Regions7525_DA5: Harnessing Adaptations of Existing nnU-Net Features to Improve Generalizability of Brain Tumor Segmentation

Sherman Chan[1*], Nihal Patil[1*], Hari Girish[1], and Simeng Zhu[1,2]

[1] The Ohio State University, Columbus, OH, USA
[2] The Ohio State University Comprehensive Cancer Center – Arthur G. James Cancer Hospital and Richard J. Solove Research Institute, Columbus, OH, USA
[*]These authors contributed equally to this work.
simengzhu@gmail.com

Abstract. Brain tumors involve the formation of masses within brain tissue and are associated with poor survival outcomes, emphasizing the need for a more standardized approach to early detection and patient-oriented treatment. BraTS-GoAT 2026 challenge aims to develop a model with the ability to accurately and efficiently generalize brain tumor segmentation of MRI images across different patient demographics and brain tumor classifications. Regions7525_DA5 was developed as a model that leveraged existing features and structures within the nnU-Net repository. This model incorporates region-based training, allowing the model to directly learn the regions of interest for this task and utilizes the existing nnUNetTrainerDA5 variant for more robust data augmentation. Additionally, Regions7525_DA5 also includes a 0.75 Binary Cross-Entropy to 0.25 Dice loss function, prioritizing voxel-by-voxel classification accuracy over region overlap and class imbalances. With Regions7525_DA5, there were improvements in both Global Dice Similarity Coefficient and Global Normalized Surface Distance metrics relative to nnU-Net baseline. Leveraging adaptations of existing nnU-Net structures, Regions7525_DA5 shows promise looking ahead towards a more standardized and effective approach to generalizing brain tumor segmentation tasks. The codes are available at https://github.com/nukanuka/buckeye-brats-goat-2026-region7525-da5/tree/main

Keywords: Deep Learning · nnU-Net · Brain Tumor Segmentation · BraTS-GoAT.

1 Introduction

Brain tumors are masses that form in brain tissue as a result of dysregulation of cell growth and proliferation [1]. Brain tumors are characteristically known to be heterogeneous, with distinctive tumor biology among different types of tumors [2,3,4,5]. Brain tumors can be classified by the number of lesions, the size of the lesions, as well as their anatomical location within the brain [6]. These brain

tumors are also distinct on a patient-by-patient basis, with differences in cancer biology also evident among different patient demographics groups, such as age and sex [3]. Despite recent advances in our understanding of the biology of these distinct classifications of brain tumors, challenges still remain in current therapeutic strategies when taking into account variations in treatment plans based on differences in tumor biology within these tumor types, as well as tumor heterogeneity between patient populations.

Due to the proliferative nature of brain tumors and the modern day challenges therapeutic strategies face when targeting these different brain tumor types and patient population demographics, there is an emphasis on the need to accurately and efficiently diagnose these brain tumors early. The early diagnosis of these brain tumors can prove to be critical in increasing our ability to implement effective patient-oriented treatment strategies in the clinical setting and ultimately improve the odds of patient survival [7]. Magnetic resonance imaging (MRI) has changed the landscape when it comes to soft tissue brain imaging and has proven to be significantly useful for early onset brain tumor diagnosis [5,7,8]. MRI images have also been shown to be a valuable tool in patient care, and the assessment of these images is taken into account in the development of a patient-specific treatment plan and serves as an effective metric to assess the patients response to treatment [5,8].

Although MRI has become invaluable in its use as a tool for the early detection of brain tumors and evaluation of the efficacy of the treatment plan in clinical care settings, MRI analysis is often a time-consuming process [2,9]. It is also susceptible to variability between different institutional MRI protocols, as well as in analysis from physician to physician [2,3,9,10]. In Brain Tumor Segmentation 2026 Challenge Generalizability Across Tumors (BraTS-GoAT) task, the aim is to improve upon the performance of existing brain tumor segmentation models by not only accounting for the diversity of different brain tumor cases, but also offering a more standardized and efficient approach to assessing MRI images. In doing so, MRI analysis with AI learning models can be leveraged to develop an appropriate corresponding treatment plan specific to the patients needs.

Convolutional Neural Networks (CNNs) have shown significant success in medical imaging segmentation tasks in recent years [9]. One such model employing CNNs, nnU-Net, has shown remarkable performance when it comes to brain segmentation tasks [11], making this model a perfect candidate for training a baseline model. In addition, nnU-Net has been employed and improved in numerous previous successful BraTS challenge tasks [3,5,12]. Our model seeks to leverage existing structures within the nnU-Net repository and enhance its brain tumor segmentation performance capabilities through fine modifications to its existing framework.

2 Methods

2.1 BraTS-GoAT 2026 Data Set

The BraTS-GoAT challenge highlights the task of developing a model that can generalize brain tumor segmentation across different tumor classifications and patient demographics. This challenge features a data set that includes cases of various brain tumor classifications such as gliomas, meningiomas, and metastases, as well as cases of pediatric brain tumors and adult populations of Sub-Saharan Africa glioma [12]. Each case in the data set includes four clinically-obtained MRI modalities (T1-weighted, T1-weighted with contrast, T2-weighted, and T2-FLAIR) originating from multiple institutions.

The segmentation labels for each of the different components of the brain tumor for the BraTS-GoAT challenge are as follows:

1: Necrosis (NCR)
2: Edema (ED)
3: Enhancing Tumor (ET)
0: Background and Non-Tumor Areas

The relevant sub-regions for this task are also labeled as follows:

(1 + 3): Tumor Core (TC = NCR + ET)
(1 + 2 + 3): Whole Tumor (WT = NCR + ET + ED)

2.2 Hardware Implementation

All work was done using resources allocated from the Ohio Supercomputer Center (OSC) server. The experiments were implemented using a PyTorch framework on an NVIDIA A100 GPU (40 GB). All aspects of the experimental pipeline involve the submission of sbatch scripts to the OSC queue. These scripts remain pending in the queue until the necessary resources needed are allocated by the OSC server, at which point the scripts would run.

2.3 nnU-Net Baseline Model

The data set was first divided into 1,080 training cases and 271 validation cases. To establish a baseline model, nnU-Net's PlainConvUNet framework was utilized with the standard 3D full-resolution configuration and trained with 5-fold cross-validation over 1,000 epochs. The network architecture consists of six encoder-decoder stages with two convolutions per stage as well as enabled deep supervision. The default nnU-Net framework has an initial learning rate of 0.01 with the learning rate decayed throughout training through a poly learning decay scheduler. The framework has four input channels corresponding to the four MRI modalities and four output channels corresponding to the individual component labels used for brain segmentation: NCR. ED, ET, and background.

2.4 Regions5050 Model

Region-based Training. Region-based training, a feature existing as part of the nnU-Net repository, was employed to enhance segmentation of the required labels of the BraTS-GoAT task: ET, TC, and WT. Region-based training allows for the model to learn not the individual component labels but rather directly from the three regions of interest for the BraTS-GoAT task (ET, TC, and WT). The NCR label and ET labels were able to be merged to obtain the TC label, and the NCR, ET, and ED labels were merged together to obtain the WT label. Region-based training leverages the ability to combine individual tumor component labels into the desired tumor subregion labels in order to optimize direct segmentation of the clinically evaluated regions of interest.

Re-establishing Baseline with Regions5050. To establish a new baseline model, the nnU-Net framework with region-based training enabled was utilized with 3D full-resolution configuration and re-trained with 5-fold cross validation over 1,000 epochs. Due to the use of region-based training in the Regions5050 model, the architecture instead consists of four output channels corresponding to the three brain segmentation regions of interest in this task (ET, TC, and WT) and background. All other conditions of the default nnU-Net framework were maintained otherwise.

2.5 Regions7525 Model

Loss Function. The nnU-Net trainer overrides the existing loss function of the nnU-Net baseline framework. The trainer maintains the loss function with the combination of Binary Cross-Entropy loss and Dice loss from the Regions5050 model, but the weights were adjusted so that the Binary Cross-Entropy and Dice loss were weighted 0.75 and 0.25, respectively. The adjustment in the weights of the loss function was made in order to prioritize voxel-by-voxel classification accuracy over region overlap and class-imbalance robustness with the goal of improving overall tumor segmentation accuracy.

$$L_{total} = 0.25L_{Dice} + 0.75L_{BCE} \tag{1}$$

Re-establishing Baseline with Regions7525. To establish a new baseline model, the nnU-Net framework was utilized with 3D full-resolution configuration and re-trained with 5-fold cross validation over 1,000 epochs. This new baseline model maintains the region-based training from the Regions5050 model while integrating in the adjusted 0.75 Binary Cross-Entropy to 0.25 Dice loss function. All other conditions of the default nnU-Net framework were maintained otherwise.

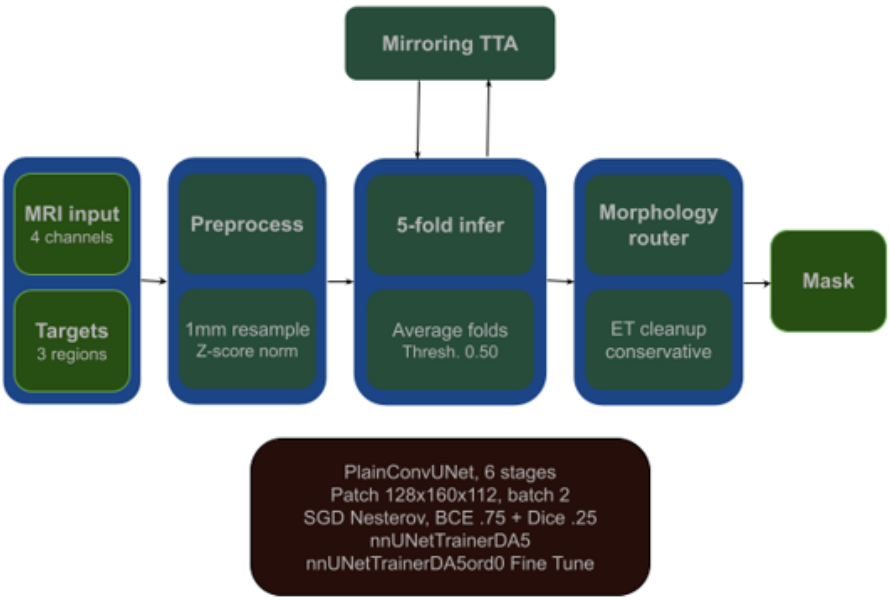

Fig. 1. Regions7525_DA5 Brain Tumor Segmentation model pipeline. This model incorporates existing features within the nnU-Net repository such as the use of region-based training as well as the nnUNetTrainerDA5 trainer variant. The model also overrides the default loss function with a 0.75 Binary Cross-Entropy to 0.25 Dice loss function.

2.6 Regions7525_DA5 Model

nnUNetTrainerDA5. nnUNetTrainerDA5 is a variant trainer within the nnU-Net repository that incorporates a more robust data augmentation compared to the default nnU-Net framework [11]. This variant trainer includes an increased frequency of rotation and scaling transforms, as well as features not included in the default trainer such as local brightness, local gamma, and sharpening that induce varying image intensities to help the model distinguish artifacts. The trainer variant also includes the BlankRectangle transform that reduces overfitting of the model. These features will help the model adapt to variability in MRI images during training, which should, in turn, aid in the model's ability to more accurately generalize brain tumor segmentation tasks.

Re-establishing Baseline with nnUNetTrainerDA5. All of the existing modifications stemming from the Regions7525 baseline model, including region-based training as well as 0.75 Binary Cross-Entropy to 0.25 Dice loss function, were maintained in this model. A new baseline model was trained using the nnUNetTrainerDA5 variant from the nnU-Net repository (see Figure 1). This

new baseline was trained with the 3D full-resolution configuration with 5-fold cross validation over 1,000 epochs.

2.7 Post-Processing

Morphology Router. The router examines the number of WT components, the size of the WT components, as well as the location of the largest WT component to assign each case to one of three cleanup modes, from which the router removes WT components smaller than a fixed-size threshold. For cases of single or surface-attached lesions, where predictions are more prone to contain isolated artifacts, the router uses a 25-voxel threshold. However, for cases with multifocal tumors, which can contain positive small lesions, the router uses a less aggressive 10-voxel threshold.

Table 1. A summary of each model stage, including nnU-Net baseline, Regions5050, Regions7525, and Regions7525_DA5, and what modifications were made between each model stage.

Model	Modifications with Each Model Stage
nnU-Net Baseline	Usage of the default nnU-Net framework. Included only the conservative ET cleanup as a post-processing feature.
Regions5050	Incorporation of region-based training. The loss function features 0.5 Binary Cross-Entropy (BCE) to 0.5 Dice loss function. The morphology router was also added as a post-processing feature.
Regions7525	Integration of the 0.75 Binary Cross-Entropy (BCE) to 0.25 Dice loss function. This was made in order to emphasize voxel-by-voxel classification accuracy over region overlap.
Regions7525_DA5	Inclusion of the nnUNetTrainerDA5 trainer variant as a part of training to provide a more robust data augmentation.

The 25-voxel threshold for single and surface attached lesion cases was chosen in order to still maintain small lesions that are tumor positive while eliminating those that are more likely to be just artifact, but a lower 10-voxel threshold was chosen specifically for multi-lesion cases due to these cases being prone to very small positive lesions and the need for increased sensitivity to maintain these lesions in the final prediction. This post-processing feature was only integrated starting with the region-based models.

Conservative ET cleanup. The conservative ET cleanup first includes hierarchy enforcement, which ensures that the ET label is fully within the TC label and that the TC label is fully within the WT label. Any ET components that were smaller than 15-voxels were removed, as they were more likely to be artifacts than actual ET. Additionally, the ET/TC overlap requirement was set to

be 0.25, and connected components use 26-connectivity. This component of the post-processing involves more ET-oriented cleanup of the model's predictions and helps to further refine the accuracy of the ET segmentation label.

3 Results

3.1 Training Progression

Training loss and validation loss were monitored over the course of 1,000 epoch training for the Regions7525_DA5 model. The training and validation loss rapidly decreased early on in training ,and both eventually stabilized at around -0.21 to -0.22, indicative of convergence and the model reaching its optimal point of performance. A similar training progression was observed across each of the different model stages as well.

3.2 Quantitative Performance

Performance evaluation of the different models was done using the 271 cases from the internal cross-validation fold data set as well as the 451 cases as part of the BraTS-GoAT challenge validation data set. The performance of each model was evaluated using the Global Dice Similarity Coefficient (DSC), a metric that measures how well the models prediction of the tumor component label overlaps with the ground truth. Global Normalized Surface Distance (NSD) was a second metric used to determine how well the edges of each tumor component label align with the ground truth.

Table 2. Global Dice Similarity Coefficient (DSC) and Global Normalized Surface Distance (NSD) evaluation scores of the different models using the internal cross-validation fold data set. Models evaluated using this validation set include nnU-Net baseline, Regions5050, Regions7525.

Model	Global Dice Similarity Coefficient (DSC)			Global Normalized Surface Distance (NSD)		
	ET	TC	WT	ET	TC	WT
nnU-Net Baseline	0.868	0.909	0.925	0.889	0.864	0.856
Regions5050	0.869	0.912	0.926	0.890	0.867	0.855
Regions7525	0.873	0.915	0.927	0.894	0.867	0.855

The scoring metrics from the DSC and NSD scores for each model stage using the internal cross-validation fold data set are shown (see Table 2). The nnU-Net baseline model demonstrates strong performance with brain tumor segmentation tasks, which was only modestly improved with the incorporation of region-based training in the Regions5050 model. When comparing the performance of the

Regions5050 model to that of the Regions7525 model, there were further slight improvements for the evaluation metrics across each segmentation label, with increases in DSC scores (+0.004, +0.003, +0.001) for the ET, TC, and WT labels, respectively, and increase in the NSD score (+0.004) for the ET label. This correlates with a stronger ability from the model to perform brain tumor segmentation tasks with prioritization of voxel-by-voxel classification accuracy over region overlap.

Table 3. Global Dice Similarity Coefficient (DSC) and Global Normalized Surface Distance (NSD) evaluation scores of the official BraTS-GoAT 2026 validation data set. Models evaluated using this validation data set include nnU-Net baseline, Regions7525, and Regions7525_DA5.

Model	Global Dice Similarity Coefficient (DSC)			Global Normalized Surface Distance (NSD)		
	ET	TC	WT	ET	TC	WT
nnU-Net Baseline	0.777	0.816	0.875	0.544	0.496	0.479
Regions7525	0.803	0.822	0.885	0.564	0.505	0.491
Regions7225_DA5	0.813	0.835	0.891	0.567	0.506	0.491

The scoring metrics from the DSC and NSD scores for each of the model stages using the official BraTS-GoAT 2026 validation data set are shown (see Table 3). The nnU-Net baseline model performance based on the evaluation metrics similarly shows promising performance, but notably shows the weakest performance when it comes to the ET label. The model does slightly better with the TC label, but its best segmentation performance is noted with the WT label. This trend continues to be observed across the other model stages. Compared to nnU-Net baseline, Regions7525 improves the evaluation metrics across all three labels, with increases in DSC scores (+0.026, +0.006, and +0.010) and NSD scores (+0.020, +0.009, and +0.012) for the ET, TC, and WT labels, respectively. This suggests an improvement in the model's performance as a result of the changes to a 0.75 Binary Cross-Entropy loss to 0.25 Dice loss function as well as the incorporation of region-based training. The development of Regions7525_DA5 further improved upon these evaluation metrics, with increases in DSC scores (+0.010, +0.013, and +0.006) and NSD scores (+0.003, +0.001, and +0.000) for the ET, TC, and WT labels, respectively, relative to Regions7525. It is notable that most of the improvements driven by the incorporation of the nnUNetTrainerDA5 variant was to the DSC scores, with minimal changes to the NSD scores. The improvement of the evaluation metrics across the different model stages is an effect that was conserved with the use of the official BraTS-GoAT 2026 challenge validation data set in addition to the use of the internal cross-validation fold data set.

3.3 Qualitative Performance

Performance of the Regions7525_DA5 model was also assessed through visualization of the prediction labels of the different cases within the BraTS-GoAT challenge validation data set. The four MRI modalities T1-weighted, T1-weighted with contrast, T2-weighted, and T2-FLAIR are each individually leveraged to offer insight into particular segmentation labels. For example, T1-weighted with contrast is best used to observe enhancing tumor and necrosis regions of brain tumors while T2-FLAIR is best used to observe edema regions of the tumor [13]. Together, these four MRI modalities offer a picture of the whole tumor and its individual components. Representative images for the T1-weighted with contrast and T2-FLAIR as well as the nnU-Net baseline, Regions7525, and Regions7525_DA5 prediction mask are shown for each distinct brain tumor classification: gliomas, meningiomas, and brain metastases (see Figure 2).

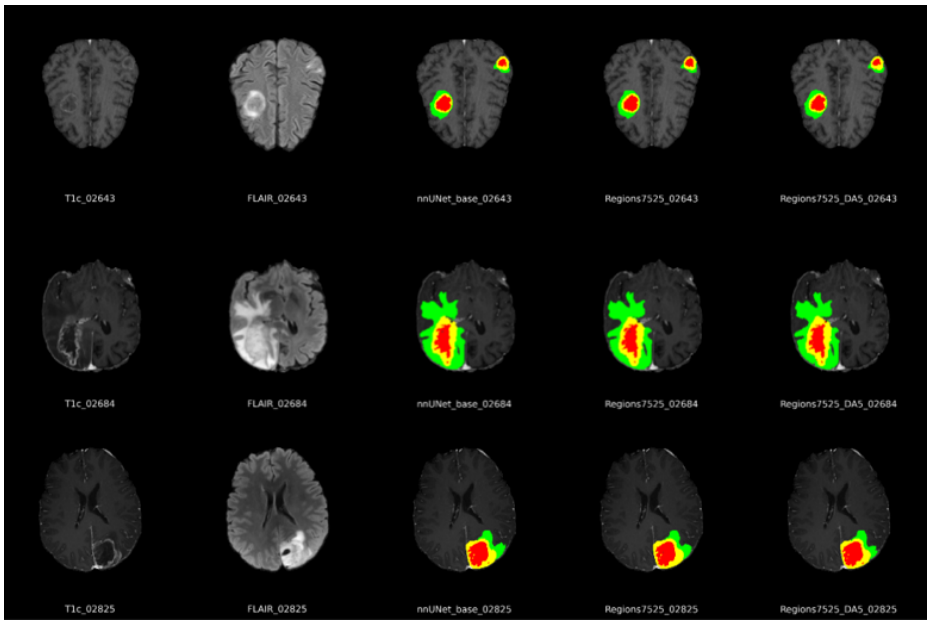

Fig. 2. T1-weighted with contrast and T2-FLAIR, as well as nnU-Net baseline, Regions7525, and Regions7525_DA5 prediction labels for three distinct cases within the BraTS-GoAT challenge validation data set. These cases are representative images of each distinct brain tumor classification: gliomas, meningiomas, and brain metastases. Red indicates necrotic, yellow indicates enhancing tumor, and green indicates edema.

4 Discussion

The ability to leverage AI learning models to generalize brain tumor segmentation across distinct classifications of brain tumors, as well as to the heterogeneity of brain tumors across different patient demographics, has become increasingly important. Different brain tumor classifications, from gliomas, to meningiomas, to brain metastases, require quick and accurate analysis that aligns with the appropriate treatment. The ability to effectively standardize brain tumor segmentation using magnetic resonance imaging is not just imperative for early detection and diagnosis of brain tumors, but also useful as a measure to design and evaluate a patient-oriented treatment plan.

Regions7525_DA5 demonstrate how existing structures within the nnU-Net repository can be leveraged and modified to improve brain segmentation performance. Incorporating region-based training allowed the model to directly learn the brain tumor segmentation labels of interest and when combined with the adjusted 0.75 BCE to 0.25 Dice loss function helped to improve the model's aim to improve generalization of brain tumor segmentation tasks as demonstrated through improved DSC and NSD evaluation metrics relative to nnU-Net baseline. This effect was further improved on with the integration of the nnUNetTrainerDA5 trainer variant into the training pipeline, providing a more robust data augmentation that leveraged feature to enable improved generalization of these brain tumor segmentation tasks.

While this model has demonstrated improvements in brain segmentation tasks relative to the default nnU-Net framework, challenges still remain, such as in the model's fine ability to delineate boundaries of the different region labels. Future work will be directed towards leveraging other existing U-Net based models with similar adaptions made to the nnU-Net framework with the aim of improving upon the current model's performance of generalizing brain tumor segmentation tasks. In addition to leveraging these U-Net frameworks with the modifications made to nnU-Net, future studies can be made into how the models perform with each of the different brain tumor classifications and demographics, not only to address deficits in our model's predictions of certain brain tumor types, but also to specifically fine-tune the model to increase segmentation performance of these particular brain tumor types.

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

References

1. Dirks, P.B., Rutka, J.T. (1997). Current concepts in neuro-oncology: The cell cyclea review. *Neurosurgery* 40(5), 10001015. https://doi.org/10.1097/00006123-199705000-00025
2. Bendazzoli, S., Moreno, R. (2026). BraTS-FL: Enhancing Generalization in Brain Tumor Segmentation via Federated Learning. In: Bakas, S., *et al.* Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries.

MICCAI 2025. Lecture Notes in Computer Science, vol 16376. Springer, Cham. https://doi.org/10.1007/978-3-032-16365-3_43
3. Chen, MY., Liang, HK.T. (2026). Scaling High-Capacity ResUNet with Dynamic Batch for Universal Brain Tumor Segmentation. In: Bakas, S., *et al.* Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries. MICCAI 2025. Lecture Notes in Computer Science, vol 16376. Springer, Cham. https://doi.org/10.1007/978-3-032-16365-3_44
4. Garg, N., Dewangan, H.K. (2025). Emerging approaches to brain cancer treatment: A review of targeting modalities. *Current Pharmaceutical Design*. Advance online publication. https://doi.org/10.2174/0113816128385127250828160927
5. Wang, H., Zhang, Y., Sun, J., An, X., Zhu, L., Li, C. (2026). ADMFNet: Enhancing Cross-Tumor Generalization inăMulti-Modal MRI Segmentation. In: Bakas, S., *et al.* Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries. MICCAI 2025. Lecture Notes in Computer Science, vol 16376. Springer, Cham. https://doi.org/10.1007/978-3-032-16365-3_46
6. Menze, B.H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M.A., Arbel, T., Avants, B.B., Ayache, N., Buendia, P., Collins, D.L., Cordier, N., Corso, J.J., Van Leemput, K. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging* 34(10), 19932024. https://doi.org/10.1109/TMI.2014.2377694
7. Rastogi, D., Johri, P., Donelli, M., et al. (2025). Deep learning-integrated MRI brain tumor analysis: feature extraction, segmentation, and survival prediction using Replicator and volumetric networks. *Scientific Reports*, 15, 1437. https://doi.org/10.1038/s41598-024-84386-0
8. Martucci, M., Russo, R., Schimperna, F., D'Apolito, G., Panfili, M., Grimaldi, A., Perna, A., Ferranti, A.M., Varcasia, G., Giordano, C., Gaudino, S. (2023). Magnetic resonance imaging of primary adult brain tumors: State of the art and future perspectives. *Biomedicines* 11(2), 364. https://doi.org/10.3390/biomedicines11020364
9. Aktar, M., Nasser, T., Souza, R. (2026). A Multitask Learning Approach for Segmenting Brain Tumor Sub-regions: Towards Better Generalization. In: Bakas, S., *et al.* Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries. MICCAI 2025. Lecture Notes in Computer Science, vol 16376. Springer, Cham. https://doi.org/10.1007/978-3-032-16365-3_42
10. Kaifi, R. (2023). A review of recent advances in brain tumor diagnosis based on AI-based classification. *Diagnostics* 13(18), 3007. https://doi.org/10.3390/diagnostics13183007
11. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* 18(2), 203211. https://doi.org/10.1038/s41592-020-01008-z
12. Hsu, TL., Nguyen, D.K., Lin, P., Lin, CT., Wang, WC. (2026). Enhancing Brain Tumor Segmentation Generalizability via Pseudo-Labeling and Ratio-Adaptive Postprocessing. In: Bakas, S., *et al.* Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries. MICCAI 2025. Lecture Notes in Computer Science, vol 16376. Springer, Cham. https://doi.org/10.1007/978-3-032-16365-3_47
13. Chou, YS. et al. (2026). μPUA-Net: PowerMLP Model Size Shrinking Method with Accuracy Maintaining. In: Bakas, S., et al. Segmentation, Classification, and Synthesis for Brain Tumors and Traumatic Brain Injuries. MICCAI 2025. Lecture Notes in Computer Science, vol 16376. Springer, Cham. https://doi.org/10.1007/978-3-032-16365-3_4

14. Karargyris, R., Umeton, R., Sheller, M.J., Aristizabal, A., George, J., Wuest, A., Pati, S., et al. (2023). Federated benchmarking of medical artificial intelligence with MedPerf. *Nature Machine Intelligence* 5, 799810. https://doi.org/10.1038/s42256-023-00652-2