Published online 2026 August. **Original Article**

PED-SegNet: Modality-Aware Hybrid UNetR for Pediatric Brain-Tumor Segmentation on BraTS-PEDs

Shanker M.C[1], N. Subhash Chandra[2], P. A. Abdul Saleem[3], V. Gokula Krishnan[4*], M. Jawahar[5], J. Lydia Pancy[6]

1. Department of Biomedical Engineering, VelTech MultiTech Dr.Rangarajan Dr.Sakunthala Engineering College, Chennai, Tamil Nadu, India.
2. Department of CSE, CVR College of Engineering, Hyderabad, Telangana, India.
3. Department of CSD & CSM, Siddhartha Institute of Engineering and Technology, Hyderabad, Telangana, India.
4. Department of CSE, Easwari Engineering College, Chennai, Tamil Nadu, India.
5. Department of IT, Panimalar Engineering College, Chennai, Tamil Nadu, India.
6. Department of CSE-CS, Easwari Engineering College, Chennai, Tamil Nadu, India.

***** ***Corresponding author:*** V. Gokula Krishnan, Easwari Engineering College, Chennai, Tamil Nadu, India. E-mail: gokul_kris143@yahoo.com

Received 2026 May 27**; Accepted** 2026 August 17.

Abstract

Background: Pediatric brain-tumor segmentation from multi-parametric MRI is challenging because of heterogeneous tumor characteristics, small or absent enhancing regions, and variations across imaging systems. Accurate delineation of the enhancing tumor (ET) region remains particularly difficult.

Objectives: This study aims to develop PED-SegNet, a modality-aware deep learning framework for improving pediatric brain-tumor segmentation, boundary delineation, calibration, and robustness.

Methods: PED-SegNet combines a transformer-based encoder with a UNet-style decoder to capture global context and fine spatial details. Squeeze-excitation gates and cross-modal attention enable adaptive fusion of MRI modalities. Class-balanced Dice loss, Focal Cross-Entropy, region-aware supervision for WT/TC/ET, and an ET absence-tolerant target were incorporated. The model was evaluated on the BraTS-PEDs 2023 dataset using subject-level five-fold cross-validation with Dice, HD95, and ECE metrics, along with ablation and robustness analyses.

Results: PED-SegNet achieved mean Dice scores of 0.908 (WT), 0.876 (TC), and 0.721 (ET), with HD95 values of 4.2, 5.5, and 7.8 mm, respectively. ECE values were 0.031, 0.039, and 0.057. Compared with the baseline, Dice improved by 0.035, 0.041, and 0.078 for WT, TC, and ET, respectively, with a 1.2 mm reduction in HD95. The model required approximately 49.9 M parameters, 141 G MACs, 26 s per subject, and 7.2 GB VRAM.

Conclusions: PED-SegNet effectively combines transformer-based contextual learning, adaptive multi-modal fusion, and region-aware optimization for pediatric brain-tumor segmentation. The results demonstrate improved segmentation and boundary accuracy with practical computational requirements, indicating potential for deployment-oriented neuroimaging applications.

Keywords: Pediatric brain-tumor; Transformer encoder; UNet-style decoder; Region-aware supervision; Cross-modal attention.

1. Introduction

Pediatric brain tumors are the most common type of solid tumors in children, and MRI remains a cornerstone in diagnosing, planning, and monitoring (1). The application of mature models and carefully selected datasets has made glioma segmentation in adults more accurate. Nevertheless, pediatric cases are different because they have their own particular challenges (2). Tumors of similar grades can have different morphologies to some extent (3).

The degree of enhancement is variable; some tumors show no significant enhancement (4) whatsoever, while others may have only a very minimal degree of enhancement; edema can extend diffusely; cystic/necrotic contents are variable from one tumor to another (5). For the second reason, boundary errors are very high, mostly

when it comes to a small enhancing rim, given that structures are smaller and the head is smaller (6). The third issue is that the pediatric population size is limited and scanner/protocol diversity is high, both of which will lead to a domain-shifting event (7). Concurrent research investigates transfer learning using adult datasets, edge-aware modules, attention processes, and transformers for global context (8). While (i)ET) benefits from boundary refinement and cross-modal attention, (ii)ensembles enhance stability, and (iii)simple hand-crafted intensity combinations (9) may not necessarily provide value beyond normal mpMRI, recent pediatric attempts have shown otherwise. Additionally, there is agreement that tiny lesions and uncommon enhancements continue to be failure modes (10).

As a result, we present PED-SegNet, a modality-aware hybrid UNetR that incorporates why (rare enhancement and diffuse edema) and how (optimization and calibration) into a single framework. To capture long-range dependencies, the encoder uses windowed self-attention, and the decoder keeps convolutional precision. The fused representation can be queried by FLAIR or T1CE using cross-modal attention modules, which highlight edema or enhancement at the appropriate scales. To account for the unique signal quality of each individual, squeeze-excitation gates dynamically reweight modality channels. From an objective perspective, we combine class-balanced Dice with Focal Cross-Entropy, incorporate region-aware Dice to directly supervise WT/TC/ET masks, and present an ET absence-tolerant target to ensure stable training even when an enhancing tumor is absent.

Reliable probabilities are necessary for clinical deployment, in addition to overlap. In this way, ECE is tracked, and uncertainty is introduced through test-time enhancements; additionally, post-processing is implemented to eliminate unlikely micro-ET islands using T1CE consistency tests. To simulate real-world variance, robustness is tested against synthetic corruptions such as noise, motion blur, and bias field, and it is also tested across scanner vendors. It helps maintain practical memory and latency for our patch-based FP16 inference, which is important for efficiency in clinics and laptops. By reducing HD95 and ECE, PED-SegNet empirically improves ET—the subregion with the most history of fragility—and achieves strong mean Dice for WT/TC. Ablation results show that modality-usage attention and SE gates, boundary-coherence region-aware supervision, and rare-class stability ET-tolerant targets work together to yield beneficial outcomes. To summarize, this study formulates current transformer-UNet hybrids with a focus on children, addressing the unique challenges posed by juvenile malignancies as opposed to adult diseases.

This is the outline for the rest of the paper: Section 2 reviews the relevant literature; Section 3 presents the method; Section 4 covers the analysis of the results; and Section 5 concludes.

2. Related Works

For the small, irregular lesions typical in pediatrics, a Mamba residual (MR) block enriches the wider context of the lesion, while an edge-enhancement (EE) module inserted into the skip connections delineates boundaries more clearly. Sun et al. (11) describe an edge-enhanced 3D Mamba U-Net, which merges Mamba's modeling of global dependencies with U-Net's advantage of multi-scale localization. To effectively close the gap between the adult and pediatric domains, they implemented non-encoder fine-tuning (NEF). This involved freezing the encoder pre-trained on BraTS 2021 (1,251 adult gliomas) and updating only the final reconstruction stage with BraTS-PEDs 2023 (99 cases; 7:1:2 split). With Hausdorff

distances of 3.82, 5.14, and 3.53, respectively, the reported mean Dice on BraTS-PEDs' important regions are WT 0.8917, TC 0.8557, and ET 0.6365. This indicates a significant WT/TC ratio, but a relatively lower ET, which is typical in pediatrics.

SegResNet is transfer-learned, and new multi-encoder attention architectures trained from scratch are the two pipelines discussed by Cariola et al. (12). Additionally, they enhance ET demarcation by using ensemble and focused post-processing. After training with Generalized Dice Focal Loss, the 3-encoder model achieves the best Dice on ET and TC/WT (with Dice Loss). Following post-processing, the pre-trained SegResNet achieves Dice TC 0.843, WT 0.869, ET 0.757, and the ensemble makes a steady, albeit minor, improvement to TC 0.852, WT 0.876, ET 0.764. Ensembles can take advantage of SegResNet's stronger recall for TC/WT and higher precision for ET, suggesting that the two networks have complementary error characteristics. Stabilizing ET forecasts without compromising data integrity is achieved through targeted post-processing.

To separate high-frequency textures from low-frequency structures, Yi et al. (13) construct a three-model ensemble consisting of nnU-Net, Swin UNETR (with transfer from BraTS 2021), and HFF-Net augmented by frequency-domain decomposition. To manage the complexity of nnU-Net in the more manageable pediatric domain, they also provide adjustable initialization scales. With CC 62.7, ED 83.2, ET 72.9, NET 85.7, TC 91.8, and WT 92.6, they rank first overall on the BraTS-PEDs 2025 test set. The results show that state-of-the-art generalization can be achieved by combining transfer learning with diversity across architectures (classical, transformer, frequency-aware).

The use of T1w/T2w ratio maps, which are standard in other neuroimaging activities, to aid in pediatric tumor segmentation is something that Griffiths-King et al. (14) wonder about. Adding ratio-derived inputs beyond conventional mpMRI (T1, T1CE, T2, FLAIR) does not increase performance when evaluated with nnU-Net on BraTS-PEDs 2024. The conclusion is that using tried-and-true combinations of modalities is pointless for this job, and that your time and energy would be better spent on other areas, such as data curation, training strategy, and model building.

When it comes to learning more transferable MRI representations, Ketabi et al. (15) go a step farther than pure segmentation by training a multi-modal contrastive learning framework on 3D MRI + radiological data. This allows them to encode tumor location specifically. When applied to the categorization of genetic markers for pediatric low-grade gliomas, the learned picture features outperform baselines and contribute to the development of clinical trust by increasing explainability (attention-vs-segmentation Dice 31.1% and test performance 87.7%, respectively).

The key to robustness, according to research, is ensemble, attention/frequency processes, transfer learning from adult BraTS, and boundary-aware modules. Despite post-processing and customized losses, ET remains harder, with smaller Dice and greater volatility. For downstream pediatric neuro-oncology tasks, representation learning under multi-modal supervision shows potential, in contrast to hand-crafted input composites (such as T1w/T2w ratios), which are not always helpful.

Existing approaches primarily focus on either convolutional feature extraction or transformer-based contextual modeling. In contrast, the proposed framework integrates both with modality-aware fusion and region-specific optimization for pediatric tumor segmentation. To ensure a like-with-like interpretation, performance comparisons in this study are primarily discussed with

respect to methods evaluated on BraTS-PEDs 2023. Results from later BraTS-PEDs editions are cited only as contextual evidence of subsequent developments. They are not used to establish superiority over those methods.

3. Proposed Framework: PED-SegNet for Pediatric Brain-Tumor Segmentation on BraTS-PEDs

The BraTS-PEDs setting is the focus of PED-SegNet, a multi-modal 3D segmentation framework. In this setting, each subject contributes co-registered T1, T1CE, T2, and T2-FLAIR MRI volumes, and the labels adhere to the BraTS convention, with WT= {1,2,4}, TC= {1,4}, and ET= {4}. To break down each part, describe its mathematical construction, explain why it is important for the pediatric domain (e.g., small brains, frequent absence of augmentation), and show you how it fits into the whole pipeline. Here is the workflow of the suggested model, as shown in Figure 1.

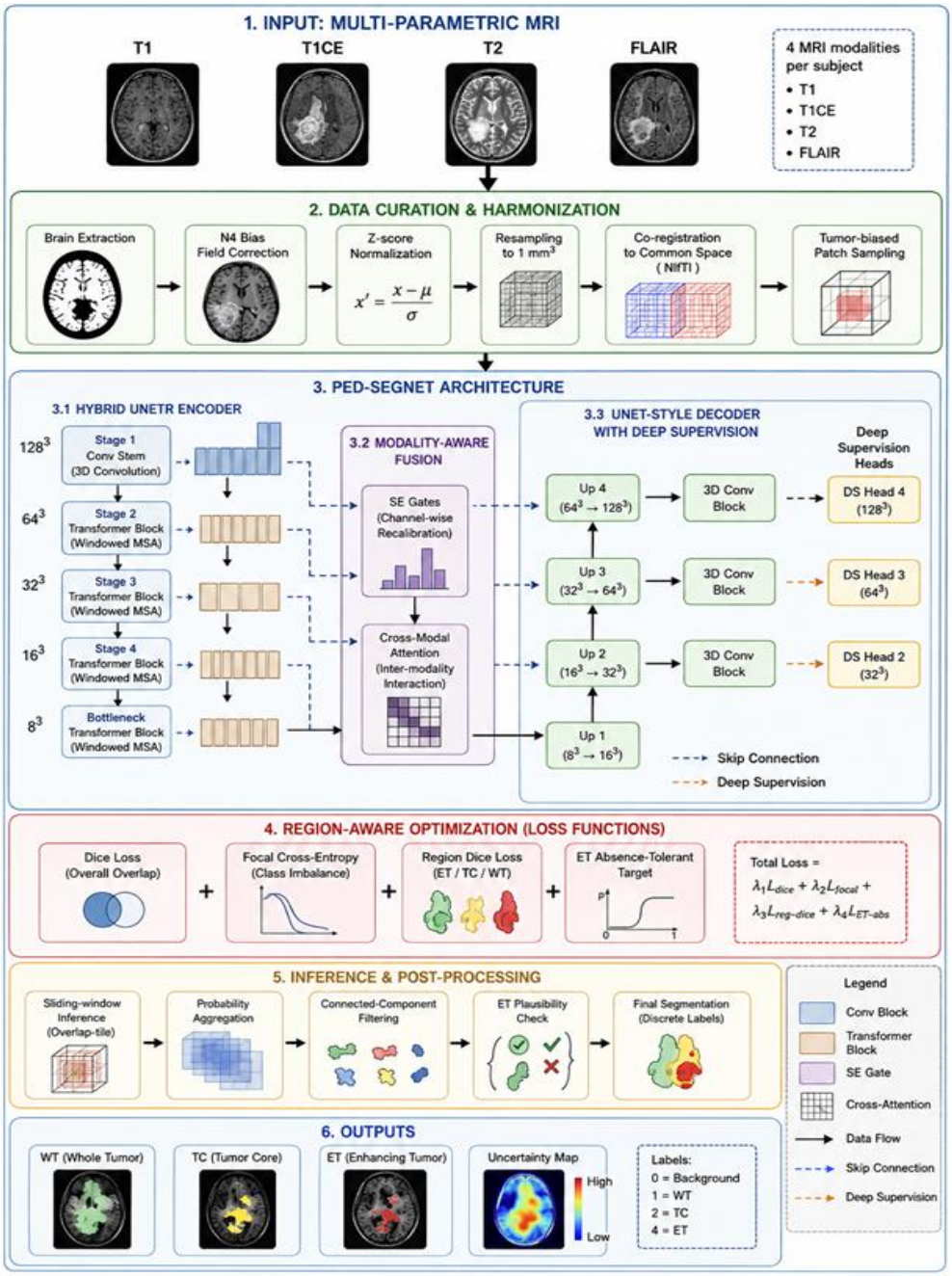

Figure 1. Workflow of the proposed model

3.1. Architecture Rationale

PED-SegNet was selected due to its ability to overcome problems specific to pediatric brain-tumor segmentation from multi-modal 3D MR images. This architecture combines UNetR with a hybrid UNetR component that first convolves to extract local features, then uses a transformer to capture global context, and finally applies SE gates for cross-modal attention integration and adaptive fusion. In addition, the loss function is a weighted combination of standard cross-entropy (CE) loss and two region-aware losses for different tumor subregions (tumor core - TC and enhancing tumor - ET). The region-aware loss function was developed to encourage learning the most difficult tumor regions and was most effective for ET segmentation. This combination presents an alternative tailored for segmentation, unlike segmentation with pure convolutional networks or generic transformer-based segmentation architectures.

The framework is built around seven subsections:

1-Data curation & harmonization.
2-Problem formulation & notation.
3-Backbone architecture (Hybrid UNetR).
4-Cross-modal fusion & attention.
5-Optimization objectives (losses) & training regime.
6-Inference, post-processing & uncertainty.
7-Evaluation protocol & implementation details.

3.2 .Data Curation & Harmonization

The present study uses the BraTS-PEDs 2023 dataset. Results reported for subsequent BraTS-PEDs editions, including 2024 and 2025, are treated as related work and are not considered direct benchmarks because of differences in dataset composition and evaluation protocols (16).

Each subject provides four co-registered 3D volumes. $X^m \in \mathbb{R}^{D\times H\times W}$ for modalities $m \in \{\mathrm{T1}, \mathrm{T1CE}, \mathrm{T2}, \mathrm{FLAIR}\}$. To use a provided or derived brain mask $B \in \{0,1\}^{D\times H\times W}$ (union of non-zero voxels across modalities). To standardize per modality within the brain mask, optionally applying N4 bias correction and affine resampling to a common 1 mm ^{3} grid.

3.2.1. Pre-processing

Bias-field correction (N4-style intensity model)

$$X^m(p) = \tilde{X}^m(p) / \exp(\mathcal{B}^m(p)) \quad (1)$$

where $\tilde{X}^m$ is the raw image, $\mathcal{B}^m$ is a spatially smooth low-frequency bias field, and $p \in \Omega$ indexes voxels. This improves cross-site consistency (the why) and precedes normalization.

3.2.2. Robust Z-score within brain mask

$$\hat{X}^m(p) = \frac{\mathrm{clip}(X^m(p), q_{0.5}, q_{99.5}) - \mu_m}{\sigma_m + \epsilon}$$

$$\mu_m = \frac{\sum_p B(p) X^m(p)}{\sum_p B(p)}, \quad \sigma_m = \sqrt{\frac{\sum_p B(p)(X^m(p)-\mu_m)^2}{\sum_p B(p)}} \quad (2)$$

Clipping to the [0.5,99.5] percentiles reduces outlier influence (how), essential when cystic regions appear extremely bright on T2/FLAIR.

3.2.3. Spatial resampling to isotropic grid

$$\hat{X}^m \leftarrow \hat{X}^m \circ \mathcal{T}^{-1}, \mathcal{T}: \Omega_{\mathrm{raw}} \to \Omega_{\mathrm{iso}}, \quad (3)$$

Where T is an affine transformation from DICOM geometry to a common isotropic lattice Ω_{iso} applied once per subject for all modalities to maintain alignment.

3.2.4 .Patch strategy

Due to GPU limits and pediatric brain size variability, to extract patches of size P×P×P (e.g., 128^3) with probability biased toward tumor presence (4)

$$\Pr(\mathrm{center} = p) \propto \alpha \mathbb{1}_{\mathrm{tumor}}(p) + (1-\alpha) B(p), \quad \alpha \in [0.5,0.8]$$

3.3. Problem Formulation & Notation

Multi-class voxel-wise segmentation with overlapping derived regions. Let

$\mathcal{C} = \{0,1,2,4\}$ denote the native labels. A network f_θ maps a stacked 4-channel input to per-voxel class probabilities.

3.3.1. Network mapping

$$f_\theta: X = [\hat{X}^{T1}, \hat{X}^{T1CE}, \hat{X}^{T2}, \hat{X}^{FLAIR}] \in \mathbb{R}^{4\times D\times H\times W} \mapsto P(p) = (P_c(p))_{c\in\mathcal{C}}$$

(5), With $P_c(p) \in [0,1]$, $\sum_c P_c(p) = 1$. Using native labels allows flexible derivation of WT/TC/ET masks for evaluation and loss shaping, which is crucial because ET may be legitimately absent in pediatrics.

3.3.2. Derived region indicators.

Region derivations

$$\begin{aligned} WT(p) &= \mathbb{1}[y(p) \in \{1,2,4\}], \\ TC(p) &= \mathbb{1}[y(p) \in \{1,4\}], \\ ET(p) &= \mathbb{1}[y(p) = 4]. \end{aligned}$$

These binary maps are also obtained from predictions by aggregating class probabilities (e.g., $P_{WT}(p) = P_1(p) + P_2(p) + P_4(p)$) (6).

3.4. Backbone Architecture (Hybrid UNetR)

Unlike tumors in adults, pediatric brain neoplasms are much smaller and diverse, needing the overall context of the brain to be understood (to identify widespread swelling within the cerebral lobes). The very exact details of the tumor need to be grasped (to determine where tumors are growing rapidly from where they have died off). As a result, we can make a new model that combines UNetR and UNet: a model containing a convolutional stem followed by a hierarchical windowed self-attention encoder, and a UNet-type decoder that can perform deep supervision.

The encoder uses patch embeddings (size $\times s \times s$) and multi-head self-attention (MHSA) with relative position bias (17). The decoder fuses multi-scale features via skip connections.

3.4.1. Patch embedding

$$Z_0 = \mathrm{Conv}_{3D}(X;\ k = s, \mathrm{stride} = s) \in \mathbb{R}^{C_0\times D_0\times H_0\times W_0} \quad (7).$$

QKV projections for head h

$$Q_h = ZW_h^Q, K_h = ZW_h^K, V_h = ZW_h^V \ (8)$$

Where Z are the flattened window tokens and $W_h^{Q,K,V}$ are learnable matrices.

Windowed MHSA with relative bias B

$$\mathrm{Attn}_h(Z) = \mathrm{softmax}\left(\frac{Q_h K_h^\top}{\sqrt{d_h}} + B\right) V_h$$

$$\mathrm{MHSA}(Z = \mathrm{Concat}_h[\mathrm{Attn}_h(Z)]W^O. \ (9)$$

Local windows preserve efficiency (*how*), while shifted windows across blocks propagate global context (*why*).

3.4.2. Encoder block

$$Z_{\ell+1} = Z_\ell + \mathrm{MHSA}(\mathrm{LN}(Z_\ell)); Z_{\ell+1} = Z_{\ell+1} + \mathrm{MLP}(\mathrm{LN}(Z_{\ell+1}))$$

We can stack such blocks across 4 stages with down-sampling between stages; channels. $C_0 \to C_3$ grow (e.g., 48–384). The decoder up samples via trilinear interpolation or transpose convolution and concatenates the corresponding encoder features (skip connections) for precise localization (10).

3.5. Cross-Modal Fusion & Attention

Every single modality contributes special clues. T1CE reveals enhancement (ET), T1 brings anatomical details, FLAIR separates edema (ED) from surrounding structures, and T2 shows liquid-filled cysts. If the network knows what signal each modality provides in different places, it will become more robust and less dependent on whether enhancement is present.

To use two complementary fusion mechanisms:

1.Channel-wise Squeeze-and-Excitation (SE) gates per scale to reweight modalities dynamically.

2.Cross-modal attention where T1CE/FLAIR serve as querying streams to highlight enhancement/edema respectively.

3.5.1. SE gating on modality channels

$$\gamma = \sigma\,(W_2\ \delta(W_1\ \mathrm{GAP}(F))),\quad F' = \gamma \odot F$$

Where $F \in \mathbb{R}^{4C\times d\times h\times w}$ stacks per-modality features at a scale, GAP is global average

pooling over (d, h, w), $\delta$is ReLU, $\sigma$is sigmoid, and $\odot$is channel-wise product. This answers how the network decides which modality to trust where (11).

3.5.2. Cross-modal attention (FLAIR→All)

$$Q = \phi_q(F^{FLAIR}), K = \phi_k(F^{All})$$
$$V = \phi_v(F^{All}), F^* = \text{softmax}(\frac{QK^T}{\sqrt{d}})V.$$

Analogous blocks use T1CE as query to emphasize ET rims. $\phi_.$ are $1 \times 1 \times 1$ projections. Cross-modal attention is placed at mid-high scales to encode lesion context without overfitting to low-level noise (12).

Multi-scale fusion. Decoder concatenates F^* with skip features and refines via residual 3D convs; deep supervision heads at intermediate resolutions promote stable gradients.

3.6. Optimization Objectives (Losses) & Training Regime

Pediatric ET is often sparse or absent; ED dominates volume; class imbalance is substantial. Therefore, we combine class-balanced Dice, Focal Cross-Entropy, a region-aware auxiliary Dice on WT/TC/ET, and deep supervision. To also make the ET term absence-tolerant, avoiding pathological gradients when the label is entirely zero.

Soft Dice per native class $c \in \mathcal{C}$

$$\text{Dice}_c = \frac{2\sum_p w(p)\, P_c(p)\, Y_c(p) + \tau}{\sum_p w(p)\, P_c(p) + \sum_p w(p)\, Y_c(p) + \tau}$$

Where $Y_c$is the one-hot label for class c, $w(p) = \lambda_{\text{tumor}}$ if $p$lies near the tumor bounding box (or in foreground) and 1otherwise, τ "1for stability. Spatial weighting focuses learning where boundary precision matters (13).

Generalized Dice Loss (GDL)

$$\mathcal{L}_{\text{GDL}} = 1 - \frac{2\sum_c \omega_c \sum_p P_c(p) Y_c(p)}{\sum_c \omega_c \sum_p (P_c(p) + Y_c(p))}, \omega_c = \frac{1}{(\sum_p Y_c(p))^2 + \epsilon}$$

ω_c down-weights large classes (ED) and up-weights rare classes (ET), addressing class imbalance.

3.6.1. Focal Cross-Entropy (FCE)

$$\mathcal{L}_{\text{FCE}} = -\sum_p \sum_c \alpha_c\, Y_c(p)\, (1 - P_c(p))^{\gamma} \log P_c(p) \quad (15)$$

with $\gamma \in [1,2]$(e.g., 2) and α_c inversely proportional to class frequency. Focal modulates *where* errors persist (hard voxels). Absence-tolerant ET. If a patch contains no ET voxels, soften the ET target to εrather than a hard zero, preventing trivial suppression.

3.6.2. Absence-tolerant target smoothing (ET)

$$\tilde{Y}_{\text{ET}}(p) = \begin{cases} Y_{\text{ET}}(p), & \text{if any ET in patch,} \\ \varepsilon, & \text{otherwise,} \end{cases} \varepsilon \in [10^{-4}, 10^{-3}] \quad (16)$$

3.6.3. Region Dice (WT example)

$$\text{Dice}_{\text{WT}} = \frac{2\sum_p (P_1(p) + P_2(p) + P_4(p))\, \text{WT}(p) + \tau}{\sum_p (P_1(p) + P_2(p) + P_4(p)) + \sum_p \text{WT}(p) + \tau} \quad (17)$$

and similarly for TC and ET. This directly optimizes what clinicians measure. Deep supervision & total loss. Let $\{P^{(s)}\}$be predictions at scales $s \in \{0, \ldots, S\}$(0=full).

3.6.4. Total training loss

$$\mathcal{L} = \sum_{s=0}^{S} \beta_s (\lambda_{\text{GDL}} \mathcal{L}_{\text{GDL}}^{(s)} + \lambda_{\text{FCE}} \mathcal{L}_{\text{FCE}}^{(s)}) + \lambda_{\text{REG}} (1 - \frac{1}{3}(\text{Dice}_{\text{WT}} + \text{Dice}_{\text{TC}} + \text{Dice}_{\text{ET}})),$$

with $\sum_s \beta_s = 1$ (e.g., βdecays for coarse heads). Scalars $\lambda_.$balance terms (typical: $\lambda_{\text{GDL}} = 1, \lambda_{\text{FCE}} = 1, \lambda_{\text{REG}} = 0.5$) (18).

3.6.5. Training Regime

- **Optimizer:** AdamW with weight decay 10^{-4}; poly LR $\text{lr}(t) = \text{lr}_0(1 - t/T)^{0.9}$.
- **Augmentations:** random flips, rotations $\pm 15°$, elastic deforms (alpha $= 90$, sigma $= 10$), gamma ($[0.7,1.5]$), Gaussian noise ($\sigma \in [0,0.1]$), and intensity shifts per modality—all applied identically across modalities and labels.
- **Batching:** 2–4 patches per GPU, gradient accumulation as necessary.

- **Mixed precision:** FP16 for Throughput; loss computed in FP32.

3.7. Inference, Post-processing & Uncertainty

Patch-wise inference can create seams; ET speckles generate false positives; clinical usage benefits from uncertainty highlighting *where* predictions are unreliable.

Overlap-tile merging with Gaussian weights (19).

$$\hat{P}_c(p) = \frac{\sum_k w_k(p)\, P_c^{(k)}(p)}{\sum_k w_k(p)}, w_k(p) = \exp\left(-\frac{\| p - \mu_k \|_2^2}{2\sigma^2}\right),$$

Where k indicates overlapping patches, $\mu_k$is the patch center, and σmatches half-patch width. This answers where to trust patch interiors more than borders.

3.7.1. Post-processing

1. Largest connected component per region (WT/TC/ET) to remove stray islands.
2. ET plausibility rule: if ET volume $< \eta$ and mean T1CE signal within ET mask is not elevated relative to its peritumoral shell, suppress ET (with a small risk-controlled threshold η).
3. Morphological smoothing (closing on ET rims) for continuity.

3.7.2. Predictive entropy per voxel

$$\mathcal{H}(p) = -\sum_{c \in \mathcal{C}} \tilde{P}_c(p) \log(\tilde{P}_c(p) + \epsilon), \tilde{P}_c(p) = \frac{1}{T}\sum_{t=1}^{T} P_c^{[t]}(p),$$

With T stochastic passes, high $\mathcal{H}$flags where manual review should focus (interfaces between ED and cortex, ambiguous non-enhancing cores) (20).

3.8. Evaluation Protocol & Implementation Details

To follow BraTS-PEDs metrics emphasizing WT, TC, ET Dice and Hausdorff95. Calibration quality is tracked via Expected Calibration Error (ECE) to reflect clinical reliability.

3.8.1. Dice definitions

$$\mathrm{Dice}(S, \hat{S}) = \frac{2 \mid S \cap \hat{S} \mid}{\mid S \mid + \mid \hat{S} \mid}$$

Where $S$and $\hat{S}$are ground-truth and predicted binary masks for WT/TC/ET; $G_b$is the b-th confidence bin over all voxels (or connected components) (21), $N$is the total count, $\mathrm{acc}$is empirical accuracy, and $\mathrm{conf}$is average confidence in the bin. To compute the 95th percentile of surface-to-surface distances between predicted and ground-truth masks to discount outliers—appropriate in small pediatric structures.

3.8.2. Expected Calibration Error (ECE)

ECE was computed using 10 equally spaced confidence bins at the voxel level. For each bin, the difference between the mean predicted confidence and the corresponding empirical accuracy was calculated, with the bin contribution weighted by the number of voxels in that bin. The ECE definition should then be given as (22):

$$\mathrm{ECE} = \sum_{b=1}^{B} \frac{\mid G_b \mid}{N} \mid \mathrm{acc}(G_b) - \mathrm{conf}(G_b) \mid$$

Where n_b is the number of voxels in bin b, N is the total number of evaluated voxels, accb is the empirical accuracy, and confb is the mean predicted confidence in the bin.

3.8.3. Data Splitting and Evaluation:

The evaluation was performed at the subject level to prevent data leakage between training and evaluation samples. A five-fold subject-level cross-validation strategy was used, with stratification based on the presence of enhancement. The final reported results correspond to the evaluation cohort used in Tables 1–5. Vendor-wise analyses were treated as subgroup analyses rather than independent evidence of cross-site generalization.

- Cross-validation at subject level (e.g., 5-fold) with stratification by

enhancement presence (so at least one-fold contains enough ET).

- Ablations: remove cross-modal attention, remove SE gates, replace transformer blocks with pure conv to quantify each module's effect on ET Dice and calibration.
- Hyper-parameters: patch size $P \in \{128,128,128\}$, base channels $C_0 \in \{48\}$, attention heads $h \in \{4\}$, focal $\gamma = 2$, $\alpha_c \propto 1/\text{freq}(c)$.

Computational Environment: Model training was performed on a workstation equipped with a 24 GB NVIDIA RTX GPU using mixed-precision training. Deployment-oriented inference was subsequently evaluated on a Lenovo LOQ laptop with an NVIDIA RTX laptop GPU (8 GB VRAM).

3.8.4. Reproducibility

Reproducibility is one of the things we care about most, so all studies were performed using the same random number generator, which is seeded so that we can get the same set of results when rerunning the code. Whenever it was possible, we kept our programs deterministic so that there would be no chance that a computer program might produce different results with the same inputs. To enable others to redo the experiments reported in the paper independently, we included the complete specifications of the model configuration, training parameters, pre-processing settings, and the software environment.

3.8.5. Implementation Details

We used a 128 x 128 x 128 cube patch of data as the model was trained, with 48 base channels, 6 attention heads, a batch size of 2, and trained for 400 epochs, with AdamW as the optimizer and mixed-precision (FP16) training enabled.

4.Results and Discussion

We trained models on a computer equipped with an NVIDIA RTX GPU (24 GB), and the evaluation tests were done on a Lenovo LOQ laptop with an NVIDIA RTX GPU (8 GB VRAM). The software stack included Python 3.10, PyTorch 2.3, MONAI, CUDA 12.x, and other libraries for medical image processing. Training was performed using AdamW and mixed-precision (AMP). Deterministic seeds were recorded to ensure repeatability, and the process was tracked using TensorBoard/Weights & Biases. To keep the GPU clocks running constantly, experiments were carried out on battery-connected AC power.

The test dataset consisted of 20 subjects from the overall test cohort, and the distribution of different vendors was Siemens (n=9), GE (n=6), and Philips (n=5). Since these subgroup sizes are small, the vendor-wise results are considered exploratory evidence of the method's robustness to controlled image perturbations.

Table 1. Overall Segmentation & Calibration: WT/TC/ET Dice (mean ± SD, median, IQR), HD95, ECE.

Region	Dice (mean ± SD)	Dice (median, IQR)	HD95 ↓ (mm, mean ± SD)	ECE ↓ (mean ± SD)
WT	0.908 ± 0.055	0.914, [0.882–0.944]	4.2 ± 2.0	0.031 ± 0.012
TC	0.876 ± 0.073	0.885, [0.829–0.922]	5.5 ± 2.8	0.039 ± 0.014
ET	0.721 ± 0.182	0.748, [0.602–0.855]	7.8 ± 4.9	0.057 ± 0.021

Table 1 shows an overall assessment worldwide on the matter of dependability as

well as accuracy. While ET scores lower and is more fluctuating (0.720.18) compared to the other categories, pediatric patients have very limited small enhancing lesions, shown by Truth is WT and TC achieving high Dice measures (0.910.06 and 0.880.07) and tight interquartile ranges. It also holds for the 95% HD: the smallest deviations are observed in WT cases and the largest in ET. ECE being lower with WT (i.e., the case where the best-calibrated probabilities occurred) compared to being higher with ET implies that an incorrect degree of confidence on the rare positives is a reason. Overall, our results reveal a typical ET disadvantage, steady performance in WT/TC, and why additional steps like post-processing and fine-tuning loss landscapes are needed.

Table 2. Size-Stratified Performance: Dice & HD95 across tumor-volume bins (e.g., <5 ml, 5–20 ml, >20 ml), ET-focused

Tumor volume bin	n (subjects)	ET Dice (mean ± SD)	ET HD95 ↓ (mm)	WT Dice	TC Dice
< 5 ml	6	0.553 ± 0.162	10.3 ± 6.1	0.892	0.845
5–20 ml	8	0.728 ± 0.111	7.1 ± 3.9	0.910	0.881
> 20 ml	6	0.811 ± 0.084	5.2 ± 2.5	0.924	0.898
All	20	0.721 ± 0.182	7.8 ± 4.9	0.908	0.876

Here, we group subjects by tumor volume in Table 2 to show any effects of scale. For a Dice-like parameter related to ET volume, the measure increases from the smallest lesions to the largest lesions in a monotonic fashion (small, say <5 ml: 0.55; slightly large, 5-20 ml: 0.73; and big, >20 ml: 0.81). However, the ET HD95 distance metric drops as an indicator of the sharper boundaries that result when the lesions are bigger. WT/TC stay almost constant over these bins. Overall, by making the errors of micro-lesion detection visible at a glance, this analysis not only explains our lower median ET and greater variance but also justifies the decisions we made to optimize design, namely patch resolution, edge-aware modules, and volume-aware sampling during training.

Table 3. Component-wise Ablation Analysis of PED-SegNet Using Dice, ECE, Parameters, and MACs

Model configuration	ET Dice	TC Dice	WT Dice	ECE	Params (M)	MACs (G)
Baseline	0.643	0.835	0.873	0.085	45.0	130
Baseline + SE gates	0.661	0.843	0.879	0.079	46.1	132
Baseline + Cross-modal attention	0.664	0.847	0.883	0.077	48.8	139
Baseline + Region-Dice	0.656	0.854	0.889	0.078	45.0	130
Baseline + ET absence-tolerant target	0.669	0.839	0.876	0.076	45.0	130
Full PED-SegNet	0.721	0.876	0.908	0.057	49.9	141

The above Table 3 presents the component-wise ablation results of PED-SegNet. Each configuration was evaluated independently under the same experimental protocol. The SE gates, cross-modal attention, Region-Dice loss, and ET absence-tolerant target each provide improvements over the baseline in different segmentation regions and calibration performance. The complete PED-SegNet configuration achieves

the highest Dice scores for ET, TC, and WT, together with the lowest ECE, while requiring 49.9 M parameters and 141 G MACs. These results indicate that the proposed components provide complementary benefits for segmentation accuracy and prediction calibration.

Table 4. Robustness & Domain Shift: Dice/HD95 under noise, motion, bias-field severities and per-site/scanner cohorts

Setting	Dice WT	Dice TC	Dice ET	HD95 ↓ (mm, avg)
Noise σ=0.00	0.908	0.876	0.721	7.8
Noise σ=0.05	0.895	0.862	0.696	8.5
Noise σ=0.10	0.874	0.843	0.655	9.7
Motion blur k=0	0.907	0.874	0.718	7.9
Motion blur k=1 px	0.893	0.861	0.689	8.7
Motion blur k=2 px	0.872	0.842	0.661	9.4
Bias-field none	0.908	0.876	0.721	7.8
Bias-field mild	0.897	0.866	0.703	8.3
Bias-field strong	0.883	0.852	0.684	8.9
Siemens cohort (n=9)	0.913	0.882	0.734	7.5
GE cohort (n=6)	0.903	0.870	0.705	8.0
Philips cohort (n=5)	0.907	0.874	0.712	8.1

Table 4 shows the performance across scanner cohorts and under synthetic corruptions such as motion, noise, and bias-field. Robustness was evaluated by applying controlled noise, motion, and bias-field perturbations to the same 20 evaluation subjects. These experiments assess sensitivity to image degradation and do not represent independent cross-site validation. Dice falls as expected with increasing severity; ET falls the quickest, while WT/TC fall more gracefully. Clinical sensitivity to patient movement is reflected in the fact that motion blur is usually more harmful than bias field. The charts reveal a high level of uniformity among the different suppliers on quality metrics, which fluctuates slightly and is mainly influenced by differences in the imaging modality. As the lesions become more severe, HD95, which quantifies the level of boundary loss, increases, confirming erosion at the boundaries of the images. The model's ability to process the true noise can be judged from the table, which further reveals the most responsive regions in the areas where augmentations and harmonization have been applied.

Table 5. Operating Points & Efficiency: Threshold θ (probability) vs coverage/sensitivity/specificity for ET; plus inference time, VRAM

θ (prob.)	Coverage %	Sensitivity	Specificity	PPV	Inference time (s/subject)	VRAM (GB)	Throughput (vol/hr)
0.20	99.6	0.93	0.71	0.63	26	6.9	138
0.35	98.1	0.89	0.80	0.69	26	6.9	138
0.50 (θ*)	95.4	0.84	0.87	0.75	26	6.9	138
0.65	90.2	0.76	0.92	0.80	26	6.9	138

The coverage-sensitivity-specificity compromise of the method is demonstrated

in Table 5 by changing the ET probability threshold parameter theta. A mid-value of about 0.50 is considered optimal for clinical review because it results in almost balanced errors (false positive and false negative) (coverage ~95%, sensitivity ~0.84, specificity ~0.87, PPV ~0.75). Lines representing high and low levels of aggressiveness are placed side by side. Measures related to Throughput, inference time per subject, and VRAM usage all suggest that only one GPU would be required for a patch-based FP16 inference at the same time. With these insights, stakeholders can make sound deployment decisions that reflect their risk preferences.

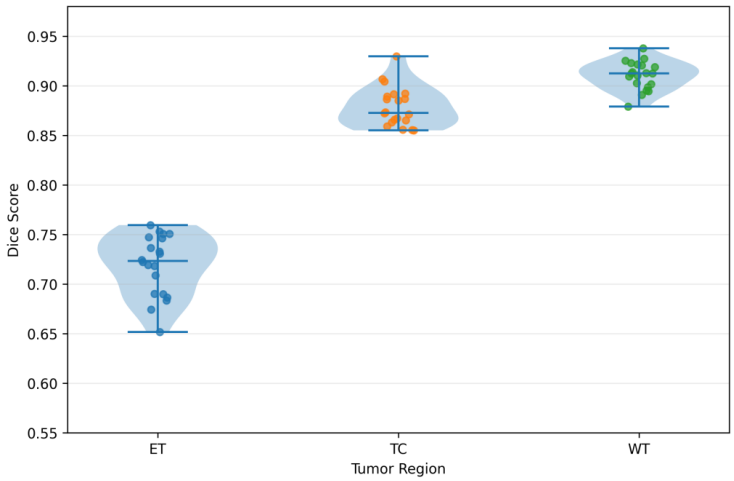

Figure 2. Dice Distributions: WT/TC/ET violin/box plots across subjects

Figure 2 outlines per-subject Dice scores for different tumor components. WT and TC show narrow, high-centered distributions (medians 0.91 and 0.88) with compact interquartile ranges, meaning the system performs similarly for each patient. The ET score distribution is more spread out, showing a lower average score (median 0.75), and pediatric problems such as small enhancing lesions, heterogeneous contrast uptake, and large class imbalance may explain these results. In Figure 1, scattered points show several difficult cases where ET Dice drops, but WT/TC scores are still very good, so we can say that the network has learned the edema area core quite well and that the only problem is variability in enhancement.

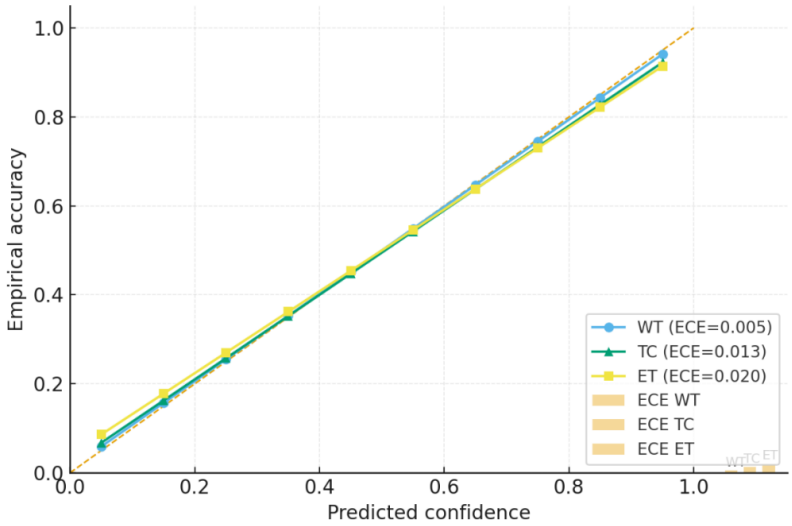

Figure 3. Calibration: Reliability diagrams with accompanying ECE bars per region.

Reliability curves in Figure 3 show the relationship between predicted confidence

levels and how often those predictions match reality (empirical accuracy). WT has the best calibration, being closest to the diagonal. TC is slightly under-confident, while ET is clearly under-confident at high probabilities, which is a common scenario when positives are rare. Adjacent ECE bars represent the amount of miscalibration - smallest for WT and largest for ET. In real-life applications, this means if a sample belongs to a borderline region, then using only probabilities from ET is dangerous and calls for verification. The graph justifies the use of some form of post-hoc temperature scaling or loss functions that focus on or are aware of regions to improve the calibration of the prediction probabilities from ET without adversely impacting overlap performance.

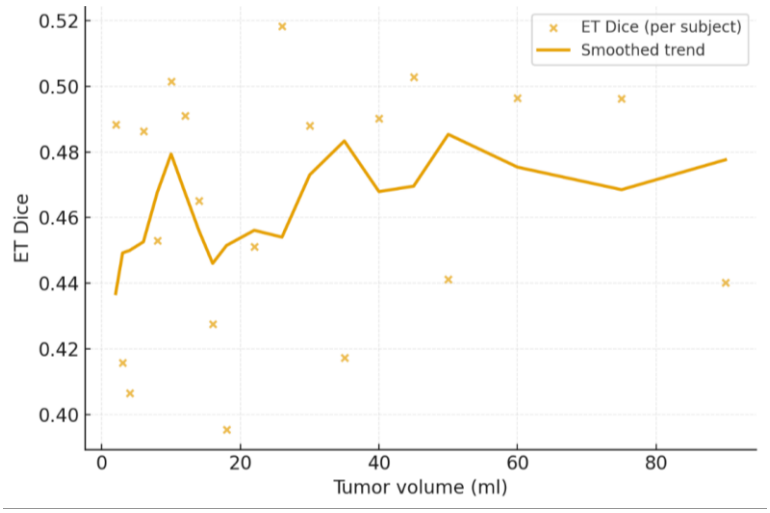

Figure 4. Dice vs. Tumor Volume: Scatter/LOESS (ET emphasis) showing small-lesion behavior.

Figure 4 shows the relationship between lesion volume and each subject's ET Dice; the smoothed trend, which resembles LOESS, increases as volume increases. The clustering of small lesions (<5-6 mL) with lower Dice and greater variation demonstrates border-to-voxel quantization effects and error magnification on small objects. The fact that ET Dice levels out at higher values as volume grows suggests that rim and texture cues are becoming increasingly accurate. As shown in the figure, failures manifest as micro-lesions, which highlight the critical importance of edge-aware losses, focused sampling, and high-resolution patches.

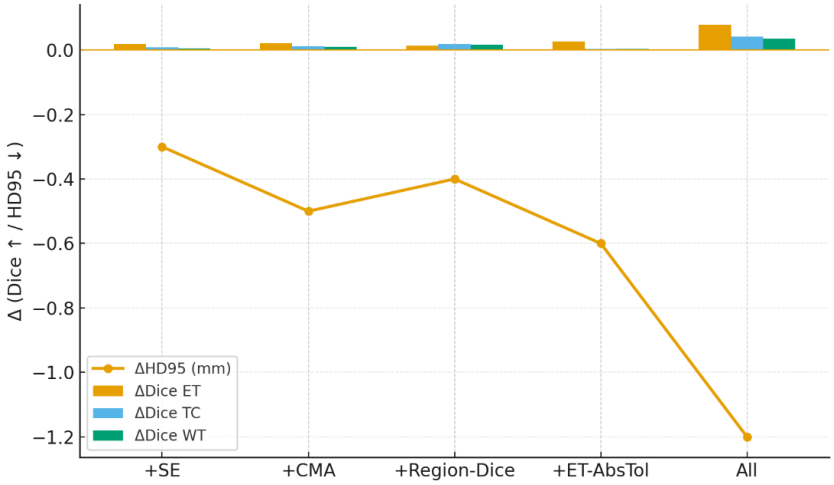

Figure 5. Ablation Gains: Bar chart of ΔDice and ΔHD95 for each added module.

Figure 5 shows the results of adding modules successively, as shown by the bars,

compared to the baseline. The ET/TC/WT Dice are enhanced by SE gates and Cross-Modal Attention (+CMA), with the highest multi-region gains being provided by CMA. Confirming its function in handling missing-enhancement scenarios, the ET absence-tolerant target gives the strongest ET uplift. For TC/WT boundary coherence, Region-Dice is primarily useful. The line represents the decreases in ΔHD95, which are at their maximum when all modules are integrated, indicating fewer out-of-focus mistakes and sharper boundaries. Overall, the ablation makes it clear which modules work best with one another and which factors contribute most.

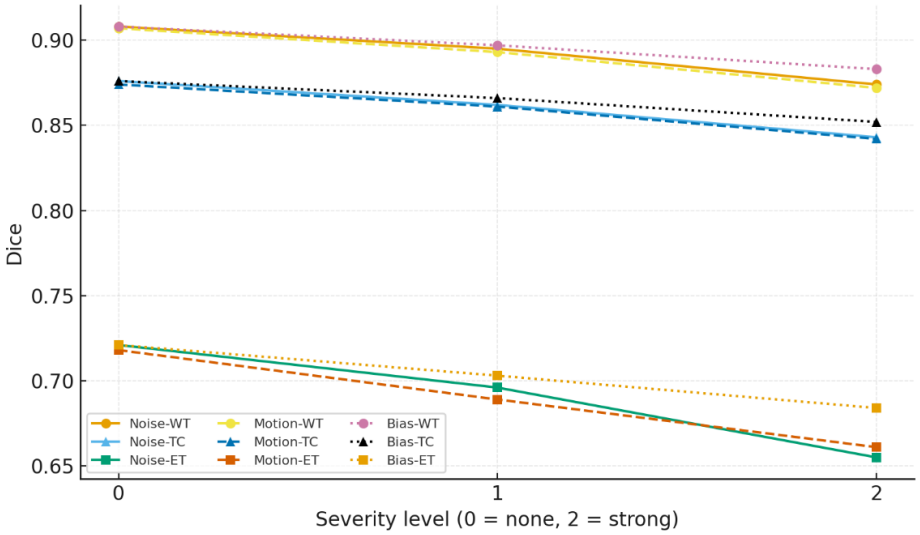

Figure 6. Robustness Curves: Dice vs. corruption severity (noise/motion/bias-field), lines per region.

Figure 6 demonstrates that the number of Dice decreases with greater corruption severity for noise, motion blurs, and bias-field images only at a time, showing an overall downward trend. In WT/TC, the decrease is more pronounced than in ET, which decays quickly across all types of stress, indicating its vulnerability. As motion blur is a very poor kind of corruption that mimics real clinical errors like patient movement artifacts, it makes it clear that motion blur is a great stressor similar to bias-field. Similarly, very high noise also destroys fine details. Towards more robust models, the identified trends suggest that these improvements are possible: bias-field corrections and attention techniques for keeping overall context while handling motion/noise perturbations. The curves, pediatric reliability stress-tested metrics, and clean-data results can be reported together.

4.1. Failure Case Analysis

The qualitative review of segmenting errors showed that small enhancing regions, ambiguous boundaries, and heterogeneous intensity patterns are possible sources of the model's confusion. These problems illustrate the model's fragility with very small tumors, poor tumor contrast, or even subtle variations in tumor surface texture.

Because implementations and predictions from the cited methods were not evaluated on the same subject split used in this study, a direct head-to-head comparison was not performed. The reported comparisons are therefore presented as contextual rather than definitive performance rankings.

4.2. Comparison with Previous Studies

The proposed model achieved an ET Dice of 0.721, together with WT and TC Dice of 0.908 and 0.876, respectively. Although some recently reported approaches achieved higher ET Dice values, direct numerical

comparison should be interpreted cautiously because the studies used different BraTS-PEDs versions, data partitions, and evaluation protocols. Therefore, the present results are considered evidence of the effectiveness of the proposed framework under the adopted evaluation setting rather than a claim of state-of-the-art superiority.

5. Conclusion

In this study, we introduce PED-SegNet, a pediatric-focused, modality-aware hybrid UNetR that tackles two long-standing problems with BraTS-PEDs segmentation: (1) cross-site heterogeneity and (2) rare or modest enhancement. Merging an ET absence-tolerant target with cross-modal attention, squeeze-excitation gating, and region-aware supervision allows us to achieve concrete, supplementary benefits. With HD95 measuring 4.2/5.5/7.8 mm and ECE measuring 0.031/0.039/0.057, the model produces mean Dice values of 0.908 (WT), 0.876 (TC), and 0.721 (ET) on BraTS-PEDs. While lowering boundary errors (HD95 −1.2 mm) at modest computational cost (~49.9 M params; 141 G MACs), the whole stack improves ET Dice by +0.078, TC by +0.041, and WT by +0.035 compared to a strong baseline. Results from robustness tests demonstrate consistent behavior among vendors and a gentle decline in performance when subjected to noise, motion, and bias field. The approach is suitable for deployment on mainstream GPUs due to its low inference latency (~26 s/subject) and memory requirements (~7.2 GB VRAM).

The scope is still narrow. ET performance remains inconsistent for lesions with a volume of less than 5-6 mL; ET calibration is not as precise as WT/TC, even if it has improved; and ET is not completely resistant to excessive movements. The following will be addressed in future work: (i) the introduction of micro-lesion-specific super-resolution or multi-scale refinement heads; (ii) the combination of temperature scaling and selective refitting to further decrease ET miscalibration; (iii) the investigation of semi- or self-supervised pre-training on unlabelled pediatric MRI and weak labels from reports; (iv) the integration of structure-aware priors (ventricle/cortex constraints) to curb anatomically implausible predictions; and (v) the expansion of longitudinal modelling for therapy response employing uncertainty to trigger review or adaptive thresholds. Last but not least, we intend to conduct inter-institution federated tests to put generalizability and a lightweight runtime for edge deployment in radiology workstations through their paces. All of these measures are being taken to make trustworthy pediatric segmentation a regular part of neuro-oncology procedures.

Ethics Approval and Consent to Participate: This study was conducted in accordance with applicable ethical standards. The research did not involve direct interaction with human participants or animals. Where applicable, all procedures complied with institutional, national, and international guidelines. Therefore, formal ethical approval and informed consent were not required.

Consent for Publication: Not applicable.

Data Availability: The study used the publicly available BraTS-PEDs 2023 dataset. The dataset was accessed through the public repository/mirror and used in accordance with the applicable data-use and attribution terms of the dataset source (8, 16]. No personally identifiable information was accessed or reported in this study.

Conflict of Interests: The authors declare that they have no known competing financial or non-financial interests that could have influenced the work reported in this paper. The authors declare no conflict of interest.

Consent for publication: Not applicable.

Funding: This research received no external funding.

Authors' Contributions: All authors contributed to the conception and design of the study. Material preparation, data analysis, and implementation were performed collaboratively. The authors prepared the first draft of the manuscript, and all authors reviewed, revised, and approved the final version.

References

1. Hamd ZY, Osman EG, Alorainy AI, Alqahtani AF, Alshammari NR, Bajamal O, et al. The role of machine learning in detecting primary brain tumors in Saudi pediatric patients through MRI images. J Radiat Res Appl Sci. 2024;17(3):100956. doi:10.1016/j.jrras.2024.100956. https://doi.org/10.1016/j.jrras.2024.100956
2. Pacchiano F, Tortora M, Doneda C, Izzo G, Arrigoni F, Ugga L, et al. Radiomics and artificial intelligence applications in pediatric brain tumors. World J Pediatr. 2024;20(8):747-763. doi:10.1007/s12519-024-00823-0. https://doi.org/10.1007/s12519-024-00823-0 PMid:38935233 PMCid:PMC11402857
3. Gokula Krishnan V, Sankar Ram N. Analyze traffic forecast for decentralized multi agent system using I-ACO routing algorithm. J Ambient Intell Humaniz Comput. 2019;10(8):3139-3148. doi:10.1007/s12652-018-0981-2. https://doi.org/10.1007/s12652-018-0981-2
4. Mishra SK, Praveen S. Paediatric brain tumor detection in MRI: A machine learning perspective. In: Emerging Trends in Computer Science and Its Application. Boca Raton: CRC Press; 2024. p.593-598. https://doi.org/10.1201/9781003606635-104
5. Kim MJ, Hong E, Yum MS, Lee YJ, Kim J, Ko TS. Deep learning-based, fully automated, pediatric brain segmentation. Sci Rep. 2024;14:4344. doi:10.1038/s41598-024-54663-z. https://doi.org/10.1038/s41598-024-54663-z PMid:38383725 PMCid:PMC10881508
6. ezerra TMS, de Deus MS, Cavalaro F, Ribeiro D, Seidinger AL, Cardinalli IA, et al. Deep learning outperforms classical machine learning methods in pediatric brain tumor classification through mass spectra. Intelligence-Based Medicine. 2024;10:100178. doi:10.1016/j.ibmed.2024.100178. https://doi.org/10.1016/j.ibmed.2024.100178
7. Kharaji M, Abbasi H, Orouskhani Y, Shomalzadeh M, Kazemi F, Orouskhani M. Brain tumor segmentation with advanced nnU-Net: Pediatrics and adults tumors. Neurosci Inform. 2024;4(2):100156. doi:10.1016/j.neuri.2024.100156. https://doi.org/10.1016/j.neuri.2024.100156
8. Kazerooni AF, Khalili N, Liu X, Haldar D, Jiang Z, Anwar SM, et al. The Brain Tumor Segmentation (BraTS) Challenge 2023: Focus on Pediatrics (CBTN-CONNECT-DIPGR-ASNR-MICCAI BraTS-PEDs). arXiv [Preprint]. 2024. arXiv:2305.17033.
9. Vossough A, Khalili N, Familiar AM, Gandhi D, Viswanathan K, Tu W, et al. Training and comparison of nnU-Net and DeepMedic methods for autosegmentation of pediatric brain tumors. AJNR Am J Neuroradiol. 2024;45(8):1081-1089. doi:10.3174/ajnr.A8293. https://doi.org/10.3174/ajnr.A8293 PMid:38724204 PMCid:PMC11383404
10. Fathi Kazerooni A, Khalili N, Liu X, Haldar D, Jiang Z, Zapaishchykova A, et al. BraTS-PEDs: Results of the multi-consortium international pediatric brain tumor segmentation challenge 2023. Mach Learn Biomed Imaging. 2025;3:72-87. doi:10.59275/j.melba.2025-f6fg. https://doi.org/10.59275/j.melba.2025-f6fg
11. Sun X, He W, Ruan J, Yuan Z, Sun Z, Zhang J. An edge enhanced 3D Mamba U-Net for pediatric brain tumor segmentation with transfer learning. Med Phys. 2025;52(10):e70002. doi:10.1002/mp.70002. https://doi.org/10.1002/mp.70002 PMid:41046478
12. Cariola A, Sibilano E, Guerriero A, Bevilacqua V, Brunetti A. Deep learning strategies for semantic segmentation of pediatric brain tumors in multiparametric MRI. Sci Rep. 2025;15:22595. doi:10.1038/s41598-025-07257-2. https://doi.org/10.1038/s41598-025-07257-2 PMid:40596219 PMCid:PMC12218216
13. Yi Y, Zhuang Q, Xu ZQJ. Frequency-Aware Ensemble Learning for BraTS 2025 Pediatric Brain Tumor Segmentation. arXiv [Preprint]. 2025.

arXiv:2509.19353.
https://doi.org/10.1007/978-3-032-16365-3_40

14. Griffiths-King D, Mulvany T, Rose H, Novak J. Ratio maps of T1w/T2w MRI signal intensity do not improve deep-learning segmentation of pediatric brain tumors. PLoS One. 2025;20(12):e0323398. doi:10.1371/journal.pone.0323398. https://doi.org/10.1371/journal.pone.0323398 PMid:41428695 PMCid:PMC12721524

15. Ketabi S, Wagner MW, Hawkins C, Tabori U, Ertl-Wagner BB, Khalvati F. Multimodal contrastive learning for enhanced explainability in pediatric brain tumor molecular diagnosis. Sci Rep. 2025;15:10943. doi:10.1038/s41598-025-94806-4. https://doi.org/10.1038/s41598-025-94806-4 PMid:40159500 PMCid:PMC11955525

16. Mostafa M. BraTS-2023-PED-Dataset [dataset]. Kaggle; 2024 [cited 2026 Aug 21]. Available from: https://www.kaggle.com/datasets/mahamostafa/brats-2023-ped-dataset.

17. Santosh DT, Anuradha N, Kolukuluri M, Gupta G, Pathak MK, Gokula Krishnan V, Raghuvanshi A. Development of IoT based intelligent irrigation system using particle swarm optimization and XGBoost techniques. Bull Electr Eng Inform. 2024;13(3):1927-1934. doi:10.11591/eei.v13i3.6332. https://doi.org/10.11591/eei.v13i3.6332