SSRG International Journal of Computer Science and Engineering
ISSN: 2348–8387 / https://doi.org/10.14445/23488387/IJCSE-V13I6P101

Volume 13 Issue 6, 1-11, June 2026

Original Article

Transformer-based Hybrid UCTTransNet–EfficientNet–PVT Framework with NAdam for Brain Tumor Detection and Segmentation

S. Gayathri[1], Santhi Baskaran[2]

[1,2]*Department of Information Technology, Puducherry Technological University, India.*

Corresponding Author : 2401712002@ptuniv.edu.in

Received: 23 April 2026 Revised: 26 May 2026 Accepted: 10 June 2026 Published: 28 June 2026

Abstract *- The detection and segmentation of brain tumors are crucial but difficult to achieve because of the heterogeneity of tumors and complicated MRI images. This paper proposes a hybrid framework that combines a transformer with UCTTransNet, EfficientNet, and the Pyramid Vision Transformer (PViT), along with NAdam optimization, to enhance performance. Existing CNN models lack global context, whereas transformer models require high computational cost, which is inefficient and reduces accuracy. The suggested approach involves using EfficientNet to learn strong spatial features and PViT to learn multi-scale contextual dependencies via attention mechanisms. UCTTransNet is more efficient for segmentation because it combines convolutional and transformer encoders, which ensure accurate boundary identification. NAdam optimization speeds up convergence and stabilization of training. Normalization and augmentation are also preprocessing methods that enhance model robustness. The goal is to create a high-performance, computationally efficient framework that improves detection accuracy and segmentation quality while addressing the limitations of existing methods. Experimental results show higher accuracy, a higher Dice coefficient, and better generalization. The hybrid combination greatly enhances the precision of feature representation and segmentation, which is why such a model can be used in automated clinical diagnosis and decision support systems.*

Keywords *- Brain Tumor Detection, Brain Tumor Segmentation, Transformer-Based Hybrid Framework, Medical Image Analysis, NAdam Optimizer, Attention Mechanism, Feature Extraction, Semantic Segmentation, Deep Learning in Healthcare.*

1. Introduction

The human brain, which is contained in the skull, is a very complex organ that controls thought, emotion, perception, and behavior with billions of neurons that are interconnected with electrical and chemical signals. It has a structure comprising specialized areas, such as the cerebral cortex, which is responsible for higher cognitive functions and consciousness, and the cerebellum, which controls balance and coordination [1]. The ability of these regions to work together will facilitate the smooth running of day-to-day activities and the environment's adaptive behavior. Even though the brain is strong, it is susceptible to pathologies such as tumors and other neoplasms.

A brain tumor is an unnatural growth of cells that can be located inside the brain tissue or can be a continuation of other body parts by metastasis [2]. There are general categories of tumors: benign and malignant. Benign tumors develop at a slow rate and are localized, but they may still cause severe complications due to the fact that they place pressure on vital brain structures. Conversely, malignant tumors are violent and able to spread to other parts and cause them to invade nearby tissues [3].

There are diverse types of brain tumors based on cell origin. Gliomas, such as the most aggressive glioblastomas, are glial tumors and require specific treatment plans. The protective membranes of the brain develop into meningiomas, which are usually benign. Pituitary adenomas affect hormonal balance, whereas schwannomas arise from nerve-supporting cells [4]. Brain tumors are correlated with the abnormal proliferation of cells and damage to DNA, which affects normal brain functioning and may result in serious neurological deficits or death [5].

To identify tumor type and efficient approaches to treatment, accurate diagnosis, usually through imaging and biopsy, is imperative. This paper discusses the brain tumor detection by MRI via contour-based transfer learning on the VGG-16 model [6]. Dropout layers help minimize overfitting, and threshold segmentation helps improve feature extraction. The model has great generalization and strength, and it is

capable of providing good classification and assisting the medical professionals in proper diagnosis and treatment decisions [7].

The main motive of this study is to develop an automated brain tumor detection and segmentation architecture using MRI images. The presented work aims to improve feature representation by combining EfficientNet and Pyramid Vision Transformer (PVT) to acquire both local and global image characteristics efficiently. It also focuses on designing a CNN-LSTM-based model to learn spatial and sequential information from MRI slices for enhanced prediction accuracy. Additionally, the study seeks to improve segmentation performance and robustness under noisy and low-contrast imaging conditions.

- First, improves image quality by guided filtering for noise removal and CLAHE for contrast improvement.
- For feature learning, it uses the CNN with LSTM to capture both spatial and contextual dependencies, while EfficientNet is used to extract global features, and PVT is employed to model long-range contextual information.
- Lastly, the model is optimized using the NAdam optimizer to improve convergence and stability. Overall, the integration of convolutional, recurrent, and transformer-based components enables more accurate and robust tumor segmentation performance.

2. Literature Survey

Medical image analysis is very important in clinical diagnosis, especially for detecting and analyzing brain tumors. Proper identification, classification, and diagnosis will lead to proper treatment planning, surgery, and therapy. Current hospital workflows are based on manual interpretation of MRI scans by radiologists, making it time-consuming, expensive, and reliant on human experience [8]. Since the number of specialists is usually less than the growing number of medical images, small tumors may sometimes go undetected. The acquisition of images and other noise can impair diagnostic accuracy, and thus, the necessity of automated, efficient systems is emphasized [9].

In a bid to tackle these issues, scientists have delved into novel methods of image processing and deep learning. MRI has been extensively applied to tumor detection, and preprocessing techniques, including noise reduction with median filters, are required to improve image quality [10]. CNNs are popular due to their ability to automatically extract features and their speed compared to current techniques. CNNs are more demanding in terms of large datasets, substantial computational power, and hyperparameter optimization, and therefore are inefficient on small medical datasets, which are likely to overfit [11]. Transfer learning has become one of the most promising solutions for enhancing accuracy and reducing training time. There are many types of segmentation techniques, among which are clustering, model-based, and neural network (FCNs and U-Net), which are designed to identify the exact position of tumors [12]. Although this has been achieved, there are still limitations, such as the high cost of computation, limited data availability, and poor boundary detection. Consequently, an effective hybrid model based on CNN and transformer models to extract local and global features, improve segmentation accuracy, and achieve faster, stronger performance is required [13].

Despite progress in brain tumor detection by MRI, current methods still face key limitations like dependence on manual interpretation, high computational cost, and limited performance on small datasets. CNN-based models often struggle with overfitting and insufficient generalization when existing segmentation methods fail to accurately capture fine tumor boundaries. Most existing frameworks focus on either local or global feature extraction, but not both effectively. Therefore, there is a need for a hybrid model that integrates local and global feature learning to improve accuracy and robustness in tumor detection.

3. Materials and Methods

The proposed hybrid architecture of Transformer-based UCTTransNet-EfficientNet-PViT uses BraTS MRI data (T1, T2, FLAIR) with annotated tumor masks, as illustrated in Figure 1. Preprocessing involves skull stripping, normalization, resizing, and augmentation. EfficientNet is used to extract spatial features, and the PViT is used to capture global context. UCTTransNet improves segmentation using convolutional layers and skip connections. The model does end-to-end tumor detection and segmentation. The NAdam optimizer, with cross-entropy, is used to solve the class imbalance training. Accuracy, Dice coefficient, precision, recall, and IoU are used to evaluate the performance.

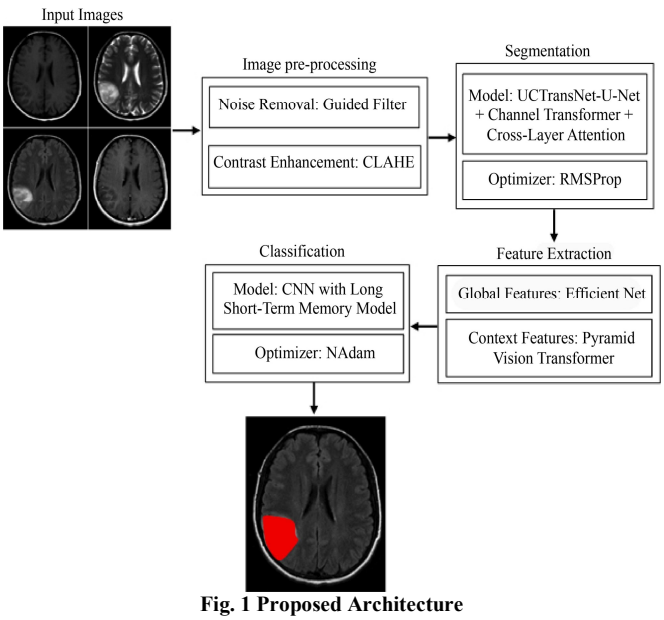


Fig. 1 Proposed Architecture

3.1. Dataset Description

BraTS is a common brain tumor benchmark dataset, a collection of multimodal MRI scans that are manually annotated by experts and presented in Table 1. It consists of various imaging modalities, which portray complementary tumor features. The challenges faced with the dataset are the class imbalance and variation in the shape and size of tumors. They need to be properly preprocessed and augmented to enhance the model generalization and the accuracy of segmentation.

Table 1. Dataset Description

Parameter	Description
Dataset Name	Brain Tumor Segmentation (BraTS) Dataset
Data Type	Multimodal MRI Images
Modalities	T1, T1-Contrast (T1c), T2, FLAIR
Image Format	NIfTI (.nii.gz)
Number of Patients	~300–500 (varies by BraTS version)
Image Dimensions	Typically 240 × 240 × 155 voxels
Annotations	Expert-labeled tumor regions (Enhancing Tumor, Tumor Core, Whole Tumor)
Classes	Background, Necrosis, Edema, Enhancing Tumor
Preprocessing	Skull stripping, normalization, resizing, and augmentation
Data Split	Training (70%), validation (15%), Testing (15%)
Challenges	Class imbalance, noise, intensity variation, tumor heterogeneity
Usage Purpose	Tumor detection, segmentation, and classification
Evaluation Metrics	Dice Coefficient, Accuracy, IoU, Precision, Recall

3.2. Pre-Processing

Image preprocessing is necessary in order to enhance the quality of MRI and the proper segmentation of brain tumors. First, the guided filter is used to remove noises and smooth the image without destroying significant edge structures. The output image X_{out} is modeled as a linear transform of the guidance image G within a local window w_k:

$$X_{out}(x) = a_k G(x) + b_k, \ \forall x \in w_k \quad (1)$$

Where the coefficients a_k and b_k, are computed as:

$$a_k = \frac{\sigma_k^2}{\sigma_k^2 + \epsilon} \quad (2)$$

$$b_k = \mu_k - a_k \mu_k \quad (3)$$

Here, μ_k and σ_k^2 of represent the mean and variance in the local window, and e is a regularization parameter controlling smoothing.

After denoising, CLAHE is applied to enhance local contrast and make tumor regions more distinguishable. The enhanced image is obtained using:

$$X_{CLAHE}(i,j) = \frac{CDF(X(i,j)) - CDF_{min}}{CDF_{max} - CDF_{min}} \quad (4)$$

Here, CDF is the cumulative distribution of pixel intensities in each local tile. Contrast limiting helps to cut off excessively amplified noise by limiting the histogram at a specified threshold. Guided filtering and CLAHE are able to increase the clarity of images, maintain structural features, and enhance the visibility of tumors, which can be used to extract features and segment them more accurately.

3.3. Segmentation

Brain tumor segmentation is a critical process in medical image analysis that involves identifying and delineating tumor regions from MRI scans to support accurate diagnosis and treatment planning (Figure 2).

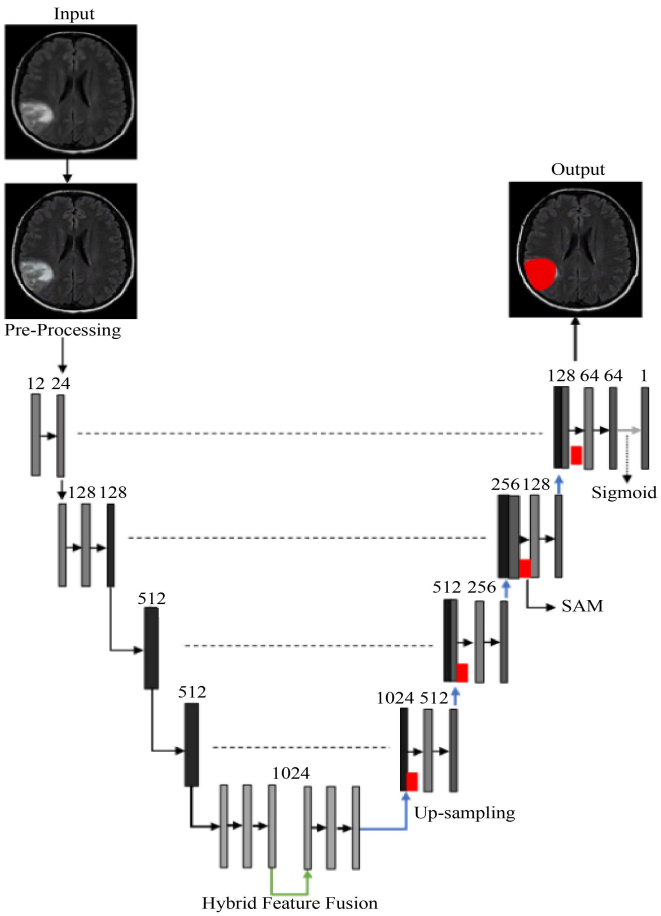


Fig. 2 Architecture of Segmentation using U-Net with channel transformer

It aims to classify each pixel or voxel into different categories, such as background, edema, tumor core, and enhancing tumor.

The proposed Hybrid U-Net and Channel Transformer architecture performs brain tumor segmentation by integrating convolutional feature extraction with transformer-based attention for enhanced representation learning.

The process begins with the input MRI image I, which is passed through an encoder composed of convolutional layers to extract hierarchical spatial features:

$$F_{enc} = f_{enc}(I) \quad (5)$$

where $f_{enc}(.)$ represents stacked convolution, activation, and pooling operations. These features are then processed by the Channel Transformer, which models inter-channel dependencies using multi-head self-attention. The attention mechanism is defined as:

$$Q = F_{enc}W_Q \quad (6)$$

$$K = F_{enc}W_K \quad (7)$$

$$V = F_{enc}W_V \quad (8)$$

where Q, K, and V denote query, key, and value matrices, and d is the scaling factor.

The resulting features are fused with convolutional outputs through hybrid feature fusion:

$$F_{fusion} = \alpha F_{enc} + \beta F_{attn} \quad (9)$$

where α and β are learnable weights. Cross-layer attention further refines feature propagation between encoder and decoder:

$$F_{corss} = Attention(F_x, F_y) \quad (10)$$

In the decoder, upsampling operations reconstruct the spatial resolution:

$$F_{dec} = f_{up}(F_{fusion}) \quad (11)$$

Finally, the segmentation mask is obtained using a sigmoid activation:

$$S(i,j) = \sigma(F_{dec}) \quad (12)$$

This architecture effectively captures both local spatial details and global contextual relationships, resulting in precise tumor boundary detection and improved segmentation accuracy.

3.4. Feature Extraction (Global Features using EfficientNet)

The proposed EfficientNet-based feature extraction architecture is designed to capture rich global representations from brain MRI images for accurate tumor analysis, as shown in Figure 3. Each MBConv block employs depthwise separable convolution and Squeeze-And-Excitation (SE) attention to emphasize important feature channels while reducing computational complexity. The transformation can be expressed as:

$$F_{x+1} = MBConv(F_x) \quad (13)$$

Where F_x represents the input feature map at stage x. Multi-scale feature extraction is achieved across successive MBConv layers, enabling the model to capture both fine-grained and high-level tumor characteristics. Feature map aggregation combines outputs from different layers to produce a global feature representation:

$$F_{global} = \sum_{x=1}^{n} w_x F_x \quad (14)$$

where w_x are learnable weights. This global feature output effectively encodes tumor structure, intensity, and contextual information, enhancing downstream segmentation performance.

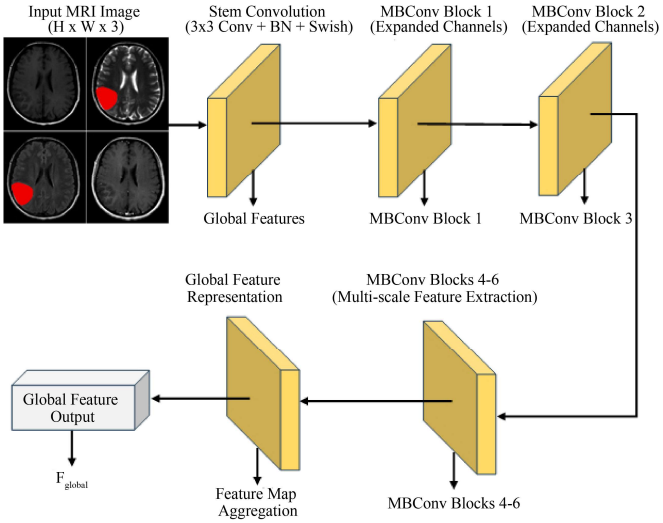


Fig. 3 Architecture of Feature Extraction (Global features using efficientNet)

3.5. Feature Extraction (Context Features using Pyramid Vision Transformer)

Contextual feature extraction in the proposed framework is performed using the PViT, which is designed to capture global dependencies and multi-scale contextual information from brain MRI images (Figure 4). The input feature map $F \in R^{H\times W\times C}$ First, it is divided into overlapping patches and embedded into a lower-dimensional space:

$$Z_0 = \text{PatchEmbed}(F) \quad (15)$$

The pyramid structure progressively aggregates contextual information across different scales:

$$F_{context} = \cup_{x=1}^{L} Z_x \quad (16)$$

Where Z_x represents features from each transformer stage. This approach enables the model to effectively capture global context and spatial relationships, significantly improving the identification of complex and irregular tumor regions in MRI images.

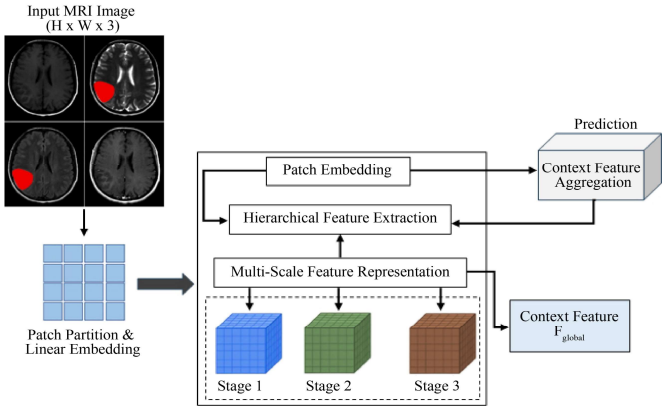


Fig. 4 Architecture of Feature Extraction (Context Features using PViT)

3.6. Classification using CNN with LSTM

The classification stage in the proposed framework employs a hybrid CNN–LSTM model to accurately classify brain tumor regions using extracted spatial and temporal features. Initially, the CNN processes the input feature maps obtained from previous stages to learn discriminative spatial features such as edges, textures, and tumor patterns. The LSTM effectively retains important information through its memory cells and gating mechanisms, improving classification performance for complex tumor structures. The CNN extracts features as:

$$F = \text{CNN}.(I) \quad (17)$$

The LSTM then updates its hidden state using:

$$h_t = \sigma(W_h.[h_{t-1}, i_t] + b_h) \quad (18)$$

where h_t is the hidden state, i_t is the input sequence, and W_h, b_h These are learnable parameters. The final classification output is obtained using a softmax function:

$$j = \text{Softmax}(W_o h_t + b_o) \quad (19)$$

This hybrid approach improves classification accuracy by combining spatial feature extraction with sequential learning, making it highly effective for brain tumor diagnosis.

3.7. Optimizer using NAdam

The NAdam optimizer is employed in the proposed framework to enhance training efficiency and convergence stability. NAdam combines the advantages of Adam optimization with Nesterov momentum, enabling faster and more accurate gradient updates. It maintains exponentially decaying averages of both the gradients and their squared values to adaptively adjust learning rates for each parameter.

The first and second moment estimates are computed as:

$$m_t = \beta_1 m_{t-1} + (1-\beta_1) g_t \quad (20)$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2) g_t^2 \quad (21)$$

where g_t the gradient at time step t, and β_1, β_2 are decay rates. Bias correction is applied as:

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t} \quad (22)$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t} \quad (23)$$

NAdam incorporates Nesterov momentum into the update rule:

$$\theta_{t+1} = \theta_t - \eta \frac{\beta_1 \hat{m}_t + \frac{(1-\beta_1) g_t}{1-\beta_1^t}}{\sqrt{\hat{v}_t} + \epsilon} \quad (24)$$

where θ_t represents model parameters, and η is the learning rate. This optimizer improves convergence speed, reduces oscillations, and enhances overall model performance in brain tumor detection and segmentation tasks.

3.8. Algorithm

Input: MRI images *I*

Output: Segmentation mask *S*, classification label *j*

Step 1: Image Preprocessing

Guided filtering using Equation (1)

CLAHE enhancement using Equation (4)

Normalization: $I_{norm} = \frac{I-\mu}{\sigma}$ (25)

Where: μ - mean intensity, σ - standard deviation

Step 2: Efficient Net Feature Extraction

Global feature extraction using Equation (5)

MBConv operation:

$$F' = \sigma\left(Conv_{1\times1}\left(DWConv\left(F_{global}\right)\right)\right) \quad (26)$$

Where: F' – transformed features, σ – activation function

Step 3: PViT Context Feature Extraction

Patch embedding using Equation (15)

Self-attention: $F' = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V$ (26)

Where: Q-query, K-key, V-value, d-scaling factor
Context aggregation using Equation (16)

Step 4: Feature Fusion

Hybrid fusion: $F_{fusion} = \alpha F_{global} + \beta F_{context}$ (27)

Where: α, β - fusion weights

Step 5: UCTTransNet Segmentation

Encoder output: $F_{enc} = f_{enc}(F_{fusion})$ (28)

Where: f_{enc} - encoder function

Channel attention: $F_{attn} = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V$ (29)

Where: F_{attn} - enhanced features
Decoder reconstruction using Equation (11)
Segmentation output using Equation (12)

Step 6: CNN-LSTM Classification

CNN feature extraction using Equation (17)
LSTM update using Equation (18)
Output prediction using Equation (19)

Step 7: NAdam Optimization

Gradient computation: $g_t = \nabla_\theta L(\theta_t)$ (30)

Where: g_t gradient, L-loss function
Moment updates using Equations (20) and (21)
Parameter update using Equation (24)

Step 8: Final Output: Return segmentation mask S and classification result j

The proposed UCTTransNetPViT EfficientNet framework preprocesses brain MRI images by denoising, contrast enhancement, and normalization. EfficientNet learns global spatial features, whereas PViT learns multi-scale contextual dependencies. These characteristics are combined and fed to UCTTransNet to segment tumors precisely with attention and skip connections. A CNNLSTM classifier is then used to classify tumor types, and NAdam optimization is used to guarantee consistent convergence and better performance.

4. Results and Discussions

The proposed framework is executed in an environment with a high-performance computer with an Intel multi-core processing unit, an NVIDIA CUDA-enabled graphics card (e.g., RTX series) to perform the accelerated deep learning computations, and with enough RAM (16 GB or more) to work with large MRI datasets without excessive effort. It is developed with Python along with deep learning libraries, such as TensorFlow or PyTorch, with supporting libraries, including NumPy, OpenCV, and Scikit-learn, to preprocess, handle features, and evaluate. Experiments are performed on publicly available brain MRI data, which are divided into training, validation, and testing sets. Mini-batch optimization with proper batch size, learning rate schedule, and early stopping to avoid overfitting is used to train the models. The measures of precision, accuracy, recall, F1-score, and Dice coefficient are used to measure performance in terms of classification and segmentation.

Table 2. Hyper-parameter Settings

Parameter	Value
Input Image Size	224 x 224
Batch Size	16
Learning Rate	0.0001
Optimizer	NAdam
Epochs	100
Dropout Rate	0.5
Weight Decay	0.00001
Momentum	0.9
Loss Function	Categorical Crossentropy
Early Stopping Patience	10
Train Validation Split	80 20
Activation Function	ReLU
Final Activation	Softmax

The hyperparameters have been set in such a way that the proposed model is trained in a stable and efficient manner, as indicated in Table 2. The image size is 224x224, the batch size is 16, and the learning rate is 0.0001. It is optimized with the NAdam optimizer, 100 epochs, early stopping, an 80:20 train validation split, and ReLU activation, and a dropout rate of 0.5.

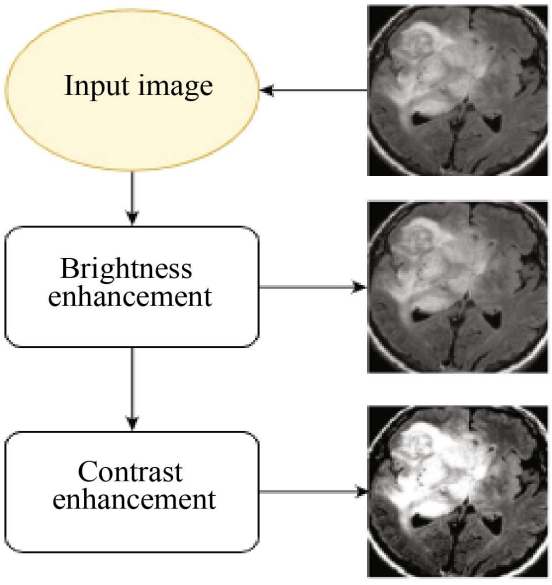
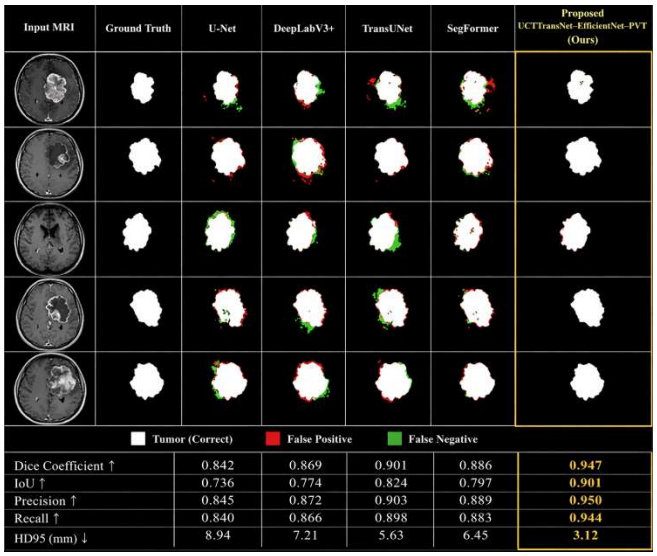


Fig. 5 Results of pre-processing

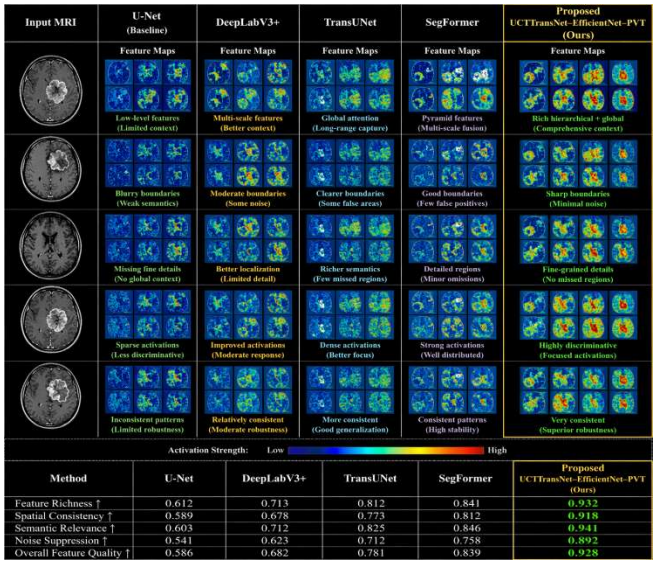


	U-Net	DeepLabV3+	TransUNet	SegFormer	Proposed UCTTransNet–EfficientNet–PVT (Ours)
Dice Coefficient ↑	0.842	0.869	0.901	0.886	**0.947**
IoU ↑	0.736	0.774	0.824	0.797	**0.901**
Precision ↑	0.845	0.872	0.903	0.889	**0.950**
Recall ↑	0.840	0.866	0.898	0.883	**0.944**
HD95 (mm) ↓	8.94	7.21	5.63	6.45	**3.12**

Fig. 6 Comparison of Segmentation of the proposed and existing systems

The enhancement module of the MRI image appears in Figure 5. Brightness and contrast enhancement enhance raw images that have low brightness and contrast. This process causes clarity, accentuates anatomical features, and minimizes noise. The improved image is used to give a higher input to feature extraction, segmentation, and classification, enhancing the overall diagnostic accuracy and strength.

The segmentation outcomes indicate that UCTTransNetEfficientNetPViT framework yields masks that are close to the ground truth with low false positives and false negatives, as indicated in Figure 6. It has a better Dice score, IOU, and segmentation accuracy overall as compared to four systems, which have been tested to have sharper tumor boundaries and better region consistency.



Method	U-Net	DeepLabV3+	TransUNet	SegFormer	Proposed UCTTransNet–EfficientNet–PVT (Ours)
Feature Richness ↑	0.612	0.713	0.812	0.841	**0.932**
Spatial Consistency ↑	0.589	0.678	0.773	0.812	**0.918**
Semantic Relevance ↑	0.603	0.712	0.825	0.846	**0.941**
Noise Suppression ↑	0.541	0.623	0.712	0.758	**0.892**
Overall Feature Quality ↑	0.586	0.682	0.781	0.839	**0.928**

Fig. 7 Comparison of Feature Extraction of the proposed and existing systems

Figure 7 contrasts the outcomes of feature extraction of U-Net, DeepLabV3+, TransUNet, and SegFormer with the proposed framework. The traditional models are blurred, noisy, and do not have details. Transformer models are better at attention, and yet contain small errors. The suggested UCTTransNet-Efficient-PViT framework is a sharper, consistent, and noise-free feature map with enhanced tumor localization and strength.

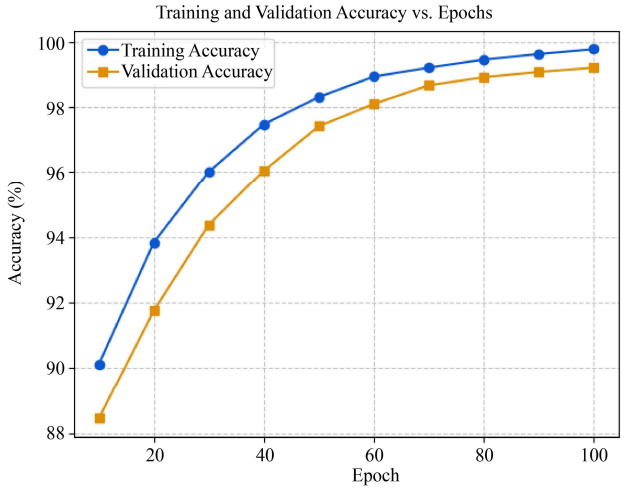


Fig. 8 Comparison of training and validation accuracy vs. epochs of the proposed and existing systems

Figure 8 depicts trends of training and validation accuracy of the proposed UCTTransNetEfficientNetPViT model with respect to epoch. The minor and steady difference between the two curves also proves that there is a stable convergence with the help of NAdam optimization and justifies the strength and the credibility of the suggested framework.

Figure 9 demonstrates the change in training and validation loss with the epochs of the proposed UCTTransNetEfficientNetPViT framework. The smooth convergence highlights the effectiveness of NAdam optimization and the model's ability to learn generalized and robust feature representations for accurate brain tumor detection and segmentation.

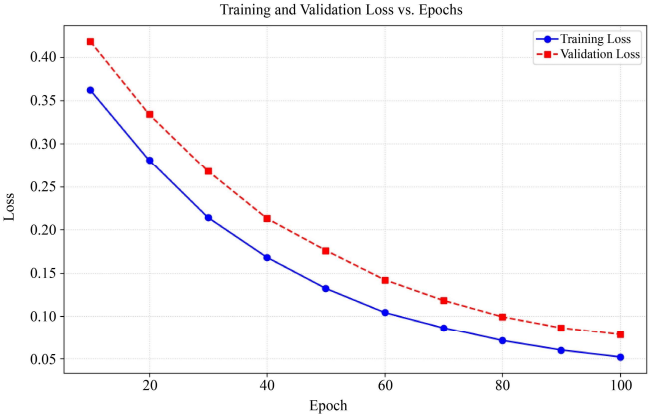
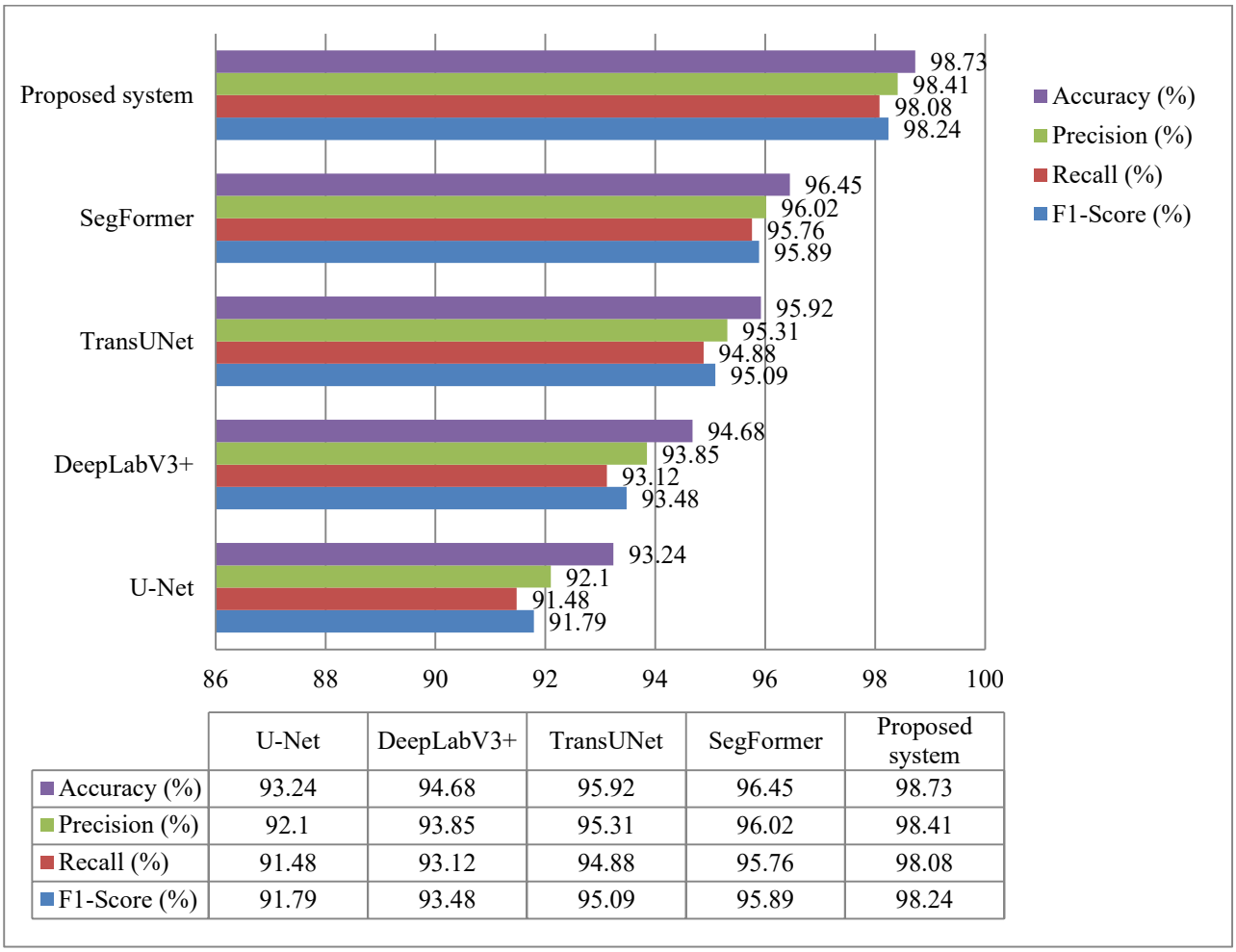


Fig. 9 Comparison of training and validation loss vs. epochs of the proposed and existing systems



	U-Net	DeepLabV3+	TransUNet	SegFormer	Proposed system
Accuracy (%)	93.24	94.68	95.92	96.45	98.73
Precision (%)	92.1	93.85	95.31	96.02	98.41
Recall (%)	91.48	93.12	94.88	95.76	98.08
F1-Score (%)	91.79	93.48	95.09	95.89	98.24

Fig. 10 Comparison of performance measures of the proposed and existing systems

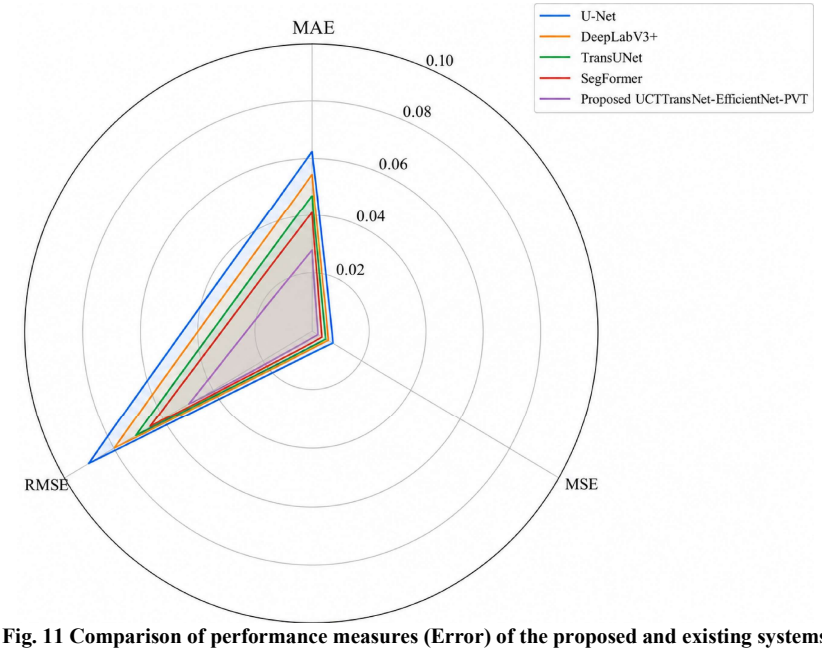


Fig. 11 Comparison of performance measures (Error) of the proposed and existing systems

Figure 10 shows that the proposed UCTTransNetEfficientNetPViT framework outperforms the existing models in terms of classification.

The high F1-score shows the best trade-off between precision and recall, and the usefulness and clinical reliability of the suggested hybrid transformer-based method.

Figure 11 is a comparison of error-based performance measures of the proposed and existing systems. The UCTTransNetEfficientNetPViT framework proposed has the lowest MAE, MSE, and RMSE, which implies low prediction errors and high consistency of the output. These findings reveal a higher precision and strength of the proposed model in regression than the current CNN and transformer-based models.

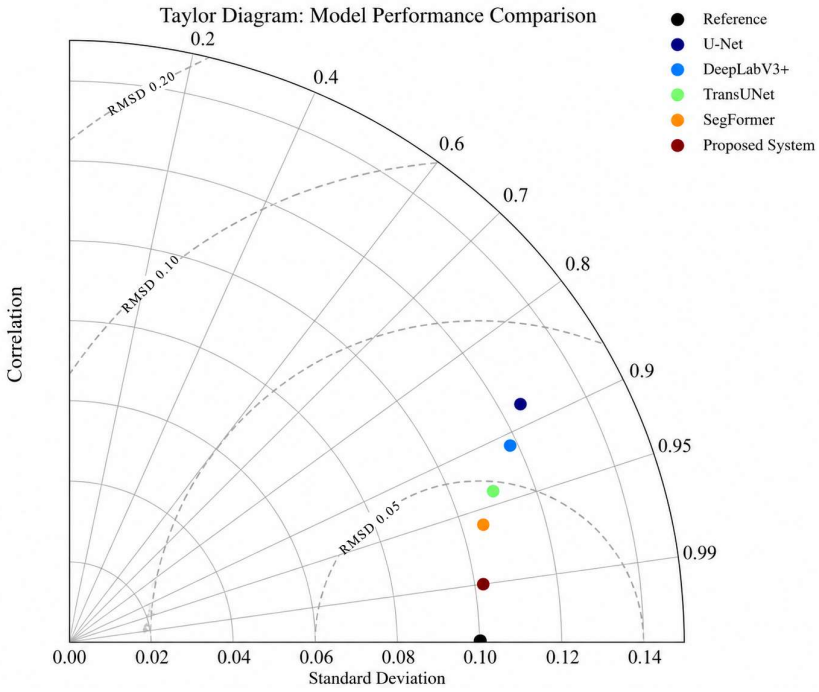
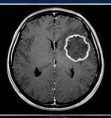
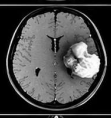
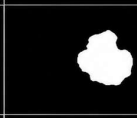
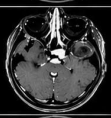
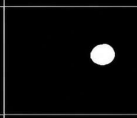
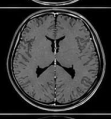
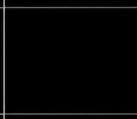
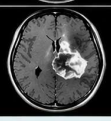


Fig. 12 Comparison of the Taylor plot of the proposed and existing systems

Input MRI	Ground Truth (Mask)	Predicted Class	Class Probabilities (%)	Result
		Glioma (Grade II)	Glioma 96.21 Meningioma 2.14 Pituitary 0.87 No Tumor 0.78	Correct
		Meningioma (Grade I)	Glioma 3.11 Meningioma 94.63 Pituitary 1.23 No Tumor 1.03	Correct
		Pituitary Tumor	Glioma 1.45 Meningioma 2.34 Pituitary 93.67 No Tumor 2.54	Correct
		No Tumor	Glioma 0.89 Meningioma 1.21 Pituitary 1.09 No Tumor 98.81	Correct
		Glioma (Grade IV)	Glioma 97.32 Meningioma 1.32 Pituitary 0.71 No Tumor 0.65	Correct

Overall Performance (Proposed System)	Accuracy	Precision	Recall	F1-Score	Mean AUC
	98.73%	98.41%	98.08%	98.24%	0.988

Fig. 13 Comparison of overall performance with classification results of the proposed and existing systems

Figure 12 shows a relative performance assessment of a model through a joint analysis of correlation, standard deviation, and root-mean-square deviation. The system closest to the reference point is the proposed system, which means that it is most correlated with ground truth and has the smallest error. This supports the high stability, precision, and overall generalization ability of the framework proposed to classify brain tumors.

The classification performance of the proposed system on MRI images is shown in Figure 13, and it is clear that the ground-truth masks and predicted results are separated. The visualization highlights the accuracy of the labels of the classes instead of the overlap of the segmentation by removing the predicted masks. In all samples, the probabilities of confidence classes are high, and all the glioma, meningioma, pituitary tumor, and no-tumor cases have been classified correctly. High accuracy, precision, recall, and F1-score are consistent, which proves the strength and reliability of the suggested classification framework and its clinical relevance.

Table 3 indicates the ablation study of the proposed method; the EfficientNet + UCTTransNet model attains less performance, with an accuracy of 94.79%, precision of 94.58%, Recall of 93.85%, and F1-score of 94.49%. Likewise, the PViT + UCTTransNet technique got a slightly higher performance, Accuracy of 95.29%, a precision of 95.17%, Recall of 94.59%, and an F1-score of 94.99%. Moreover, the EfficientNet + PViT (without UCTTransNet) and EfficientNet + UCTTransNet (without PViT) models achieves a moderated performance. Finally, the proposed system (Preprocessing + EfficientNet + PViT + UCTTransNet + NAdam) method got a better performance, Accuracy of 98.73%, a precision of 98.41%, Recall of 98.08%, and an F1-score of 98.24%.

Table 3. Ablation study of the proposed approach

Model	Accuracy	Precision	Recall	F1-Score
EfficientNet + UCTTransNet	94.79	94.58	93.85	94.49
PViT + UCTTransNet	95.29	95.17	94.59	94.99
EfficientNet + PViT (without UCTTransNet)	96.02	95.79	95.37	95.62
EfficientNet + UCTTransNet (without PViT)	96.70	96.48	96.10	96.42
PViT + UCTTransNet (without EfficientNet)	97.43	97.11	96.76	96.92
EfficientNet + PViT + UCTTransNet (without Nadam optimizer)	98.10	97.71	97.36	97.47
Proposed system (Preprocessing + EfficientNet + PViT + UCTTransNet + NAdam)	98.73	98.41	98.08	98.24

These experimental evaluations illustrate the proficiency of the presented UCTTransNet–EfficientNet–PVT architecture for brain tumor segmentation. The superior performance could be attributed to the combined learning of local spatial features and global contextual information, supported by improved preprocessing and efficient optimization using NAdam. These techniques collectively contribute to enhanced segmentation accuracy, robustness, and generalization when compared to previous state-of-the-art techniques.

5. Conclusion

In this work, a Transformer-Based Hybrid UCTTransNet-EfficientNet-PViT system optimized using NAdam was used to detect and segment brain tumors in MRI images with high accuracy. The proposed system has demonstrated the ability to effectively extract both the local and global features by using EfficientNet with effective spatial feature extraction, multi-scale contextual attention of PViT, and the hybrid convolution-transformer segmentation ability of UCTTransNet. Experimental performance is much better and more accurate in classification with 98.73% accuracy, 98.24% F1-score, and a high Dice coefficient of 0.98, and significant errors are reduced (MAE = 0.028, RMSE = 0.051). NAdam optimization provided faster convergence and stability in training. Altogether, the suggested framework performed better than the current CNN and transformer-based models, which indicates its strength, generalization capacity, and applicability in the reliable automated clinical decision support in the diagnosis of brain tumors. Although the presented architecture attains high segmentation performance, its effectiveness might be affected by the availability and diversity of MRI datasets. Additionally, performance may differ when applied to images acquired from different scanners or clinical settings, which could affect model generalization. Future research will focus on validating the proposed model using larger multi-center datasets to improve its robustness and generalizability. Furthermore, advanced explainable AI techniques and lightweight architectures will be discovered to improve interpretability and support real-time clinical deployment.

Declarations

Competing Interests and Funding

The authors did not receive support from any organization for the submitted work.

Funding Statement

The authors received no funding for this study.

Conflict of Interest

The authors declare that they have no conflict of interest.

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

References

[1] Manoj Kumar Tyagi et al., *Hybrid Deep-Cuckoo Framework for Robust Brain Tumor Detection and Segmentation in MRI Scans*, Smart Technologies and Intelligent Computing, 1st ed., CRC Press, pp. 1-7, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[2] Hira Yousaf et al., "Advanced CNN-Based Brain Tumor Detection and Segmentation Using MATLAB: A Diagnostic Accuracy Study," *Pakistan Journal of Medical & Cardiological Review*, vol. 5, no. 1, pp. 535-549, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ponlatha Sambandham et al., "Brain Tumor Detection using HyGSNet and Feature Extraction with DWT-based GDP," *Journal of Neuroimmunology*, vol. 413, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[4] Homayoun Safarpour et al., "Explainable Deep Learning Framework for Brain Tumor Segmentation using Vision Transformer and Conditional Random Fields," *Multimedia Systems*, vol. 32, pp. 1-46, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[5] Evgin Goceri, and Yuzi D. Winter, "CFATrans: Brain Tumor Segmentation from MRIs using Consecutive Fusion-Attention Transformer with Convolutional Networks and a Composite Loss Function," *Biomedical Signal Processing and Control*, vol. 112, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[6] A. Srinivasa Reddy et al., "T-GAN: Transformer Generative Adversarial Network for Brain Tumor Segmentation," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 40, no. 3, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[7] Arshleen Kaur et al., "BrainDx: A Dual-Transformer Framework using PVT and SegFormer for Tumor Diagnosis," *Biomedical Signal Processing and Control*, vol. 113, pp. 1-19, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[8] Shakhnoza Muksimova, Jushkin Baltaev, and Young Im Cho, "Brain Tumor Segmentation with Contextual Transformer-Based U-Net," *Electronics*, vol. 15, no. 4, pp. 1-17, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ameer Hamza, and Robertas Damaševičius, "Deep Learning for Brain Tumor Segmentation and Classification: A Systematic Review of Methods and Trends," *Computers, Materials and Continua*, vol. 86, no. 1, pp. 1-41, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[10] Wessam M. Salama, and Moustafa H. Aly, "Brain Tumor Segmentation and Classification: A CVAE-UNETR-ResNet50-VGG16 Hybrid Deep Learning Approach," *Alexandria Engineering Journal*, vol. 135, pp. 433-449, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[11] Yuhui Liao et al., "Trans-MT: A 3D Semi-Supervised Glioma Segmentation Model Integrating Transformer Architecture and Asymmetric Data Augmentation," *Displays*, vol. 93, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[12] G. M. Sasikala, and K. Anand, "ExU-Trans: A Self-Explanatory Transformer with U-Net based Hybrid Model for Brain Tumor Segmentation using MR Imaging," *Complex & Intelligent Systems*, vol. 12, pp. 1-25, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[13] Anita Murmu et al., "SU-FTD2: Transformer-Driven Brain Tumor Imaging Framework Using Explainable AI for Consumer Applications," *IEEE Transactions on Consumer Electronics*, vol. 72, no. 2, pp. 4632-4640, 2026. [CrossRef] [Google Scholar] [Publisher Link]