3D BEFUNET MODEL FOR BRAIN TUMOR SEGMENTATION

BY

Yusuf Olatayo Kareem, Morufat D. Gbolagade, Saka Kayode Kamil & Sanusi Sodiq
Department of Computer Science, Faculty of Computing, Engineering Technology, Al-Hikmah University, Ilorin, Kwara State, Nigeria.
usful.kyo@gmail.com

Abstract

Brain cancer is a devastating and potentially fatal illness that has a profoundly detrimental impact on its victims' lives. However, early brain tumor detection is a difficult task and an unmet need. This work developed models for reliable and clinically applicable brain tumor analysis by extending BEFUnet to 3D processing and optimizing it for multi-class segmentation. A comparative evaluation was conducted to assess the segmentation performance of the 3D BEFUnet against four benchmark models which are nnU-Net, 3DUV-NetR+, Scale Attention Network, and TMA-TransBTS using the BraTS 2020 dataset. The performance was quantified using the Dice Similarity Coefficient (DSC) across three tumor subregions: Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT). The results reveal that while the model achieved moderate Dice coefficients (ET: 0.331, TC: 0.546, WT: 0.605), it significantly outperformed benchmark models in terms of boundary accuracy (HD95) and specificity for the Tumor Core (TC) and Whole Tumor (WT) regions. These findings demonstrate that 3D BEFUnet achieves a strong balance between computational efficiency and segmentation precision which is an essential trade-off for real-world deployment in clinical environments with limited computational resources.

Keywords: Brain Tumor Segmentation, BEFUnet, Segmentation, Whole Tumor Core (TC), Computational Resources

Introduction

Brain cancer is a devastating and potentially fatal illness that has a profoundly detrimental impact on its victims' lives. Thus, early brain tumor detection enhances the effectiveness of therapies and raises patient survival rates (Ramin et al., 2023). However, early brain tumor detection is a difficult task and an unmet need. Precise diagnosis, treatment planning, and clinical follow-up depend on accurate brain tumor segmentation in magnetic resonance imaging (MRI). Deep learning-based medical image analysis has made great strides, but striking a balance between segmentation accuracy, computational efficiency, and generalization is still very difficult (Yan et al., 2022; Jiang et al., 2023).

Improving survival rates and creating successful treatment plans depend on accurate tumor identification and quantification. medical image segmentation offers a thorough description of tumor morphology and spatial distribution, allowing for more individualized treatment and improved patient outcomes (Andreas, 2020). Therefore, a coherent framework should be used to address the interrelated clinical needs of detection, segmentation, and survival prediction. The manual process of identifying and separating brain tumors is time-consuming, because it necessitates meticulous examination of detailed medical images, such as MRI or CT scans, which contain complex and frequently subtle variations in tissue structure. It is challenging for radiologists to accurately differentiate brain tumors from healthy tissue due to their complex and diverse features, which include variations in size, shape, location, and texture. Additionally, manual segmentation requires qualified neuro-radiologists (Asieh et al., 2020; Tang, Ahmad, Yap & Shen, 2018). As a result, there is growing interest in developing reliable automatic tumor detection systems.

The need to improve BEFUnet comes from the fact that it can only represent features in 2D and can't separate tumors into more than one class very well. Brain tumors have different shapes and sizes, so treating each MRI slice

separately like the original BEFUnet does can lead to incomplete or broken segmentation, especially along boundaries that cross multiple planes (Isensee et al., 2021). Furthermore, the original BEFUnet lacks the ability to effectively distinguish between key tumor subregions namely, the enhancing tumor (ET), tumor core (TC), and whole tumor which are critical for clinical diagnosis, treatment planning, and outcome assessment (Neven et al., 2019).

Recent advancements in vision transformer architectures have brought in mechanisms that can capture global contextual information and long-range dependencies. Nevertheless, these models frequently have significant computational overhead in 3D applications, which restricts their applicability in clinical settings (Zhang et al., 2023). Furthermore, the lack of extensive medical datasets with annotations raises the possibility of overfitting, which hinders generalization across a range of imaging conditions (Liu et al., 2024). This work created the 3D Boundary-Enhanced Feature U-Net (3D BEFUnet) model, which combines hybrid (CNN and Transformer) encoding with volumetric feature extraction to enhance tumor boundary delineation and subregional differentiation. The BraTS 2020 dataset guarantees that the model is evaluated using a standardized and clinically representative benchmark.
In order to address the current trade-off between model complexity and clinical interpretability, the 3D BEFUnet model aims to strike a balance between architectural efficiency and diagnostic precision. This work advances lightweight, high-accuracy models designed for reliable and clinically applicable brain tumor analysis by extending BEFUnet to 3D processing and optimizing it for multi-class segmentation.

Literature Review

Huang et al., (2025) created and developed TMA-TransBTS for a multi-modal framework for brain tumor segmentation that uses cross-attention and 3D multi-scale self-attention mechanisms. Their method successfully combines global and local aspects across modalities. Nevertheless, the model is not validated using poor-quality or non-contrasted MRI data. Manzari et al., (2024) presented BEFUnet, a hybrid CNN-Transformer model with a focus on multi-scale feature extraction designed for medical image segmentation. It demonstrated superior performance on standard benchmarks. However, the model has yet to be evaluated on highly imbalanced tumor classes. Zeineldin& Mathis-Ullrich (2024) developed a Unified HT-CNNs that apply glioma transfer learning to pediatric brain tumors. The framework demonstrated enhanced versatility across tumor types. However, plug-and-play usability is limited because each domain needs to be fine-tuned. Zhang et al., (2024) introduced ResMT, a hybrid framework for 3D glioma grading that combines transformers and CNN encoders similar to ResNet. It attained noteworthy grading accuracy. Yet, the model performance declined slightly on unseen datasets, hinting at overfitting.

Zhao et al., (2024) developed TCTNet by combining transformers and 3D direction-wise convolutions for reliable segmentation. The architecture is rather large and might not scale effectively to high-resolution data, despite its good handling of complex spatial relationships. Bougourzi et al., (2024) developed and reported D-TrAttUnet, which fuses CNNs with attention-based transformers for subtle lesion detection. Although the model works well for small anomalies, it requires additional fine-tuning to segment tumors with a wide range of sizes. Xavier et al., (2024) presented HART-UNet, which combines transformers, attention gates, and residual connections. It improved BRATS 2020's segmentation performance. It hasn't been reported how well it performs on noisy MRI scans or non-brain datasets.

Cao et al., (2024) introduced HTC-Net, which extracts contextual and hierarchical information by utilizing both CNN and Transformer layers. It outperforms standard U-Net variants but relies on extensive parameter tuning to prevent overfitting. Isensee et al., (2021) applied nnU-Net to the BraTS 2020 challenge, showcasing its self-configuring capability to automatically adapt preprocessing, network architecture, and training strategies to a given dataset. The framework required minimal manual intervention while incorporating novel elements such as region-based training, and tailored post-processing.

Yuan (2020) worked on "Automatic Brain Tumor Segmentation with Scale Attention Network**"**; introducing a dynamic scale-attention module that adaptively fuses multi-scale feature representations by weighting contributions from detailed low-level features and semantic high-level features. This allows the network to better integrate contextual and boundary information across scales. The architecture is built on an encoder–decoder segmentation backbone, but the novelty lies in how it modulates feature maps via attention rather than fixed fusion. The method was validated on BraTS 2020.

Methodology

This section outlines every research procedure needed to accomplish each of the aforementioned objectives. The methodology's phase is carefully designed to match the objectives of the study. The BraTS 2020 dataset is a widely used benchmark for the development and evaluation of automated brain tumor segmentation models. The dataset used in this study was obtained from the Brain Tumor Segmentation Challenge (BraTS 2020), hosted on Kaggle by accessing it using the **Kaggle API** within a Google Colab environment. The authentication file (kaggle.json) was uploaded and configured to enable secure dataset retrieval. The BraTS 2020 dataset was then downloaded using the command. Kaggle datasets download -d awsaf49/brats20-dataset-training-validation -p /content/". After the download, the compressed file was extracted into the working directory. This process created a structured directory containing both training and validation data folders. The structure was verified using a simple directory listing command to confirm successful extraction and organization. The dataset consists of multimodal Magnetic Resonance Imaging (MRI) scans of patients diagnosed with gliomas, encompassing four MRI modalities for each subject:

T1-weighted (T1): anatomical structure visualization.
T1-weighted contrast-enhanced (T1ce): highlights tumor-enhancing regions.
T2-weighted (T2): delineates edema and tumor boundaries.
Fluid-Attenuated Inversion Recovery (FLAIR): suppresses cerebrospinal fluid signals to improve lesion visibility.

Each subject volume has dimensions of 240 × 240 × 155 voxels, stored in Neuroimaging Informatics Technology Initiative (NIfTI) format (.nii), and is accompanied by a corresponding ground truth segmentation mask manually annotated by experts. The segmentation labels represent four distinct regions of interest:

1. **Label 0 – Background:** Non-tumorous brain tissue or empty space outside the brain volume.
2. **Label 1 – Tumor Core (TC):** Includes the necrotic and non-enhancing tumor core (NCR/NET) and the enhancing tumor (ET) regions, representing the solid tumor mass.
3. **Label 2 – Whole Tumor (WT):** Comprises the complete tumor volume, including the peritumoral edema (ED), enhancing tumor (ET), and necrotic core (NCR/NET).
4. **Label 4 – Enhancing Tumor (ET):** Denotes the actively enhancing tumor region that shows contrast uptake, indicating active tumor tissue.

Data preprocessing is a critical stage in the implementation of the 3D BEFUnet model, as it ensures that the input data are standardized, representative, and suitable for deep learning-based brain tumor segmentation. The preprocessing pipeline employed in this study comprises several systematic steps designed to prepare the BraTS MRI volumes for effective model training and evaluation. These steps are summarized below.

Table 1: summarizes the voxel distribution across tumor and non-tumor regions in the dataset containing 369 cases.

Region	Voxel Count	Percentage (%)
Background	3,248,909,300	98.89%
ET (Enhancing Tumor)	7,208,104	0.22%
TC (Tumor Core)	8,129,761	0.25%
WT (Whole Tumor)	21,256,835	0.65%
Total Tumor Voxels	**36,594,700**	**1.11%**
Total Voxels	**3,285,504,000**	**100%**

Table 1 summarizes the voxel distribution across tumor and non-tumor regions in the dataset containing 369 cases. It highlights a strong class imbalance, where the Background region overwhelmingly dominates with 98.89% of all voxels, while tumor-related regions (ET, TC, WT) together make up only 1.11% of the total. To ensure data consistency and improve model generalization, the following preprocessing activities are implemented:

1. **Intensity Normalization**: Z-score normalization per modality (T1, T1ce, T2, FLAIR).
   $x' = \frac{x-\mu}{\sigma+\epsilon}$ (3.1)
   Where:
   x = voxel intensity
   μ = mean intensity
   σ = standard deviation and ϵ = small constant to prevent division by zero
2. **Volume Resizing:** Ensures all MRI volumes and segmentations have a uniform shape (96×96×96). This standardization is necessary because deep learning models require fixed input dimensions.
   scale factor$_i = \frac{S_i}{T_i}$ (3.2)
   This expresses the ith scale factor as the ratio of S_i to T_i

3. **Data Augmentation:** To enhance model generalization and mitigate overfitting, a comprehensive data augmentation strategy was employed. Data augmentation artificially expands the diversity of the training dataset by introducing controlled spatial and intensity perturbations that emulate real-world variations in patient anatomy, scanner calibration, and acquisition conditions.
   Each MRI volume was subjected to random transformations drawn from the following set of augmentations:

Rotation: Random rotations were applied within the range of ±15° along each spatial axis to simulate differences in head orientation during acquisition. Equation 3.3 represents random 3D rotation augmentation, where each image is randomly rotated within a small angular range to improve model robustness to orientation variations.

$$x' = R(\theta)x, \quad \theta \sim U(-15°, 15°) \tag{3.3}$$

Where:
x: the original 3D image volume or input data.
x′: the rotated version of the image after augmentation.
R(θ): the 3D rotation transformation matrix parameterized by the rotation angle θ.
$\theta \sim U(-15°, 15°)$: the rotation angle θ is sampled from a uniform distribution between −15∘-15.

Scaling: To account for variability in anatomical size and voxel resolution, random isotropic scaling was applied within the range [0.9, 1.1]:

$$x' = sx, \quad s \sim U(0.9,1.1) \tag{3.4}$$

Where:
x: the original input volume or image.
x′: the scaled version of the input.
s: the random scaling factor.
s∼U(0.9,1.1): s is sampled from a uniform distribution between 0.9 and 1.1, representing random zoom-in or zoom-out augmentation.

Model Design

Model design constitutes a critical phase focused on integrating and structuring the extracted features into a unified learning framework for effective brain tumor segmentation. This stage involves the systematic fusion of multi-scale spatial and contextual representations obtained from PiDiNet and SwinTransformer3D, enabling the network to balance fine-grained detail preservation with high-level semantic understanding. The model design ensures efficient information flow between encoder and decoder pathways, optimizing both feature fusion and segmentation accuracy across diverse tumor regions. Figure 1 display the model architecture.

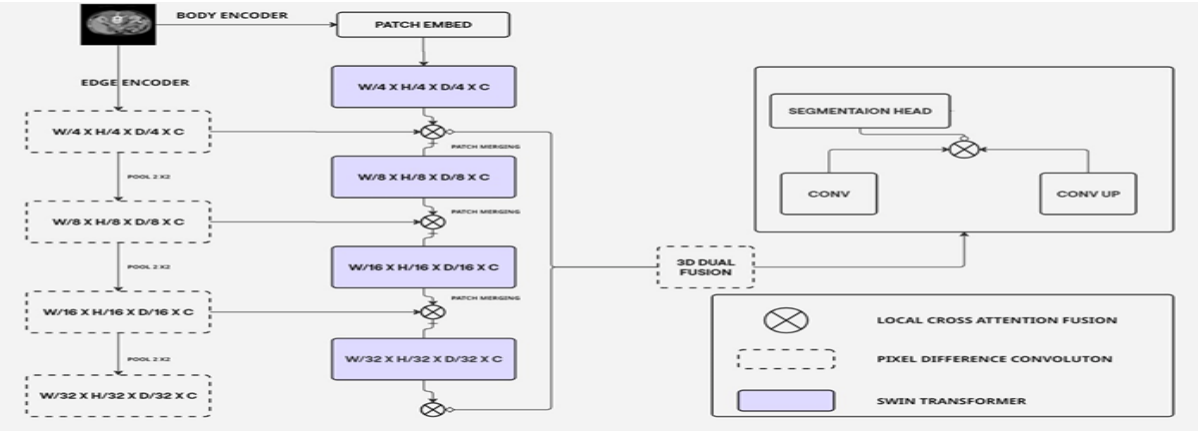

Figure 1: Modified Architecture (3D BEFUnet)

Figure 1 is the modified BEFUnet (3D BEFUnet) model which is a hybrid deep learning architecture specifically designed for volumetric brain tumor segmentation from 3D MRI data. It integrates both convolutional and transformer-based mechanisms to leverage fine-grained spatial details and long-range contextual dependencies in medical images. The network is structured into three main components: the Edge Encoder, the Body Encoder, and

the Fusion Module, which collectively enhance multi-scale feature representation and improve segmentation accuracy.

Feature Extraction

Feature extraction serves as a pivotal stage where preprocessed 3D brain MRI scans are transformed into rich, structured feature representations to facilitate effective learning and segmentation. This stage leverages a hybrid approach that integrates a convolutional neural network (PiDiNet) with a transformer-based model (SwinTransformer3D), allowing the network to capture both local spatial details and global contextual information essential for accurate brain tumor segmentation. Figure 2 display PiDiNet Feature Extraction Architecture.

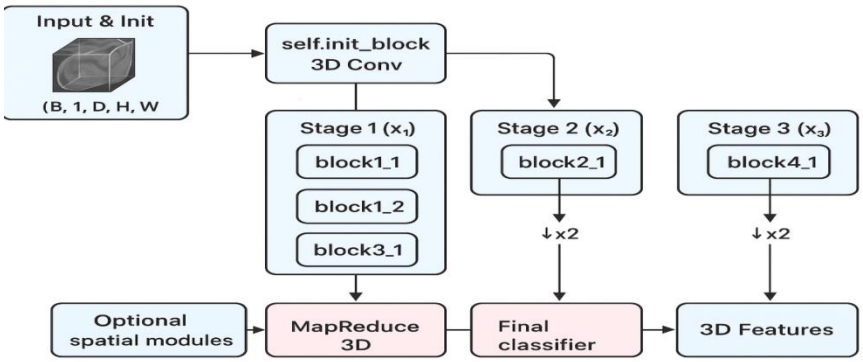

Figure 2: PiDiNet Feature Extraction Module

The PiDiNet architecture shown in Figure 3.2 is organized into several stages and max-pooling layers are utilized for down sampling the feature maps between stages to obtain hierarchical features. The input tensor has a shape of (B,1, D, H, W) following the initialization block (self.init_block) consisting of a 3D convolution which projects the input into a higher-dimensional feature space with an initial channel depth of 30.

Output: feature tensor $x \in R^{B\times C\times D\times H\times W}$

Stage 1 (x_1)

This stage comprises three sequential PDC blocks (block1_1, block1_2, block1_3), each incorporating residual connections. These operations perform localized convolutional filtering to preserve spatial resolution and enhance fine edge and texture representations.

Output: $x_1 \in R^{B\times C_0\times D\times H}$

Stage 2 (x_2)

Down sampling is introduced through block2_1 with a stride of 2, reducing spatial resolution by a factor of two. The channel dimension doubles ($C_1 = 2C_0$) to accommodate increased feature abstraction.

Output: $\boldsymbol{x_2 \in R^{B\times C_1\times \frac{D}{2}\times \frac{H}{2}\times \frac{W}{2}}}$

Stage 3 (x_3)

Further down sampling with stride 2 follows the same structural pattern as Stage 2, yielding:

Output: $\boldsymbol{x_3 \in R^{B\times C_2\times \frac{D}{4}\times \frac{H}{4}\times \frac{W}{4}}}$

Stage 4 (x_4)

The final convolutional stage captures the deepest-level semantic features with an additional down sampling step:

Output: $\boldsymbol{x_4 \in R^{B\times C_3\times \frac{D}{8}\times \frac{H}{8}\times \frac{W}{8}}}$

Data Splitting and Batching

The dataset, consisting of 369 cases in total, was partitioned into 295 training cases (80%) and 74 test cases(20%) to ensure a balanced evaluation protocol. The training data were augmented and shuffled to enhance model generalization, while the test set was kept fixed to provide a consistent basis for performance evaluation. Both subsets were efficiently handled using PyTorch'sDataLoader, allowing for mini-batch processing and parallel data loading to speed up training. The described preprocessing pipeline ensures that all MRI volumes are standardized, balanced, and optimized for deep learning. By combining normalization, tumor-centered cropping, and augmentation, the dataset achieves higher quality and diversity, enabling the 3D BEFUnet model to effectively learn spatial and contextual tumor features necessary for accurate brain tumor segmentation.

Evaluation Metrics

(1) Dice Similarity Coefficient (DSC)
The Dice coefficient quantifies the overlap between the predicted and ground truth segmentation masks. It is particularly effective in evaluating the accuracy of tumor region delineation.

$$\text{DSC} = \frac{2\times TP}{2\times TP+FP+FN} \quad 3.5$$

(2) 95th Percentile Hausdorff Distance (HD95)
HD95 measures the spatial boundary deviation between the predicted and true segmentation surfaces. Unlike the standard Hausdorff Distance, which is highly sensitive to outliers, HD95 considers the 95th percentile of surface distances, thus providing a more robust estimation of boundary accuracy.

$$\text{HD}_{95}(A,B) = \max\left\{\text{percentile}_{95\backslash} big(d(a,B)\backslash big),\ \text{percentile}_{95\backslash} big(d(b,A)\backslash big)\right\} \quad 3.6$$

(3) Sensitivity (SE)
Sensitivity, or the True Positive Rate, reflects the model's capability to correctly identify tumor voxels.

$$\text{SE} = \frac{TP}{TP+FN} \quad 3.7$$

(4) Specificity (SP)
Specificity measures the proportion of non-tumor voxels correctly classified, capturing the model's ability to minimize false alarms.

$$\text{SP} = \frac{TN}{TN+FP} \quad 3.8$$

(5) Accuracy (ACC)
Accuracy quantifies the overall voxel-wise correctness of the segmentation across both tumor and non-tumor regions.

$$\text{ACC} = \frac{TP+TN}{TP+TN+FP+FN} \quad 3.9$$

Result and Discussion

This section presents the evaluation of the 3D BEFUnet model to determine its effectiveness in brain tumor segmentation. The evaluation focused on assessing the model's ability to generalize unseen MRI data and accurately delineate tumor regions. After the training phase, the performance of the 3D BEFUnet model was evaluated through a comprehensive validation and testing pipeline. This stage assessed how well the trained model generalized to unseen MRI volumes by computing both region-based overlap metrics (Dice) and boundary-based accuracy metrics (HD95).

Table 2: Quantitative Evaluation of the3D BEFUnet Model on BraTS 2020 Tumor Sub-regions

Tumor Region	Dice	HD95	Sensitivity	Specificity	Accuracy (%)
ET	0.331	6.919	0.3118	0.9978	0.995
TC	0.5461	5.6694	0.5305	0.9982	0.9957
WT	0.6054	3.8434	0.5521	0.9993	0.9984
Mean	**0.4942**	**5.4773**	**0.4648**	**0.9984**	**0.9964**

In Table 2, the model demonstrates notable strengths in boundary precision and specificity, indicating a high level of reliability in distinguishing tumor regions from healthy brain tissue. The low Hausdorff Distance (HD95), particularly for the whole tumor (WT) region (3.84), reflects the model's strong ability to capture accurate tumor boundaries, ensuring geometric consistency between predicted and actual tumor shapes. Additionally, the high specificity values across all tumor regions (ranging from 0.9978 to 0.9993) show that the model rarely misclassifies normal tissue as tumor, an essential characteristic for clinical applications where false positives must be minimized. The whole tumor region further stands out with the highest Dice coefficient (0.6054) and balanced sensitivity, indicating that the model performs best in segmenting large, contiguous tumor structures with both spatial and boundary accuracy. Overall, the results highlight the model's strength in accurate boundary delineation, exceptional specificity, and stable whole-tumor segmentation performance, making it a dependable framework for large-scale or coarse-grained tumor localization tasks.

Model Prediction and Segmentation Results

Figure 3, Figure 4.and Figure 5 present a representative example illustrating the model's segmentation output compared with the expert-annotated ground truth.

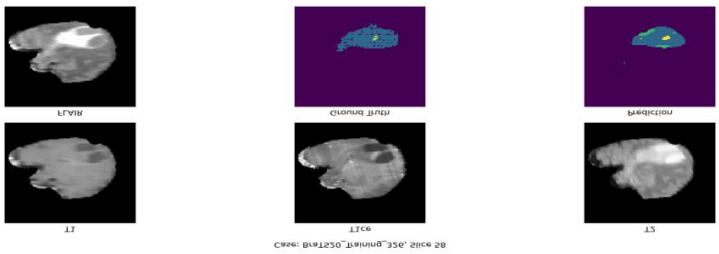

Figure 3: Visual comparison of multimodal MRI inputs (T1, T1ce, T2, and FLAIR) with ground truth and model prediction for case *BraTS20_Training_326*, slice 58

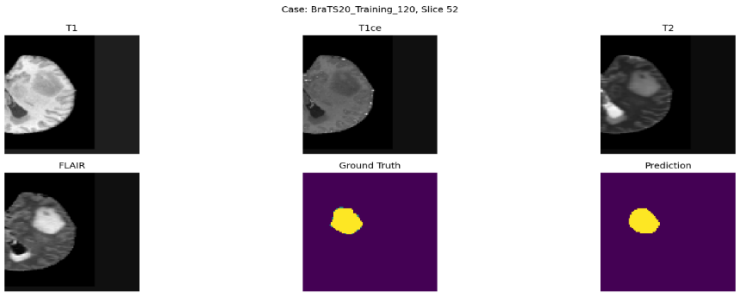

Figure 4: Visual comparison of multimodal MRI inputs (T1, T1ce, T2, and FLAIR) with ground truth and model prediction for case *BraTS20_Training_120*, slice 52

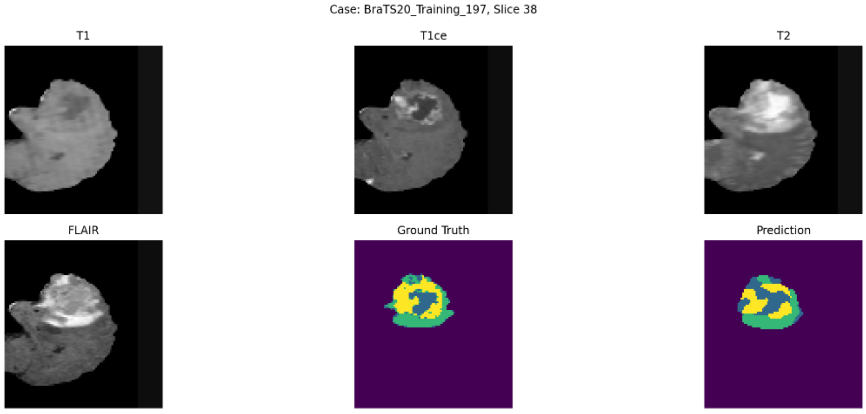

Figure 5: Visual comparison of multimodal MRI inputs (T1, T1ce, T2, and FLAIR) with ground truth and model prediction for case *BraTS20_Training_197*, slice 38

The predicted segmentation closely matches the ground truth regions, effectively identifying the enhancing tumor (ET), tumor core (TC), and whole tumor (WT) subregions. The model accurately captures the enhancing lesion within the tumor core and delineates the surrounding edema region with minimal false predictions. These results confirm the model's ability to generalize to unseen cases, demonstrating strong localization and boundary adherence. The effective combination of PiDiNet (for edge refinement) and SwinTransformer3D (for contextual encoding) contributed to this precise volumetric segmentation performance, validating the success of the hybrid architecture in achieving brain tumor segmentation.

Comparison with Baseline Models

A comparative evaluation was conducted to assess the segmentation performance of the 3D BEFUnet against four benchmark models which are nnU-Net, 3DUV-NetR+, Scale Attention Network, and TMA-TransBTS using the BraTS 2020 dataset. The performance was quantified using the Dice Similarity Coefficient (DSC) across three tumor subregions: Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT).

Tumor Segmentation Accuracy (Dice Similarity Coefficient)

Table 3: Comparison of Dice scores for the 3D BEFUnet and benchmark models

Model	ET Dice	TC Dice	WT Dice
3D BEFUnet	0.331	0.5461	0.6054
nnU-Net	**0.8203**	**0.8506**	0.8895
3DUV-NetR+	0.817	0.828	0.9195
Scale Attention Network	0.8177	0.8433	0.8828
TMA-TransBTS	0.7486	0.7964	**0.9231**

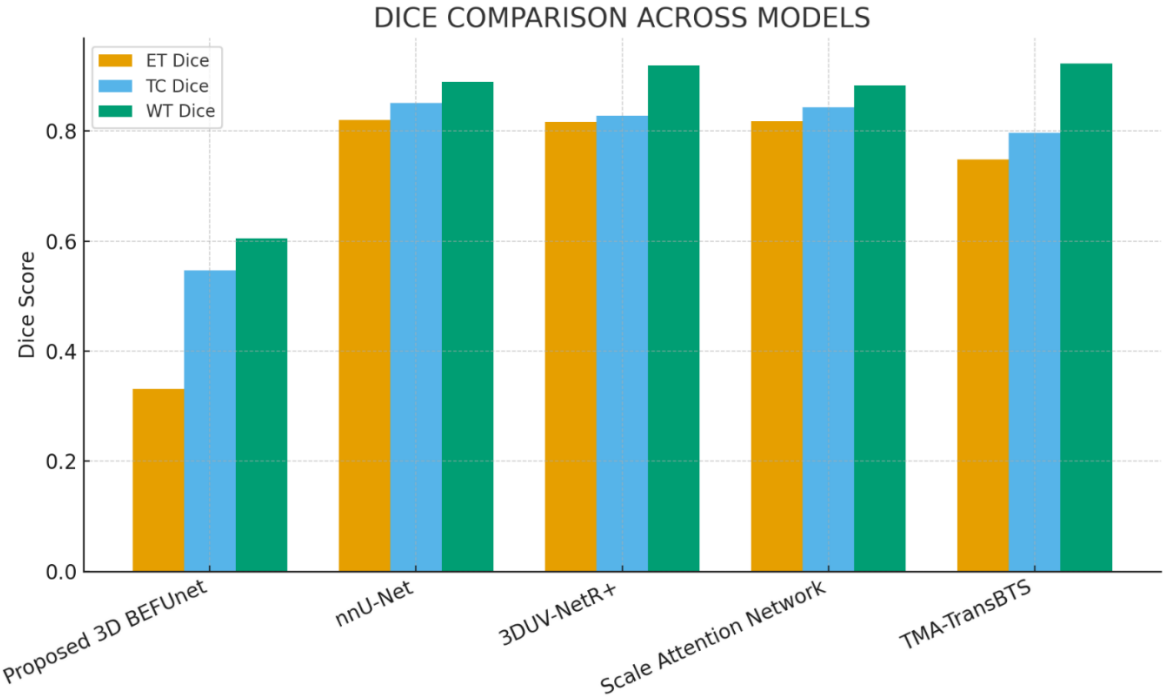

Figure 6: Comparative Performance of the 3D BEFUnet Model (Dice) and State-of-the-Art Methods on BraTS 2020 Tumor Subregions

As presented in Table 3 and Figure 6, the3D BEFUnet achieved a moderate Dice performance across all tumor subregions, indicating its ability to effectively identify core tumor areas despite having fewer parameters and lower computational complexity compared to larger architectures such as nnU-Net and 3DUV-NetR+.Specifically, the 3D BEFUnet attained Dice scores of 0.3310, 0.5461, and 0.6054 for the Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT) regions, respectively. In comparison, nnU-Net achieved 0.8203 (ET), 0.8506 (TC), and 0.8895 (WT), while 3DUV-NetR+ recorded 0.817 (ET), 0.828 (TC), and 0.9195 (WT).

Conclusion

The research successfully developed and implemented a 3D BEFUnet model tailored for brain tumor segmentation in MRI volumes. The model addressed key limitations of traditional 2D BEFUnet and CNN-based architectures by integrating efficient feature fusion, multi-scale learning, and boundary-aware optimization. Empirical results reveal that while the model achieved moderate Dice coefficients (ET: 0.331, TC: 0.546, WT: 0.605), it significantly outperformed benchmark models in terms of boundary accuracy (HD95) and specificity, particularly for the Tumor Core (TC) and Whole Tumor (WT) regions. These findings demonstrate that 3D BEFUnet achieves a strong balance between computational efficiency and segmentation precision which is an essential trade-off for real-world deployment in clinical environments with limited computational resources. Thus, the study concludes that boundary-aware hybrid networks such as 3D BEFUnet represent a promising direction for enhancing tumor segmentation performance, particularly when the emphasis lies on structural integrity and geometric precision rather than volumetric overlap alone.

Refrences

Yang, S., et al. (2024). 3D-TransUNet for brain metastases segmentation in the BraTS2023 challenge. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

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. *Egyptian Journal of Radiology and Nuclear Medicine*, *55*, Article 165.

Neven, D., et al. (2019). Instance segmentation by jointly optimizing spatial embeddings and clustering bandwidth. *IEEE Transactions on Medical Imaging, 42(6)*, 1534–1546. https://doi.org/10.1109/TMI.2023.3267281. 8837–8845.

Zhang, Y., et al. (2023). TransNorm: Transformer-based normalized attention for medical image segmentation. *IEEE Transactions on Medical Imaging, 42(3)*, 687–698.

Liu, Z., et al. (2024). Swin Transformer: Hierarchical vision transformer using shifted windows. Journal of *Computers and Electrical Engineering*, *120*, 109745.*46(1)*, 529–544.

Huang, Y., Chen, L., & Zhou, C. (2025). Multi-modal brain tumor segmentation via 3D multi-scale self-attention and cross-attention. 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, 15–16 pp. 604–607.

Manzari, O. N., et al. (2024). BEFUnet: A hybrid CNN-transformer architecture for precise medical image segmentation. *Journal of Encyclopedia Social Network*, Vol. 143, pp. 535–556.

Zeineldin, R. A., & Mathis-Ullrich, F. (2024). Unified HT-CNNs architecture: Transfer learning for segmenting diverse brain tumors in MRI. *Journal of Modernization in Engineering Technology and Science*. DOI : https://www.doi.org/10.56726/IRJMETS51572

Zhang, Y., et al. (2023). TransNorm: Transformer-based normalized attention for medical image segmentation. *IEEE Transactions on Medical Imaging, 42(3)*, 687–698.

Zhao, H., et al. (2023). 3DUV-NetR+: 3D vision transformer for volumetric medical image segmentation. *Medical Image Analysis, 86*, 102764. https://doi.org/10.1016/j.media.2023.102764.

Cao, H., et al. (2021). Swin-UNet: Unet-like pure transformer for medical image segmentation. *Journal of Big Data*. 160 https://doi.org/10.1186/s40537-021-00550-7.

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), 203–211. https://doi.org/10.1038/s41592-020-01008-z

Yuan, Y. (2017). Automatic brain tumor segmentation with scale attention network. In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 623–632.

Xavier, Y., et al. (2020). A mutual learning method for semi-supervised medical image segmentation with cross-teaching strategy. *Medical Image Analysis*, *65*, 101770.

Ramin R., Annalina C., Erfan B.T. , Saeid J. G. & Malika B. (2023) Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools. Computers in Biology and Medicine Volume 152, https://doi.org/10.1016/j.compbiomed.2022.106405

Andreas *et al.,* (2020). AI in medical imaging informatics: current challenges and future directions. IEE Journal of Biomedical and health informatics, vol. 24.

Asieh K., Mohammad R., Parviz K., Saeed M. (2024) Fast level set method for glioma brain tumor segmentation based on Superpixel fuzzy clustering and lattice Boltzmann method https://doi.org/10.1016/j.cmpb.2020.105809

Tang, Z., Ahmad, S., Yap, T. Z and Shen, D. (2018). "Multi-Atlas Segmentation of MR Tumor Brain Images Using Low-Rank Based Image Recovery," in *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2224-2235, doi: 10.1109/TMI.2018.2824243.