



ARTICLE

MambaFNO-NET: A Dual-Domain Framework Integrating State Space Models and Fourier Neural Operators for Brain Tumor Segmentation

Ronak Patel¹, Miral Patel², Deep Kothadiya³, Noor A. Khan⁴, Shaha Al-Otaibi^{5,*},
Roaa Khalil Mohamed Ali Abed⁶ and Tanzila Saba⁷

¹U & P U. Patel Department of Computer Engineering, Chandubhai S. Patel Institute of Technology (CSPIT), Faculty of Technology (FTE), Charotar University of Science and Technology (CHARUSAT), Changa, India

²G H Patel College of Engineering and Technology, CVM University, V.V. Nagar, Anand, Gujarat, India

³Symbiosis Centre for Information Technology, Symbiosis International (Deemed University), Pune, India

⁴Center of Excellence in Cyber Security (CYBEX), Prince Sultan University, Riyadh, Saudi Arabia

⁵Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh, Saudi Arabia

⁶College of Sciences and Humanities (CSH), Prince Sultan University, Riyadh, Saudi Arabia

⁷AIDA Lab. CCIS, Prince Sultan University, Riyadh, Saudi Arabia

*Corresponding Author: Shaha Al-Otaibi. Email: stalotaibi@pnu.edu.sa

Received: 15 February 2026; Accepted: 15 April 2026; Published: 27 May 2026

ABSTRACT: Magnetic resonance imaging (MRI) is widely utilized for brain tumor segmentation, yet significant challenges persist due to intensity variations, irregular boundaries, and substantial morphological heterogeneity. Current state-of-the-art deep learning methods often struggle to capture long-range spatial dependencies, delineate fine boundary details, and efficiently process 3D volumetric data. This study introduces a novel hybrid framework that integrates state-space models with frequency-domain learning to address these limitations. The proposed model offers four primary contributions: (1) incorporation of a morphological attention block in the encoder to enhance boundary localization via dilation-erosion gradient modeling; (2) a dual-domain bottleneck module that combines Mamba-inspired sequential modeling with the Fourier Neural Operator (FNO) for efficient local and global pattern modeling with linear complexity; (3) a Feature Pyramid Network (FPN) augmented with Feature-Guided Learning (FGL) for adaptive multi-scale semantic fusion; and (4) Laplacian Pyramid decomposition to preserve high-frequency edge details. The model demonstrates state-of-the-art performance, achieving Dice Similarity Coefficients of 0.81 ± 0.05 , 0.92 ± 0.02 , and 0.86 ± 0.03 for Enhancing Tumor (ET), Whole Tumor (WT), and Tumor Core (TC) on the BraTS2020 dataset, respectively, with a mean Dice of 0.86 ± 0.05 . MambaFNO-NET attains Hausdorff distances (HD95) of 4.21, 6.22, and 6.85 mm for ET, WT, and TC, respectively, resulting in a mean HD95 of 5.76 mm, which underscores its superior boundary localization accuracy. Overall, MambaFNO-NET delivers an efficient and accurate solution for 3D brain tumor segmentation, balancing volumetric precision with spatial boundary alignment for practical clinical deployment.

KEYWORDS: Mamba; fourier neural network; feature pyramid network; Laplacian pyramid; BraTS2020; healthcare

1 Introduction

According to the Central Brain Tumor Registry of the United States (CBTRUS) Statistical Report published in 2025, 107,100 new cases are projected to be diagnosed with brain issues, with an average age of 26.05 per 100,000 cases. Of these, 17,637 deaths were attributed to malignant tumors (cancerous brain

growths), a rate of 4.41 per 100,000. The survival rate over the past 5 years was 34.8% [1]. Abnormal growth of brain tissue, which can significantly impact well-being, underscores the importance of early detection for survival. Despite advances, distinguishing between gliomas (tumors arising from glial cells in the brain), meningiomas (tumors arising from the membranes covering the brain and spinal cord), and pituitary tumors (tumors arising from the pituitary gland) remains challenging. Research has focused more on gliomas due to their aggressiveness and complex treatment needs [2]. Magnetic resonance imaging (MRI) is commonly used to evaluate these tumors. However, diagnosis can be complicated by the affected area's complex appearance, variable size and shape, intensity differences, and blurred boundaries and locations [3]. As a result, automatic segmentation solutions are increasingly important for improving the accuracy of clinical diagnosis and treatment planning [4].

Over the past decade, remarkable progress has been made through the Brain Tumor Segmentation (BraTS) challenge, organized by MICCAI each year [5]. Many hybrids segmentation approaches have used the U-Net [6] as a baseline architecture. Attention-based approaches, when aligned with the U-Net architecture, suppress irrelevant background information and highlight important features. Attention-UNet [7] integrates an attention mechanism on the decoder side as an excitation module. A ResU-Net [8] strengthens the U-Net using residual convolutional blocks and an attention gate for 2D tumor segmentation. MAU-Net [9] introduces spatial attention, helping focus on important feature locations. 3D AIR-UNet [10] adds attention not only to the encoder but also in the bottleneck. Integrating attention blocks thus helps automatically learn and suppress unwanted information when the dataset is imbalanced.

In recent years, many studies have examined attentive segmentation using transformers. TansBTS [11], SwinUnet [12], and VGX [13] are advanced architectures focused on long-range dependencies. In some aspects, different FPN-based approaches are used. For example, ResAtt-NASFPN [14] helps address variation issues in tumors, which are often very small regions. SSMs (state-space models) have reduced transformer complexity for long sequences. A recent study showed BraTS-UMamba [15] handled 3D MRI volumes faster than transformers and achieved superior TC identification. Integrating attention, transformers, and FPN at different levels still outperforms standard segmentation, prioritizing features and capturing global context.

Research Motivation and Contributions. This article presents MambaFNO-NET, a hybrid deep learning framework that integrates state-space models with frequency-domain learning. Experimental results demonstrate that this approach outperforms state-of-the-art segmentation networks.

Research Motivation and Contributions. In this article, we propose MambaFNO-NET, a hybrid deep learning framework that combines state-space models and frequency domain learning. The experiments show the superiority of our approach compared to the state-of-the-art segmentation networks.

Our main findings are summarized as follows:

- Proposed a feature-guided architecture that combines Mamba and Fourier Neural Operators (FNO) in the bottleneck to effectively learn long-range dependencies and global context.
- Combining Morphological Attention (MA) and Laplacian Pyramids (LP) in the encoder-decoder architecture to improve structural feature learning and retain the fine details of edges.
- Simulation results for MambaFNO-NET show superiority over state-of-the-art deep learning models on the BraTS2020 benchmark dataset.

The rest of this article is structured as follows. [Section 2](#) presents a literature review, discussing segmentation methods and the challenges of accurate tumor segmentation. [Section 3](#) then introduces the methodology, detailing each architectural component and its design for BraTS dataset evaluation. Building on this, [Section 4](#) explains the experimental setup, covering dataset details, training and validation processes, and evaluation metrics. [Section 5](#) analyzes the results by comparing our method with existing

ones. [Section 6](#) explores the ablation study and provides visual improvements for the components. Finally, the discussion and conclusion summarize the study's findings and contributions to overcoming brain tumor segmentation challenges.

2 Literature Analysis

2.1 Encoder-Decoder Architecture

Encoder-decoder architectures dominate brain tumor segmentation. U-Net's [6] symmetric design captures multi-scale context using contracting encoders, such as ResNet [16] as backbones, and expansive decoders with skip connections for localization on BraTS [17] MRI modalities (T1, T1c, T2, FLAIR). Variants like nnU-Net [18] and hybrid CNN (Convolution Neural Network)-Transformer models such as EfficientNet-Swin [19] aggregate local details and global semantics. These models achieve Dice scores of 91%–94% for whole-tumor/core regions through feature upsampling and fusion. They handle class imbalance with weighted losses but require extensive hyperparameter tuning for 3D volumes. Many are pre-trained on ImageNet [20] and fine-tuned on BraTS. Lairedj et al. [21] proposed a 3D U-Net using a Gaussian mixture model based on intensity preprocessing to identify low-contrast MRI regions.

Challenges

An encoder-decoder-based framework requires significant computational power, limiting its use in real-time applications. Additionally, Tumor heterogeneity in the encoder network causes boundary blurring, leading to misalignment between modalities.

2.2 Feature Pyramid Network

Feature Pyramid Networks (FPN) [22] use top-down pathways and lateral connections in U-Net decoders. This allows them to fuse shallow, high-resolution details with deep semantic features, addressing scale variance in irregular tumors. Enhanced FPN variants employ path aggregation and weighted fusion, improving Hausdorff distances by 15%–20% on BraTS 2020. This use enriched receptive fields across scales. BiFPN [23] adds learnable weights for efficiency. As a result, it yields 92% Dice while reducing parameters.

Challenges

The FPN approach has limitations in handling fine-grained details, particularly for microtumors near the ventricles, where inadequate feature fusion fails to capture subtle contrast variations. The use of fixed pyramid ratios does not work for modality-specific depth priorities, leading to suboptimal representation and integration of multiscale features across different imaging modalities.

2.3 Attention Gate Based Architecture

Attention gates calibrate skip connections with additive gating signals from the decoder layers, suppressing irrelevant activations via spatial/channel maps (sigmoid pooling) to focus on tumor boundaries. AG-U-Net [24] and multi-scale CBAM [25] variants refined FLAIR/T2 noise, boosting the tumor Dice to 89% by emphasizing discriminative regions over backgrounds. Self-attention extensions, such as AMSU-Net [26], model long-range dependencies, with 3D adaptations that handle volumetric inconsistencies across BraTS challenges. Chen et al. [27] proposed a 3D U-Net framework with Squeeze-and-Excitation attention to combine residual connections and improve feature recalibration via MRI.

Challenges

When processing a high-resolution MRI image, an attention-based framework increases computation complexity and training time. The presence of imbalanced datasets leads to overfitting toward the dominant classes, reducing the model's ability to make accurate predictions.

2.4 Mamba-Based Segmentation

Mamba, a state-space model (SSM) [28], replaces quadratic Transformer attention with linear-time selective scans for long-range dependencies, which is ideal for 3D MRI sequences. SF-SSM UNet [29] serializes slices for spatiotemporal capture, achieving 88% Dice and low Hausdorff (1.3) on BraTS-2019 via frequency-domain enhancements. DRBD-Mamba [30] uses bidirectional, dual-resolution scans with space-filling curves for robust, efficient segmentation in the presence of heterogeneity. Mamba leverages selective State-Space Models (SSMs) for linear-time sequence modeling, serializing 3D MRI patches via bi-directional scans in UNet-SSM [31] frameworks to capture long-range dependencies without the Transformer's cost. VM-UNet [32] and Mamba Fusion (MF) [30] variants use cross-level MF blocks to enable accurate segmentation from incomplete MRI modalities, with the proposed method achieving a mean dice score near 82%.

Challenges

The SSM approach is limited by its reliance on local inductive biases. It can weaken the representation of subpixel-level edges and fine structural details. The model's sensitivity to scan order can introduce inconsistencies in feature extraction, potentially hindering the reliable detection of heterogeneous tumors.

2.5 Feature Guided Learning

Feature-guided learning deploys instructive modules or priors to supervise decoder fusion, enhancing discriminative cues via multimodal interaction by guiding CNN paths with edge maps. Hybrids like IFEM [33] integrate handcrafted filters with deep learning (DL), refining BraTS boundaries by 10% Dice through adaptive enhancement, bridging supervision gaps in low-data regimes, unlike pure attention.

Challenges

A feature-guided framework is limited by its reliance on pooling operations, which can discard subtle lesion textures and fine-grained details. Reliance on prior information introduces the risk of domain shift and of performance differences across datasets. Table 1 represent the comparative analysis with the previous SOTA approach which evaluated on BraTS2020 dataset.

Table 1: Comparative analysis with previous SOTA work on BraTS2020.

Model	ET Dice (%)	TC Dice (%)	WT Dice (%)
AResU-Net [8]	78.45	89.20	92.10
nnU-Net [18]	74.30	88.50	91.20
FPN-U-Net [22]	76.80	90.15	91.85
SF-SSM UNet [29]	82.10	88.90	91.95
Mamba Fusion [30]	72.08	82.73	89.32
Multi-scale Attn U-Net [34]	80.12	89.75	92.40
EfficientNet-Swin Hybrid [35]	81.50	90.30	92.80
Feature-Guided CNN [36]	79.90	90.05	92.15

Despite these advances, review methods collectively suffer from inefficient multi-scale integration and are vulnerable to MRI scan and class imbalance, necessitating hybrid, lightweight models that combine Mamba efficiency, FPN fusion, guided priors, and calibrated attention for robust, real-time segmentation.

3 Methods

The proposed architecture is an enhancement of the traditional encoder-decoder for better segmentation of tumor tissues. A morphological attention block is integrated for accurate boundary identification on the encoder side. In the V-shape architecture, bottlenecks are always useful for capturing high-level, abstract features; to this end, the study incorporates Mamba and the Fourier Neural Operator (FNO). Transforming preserves contextual information via skip connections, as in FPN, followed by feature-guided learning and a Laplacian block. The overall impact of the proposed architecture targets small regions within the affected area, with varying shapes and locations. The overall architecture is shown in Fig. 1.

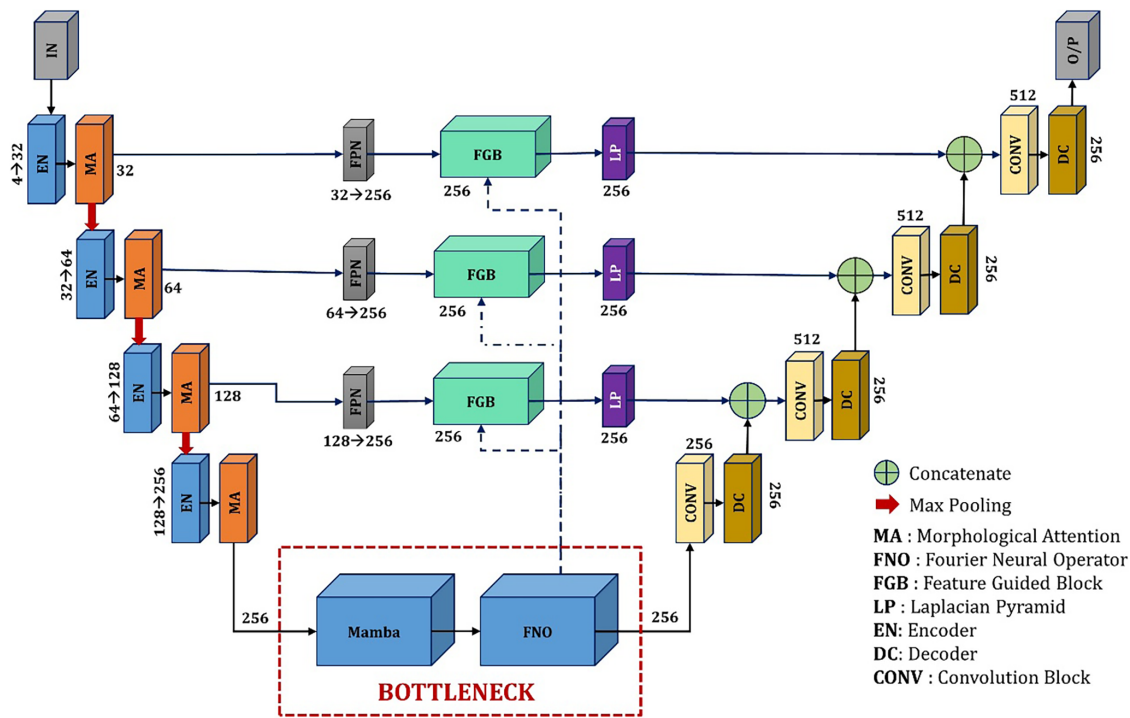


Figure 1: Conceptual architecture of the proposed encoder–decoder network with a hybrid Mamba–FNO bottleneck for joint spatial and frequency-domain feature modeling.

3.1 Encoder with Morphological Attention Block

Four consecutive encoder stages maintain spatial down-sampling and channel expansion for hierarchical feature extraction. Each encoder block consisted of convolution layers followed by batch normalization and the ReLU activation function. Channel expansion helps identify shallow layers (simple features) and deep layers (rich features) from raw MRI intensities (edges, textures, and simple contrasts). Expansions were performed gradually: $4 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 256$. Spatial down-sampling $[64^3 \rightarrow 32^3 \rightarrow 16^3 \rightarrow 8^3]$ helps our architecture to focus on local-voxel-level patterns, such as intensity variation, edges, and boundaries, based on Eq. (1).

$$C_i = \varphi_i \left(\text{MaxPool} \left(C_{\{i-1\}} \right) \right), i \in \{2, 3, 4\}, C_1 = \varphi_1(X), \quad (1)$$

where φ_i represents the convolutional block at level i , and MaxPool denotes 3D max pooling with a stride of 2, and C denote the convolution blocks.

Traditional encoder behavior still struggles to identify enhancing tumors and edema due to sharp intensity gradients, which must be preserved during feature extraction. Morphological attention, shown in Fig. 2A, helps to target shapes and boundary detections using dilation blocks and erosion blocks. The dilation block expands bright regions in the images, calculated using Eq. (2).

$$(f \oplus g)(v) = \max_{u \in N(v)} \{f(v - u)\}, \quad (2)$$

where, $(f \oplus g)$ represents the dilatation of ' f ' by the structuring element of ' g '. The same erosion block $(f \ominus g)$ is used to contract bright regions, where $N(v)$ denotes the neighborhood of voxel v defined by g , as calculated in Eq. (3). The Morphological gradients in back propagation were calculated using Eq. (4).

$$(f \ominus g)(v) = \min_{u \in N(v)} \{f(v + u)\}, \quad (3)$$

$$\nabla_{morph}(f) = (f \oplus g) - (f \ominus g). \quad (4)$$

Gradient highlights image locations where the local intensity range is maximal, thereby producing a strong response at the tissue boundaries. Morphological gradients are concatenated with the attention weight to emphasize boundary regions. The sigmoid activation function ensures that the attention weight is in the range $[0, 1]$.

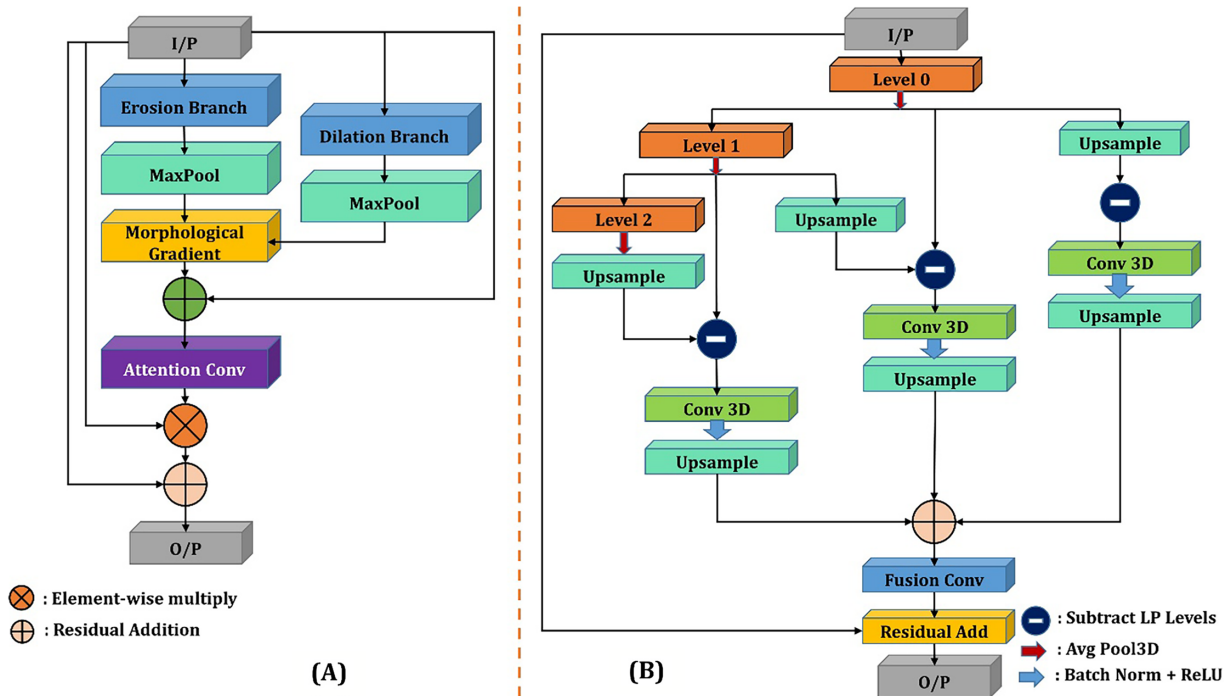


Figure 2: (A) Morphological Attention Block combining erosion, dilation, and gradient operations to enhance edge-aware feature representation. (B) Laplacian Pyramid module for multi-scale edge prediction with hierarchical feature fusion and residual refinement.

3.2 Bottleneck for Spatial Context Enhancement Using Mamba & FNO

Local feature extraction from the BraTS MRI scans is a major challenge due to their complex spatial characteristics. Each MRI modality identifies different regions to better understand tumor spread in the entire hemisphere. First, a Mamba-inspired sequential processing block that treats 3D volumetric features as temporal sequences is introduced, enabling efficient modelling of long-range spatial dependencies through selective state-space modelling. Second, Fourier Neural Operators (FNO) that process features in the frequency domain and capture global structural patterns through spectral convolution are introduced. This dual-domain approach, which combines spatial-sequential processing with frequency-domain analysis, provides comprehensive global context modelling while maintaining tractability.

3.2.1 Mamba-Inspired 3D Block

Mamba targets sequence processing to handle linear complexity and maintain global context through gating and temporal channel mixing, as shown in Fig. 3.

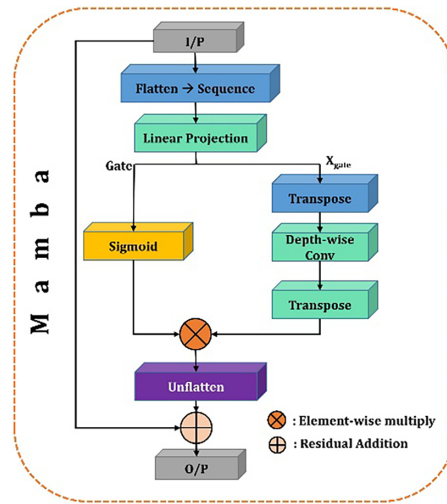


Figure 3: Architecture of the Mamba block illustrating sequence flattening, gated linear projection, depth-wise convolution, and residual connections for efficient long-range dependency modeling.

An earlier spatial-to-sequential transformation has been performed, converting 3D MRI into a sequence for simultaneous processing of the entire slice for ET, TC, and WT identification. The conversion of the sequences is based on the $X_{seq} = Reshape(X) : R^{(B \times C \times D \times H \times W)} \rightarrow R^{(B \times L \times C)}$, where the observations are $L = D \times H \times W = 4096$, $B = 2$, $C = 56$ and $D, H, W = 16$. At the next level, normalization helps identify 256 features from the 4096 tokens received. It targets heterogeneous intensity distributions across modalities and stabilizes the features. Overall calculation of sequence processing, which deals with linear complexity through gating and temporal channels, is expressed as follows, Eq. (5).

$$\hat{X}_{seq}^{b,l,c} = \gamma_c \frac{X_{seq}^{b,l,c} - \mu^{[b,l]}}{\sqrt{(\sigma^2)^{b,l} + \epsilon}} + \beta_c, \quad (5)$$

where $\gamma, \beta \in \mathbb{R}^C$ are learnable parameters, and $\epsilon = 10^{-5}$ is used for numerical stability. μ denotes the average of all channels. σ^2 denotes variance. \hat{X}_{seq} represent layer-normalized features.

After Linear Projection, for dual-path execution, we split the projection feature into two sets, Xgate and Gate. Feature identification is performed through, $X_{proj} = \hat{X}_{seq} \cdot W_{in}^T + b_{in}$, which helps to target “what features are present” like enhancement of tumor and edema intensity through Xgate, and “how important these features are” through Gate. Xgate is further processed by a depth-wise convolutional block to better identify tissues spreading along white matter tracts and capture smooth edema transitions, as defined by Eq. (6).

$$X_{conv}^{b,l,c} = \sum_{i=-K}^K w_c [i + K] X_{gate}^{b,l+i,c}, K = 4 \quad (6)$$

where b : batch index, l : sequence number, c : channel index, W_c : channel-specific weight. The gating information is selected from the depth-wise convolution and the direct gate information after the split by $X_{gated} = X_{conv} \odot \sigma(Gate)$, where σ is used to control the information, $\sigma \rightarrow 1$ passes tumor features, and $\sigma \rightarrow 0$ suppresses background. In the last phase of the mamba, projection and residual are used to integrate the multi-modal information with spatial relation for accurate decoder reconstruction.

3.2.2 Fourier Neural Operator (FNO)

Capturing long-range dependencies with Mamba still leaves the bottleneck; for this reason, we focus on global contextual information. The Mamba’s working behavior is sequence-wise selective mixing; thereafter, the FNO directly evaluates the parameters via the global operator, exploiting spectral representations in an orthogonal and complementary manner, as shown in Fig. 4.

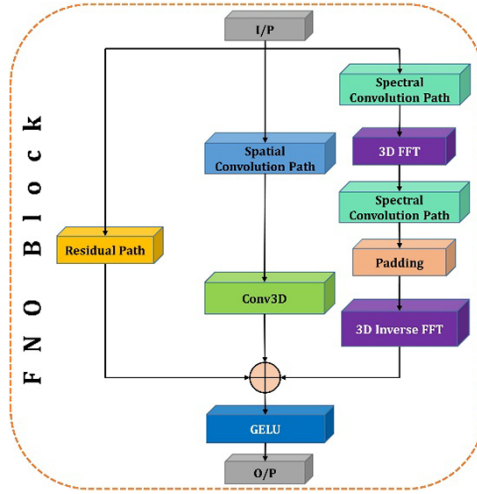


Figure 4: Structure of the Fourier Neural Operator (FNO) block integrating spatial convolution with spectral convolution paths via Fast Fourier Transform (FFT) and inverse FFT for global feature modeling.

Features are represented by $X \in R^{C \times D \times H \times W}$, and 3D Fourier transformation is used to transform the features, which enables global spatial information with respect to fields that are considered as $X = F(x)$. After converting the feature map into the frequency domain, the learning of the weight on different frequency components is calculated based on spectral convolution. Spectral convolution helps modulate low-frequency Fourier modes with learnable weights, as expressed in Eq. (7).

$$y_k^{(c)} = \frac{1}{C} X_c W_{c,k}, \quad (7)$$

here, W represents the trainable spectral coefficients, and truncation enforces efficiency while preserving the dominant anatomical structure. For refined weights, back-mapping is performed through the spatial domain via the inverse Fourier transform, which helps in optimization by combining point-wise convolution and residual connections denoted by $Y = F^{-1}(y)$. After execution the FNO operates in the frequency domain to enforce global consistency and translation-invariant interaction. A combination of Mamba and FNO helps to identify sequential voxel-wise dependencies and holistic anatomical patterns for complex tumor morphology and long-range contextual.

A combination of Mamba and FNO helps identify sequential voxel-wise dependencies and holistic anatomical patterns for complex tumor morphology and long-range contextual correlations. The Mamba component handles spatially continuous relationships with bounded sequential processing, and the FNO component uses global shape priors with spectral convolution, which operates over the entire volume at once. The two blocks operate in different spaces; the Mamba component is based on spatial and sequential structure, and the FNO component is based on spectral structure. This provides the combined model with a richer feature space than either component alone. This is evidenced by improvements in the HD95 metric, as the spectral step in the FNO component preserves high-frequency boundary features that are normally lost with local convolution. The ablation study represents the removal of each component of the MambaFNO-NET bottleneck, which affects overall Dice and HD95 metrics.

3.3 Skip Connection Enhancement Using FPN, FGL & LP

3.3.1 Feature Pyramid Network + Feature Guided Learning

On the decoder side, recovering spatial precision is needed to capture semantic information via the skip connection. The proposed approach focuses more on the skip connection. First, the FPN operates in two directions: a top-down pathway for propagating semantic information and a lateral connection for recovering spatial detail.

Encoder feature maps $\{C_1, C_2, C_3, C_4\}$ and, lateral connections first project them into a unified feature space using $L_i = \mathcal{O}_{1 \times 1 \times 1}(C_i)$. This helps identify semantic information through a top-down pathway, where high-level features are up-sampled and fused with lateral features. To preserve fine spatial details by enabling high-level contextual representation that guides shallow layers through $P_i = \mathcal{O}_{1 \times 1 \times 1}(L_i \oplus \uparrow(P_{i+1}))$. This enhances object size variations and boundaries for 3D MRI.

To clarify the semantic features further, FGL (feature guided learning) is used to handle adaptive attention weights at each pyramid level, represented by $\alpha_i = \sigma(\mathcal{O}_{1 \times 1 \times 1}(P_i))$. Refined features actively identify shallow layers to focus on relevant regions at deeper levels through $\bar{P}_i = P_i \odot \alpha_i + P_i$. This combination helps to combine global semantic context with precise spatial information.

3.3.2 Laplacian Pyramid for Multi-Scale Edge Preservation

Laplacian pyramid is a multi-resolution decomposition that explicitly preserves edge information across scales by computing $L_i = I_i - G(I_i)$, wherein G represents the Gaussian blur and downsample-upsample operations. This operation helps to capture the information discarded by standard pooling operations, such as high-frequency boundaries. Laplacian helps extract the process edges at each scale individually. A pyramid naturally performs band pass filtering: Level 0 isolates fine details, Level 1 captures medium structures, and Level 2 encodes coarse shapes. It is calculated based on Eq. (8).

$$I = \sum_{i=0}^N L_i + G_N \quad (8)$$

L_i is the Laplacian image at level i , G_N Gaussian image at the top level. Each level undergoes separate convolution processing before fusion, enabling scale—specific feature learning while maintaining meaningful information during down sampling, as shown in Fig. 2B.

4 Results

4.1 Dataset

The BraTS2020 [17] dataset was used to analyze the MambaFNO-NET approach, which was provided by the MICCAI. MICCAI is a top-tier conference globally focused on medical image analysis, biomedical computer vision, and AI for healthcare. The BraTS dataset was used for High-Grade Glioma (HGG) and Low-Grade Glioma (LGG) glioma analysis, which is the most aggressive tumor. The BraTS dataset provides an MRI scan that has already been preprocessed by the organizers. Every MRI scan was captured on a 2.5 Tesla machine, which provided 175 slices per scan. The benefits are that tumor boundaries are clearly visible, small tumor regions are easily identified, and the images are less grainy.

The BraTS2020 dataset contains 369 MRI scans, of which 280 are used for training, 50 for validation, and 39 for testing. All MRI scans are in NIfTI (.nii.gz) format with a uniform voxel size of $240 \times 240 \times 155$. Each MRI scan has five different modalities: T1-weighted (T1), T1-contrast-enhanced (T1CE), T2-weighted (T2), Fluid-Attenuated Inversion Recovery (FLAIR), and Ground Truth (SEG), as shown in Fig. 5. The BraTS dataset provides voxel-wise annotations for clear identification of Whole Tumor (WT), Enhanced Tumor (ET), and Tumor Core (TC).

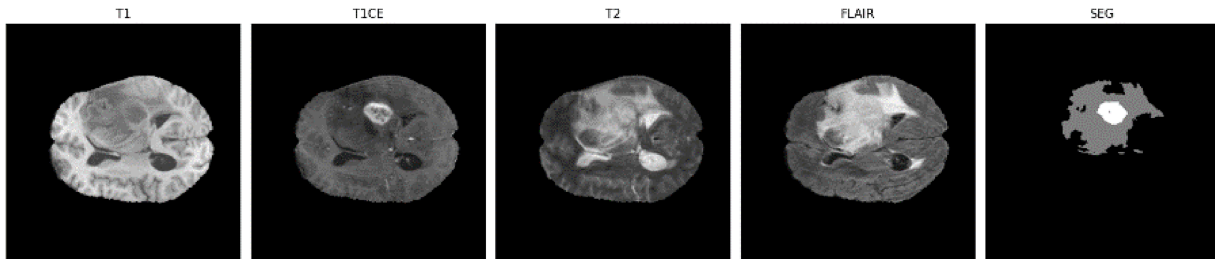


Figure 5: Different modalities of the BraTS2020.

4.2 Training & Validation Setup

The proposed approach was implemented using PyTorch 2.4.0, and an RTX 4090 GPU was used as the processing architecture. For accurate class segmentation, hyperparameters such as the learning rate, batch size, and number of epochs were used. MRI Scan with 3D volumetric multi-modal inputs; the batch size is set to 1 to satisfy the high memory requirements for processing. The final learning rate was 0.01 after evaluating the range of [0.0001, 0.01]. Based on the complexity of the models, 50 epochs were used to cover all aspects of the scans for better visualization and identification. Table 2 lists the hyperparameters.

To avoid overfitting on the BraTS2020 training set, we applied Adam with a weight decay of 1×10^{-4} as L2 regularization and selected the model with the best mean validation Dice score over 50 epochs. The difference between the training and validation Dice scores remained small during the entire training process. Detailed analysis has been discussed in Section 5, which also confirmed by the low standard deviation of the test Dice scores, which were all close to the mean: ET: ± 0.05 , WT: ± 0.02 , and TC: ± 0.03 .

To evaluate the practical deployment of MambaFNO-NET, Table 3 presents the total trainable parameters and floating point operations (FLOPs) measured under the above-mentioned conditions.

Table 2: Hyperparameter settings used for training the proposed approach with range and best value.

Parameters	Best Value	Parameter Range
Epoch	50	30, 40, 50, 60
Optimizer	Adam	Adam, AdamW, RAdam
Learning Rate	0.01	0.01, 0.001, 0.0001
Dice Loss Weight	0.75	0.75, 0.65
Focal Loss Weight	0.25	0.25, 0.35
Precision Mode	FP16	FP16, FP32
Weight Decay	1×10^{-4}	1×10^{-3} , 1×10^{-4} , 1×10^{-5}

Table 3: Computation comparison of MambaFNO-NET with SOTA.

Model	Parameters (M)	FLOPs (G)
UNETR [13]	92.78	82.60
FPN-UNet [22]	1.43	102.24
nnU-Net [37]	34.33	1405.78
SegNet 3D [38]	31.50	178.80
U-SegNet [39]	28.50	35.80
dResU-Net [40]	15.83	56.84
MambaFNO-NET	25.64	77.98

MambaFNO-NET achieves a favorable balance between the model capacity and computational cost. With 25.64M parameters, it remains significantly lighter than UNETR (92.78M) and nnU-Net (34.33M), which are the two most widely adopted transformer and CNN baselines, respectively. Notably, its FLOPs of 77.98 G represent a reduction of approximately 18× compared to nnU-Net (1405.78 G). The GPU memory requirement of 6.54 GB was well within the capacity of a single clinical-grade GPU, confirming practical deploy ability without a multi-GPU infrastructure.

4.3 Evaluation Parameters

The results of brain tumor segmentation are based on two measurements: Dice Similarity Coefficient (DSC) and Hausdorff. The DSC measures the spatial overlap between the ground truth and the results predicted by the models. This can be calculated using Eq. (9). A higher DSC [41] represents the exact match of the ground truth which is manually identified by a neurologist.

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (9)$$

In Eq. (9), False Positive, False Negative, and True Positive are represented by FP , FN , and TP , respectively.

Hausdorff distances were used to evaluate the boundary discrepancy between the predicted truth and ground truth. A lower HD95 [42] indicates better results, as calculated using Eq. (10).

$$HD_{95}(A, B) = \max \left\{ percentile_{95} \left(\min_{b \in B} \|a - b\| \right), percentile_{95} \left(\min_{a \in A} \|b - a\| \right) \right\}, \quad (10)$$

where predicted segmentation and ground truth are represented as A and B . Euclidean distance between two points of the predicted and ground truth is described by $d(a, b)$. Based on the above evaluation parameters, High Dice scores and low HD95 values indicate accurate region overlap and precise boundary alignment between the predicted and ground-truth segmentations.

5 Results & Analysis

The MambaFNO-NET approach yielded a mean DSC of 0.86, outperforming most competing approaches by 0.02–0.19 shown in Table 4. In particular, compared with Nguyen et al. (0.82) [43] and Zhao et al. (0.84) [44], the proposed approach shows absolute advantages of +0.04, +0.06, and +0.02, respectively. In contrast, the overlap accuracies of the previous convolution-based models Luo et al. (0.75) [45] and Ding et al. (0.70) [46] were significantly lower than that of the proposed approach, with absolute differences of -0.19 and -0.16 , respectively. Compared with the best models available in the BraTS 2020 challenge, Isensee et al. [37] and Jia et al. [47], which reported a mean DSC of 0.85, the proposed model holds a positive absolute margin of +0.01. This shows that the proposed model does not overfit a single sub-region of the tumor at the cost of other regions.

Table 4: Quantitative comparison of dice score & HD95 across the baseline and proposed approach.

Model	DSC				HD95			
	ET	WT	TC	Mean	ET	WT	TC	Mean
MambaFNO-NET	0.81	0.92	0.86	0.86	4.21	6.22	6.85	5.76
Peng & Sun [48]	0.76	0.9	0.8	0.82	35.2	7.22	15.3	19.24
Wang et al. [49]	0.6	0.81	0.61	0.67	56.9	22.3	25.7	34.95
Ding et al. [46]	0.66	0.88	0.73	0.7	46.5	7.95	16	23.48
Liu et al. [50]	0.68	0.88	0.76	0.77	38.7	5.37	15.3	19.79
Luo et al. [45]	0.67	0.84	0.73	0.75	53.2	34.2	33.5	40.3
Zhao et al. [44]	0.78	0.9	0.83	0.84	26.35	4.02	6.51	12.29
Ding et al. [51]	0.61	0.86	0.78	0.75	–	–	–	–
Agravat and Raval [52]	0.76	0.87	0.75	0.79	–	–	–	–
Nguyen et al. [43]	0.8	0.85	0.81	0.82	–	–	–	–
Ghaffari et al. [53]	0.78	0.9	0.83	0.83	–	–	–	–
Guan et al. [54]	0.68	0.85	0.69	0.74	47.4	8.44	31.6	29.14
Zhu et al. [55]	0.80	0.87	0.93	0.86	–	–	–	–
Zhang et al. [56]	0.77	0.91	0.84	0.84	–	–	–	–
Lu et al. [29]	0.91	0.90	0.88	0.89	–	–	–	–
Isensee et al. [37] (2020 winner)	0.79	0.91	0.85	0.85	19.2	3.79	7.77	13.59
Jia et al. [47] (2020 2nd place)	0.78	0.91	0.85	0.85	26.57	4.97	11.91	14.48

In terms of region-wise performance, the proposed approach achieved a DSC of 0.81 (ET), 0.92 (WT), and 0.86 (TC). In the case of the ET region, known for its high spatial variability and low volumetric consistency, the proposed approach outperformed the most recent methods (Zhao et al., 0.78 [44]; Nguyen et al., 0.80 [43]) by absolute margins of +0.03 to +0.05. In the TC region, the proposed approach also achieved absolute gains of +0.02 to +0.06 across most comparative methods, indicating better structural discrimination between the necrotic and non-enhancing tumor regions.

Regarding the surface distance error of HD95, the proposed method had a mean value of 5.76 mm, which was significantly smaller than that of most existing methods. Compared with Luo et al. (40.3 mm) [45], and Guan et al. (29.14 mm) [54], the proposed method can reduce the absolute boundary error by -29.19 , -34.54 , and -23.38 mm, respectively. The large margins indicate that the boundaries are significantly better aligned in space.

Even when compared with strong baselines such as Isensee et al. (13.59 mm) [37] and Jia et al. (14.48 mm) [47], the proposed method shows absolute reductions of -7.83 and -8.72 mm, respectively. It is also important to note that the proposed method showed low HD95 values for all tumor sub-regions (ET: 4.21 mm, WT: 6.22 mm, TC: 6.85 mm).

The observed stability in the DSC and HD95 values is reflected in the visuals in Fig. 6. The predicted whole-tumor (WT) region boundaries were smooth and continuous with very little leakage outside, indicating an excellent understanding of the global context. Within the WT region, the tumor core (TC) segmentation was also seamless, with no gaps, indicating stable region boundary extraction. The enhancing tumor (ET) region, although small and irregularly shaped, remained well-contained within the TC boundaries with good adherence to their edges, and very few false positives. The visual overlap indicates that the improvement in overlap accuracy is due to precise localization rather than boundary expansion, which helps mitigate surface-distance outliers. On selected slices, the predicted masks aligned with the ground truth in a stable spatial manner, indicating reduced prediction variability and successful exploration of the trade-off between overlap accuracy and boundary accuracy.

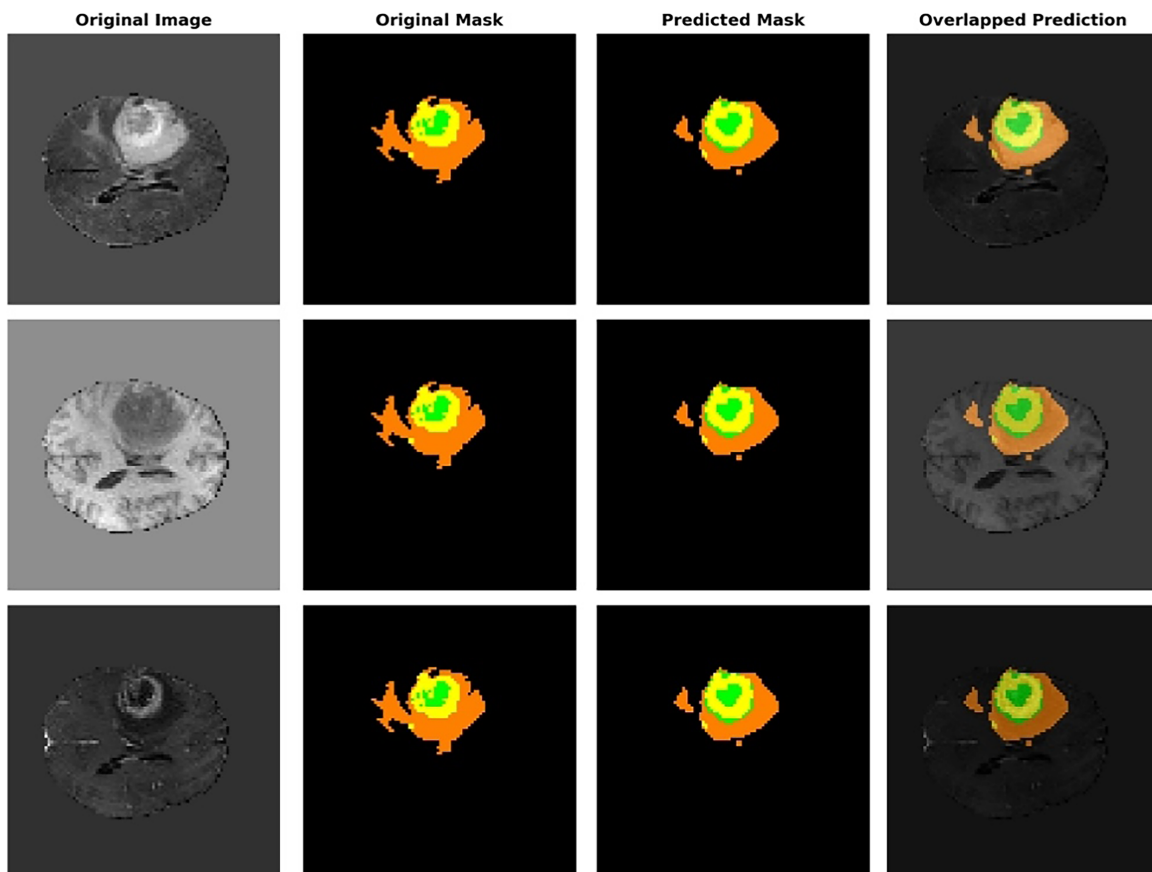


Figure 6: (Continued)

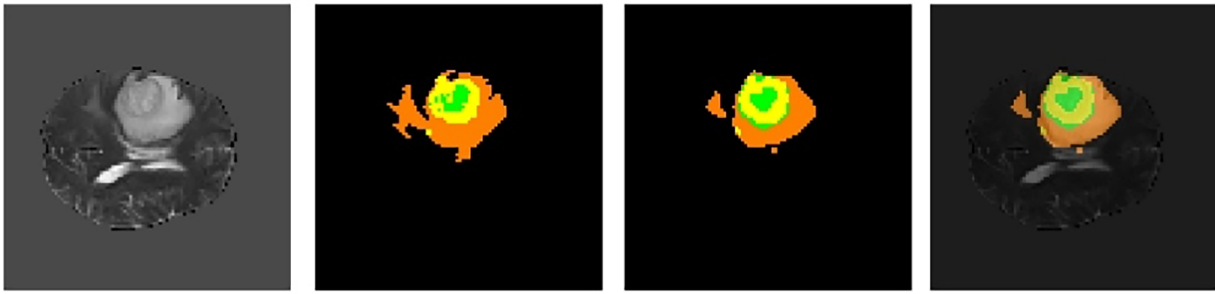


Figure 6: Qualitative comparison of segmentation results showing the original input images, ground truth masks, predicted masks, and overlapped predictions, demonstrating the proposed model’s ability to accurately localize and delineate lesion regions across different samples.

Fig. 7 shows two training metrics for the MambaFNO-NET architecture on BraTS2020: the left graph shows the training loss going from 0.8 to 0.1 over 50 epochs, which shows that the model is training well, while the right graph shows the validation Dice values for the three classes (TC, WT, and ET—possibly different regions of the tumor) increasing from 0 to 0.9, and the stacked area chart showing that all segmentation classes are being improved simultaneously. The two graphs in this image show that the model is training well, with both the loss and validation performance improving.

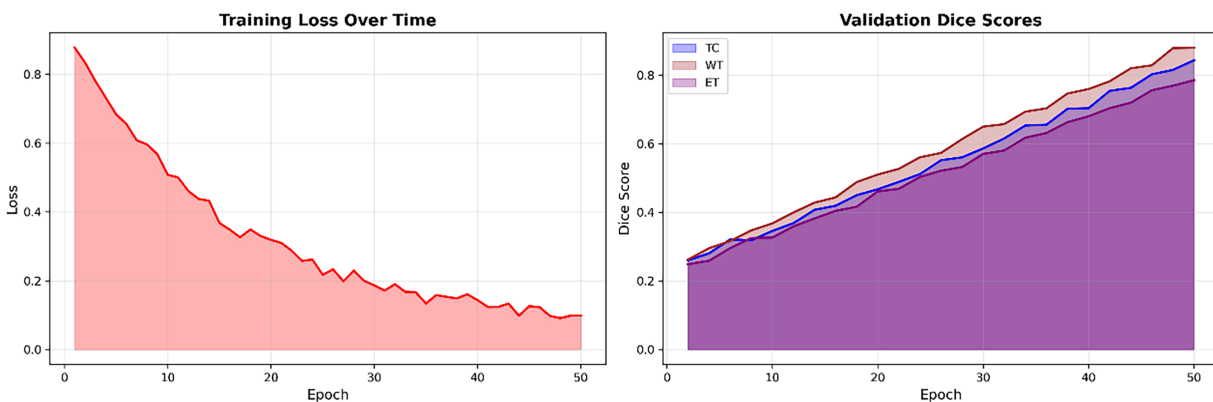


Figure 7: Training loss convergence and validation Dice score evolution of the proposed model over training epochs for Tumor Core (TC), Whole Tumor (WT), and Enhancing Tumor (ET), illustrating stable optimization and consistent generalization performance across tumor sub regions.

5.1 Comparison with U-Net

Fig. 8 shows a qualitative comparison of the proposed approach with the U-Net baseline for representative patient cases. The U-Net predictions were characterized by fragmented tumor areas, incomplete delineation of the tumor core (TC) and enhancing tumor (ET), and irregular boundaries, particularly in low-contrast and heterogeneous areas. In contrast, the proposed approach generates more compact, spatially consistent, and anatomically valid segmentations of all the tumor areas. The visual improvements are reflected in higher region-wise Dice scores, suggesting that the proposed approach offers better feature representation and contextual modeling than the U-Net baseline.

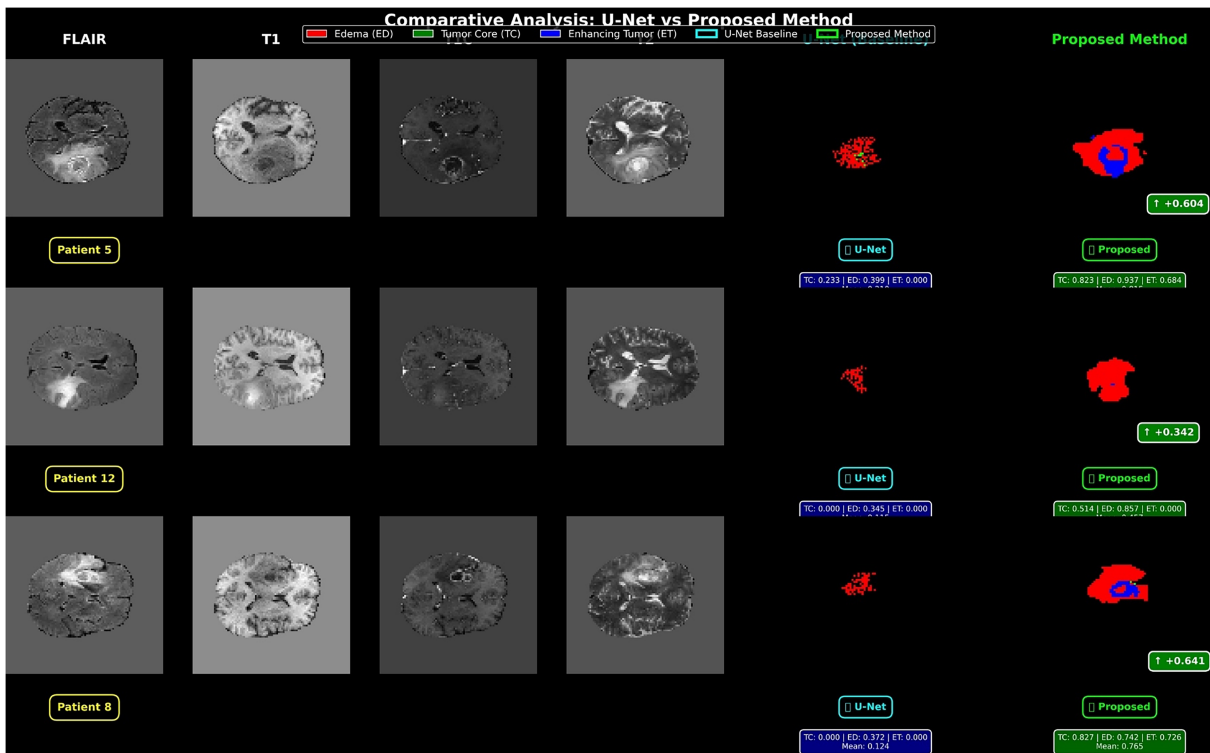


Figure 8: Qualitative comparative analysis between U-Net and the proposed method across different MRI modalities (FLAIR, T1, T1ce, T2), illustrating improved lesion localization and boundary delineation achieved by the proposed model.

5.2 Error Analysis

Fig. 9 presents quantitative analysis of three cases, showing the predicted segmentation overlaid on the ground truth, along with per-voxel error maps for True Positive (TP), False Positive (FP), and False Negative (FN) across three regions. In Case 10, Statistics of the voxels (TP: 649, FP: 58, FN: 139) in all three sub-regions are consistent to achieve a high dice score (WT Dice: 0.966, TC: 0.908, ET: 0.932), which acknowledges the morphological attention block, and FNO produces well-contained and precise locations in high-contrast cases. About Cases 5 and 0, higher FP count at WT periphery (FP: 47 and 43, respectively) directly affects dice drop to 0.784 for WT—the model over-segments the diffuse edema boundary. Importantly, TC and ET robustly segmented even in this difficult case, which corresponds to the ablation finding in Table 5 that the Mamba block contributes well in TC Dice gain (+1.6, from B2: 82.1 to B3: 83.7), which confirms that long-range sequential dependency modelling is particularly critical for the structurally complex necrotic core. Overall, in difficult cases, core tumor identification is accurate, but on low-contrast scans, WT boundaries remain challenging to identify.

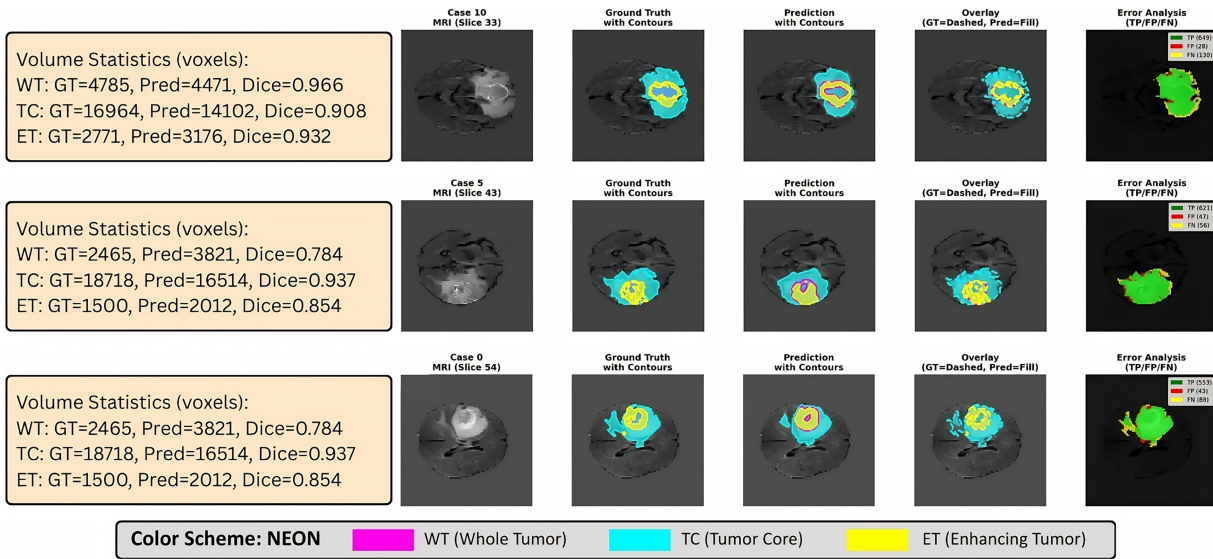


Figure 9: Voxel-wise—error analysis represent the WT, TC and ET ground truth and prediction voxel with dice similarity.

Table 5: Quantitative ablation results showing the effect of progressive enabling FNO, morphological attention, laplacian pyramid, mamba and feature guided learning (FGL) modules.

Model	FNO	Morph	Laplacian	Mamba	FGL	DSC(WT)	DSC(TC)	DSC(ET)
B0	✗	✗	✗	✗	✗	86.1	78.4	72.3
B1	✗	✓	✗	✗	✗	87.4	80.2	74.1
B2	✗	✓	✓	✗	✗	88.6	82.1	75.9
B3	✗	✓	✓	✓	✗	89.8	83.7	77.3
B4	✓	✓	✓	✓	✗	90.6	85.2	78.8
Final	✓	✓	✓	✓	✓	91.4	86.5	80.1

6 Comparative Analysis Based on Different Components

Ablation are conducted based on the adding the component one by one in to the base line model which is shown in Table 6. The ablation study in Table 5 shows a gradual increase in segmentation accuracy with the addition of each proposed module, and these quantitative gains are reflected in the visual analysis in Fig. 10. Improving ET Dice from 72.3 to 74.1, shown in Fig. 10, which sharpens tumor contours and reduces boundary leakage around the enhancing region, is due to adding Morphological attention (B0→B1). The Laplacian Pyramid(B1→B2) helps to recover fine edge detail and improve ET from 74.1 to 75.9. The Mamba block (B2→B3) improves the TC (+1.6:82.1→83.7), aligning with Fig. 10, where it shows the most complete recovery of the tumor core. Addition of FNO(B3→B4) helps for the largest WT improvement (+0.8), where FNO helps to capture the effective whole-tumor region through global frequency-domain shape encoding. Feature-Guided Learning (B4→Final) helps to overall gain in all the sub-regions (WT: +0.8, TC: +1.3, ET: +1.3). The role of the FGL is to suppress the false positives across heterogeneous regions. The residual WT boundary failure identified in the error analysis (Section 5.2, Fig. 9) is consistent with the gap between B0 and the final WT model, confirming that the combination of all modules accurately handles diffuse low-contrast edema boundaries.

Table 6: Ablation analysis of MambaFNO-NET architecture, illustrates the progressive addition of key components from B0 to the final model.

Model ID	Component Insertion	Purpose
B0	Baseline FPN-UNet3D	Base Model
B1	B0 + Morphological Attention	Improves boundary delineation by explicitly modeling dilation–erosion gradients, leading to sharper tumor contours and reduced boundary leakage.
B2	B1 + Laplacian Pyramid	Enables explicit multi-scale edge preservation, improving sensitivity to small and medium-sized tumor regions while reducing blurring artifacts.
B3	B2+ Mamba Block	Captures long-range dependencies across the 3D volume, resulting in improved global consistency of tumor predictions.
B4	B3+ FNO Block	In the frequency domain, enabling efficient modeling of global anatomical patterns, which enhances whole-tumor segmentation accuracy.
Final	B4+ Feature Guided Learning	Aligns deep semantic features with shallow spatial features, improving cross-scale consistency and final segmentation quality.

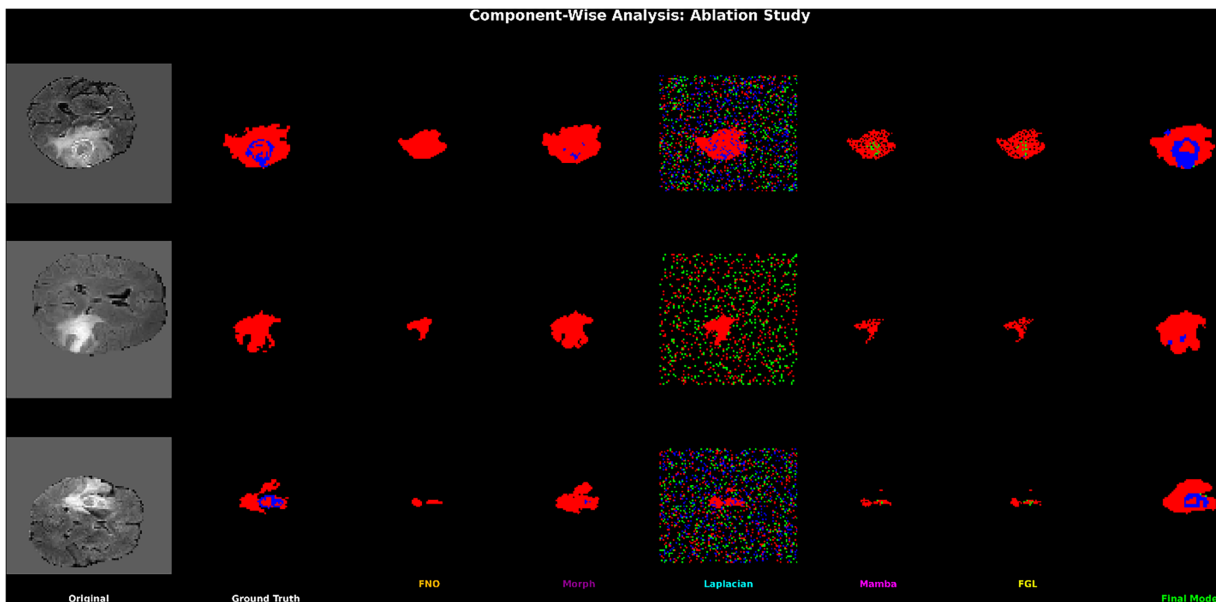


Figure 10: Qualitative component-wise ablation study to demonstrate that the individual modules are capable of partial lesion localization, and the final integrated model performs better than the ground truth.

Component-wise evaluation is shown in Fig. 10, which presents the ablation analysis for brain tumor segmentation. Each row corresponds to a patient, and the columns show the Original FLAIR MRI, Ground Truth, predictions of individual components (FNO, Morph, Laplacian, Mamba, FGL), and the Final Model.

The tumor regions are marked using color codes: Tumor Core (green), Edema (red), and Enhancing Tumor (blue). This visualization clearly shows the unique contribution of each component in the final model, where FNO focuses on the global tumor shape, Morph helps with boundary refinement, Laplacian highlights the edges, Mamba maintains the contextual information, and FGL helps with feature localization. In this process, the above components work at a single resolution level, and the FPN helps create a multi-scale semantic bridge to ensure that the global tumor shape is learned by FNO and Mamba. The Final Model, which combined all components, achieved the best possible segmentation, very close to the ground truth, confirming the efficacy of combining components for improved tumor segmentation.

7 Discussion

The strength of the MambaFNO-NET framework lies in the synergistic combination of complementary mechanisms that address the challenges of brain tumor segmentation. Morphological Attention is specifically designed to model boundary gradients that are not well captured by conventional convolution, and Laplacian Pyramids are used to preserve multi-scale edge information, which is essential for detecting irregular tumors. The Mamba-FNO bottleneck is a unique combination of linear-time sequential processing and global frequency-domain pattern recognition that can address the computational constraints of transformer-based models. Feature-Guided Learning helps maintain semantic-spatial consistency across different scales of the decoder, thereby helping suppress false positives in heterogeneous areas. The proposed work was evaluated on the BraTS 2020 dataset, a commonly used, highly controlled dataset for brain tumor segmentation. This will allow us to compare the current state of the art on the same data splits and evaluation procedures. BraTS 2020 provides both High Grade Glioma (HGG) and Low-Grade Glioma (LGG) cases collected from various institutions with varying imaging machines. The small variation in Dice scores across models (ET: ± 0.05 , WT: ± 0.02 , TC: ± 0.03) indicates that the model is working consistently across cases in the dataset rather than fitting to a small subset. Testing on additional datasets, such as BraTS2021 and different multi-center clinical datasets, will be a key step in the future to verify domain generalization.

8 Conclusion

This work presents a hybrid deep learning architecture for automatic brain tumor segmentation, capable of addressing the variability in brain tumor morphology, irregular boundaries, and the need for efficient 3D MRI analysis. By carefully combining Morphological Attention modules, Mamba-based sequential modeling, Fourier Neural Operators, Feature Pyramid Networks with Feature-Guided Learning, and Laplacian Pyramids, this architecture shows significant improvements over state-of-the-art solutions on the BraTS 2020 dataset. The Mamba FNO-NET achieved a mean Dice Similarity Coefficient of 0.86 for tumor sub-regions with outstanding boundary accuracy (mean HD95 of 5.76 mm), showing significant gains over the BraTS 2020 challenge winners. Well-balanced performance on Enhancing Tumor (Dice 0.81, HD95 4.21 mm), Whole Tumor (Dice 0.92, HD95 6.22 mm), and Tumor Core (Dice 0.86, HD95 6.85 mm) shows robustness to different tumor types. Systematic ablation experiments confirmed the complementary roles of each component: Morphological Attention improved boundaries, Laplacian Pyramids preserved edges, Mamba enabled efficient long-range modeling, and FNO captured global patterns. From a clinical perspective, the MambaFNO-NET architecture, with its volumetric and boundary accuracy, can be used for treatment planning, surgical planning, and radiation therapy planning. While the current study demonstrates strong performance on the BraTS dataset, further validation across different MRI scans and imaging parameters is still needed for clinical deployment. Future work will focus on uncertainty estimation, multi-center validation, and handling incomplete modalities.

Acknowledgement: This research was supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R136). The authors would also like to acknowledge the APC support of Prince Sultan University, Riyadh, Saudi Arabia.

Funding Statement: This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R136), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Author Contributions: Ronak Patel, Tanzila Saba, Noor A. Khan contributed to Methodology, Implementation, Investigation, and Writing—Original Draft; Miral Patel, Shaha Al-Otaibi, Roaa Khalil Mohamed Ali Abed, Deep Kothadiya contributed to Visualization and Supervision and Writing—Review & Editing. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The BraTS 2020 dataset used in this study is publicly available and can be accessed through the official Brain Tumor Segmentation (BraTS) Challenge repository at <https://www.med.upenn.edu/cbica/brats2020/data.html>.

Ethics Approval: This article does not contain any studies with human participants or animals performed by any of the authors. All data used in this work are obtained from publicly available datasets.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Price M, Ballard CAP, Benedetti JR, Kruchko C, Barnholtz-Sloan JS, Ostrom QT. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2018–2022. *Neuro Oncol.* 2025;27(Supplement_4):iv1–66. doi:10.1093/neuonc/noaf194.
2. Huang M, Zou J, Zhang Y, Bhatti UA, Chen J. Efficient click-based interactive segmentation for medical image with improved plain-ViT. *IEEE J Biomed Health Inform.* 2025;29(12):8904–16. doi:10.1109/JBHI.2024.3392893.
3. Ali H. A meta-review of computational intelligence techniques for early autism disorder diagnosis. *Int J Theor Appl Comput Intell.* 2025;1–21. doi:10.65278/ijtaci.2025.1:
4. Rehman A. Brain stroke prediction through deep learning techniques with ADASYN strategy. In: 2023 16th International Conference on Developments in eSystems Engineering (DeSE); 2023 Dec 18–20; Istanbul, Turkiye. p. 679–84. doi:10.1109/DeSE60595.2023.10469013.
5. MICCAI BRATS—the multimodal brain tumor segmentation challenge [Internet]. [cited 2026 Jan 1]. Available from: <http://braintumorsegmentation.org/>.
6. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-assisted Intervention—MICCAI 2015*. Cham, Switzerland: Springer International Publishing; 2015. p. 234–41. doi:10.1007/978-3-319-24574-4_28.
7. Ramzan F, Khan MUG, Iqbal S, Saba T, Rehman A. Volumetric segmentation of brain regions from MRI scans using 3D convolutional neural networks. *IEEE Access.* 2020;8:103697–709. doi:10.1109/ACCESS.2020.2998901.
8. Zhang J, Lv X, Zhang H, Liu B. AResU-Net: attention residual U-Net for brain tumor segmentation. *Symmetry.* 2020;12(5):721. doi:10.3390/sym12050721.
9. Zhang Y, Han Y, Zhang J. MAU-Net: mixed attention U-Net for MRI brain tumor segmentation. *Math Biosci Eng.* 2023;20(12):20510–27. doi:10.3934/mbe.2023907.
10. Sharma V, Kumar M, Yadav AK. 3D AIR-UNet: attention-inception-residual-based U-Net for brain tumor segmentation from multimodal MRI. *Neural Comput Appl.* 2025;37(16):9969–90. doi:10.1007/s00521-025-11105-9.
11. Wang W, Chen C, Ding M, Yu H, Zha S, Li J. TransBTS: multimodal brain tumor segmentation using transformer. In: *Medical image computing and computer assisted intervention—MICCAI 2021*. Cham, Switzerland: Springer International Publishing; 2021. p. 109–19. doi:10.1007/978-3-030-87193-2_11.
12. Ben Gara Ali M, Smiti A. Dynamic Swin-UNet: a transformer-based adaptive framework for precise and efficient Alzheimer’s disease brain segmentation. *Multimed Tools Appl.* 2026;85(2):88. doi:10.1007/s11042-026-21266-x.

13. Kothadiya D, Rehman A, AlGhofaily B, Bhatt C, Ayesha N, Saba T. VGX: VGG19-based gradient explainer interpretable architecture for brain tumor detection in microscopy magnetic resonance imaging (MMRI). *Microsc Res Tech.* 2025;88(5):1544–54. doi:10.1002/jemt.24809.
14. Patel RR, Patel M, Kothadiya D. ResAtt-NASFPN: a residual attention driven NAS-FPN framework for robust 3D brain tumor segmentation. *J Innov Image Process.* 2026;8(1):34–53. doi:10.36548/jiip.2026.1.003.
15. Yao H, Xiong H, Liu D, Shen H, Berkovsky S. BraTS-UMamba: adaptive mamba UNet with dual-band frequency based feature enhancement for brain tumor segmentation. In: *Medical Image Computing and Computer Assisted Intervention—MICCAI 2025*. Cham, Switzerland: Springer Nature; 2025. p. 98–107. doi:10.1007/978-3-032-05325-1_10.
16. Xu W, Fu YL, Zhu D. ResNet and its application to medical image processing: research progress and challenges. *Comput Methods Programs Biomed.* 2023;240(9):107660. doi:10.1016/j.cmpb.2023.107660.
17. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging.* 2015;34(10):1993–2024. doi:10.1109/tmi.2014.2377694.
18. Magadza T, Viriri S. Efficient nnU-net for brain tumor segmentation. *IEEE Access.* 2023;11:126386–97. doi:10.1109/ACCESS.2023.3329517.
19. Sarker L, Yeafi A. EF-SwinNet: a hybrid EfficientNet-swin transformer model for skin cancer classification. In: *2024 International Conference on Recent Progresses in Science, Engineering and Technology (ICRPSET); 2024 Dec 7–8; Rajshahi, Bangladesh.* p. 1–4. doi:10.1109/ICRPSET64863.2024.10955919.
20. Gao S, Li ZY, Yang MH, Cheng MM, Han J, Torr P. Large-scale unsupervised semantic segmentation. *IEEE Trans Pattern Anal Mach Intell.* 2023;45(6):7457–76. doi:10.1109/tpami.2022.3218275.
21. Lairedj KI, Chama Z, Bagdaoui A, Laguech S, Menni Y, Becheikh N, et al. Advanced brain tumor segmentation in magnetic resonance imaging via 3D U-Net and generalized Gaussian mixture model-based preprocessing. *Comput Model Eng Sci.* 2025;144(2):2419–43. doi:10.32604/cmesci.2025.069396.
22. Sun H, Yang S, Chen L, Liao P, Liu X, Liu Y, et al. Brain tumor image segmentation based on improved FPN. *BMC Med Imaging.* 2023;23(1):172. doi:10.1186/s12880-023-01131-1.
23. Annavarapu CSR, Parisapogu SAB, Keetha NV, Donta PK, Rajita G. A Bi-FPN-based encoder-decoder model for lung nodule image segmentation. *Diagnostics.* 2023;13(8):1406. doi:10.3390/diagnostics13081406.
24. Zhang J, Jiang Z, Dong J, Hou Y, Liu B. Attention gate ResU-net for automatic MRI brain tumor segmentation. *IEEE Access.* 2020;8:58533–45. doi:10.1109/ACCESS.2020.2983075.
25. Jiao L, Liu Y, Gu Y, Wu J, Zhang F. Multi-scale TransUnet combined with CBAM for nuclear image segmentation. In: *2023 12th International Conference on Computing and Pattern Recognition.* New York, NY, USA: ACM; 2023. p. 395–401. doi:10.1145/3633637.3633699.
26. Yin Y, Han Z, Jian M, Wang GG, Chen L, Wang R. AMSUnet: a neural network using atrous multi-scale convolution for medical image segmentation. *Comput Biol Med.* 2023;162(9):107120. doi:10.1016/j.combiomed.2023.107120.
27. Chen YT, Ahmad N, Aurangzeb K. Enhancing 3D U-Net with residual and squeeze-and-excitation attention mechanisms for improved brain tumor segmentation in multimodal MRI. *Comput Model Eng Sci.* 2025;144(1):1197–224. doi:10.32604/cmesci.2025.066580.
28. Patro BN, Agneeswaran VS. Mamba-360: survey of state space models as transformer alternative for long sequence modelling: methods, applications, and challenges. *Eng Appl Artif Intell.* 2025;159(12):111279. doi:10.1016/j.engappai.2025.111279.
29. Lu J, Ding H, Huo Q, Wang K, Sun X, Zhang S. A sequential flow UNet for MRI brain tumor segmentation based on state-space-model. *Appl Soft Comput.* 2026;186(4):114069. doi:10.1016/j.asoc.2025.114069.
30. Liu C, Li XL, Xu D, Wang H, Jiang J. Mamba-based brain tumor segmentation of incomplete multi-modal MR images. *Quant Imaging Med Surg.* 2026;16(2):142. doi:10.21037/qims-2025-1913.
31. Meng W, Mu A, Wang H. Efficient UNet fusion of convolutional neural networks and state space models for medical image segmentation. *Digit Signal Process.* 2025;158:104937. doi:10.1016/j.dsp.2024.104937.
32. Ruan J, Li J, Xiang S. VM-UNet: vision mamba UNet for medical image segmentation. *ACM Trans Multimedia Comput Commun Appl.* 2025;3767748. doi:10.1145/3767748.

33. Elbachir YM, Makhoulouf D, Mohamed G, Bouhamed MM, Abdellah K. Federated learning for multi-institutional on 3D brain tumor segmentation. In: 2024 6th International Conference on Pattern Analysis and Intelligent Systems (PAIS); 2024 Apr 24–25; El Oued, Algeria; 2024. p. 1–8. doi:10.1109/PAIS62114.2024.10541292.
34. Preetha R, Jasmine Pemeena Priyadarsini M, Nisha JS. Brain tumor segmentation using multi-scale attention U-Net with EfficientNetB4 encoder for enhanced MRI analysis. *Sci Rep.* 2025;15(1):9914. doi:10.1038/s41598-025-94267-9.
35. Kunekar P, Yadav A, Yadav A, Dusankar Y, Nalawade Y, Yawale S. Hybrid swin transformer EfficientNet U-Net model for enhanced brain tumor segmentation. *Res Sq.* 2025. doi:10.21203/rs.3.rs-6964779/v1.
36. Liu Z, Liu X, Qu L, Shi Y. FANCL: feature-guided attention network with curriculum learning for brain metastases segmentation. *Neurocomputing.* 2025;655(4):131369. doi:10.1016/j.neucom.2025.131369.
37. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Meth.* 2021;18(2):203–11. doi:10.1038/s41592-020-01008-z.
38. Zhang C, Lu W, Wu J, Ni C, Wang H. SegNet network architecture for deep learning image segmentation and its integrated applications and prospects. *Acad J Sci Technol.* 2024;9(2):224–9. doi:10.54097/rfa5x119.
39. Kumar P, Nagar P, Arora C, Gupta A. U-segnet: fully convolutional neural network based automated brain tissue segmentation tool. In: 2018 25th IEEE International Conference on Image Processing (ICIP); 2018 Oct 7–10; Athens, Greece. p. 3503–7. doi:10.1109/ICIP.2018.8451295.
40. Raza R, Ijaz Bajwa U, Mehmood Y, Waqas Anwar M, Hassan Jamal M. dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI. *Biomed Signal Process Control.* 2023;79(4):103861. doi:10.1016/j.bspc.2022.103861.
41. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index1 scientific reports. *Acad Radiol.* 2004;11(2):178–89. doi:10.1016/S1076-6332(03)00671-8.
42. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell.* 1993;15(9):850–63. doi:10.1109/34.232073.
43. Nguyen HT, Le TT, Nguyen TV, Nguyen NT. Enhancing MRI brain tumor segmentation with an additional classification network. In: *Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries.* Cham, Switzerland: Springer International Publishing; 2021. p. 503–13. doi:10.1007/978-3-030-72084-1_45.
44. Zhao X, Zhang P, Song F, Ma C, Fan G, Sun Y, et al. Prior attention network for multi-lesion segmentation in medical images. *IEEE Trans Med Imaging.* 2022;41(12):3812–23. doi:10.1109/TMI.2022.3197180.
45. Luo Z, Jia Z, Yuan Z, Peng J. HDC-net: hierarchical decoupled convolution network for brain tumor segmentation. *IEEE J Biomed Health Inform.* 2021;25(3):737–45. doi:10.1109/JBHI.2020.2998146.
46. Ding Y, Gong L, Zhang M, Li C, Qin Z. A multi-path adaptive fusion network for multimodal brain tumor segmentation. *Neurocomputing.* 2020;412(1):19–30. doi:10.1016/j.neucom.2020.06.078.
47. Jia H, Cai W, Huang H, Xia Y. H²NF-net for brain tumor segmentation using multimodal MR imaging: 2nd place solution to BraTS challenge 2020 segmentation task. In: *Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries.* Cham, Switzerland: Springer International Publishing; 2021. p. 58–68. doi:10.1007/978-3-030-72087-2_6.
48. Peng Y, Sun J. The multimodal MRI brain tumor segmentation based on AD-Net. *Biomed Signal Process Control.* 2023;80(5):104336. doi:10.1016/j.bspc.2022.104336.
49. Wang E, Hu Y, Yang X, Tian X. TransUNet with attention mechanism for brain tumor segmentation on MR images. In: 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA); 2022 Jun 24–26; Dalian, China. p. 573–7. doi:10.1109/ICAICA54878.2022.9844551.
50. Liu L, Cheng J, Quan Q, Wu FX, Wang YP, Wang J. A survey on U-shaped networks in medical image segmentations. *Neurocomputing.* 2020;409(2):244–58. doi:10.1016/j.neucom.2020.05.070.
51. Ding Y, Yu X, Yang Y. RFNet: region-aware fusion network for incomplete multi-modal brain tumor segmentation. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV); 2021 Oct 10–17; Montreal, QC, Canada. p. 3955–64. doi:10.1109/ICCV48922.2021.00394.

52. Agravat RR, Raval MS. 3D semantic segmentation of brain tumor for overall survival prediction. In: Brainlesion: glioma, multiple sclerosis, stroke and traumatic brain injuries. Cham, Switzerland: Springer International Publishing; 2021. p. 215–27. doi:10.1007/978-3-030-72087-2_19.
53. Ghaffari M, Samarasinghe G, Jameson M, Aly F, Holloway L, Chlap P, et al. Automated post-operative brain tumour segmentation: a deep learning model based on transfer learning from pre-operative images. *Magn Reson Imaging*. 2022;86:28–36. doi:10.1016/j.mri.2021.10.012.
54. Guan X, Yang G, Ye J, Yang W, Xu X, Jiang W, et al. 3D AGSE-VNet: an automatic brain tumor MRI data segmentation framework. *BMC Med Imaging*. 2022;22(1):6. doi:10.1186/s12880-021-00728-8.
55. Zhu Z, Wang Z, Qi G, Zhao Y, Liu Y. Visually stabilized mamba U-shaped network with strong inductive bias for 3-D brain tumor segmentation. *IEEE Trans Instrum Meas*. 2025;74(6):2518511. doi:10.1109/TIM.2025.3551581.
56. Zhang M, Sun Q, Han Y, Zhang J. Edge-interaction mamba network for MRI brain tumor segmentation. In: ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2025 Apr 6–11; Hyderabad, India. p. 1–5. doi:10.1109/ICASSP49660.2025.10889470.