences collected routinely: precontrast T1-weighted (T1), contrast-enhanced T1-weighted (T1CE), T2-weighted (T2), and T2-weighted fluid-attenuated inversion recovery (T2-FLAIR). The current version contains 1,251 GLI-pre and 1,350 GLI-post training cases. For the validation set, there are 219 GLI-pre and 188 GLI-post cases. Ground truth annotations cover six clinically relevant regions: non-enhancing tumor core (NETC, label 1), surrounding non-enhancing FLAIR hyperintensity (SNFH, label 2), enhancing tumor tissue (ET, label 3), resection cavity (RC, label 4), the tumor core (TC = ET + NETC), and the whole tumor (WT = ET + SNFH + NETC). The challenge ranking is based on the Dice and the normalized surface distance (NSD) computed per lesion.

BraTS 2024 SSA [2, 3] is the largest public collection of annotated pretreatment glioma scans from adult African patients, including low-grade gliomas and glioblastomas. It uses four MRI sequences (T1, T1CE, T2, T2-FLAIR) and the same label taxonomy (NETC, SNFH, ET, TC, WT) as the GLI task. The data set comprises 60 training and 35 validation cases; test labels are withheld. Evaluation employs identical Dice/NSD metrics.

4 Methods

We built on large-scale pre-trained segmentation networks and included a task specific, three stage post-processing pipeline (Fig. 1). The stages were (i) *radiomic feature extraction and case clustering*, (ii) *thresholding to delete small, isolated components*, and (iii) *thresholding to fix label mix-ups*.

We optimized the thresholds and chose the best model with the ranking approach proposed by the BraTS team LaBella *et al.* [20]. LaBella *et al.* stated - “The evaluation is done in a hidden test set, computing lesion-wise metrics in all regions and comparing ranks of all the submissions rather than the metrics”. We followed the idea and replicated this procedure, we built an internal ranking metric that produced a single score. The ranking metric was optimized for a lower score where a smaller value means better performance on Dice and NSD at the 0.5 and 1 mm thresholds for each of the regions. The code is available on GitHub ².

The ranking metric approach was chosen because it was robust to outlier predictions. Further, the ranking metric allowed us to optimize for a single value while aligning with the contest evaluation pipeline.

Radiomic Feature Extraction and Clustering: For the prediction of the WT ensemble, we calculated 386 radiomic features using *PyRadiomics* [28], following the protocol of Jiang *et al.* [17]. Therefore, each case had- 14 shape descriptors that capture tumor geometry, and 93 intensity & texture descriptors- for each of the four MRI sequences (T1, T1Gd, T2, FLAIR). Using principal component analysis (PCA), we retained the principal components that explain 90% of the variance and then partitioned the cases with k means clustering [15,16]. The optimal number of clusters was determined by maximizing the silhouette coefficient in the training folds. Each new case was assigned to the nearest cluster at test time, allowing for the post-processing to be applied on the basis of its radiomic signature.

Threshold Identification for Removing Small Components (p_{cc}): Within each cluster and for each label (NETC, SNFH, ET, and RC), we performed a grid search over minimum size thresholds to retain the lesion [23]. Each threshold was evaluated on a cross-validated ranking metric; the threshold that minimized the ranking metrics was selected. The application of a sample-specific cluster p_{cc} removed tiny disconnected islands, noise that would otherwise inflate the false positive count.

Threshold Identification for Label Redefinition (lbl_{redef}): A second adaptive search fine-tuned the consistency between labels. Jiang *et al.* [15] suggested a similar approach after the removal of the noisy component. For example, if the enhancing tumor fraction fell below its cluster-specific cut-off point, ET voxels were relabeled as non-enhancing core or surrounding edema; an analogous

² <https://github.com/Pediatric-Accelerated-Intelligence-Lab/BraTS-Unofficial-Ranker>

rule was applied to ED when its share of WT was too small. We propose correcting systematic label confusions in a data-driven way. After removing small components, we built a confusion matrix over all cross-validated predictions to identify pairs of frequently swapped labels. For every such pair (lbl_x, lbl_y) , we searched, within each cluster, for the cut-off point on the ratio lbl_x/WT that minimizes the ranking metric. If a new case fell below this cutoff point, all lbl_x voxels are converted to lbl_y . This ratio-based *lblredef* step enforced anatomically plausible label volumes and improved the performance on the BraTS metrics at the lesion level.

5 Implementation Details

For each task, we used the best segmentation models from the previous BraTS edition (see Sect. 2.1). These models were trained on BraTS 2024 datasets, which is identical to the current edition BraTS 2025 data. GLI-pre and GLI-post were treated as separate tasks for post-processing (GLI-post included RC). The labels being the same, the SSA task used the same pipeline for hyperparameter search as GLI-pre. The post-processing used the binary labels produced by the segmentation models as input.

6 Results

Table 1 provides an overview of the performance evaluation of our post-processed models for the validation set of the GLI and SSA tasks. Additionally, results on the testing set are included in Table 2. The reported numbers are obtained from the automatic pipeline setup in the BraTS 2025 digital platform with no access to the ground truth of the validation set and no access to any testing data including

Table 1. Quantitative results on the validation subset of GLI and SSA. Lesion-wise (LW) Dice coefficients and Normalized Surface Distance (NSD) with thresholds of 1.0 mm were computed for enhancing tumor (ET), tumor core (TC), whole tumor (WT), non-enhancing tumor core (NETC), surrounding non-enhancing FLAIR hyperintensity (SNFH), and resection cavity (RC), respectively. (\uparrow) represents a metric where a higher value is better and (\downarrow) represents a metric where a lower value is better. The best performing ranking metric for each task is highlighted in **bold**. SM: segmentation model.

Task	Model	GPU Time(hrs)	Ranking Metric(\downarrow)	LW Dice(\uparrow)					LW NSD(\uparrow)						
				ET	TC	WT	NETC	SNFH	RC	ET	TC	WT	NETC	SNFH	RC
GLI (n = 407)	SM	401 [§]	1.137	0.794	0.798	0.881	0.756	0.823	0.858	0.838	0.782	0.832	0.770	0.819	0.860
	SM + p_{cc}	0	1.129	0.794	0.798	0.881	0.756	0.824	0.859	0.838	0.782	0.832	0.770	0.819	0.862
	SM + p_{cc} + <i>lblredef</i>	0	1.127	0.794	0.798	0.881	0.757	0.824	0.859	0.838	0.782	0.832	0.771	0.819	0.862
SSA (n = 35)	SM	168	1.729	0.870	0.865	0.926					0.871	0.812	0.849		
	SM + p_{cc}	0	1.629	0.870	0.865	0.927					0.872	0.812	0.849		
	SM + p_{cc} + <i>lblredef</i>	0	1.471	0.870	0.865	0.927					0.872	0.812	0.849		

[§] estimated based on the segmentation models used; does not include synthetic data generation training

images and labels. Submission-related CSVs were downloaded from the platform and used to obtain the ranking metric. Figure 2 illustrated qualitative results on validation cases for SSA and GLI.

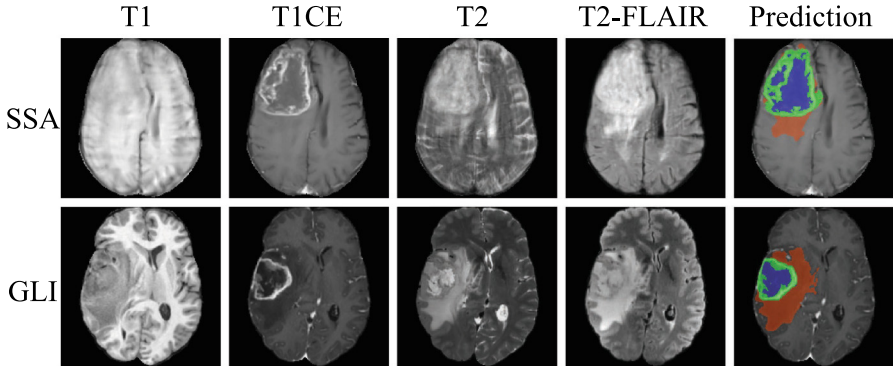


Fig. 2. Qualitative results showing median lesion-wise Dice of the whole tumor. SSA: 0.959, orange = ED, blue = NCR, green = ET; GLI: 0.947, orange = SNFH, blue = NETC, green = ET. (Color figure online)

Table 2. Quantitative results on the test subset of GLI and SSA. These results were calculated on a hidden test set and provided by the organizers post-challenge. Lesion-wise (LW) Dice coefficients and Normalized Surface Distance (NSD) with thresholds of 1.0mm were computed for enhancing tumor (ET), tumor core (TC), whole tumor (WT), non-enhancing tumor core (NETC), surrounding non-enhancing FLAIR hyper-intensity (SNFH), and resection cavity (RC), respectively. (\uparrow) represents a metric where a higher value is better and (\downarrow) represents a metric where a lower value is better. SM: segmentation model.

Task	Model	LW Dice(\uparrow)				LW NSD(\downarrow)			
		ET	TC	WT	RC	ET	TC	WT	RC
GLI	SM + p_{cc} + <i>lblredef</i>	0.813	0.813	0.882	0.894	0.855	0.821	0.851	0.892
SSA	SM + p_{cc} + <i>lblredef</i>	0.900	0.912	0.934		0.904	0.865	0.872	

For the GLI task in Table 1, the lesion-wise Dice scores increase from 0.756 to 0.757 for NETC, 0.823 to 0.824 for SNFH and 0.858 to 0.859 for RC due to post-processing. Lesion-wise NSD also improves, from 0.770 to 0.771 for NETC and 0.860 to 0.862 for RC. Together, these refinements reduce the overall ranking metric from 1.137 to 1.127, showing that the post-processed results perform better than the original segmentation model.

TRUE LABEL	BG	1.00	0.00	0.03	0.01
	LBL1	0.00	0.93	0.01	0.02
	LBL2	0.00	0.02	0.95	0.01
	LBL3	0.00	0.05	0.01	0.96
		BG	LBL1	LBL2	LBL3
		PREDICTED LABEL			

(a) GLI-pre

TRUE LABEL	BG	1.00	0.00	0.05	0.02	0.03
	LBL1	0.00	0.89	0.00	0.01	0.01
	LBL2	0.00	0.03	0.95	0.01	0.01
	LBL3	0.00	0.05	0.00	0.95	0.00
	LBL4	0.00	0.03	0.00	0.01	0.95
		BG	LBL1	LBL2	LBL3	LBL4
		PREDICTED LABEL				

(b) GLI-post

Fig. 3. Confusion matrices illustrating systematic label confusions across cross-validated predictions of GLI-pre and GLI-post after pp_{cc} . These confusion matrices are used to identify the labels that are redefined as part of lbl_{redef} . For example in GLI-pre lbl_1 is redefined to lbl_3 based on the lbl_1/WT ratio.

Also, in Table 1 for the SSA task, post-processing improves the lesion-wise Dice of the WT from 0.926 to 0.927 and the lesion-wise NSD for the ET from 0.871 to 0.872. The ranking metric improved from 1.729 to 1.471.

In Fig. 3a, we see the confusion matrix for the GLI-pre cross-validated set after the pp_{cc} step. We see that the biggest error is lbl_3 falsely predicted as lbl_1 . Similarly, in Fig. 3b for GLI-post the biggest errors are lbl_3 falsely predicted as lbl_1 . These error regions are the focus of lbl_{redef} , where lbl_1/WT are identified as $lbl_1 \rightarrow lbl_3$ for each of the clusters.

7 Discussion

For the GLI validation set, adaptive post-processing nudged metrics only slightly (NETC Dice +0.001, SNFH Dice +0.001, RC NSD +0.002), improving the BraTS ranking score by just 0.9 % (1.137 \rightarrow 1.127). The ensemble is therefore already near the task ceiling; further gains will likely need (i) stronger anatomical priors for smoother boundaries or (ii) fine-tuning on the few failure cases.

In the resource-limited SSA cohort, the Dice scores were unchanged (≤ 0.001), but the ranking score improved by 14.9 % (1.729 \rightarrow 1.471). Because this metric penalizes severely poor segmented subjects, radiomics-guided threshold tuning helped improve the segmentation scores for outliers. This could be largely attributed to the poor quality of SSA data acquisition, leading to more false positives in the segmentations from the pre-trained model. This leads to better result improvement after applying the post-processing strategies.

Training the full ensemble models costs 401 GPU hours (GLI) and 168 GPU hours (SSA). In contrast, the entire post-processing pipeline—PyRadiomics feature extraction, k means clustering, and a grid search over p_{cc} and $lblrdef$ thresholds—ran on CPUs and did not use any GPU time.

It is important to note that the optimization for the tasks was tailored to improve the BraTS ranking score, which does not necessarily translate into clinically better segmentations. For medical relevance, optimizing directly for Dice or NSD, or a combination of both, may be more meaningful than maximizing challenge ranking, as these metrics better capture overlap and boundary accuracy in a clinical setting.

The rules of connected component and volume ratio appear saturated in GLI; integrating shape descriptors, vascular atlases, or uncertainty maps could further smooth out the boundaries. Because radiomics are derived from predicted masks, an iterative loop that alternates segmentation and feature extraction is another avenue for improvement. Further improvement in radiomics extraction can be improved by calculating the radiomics on the WT of a lesion instead of the WT of the entire case.

Finally, to facilitate reproducibility and extend the utility of our adaptive post-processing approach, we have made the complete pipeline publicly available as easy-to-use Docker containers and a webapp. This enables researchers and clinicians to easily deploy, test, and build on our methods without the need for complex setups. The Docker images are hosted at: <https://hub.docker.com/r/aparida12/brats2025> and the webapp is accessible at: <https://segmenter.hope4kids.io/>.

8 Conclusion

We demonstrated that an adaptive, radiomic-based postprocessing pipeline improved the accuracy of brain segmentation models on multiple BraTS 2025 glioma segmentation tasks with zero additional GPU hours. These findings promote shifting the focus from building increasingly larger models or further resource intensive training schemes, towards data-driven efficient post-processing strategies. By fine-tuning large-scale segmentation models using adaptive post-processing, we can make automated brain tumor segmentation both accurate and equitable.

Acknowledgments. This work was supported by the National Cancer Institute (UG3 CA236536), the Spanish Ministerio de Ciencia e Innovación, the Agencia Estatal de Investigación, NextGenerationEU grants PDC2022-133865-I00 and PID2022-141493OB-I00, and the EUCAIM project co-funded by the European Union (Grant Agreement #101100633). The authors acknowledge the Universidad Politécnica de Madrid for providing computing resources on the Magerit Supercomputer.

References

1. Abidin, Z.U., Naqvi, R.A., Haider, A., Kim, H.S., Jeong, D., Lee, S.W.: Recent deep learning-based brain tumor segmentation models using multi-modality magnetic resonance imaging: a prospective survey. *Front. Bioeng. Biotechnol.* **12**, 1392807 (2024)
2. Adewole, M., et al.: The BraTS-Africa dataset: expanding the brain tumor segmentation data to capture African populations. *Radiol. Artif. Intell.* **7**(4), e240528 (2025)
3. Adewole, M., et al.: The brain tumor segmentation (BraTS) challenge 2023: Glioma segmentation in sub-saharan Africa patient population (BraTS-Africa). *ArXiv pp. arXiv-2305* (2023)
4. Anantharajan, S., Gunasekaran, S., Subramanian, T., R, V.: MRI brain tumor detection using deep learning and machine learning approaches. *Meas. Sensors* **31**, 101026 (2024). <https://doi.org/10.1016/j.measen.2024.101026>. <https://www.sciencedirect.com/science/article/pii/S2665917424000023>
5. Baid, U., et al.: The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. *CoRR* [arxiv:2107.02314](https://arxiv.org/abs/2107.02314) (2021)
6. Baid, U., et al.: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification (2021). <https://arxiv.org/abs/2107.02314>
7. Bakas, S., et al.: Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Sci. Data* **4**(1), 1–13 (2017)
8. Brighi, C., et al.: Repeatability of radiotherapy dose-painting prescriptions derived from a multiparametric magnetic resonance imaging model of glioblastoma infiltration. *Phys. Imaging Radiat. Oncol.* **23**, 8–15 (2022)
9. Capellán-Martín, D., et al.: Model ensemble for brain tumor segmentation in magnetic resonance imaging. In: *International Challenge on Cross-Modality Domain Adaptation for Medical Image Segmentation*, pp. 221–232. Springer, Heidelberg (2023). https://doi.org/10.1007/978-3-031-76163-8_20
10. Climent Pardo, J.C., et al.: Deep learning volumetrics reveal distinct clinical trajectories for pediatric low-grade gliomas under surveillance: a multicenter study. *Neuro-Oncol. Adva. vdafl45* (2025)
11. Ferreira, A., Solak, N., Li, J., Dammann, P., Kleesiek, J., Alves, V., Egger, J.: How we won BraTS 2023 adult glioma challenge? Just faking it! enhanced synthetic data augmentation and model ensemble for brain tumour segmentation. *arXiv preprint* [arXiv:2402.17317](https://arxiv.org/abs/2402.17317) (2024)
12. Fu, G., et al.: Comparing foundation models and nnU-Net for segmentation of primary brain lymphoma on clinical routine post-contrast T1-weighted MRI. In: Gimi, B.S., Krol, A. (eds.) *Medical Imaging 2025: Clinical and Biomedical Imaging*, vol. 13410, p. 1341019. International Society for Optics and Photonics, SPIE (2025). <https://doi.org/10.1117/12.3044679>
13. 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: *International MICCAI Brainlesion Workshop*, pp. 272–284. Springer, Heidelberg (2021). https://doi.org/10.1007/978-3-031-76163-8_20
14. Isensee, F., Jaeger, P.F., Kohl, S.A., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211 (2021)

15. Jiang, Z., et al.: Enhancing generalizability in brain tumor segmentation: model ensemble with adaptive post-processing. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI), pp. 1–4. IEEE (2024)
16. Jiang, Z., Capellán-Martín, D., Parida, A., Tapp, A., Liu, X., Ledesma-Carbayo, M.J., Anwar, S.M., Linguraru, M.G.: Magnetic resonance imaging feature-based subtyping and model ensemble for enhanced brain tumor segmentation. arXiv preprint [arXiv:2412.04094](https://arxiv.org/abs/2412.04094) (2024)
17. Jiang, Z., et al.: Automatic visual acuity loss prediction in children with optic pathway gliomas using magnetic resonance imaging. In: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1–5. IEEE (2023)
18. Kofler, F., et al.: BraTS orchestrator: democratizing and disseminating state-of-the-art brain tumor image analysis. arXiv preprint [arXiv:2506.13807](https://arxiv.org/abs/2506.13807) (2025)
19. Kudiabor, H.: AI's computing gap: academics lack access to powerful chips needed for research. *Nature* **636**(8041), 16–17 (2024)
20. LaBella, D., et al.: Analysis of the BraTS 2023 intracranial meningioma segmentation challenge. *J. Mach. Learn. Biomed. Imaging* **3**(March 2025), 38–58 (2025)
21. Maani, F., Hashmi, A.U.R., Aljuboory, M., Saeed, N., Sobirov, I., Yaqub, M.: Advanced tumor segmentation in medical imaging: an ensemble approach for BraTS 2023 adult glioma and pediatric tumor tasks. In: International Challenge on Cross-Modality Domain Adaptation for Medical Image Segmentation, pp. 264–277. Springer, Heidelberg (2023). https://doi.org/10.1007/978-3-031-76163-8_24
22. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BraTS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2015). <https://doi.org/10.1109/TMI.2014.2377694>
23. Parida, A., et al.: Adult Glioma Segmentation in Sub-Saharan Africa using Transfer Learning on Stratified Finetuning Data. arXiv preprint [arXiv:2412.04111](https://arxiv.org/abs/2412.04111) (2024)
24. 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.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
25. Roy, S., et al.: Mednext: transformer-driven scaling of convnets for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 405–415. Springer, Heidelberg (2023). https://doi.org/10.1007/978-3-031-43901-8_39
26. Sabouri, M., Dogonchi, A.F., Shafiei, M., Tehrani, D.S.: Survival rate of patient with glioblastoma: a population-based study. *Egypt. J. Neurosurg.* **39**(1), 42 (2024)
27. Shaver, M.M., et al.: Optimizing neuro-oncology imaging: a review of deep learning approaches for glioma imaging. *Cancers* **11**(6), 829 (2019)
28. Van Griethuysen, J.J., et al.: Computational radiomics system to decode the radiographic phenotype. *Can. Res.* **77**(21), e104–e107 (2017)
29. de Verdier, M.C., et al.: The 2024 brain tumor segmentation (BraTS) challenge: glioma segmentation on post-treatment MRI. arXiv preprint [arXiv:2405.18368](https://arxiv.org/abs/2405.18368) (2024)
30. Verma, A., Yadav, A.K.: Brain tumor segmentation with deep learning: current approaches and future perspectives. *J. Neurosci. Methods* 110424 (2025)