

Explainable Deep Learning Framework for Brain Tumor Segmentation Using Vision Transformer and Conditional Random Fields

Homayoun Safarpour

University of Szeged

Soroush Sadeghi

University of Tehran

Payam Zarbakhsh

Ada kent University

Mohammadreza Kamsari

Malek-Ashtar University of Technology(MUT)

Marjan Kia

Park University

Ramin Ranjbarzadeh

`ramin.ranjbarzadehkondrood@universityofgalway.ie`

University of Galway

Research Article

Keywords: Brain Tumor Segmentation; Vision Transformer; Grad-CAM; Conditional Random Fields

Posted Date: June 9th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-6404734/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Multimedia Systems on December 4th, 2025. See the published version at <https://doi.org/10.1007/s00530-025-02075-y>.

Explainable Deep Learning Framework for Brain Tumor Segmentation Using Vision Transformer and Conditional Random Fields

Homayoun Safarpour ¹, Soroush Sadeghi ², Payam Zarbakhsh ³, Mohammadreza Kamsari ⁴, Marjan Kia ⁵,
Ramin Ranjbarzadeh ^{6*} / ✉

¹ Department of Software Engineering, University of Szeged, Szeged, Hungary.

safarpour@inf.u-szeged.hu

² School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran.

Soroush.sadeghi@ut.ac.ir

³ Software Engineering Department, Ada kent University, Famagusta, North cyprus, Mersin 10 Turkey.

payam.zarbakhsh@adakent.edu.tr

⁴ Faculty of Electrical Engineering, Malek-Ashtar University of Technology(MUT), Esfahan, Iran.

mohammadrezakamsari@gmail.com

⁵ Department of Business, Information System Business Analysis (ISBA), Park university, United States.

MK.MARJAN.KIA@gmail.com

⁶ School of Computer Science, University of Galway, Ireland.

ramin.ranjbarzadehkondrood@universityofgalway.ie

* corresponding author

Abstract

Background/Objectives: Brain tumor segmentation is critical for accurate identification and treatment planning. However, delineating intricate tumor areas like Edema, Tumor Core, and Enhancing Tumor remains challenging owing to diversity in dimensions, morphology, and delineation. **Methods:** This introduces an innovative a novel deep learning model integrating a Pretrained Vision Transformer (ViT) for global context representation, Atrous Spatial Pyramid Pooling (ASPP) in both encoder and merging stages, and ReLU6. Additionally, attention mechanisms and Conditional Random Fields (CRF) were incorporated for refining feature emphasis and segmentation boundaries. **Results:** The model was trained and tested on the BraTS 2020 dataset and compared with state-of-the-art methods. The suggested model obtained the highest Dice coefficients across all tumor regions, with improvements of up to 4% for Enhancing Tumor and Edema. Grad-CAM heatmaps illustrated the model's proficiency in precisely identifying tumor areas, hence enhancing the interpretability of predictions. **Conclusions:** The incorporation of multi-scale feature extraction, attention mechanism, and global context representation markedly improved segmentation accuracy and interpretability. The findings underscore the proposed model's capacity to enhance clinical processes and establish a standard for brain tumor segmentation tasks. Future endeavors involve testing on varied datasets and investigating lightweight structures for real-time applications.

Keywords: Brain Tumor Segmentation; Vision Transformer; Grad-CAM; Conditional Random Fields.

1. Introduction

Brain tumors are one of the most critical and life-threatening conditions in humans, marked by atypical cellular proliferation within the brain that interferes with its regular operations [1], [2], [3]. They can be categorized as benign or malignant, with the latter presenting considerable hazards due to their aggressive characteristics. Timely and precise diagnosis is crucial for efficient treatment planning, rendering brain tumor segmentation a vital role in medical imaging [1], [4], [5].

Brain tumor segmentation entails the accurate detection and delineation of tumor areas using brain imaging data, including magnetic resonance imaging (MRI). This method is essential for multiple reasons: it aids in assessing the tumor's dimensions, location, and advancement, which

are critical factors for strategizing surgical procedures, radiation therapy, and evaluating treatment efficacy. Precise segmentation facilitates personalized treatment strategies, enhancing patient outcomes [6], [7], [8].

Traditionally, brain tumor segmentation has been conducted manually by radiologists, a process that is both time-intensive and susceptible to inter-observer variability. The growing accessibility of medical imaging data has highlighted the necessity for automated methods that yield consistent and reproducible outcomes. Automated segmentation methods seek to alleviate the burden on medical practitioners while maintaining uniform precision, even in intricate scenarios [9], [10], [11].

Recent developments in deep learning have transformed medical image analysis, particularly in brain tumor segmentation [12], [13], [14]. Deep learning models, especially convolutional neural networks (CNNs), have exhibited remarkable efficacy in extracting intricate information from medical images. Architectures such as U-Net, 3D CNNs, and encoder-decoder networks are frequently employed for tumor segmentation owing to their proficiency in effectively capturing spatial and contextual information. Nevertheless, current models continue to encounter numerous problems that limit their clinical usefulness [7], [15], [16], [17].

A significant constraint is the dependence on extensive labeled datasets for training, which can be challenging to acquire in the medical field. Moreover, deep learning models frequently encounter challenges due to the heterogeneity in tumor morphology, dimensions, and anatomical placements among patients, potentially resulting in diminished performance [18], [19], [20], [21]. Overfitting to the training data is a significant concern, particularly when working with small datasets, which are prevalent in medical imaging. Furthermore, numerous models approach tumor segmentation as a single-modality issue, depending on particular MRI sequences, while inadequately leveraging the advantages of multi-modality data that offer supplementary insights into the tumor [22], [23], [24], [25].

We present a U-Net-based segmentation model that integrates residual blocks, attention processes, and an Atrous Spatial Pyramid Pooling (ASPP) module to overcome these limitations. The residual blocks improve the model's capacity to capture hierarchical information and alleviate the vanishing gradient issue, facilitating superior performance on intricate tumor segmentation tasks [26], [27], [28]. Attention methods are utilized to direct the model's focus towards the most pertinent areas of the image, hence enhancing accuracy and resilience across diverse tumor

attributes. The ASPP module enhances the model by acquiring multi-scale contextual information, rendering it especially effective for tumors with irregular shapes and diverse sizes [13], [29].

Furthermore, we utilize Gradient-weighted Class Activation Mapping (Grad-CAM) to elucidate the model's decision-making process, thereby improving interpretability and fostering clinician confidence in the model's predictions. Grad-CAM generates heatmaps that emphasize the areas of the image that affect the model's output, facilitating a deeper comprehension of its functionality.

This research employs the BraTS 2020 dataset, comprising multi-modal MRI images, to validate our suggested methodology. Utilizing the dataset's extensive information and incorporating sophisticated architectural elements such as residual blocks, attention mechanisms, and the ASPP module, alongside explainability tools like Grad-CAM, our objective is to create a resilient and interpretable model for brain tumor segmentation. This methodology is expected to markedly enhance clinical decision-making and patient care results.

2. Related works

Zhu et al. [30] underscore the significance of exact 3D MRI brain tumor segmentation for clinical diagnosis and treatment, highlighting the necessity for correct localization and spatial distribution of brain tumors. To overcome the shortcomings of current methodologies that frequently neglect voxel-level intricacies, inter-layer linkages, and detailed attributes, they provide an innovative segmentation network called SDV-TUNet (Sparse Dynamic Volume TransUNet). This 3D encoder-decoder architecture utilizes volumetric data to produce dense predictions, integrating voxel information, inter-layer feature connections, and intra-axis information. SDV-TUNet utilizes a Sparse Dynamic (SD) encoder-decoder module that combines multi-head self-attention with a 3D extended shifting window approach, facilitating the extraction of global spatial semantic characteristics. The encoding phase integrates both local strong correlations and long-range sparse connections, improving the understanding of regional interactions. A Multi-Level Edge Feature Fusion (MEFF) module is introduced, which progressively integrates multi-level edge characteristics layer by layer. This module employs skip connections to interface with the decoder, enhancing the transmission of spatial edge information and boosting segmentation precision. Experimental results indicate that SDV-TUNet proficiently integrates global and local data, surpassing conventional approaches in addressing the intricacies of 3D MRI brain tumor segmentation tasks.

Sharif et al. [31] introduced a thorough methodology for identifying malignant brain tumors and precisely localizing aberrant regions in Magnetic Resonance Imaging (MRI) scans. Their methodology consists of a four-stage workflow that includes tumor preprocessing, segmentation, feature extraction, and classification. During the preprocessing phase, the region of interest (ROI) is delineated via manual skull stripping, and noise effects are mitigated through the implementation of a Gaussian filter. The second stage concentrates on tumor segmentation, employing an enhanced thresholding technique that integrates binomial mean, variance, and standard deviation for accurate identification of tumor areas. During the third stage, a collection of geometric and texture features is retrieved to delineate the essential characteristics of the segmented tumor. The features are subsequently integrated using a serial-based approach, and the optimal subset is identified with the assistance of a Genetic Algorithm (GA) to improve classification performance. Classification is ultimately performed using a Support Vector Machine (SVM) with a linear kernel function, effectively differentiating between malignant and normal tissues. The incorporation of enhanced segmentation methods and feature optimization allows this strategy to attain reliable and precise tumor identification and classification. This methodology highlights its potential in enhancing MRI-based brain tumor analysis.

Feng et al. [32] examined the difficulties of brain tumor segmentation in MRI images, noting that substantial variability in tumor areas among people may result in misdiagnosis if dependent exclusively on clinical judgment. They introduced DAUnet, a U-shaped architecture aimed at accurate segmentation of brain tumor MRI images, utilizing deep supervision and convolutional attention. The architecture integrates advanced modules to address the shortcomings of conventional segmentation techniques that frequently neglect essential 3D information. The primary element of DAUnet is a Bottle-neck and Attention (BA) module, which combines spatial and channel (SC) attention with residual connections, termed 3D SC attention. This combination guarantees effective concentration on essential areas while maintaining crucial details over layers. A Convolution and Atrous Spatial Pyramid (CASP) module is developed to augment feature extraction. The CASP module enlarges the receptive field without sacrificing resolution, employing downsampling and upsampling processes to enhance layer-wise correlations. The network utilizes deep supervision as an auxiliary branch, integrating deep learning and regularization methods to enhance parameter refinement and model fitting. This approach allows DAUnet to get precise segmentation while managing the intricacies of 3D brain tumor MRI

images, providing a dependable resource for physicians in diagnosis and treatment planning.

Ullah et al. [11] suggested a hybrid approach that combines handcrafted features with CNNs to enhance segmentation performance. Handcrafted features, encompassing texture-based, intensity-based, and shape-based attributes, were retrieved from MRI scans to capture essential tumor characteristics. A unique CNN architecture was concurrently developed to autonomously extract features from the data, therefore effectively augmenting the handcrafted features. The suggested hybrid technique integrates these two paths into a novel CNN architecture, utilizing the advantages of both manually produced and automatically acquired features. The methodology was assessed using the Brain Tumor Segmentation (BraTS) challenge dataset with measures like segmentation sensitivity, accuracy, Dice score, and specificity.

Yang et al. [33] acknowledged the significance of multi-modal magnetic resonance imaging (MRI) and created F²Net, a versatile fusion network aimed at improving segmentation performance by utilizing the complementary information in multi-modal images while maintaining modality-specific traits. In contrast to traditional methods that concentrate exclusively on common representations, F²Net adeptly tackles the difficulties of efficiently integrating a variable number of modalities while preserving modality-specific characteristics. F²Net is constructed on an encoder-decoder architecture that integrates two Transformer-based feature extraction streams with a cross-modal shared learning network. To effectively integrate multi-modal information, the network utilizes a cross-modal feature-enhanced module (CFM) to merge modality-specific features into a shared representation and a multi-modal collaboration module (MCM) to include encoder-derived features into the shared decoder. This approach allows F²Net to adapt flexibly to varying numbers of modalities, extracting both unique and common feature representations.

3. Materials and Methods

This section systematically delineates the proposed methodology. The visual summary of the proposed model is depicted in [Figure 2](#).

3.1. Dataset

The research employs the BraTS 2020 dataset, a well-established standard for brain tumor segmentation tasks [34], [35], [36], [37], [38]. This dataset consists of multi-modal MRI scans, including four modalities: T1, FLAIR, T1ce, and T2. Each modality offers complementary insights regarding the tumor and adjacent brain tissue, facilitating precise segmentation. The dataset

includes voxel-level ground truth masks that categorize each voxel into four classifications: normal tissue, edema (ED), tumor core (NCR/NET), and enhancing tumor (ET) [34], [39], [40]. These annotations were meticulously verified by medical specialists, ensuring exceptional correctness and dependability. The dataset includes a variety of tumor shapes, sizes, and locations, rendering it an optimal selection for assessing the generalizability of segmentation methods [35], [37], [40].

We divided the dataset into training, validation, and testing subsets to enhance model development and assessment. The training dataset comprises labeled MRI scans, whilst the validation dataset is utilized for hyperparameter optimization and model selection. The testing set comprises unlabeled data for autonomous performance evaluation. Each MRI volume undergoes preprocessing to maintain uniformity, which includes scaling to a standardized shape, normalization of pixel intensities, and one-hot encoding of segmentation masks. This configuration guarantees that the dataset tests the model's capacity to generalize across various tumor topologies while simultaneously offering a standardized benchmark for evaluating segmentation methods.

3.2. Data Preprocessing

Data preprocessing is an essential phase to guarantee that the MRI volumes are adequately prepared for input into the deep learning model. In this work, all MRI volumes were standardized to a consistent dimension of $128 \times 128 \times 128$ to provide consistency across the dataset. The scaling was executed to address computational constraints while maintaining adequate spatial resolution for efficient segmentation. The pixel intensity values of the MRI scans were normalized to a range of $[0,1]$ by scaling each volume's intensity values, so assuring consistency in data representation and enhancing model convergence during training.

To prepare the segmentation masks for model training, the ground truth annotations were encoded using one-hot encoding. This procedure entails transforming the multi-class voxel labels (normal tissue, edema, tumor core, and enhancing tumor) into binary format over four distinct channels. Furthermore, data consistency was kept by ensuring the accurate alignment of all modalities inside an MRI volume. Any absent slices in the volumes were supplemented to preserve the uniform shape. The preprocessing processes were essential for preparing the dataset for efficient training, minimizing noise, and ensuring compatibility with the architecture of the proposed deep learning model.

3.3. Data Augmentation

Data augmentation is essential for improving the diversity and robustness of training data,

particularly in medical imaging datasets that frequently contain restricted samples [41], [42], [43], [44], [45]. This work utilized various 3D augmentation approaches to replicate real-world variability in MRI images and enhance the model's generalizability. The approaches encompassed elastic deformations, random rotations, flipping, noise additions, and intensity scaling. Random rotations between -20 degrees and 20 degrees were implemented to provide changes in imaging orientations, while horizontal and vertical flipping replicated mirror-like transformations of brain anatomy, assuring invariance to these variations [46], [47], [48], [49].

Elastic deformation was notably effective in producing non-linear changes by the application of Gaussian-filtered random displacements to MRI volumes. This method emulates real anatomical distortions and drives the model to acquire more resilient features. Gaussian noise was used to replicate imaging artifacts and acquisition noise typically observed in actual clinical environments. Furthermore, intensity scaling was executed by multiplying voxel intensities by random scaling factors, enabling the model to adjust to fluctuations in brightness and contrast resulting from disparities in MRI scanners or acquisition techniques [45], [48], [49]. [Table 1](#) delineates the data augmentation techniques utilized in this investigation.

Table 1. Summarizing the data augmentation techniques employed in this study.

Technique	Description	Purpose	Parameters
Elastic Deformation	Applies Gaussian-filtered random displacements to MRI volumes.	Mimics non-linear anatomical distortions for better generalization.	Gaussian filter: $\alpha=34 \setminus \alpha=34$ $\sigma=4 \setminus \sigma=4$
Rotation	Randomly rotates MRI volumes within a specified angle range.	Simulates different imaging orientations to enhance rotational invariance.	Angle range: -20 to 20
Intensity Scaling	Multiplies voxel intensities by random scaling factors.	Simulates mirror-like transformations for spatial invariance.	Scaling factor: 0.4 to 1.2
Flipping	Flips MRI volumes along vertical and horizontal axes.		Axes: horizontal, vertical
Noise Addition	Adds Gaussian noise to the voxel intensities of MRI volumes.	Simulates imaging artifacts and acquisition noise.	Standard Deviation: 0.05 , Mean: 0

The augmentation strategies markedly enhanced the diversity of the training data, diminishing the probability of overfitting and facilitating the model's ability to generalize effectively to novel instances [47], [48]. By dynamically producing augmented data during training, the model

experienced distinct variances in each epoch, hence improving its ability to learn invariant properties across various imaging settings. All augmentations were meticulously crafted to maintain the anatomical and spatial integrity of the data, so assuring that the enhanced samples remained clinically valid and pertinent for the brain tumor segmentation work. This extensive enhancement method improved the resilience and dependability of the proposed deep learning model [47], [48], [50]. Figure 1 illustrates the outcomes of implementing the augmentation techniques on an image.

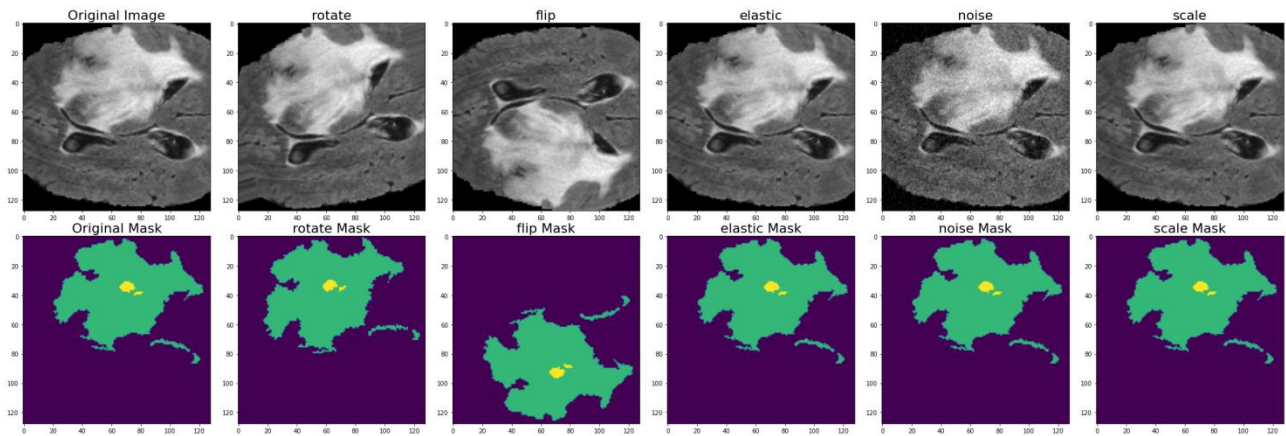


Figure 1. The results of applying the augmentation techniques to an image.

3.4. Proposed Model Architecture

The U-Net backbone functions as the core architecture for the proposed brain tumor segmentation model, utilizing an encoder-decoder framework with skip connections to capture both global context and intricate details [7], [16], [51]. The encoder path derives hierarchical features from the input MRI volumes via successive convolutional and pooling layers, systematically diminishing spatial dimensions while enhancing feature richness. The decoder path reconstructs the segmentation mask by progressively upsampling the encoded features to recover the original spatial dimensions. Skip connections link the encoder and decoder pathways, conveying high-resolution characteristics from prior encoder layers to their corresponding decoder layers. This technique maintains spatial details and guarantees accurate localization, which is essential for medical image segmentation tasks [52], [53], [54].

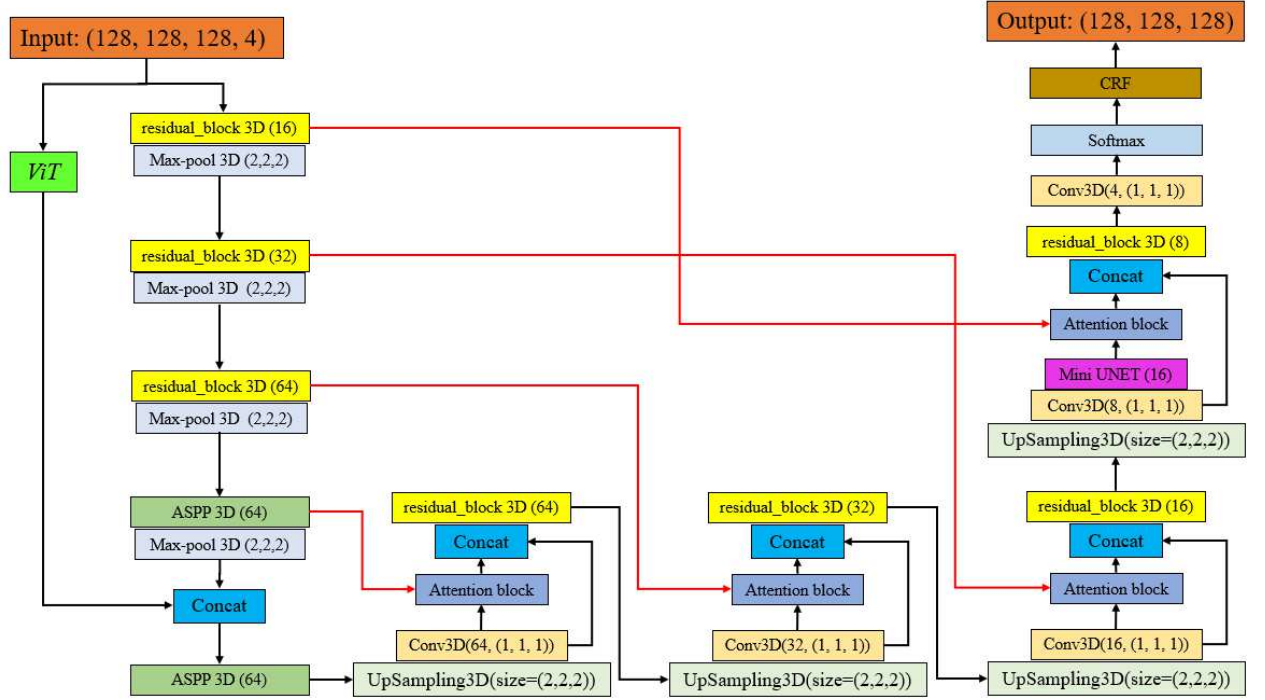


Figure 2. Graphical representation of our model.

Residual blocks are integrated at every stage of the encoder and decoder to augment the feature extraction capability of the U-Net backbone. A residual block comprises convolutional layers interleaved with batch normalization and activation functions, accompanied by shortcut connections that bypass these layers. This architecture alleviates the vanishing gradient issue, enhances gradient flow during backpropagation, and facilitates the acquisition of deeper and more resilient feature representations. The residual blocks enhance the model's capacity to accurately identify complicated and changeable tumor areas by efficiently capturing both low- and high-level information [55], [56], [57], [58].

The employed residual block, depicted in Figure 3, is intended to improve feature extraction and gradient flow inside the U-Net architecture. It has two 3D convolutional layers: the initial layer has a kernel size of $3 \times 3 \times 3$, while the subsequent layer has a kernel size of $1 \times 1 \times 1$. Both layers employ the ReLU6 activation function, which successfully mitigates gradient saturation problems and enhances computational efficiency. Every convolutional operation is succeeded by batch normalization, which stabilizes the learning process by normalizing the feature maps. In contrast to conventional residual blocks, this block does not directly transmit the input to the output. The input is subjected to an independent $3 \times 3 \times 3$ convolutional transformation prior to being combined

with the result of the two-layer transformation. This method guarantees that the input's dimensionality aligns with the output, allowing the block to successfully include both the converted features and the enhanced residual information. The architecture is especially beneficial for detecting intricate patterns in medical images, as it harmonizes feature enhancement with computing efficiency.

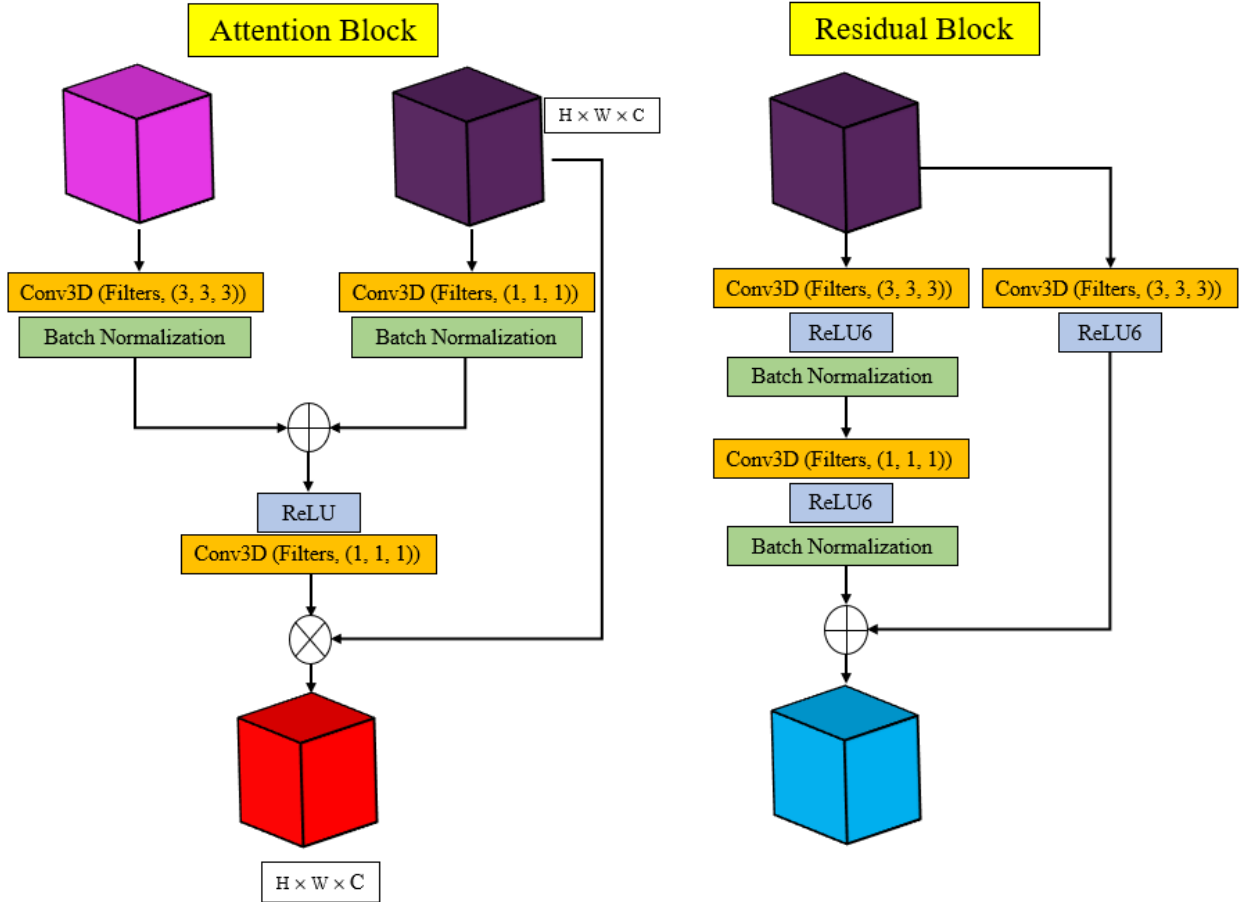


Figure 3. Architecture of the Attention and Residual Blocks.

The Vision Transformer (ViT) improves the proposed model by incorporating global feature representation through self-attention mechanisms [59], [60], [61], [62]. In contrast to traditional CNNs, which are fundamentally localized because of their kernel-based architecture, ViT effectively capture long-range connections and relationships within the input data. This capability is especially significant for medical image segmentation, as tumors may display intricate spatial interactions that extend across several parts of the brain. By incorporating ViT into the architecture, the model utilizes global contextual information in conjunction with localized features, facilitating more accurate segmentation of brain tumors [63], [64], [65].

The suggested method employs a customized 3D Vision Transformer to analyze volumetric MRI data. The input MRI volume is segmented into non-overlapping 3D patches of $16 \times 16 \times 16$, with each patch representing a small brain region. The patches are flattened and mapped into an embedding space via a dense layer, thereby encoding each patch into a feature vector. Positional embeddings are incorporated into the patch embeddings to preserve spatial information, enabling the transformer to differentiate the relative positions of patches within the 3D volume. A trainable class token is added to the beginning of the sequence of patch embeddings, aggregating global information from all patches and functioning as a summary representation for downstream tasks.

The transformer consists of several blocks of multi-head self-attention and feed-forward neural networks, each block being preceded by layer normalization. The multi-head self-attention method allows the model to assess the significance of one patch in relation to others, thereby capturing significant interactions across isolated areas in the MRI volume. Each transformer block iteratively changes the patch embeddings, enhancing their representations by considering both local and global contexts. This iterative refinement enables the ViT to simulate intricate spatial interactions throughout the brain, rendering it especially useful for segmenting tumor regions that may exhibit irregular forms or overlap with adjacent structures.

Upon processing the input through the transformer blocks, the resultant embeddings are reconfigured into a 3D feature map. The feature map, enhanced with global contextual information, is included into the U-Net backbone to aid in the segmentation process. By integrating the ViT's capacity to capture global dependencies with the U-Net's proficiency in maintaining local spatial details, the model attains a robust equilibrium between global and local feature representation. This integration markedly enhances the precision of brain tumor segmentation, allowing the model to identify tumor boundaries more efficiently in intricate MRI data.

Attention mechanisms are essential to the proposed model, especially for enhancing the skip connections in the U-Net design. In traditional U-Net models, skip connections convey high-resolution spatial data from the encoder to the decoder, thereby preserving fine details throughout the reconstruction of segmentation masks. Nonetheless, not all attributes in these linkages possess equal significance, particularly for tasks such as brain tumor segmentation, where the emphasis should be on tumor regions. To resolve this, attention blocks are implemented to selectively highlight the most pertinent elements within the skip connections, enhancing the model's capacity to focus on tumor-specific areas while diminishing extraneous background information [2], [19],

[66], [67], [68], [69].

The attention block functions by dynamically assigning weights to the features in the skip connection according to their significance for the current decoder phase. This is accomplished by calculating an attention map that aligns the features from the skip connection with those from the decoder. The block initiates by applying a $3 \times 3 \times 3$ convolution to the input decoder features, succeeded by batch normalization to produce a feature map. The skip connection features are concurrently analyzed using a $1 \times 1 \times 1$ convolution and batch normalization. The processed features are subsequently merged through element-wise addition and activated using a ReLU function. A concluding $1 \times 1 \times 1$ convolution with a sigmoid activation produces the attention map, emphasizing the significance of each feature in the skip connection.

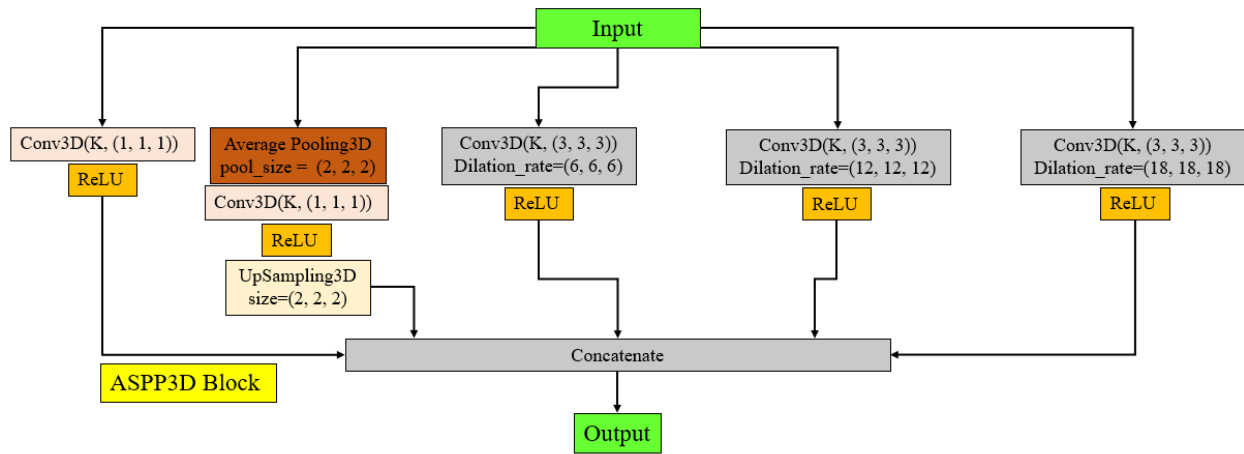


Figure 4. Architecture of the 3D Atrous Spatial Pyramid Pooling (ASPP) block.

The produced attention map is utilized on the skip connection via element-wise multiplication, enabling the network to concentrate solely on the most pertinent aspects for the current decoding phase. This selective refinement improves the transmission of significant spatial and contextual information, ensuring that the decoder concentrates on tumor locations with more precision. The attention mechanism allows the network to dynamically adjust its focus, hence improving its efficacy in segmenting complicated and irregular tumor formations.

The incorporation of attention blocks into the skip connections enhances the model's precision and robustness in brain tumor segmentation. These blocks not only augment the model's capacity to identify tumor locations but also mitigate the influence of extraneous characteristics and background noise. This targeted refinement is essential for guaranteeing that the model produces

high-quality segmentation results, especially in complex scenarios with varied tumor shapes and overlapping regions.

The Atrous Spatial Pyramid Pooling (ASPP) module is a crucial element of the proposed architecture, intended to capture multi-scale contextual information vital for segmenting brain tumors with intricate shapes and varying sizes [68], [70], [71], [72]. The ASPP module utilizes atrous convolutions with varying dilation rates, allowing the model to extract features from diverse receptive field widths while maintaining the spatial precision of the feature maps. This multi-scale feature extraction is essential for comprehending both intricate features, such as tumor boundaries, and overarching structures, such as the tumor's interaction with adjacent tissue [68], [71], [73], [74]. The architecture of the ASPP block is depicted in [Figure 4](#).

The ASPP module consists of multiple parallel layers that manipulate the input feature map through various actions. The architecture comprises a $1 \times 1 \times 1$ convolutional layer for the extraction of intricate local features, alongside three $3 \times 3 \times 3$ atrous convolutions with dilation rates of 6, 12, and 18. The diverse dilation rates enable the model to identify characteristics at several scales, with smaller dilation rates emphasizing local patterns and bigger dilation rates including wider spatial contexts. Furthermore, a global average pooling operation consolidates characteristics from the complete input volume, guaranteeing that the module encodes comprehensive contextual information. The aggregated features are further reshaped, processed via a $1 \times 1 \times 1$ convolutional layer, and upsampled to align with the spatial dimensions of the other outputs.

The outputs from these parallel layers are concatenated along the channel dimension to merge multi-scale characteristics into a singular enriched representation. This integrated feature map amalgamates data from several receptive field dimensions, enhancing the model's comprehension of the spatial complexity of brain tumors. The ASPP module improves the model's capacity to collect both local and global spatial features by utilizing multi-scale atrous convolutions and global pooling. This is especially advantageous for brain tumor segmentation, as it guarantees accurate delineation of tumor boundaries while considering the wider anatomical context. The integration of varied dilation rates and global context aggregation enables the model to effectively manage the heterogeneous and irregular characteristics of brain tumors in medical imaging tasks.

Conditional Random Fields (CRFs) are integrated into the proposed model as a post-processing phase to enhance segmentation boundaries and augment the precision of the projected masks [75], [76], [77]. The U-Net-based architecture, although effective in generating detailed

segmentation predictions, may yield excessively smooth or coarse boundaries, especially in instances when tumor regions are intricate or intersect with adjacent tissues. CRFs mitigate this disadvantage by functioning as a probabilistic graphical model that ensures spatial consistency and aligns segmentation predictions with the actual anatomical structures in the MRI volumes [78], [79], [80], [81].

The CRF layer functions by enhancing the pixel-level predictions generated by the network. It specifically examines the associations between adjacent pixels or voxels, employing spatial and intensity similarities to ascertain the probability of a voxel being classified into a particular category [76], [78], [81]. The CRF incorporates spatial information into the segmentation map, ensuring that the projected boundaries are more accurate and closely coincide with the real tumor edges. This technique is crucial for differentiating areas of tumor from adjacent tissues, where intensity variations may be minimal [76], [77], [80].

In the suggested design, the CRF is included as a customized layer that processes the output of the final segmentation layer. The existing implementation functions as a placeholder, but the CRF logic can be augmented to incorporate widely utilized techniques, such as fully linked CRFs, which utilize pairwise potentials to enhance segmentation accuracy. These potentials encompass both the unary components (network predictions) and the pairwise components (spatial consistency). Utilizing this information, the CRF rectifies segmentation inaccuracies resulting from network misclassifications and yields a more anatomically coherent outcome.

The incorporation of a CRF layer improves the model's overall performance, especially in difficult scenarios where precise border delineation is essential. The proposed model integrates the advantages of deep learning for feature extraction with CRF for boundary refinement, resulting in enhanced segmentation accuracy and reliability, hence rendering it suitable for clinical applications in brain tumor analysis.

3.5. Explainability via Grad-CAM

Explainability is a crucial component of deep learning models, especially in medical applications, where comprehending the reasoning behind predictions is vital for clinical credibility and acceptance [5], [9], [82], [83]. To improve the interpretability of the proposed brain tumor segmentation model, Gradient-weighted Class Activation Mapping (Grad-CAM) was utilized to produce heatmaps that illustrate the areas of the input MRI volumes that significantly impacted the model's predictions. Grad-CAM emphasizes significant spatial features by calculating the

gradients of the target class (e.g., tumor region) relative to the activations of a chosen convolutional layer, thereby indicating which areas of the MRI most influenced the model's judgments [9], [52], [84].

The final convolutional layer of the model was chosen for the creation of Grad-CAM heatmaps. This layer was selected due to its ability to preserve spatial information while abstracting essential high-level elements for prediction. The gradients of the target class concerning the output of this layer were backpropagated to ascertain the significance of each feature map. The significance values were subsequently weighted and aggregated over the layer's channels, resulting in a spatial activation map. The resultant heatmap was superimposed onto the original MRI slice, emphasizing the areas of emphasis for the model during segmentation.

Incorporating Grad-CAM into the workflow yielded significant insights into the model's decision-making process [5], [84], [85]. The heatmaps illustrated the model's ability to distinguish between tumor locations and healthy tissues, providing a visual explanation of its predictions. This capability is especially advantageous in detecting potential failure instances, such as when the model inaccurately forecasts tumor margins or overlooks subtle tumor areas. Through the examination of heatmaps, doctors can ascertain if the model's predictions correspond with their anticipations and comprehend the rationale behind misclassifications, so enhancing the transparency and reliability of the segmentation process [5], [82], [85].

Grad-CAM functioned as an effective instrument for comprehending and corroborating the predictions of the presented model. Providing voxel-level explainability not only boosted clinical interpretability but also facilitated debugging and improvement of the model's performance. This integration ensures that the model's predictions are reliable and based on pertinent features, thereby bridging the gap between the complexity of deep learning models and the necessity for actionable insights in medical imaging.

3.6. Evaluation metrics

Evaluation metrics are essential for analyzing the efficacy of a segmentation model, especially in intricate tasks such as brain tumor segmentation. This study included multiple metrics to assess the model's proficiency in accurately delineating and classifying tumor areas. Accuracy served as a comprehensive indicator to assess the ratio of accurately categorized pixels throughout the full segmentation map. Although accuracy offers a general assessment of the model's efficacy, it may inadequately reflect the variations within specific tumor patches, particularly in datasets

characterized by uneven class distributions, where small tumor areas may be shadowed by extensive background regions [68], [86], [87].

In addition to accuracy, precision and recall were employed to evaluate class-specific performance. Precision emphasizes the reduction of false positives by measuring the ratio of accurately identified tumor pixels to the total pixels indicated as tumor class. This parameter is essential for preventing the erroneous classification of healthy tissue as tumor, which is vital in therapeutic applications. Recall, conversely, assesses the model's capacity to detect all pertinent tumor pixels, hence reducing false negatives. Elevated recall guarantees that no segment of the tumor is overlooked, which is crucial for treatment strategizing and surveillance [6], [7], [24], [88].

The Dice coefficient, a widely used statistic in medical image segmentation, was utilized to assess the overlap between the predicted segmentation and the ground truth. This score spans from 0 (no overlap) to 1 (complete overlap) and offers a balanced assessment of precision and recall. This is especially beneficial for assessing the model's efficacy in delineating tumor boundaries, rendering it optimal for medical applications where accuracy in boundary detection is essential. Furthermore, ROC-AUC was employed to evaluate the model's capacity to differentiate between classes based on predicted probabilities, providing insights into its performance across diverse confidence levels [3], [24], [25], [89].

3.7. Model Training

The model was developed in Python and implemented utilizing tools like TensorFlow and Keras for architecture, training, and assessment. OpenCV was employed for data preprocessing activities, encompassing the scaling of MRI volumes and the normalization of intensity values to provide consistency throughout the dataset. The Vision Transformer (ViT) encoder was incorporated utilizing the timm library, notably the 'vit_base_patch16_224' pretrained model, which utilized knowledge from extensive datasets. The studies were performed on a high-performance machine featuring an NVIDIA GeForce RTX 3080 GPU, facilitating effective parallel computation, complemented by a 12th Gen Intel(R) Core(TM) i7-12700H processor and 32 GB of RAM for uninterrupted data processing and model training.

The proposed model comprises about 7 million parameters, including approximately 3.78 million trainable parameters and 3.27 million non-trainable parameters. The non-trainable parameters were obtained from the pretrained Vision Transformer, initialized with weights from ImageNet training. By freezing these weights during training, the overall feature extraction

capabilities of the transformer were preserved, hence providing strong global feature representations. This method significantly decreased computational expenses and alleviated the risk of overfitting, particularly considering the relatively small size of the BraTS 2020 dataset. Trainable parameters were predominantly concentrated on acquiring task-specific characteristics for delineating brain tumor areas.

The training procedure was set for a maximum of 50 epochs with a batch size of 1, employing a learning rate scheduler to dynamically modify the learning rate over time, hence facilitating stable convergence. Early stopping was implemented to track validation loss and terminate training if no enhancement was detected for three successive epochs, hence ensuring optimal performance without overfitting. The categorical cross-entropy loss function directed the model's predictions in this multi-class segmentation challenge, facilitating precise separation between tumor kinds and normal tissue.

In order to optimize computational efficiency, a custom time-tracking callback was implemented to monitor the time taken per epoch, while model checkpointing was implemented to protect the model weights with the greatest validation performance during the training process. The combination of sophisticated methodologies, pretrained transformer integration, and robust hardware setup facilitated the model's attainment of elevated segmentation accuracy while preserving computing efficiency.

4. Results and Discussion

The suggested model was assessed using the BraTS 2020 dataset in conjunction with various cutting-edge models, including nnU-Net [90], SwinBTS [18], Edge U-Net [53], and 3DUV-NetR+ [91]. All models were trained and evaluated on identical dataset samples to guarantee a fair comparison. The evaluation measures were accuracy, precision, recall, and the Dice coefficient for three principal tumor regions: Tumor Core, Edema, and Enhancing Tumor. The findings indicate that the suggested model surpassed alternative methods across multiple important criteria, illustrating its efficacy in brain tumor segmentation. The results of applying five models to the dataset are presented in [Tables 2 and 3](#) and [Figure 5](#).

The suggested model for the Tumor Core attained the highest precision (0.55) and Dice coefficient (0.39) compared to all models, while maintaining an accuracy of 0.99, comparable to nnU-Net. The results demonstrate that the proposed approach accurately identifies tumor areas while reducing false positives, an essential consideration for clinical applications. Significantly,

none of the other models attained a greater recall for Tumor Core than the suggested method, hence emphasizing its enhanced segmentation proficiency.

In the Edema segmentation challenge, the suggested model exhibited remarkable superiority, attaining the greatest precision (0.70) and Dice coefficient (0.62). The results demonstrate the model's proficiency in accurately segmenting Edema areas, even when the structure is uneven and diffuse. In contrast to nnU-Net, which attained a Dice coefficient of 0.58, and 3DUV-NetR+, which achieved 0.56, the proposed model's attention mechanisms and ASPP module markedly improved its ability to capture multi-scale features and efficiently concentrate on the Edema regions.

The suggested model attained a Dice coefficient of 0.54 for the Enhancing Tumor region, above the performance of nnU-Net (0.51), SwinBTS (0.50), and 3DUV-NetR+ (0.50). The accuracy for Enhancing Tumor segmentation consistently stayed at 0.99 across the models; however, the proposed model's superior Dice score underscores its enhanced equilibrium between precision and recall. The Enhancing Tumor areas present significant challenges due to the small size and irregular morphology, and the suggested model's superior performance compared to other techniques in this domain highlights its robustness.

The higher efficacy of the proposed model is due to its attention-enhanced skip connections, which optimize the transfer of spatial features between the encoder and decoder. The attention methods enabled the model to dynamically concentrate on pertinent features in the input data while attenuating background noise, resulting in enhanced precision and superior segmentation outcomes for all tumor regions.

The ASPP module enhanced the model's capabilities by extracting multi-scale contextual information through atrous convolutions with diverse dilation rates. This function was especially advantageous for delineating Edema and Enhancing Tumor areas, where diversity in size and spatial distribution frequently presents considerable challenges. The multi-scale feature extraction allowed the model to identify both intricate details and overarching architecture, facilitating effective segmentation across various tumor types.

All models evaluated in this study were trained and assessed on identical dataset samples, guaranteeing a uniform and impartial comparison. Although several models, such as nnU-Net and SwinBTS, demonstrated competitive performance in specific measures, the suggested model consistently achieved higher results across all tumor regions. For instance, although nnU-Net

attained excellent accuracy, its precision and Dice scores were inferior to those of the suggested model for Edema and Enhancing Tumor regions, underscoring the latter's proficiency in properly balancing several facets of segmentation performance.

Despite all models attaining comparable high accuracy (0.98 or 0.99), accuracy alone is an inadequate metric for assessing segmentation quality in this context, owing to the dataset's imbalanced nature, characterized by a predominance of background pixels. Metrics like the Dice coefficient and precision, which concentrate on tumor-specific areas, offer a more refined assessment and were essential in differentiating the performance of the proposed model from its rivals.

Table 2. Quantitative results of implementing five models on the BraTS 2020 dataset.

<i>Models tested on BraTS 2020 dataset</i>	<i>Class</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>Dice</i>
nnU-Net [90]	Tumor Core	0.99	0.52	0.40	0.35
	Edema	0.97	0.66	0.60	0.58
	Enhancing Tumor	0.98	0.63	0.51	0.51
SwinBTS [18]	Tumor Core	0.98	0.54	0.39	0.37
	Edema	0.98	0.68	0.61	0.59
	Enhancing Tumor	0.99	0.64	0.53	0.50
Edge U-Net [53]	Tumor Core	0.98	0.52	0.37	0.36
	Edema	0.98	0.65	0.58	0.57
	Enhancing Tumor	0.98	0.64	0.51	0.52
3DUV-NetR+ [91]	Tumor Core	0.98	0.53	0.37	0.35
	Edema	0.98	0.67	0.60	0.56
	Enhancing Tumor	0.99	0.62	0.53	0.50
Ours	Tumor Core	0.99	0.55	0.41	0.39
	Edema	0.98	0.70	0.63	0.62
	Enhancing Tumor	0.99	0.66	0.55	0.54

The incorporation of the Vision Transformer (ViT) and the CRF post-processing layer markedly improved the performance of the proposed model. The ViT identified long-range dependencies in the MRI volumes, enhancing the representation of global properties.

Simultaneously, the CRF layer enhanced the segmentation boundaries, ensuring that the predicted masks accurately corresponded to the actual tumor shapes. These enhancements enabled the model to surpass current methodologies and attain an improved equilibrium between sensitivity and specificity.

The suggested model attained state-of-the-art performance on the BraTS 2020 dataset, showcasing its robustness and efficacy in brain tumor segmentation. This study establishes a new benchmark for brain tumor segmentation by utilizing advanced architectural components, including attention mechanisms, ASPP, and CRF, while assuring a fair comparison through the training and testing of all models on identical dataset samples.

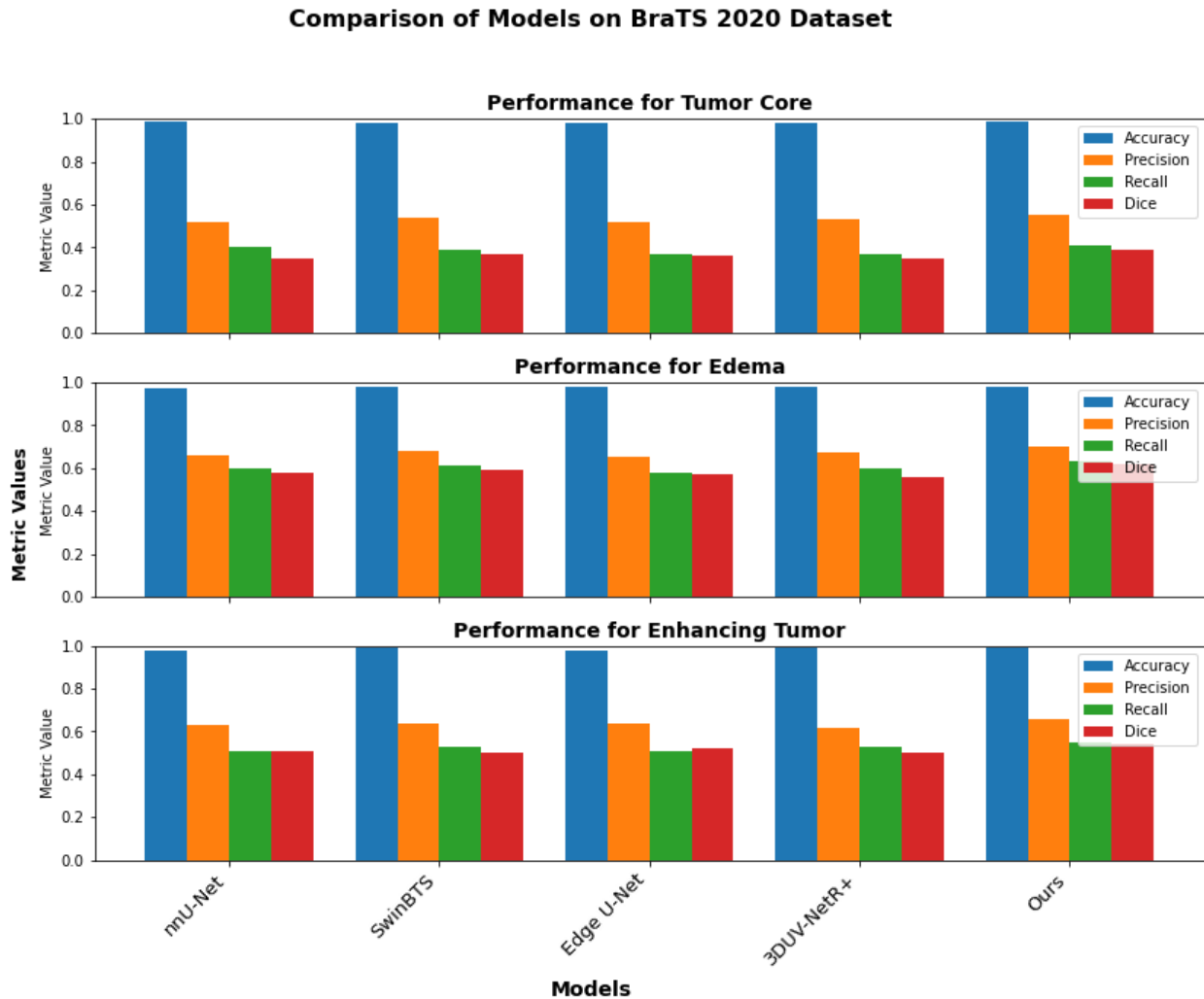
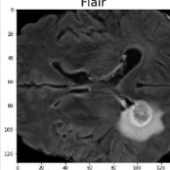
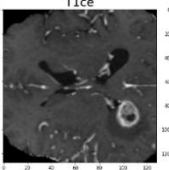
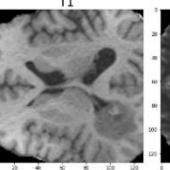
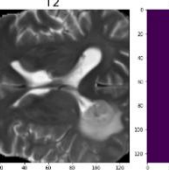
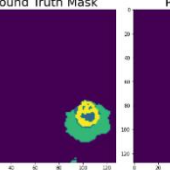
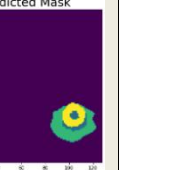
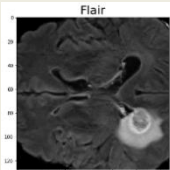
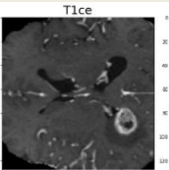
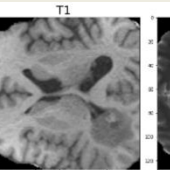
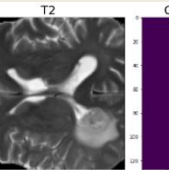
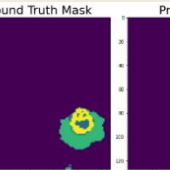
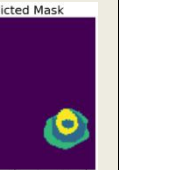
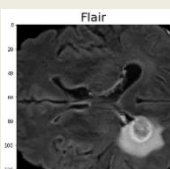
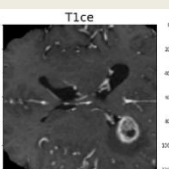
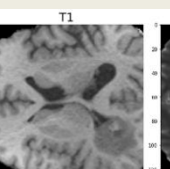
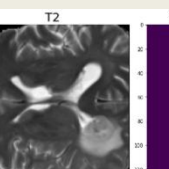
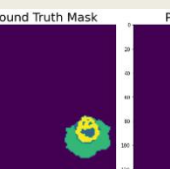
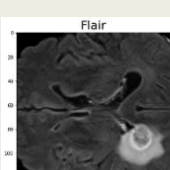
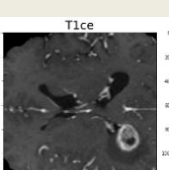
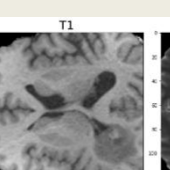
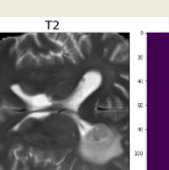
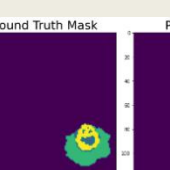
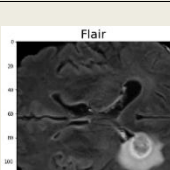
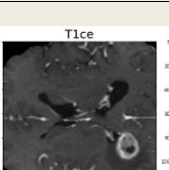
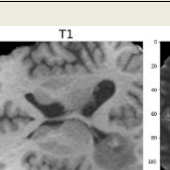
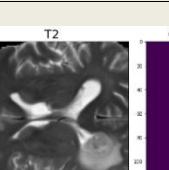
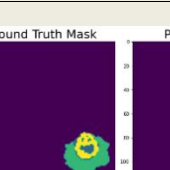


Figure 5. Evaluation of models using the BraTS 2020 dataset across three tumor regions: Tumor Core, Edema, and Enhancing Tumor. The metrics encompass Accuracy, Recall, Precision, and Dice Coefficient for nnU-Net, Edge U-Net, SwinBTS, 3DUNetR+, and the suggested model (Ours).

Table 3. Visual comparison of segmentation outcomes for nnU-Net [90], SwinBTS [18], Edge U-Net [53], 3DUV-NetR+ [91], and the suggested model. The outcomes demonstrate the ability of each model to delineate tumor areas, including Edema, Tumor Core, and Enhancing Tumor.

<i>Models</i>	<i>Results</i>					
nnU-Net [90]						
SwinBTS [18]						
Edge U-Net [53]						
3DUV-NetR+ [91]						
Ours						

5. Ablation study

The ablation study assesses alternative model configurations to determine the contributions of different components in improving segmentation performance on the BraTS 2020 dataset. Six models featuring diverse configurations of ReLU, ReLU6, ASPP (in both encoder and merging phases), variations in kernel size, and the inclusion of Pretrained ViT were evaluated. This analysis

compares the performance of these setups based on critical assessment measures like as accuracy, precision, recall, and the Dice coefficient.

Model 1 (ReLU + ASPP after merging + ASPP in encoder + pretrained ViT), as outlined in Table 4, functioned as a benchmark for enhanced setups. Although it demonstrated commendable overall accuracy and precision, its recall and Dice coefficients were marginally inferior for intricate tumor locations such as Edema and Enhancing Tumor, indicating potential for enhancement in the equilibrium of false negatives and positives. The use of ASPP at various stages and a pretrained ViT facilitated the acquisition of multi-scale features and global dependencies; nevertheless, the dependence on ReLU constrained gradient flow efficiency in deeper layers.

Model 2 (ReLU6 + ASPP in encoder + pretrained ViT) employs ReLU6 instead of ReLU, hence augmenting gradient flow and enhancing recall and Dice coefficients for Edema segmentation. Nevertheless, in the absence of ASPP post-merger, the model encountered difficulties in capturing intricate details, especially inside Enhancing Tumor areas. The findings suggest that ASPP in the encoder alone is inadequate for efficient multi-scale feature extraction.

Model 3 (ReLU + ASPP in the encoder + pretrained ViT) exhibited inferior overall performance relative to Models 1 and 2. The lack of ASPP following the merger resulted in diminished capacity to grasp global context, as indicated by decreased Dice scores in tumor regions. This approach emphasized the significance of including ASPP post-merging stages to enhance segmentation boundaries efficiently.

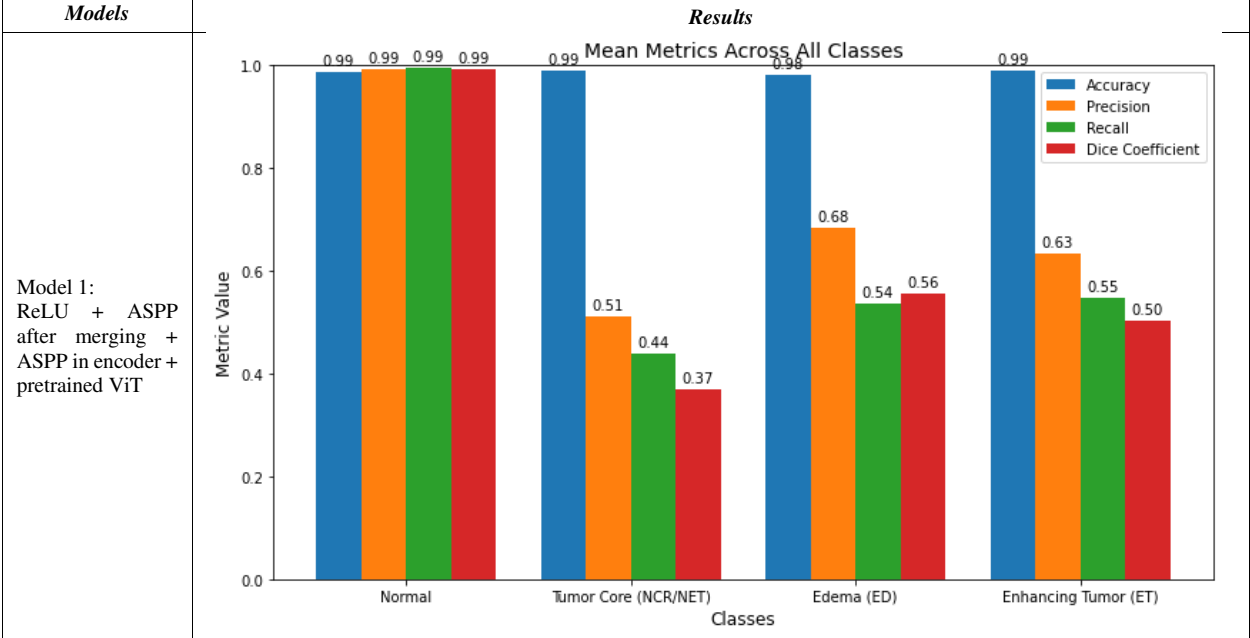
Model 4 (ReLU + ASPP post-merger + ASPP in encoder + kernel_size=5 in ASPP + pretrained ViT) incorporated an enlarged kernel size in ASPP to capture more extensive spatial features. This arrangement enhanced precision and Dice scores for bigger tumor areas such as Edema, but exhibited somewhat diminished performance for tiny regions like Enhancing Tumor. This indicates that augmenting the kernel size in ASPP can enhance performance for particular tumor types, although may not be applicable to other instances.

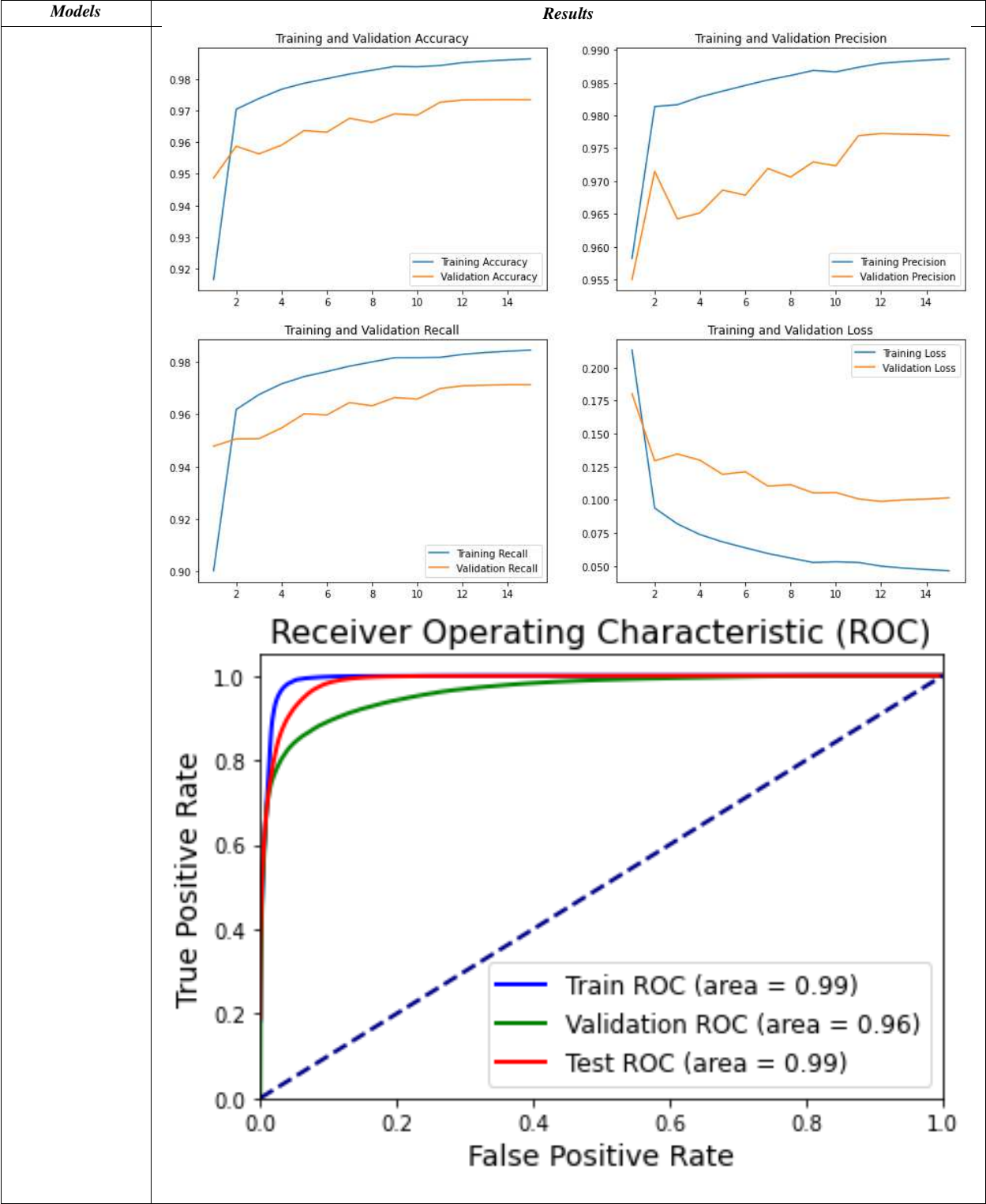
Model 5 (ReLU6 + ASPP post-merger + ASPP in encoder) substituted the pretrained ViT with an exclusively convolutional backbone. Although it exhibited satisfactory accuracy and recall, the absence of comprehensive feature representation resulted in marginally reduced precision and Dice coefficients, especially for Edema and Enhancing Tumor areas. This underscores the essential function of Pretrained ViT in capturing long-range dependencies and enhancing overall segmentation quality.

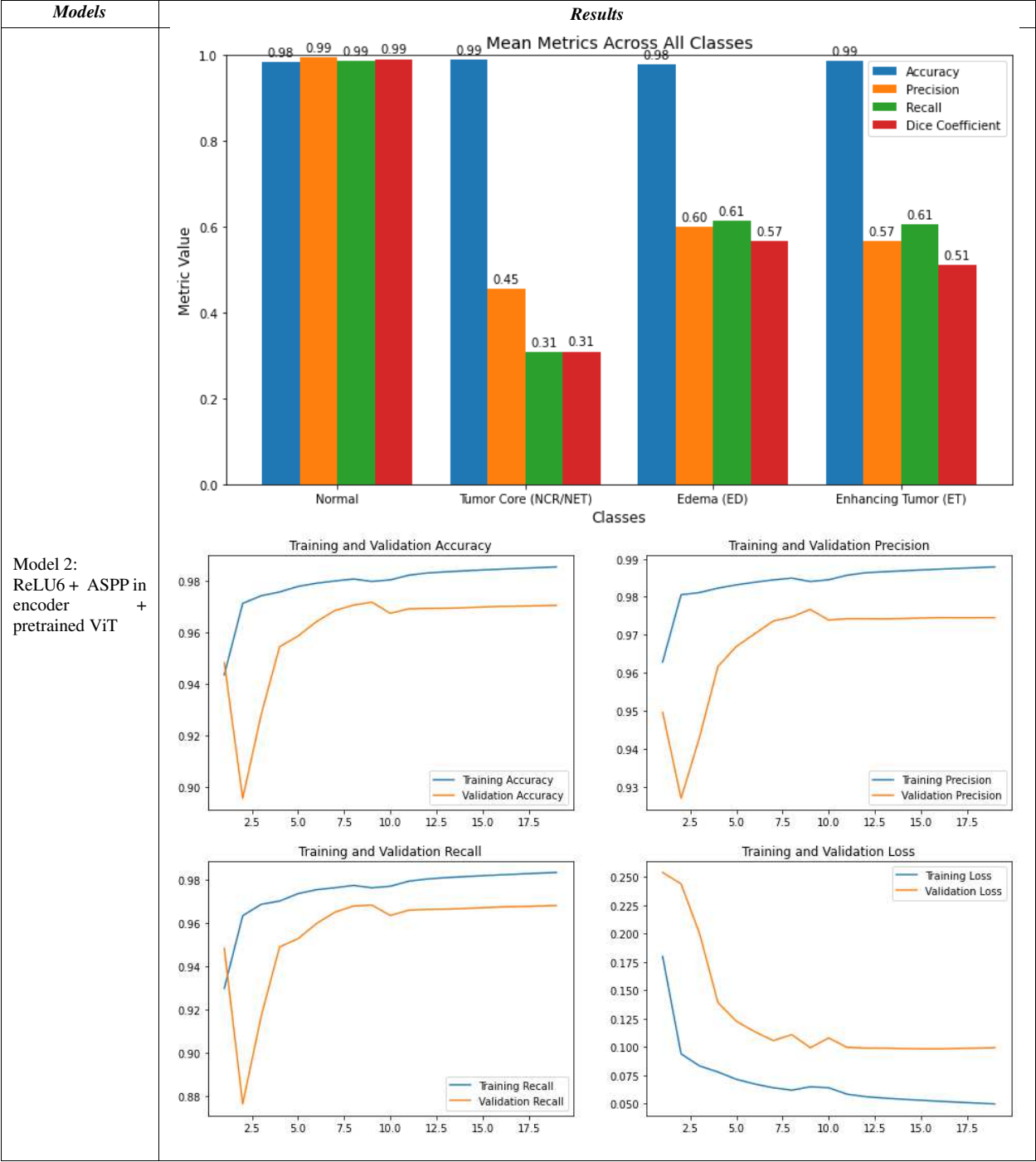
Model 6 (Ours: ReLU6 + ASPP post-merging + ASPP in encoder + pretrained ViT) demonstrated superior performance, attaining the highest precision and Dice coefficients across all tumor locations. The integration of ReLU6 for gradient optimization, ASPP at both the encoder and merging phases for multi-scale feature extraction, and a pre-trained ViT for global context representation established a resilient architecture for brain tumor segmentation. This arrangement effectively balanced precision and recall, especially in tough areas such as Enhancing Tumor, where the Dice coefficient was markedly superior to those of previous models.

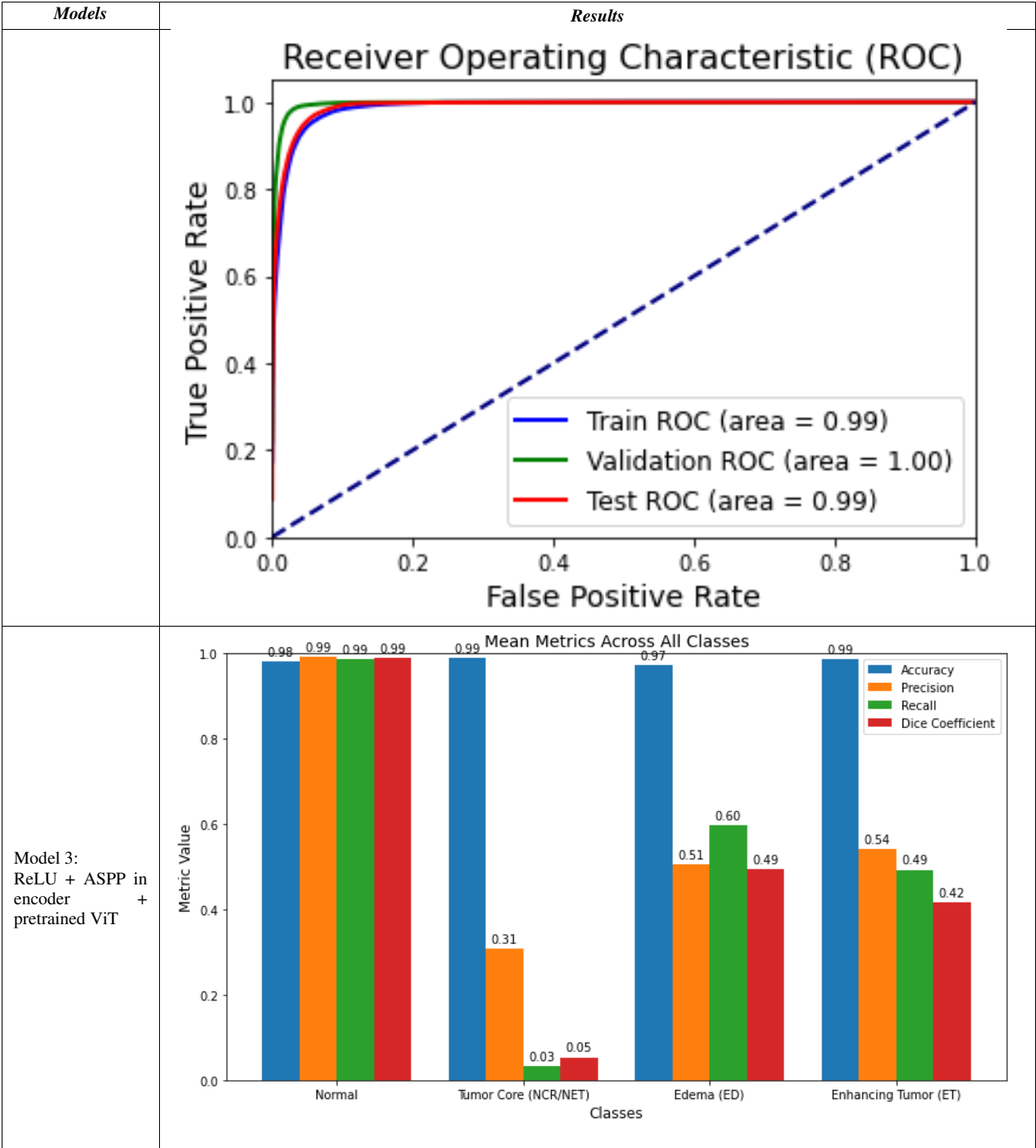
The ablation investigation underscores the significant roles of ReLU6, ASPP location, and the pretrained ViT in enhancing segmentation performance. The separate components, such as ASPP and pretrained ViT, function individually; however, their integration in the proposed model (Model 6) yields a synergistic enhancement, establishing a new standard for brain tumor segmentation on the BraTS 2020 dataset.

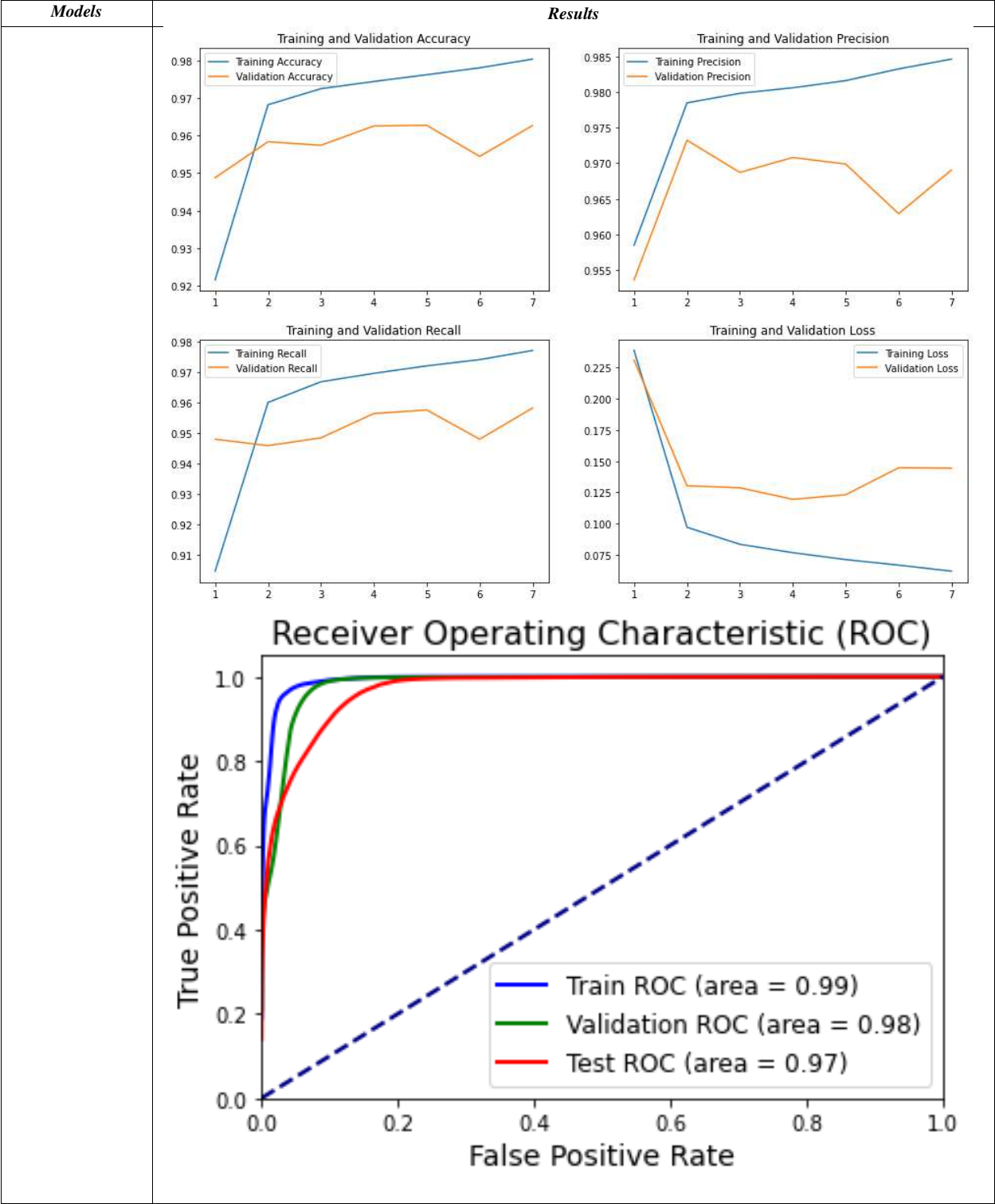
Table 4. Evaluation metrics, ROC curves, and training-validation performance curves from the ablation study. The bar charts highlight the accuracy, precision, recall, and Dice coefficients for each tumor class, while the training curves show the progression of precision, accuracy, recall, and loss, providing insights into the convergence and generalization of each model structure. The ROC curves illustrate the model's ability to balance sensitivity and specificity across training, validation, and test sets.

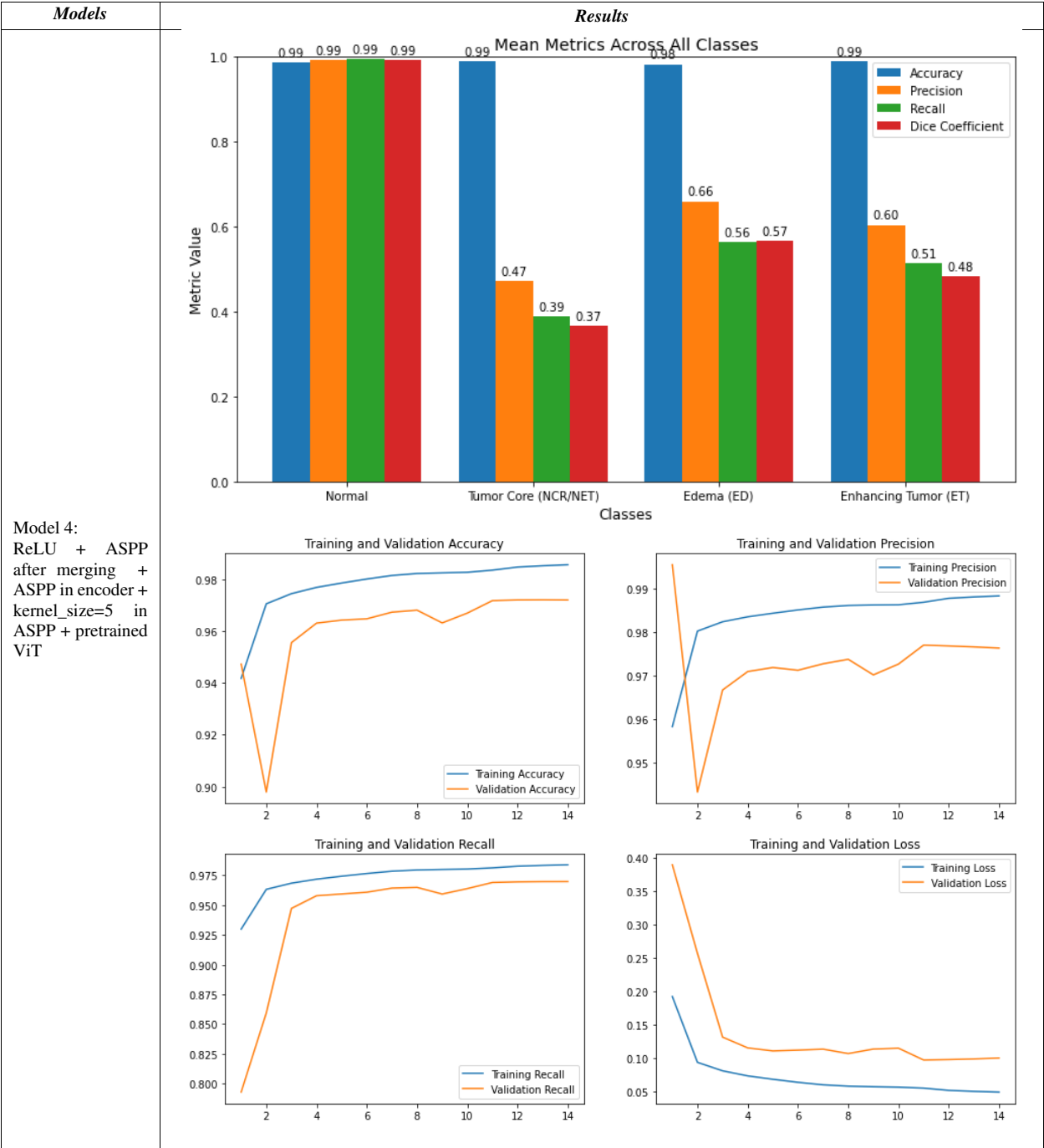




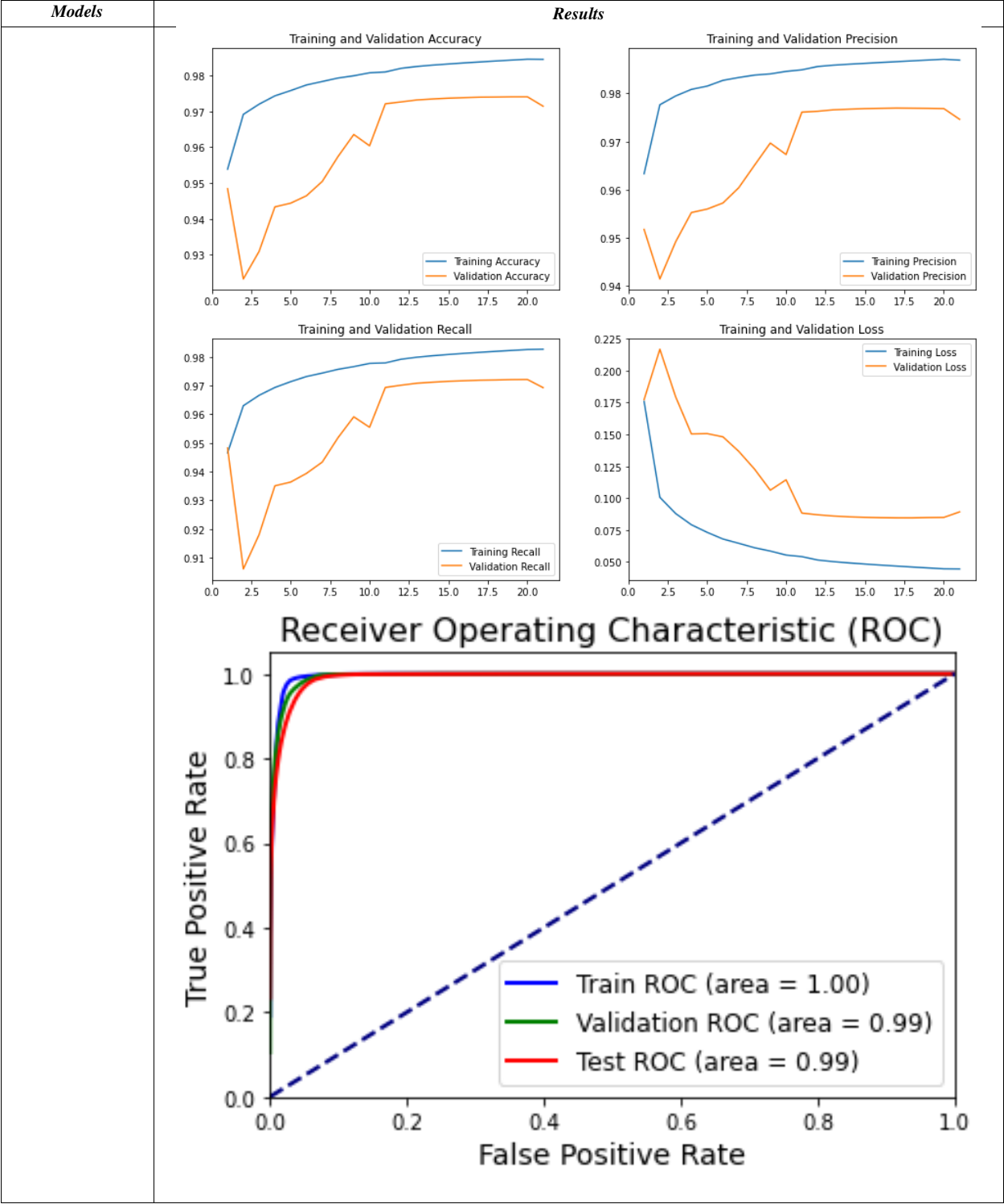


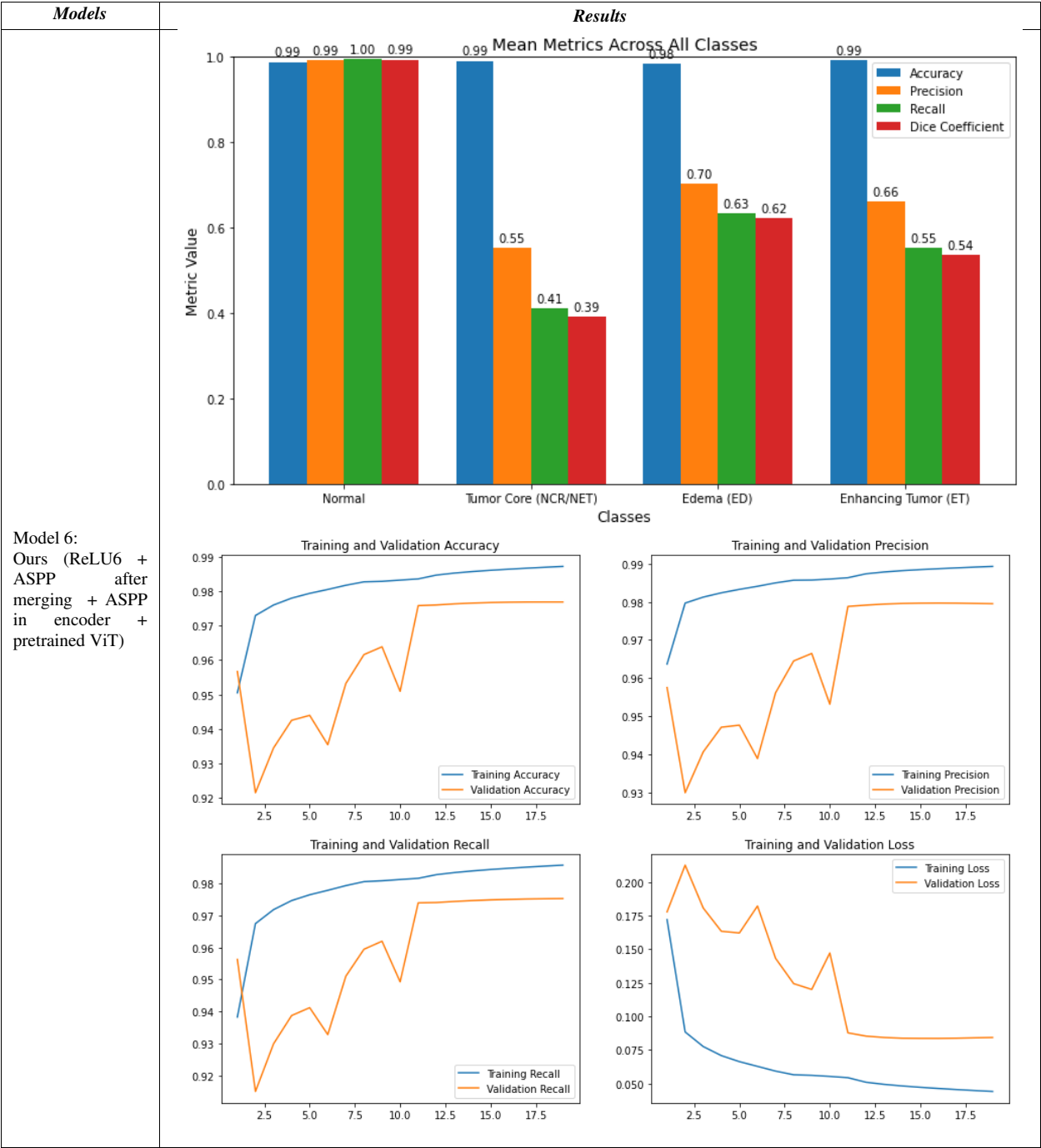


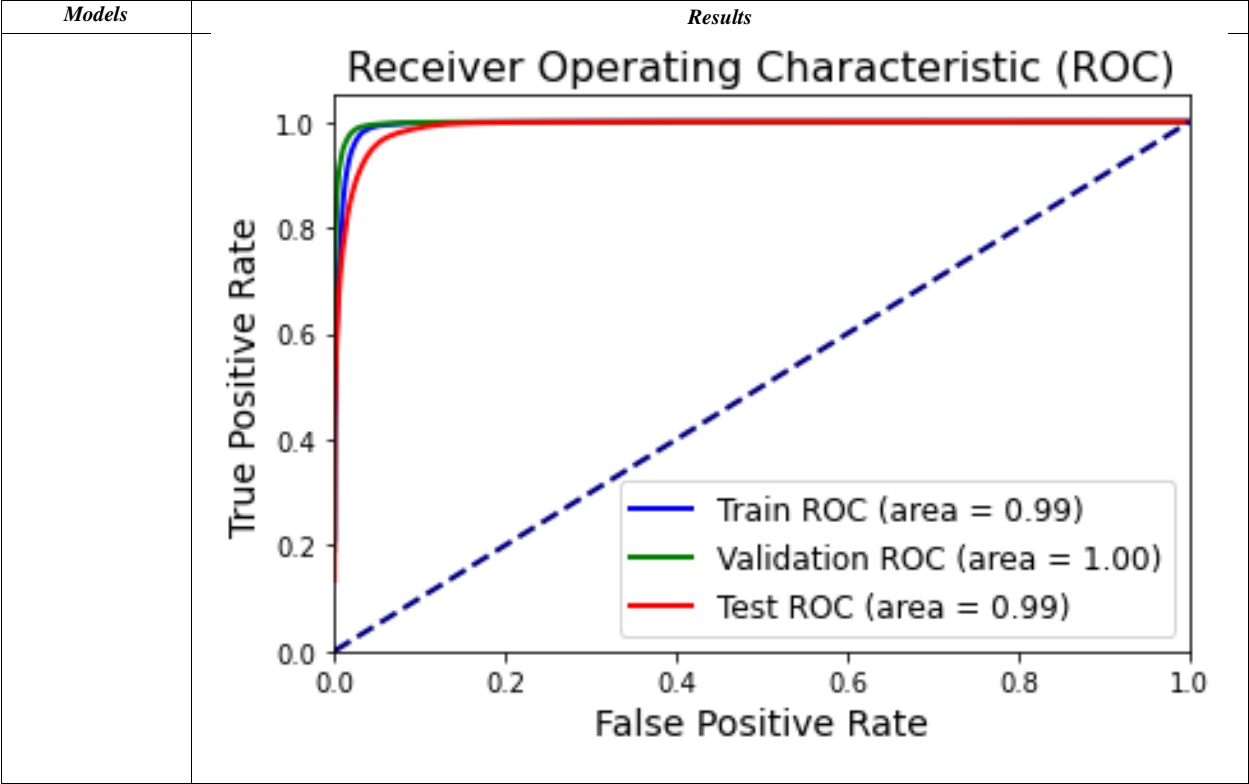




Models	Results																									
	Receiver Operating Characteristic (ROC) True Positive Rate False Positive Rate Train ROC (area = 0.99) Validation ROC (area = 0.99) Test ROC (area = 0.99)																									
Model 5: ReLU6 + ASPP after merging + ASPP in encoder	Mean Metrics Across All Classes Metric Value Classes Accuracy Precision Recall Dice Coefficient <table><tr><th>Class</th><th>Accuracy</th><th>Precision</th><th>Recall</th><th>Dice Coefficient</th></tr><tr><td>Normal</td><td>0.99</td><td>0.99</td><td>1.00</td><td>0.99</td></tr><tr><td>Tumor Core (NCR/NET)</td><td>0.99</td><td>0.50</td><td>0.34</td><td>0.30</td></tr><tr><td>Edema (ED)</td><td>0.98</td><td>0.68</td><td>0.64</td><td>0.62</td></tr><tr><td>Enhancing Tumor (ET)</td><td>0.99</td><td>0.67</td><td>0.52</td><td>0.51</td></tr></table>	Class	Accuracy	Precision	Recall	Dice Coefficient	Normal	0.99	0.99	1.00	0.99	Tumor Core (NCR/NET)	0.99	0.50	0.34	0.30	Edema (ED)	0.98	0.68	0.64	0.62	Enhancing Tumor (ET)	0.99	0.67	0.52	0.51
Class	Accuracy	Precision	Recall	Dice Coefficient																						
Normal	0.99	0.99	1.00	0.99																						
Tumor Core (NCR/NET)	0.99	0.50	0.34	0.30																						
Edema (ED)	0.98	0.68	0.64	0.62																						
Enhancing Tumor (ET)	0.99	0.67	0.52	0.51																						





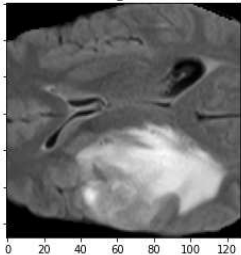
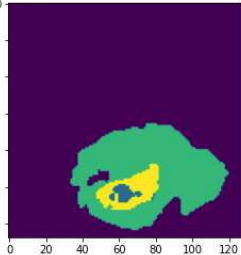
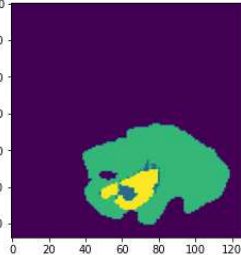
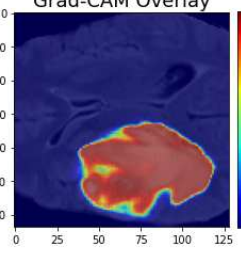
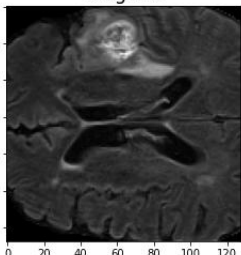
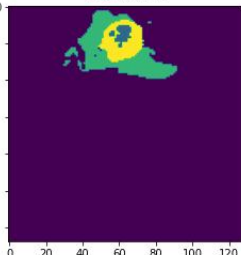
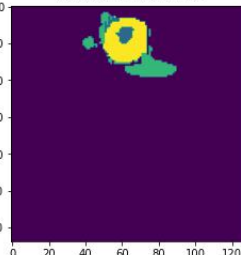
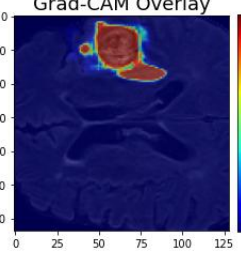
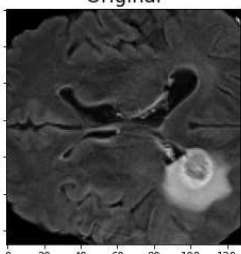
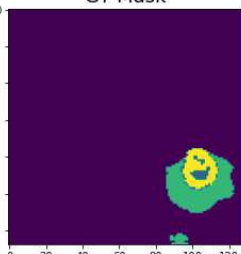
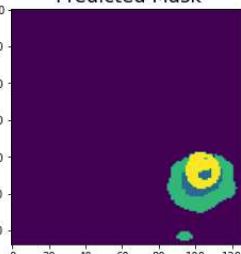
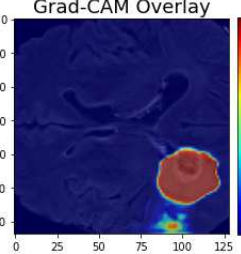
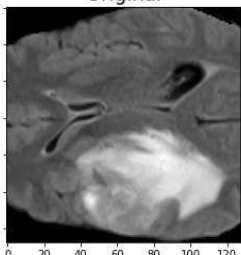
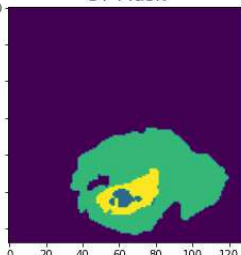
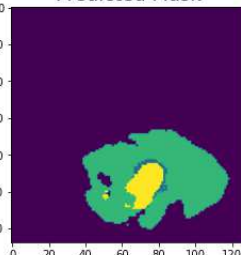
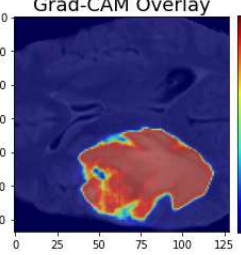
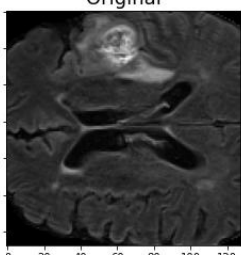
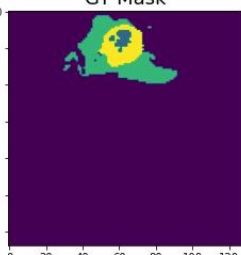
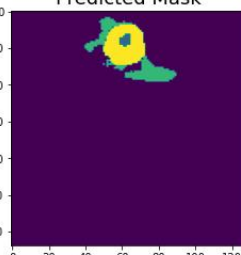
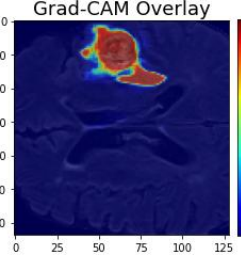


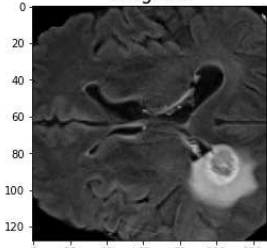
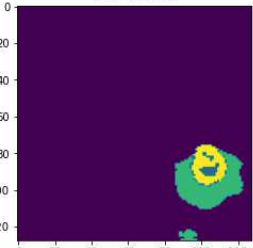
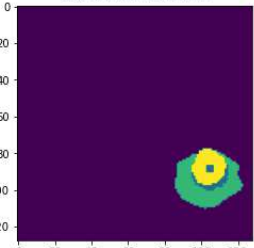
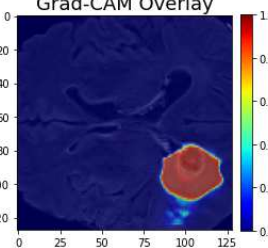
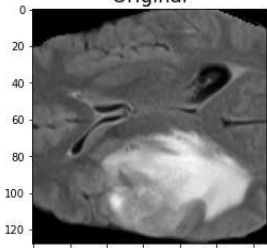
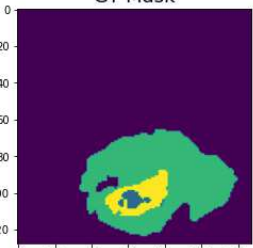
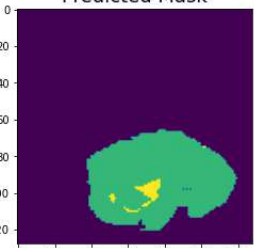
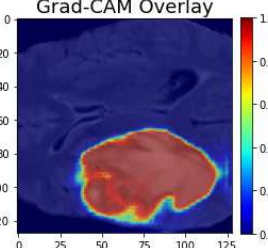
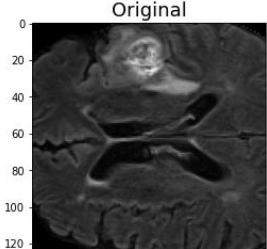
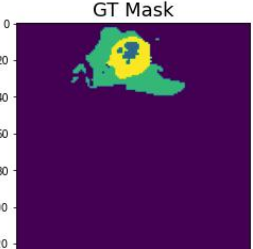
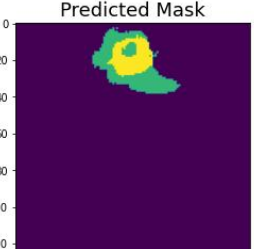
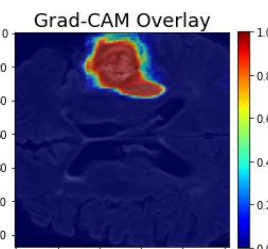
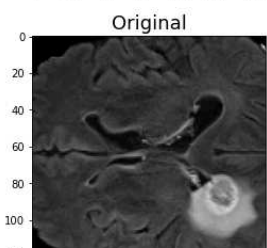
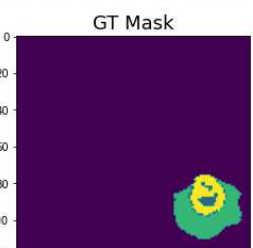
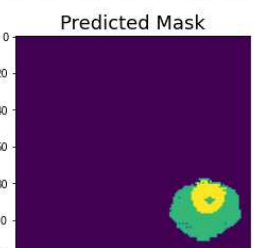
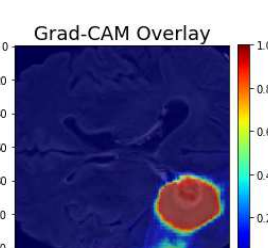
To assess the interpretability and decision-making processes of the models in the ablation study, Grad-CAM heatmaps were produced for each structure to illustrate the regions influencing the predictions (refer to [Table 5](#)). The findings indicated that the proposed model (Model 6) consistently concentrated on pertinent tumor regions with accurate attention to borders, in contrast to other models. Structures lacking ASPP post-merger (e.g., Model 3) exhibited diffuse and less distinct activations, signifying diminished localization efficacy. Likewise, models devoid of pretrained ViT (e.g., Model 5) demonstrated less global context, leading to overlooked or irrelevant activations for extensive tumor areas such as Edema. These discoveries underscore the synergistic impact of using ASPP at both the encoder and merging phases, ReLU6 for effective feature extraction, and pretrained ViT for comprehensive feature representation in enhancing segmentation accuracy and interpretability.

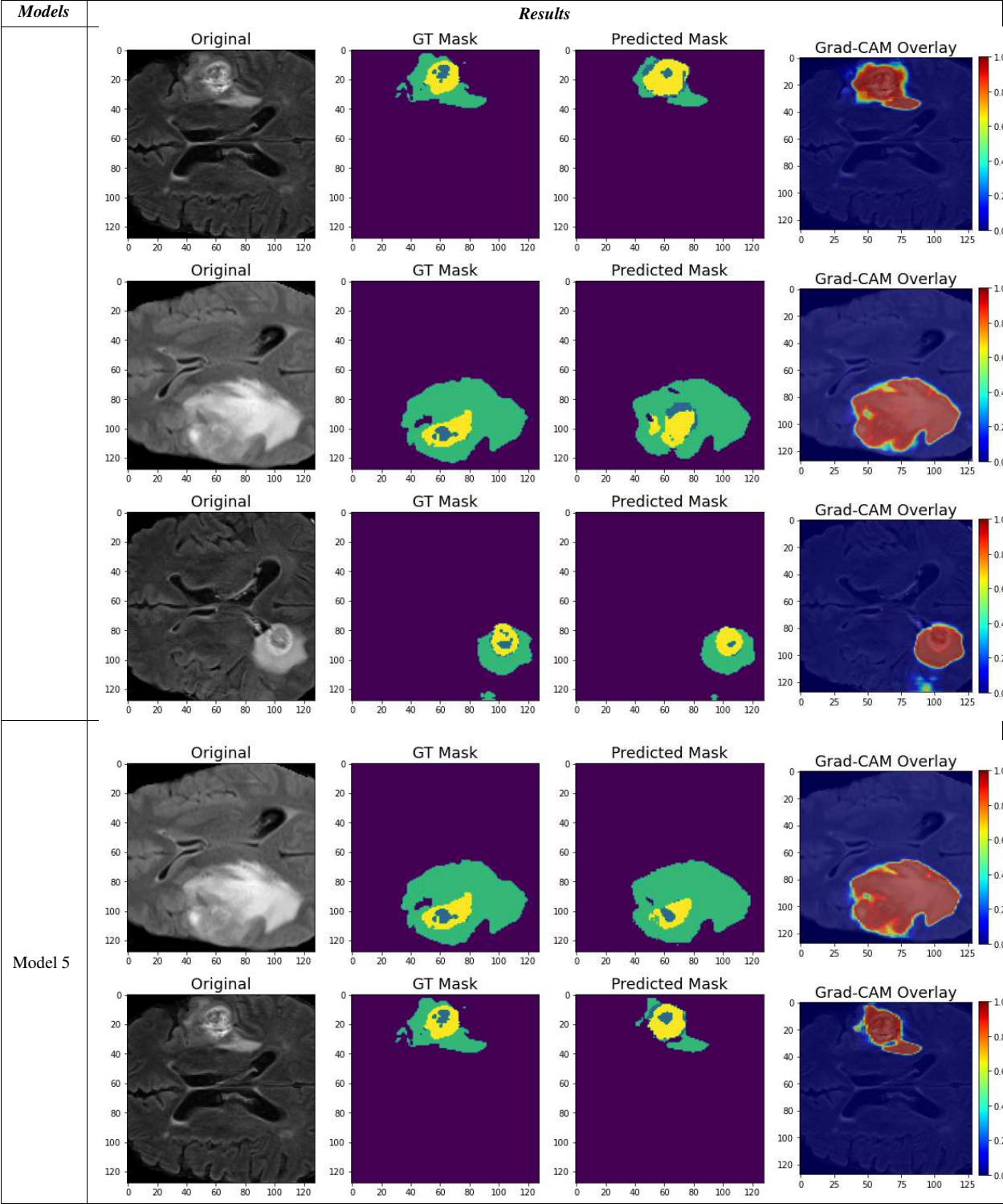
637

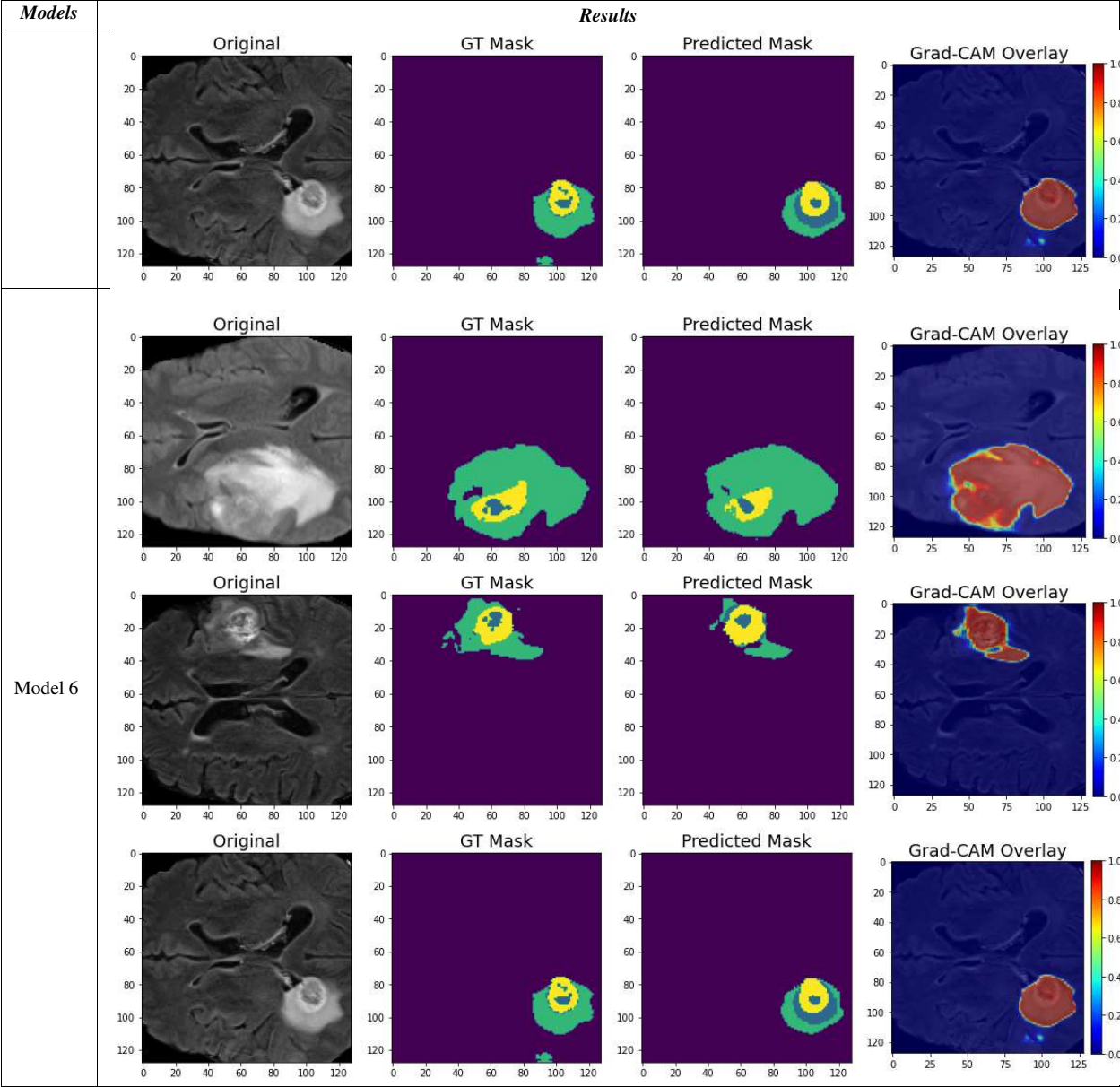
638

Table 5. Comparison of Grad-CAM heatmaps across models from the ablation study, revealing the ability of diverse architectures to focus on tumor areas and refine segmentation predictions.

Models	Results			
Model 1	Original	GT Mask	Predicted Mask	Grad-CAM Overlay
				
	Original	GT Mask	Predicted Mask	Grad-CAM Overlay
				
	Original	GT Mask	Predicted Mask	Grad-CAM Overlay
				
Model 2	Original	GT Mask	Predicted Mask	Grad-CAM Overlay
				
	Original	GT Mask	Predicted Mask	Grad-CAM Overlay
				

Models	Results			
	Original 	GT Mask 	Predicted Mask 	Grad-CAM Overlay 
Model 3	Original 	GT Mask 	Predicted Mask 	Grad-CAM Overlay 
	Original 	GT Mask 	Predicted Mask 	Grad-CAM Overlay 
	Original 	GT Mask 	Predicted Mask 	Grad-CAM Overlay 
Model 4				





6. Conclusion and Outlook

This paper presents an advanced deep learning model for brain tumor segmentation utilizing the BraTS 2020 dataset. Our model incorporates ReLU6 for effective gradient propagation, an Atrous Spatial Pyramid Pooling (ASPP) module for multi-scale feature extraction during both the encoder and merging phases, and a pretrained ViT for the acquisition of global context. The incorporation of these elements resulted in superior performance, especially in intricate tumor areas such as Edema and Enhancing Tumor, as seen by elevated Dice coefficients and balanced

precision-recall measures relative to other models. We conducted a comprehensive ablation investigation that emphasized the essential contributions of each component, demonstrating their individual and collective advantages in enhancing segmentation accuracy and robustness.

The findings underscore the significance of multi-scale feature representation and global dependency modeling in medical image segmentation. The attention mechanisms and post-processing approaches, including the CRF layer, enhanced segmentation boundaries, ensuring precise localization of tumor locations. The model's performance across critical measures highlights its potential as a dependable instrument for aiding doctors in brain tumor diagnosis and treatment planning. Furthermore, the application of Grad-CAM for interpretability revealed the model's capacity to deliver comprehensible predictions, which is crucial for clinical implementation.

Several interesting approaches exist to improve the presented methodology. Integrating supplementary datasets featuring varied patient demographics and imaging techniques may enhance the model's generalizability and resilience. Investigating lightweight and efficient model versions for deployment in real-time clinical environments, including edge devices, represents a further direction for future research. Moreover, incorporating domain adaptation strategies could further improve the model's efficacy on unfamiliar datasets, mitigating possible issues related to domain shift in medical imaging.

The use of multimodal data, including radiology reports and genetic information, in conjunction with MRI images, may reveal new opportunities for tumor characterization and classification. Subsequent research may investigate self-supervised or semi-supervised learning methodologies to utilize unlabeled data, hence diminishing reliance on annotated datasets. The gains in this study establish a basis for creating more precise, interpretable, and scalable AI-driven solutions for brain tumor segmentation and other applications.

Declaration of competing interests

The authors declare that there are no conflicting or financial interests.

Declarations Ethics Approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Availability of data and materials

Data available on request due to privacy/ethical restrictions.

Author Contributions:

Homayoun Safarpour: Software, Conceptualization, Methodology, Investigation;

Soroush Sadeghi: Software, Conceptualization, Methodology, Investigation;

Payam Zarbakhsh: Methodology, Software, Investigation;

Mohammadreza Kamsari: Validation, Methodology, Formal analysis;

Marjan Kia: Writing- Reviewing and Editing, Formal analysis;

Ramin Ranjbarzadeh: Supervisor, Writing- Reviewing and Editing, Validation;

References:

M. S. Ullah, M. A. Khan, N. A. Almujaally, M. Alhaisoni, T. Akram, and M. Shabaz, "BrainNet: a fusion assisted novel optimal framework of residual blocks and stacked autoencoders for multimodal brain tumor classification," *Scientific Reports* 2024 14:1, vol. 14, no. 1, pp. 1–20, Mar. 2024, doi: 10.1038/s41598-024-56657-3.

R. Ranjbarzadeh, A. Bagherian Kasgari, S. Jafarzadeh Ghouschi, S. Anari, M. Naseri, and M. Bendeche, "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," *Scientific Reports* 2021 11:1, vol. 11, no. 1, pp. 1–17, May 2021, doi: 10.1038/s41598-021-90428-8.

R. Ranjbarzadeh, A. Caputo, E. B. Tirkolaee, S. Jafarzadeh Ghouschi, and M. Bendeche, "Brain tumor segmentation of MRI images: A comprehensive review on the application of artificial intelligence tools," *Comput Biol Med*, vol. 152, p. 106405, Jan. 2023, doi: 10.1016/J.COMPBIOMED.2022.106405.

A. Bataineh, J. Transfer-Learning, A. Amarnath, A. Al Bataineh, and J. A. Hansen, "Transfer-Learning Approach for Enhanced Brain Tumor Classification in MRI Imaging," *BioMedInformatics* 2024, Vol. 4, Pages 1745-1756, vol. 4, no. 3, pp. 1745–1756, Jul. 2024, doi: 10.3390/BIOMEDINFORMATICS4030095.

R. A. Zeineldin *et al.*, "Explainable hybrid vision transformers and convolutional network for multimodal glioma segmentation in brain MRI," *Scientific Reports* 2024 14:1, vol. 14, no. 1, pp. 1–14, Feb. 2024, doi: 10.1038/s41598-024-54186-7.

W. Li, W. Huang, and Y. Zheng, "CorrDiff: Corrective Diffusion Model for Accurate MRI Brain Tumor Segmentation," *IEEE J Biomed Health Inform*, vol. 28, no. 3, pp. 1587–1598, Mar. 2024, doi: 10.1109/JBHI.2024.3353272.

W. Zhang, S. Chen, Y. Ma, Y. Liu, and X. Cao, "ETUNet: Exploring efficient transformer enhanced UNet for 3D brain tumor segmentation," *Comput Biol Med*, vol. 171, p. 108005, Mar. 2024, doi: 10.1016/J.COMPBIOMED.2024.108005.

M. Usman Akbar, M. Larsson, I. Blystad, and A. Eklund, "Brain tumor segmentation using synthetic MR images - A comparison of GANs and diffusion models," *Scientific Data* 2024 11:1, vol. 11, no. 1, pp. 1–17, Feb. 2024, doi: 10.1038/s41597-024-03073-x.

R. Ranjbarzadeh *et al.*, "Ebttag: Explainable Brain Tumor Segmentation with Efficientnetv2, Attention Mechanism, and Grad-Cam in Clinical and Public Datasets," 2024, doi: 10.2139/SSRN.5023648.

- 719] J. Chaki and M. Wozniak, "Brain Tumor Categorization and Retrieval Using Deep Brain Incep Res
720 Architecture Based Reinforcement Learning Network," *IEEE Access*, vol. 11, pp. 130584–130600, 2023,
721 doi: 10.1109/ACCESS.2023.3334434.
- 722] F. Ullah *et al.*, "Brain Tumor Segmentation from MRI Images Using Handcrafted Convolutional Neural
723 Network," *Diagnostics 2023, Vol. 13, Page 2650*, vol. 13, no. 16, p. 2650, Aug. 2023, doi:
724 10.3390/DIAGNOSTICS13162650.
- 725] A. Brunetti *et al.*, "AHANet: Adaptive Hybrid Attention Network for Alzheimer's Disease Classification
726 Using Brain Magnetic Resonance Imaging," *Bioengineering 2023, Vol. 10, Page 714*, vol. 10, no. 6, p. 714,
727 Jun. 2023, doi: 10.3390/BIOENGINEERING10060714.
- 728] R. Yousef, S. Khan, G. Gupta, B. M. Albahlal, S. A. Alajlan, and A. Ali, "Bridged-U-Net-ASPP-EVO and
729 Deep Learning Optimization for Brain Tumor Segmentation," *Diagnostics 2023, Vol. 13, Page 2633*, vol.
730 13, no. 16, p. 2633, Aug. 2023, doi: 10.3390/DIAGNOSTICS13162633.
- 731] F. Ullah *et al.*, "Evolutionary Model for Brain Cancer-Grading and Classification," *IEEE Access*, vol. 11,
732 pp. 126182–126194, 2023, doi: 10.1109/ACCESS.2023.3330919.
- 733] P. Ahmad, S. Qamar, L. Shen, and A. Saeed, "Context Aware 3D UNet for Brain Tumor Segmentation,"
734 *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture*
735 *Notes in Bioinformatics)*, vol. 12658 LNCS, pp. 207–218, 2021, doi: 10.1007/978-3-030-72084-
736 1_19/COVER/.
- 737] N. Cinar, A. Ozcan, and M. Kaya, "A hybrid DenseNet121-UNet model for brain tumor segmentation from
738 MR Images," *Biomed Signal Process Control*, vol. 76, p. 103647, Jul. 2022, doi:
739 10.1016/J.BSPC.2022.103647.
- 740] Z. Huang, Y. Zhao, Y. Liu, and G. Song, "GCAUNet: A group cross-channel attention residual UNet for
741 slice based brain tumor segmentation," *Biomed Signal Process Control*, vol. 70, p. 102958, Sep. 2021, doi:
742 10.1016/J.BSPC.2021.102958.
- 743] Y. Jiang, Y. Zhang, X. Lin, J. Dong, T. Cheng, and J. Liang, "SwinBTS: A Method for 3D Multimodal
744 Brain Tumor Segmentation Using Swin Transformer," *Brain Sciences 2022, Vol. 12, Page 797*, vol. 12, no.
745 6, p. 797, Jun. 2022, doi: 10.3390/BRAINS12060797.
- 746] N. S. Shaik and T. K. Cherukuri, "Multi-level attention network: application to brain tumor classification,"
747 *Signal Image Video Process*, vol. 16, no. 3, pp. 817–824, Apr. 2022, doi: 10.1007/S11760-021-02022-
748 0/TABLES/4.
- 749] B. Jena, S. Jain, G. K. Nayak, and S. Saxena, "Analysis of depth variation of U-NET architecture for brain
750 tumor segmentation," *Multimed Tools Appl*, vol. 82, no. 7, pp. 10723–10743, Mar. 2023, doi:
751 10.1007/S11042-022-13730-1/FIGURES/13.
- 752] S. Deepak and P. M. Ameer, "Brain tumor categorization from imbalanced MRI dataset using weighted
753 loss and deep feature fusion," *Neurocomputing*, vol. 520, pp. 94–102, Feb. 2023, doi:
754 10.1016/J.NEUCOM.2022.11.039.
- 755] M. U. Rehman, J. Ryu, I. F. Nizami, and K. T. Chong, "RAAGR2-Net: A brain tumor segmentation network
756 using parallel processing of multiple spatial frames," *Comput Biol Med*, vol. 152, p. 106426, Jan. 2023, doi:
757 10.1016/J.COMPBIOMED.2022.106426.
- 758] V. Vinay Kumar and P. Grace Kanmani Prince, "Deep belief network Assisted quadratic logit boost
759 classifier for brain tumor detection using MR images," *Biomed Signal Process Control*, vol. 81, p. 104415,
760 Mar. 2023, doi: 10.1016/J.BSPC.2022.104415.

- 761] Y. Ding and H. Liu, "Barely-supervised Brain Tumor Segmentation via Employing Segment Anything
762 Model," *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2024, doi:
763 10.1109/TCSVT.2024.3384503.
- 764] C. S. D S and J. Christopher Clement, "Enhancing brain tumor segmentation in MRI images using the IC-
765 net algorithm framework," *Scientific Reports* 2024 14:1, vol. 14, no. 1, pp. 1–19, Jul. 2024, doi:
766 10.1038/s41598-024-66314-4.
- 767] M. U. Saeed *et al.*, "RMU-Net: A Novel Residual Mobile U-Net Model for Brain Tumor Segmentation
768 from MR Images," *Electronics* 2021, Vol. 10, Page 1962, vol. 10, no. 16, p. 1962, Aug. 2021, doi:
769 10.3390/ELECTRONICS10161962.
- 770] P. Sreedevi, A. Kiran, T. Santhi Sri, E. Poornima, S. Polepaka, and Y. Supriya Reddy, "Da-resbigru -brain
771 tumor classification using Dual attention residual bi directional gated recurrent unit using MRI images,"
772 *Biomed Signal Process Control*, vol. 88, p. 105596, Feb. 2024, doi: 10.1016/J.BSPC.2023.105596.
- 773] H. Mehnatkesh, S. M. J. Jalali, A. Khosravi, and S. Nahavandi, "An intelligent driven deep residual learning
774 framework for brain tumor classification using MRI images," *Expert Syst Appl*, vol. 213, p. 119087, Mar.
775 2023, doi: 10.1016/J.ESWA.2022.119087.
- 776] F. Khodadadi Shoushtari, S. Sina, and A. N. V. Dehkordi, "Automatic segmentation of glioblastoma
777 multiform brain tumor in MRI images: Using Deeplabv3+ with pre-trained Resnet18 weights," *Physica*
778 *Medica*, vol. 100, pp. 51–63, Aug. 2022, doi: 10.1016/J.EJMP.2022.06.007.
- 779] Z. Zhu, M. Sun, G. Qi, Y. Li, X. Gao, and Y. Liu, "Sparse Dynamic Volume TransUNet with multi-level
780 edge fusion for brain tumor segmentation," *Comput Biol Med*, vol. 172, p. 108284, Apr. 2024, doi:
781 10.1016/J.COMPBIOMED.2024.108284.
- 782] M. Sharif, U. Tanvir, E. U. Munir, M. A. Khan, and M. Yasmin, "Brain tumor segmentation and
783 classification by improved binomial thresholding and multi-features selection," *J Ambient Intell Humaniz*
784 *Comput*, vol. 15, no. 1, pp. 1063–1082, Jan. 2024, doi: 10.1007/S12652-018-1075-X/TABLES/15.
- 785] Y. Feng, Y. Cao, D. An, P. Liu, X. Liao, and B. Yu, "DAUnet: A U-shaped network combining deep
786 supervision and attention for brain tumor segmentation," *Knowl Based Syst*, vol. 285, p. 111348, Feb. 2024,
787 doi: 10.1016/J.KNOSYS.2023.111348.
- 788] H. Yang, T. Zhou, Y. Zhou, Y. Zhang, and H. Fu, "Flexible Fusion Network for Multi-Modal Brain Tumor
789 Segmentation," *IEEE J Biomed Health Inform*, vol. 27, no. 7, pp. 3349–3359, Jul. 2023, doi:
790 10.1109/JBHI.2023.3271808.
- 791] J. Tang, T. Li, H. Shu, and H. Zhu, "Variational-Autoencoder Regularized 3D MultiResUNet for the BraTS
792 2020 Brain Tumor Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes*
793 *in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12659 LNCS, pp. 431–440, 2021, doi:
794 10.1007/978-3-030-72087-2_38/COVER/.
- 795] RanjbarzadehRamin, KelesAyse, CraneMartin, and BendechangeMalika, "Comparative analysis of real-
796 clinical MRI and BraTS datasets for brain tumor segmentation," *IET Conference Proceedings*, 2024, doi:
797 10.1049/ICP.2024.3274.
- 798] L. Fidon, S. Ourselin, and T. Vercauteren, "Generalized Wasserstein Dice Score, Distributionally Robust
799 Deep Learning, and Ranger for Brain Tumor Segmentation: BraTS 2020 Challenge," *Lecture Notes in*
800 *Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*
801 *Bioinformatics)*, vol. 12659 LNCS, pp. 200–214, 2021, doi: 10.1007/978-3-030-72087-2_18/COVER/.
- 802] S. Bakas *et al.*, "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation,
803 Progression Assessment, and Overall Survival Prediction in the BRATS Challenge," Nov. 2018, Accessed:
804 Sep. 14, 2021. [Online]. Available: <https://arxiv.org/abs/1811.02629v3>

- 805] T. Henry *et al.*, “Brain Tumor Segmentation with Self-ensembled, Deeply-Supervised 3D U-Net Neural
806 Networks: A BraTS 2020 Challenge Solution,” *Lecture Notes in Computer Science (including subseries*
807 *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12658 LNCS, pp. 327–
808 339, 2021, doi: 10.1007/978-3-030-72084-1_30/COVER/.
- 809] Z. Jiang, C. Ding, M. Liu, and D. Tao, “Two-stage cascaded u-net: 1st place solution to brats challenge
810 2019 segmentation task,” *Lecture Notes in Computer Science (including subseries Lecture Notes in*
811 *Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11992 LNCS, pp. 231–241, 2020, doi:
812 10.1007/978-3-030-46640-4_22/COVER/.
- 813] R. Ranjbarzadeh, P. Zarbakhsh, A. Caputo, E. B. Tirkolaee, and M. Bendeche, “Brain Tumor
814 Segmentation based on an Optimized Convolutional Neural Network and an Improved Chimp Optimization
815 Algorithm,” Oct. 2022, doi: 10.21203/RS.3.RS-2203596/V1.
- 816] T. M. Vo, T. T. Vo, T. T. Phan, H. T. Nguyen, and D. T. Tran, “Data Augmentation Techniques Evaluation
817 on Ultrasound Images for Breast Tumor Segmentation Tasks,” *Studies in Computational Intelligence*, vol.
818 1097, pp. 153–164, 2023, doi: 10.1007/978-3-031-29447-1_14.
- 819] X. Yang, E. del Rey Castillo, Y. Zou, and L. Wotherspoon, “Semantic segmentation of bridge point clouds
820 with a synthetic data augmentation strategy and graph-structured deep metric learning,” *Autom Constr*, vol.
821 150, p. 104838, Jun. 2023, doi: 10.1016/J.AUTCON.2023.104838.
- 822] F. López de la Rosa, J. L. Gómez-Sirvent, R. Sánchez-Reolid, R. Morales, and A. Fernández-Caballero,
823 “Geometric transformation-based data augmentation on defect classification of segmented images of
824 semiconductor materials using a ResNet50 convolutional neural network,” *Expert Syst Appl*, vol. 206, p.
825 117731, Nov. 2022, doi: 10.1016/J.ESWA.2022.117731.
- 826] H. Jia *et al.*, “Assessing the Potential of Data Augmentation in EEG Functional Connectivity for Early
827 Detection of Alzheimer’s Disease,” *Cognit Comput*, vol. 16, no. 1, pp. 229–242, Jan. 2024, doi:
828 10.1007/S12559-023-10188-7/TABLES/5.
- 829] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, “A systematic review of deep learning
830 data augmentation in medical imaging: Recent advances and future research directions,” *Healthcare*
831 *Analytics*, vol. 5, p. 100340, Jun. 2024, doi: 10.1016/J.HEALTH.2024.100340.
- 832] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” *J Big*
833 *Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/s40537-019-0197-0.
- 834] F. Asadi, T. Angsuwatanakul, and J. A. O’Reilly, “Evaluating synthetic neuroimaging data augmentation
835 for automatic brain tumour segmentation with a deep fully-convolutional network,” *IBRO Neurosci Rep*,
836 vol. 16, pp. 57–66, Jun. 2024, doi: 10.1016/J.IBNEUR.2023.12.002.
- 837] S. Benbakreti, M. Benouis, A. Roumane, and S. Benbakreti, “Impact of the data augmentation on the
838 detection of brain tumor from MRI images based on CNN and pretrained models,” *Multimed Tools Appl*,
839 vol. 83, no. 13, pp. 39459–39478, Apr. 2024, doi: 10.1007/S11042-023-17092-0/FIGURES/7.
- 840] Z. Xu *et al.*, “Cross-domain attention-guided generative data augmentation for medical image analysis with
841 limited data,” *Comput Biol Med*, vol. 168, p. 107744, Jan. 2024, doi:
842 10.1016/J.COMPBIOMED.2023.107744.
- 843] J. Nalepa, M. Marcinkiewicz, and M. Kawulok, “Data Augmentation for Brain-Tumor Segmentation: A
844 Review,” *Front Comput Neurosci*, vol. 13, p. 83, Dec. 2019, doi: 10.3389/FNCOM.2019.00083/BIBTEX.
- 845] S. Jiang, X. Chen, and C. Yi, “SSA-UNet: Whole brain segmentation by U-Net with squeeze-and-excitation
846 block and self-attention block from the 2.5D slice image,” *IET Image Process*, 2024, doi:
847 10.1049/IPR2.13052.

848] S. Anari, S. Sadeghi, G. Sheikhi, R. Ranjbarzadeh, and M. Bendechache, "Explainable attention based
849 breast tumor segmentation using a combination of UNet, ResNet, DenseNet, and EfficientNet models,"
850 *Scientific Reports* 2025 15:1, vol. 15, no. 1, pp. 1–39, Jan. 2025, doi: 10.1038/s41598-024-84504-y.

851] A. M. Gab Allah, A. M. Sarhan, and N. M. Elshennawy, "Edge U-Net: Brain tumor segmentation using
852 MRI based on deep U-Net model with boundary information," *Expert Syst Appl*, vol. 213, p. 118833, Mar.
853 2023, doi: 10.1016/J.ESWA.2022.118833.

854] J. Zhang, J. Zeng, P. Qin, and L. Zhao, "Brain tumor segmentation of multi-modality MR images via triple
855 intersecting U-Nets," *Neurocomputing*, vol. 421, pp. 195–209, Jan. 2021, doi:
856 10.1016/J.NEUCOM.2020.09.016.

857] Z. Li, H. Zhang, Z. Li, and Z. Ren, "Residual-Attention UNet++: A Nested Residual-Attention U-Net for
858 Medical Image Segmentation," *Applied Sciences* 2022, Vol. 12, Page 7149, vol. 12, no. 14, p. 7149, Jul.
859 2022, doi: 10.3390/APP12147149.

860] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted Residuals and
861 Linear Bottlenecks," *Proceedings of the IEEE Computer Society Conference on Computer Vision and*
862 *Pattern Recognition*, pp. 4510–4520, Jan. 2018, doi: 10.1109/CVPR.2018.00474.

863] M. Bhalerao and S. Thakur, "Brain tumor segmentation based on 3D residual U-Net," *Lecture Notes in*
864 *Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*
865 *Bioinformatics)*, vol. 11993 LNCS, pp. 218–225, 2020, doi: 10.1007/978-3-030-46643-5_21/COVER/.

866] M. M. Zahoor *et al.*, "Brain Tumor MRI Classification Using a Novel Deep Residual and Regional CNN,"
867 *Biomedicines* 2024, Vol. 12, Page 1395, vol. 12, no. 7, p. 1395, Jun. 2024, doi:
868 10.3390/BIMEDICINES12071395.

869] N. M. Hassan, S. Hamad, and K. Mahar, "YOLO-based CAD framework with ViT transformer for breast
870 mass detection and classification in CESM and FFDM images," *Neural Comput Appl*, vol. 36, no. 12, pp.
871 6467–6496, Apr. 2024, doi: 10.1007/S00521-023-09364-5/TABLES/12.

872] Z. Zhao *et al.*, "SwinHR: Hemodynamic-powered hierarchical vision transformer for breast tumor
873 segmentation," *Comput Biol Med*, vol. 169, p. 107939, Feb. 2024, doi:
874 10.1016/J.COMPBIOMED.2024.107939.

875] P. T. Krishnan, P. Krishnadoss, M. Khandelwal, D. Gupta, A. Nihaal, and T. S. Kumar, "Enhancing brain
876 tumor detection in MRI with a rotation invariant Vision Transformer," *Frontiers in Neuroinformatics*, vol.
877 18, p. 1414925, Jun. 2024, doi: 10.3389/FNINF.2024.1414925/BIBTEX.

878] Y. Bi, A. Abrol, Z. Fu, and V. D. Calhoun, "A multimodal vision transformer for interpretable fusion of
879 functional and structural neuroimaging data," *Hum Brain Mapp*, vol. 45, no. 17, p. e26783, Dec. 2024, doi:
880 10.1002/HBM.26783.

881] J. Wang, S. Y. Lu, S. H. Wang, and Y. D. Zhang, "RanMerFormer: Randomized vision transformer with
882 token merging for brain tumor classification," *Neurocomputing*, vol. 573, p. 127216, Mar. 2024, doi:
883 10.1016/J.NEUCOM.2023.127216.

884] W. Mao *et al.*, "Brain Structural Connectivity Guided Vision Transformers for Identification of Functional
885 Connectivity Characteristics in Preterm Neonates," *IEEE J Biomed Health Inform*, vol. 28, no. 4, pp. 2223–
886 2234, Apr. 2024, doi: 10.1109/JBHI.2024.3355020.

887] S. Poornam and J. J. R. Angelina, "VITALT: a robust and efficient brain tumor detection system using
888 vision transformer with attention and linear transformation," *Neural Comput Appl*, vol. 36, no. 12, pp.
889 6403–6419, Apr. 2024, doi: 10.1007/S00521-023-09306-1/TABLES/7.

890] R. Azad *et al.*, "Beyond Self-Attention: Deformable Large Kernel Attention for Medical Image
891 Segmentation," 2024.

- 892] S. Saurav, A. Sharma, R. Saini, and S. Singh, "An attention-guided convolutional neural network for
893 automated classification of brain tumor from MRI," *Neural Comput Appl*, vol. 35, no. 3, pp. 2541–2560,
894 Jan. 2023, doi: 10.1007/S00521-022-07742-Z/TABLES/11.
- 895] R. Ranjbarzadeh *et al.*, "Combining DeepLabV3 with Attention Mechanisms for Accurate Brain Tumor
896 Segmentation: Insights from BraTS 2020 and a Private Clinical Dataset," *Proceedings - European*
897 *Workshop on Visual Information Processing, EUVIP*, 2024, doi: 10.1109/EUVIP61797.2024.10772776.
- 898] M. Kia, "Attention-guided deep learning for effective customer loyalty management and multi-criteria
899 decision analysis," *Iran Journal of Computer Science*, pp. 1–22, Dec. 2024, doi: 10.1007/S42044-024-
900 00215-7/FIGURES/9.
- 901] H. Liu, Q. Zhang, and Y. Liu, "Image Segmentation of Bladder Cancer Based on DeepLabv3+," *Lecture*
902 *Notes in Electrical Engineering*, vol. 805 LNEE, pp. 614–621, 2022, doi: 10.1007/978-981-16-6320-
903 8_62/COVER.
- 904] K. M. Sunnetci, E. Kaba, F. Beyazal Çeliker, and A. Alkan, "Comparative parotid gland segmentation by
905 using ResNet-18 and MobileNetV2 based DeepLab v3+ architectures from magnetic resonance images,"
906 *Concurr Comput*, vol. 35, no. 1, p. e7405, Jan. 2023, doi: 10.1002/CPE.7405.
- 907] M. Cai *et al.*, "Image Segmentation Method for Sweetgum Leaf Spots Based on an Improved DeeplabV3+
908 Network," *Forests 2022, Vol. 13, Page 2095*, vol. 13, no. 12, p. 2095, Dec. 2022, doi: 10.3390/F13122095.
- 909] H. Polat, "Multi-task semantic segmentation of CT images for COVID-19 infections using DeepLabV3+
910 based on dilated residual network," *Phys Eng Sci Med*, vol. 45, no. 2, pp. 443–455, Jun. 2022, doi:
911 10.1007/S13246-022-01110-W/TABLES/7.
- 912] I. Vykopal, L. Hudec, M. Kveton, O. Fabian, A. Felsoova, and W. Benesova, "DeepLabV3+ Ensemble
913 for Diagnosis of Cardiac Transplant Rejection," *Lecture Notes in Computer Science (including subseries*
914 *Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 14253 LNCS, pp. 112–
915 122, 2023, doi: 10.1007/978-3-031-44137-0_10/FIGURES/3.
- 916] Y. Wang *et al.*, "Improved brain metastases segmentation using generative adversarial network and
917 conditional random field optimization mask R-CNN," *Med Phys*, vol. 51, no. 9, pp. 5990–6001, Sep. 2024,
918 doi: 10.1002/MP.17176.
- 919] J. S. Park, S. Fadnavis, and E. Garyfallidis, "Multi-scale V-net architecture with deep feature CRF layers
920 for brain extraction," *Communications Medicine 2024 4:1*, vol. 4, no. 1, pp. 1–12, Feb. 2024, doi:
921 10.1038/s43856-024-00452-8.
- 922] S. Wu *et al.*, "Pelvic bone tumor segmentation fusion algorithm based on fully convolutional neural network
923 and conditional random field," *J Bone Oncol*, vol. 45, p. 100593, Apr. 2024, doi:
924 10.1016/J.JBO.2024.100593.
- 925] Z. Liu *et al.*, "CANet: Context Aware Network for Brain Glioma Segmentation," *IEEE Trans Med Imaging*,
926 vol. 40, no. 7, pp. 1763–1777, Jul. 2021, doi: 10.1109/TMI.2021.3065918.
- 927] B. Yang, W. Wu, Y. Liu, and H. Liu, "A Novel Sleep Stage Contextual Refinement Algorithm Leveraging
928 Conditional Random Fields," *IEEE Trans Instrum Meas*, vol. 71, 2022, doi: 10.1109/TIM.2022.3154838.
- 929] B. Zhang *et al.*, "Multi-scale segmentation squeeze-and-excitation UNet with conditional random field for
930 segmenting lung tumor from CT images," *Comput Methods Programs Biomed*, vol. 222, p. 106946, Jul.
931 2022, doi: 10.1016/J.CMPB.2022.106946.
- 932] N. G. Rezk, A. F. Attia, M. A. El-Rashidy, A. El-Sayed, and E. E. D. Hemdan, "An Efficient Plant Disease
933 Recognition System Using Hybrid Convolutional Neural Networks (CNNs) and Conditional Random
934 Fields (CRFs) for Smart IoT Applications in Agriculture," *International Journal of Computational*
935 *Intelligence Systems*, vol. 15, no. 1, pp. 1–21, Dec. 2022, doi: 10.1007/S44196-022-00129-X/TABLES/4.

- 936] A. Keles, O. Akcay, H. Kul, and M. Bendeche, "Saliency Maps as an Explainable AI Method in Medical
937 Imaging: A Case Study on Brain Tumor Classification," 2023, doi: 10.5281/ZENODO.8199333.
- 938] H. Mzoughi, I. Njeh, M. BenSlima, N. Farhat, and C. Mhiri, "Vision transformers (ViT) and deep
939 convolutional neural network (D-CNN)-based models for MRI brain primary tumors images multi-
940 classification supported by explainable artificial intelligence (XAI)," *Visual Computer*, pp. 1–20, Jun. 2024,
941 doi: 10.1007/S00371-024-03524-X/TABLES/9.
- 942] D. Muhammad and M. Bendeche, "Unveiling the black box: A systematic review of Explainable
943 Artificial Intelligence in medical image analysis," *Comput Struct Biotechnol J*, vol. 24, pp. 542–560, Dec.
944 2024, doi: 10.1016/J.CSBJ.2024.08.005.
- 945] MuhammadDost, Rafiullah, and BendecheMalika, "Improving diagnostic trust: an explainable deep
946 learning framework for genitourinary cancer prediction," *IET Conference Proceedings*, 2024, doi:
947 10.1049/ICP.2024.3275.
- 948] H. Liu, Z. Ni, D. Nie, D. Shen, J. Wang, and Z. Tang, "Multimodal Brain Tumor Segmentation Boosted by
949 Monomodal Normal Brain Images," *IEEE Transactions on Image Processing*, vol. 33, pp. 1199–1210,
950 2024, doi: 10.1109/TIP.2024.3359815.
- 951] A. Ali, M. Sharif, M. S. Faisal, A. Rizwan, G. Atteia, and M. Alabdulhafith, "Brain Tumor Segmentation
952 using Generative Adversarial Networks," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3450593.
- 953] K. S. Neetha and D. L. Narayan, "Segmentation and classification of brain tumour using LRIFCM and
954 LSTM," *Multimed Tools Appl*, vol. 83, no. 31, pp. 76705–76730, Sep. 2024, doi: 10.1007/S11042-024-
955 18478-4/TABLES/12.
- 956] S. R. Vinta, P. V. Chintalapati, G. R. Babu, R. Tamma, and G. Sai Chaitanya Kumar, "EDLNet: ensemble
957 deep learning network model for automatic brain tumor classification and segmentation," *J Biomol Struct*
958 *Dyn*, 2024, doi: 10.1080/07391102.2024.2311343.
- 959] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnU-Net for Brain Tumor
960 Segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*
961 *Intelligence and Lecture Notes in Bioinformatics)*, vol. 12659 LNCS, pp. 118–132, 2021, doi: 10.1007/978-
962 3-030-72087-2_11/COVER/.
- 963] I. Aboussaleh, J. Riffi, K. el Fazazy, A. M. Mahraz, and H. Tairi, "3DUV-NetR+: A 3D hybrid semantic
964 architecture using transformers for brain tumor segmentation with MultiModal MR images," *Results in*
965 *Engineering*, vol. 21, p. 101892, Mar. 2024, doi: 10.1016/J.RINENG.2024.101892.
- 966