



# Adaptable Segmentation Pipeline for Diverse Brain Tumors with Radiomic-Guided Subtyping and Lesion-Wise Model Ensemble

Daniel Capellán-Martín<sup>1,2</sup>, Abhijeet Parida<sup>1,2</sup>, Zhifan Jiang<sup>1</sup>,  
Nishad Kulkarni<sup>1</sup>, Krithika Iyer<sup>1</sup>, Austin Tapp<sup>1</sup>, Syed Muhammad Anwar<sup>1,3</sup>,  
María J. Ledesma-Carbayo<sup>2</sup>, and Marius George Linguraru<sup>1,3</sup>(✉)

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

<sup>2</sup> Universidad Politécnica de Madrid and CIBER-BBN, ISCIII, Madrid, Spain

<sup>3</sup> School of Medicine and Health Sciences, George Washington University, Washington, DC, USA

milingura@childrensnational.org

**Abstract.** Robust and generalizable segmentation of brain tumors on multi-parametric magnetic resonance imaging (MRI) remains difficult because tumor types differ widely. The BraTS 2025 Lighthouse Challenge benchmarks segmentation methods on diverse high-quality datasets of adult and pediatric tumors: multi-consortium international pediatric brain tumor segmentation (PED), preoperative meningioma tumor segmentation (MEN), meningioma radiotherapy segmentation (MEN-RT), and segmentation of pre- and post-treatment brain metastases (MET). We present a flexible, modular, and adaptable pipeline that improves segmentation performance by selecting and combining state-of-the-art models and applying tumor- and lesion-specific processing before and after training. Radiomic features extracted from MRI help detect tumor subtype, ensuring a more balanced training. Custom lesion-level performance metrics determine the influence of each model in the ensemble and optimize post-processing that further refines the predictions, enabling the workflow to tailor every step to each case. On the BraTS testing sets, our pipeline achieved performance comparable to top-ranked algorithms across multiple challenges. These findings confirm that custom lesion-aware processing and model selection yield robust segmentations yet without locking the method to a specific network architecture. Our method has the potential for quantitative tumor measurement in clinical practice, supporting diagnosis and prognosis.

**Keywords:** Brain tumor segmentation · MedNeXt · Meningiomas · Metastases · MRI · nnU-Net · Pediatric brain tumors · Tumor subtyping

D. Capellán-Martín, A. Parida and Z. Jiang—Equal contribution.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2026  
S. Bakas et al. (Eds.): MICCAI 2025, LNCS 16376, pp. 456–467, 2026.

[https://doi.org/10.1007/978-3-032-16365-3\\_41](https://doi.org/10.1007/978-3-032-16365-3_41)

# 1 Introduction

In the United States, cancer is the second leading cause of death overall and the primary cause among individuals younger than 85 years old. Brain and other central nervous system tumors are the top cause of cancer death in children and adolescents under 20, and brain tumors also lead cancer mortality in men aged 20–39. In 2025, about two million new cancer cases and 618,000 cancer-related deaths are projected in the U.S. [24]. Early, accurate diagnosis is essential to improve outcomes, but wide variation in tumor appearance across imaging devices and population makes consistent assessment difficult, highlighting the need for reliable quantitative diagnostic and prognostic tools.

In neuro-oncology, accurate segmentation of brain tumors in multi-parametric magnetic resonance imaging (mpMRI) is a fundamental step for diagnosis, treatment planning, and longitudinal monitoring. Yet manual contouring remains labor-intensive and prone to inter-observer variability, motivating robust automated solutions in the clinical workflow.

The international Brain Tumor Segmentation (BraTS) Challenge [1–4, 13, 20], organized in conjunction with the Medical Image Computing and Computer Assisted Intervention (MICCAI) conference, has driven and benchmarked segmentation algorithmic innovation since MICCAI 2012. In its 2025 edition, BraTS expanded into a suite of challenges that cover a broader range of tumors and tasks, earning recognition as one of MICCAI’s three Lighthouse Challenges.

Recent deep learning models such as nnU-Net [8], MedNeXt [23], Swin UNETR [7, 25], and emerging Mamba-based hybrids [18, 28] have defined the current state of the art (SOTA) in medical image segmentation. As the SOTA models have become more stable and robust across diverse tasks, their reported Dice scores have converged within narrow margins [9], signaling a performance plateau. While ever more complex architectures are possible, they would offer diminishing added values.

BraTS-winning solutions have demonstrated that model-independent and GPU-free heuristics, including ensemble voting [19], size-aware connected component filtering and ET-to-NCR relabelling [5], and adaptive refinement schemes [11, 22], can substantially boost segmentation performance. Instead of pursuing ever-deeper or more intricate networks, we believe that the next gains will arise from target-specific pipeline design. Robust pre-processing such as radiomic-guided fold splitting mitigates sampling bias; lesion-aware ensemble weighting exploits complementary backbone strengths; and subtype-specific post-processing removes residual artifacts.

We therefore introduce and apply a flexible, modular, backbone-independent pipeline to multiple BraTS 2025 Lighthouse Challenges to show its versatility. This end-to-end strategy consistently outperforms any single backbone, demonstrating that thoughtful pipeline engineering remains promising once model performance plateaus.

## 2 Methods

### 2.1 Dataset

Our segmentation pipeline was applied to four BraTS 2025 Lighthouse tasks: multi-consortium international pediatric brain tumor segmentation (PED; train = 261, val = 91) [6,14], preoperative meningioma tumor segmentation (MEN; train = 1000, val = 141) [15], meningioma radiotherapy segmentation (MEN-RT; train = 500, val = 70) [17], and segmentation of pre- and post-treatment brain metastases (MET; train = 1296, val = 179) [21]. Each case contains co-registered isotropic pre-contrast T1-weighted (T1), contrast-enhanced T1-weighted (T1CE), T2-weighted (T2), and T2-weighted fluid-attenuated inversion recovery (T2-FLAIR) MRI sequences. In MEN-RT cohort, only T1CE is available and in its original image space.

Reference standard annotations of the tumor sub-regions were created and approved by expert neuroradiologists. Concretely the following tumor annotations and subregions were defined for each of the Lighthouse challenges considered in this work.

- **PED:** enhancing tumor (ET), non-enhancing tumor (NET), cystic component (CC), peritumoral edema (ED), tumor core (TC=ET+NET+CC), and whole tumor (WT=TC+ED).
- **MEN:** ET, non-enhancing tumor core (NETC), surrounding non-enhancing FLAIR hyperintensity (SNFH), TC (ET+NETC), and WT (TC+SNFH).
- **MEN-RT:** gross tumor volume (GTV).
- **MET:** ET, NETC, SNFH, resection cavity (RC; delineates the resection of region within the brain in post-treatment cases), TC (ET+NETC), and WT (TC+SNFH).

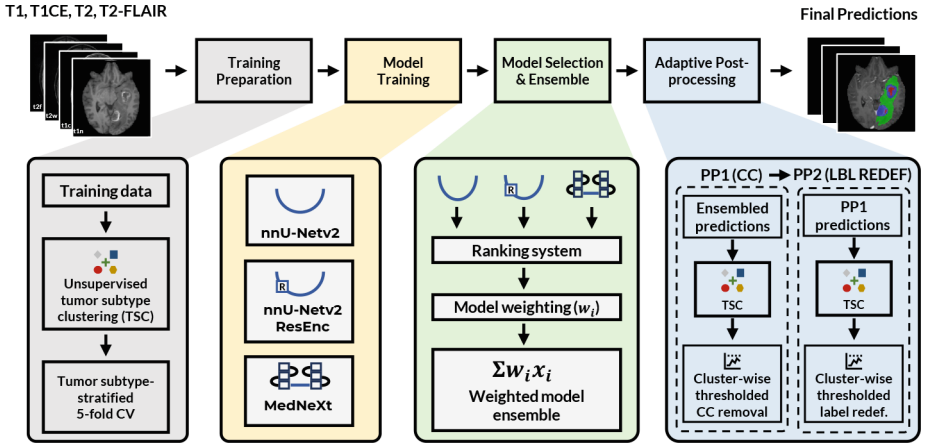
For more details, please refer to the challenge website: <https://www.synapse.org/Synapse:syn64153130/wiki/>.

### 2.2 Segmentation Pipeline

Our pipeline uses MRI radiomic features and lesion-wise metrics to adapt the workflow to each tumor subtype. As summarized in Fig. 1, it comprises four stages: (i) training data preparation: stratified fold splitting based on radiomic feature clustering; (ii) model training; (iii) model selection and ensembling; and (iv) adaptive post-processing, including optimal removal of small connected components and label redefinition.

### 2.3 Stratified Training Data Preparation

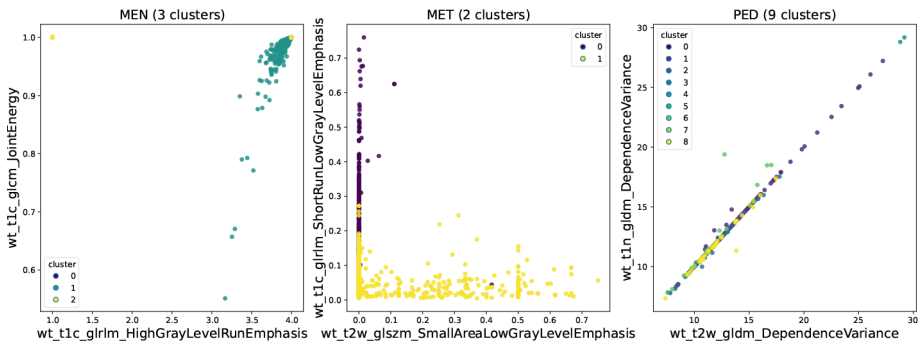
In five-fold cross-validation (5-fold CV) training, stratified and balanced folds outperform purely random fold splits. While stratification by tumor volume has been explored, we argue that radiomic features-capturing the full heterogeneity of tumor appearance-provide a more robust basis for fold assignment. For



**Fig. 1.** Overview of the segmentation pipeline. PP, CC, and LBL REDEF refer to post-processing, connected components, and label redefinition, respectively.

each case in the training set, we computed 14 shape- and 93 appearance-based radiomic features per MRI sequence based on WT mask provided by reference annotation, using PyRadiomics [26], following the protocol of Jiang *et al.* [12].

To reduce the large number of radiomic features, we keep only the principal components that explain 90% of the variance and then partition the cases with *k*-means clustering using reduced features [10,11]. The optimal number of clusters is determined by maximizing the silhouette coefficient in the training folds. Each cluster is randomly split into five folds, yielding the final CV sets for training. Figure 2 illustrates the clustering based on the two radiomic features most relevant for cluster separation.



**Fig. 2.** Clustering visualization across challenges using top two radiomic features.

## 2.4 Model Training

Based on the reported performance in [9] and our previous experience, we selected three state-of-the-art models: nnU-Net V2, nnU-Net with residual encoder (nnU-Net ResEnc M), and MedNeXt M (k=3, 17.6M parameters, 248 GFlops). Each model was trained using a 5-fold CV strategy with input image patches of  $128 \times 128 \times 128$  voxels. The nnU-Net variants and MedNeXt-M were trained using same settings, employing label-wise softmax activation, a class-weighted loss function combining Dice and cross-entropy losses, optimized by the stochastic gradient descent (SGD) with Nesterov momentum (initial learning rate=0.01, momentum=0.99, weight decay=3e-5), for 200 epochs on NVIDIA A100 (40 GB) GPUs. The implementation is available through the official frameworks repositories: <https://github.com/MIC-DKFZ/nnUNet> and <https://github.com/MIC-DKFZ/MedNeXt>.

## 2.5 Label-Wise Metrics as Objective Function $F$

As model training relies on a loss function, each downstream stage (model selection, ensembling, and post-processing) also needs a quantitative objective so the pipeline can be customized automatically to each tumor type and region. This objective function can simply be the Dice similarity coefficient (DSC). However, in reality or in the context of challenges, DSC alone is insufficient because segmentation performance is evaluated with several metrics, including DSC, normalized surface distance (NSD) and Hausdorff distance (HD), which sit on different scales and can not be added directly. Also, integrating scores across tumour sub-regions (ET, TC, WT etc.) compounds another challenge.

We therefore adopt the ranking strategy proposed by BraTS [16]. For every candidate prediction obtained on the training set using 5-fold CV, we compute lesion-wise metrics for all sub-regions, rank the predictions case-wise, and average these ranks across metrics and regions. The resulting internal CV score  $F$  is a single value (smaller is better) that combines all lesion-wise metrics, aligns with the official BraTS leaderboard and remains robust to outliers. Our implementation is open-source and available at [github.com/Pediatric-Accelerated-Intelligence-Lab/BraTS-Unofficial-Ranker](https://github.com/Pediatric-Accelerated-Intelligence-Lab/BraTS-Unofficial-Ranker).

## 2.6 Model Selection and Ensemble

Each trained model  $M_i$  is ranked by the approach in Sect. 2.5 with a score  $F_i$ . Then we used a weighted model ensemble strategy when adding the averaged 5-fold probability output from each model. The weight of each model is defined as  $W_i = \sum_{j \neq i} F_j / \sum_i F_i$ . The weights for each model are summarized in Table 1.

We also evaluated the STAPLE [27] ensemble strategy. It lagged behind the weighted ensemble when we used the three 5-fold-averaged predictions on the training data. However, STAPLE performed better on the validation set for MEN-RT task because treating each fold's model as an independent candidate yielded 15 predictions, which markedly enhanced DSC, NSD, and rank for the MEN-RT validation cohort.

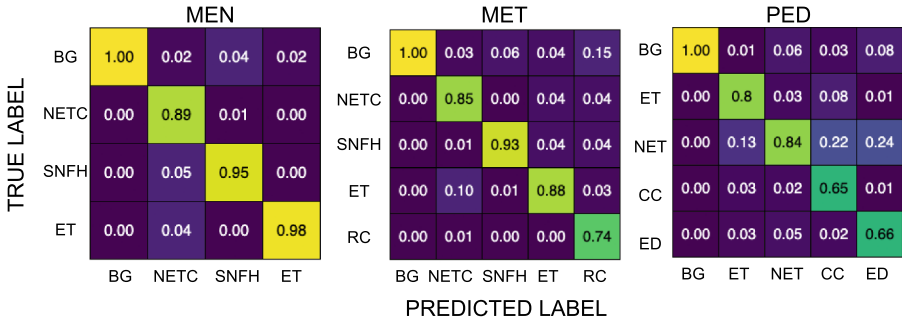
**Table 1.** Weights for each model across tasks.

	PED	MEN	MEN-RT	MET
$W_0$ (nnU-Net V2)	0.332	0.323	0.336	0.295
$W_1$ (nnU-Net ResEnc M)	0.331	0.326	0.338	0.296
$W_2$ (MedNeXt M)	0.337	0.351	0.326	0.409

## 2.7 Adaptive Post-processing

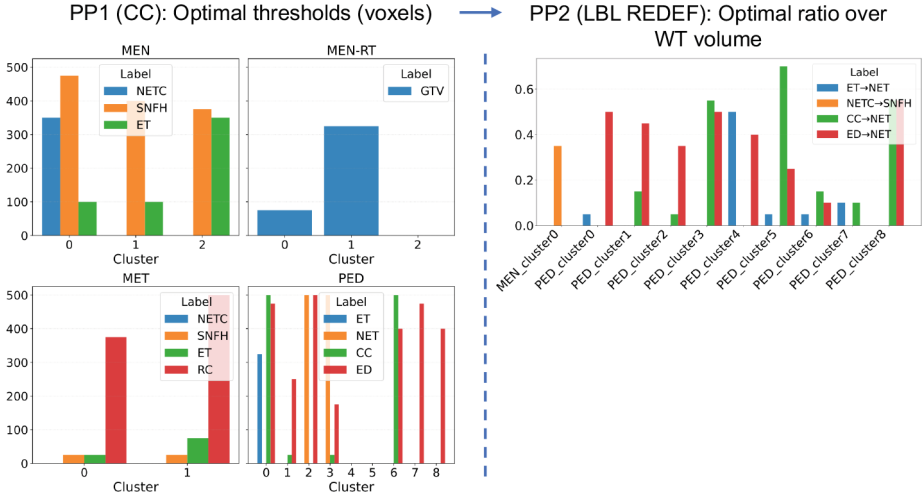
After ensembling, we recompute stratification clusters on the predicted masks, enabling cluster- and label-specific thresholds for adaptive post-processing.

**Post-processing for the Removal of Isolated Connected Components (PP1-CC):** To remove small isolated CCs that are likely false or noisy predictions, we perform a grid search across clusters and labels over volumes from 0 to 500 voxels in 25-voxel increments. For each threshold, we filter the predicted mask and compute the internal ranking score, and finally select the threshold that achieves the best averaged rank. This cluster- and label-specific filter eliminates fragments that would otherwise be counted as false-positive lesions.



**Fig. 3.** Post-PP1 confusion matrix indicating labels that need to be redefined in PP2. BG refers to background.

**Post-processing for Label Redefinition (PP2-LBLREDEF):** A second adaptive PP fine-tunes the consistency of tumor sub-regions (labels). We correct systematic label confusions in a data-driven way. After PP1-CC, we build a confusion matrix (Fig. 3) over all predicted masks again to identify pairs of frequently swapped labels. For every such pair  $(label_x, label_y)$ , we search, within each cluster, for the threshold on the volume ratio  $label_x/WT$  that maximizes internal CV metrics (*e.g.* rank). If a case with predicted mask falls below this threshold, all  $label_x$  voxels are converted to  $label_y$ . This ratio-based PP2-LBLREDEF enforces anatomically plausible label volumes and improves the performance of the BraTS metrics at the lesion level.



**Fig. 4.** Adaptive Post-processing: optimal thresholds per cluster and label across challenges.

At inference, a new test case is assigned to its nearest radiomic cluster, after model ensemble and the cluster-specific post-processing threshold values are then applied. Figure 4 shows the optimal thresholds identified by CV-based grid search.

### 3 Results

The evaluation of the model prediction on the validation set was performed on the Synapse platform. The models were assessed for each of the tumor regions using the lesion-wise DSC and NSD with boundary threshold of 1.0 mm.

Quantitative results of our models across the validation and testing data for each challenge are shown in Tables 2 and 3, for PED, MET, MEN, and MEN-RT, respectively. These evaluations were performed automatically by the challenge’s digital platform, with no access to the reference standard annotations on the validation set and no access to any testing data including images and labels. Figure 5 illustrated qualitative results on validation cases for PED, MET, MEN, and MEN-RT, respectively.

### 4 Discussion

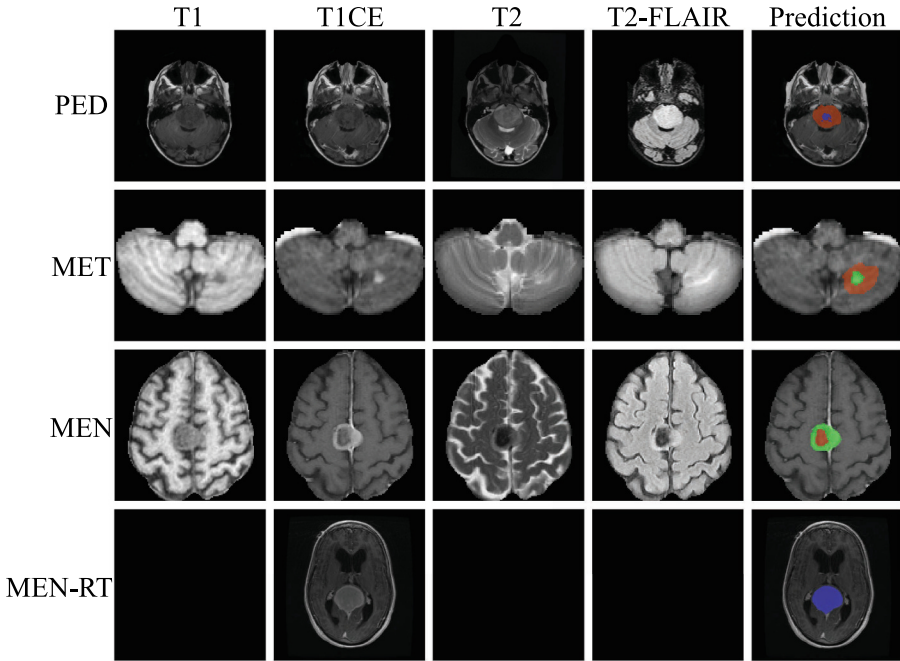
Our experiments confirm that carefully engineered pre- and post-processing can still deliver tangible gains even when base architectures have plateaued. Notably, some gains in Dice or NSD were limited on validation set but became evident under 5-fold CV, underscoring the danger of optimizing solely on a limited and sometimes skewed validation set.

**Table 2. PED quantitative results** Lesion-wise (LW) Dice coefficients and Normalized Surface Distance (NSD) at threshold 1mm were computed for enhancing tumor (ET), tumor core (TC), whole tumor (WT), non-enhancing tumor (NET), cystic components(CC), and edema (ED), respectively.

Task	Model	LW Dice					LW NSD thresh. 1mm						
		CC	ED	ET	NET	TC	WT	CC	ED	ET	NET	TC	WT
<b>PED</b>	MedNeXt(M)	0.735	0.857	0.694	0.899	0.926	0.927	0.749	0.857	0.748	0.878	0.903	0.905
	nmU-Net	0.713	0.813	0.630	0.900	0.928	0.928	0.762	0.813	0.693	0.885	0.914	0.915
<b>Validation</b>	nmU-Net-ResEnc(M)	0.719	0.857	0.692	0.902	0.931	0.932	0.735	0.857	0.755	0.883	0.910	0.913
<b>N=91</b>	Ensemble	0.704	0.868	0.677	0.906	0.933	0.933	0.736	0.868	0.731	0.891	0.916	0.917
	Post-processing	0.72	0.967	0.65	0.907	0.933	0.933	0.72	0.967	0.696	0.892	0.914	0.914
<b>PED</b>	Mean	0.669	0.892	0.672	0.851	0.908	0.915	0.67	0.892	0.714	0.815	0.82	0.83
<b>Testing</b>	Standard Deviation	0.468	0.304	0.361	0.201	0.145	0.14	0.466	0.304	0.371	0.206	0.228	0.213

**Table 3. MET, MEN, MEN-RT quantitative results.** Lesion-wise (LW) or global (G) Dice coefficients and Normalized Surface Distance (NSD) at 1mm-threshold were computed for enhancing tumor (ET), tumor core (TC), whole tumor (WT), resection cavity (RC), and gross tumor volume (GTV).

Task	Model	Dice					NSD thresh. 1mm				
		ET	TC	WT	RC	GTV	ET	TC	WT	RC	GTV
<b>MET</b>	MedNeXt(M)	0.747	0.768	0.776	0.628		0.588	0.595	0.496	0.591	
	nmU-Net	0.731	0.752	0.754	0.656		0.559	0.566	0.468	0.617	
<b>Validation (G)</b>	nmU-Net-ResEnc(M)	0.744	0.761	0.766	0.693		0.570	0.577	0.473	0.654	
<b>N=179</b>	Ensemble	0.744	0.765	0.769	0.705		0.579	0.587	0.489	0.668	
	Post-processing	0.699	0.72	0.719	0.907		0.535	0.544	0.442	0.871	
<b>MET</b>	Mean	0.544	0.555	0.561	0.86		0.606	0.607	0.589	0.858	
<b>Testing (LW)</b>	Standard Deviation	0.317	0.322	0.322	0.322		0.336	0.338	0.32	0.321	
<b>MEN</b>	MedNeXt(M)	0.841	0.863	0.839			0.853	0.87	0.842		
	nmU-Net	0.811	0.820	0.834			0.824	0.823	0.832		
<b>Validation (LW)</b>	nmU-Net-ResEnc(M)	0.807	0.822	0.825			0.818	0.825	0.826		
<b>N=141</b>	Ensemble	0.835	0.844	0.836			0.847	0.850	0.839		
	Post-processing	0.856	0.855	0.848			0.873	0.865	0.854		
<b>MEN</b>	Mean	0.881	0.878	0.868			0.885	0.878	0.86		
<b>Testing (LW)</b>	Standard Deviation	0.224	0.226	0.224			0.218	0.221	0.216		
<b>MEN-RT</b>	MedNeXt(M)					0.804				0.671	
	nmU-Net					0.764				0.618	
<b>Validation (LW)</b>	nmU-Net-ResEnc(M)					0.795				0.651	
<b>N=70</b>	Ensemble					0.796				0.654	
	Post-processing					0.796				0.654	
<b>MEN-RT</b>	Mean					0.807				0.683	
<b>Testing (LW)</b>	Standard Deviation					0.201				0.236	



**Fig. 5.** Qualitative results showing median lesion-wise Dice (L) or global Dice (G) of the whole tumor. PED: 0.979L, orange=NET, blue=ET; MET: 0.868G, orange=SNFH, green=ET; MEN: 0.961L, orange=SNFH, green=ET; MEN-RT: 0.887L, blue=GTV. (Color figure online)

Because the training set is much larger than the validation set, the latter can present a skewed tumor-type distribution and thus provide an unreliable signal for hyper-parameter and threshold tuning. We therefore recommend selecting models and thresholds by CV on the training set, using validation scores only as a sanity check. In the same spirit, the internal rank metric—aggregating Dice and NSD across tumor regions, proved more robust than any single metric, especially when candidate models were numerous. However, its sensitivity to the number of candidates (rank scores can be very close if only three candidates) suggests that alternative weighting schemes merit exploration.

We previously tested that stratified fold splitting based on radiomic clusters yielded higher CV scores than random splits, indicating that heterogeneity-aware sampling reduces overfitting. Additional ablations on multiple BraTS tasks are planned to quantify this effect. We also found that STAPLE fusion benefited the MEN-RT task when each fold was treated as a separate candidate, indicating that larger model pools can regularize ensembles.

Limitations include our reliance on simple volume thresholds for PP1 and PP2. Future work will replace these heuristics with radiomic-driven criteria anal-

ogous to those used for stratification, further tailoring post-processing to both tumor morphology and appearance.

Open-sourcing our ranker and keeping all components model-independent, we aim to make the pipeline easy to deploy in clinical workflows. When its utility is evaluated in more clinical studies, the pipeline would have the potential for larger clinical impact beyond BraTS benchmarks.

Finally, to facilitate reproducibility and extend the utility of our approaches, we have made the complete pipeline publicly available as easy-to-use Docker containers and a webapp for all the tasks. The Docker images are hosted at: <https://hub.docker.com/r/aparida12/brats2025> and the webapp is accessible at: <https://segmenter.hope4kids.io/>.

## 5 Conclusion

This work introduces a tumor- and lesion-aware ensemble pipeline that consistently segments four distinct brain tumour cohorts in the BraTS 2025 Lighthouse Challenge. By selecting and weighting models according to lesion-level accuracy and adding subtype-specific processing, the framework achieves strong segmentation performance relying on state-of-the-art network backbones. Its modular design makes it easy to adopt and extend, while the resulting quantitative tumor measurements could support clinical diagnosis, treatment planning, and longitudinal monitoring.

**Acknowledgements.** 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. Baid, U., Ghodasara, S., Mohan, S., et al.: The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint: [arXiv:2107.02314](https://arxiv.org/abs/2107.02314) (2021)
2. Bakas, S., Akbari, H., Sotiras, A., et al.: Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data* 4(1), 170117 (2017). <https://doi.org/10.1038/sdata.2017.117>
3. Bakas, S., Akbari, H., Sotiras, A., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection. *Cancer Imaging Arch.* (2017). <https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q>
4. Bakas, S., Akbari, H., Sotiras, A., et al.: Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *Cancer Imaging Arch.* (2017). <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>

5. 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 (2023)
6. Fathi Kazerooni, A., Khalili, N., Liu, X., Gandhi, D., Jiang, Z., et al.: The brain tumor segmentation in pediatrics (BraTS-PEDs) challenge: focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs) (2024). <https://arxiv.org/abs/2404.15009>
7. Hatamizadeh, A., Nath, V., Tang, Y., et al.: Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: Crimi, A., Bakas, S. (eds.) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pp. 272–284. Springer, Cham (2022)
8. 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)
9. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., et al.: nnU-Net revisited: a call for rigorous validation in 3d medical image segmentation. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, pp. 488–498. Springer Nature Switzerland, Cham (2024)
10. 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)
11. Jiang, Z., et al.: 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)
12. 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)
13. Karargyris, A., Umeton, R., Sheller, M., et al.: Federated benchmarking of medical artificial intelligence with MedPerf. *Nat. Mach. Intell.* **5**, 799–810 (2023)
14. Kazerooni, A.F., Khalili, N., Liu, X., et al.: The brain tumor segmentation (BraTS) challenge 2023: focus on pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs) (2023)
15. LaBella, D., Adewole, M., Alonso-Basanta, M., et al.: The ASNR-MICCAI brain tumor segmentation (BraTS) challenge 2023: intracranial meningioma (2023)
16. LaBella, D., Baid, U., Khanna, O., McBurney-Lin, S., McLean, R., et al.: Analysis of the BraTS 2023 intracranial meningioma segmentation challenge. *J. Mach. Learn. Biomed. Imaging* **3**(March 2025), 38–58 (2025)
17. LaBella, D., Schumacher, K., Mix, M., Leu, K., McBurney-Lin, S., et al.: Brain tumor segmentation (BraTS) challenge 2024: meningioma radiotherapy planning automated segmentation (2024). <https://arxiv.org/abs/2405.18383>
18. Ma, J., Li, F., Wang, B.: U-Mamba: enhancing long-range dependency for biomedical image segmentation. arXiv preprint: [arXiv:2401.04722](https://arxiv.org/abs/2401.04722) (2024)
19. 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 (2023)
20. Menze, B.H., Jakab, A., Bauer, S., 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>

21. Moawad, A.W., Janas, A., Baid, U., et al.: The brain tumor segmentation (BraTS-METS) challenge 2023: brain metastasis segmentation on pre-treatment MRI (2023)
22. 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)
23. 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 (2023)
24. Siegel, R.L., Kratzer, T.B., Giaquinto, A.N., Sung, H., Jemal, A.: Cancer statistics, 2025. *CA: Cancer J. Clin.* **75**(1), 10–45 (2025). <https://doi.org/10.3322/caac.21871>
25. Tang, Y., Yang, D., Li, W., Roth, H.R., et al.: Self-supervised pre-training of Swin transformers for 3D medical image analysis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20730–20740 (2022)
26. Van Griethuysen, J.J., et al.: Computational radiomics system to decode the radiographic phenotype. *Can. Res.* **77**(21), e104–e107 (2017)
27. Warfield, S., Zou, K., Wells, W.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Trans. Med. Imaging* **23**(7), 903–921 (2004). <https://doi.org/10.1109/TMI.2004.828354>
28. Xing, Z., et al.: SegMamba-V2: long-range sequential modeling mamba for general 3D medical image segmentation. *IEEE Trans. Med. Imaging*, 1–1 (2025). <https://doi.org/10.1109/TMI.2025.3589797>