RACER: A Radiomic-Adaptive Cohort Ensemble with Refinement for Cross-cohort Brain Tumor Segmentation using Deep Learning

Piyush Kumar Prajapati[1], Mohammad Tufail Sheikh[1], Arpit Choudhary[1], Megh Parikh[1], Anup Singh[1,2,3]

[1] Centre for Biomedical Engineering, Indian Institute of Technology, Delhi, India

[2] Yardi School of Artificial Intelligence, Indian Institute of Technology, Delhi, India

[3] Department of Biomedical Engineering, All India Institute of Medical Sciences, New Delhi, India

anupsm@iitd.ac.in

Abstract. The generalizability of brain tumor segmentation models is an important clinical factor in the deployment of deep learning based models in hospital settings. Usually, the segmentation models are trained and evaluated on a single tumor type, yet clinical adoption requires robustness across different tumor data cohorts that differ in age, geography, imaging protocols, and scanner, etc. The BraTS Generalizability Across Tumors (GoAT) task evaluates this setting on images from multiparametric magnetic resonance imaging, where training data has glioma, meningioma and metastasis tumor types while the held-out validation dataset includes pediatric and sub-Saharan African cases. We proposed RACER (Radiomic-Adaptive Cohort Ensemble with Refinement), a deep learning based methodology, for this task. In the proposed methodology, three convolution neural network (CNN) based encoder were trained as backbones, two of which with residual encoder were conditioned per case through a radiomic feature-wise linear modulation layer, and a self-supervised masked-image-modelling initialization together with cohort-shift augmentation was used to encourage generalization. The five-fold softmax ensemble was refined by a dual-encoder network that corrects the ensemble prediction through a zero-initialized residual, and a cohort-safe relabeling rule was applied as post-processing. On the official validation synapse portal, the final model reached a mean Dice of 0.840 ± 0.190, a normalized surface Dice of 0.497 ± 0.209 and a mean 95th percentile Hausdorff distance of 21.3 ± 59.61 mm, and a sub-region mean Dice of 0.812 ± 0.251, 0.830 ± 0.236, 0.877 ± 0.174 for enhancing tumor, tumor core, whole tumor, respectively, thus improving over the ensemble on all three region-averaged metrics. (*Source code*: https://github.com/piyushkp27/RACER)

Keywords: Brain tumor segmentation · Cross-cohort generalization · Feature-wise linear modulation · Self-supervised learning · Ensemble refinement · Deep learning · MRI

1 Introduction

Brain is a critical organ in the human body and is a part of central nervous system (CNS). Tumors can originate directly inside the brain (Primary tumors) or spread from other organs (Secondary tumors). These tumors are initially diagnosed non-invasively by different medical imaging techniques. Mainly multiparametric magnetic resonance imaging (mp-MRI) is preferred due to its excellent soft-tissue contrast and shows better

visualization of tumor subregions. MRI has different multiparametric sequences which includes T1-weighted (T1W), T1-contrast (T1C), Flair Attenuated Inverse Recovery (FLAIR) and T2-weighted (T2W), etc. Each sequence highlights distinct information which helps in better understanding of the tumor physiology. These provide information on tumor location, extent and possible effects on the surrounding tumor environment. Thus, the localization of the tumor is an important task in the medical imaging domain, which corresponds to the segmentation of tumor image pixels.

Traditionally, conventional image processing techniques like thresholding and region-based techniques were used for the segmentation of tumors but these techniques are manual and underperform due to the complex intensity variations in tumor region. Thus, automated segmentation techniques based on learning-based methods (Deep Learning, Machine Learning) are used. These methods enable a model to learn the intrinsic mapping between the input and the labels. In tumor segmentation, these models learn the complex relationship between the MRI sequences and the corresponding labels (masks). In recent years, multiple deep learning models have been proposed for the automatic segmentation of brain tumors; however, their generalizability is still limited and efforts are being made to enhance the accuracy as well as generalizability of such models.

The Brain Tumor Segmentation (BraTS) challenge [1] was introduced to advance and accelerate the development of automated segmentation algorithms. It has different sub-challenges on variety of problems associated with tumor segmentation which include segmentation of different tumor types (Glioma, Meningioma, Metastasis, Pediatric brain tumors). One of the specific tasks in the challenge focuses on the generalizability of the developed algorithms. The objective of the BraTS Generalizability Across Tumors (BraTS-GoAT) sub-challenge, is to advance the capability of the models in segmenting the tumor across different lesion types, institutions and demographics with limited information and data on the target tumor type, thus simulating the clinical scenario in the hospital settings.

Automated segmentation of gliomas and other brain tumors on mp-MRI has advanced rapidly. Many algorithms and models such as U-Net[2], nnU-Net[3], SegResNet[4] and Swin UNETR[5] provide strong and reproducible baselines. Most reported results, however, are obtained within a single tumor cohort, so the reported accuracy does not necessarily transfer to populations that were not represented during model training. Therefore, the BraTS-GoAT task [6] has a training set of preoperative tumors which contain cases from glioma and a subset of meningioma and metastasis, while the withheld evaluation set additionally contains pediatric and sub-Saharan African (SSA) cases whose intensity statistics and lesion morphology differ from the training distribution. As the tumor subregion visual properties remain consistent across different tumor types., the model trained on a particular set of tumors should generalize well on the different set of tumors, creating a robust cross-cohort system.

Therefore, the objective of this study is to develop an automatic segmentation framework using deep learning based techniques which could generalize well on different types of brain tumors, The proposed approach takes three complementary stages. First, a per-scan conditioning mechanism at bottleneck layer was added to the residual-encoder (ResEnc) U-Net so that the network can adapt its internal features to the intensity

signature of each scan. Second, a self-supervised initialization and an aggressive cohort-shift augmentation policy were used to widen the distribution seen during training. Third, the ensemble output was refined by a small, learned network and by post-processing conservative relabeling rule on small enhancing tumor lesions. Therefore, the proposed pipeline is referred to as Radiomic-Adaptive Cohort Ensemble with Refinement (RACER). Fig. 1 gives an overview of the pipeline. The main contributions to the proposed framework are:

- A pseudo-label self-training approach with cohort motivated domain randomization.
- A radiomic conditioned feature-wise linear modulation layer was integrated into a ResEnc U-Net and combined with three-dimensional masked-image-modelling pretraining,
- A dual-encoder refiner with a zero-initialized residual head was designed to correct the ensemble predictions.

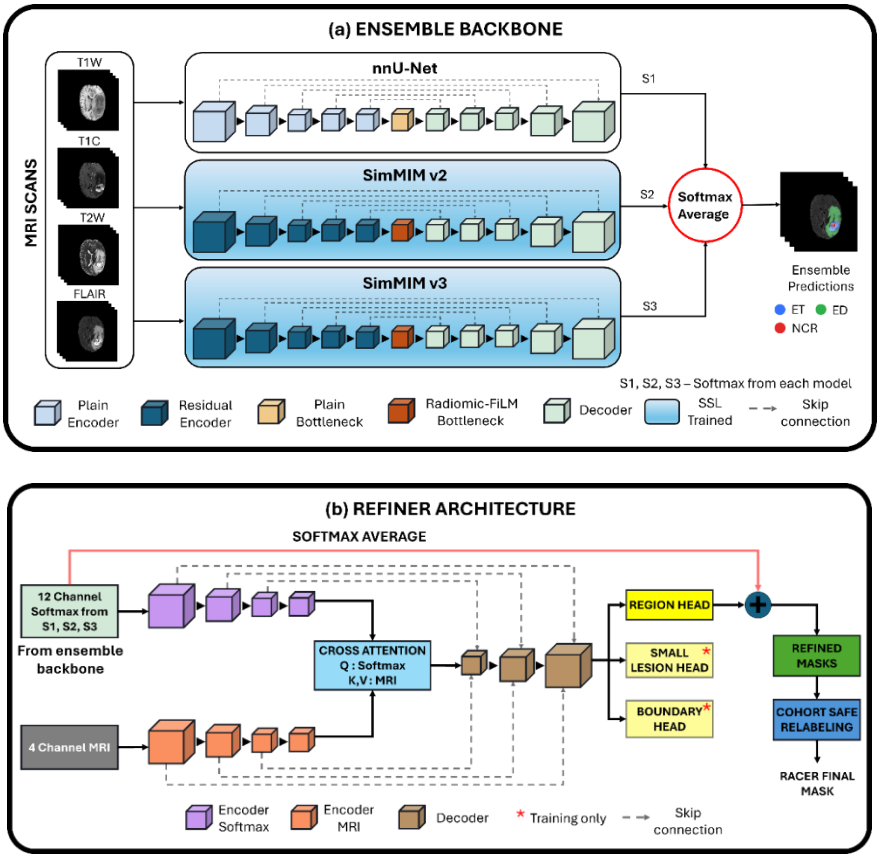

Fig. 1. Proposed Methodology of RACER (a) Ensemble backbone: The four MRI channels (T1W, T1C, T2W, FLAIR) are given to three models: a plain nnU-Net and two residual-encoder networks (SimMIM v2 and v3) that are pretrained by self-

supervised learning and use a radiomic-FiLM bottleneck. Their softmax outputs S1, S2 and S3 are averaged to form the ensemble prediction, (b) The 12-channel ensemble softmax (S1, S2, S3) and the 4-channel MRI are encoded in two branches. Cross-attention uses the softmax as the query and the MRI as key and value, and the decoder feeds three heads: region, small-lesion and boundary (the last two are used in training only). The region head is added back to the averaged softmax to give the refined masks, which pass through cohort-safe relabeling to produce the RACER final mask.

2 Materials and Methods

2.1 Data and Pre-processing

The BraTS-GoAT 2026 training set provides labelled 1351 cases of brain tumor data, each with four co-registered MRI sequences, namely T1-weighted (T1W), T1-contrast (T1C), Fluid Attenuated Inverse Recovery (FLAIR) and T2-weighted (T2W) and an expert annotation with three labels: necrotic tumor core (NEC), surrounding non-enhancing FLAIR hyperintense (SNFH), and enhancing tumor (ET) and 1138 unlabeled cases with same four MRI sequences. Fig. 2 shows representative cases from labelled dataset. The validation set provides 451 cases without annotations and is scored through the challenge portal. All images have an isotropic resolution of 1 mm on a 240x240x155 grid. Each volume was automatically cropped to include only foreground region containing maximum brain volume and each channel was normalized by z-score normalization using foreground pixels only. The normalization equation is as follows:

$$\widehat{x^c} = \frac{(x^c - \mu^c)}{\sigma^c} \qquad (1)$$

where, x^c is the c^{th} channel in our input with mean μ^c and standard deviation σ^c and $\widehat{x^c}$ is the normalized output computed over brain voxels in c^{th} channel.

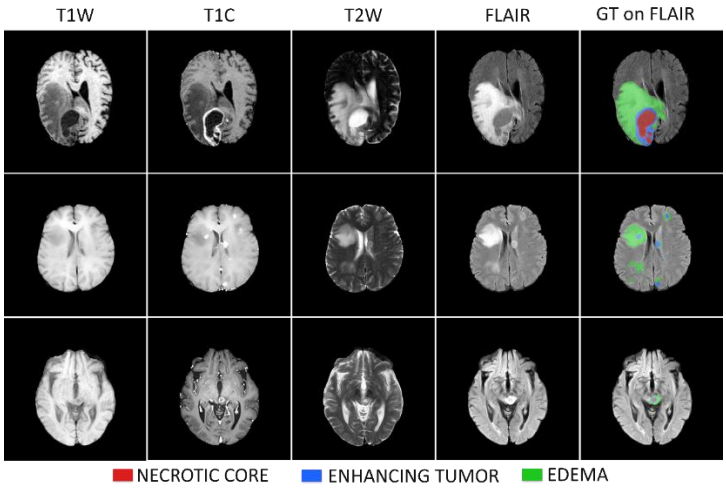

Fig. 2. Three training cases from data, shown as the four MRI channels with reference overlaid on FLAIR in the last column. Row 1, row 2 and row 3 show a large single mass, a multifocal case, and a small mainly enhancing lesion, respectively. Enhancing

tumor is red, necrotic core blue, and peritumoral signal abnormality green. NCR: Necrotic region; ED: Edema; ET: Enhancing tumor

2.2 Proposed Method: RACER

The proposed method uses nnU-Net as baseline architecture. For ablation, we have considered two approaches. One is on the labelled dataset only and the other on the labelled and unlabeled dataset. The approach on the labelled dataset (GT) has three backbones: nnU-Net, CohortAug v2, and CohortAug v3 models, while the other approach employs the labelled dataset and Pseudo Labels (PL) generated from each of above models from 739 unlabeled cases out of 1138 cases. This approach includes nnU-Net model, SimMIM v2 and SimMIM v3, each trained on GT+PL. The self-supervised learning (SSL) based approach model predictions were ensembled and refined using a refiner network for refining the missed or less accurate lesions. Both approaches use techniques which are described in the subsequent sections.

2.2.1 Cohort-Shift Augmentation

In addition to conventional augmentation techniques such as geometric transformations, three additional augmentation approaches for increasing the training distribution towards the unseen cohorts were layered on top of the default nnU-Net augmentation. The first was modality dropout, in which one of the four channels is occasionally dropped and replaced by a zero-value channel (p = 0.15). The second added strong per-channel gamma variation (from 0.5 – 2.0 with p = 0.3), additive Gaussian noise (p = 0.3, σ = [0 – 0.15]), and a low-frequency multiplicative bias field (p = 0.15, 4^3 control grid, ± 30%) to mimic missing sequences, contrast variations and scanner inhomogeneity (CohortAug v2). The third added enhancement dropout (p = 0.2), in which the contrast-enhanced channel is blended toward the native T1 channel so that the network does not rely on contrast brightness alone, together with a simulated motion artefact (p = 0.5) created by perturbing random lines in k-space (p = 0.15) and downsample-upsample low-resolution simulation (p = 0.15, scale = 0.55 – 0.85) (CohortAug v3). These operations act as inexpensive augmentations for the intensity and quality shifts expected on pediatric and SSA scans.

2.2.2 Self-Supervised Initialization (SimMIM-3D)

A three-dimensional Self-Supervised Learning (SSL) adaptation of masked image modelling (SimMIM) [7] was used to initialize the ResEnc with labelled and unlabeled dataset images before supervised training. The SSL module was trained on a cropped volume of 96x96x96. For each crop, the volume was divided into 3D patches of 16 voxels and subsequently a subset of the patches was masked to zero voxel values, and the encoder together with a light convolutional reconstruction head predicted the four original intensity channels (shown in Fig. 3). Only the masked voxels contributed to the loss,

$$L_{SSL} = \frac{\sum_i \left| \hat{x}_i - x_i \right| \times m_i}{\Sigma_i m_i}, \quad m_i \text{ is the block mask for masked patch} \qquad (2)$$

The SSL block was trained for 200 epochs over the labelled and unlabeled preprocessed images. The pretrained weights were loaded into the two ResEnc backbones and retrained over CohortAug v2 and v3 generated PL, the third backbone, plain U-Net model trained on the PL data without this initialization. During fine-tuning, FiLM layer was added as the bottleneck layer of ResEnc backbone with the encoder learning rate set to one-tenth of the decoder rate so that the pretrained features were preserved.

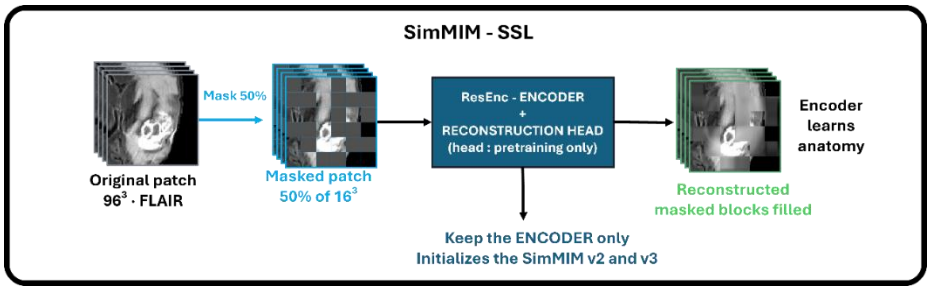

Fig. 3. Self-supervised pretraining with SimMIM. A 96x96x96 FLAIR patch is divided into 16x16x16 blocks and half of them are masked. The residual encoder with a reconstruction head learns to fill the masked blocks, so the encoder learns brain anatomy without labels. The head is used in pretraining only and the trained encoder initializes SimMIM v2 and v3.

2.2.3 Radiomic Feature-wise Linear Modulation

Two (SimMIM-based models) of the three backbones were conditioned per case by a radiomic feature-wise linear modulation layer inspired by FiLM [8]. Fig. 4, shows the radiomic based conditioning for FiLM in bottleneck layer of ResEnc. For each input patch, a 32-dimensional descriptor was computed on the fly: seven first-order statistics (mean, standard deviation, 10^{th}, 50^{th} and 90^{th} percentiles, skewness and kurtosis) for each of the four channels, together with three cross-channel contrasts ([T1C – T1W], [FLAIR – T1W], [T2W – T1W]) and the foreground fraction. A small multilayer perceptron (MLP) mapped this descriptor to a scale and a shift that modulates the 320-channel bottleneck feature x,

$$\begin{aligned} \gamma &= 1 + MLP(\gamma(r)), \\ \beta &= MLP(\beta(r)), \\ x' &= \gamma x + \beta, \end{aligned} \tag{3}$$

where γ and β are parameters learned during the training process, MLP is the multiple layer preceptron and x′ is the updated feature map.

The MLP input is 36-dimensional: the 32-value descriptor concatenated with four additional conditioning slots that are held at zero during both training and inference, so the conditioning is driven by the 32 radiomic features only. For reproducibility, as the patch (128x160x112) exceeds one million voxels, the 10^{th}, 50^{th} and 90^{th} percentiles are computed on a random one million voxel subsample that is seeded in the released code. Percentiles are computed after background voxels are replaced by the foreground mean, while the remaining first order statistics use the foreground mask only.

The final linear layer was initialized to zero so that the modulation begins as the identity and the conditioning is learned gradually. The supervised loss combined Dice and cross-entropy with a Hausdorff distance (HD) transform term [9] that penalizes errors in proportion to their distance from the boundary is

$$L = L_{Dice} + L_{CE} + 0.5 \times L_{HD} \tag{4}$$

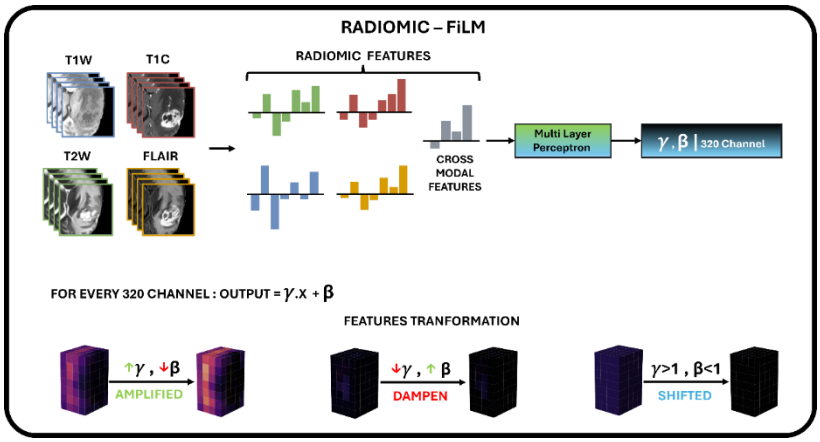

Fig. 4. Radiomic-FiLM: Radiomic features are computed from each MRI channel and joined into a cross-modal descriptor. A multi-layer perceptron maps this descriptor to a scale and a shift (γ, β) for all 320 bottleneck channels. Each channel is then changed as output = γx + β, which can amplify, dampen or shift the feature.

2.2.4 Ensemble Refiner

The three backbones were each trained with five-fold cross-validation, and the soft probability softmax maps of the five folds of each backbone were averaged. Each backbone predicts four output maps for each background, WT, Tumor core (TC) and ET labels, concatenated to a total of 12 channels for all three backbones. A refiner network was then trained to correct this prior. The refiner used two parallel encoders, one with a 12-channel prior and other with the four MRI channels, a cross-attention block at the bottleneck in which the prior queries the image features, and a shared decoder. The region head produced a residual on top of the ensemble logits,

$$\begin{aligned} region &= \text{logit}(mean\ softmax) + \Delta\ , \\ \Delta &= residual\ head\ output \end{aligned} \tag{5}$$

The residual head was initialized to zero, thus at the start of training the refiner reproduces the ensemble and only the correction is learned. This made training stable and prevented the refiner from ever performing worse than its input. The refiner network was trained on 50 epochs. Two auxiliary heads, one detecting small lesions and one regressing a boundary distance, were used during training as multi-task regularisers and were not included at inference. Fig. 1(b) shows the refiner network pipeline. For inference, model uses a sliding window of 128x128x128 with a stride of 64 and gaussian

weighting. The refiner was optimized with a combination of Dice-focal-Tversky [10,11], boundary, distance and small-lesion loss terms, which is given as:

$$L = L_{DiceFTv} + 0.5 \times L_{bd} + 0.3 \times L_{dist} + 0.4 \times L_{small} \tag{6}$$

2.3 Post-processing and Cohort-Safe Relabeling

The three region probabilities were thresholded at 0.5 each, the anatomical hierarchy that the WT contains TC, which contains ET was enforced, and small connected components were removed (ET components below 15 voxels and WT components below 50 voxels). A final cohort-safe rule, following the ratio-adaptive post-processing [12], relabeled the ET to necrotic core, termed as cohort-safe relabeling, when it was both a very small fraction of the whole tumor and small in absolute size,

$$\begin{aligned} &if\ \frac{|ET|}{|WT|} < 0.04\ \ and\ \ |ET| < 450\ voxels, \\ &then\ \ ET\ becomes\ NCR \end{aligned} \tag{7}$$

The absolute size limit of 450 voxels was introduced after inspection of the reference masks to limit the removal of original ET.

2.4 Evaluation Metrics

Segmentations were evaluated on the three standard regions: ET, TC and WT. The challenge evaluates the predictions on the Dice coefficient, the normalized surface Dice (NSD) and the 95th percentile HD, computed globally over the whole segmentation by the official BraTS-GoAT evaluation. For a predicted mask A and reference B,

$$Dice\ Coefficient = \frac{2\,|A \cap B|}{|A| + |B|} \tag{8}$$

$$HD_{95}(A,B) = \max\left(P_{95}\left[\sup_{a \in A}\left\{\inf_{b \in B} d(a,b)\right\}\right], P_{95}\left[\sup_{b \in B}\left\{\inf_{a \in A} d(b,a)\right\}\right]\right) \tag{9}$$

The NSD measures the fraction of the two surfaces that lie within a small tolerance of each other, and the 95th percentile HD measures boundary displacement while discarding the most extreme 5% of surface distances.

2.5 Training Details

All the training was done on a single Nvidia A100 GPU with 80GB VRAM. The optimizer used was stochastic gradient descent (SGD) with an initial learning rate of 1e^{-3} and a cosine learning rate scheduler. During training, a patch size of 128x160x112 voxels was used rather than the whole volume for training the nnU-Net, with a batch size of two. The backbone was the ResEnc U-Net with six stages and feature widths of 32, 64, 128, 256, 320 and 320. All ensemble backbones were trained for 500 epochs.

3 Results

The results were computed with 5-fold validation on the labelled training dataset. The ablation results for each of blocks on internal validation dataset are given in Table 1. The synapse portal submission results on external validation dataset are given in Table 2. On submission portal, the final configuration reached a mean Dice of 0.840, a

mean NSD of 0.497 and a mean 95th percentile HD of 21.3 mm, the best values in every region-averaged column. The ET Dice scores increased from 0.794 to 0.812 after re-labelling. Fig. 5 shows representative cases in which RACER prediction closely follows the reference on the first two rows and loses part of the enhancing rim on the weaker case. Fig. 6 shows the RACER predictions on unseen validation data showing results across differing tumor appearances.

Table 1. Internal five-fold cross-validation region Dice coefficient for the three backbones, computed from 5-fold predictions on the training data. SinMIM: Simple Masked Image Modelling; GT: Ground Truth; PL: Pseudo Label; ET: Enhancing Tumor; TC: Tumor Core; WT: Whole Tumor. Ensemble comprises of nnU-Net (no FiLM) + SimMIM v2 + SimMIM v3.

Backbone	Augmentation Model	Training Dataset	ET	TC	WT	Mean
nnU-Net (no FiLM)	-	GT	0.852 ± 0.209	0.907 ± 0.153	0.920 ± 0.105	0.893 ± 0.129
SwinUN-ETR	-	GT	0.853	0.895	0.919	0.889
Cohort Aug v2 (nnU-Net + FiLM)	-	GT	0.854 ± 0.205	0.907 ± 0.140	0.923 ± 0.093	0.895 ± 0.120
Cohort Aug v3 (nnU-Net + FiLM)	-	GT	0.855 ± 0.199	0.908 ± 0.134	0.924 ± 0.092	0.895 ± 0.116
nnU-Net (no FiLM)	-	GT+PL	0.861 ± 0.214	0.907 ± 0.161	0.922 ± 0.082	0.897 ± 0.124
SimMIM v2	Cohort Aug v2	GT+PL	0.852 ± 0.200	0.910 ± 0.159	0.926 ± 0.086	0.896 ± 0.121
SimMIM v3	Cohort Aug v3	GT+PL	0.851 ± 0.189	0.910 ± 0.135	0.926 ± 0.076	0.896 ± 0.106
Ensemble	Cohort Aug v2+v3	GT+PL	0.860 ± 0.196	0.911 ± 0.150	0.930 ± 0.078	0.900 ± 0.114
Ensemble + Refiner	Cohort Aug v2+v3	GT+PL	0.859 ± 0.205	0.914 ± 0.145	0.931 ± 0.083	0.901 ± 0.115
RACER (final)	**Cohort Aug v2+v3**	**GT+PL**	**0.875 ± 0.179**	**0.914 ± 0.145**	**0.931 ± 0.083**	**0.907 ± 0.109**

Table 2. Ablation on the official validation portal. Ensemble models include nnU-Net, SimMIM v2 and SimMIM v3. Enhancing Tumor; TC: Tumor Core; WT: Whole Tumor. DSC: Dice Similarity Coefficient; NSD: Normalized Surface Dice; HD95: 95th percentile Hausdorff Distance

Configuration	DSC ET	DSC TC	DSC WT	DSC mean	NSD mean	HD95 mean
Ensemble	0.772 ± 0.295	0.823 ± 0.248	0.876 ± 0.175	0.824 ± 0.203	0.494 ± 0.212	24.8 ± 62.8

Ensemble + refiner	0.794 ± 0.271	0.830 ± 0.236	0.877 ± 0.174	0.834 ± 0.194	0.496 ± 0.210	22.5 ± 60.41
RACER (final)	**0.812 ± 0.251**	**0.830 ± 0.236**	**0.877 ± 0.174**	**0.840 ± 0.190**	**0.497 ± 0.209**	**21.3 ± 59.61**

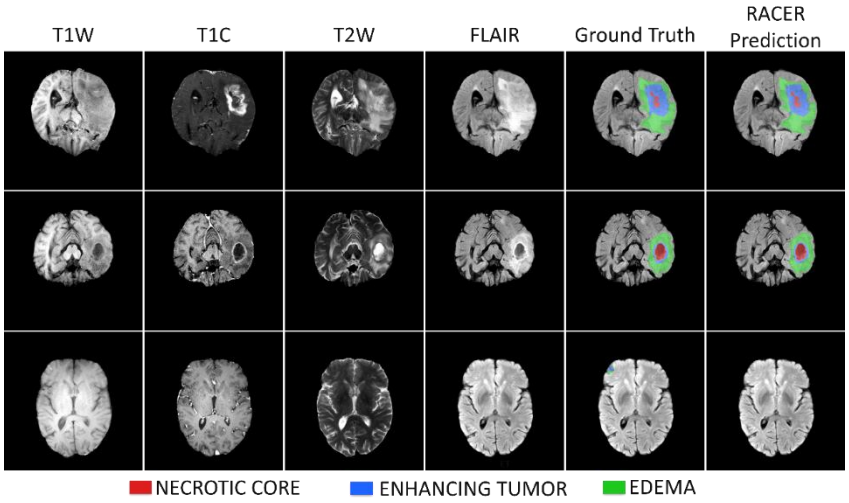

Fig. 5. Internal held-out cases with reference and RACER prediction overlaid on FLAIR. Row 1, row 2 and row 3 show a strong, an average and a weaker case, respectively.

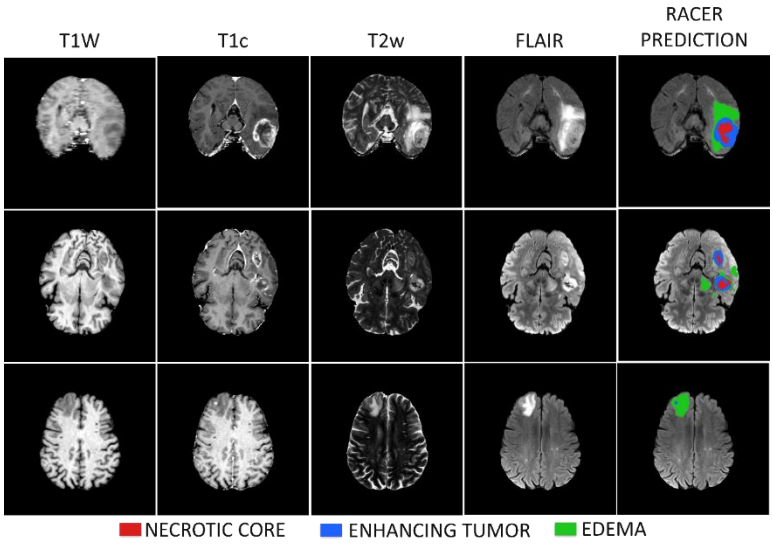

Fig. 6: Validation cases, for which no reference is available, showing the four MRI channels and the RACER prediction overlaid on FLAIR.

4 Discussion

The generalizability of models is an important clinical factor in the adoption of deep learning-based models in hospital settings where the samples from the target class may be less or not available during the model training. The proposed RACER framework establishes a solution for the generalizability in brain tumor segmentation. The training dataset contains different brain tumor types. It includes glioma, meningioma and metastasis whereas the validation dataset also contains additional tumor types like pediatric and SSA cases whose intensity statistics and lesion morphology differ from the training data distribution. The proposed framework trains a model on supervised and SSL techniques. As the annotated data in healthcare is limited, the application of SSL techniques has grown considerably for learning a diverse distribution of data from a large available unlabeled data cohort. Our method uses cohort-based augmentation to widen the sample space for model training as well as generated the PL dataset to increase the training cohort for the final model. The backbone remains the nnU-Net model which showed greater accuracy as compared to other models. The feature-wise linear modulation technique was used to refocus the learned features based on the radiomic features of the input images. The final model for the proposed framework ensembles the core nnU-Net and the SSL based pretrained SimMIM models followed by a dual plain CNN-based refiner network and a postprocessing step. Thus, the proposed framework is referred to as Radiomic-Adaptive Cohort Ensemble with Refinement (RACER).

The results support different modules used in the framework while also exposing where the method is limited. The SSL blocks improved TC and WT dice and lowered their boundary distances. The value of the self-supervised and PL representation therefore appeared after refinement rather than in the raw ensemble average, and the refiner, which reads those representations, produced the best ET Dice of any ensemble in Table 2. SSL based masked image modelling and per-case conditioning are thus best understood as representation-shaping steps whose benefit is seen in downstream task like segmentation.

The refiner had multiple segmentation heads which were used as regularizer during training. Because the region head is a zero-initialized residual on the ensemble logits, the network should not start worse than its input and only learns a correction, therefore a refiner was able to improve every region over the ensemble it refines. The cross-attention between the region prior and the image features [13] helps the refiner to check the raw MRI where the ensemble is uncertain. An ablation on the held-out data found that the region head alone gave the best segmentation, so only that head was used at inference.

The cohort-safe relabeling follows previous GoAT challenges submissions and is closely related to other generalization-oriented post-processing for this task [14]. It was a hit-and-trail method applied to enhancing tumor labels only. It's a trade-off between removing any small enhancing tumor and false positives. The dice gain was achieved on submission portal.

On the internal validation held-out data, the final pipeline reached a mean Dice of 0.907, a mean normalized surface Dice of 0.908 and a mean 95^{th} percentile Hausdorff

distance of 3.31 mm, whereas on the mixed-cohort portal the same pipeline reached 0.840, 0.497 and 21.3 mm. The large fall in the surface metrics was seen due to the missed and poorly segmented small lesions. The large lesions showed good segmentation results for all evaluation metrics but the small lesions were segmented poorly which resulted in poor metrics. Thus, boundary-oriented post-processing was therefore not expected to help, and a test-time augmentation experiment on the held-out data confirmed a negligible change, so it was not adopted.

5 Limitations

The proposed RACER framework has few limitations. The training data contain no pediatric or SSA, so all cross-cohort robustness is obtained indirectly through augmentation and self-supervision, and the unseen cohorts remain the weakest part of the results. The refiner used a single fold as refiner ensemble would help in small lesions but due to high computational cost, only single fold refining was done. Finally, the enhancing tumor improvement is in part a scoring-convention effect based on a hit-and-trail method.

6 Conclusion

A cross-cohort brain tumor segmentation pipeline RACER was proposed that combines SimMIM self-supervised initialization, radiomic per-case conditioning and cohort-shift augmentation in the backbones with a light residual refiner and a cohort-safe relabeling. On the BraTS-GoAT 2026 validation portal, the pipeline improved the mean Dice, NSD and HD over a strong ensemble baseline. The analysis of the in-distribution and cross-cohort results indicates that the poor results are due to small set of out-of-distribution cases, which points to cohort-targeted data and refinement as future work.

Acknowledgements

The authors gratefully acknowledge the HPC facility of Indian Institute of Technology Delhi for providing the computational infrastructure and resources that supported this research. The authors also thank the organizers of the BraTS Challenge for making the dataset available to the research community. This work was supported by the Indian Council for Medical Research, project number: CAR-2024-01-000187 and WIN-COE, project number: IA -0000000037.

Disclosure of Interests

The authors declare that they have no competing interests.

References

1. Baid, Ujjwal, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani et al. "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification." *ArXiv*, (2021). https://arxiv.org/abs/2107.02314.
2. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted

Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham.
3. Isensee, Fabian, Paul F. Jaeger, Simon A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. "NnU-Net: A Self-configuring Method for Deep Learning-based Biomedical Image Segmentation." *Nature Methods 18*, no. 2 (2021): 203-211. https://doi.org/10.1038/s41592-020-01008-z.
4. Myronenko, Andriy. "3D MRI Brain Tumor Segmentation Using Autoencoder Regularization." ArXiv, (2018).
5. Tang, Yucheng, Dong Yang, Wenqi Li, Holger R. Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. "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.
6. BraTS-GoAT Challenge Organizers: Brain Tumor Segmentation (BraTS) Generalizability Across Tumors (GoAT) Challenge. Synapse syn52939291 (2024).
7. Xie, Zhenda, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. "Simmim: A simple framework for masked image modeling." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 9653-9663. 2022.
8. Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. 2018. FiLM: visual reasoning with a general conditioning layer. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18). AAAI Press, Article 483, 3942–3951.
9. D. Karimi and S. E. Salcudean, "Reducing the Hausdorff Distance in Medical Image Segmentation With Convolutional Neural Networks," in IEEE Transactions on Medical Imaging, vol. 39, no. 2, pp. 499-513, Feb. 2020, doi: 10.1109/TMI.2019.2930068.
10. Salehi, S.S.M., Erdogmus, D., Gholipour, A.: Tversky Loss Function for Image Segmentation Using 3D Fully Convolutional Deep Networks. In: MLMI 2017, LNCS 10541, pp. 379-387 (2017).
11. Abraham, N., Khan, N.M.: A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation. In: IEEE ISBI 2019, pp. 683-687 (2019)
12. Hsu, T.-L., et al.: Enhancing Brain Tumor Segmentation Generalizability via Pseudo-Labeling and Ratio-Adaptive Postprocessing. In: BraTS-Lighthouse, MICCAI 2025, LNCS (2025).

13. Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. "Attention is all you need." Advances in neural information processing systems 30 (2017)
14. Nikolov, S., et al.: Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy. J. Med. Internet Res. 23(7), e26151 (2021)..