Article

PRA-Unet: Parallel Residual Attention U-Net for Real-Time Segmentation of Brain Tumors

Ali Zakaria Lebani [1,2,*], Medjeded Merati [2,3,*] and Saïd Mahmoudi [4,*]

1 Laboratoire de Génie Energétique et Génie Informatique (L2GEGI), University of Tiaret, Tiaret 14000, Algeria
2 Department of Computer Science, University of Tiaret, Tiaret 14000, Algeria
3 Laboratoire d'Informatique et Mathematique (LIM), University of Tiaret, Tiaret 14000, Algeria
4 Department of Computer Science, University of Mons, 7000 Mons, Belgium
* Correspondence: alizakaria.lebani@univ-tiaret.dz (A.Z.L.); medjeded.merati@univ-tiaret.dz (M.M.); said.mahmoudi@umons.ac.be (S.M.)

Abstract

With the increasing prevalence of brain tumors, it becomes crucial to ensure fast and reliable segmentation in MRI scans. Medical professionals struggle with manual tumor segmentation due to its exhausting and time-consuming nature. Automated segmentation speeds up decision-making and diagnosis; however, achieving an optimal balance between accuracy and computational cost remains a significant challenge. In many cases, current methods trade speed for accuracy, or vice versa, consuming substantial computing power and making them difficult to use on devices with limited resources. To address this issue, we present PRA-UNet, a lightweight deep learning model optimized for fast and accurate 2D brain tumor segmentation. Using a single 2D input, the architecture processes four types of MRI scans (FLAIR, T1, T1c, and T2). The encoder uses inverted residual blocks and bottleneck residual blocks to capture features at different scales effectively. The Convolutional Block Attention Module (CBAM) and the Spatial Attention Module (SAM) improve the bridge and skip connections by refining feature maps and making it easier to detect and localize brain tumors. The decoder uses depthwise separable convolutions, which significantly reduce computational costs without degrading accuracy. The BraTS2020 dataset shows that PRA-UNet achieves a Dice score of 95.71%, an accuracy of 99.61%, and a processing speed of 60 ms per image, enabling real-time analysis. PRA-UNet outperforms other models in segmentation while requiring less computing power, suggesting it could be suitable for deployment on lightweight edge devices in clinical settings. Its speed and reliability enable radiologists to diagnose tumors quickly and accurately, enhancing practical medical applications.

Keywords: brain tumor segmentation; MRI; U-Net; attention mechanisms; real-time; parallel residual; edge devices

1. Introduction

In neuro-oncology, brain tumors are difficult to diagnose and treat, and early detection with accurate characterization is essential for planning effective treatment and improving clinical outcomes. Magnetic Resonance Imaging (MRI) is a key part of this process; it provides detailed views of the brain's structures and any pathological changes. However, manually interpreting MRI scans is time-consuming, subjective, and heavily dependent on the clinician's expertise, which highlights the need for automated segmentation methods [1].

In the past, brain tumor segmentation relied on traditional image processing methods such as thresholding, edge detection, and region-growing algorithms [2]. Later, machine learning methods such as Support Vector Machines (SVMs) and Random Forests were introduced, but they still relied on hand-crafted features [3,4]. These methods were a good start for automated analysis, but they struggled with noise, anatomical variation, and the natural variability of tumors.

Deep Learning (DL), particularly Convolutional Neural Networks (CNNs), has substantially transformed medical image segmentation by enabling the automatic extraction of features from raw data [5]. U-Net has become the gold standard architecture among these. It uses an encoder–decoder architecture with skip connections to maintain spatial accuracy while also capturing contextual information [1]. Many variations have been developed based on its success, each aiming to improve the accuracy and speed of segmentation [6,7].

Despite these improvements, significant challenges remain:

- Tumor variability: Irregular shapes, sizes, and intensity distributions complicate precise boundary delineation [8].
- Multi-modal MRI integration: Different MRI sequences (FLAIR, T1, T1c, and T2) provide distinct diagnostic information, making it challenging to design strategies that effectively exploit their complementarity.
- Computational efficiency: Highly accurate segmentation models often require substantial computational resources, limiting their real-time application in clinical settings [5,6].
- Compatibility with edge devices: Models deployed on low-power edge devices must achieve high accuracy within stringent resource constraints, emphasizing lightweight architectures [8].

To address these interrelated challenges, we present PRA-UNet—an optimized 2D CNN architecture tailored for robust brain tumor segmentation. The model integrates architectural innovations and efficiency-oriented components, each addressing a specific issue identified above:

- Handling tumor variability: PRA-UNet incorporates attention mechanisms, including the Convolutional Block Attention Module (CBAM) in the bridge and the Spatial Attention Module (SAM) in skip connections, thereby enhancing the network's ability to accurately capture tumor boundaries and variability.
- Effective multi-modal MRI integration: The architecture accepts input patches of size $256 \times 256 \times 4$, enabling the model to jointly process multiple MRI modalities and extract complementary features, thereby improving segmentation reliability.
- Improving computational efficiency: The encoder employs bottleneck residual blocks and inverted residual blocks, carefully balancing model performance with computational requirements, thus enabling efficient segmentation suitable for clinical real-time use.
- Ensuring compatibility with edge devices: The decoder utilizes depthwise separable convolutions (DSC), significantly reducing computational complexity without sacrificing accuracy, making PRA-UNet theoretically suitable for deployment on resource-limited edge devices.

We use standard metrics such as the Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) [9] to evaluate PRA-UNet on the widely used BraTS 2020 dataset [10]. Our results show that the model can achieve high accuracy while remaining fast enough for real-time clinical applications.

The remainder of this paper is structured as follows: In Section 2, we review prior related work on CNN-based segmentation, U-Net variants, and how they can be applied

to the study of brain tumors. Section 3 provides more detail on the proposed method, the experimental setup, the dataset, and the evaluation metrics. In Section 4, you will find the results, the discussion, the ablation study, the validation on another dataset, the deployment prospects and clinical integration, and the limitations. Finally, Section 5 wraps up the paper and outlines potential areas for future research.

2. Related Work

Deep learning has significantly improved image segmentation, particularly through CNNs. Various architectures have been proposed to enhance segmentation accuracy, robustness, and efficiency. Among them, U-Net and its variants have demonstrated strong performance in medical applications.

This section reviews key CNN-based segmentation models, explores U-Net's evolution, and discusses its application in brain tumor segmentation.

2.1. Segmentation Models Based on CNN

Image segmentation is a crucial aspect of computer vision since it enables robots to comprehend visual information at the pixel level. Segmentation models have progressed significantly and are now applied to various domains, including autonomous vehicles and medical imaging. This section examines the latest models and analyzes their construction and functionality.

Chen et al. [11] proposed DeepLabv3+, an architecture for semantic segmentation that integrates an encoder–decoder framework with an Atrous Spatial Pyramid Pooling (ASPP) module. They employ depthwise separable atrous convolutions and the Xception model to enhance feature extraction, achieving an optimal balance between speed and accuracy.

Strudel et al. [12] developed Segmenter, a transformer-based approach for semantic segmentation. It employs the Vision Transformer (ViT) to obtain global context from the initial layer. The model segments images into patches, encodes them using a transformer, and subsequently employs either a linear decoder or a mask-based decoder. The model was pre-trained for picture classification.

Cheng et al. [13] developed Mask2Former, an image segmentation model that employs masked attention to concentrate analysis on certain regions. It employs a multi-scale approach to enhance accuracy and optimizes the Transformer Decoder's performance by rearranging the modules and enabling learnable queries. Assessing the loss on a sample subset of points rather than the entire mask minimizes the memory requirements for training.

Wang et al. [14] introduced HRNet, a neural network that maintains high-resolution representations alongside multi-resolution streams. It prevents detail loss by continuously integrating information from these streams, distinguishing it from conventional models. This approach enhances the precision of tasks such as human pose estimation and semantic segmentation.

PointRend is an image segmentation approach developed by Kirillov et al. [15] utilizing rendering methodologies. PointRend distinguishes itself from conventional approaches that operate on a regular grid by dynamically selecting critical spots to enhance segmentation, particularly at object borders. It employs an iterative subdivision approach to enhance precision and expedite computations. It is included in models such as Mask R-CNN and significantly enhances mask resolution without increasing processing time.

Caron et al. [16] developed DINO-seg, a method enabling Vision Transformers (ViT) to learn autonomously. To enhance training stability, they employ a momentum encoder, multi-scale learning with multi-crop training, and label-free self-distillation to refine image representations. The utilization of tiny patches facilitates the segregation of things.

Segmentation models exhibit considerable diversity. Some use CNN-based designs such as DeepLabV3+ and HRNet, while others adopt transformer-based approaches like Segmenter and DINO-Seg. These models leverage sophisticated techniques to improve precision and enhance object representation.

2.2. U-Net and Its Variants

Numerous CNN models exist for segmentation tasks; however, the U-Net architecture and its specialized variants for medical imaging have demonstrated superior reliability and effectiveness. These architectures are specifically designed to address challenges in biomedical image segmentation, rendering them superior to conventional CNN approaches. This section elaborates on U-Net and its significant iterations, highlighting their advancements and contributions.

In 2015, Ronneberger et al. presented the U-Net architecture [1]. It is crucial to biological picture segmentation, using a contracting path for context encoding and an expansive path for accurate localization. It integrates skip connections to link these pathways, resulting in optimal performance. In 2018, Zhou et al. expanded upon this concept and developed U-Net++, which employed layered dense skip connections to enhance feature fusion and representation learning [17]. Concurrently, Oktay et al. developed the Attention U-Net [18], which used attention gates to emphasize critical regions and enhance segmentation in challenging scenarios.

ResU-Net, developed by Diakogiannis et al. [19], represents a significant advancement. It relies on residual connections to enhance gradient flow, facilitating effective performance with large datasets such as remote sensing photos. Najme et al. proposed a Squeeze-and-Excitation (SE) Dense-UNet model [20], enhancing the original Dense-UNet by incorporating SE blocks with GeLU activation. This enhances the synergy among channels, reduces overfitting, and facilitates the segmentation of lung CT scans. Alom et al. devised R2U-Net [21], which combines residual and recurrent methods to exploit spatiotemporal characteristics.

Chen et al. recently integrated transformers into TransUNet [22], merging the global contextual learning capabilities of transformers with the structural benefits of U-Net. Cao et al. subsequently developed Swin-Unet [23], which employed Swin Transformers to exhibit features at various scales and established benchmarks for multi-organ segmentation. Finally, Hatamizadeh et al. developed UNETR in 2022 [24]. It uses pure transformers rather than conventional encoders for medical imaging, yielding optimal results.

U-Net and its derivatives have revolutionized biomedical image segmentation by integrating an architecture designed to address specific challenges in the domain. They have achieved enhanced efficiency and accuracy due to a few advancements, including dense connections, attention mechanisms, and transformers.

2.3. U-Net for Brain Tumor Segmentation

Leveraging U-Net's established efficacy in medical picture segmentation is particularly crucial for the challenging endeavor of brain tumor segmentation. This endeavor requires models capable of delivering exceptional accuracy and specialization to yield precise and dependable results. This section examines how various studies using U-Net and its adaptations have contributed to brain tumor segmentation.

MPB-UNet [25] utilizes multiscale parallel pathways to extract features from brain tumors and uses Atrous Spatial Pyramid Pooling (ASPP) to capture contextual information across multiple scales. MAU-Net [26] implements spatial-channel attention and self-attention techniques to enhance the representation of local and global data. TransDoubleU-

Net [27] uses two U-Nets integrated with Swin Transformers to obtain multiscale characteristics that enhance segmentation accuracy.

In [28], the authors examined a deep learning methodology that integrates tumor segmentation using UNet-based convolutional neural networks (CNNs) with tumor grading by transfer learning with a pretrained VGG16 model and a fully connected classifier. In another study, an efficient hybrid U-Net [29] with a ResNet50 encoder was implemented for brain MRI segmentation, using images resized to 256 × 256 pixels. U-Net++ [30] utilizes dense connections to mitigate vanishing gradients and improve feature reuse. The ACU-Net [31] model has improved brain tumor segmentation by integrating an attention mechanism into an optimized U-Net architecture. It effectively extracts essential features from fMRI images while reducing computational complexity, thereby surpassing traditional approaches in accuracy. Table 1 below provides a comparative overview and highlights the techniques, datasets, and segmentation performance, particularly the Dice score.

Table 1. Summary of Techniques, Datasets, and Results in Brain Tumor Segmentation with U-Net Architecture.

Study	Technique	Dataset	Dice Score
MPB-UNet (2025) [25]	Multi-scale parallel paths and ASPP.	LGG dataset (Kaggle)	99.80%
MAU-Net (2023) [26]	Attention mechanisms (spatial, channel, self).	BraTS 2019, 2020	90.00% (BraTS 2020)
TransDoubleU-Net (2023) [27]	Dual U-Nets with Swin Transformers.	BraTS 2019, 2020	92.87% (BraTS 2020)
UNet (2020) [28]	UNet and VGG16 with transfer learning.	LGG dataset (Kaggle)	84%
An efficient U-Net (2024) [29]	A hybrid U-Net with a ResNet50 encoder.	LGG dataset (Kaggle)	82%
U-Net++ with Dense (2024) [30]	Dense connections with improved skip paths.	BraTS 2020	93%
ACU-Ne (2025) [31]	U-Net with spatial attention mechanism.	BraTS 2020	98.59%

The majority of these models prioritize improving accuracy, often at the cost of increased computational complexity, hindering the achievement of an optimal balance between performance and efficiency. This raises the need for a new architecture that strikes this balance, enabling real-time segmentation without compromising precision.

3. Methodology

The proposed 2D architecture, named PRA-UNet, is designed to perform the semantic segmentation of brain tumors from MRI images, balancing accuracy and computational efficiency. To enhance segmentation precision, the input resolution is set to 256 × 256 pixels to preserve key tumor-related features. Moreover, PRA-UNet integrates four MRI modalities—T1, T1c, T2, and FLAIR—into a single multichannel tensor of shape (256 × 256 × 4), leveraging their complementary information to improve tumor delineation.

Although 3D segmentation provides valuable volumetric context, this study did not adopt it for two main reasons. First, 3D architectures require substantial GPU memory and longer training times, making them incompatible with real-time inference and deployment on resource-limited clinical systems. Second, in neuro-oncology practice, clinicians commonly evaluate tumor size and treatment response using two-dimensional measurements on the slice showing the maximal tumor extent. According to the updated RANO 2.0 criteria for glioma response assessment, the maximum cross-sectional tumor area (a 2D

measurement) remains the primary indicator of tumor burden, with complete volumetric analysis optional [32].

While retaining its 2D formulation, the architecture remains adaptable to resource-constrained environments through optimized strategies that reduce processing overhead. Figure 1 presents an overview of the main stages and components of the proposed architecture, whereas Table 2 details the corresponding building blocks, input shapes, and filter configurations at each stage.
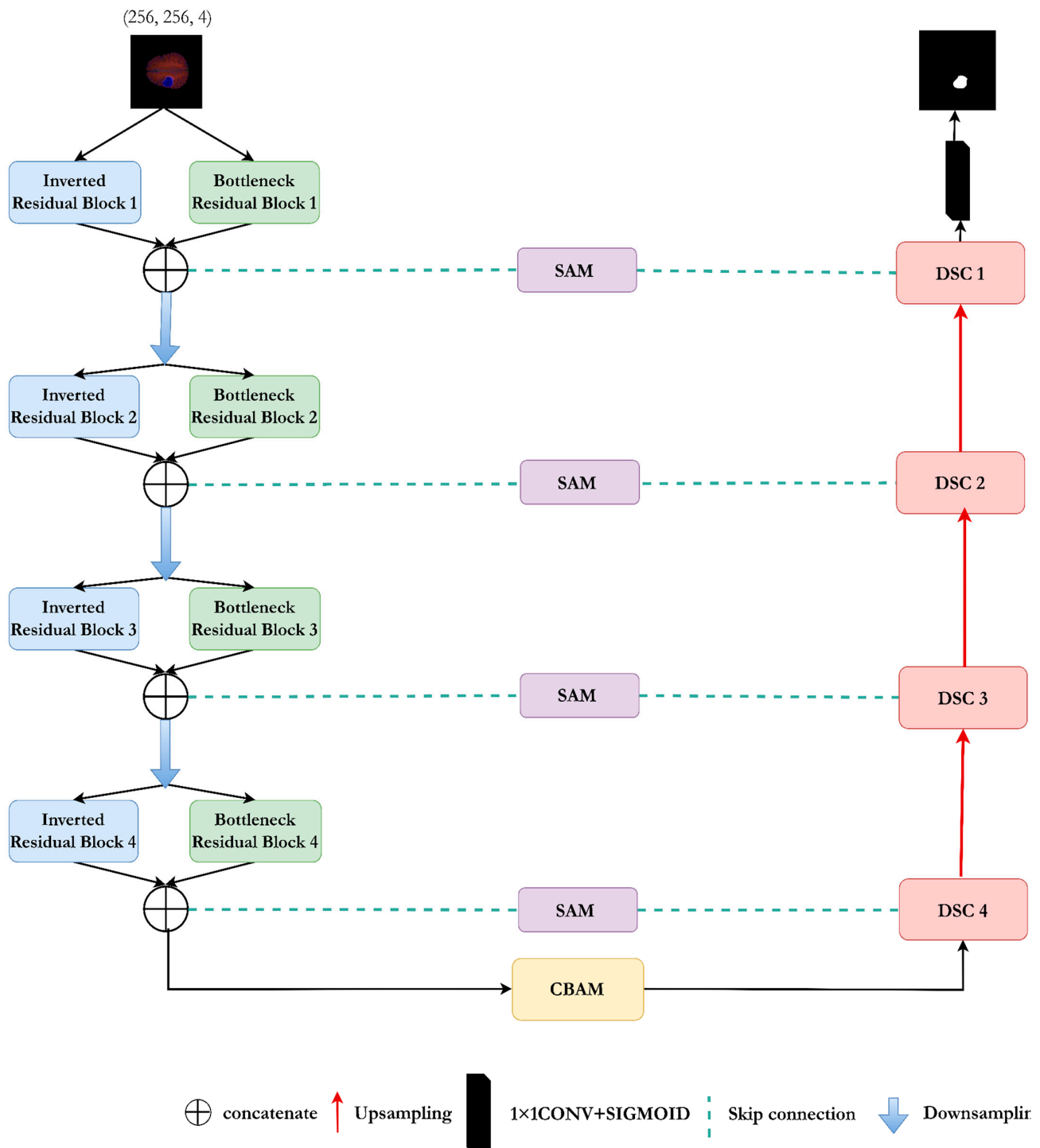

Figure 1. Architectural Overview of PRA-UNet Illustrating the Encoder Path, Bridge Attention Mechanism, and Decoder Reconstruction Stages.

Table 2. Summary of PRA-UNet Architecture: Blocks, Input Shapes, and Number of Filters.

Encoder Block	Input	Number of Filters
Residual Bottleneck 1	(256,256,4)	32
Inverted Residual 1	(256,256,4)	32
Residual Bottleneck 2	(128,128,64)	64
Inverted Residual 2	(128,128,64)	64
Residual Bottleneck 3	(64,64,128)	128
Inverted Residual 3	(64,64,128)	128
Residual Bottleneck 4	(32,32,256)	256
Inverted Residual 4	(32,32,256)	256
DSC 4	(32,32,512)	256
DSC 3	(64,64,256)	128
DSC 2	(128,128,128)	64
DSC 1	(256,256,64)	32

3.1. Encoder

The encoder consists of four hierarchical levels. At each level, feature extraction is performed using two parallel blocks: a Bottleneck Residual Block and an Inverted Residual Block [33,34]. This dual-block structure enriches the representation while keeping the computational cost low. The Inverted Residual Block focuses on global context by expanding channels and applying depthwise convolution, whereas the Bottleneck Residual Block emphasizes local details by temporarily reducing the number of channels. The outputs of both blocks are concatenated and passed through a 2×2 max-pooling operation, reducing the spatial resolution by a factor of 2 at each layer. This design ensures efficient and progressive extraction of multi-scale features.

3.1.1. Bottleneck Residual Block

The Bottleneck Residual Block module, inspired by ResNet [34], balances representational power and computational efficiency. It emphasizes local feature extraction through four independent phases:

1. Channel reduction using a 1×1 convolution: This step reduces the number of channels from C to $\frac{C}{r}$ using a 1×1 convolution. It focuses on the most critical features, reducing unnecessary information and computational load. In our work, r is empirically set to 4, as in ResNet [34].

2. Spatial feature extraction: A 3×3 convolution is then applied to the reduced channels to extract essential spatial details, such as edges and textures. Fewer channels at this stage mean fewer computations.

3. Restoration of the original dimension using a 1×1 convolution: Here, another 1×1 convolution restores the channel count to its original value, C. This step combines extracted features into a detailed feature map without significantly increasing computational cost.

4. Residual Connection: The original input features combine with the output of previous convolutions through a residual connection (Figure 2a). This improves gradient flow, speeds up learning, and enhances segmentation accuracy. If the number of input and output channels differs, an extra 1×1 convolution ensures matching dimensions (Figure 2b).

The Bottleneck Residual Block achieves its efficiency primarily through the innovative use of 1×1 convolutions, which capture fine details and efficiently manage channel interactions. Unlike 3×3 convolutions, which handle broader contexts, these smaller convolutions focus on local details. Stacking multiple such blocks progressively enhances

the ability to recognize complex spatial patterns, which are essential for accurate medical image segmentation. Figure 2 illustrates the structure of the Bottleneck Residual Block.
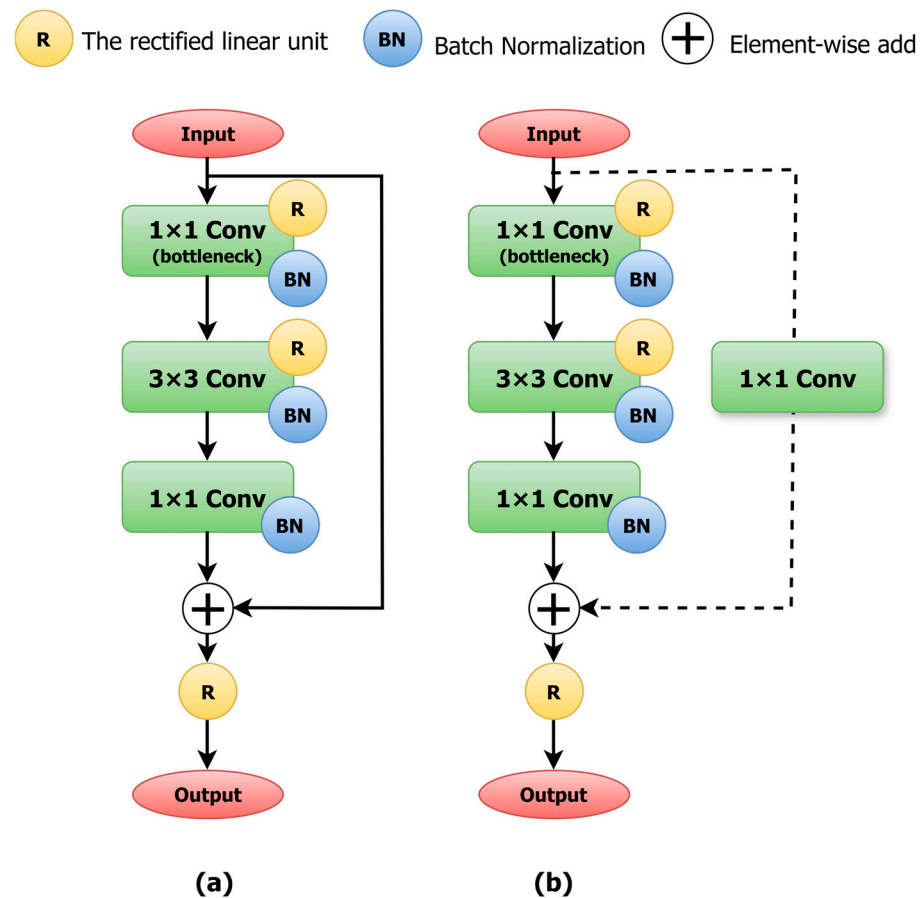

Figure 2. Bottleneck Residual Block: (**a**) Bottleneck Residual Block with $C_{in} \neq C_{out}$. (**b**) Bottleneck Residual Block with $C_{in} = C_{out}$.

3.1.2. Inverted Residual Block

The Inverted Residual Block, derived from MobileNetV2 [33], is integrated into PRA-UNet to efficiently achieve accurate brain tumor segmentation in real-time. This block consists of four stages:

1. Expansion (1×1 Convolution): A 1×1 convolution expands the number of channels from C_{in} to C_{exp}, where $C_{exp} = t \times C_{in}$ and $t = 6$, as in MobileNetV2 [33]. This expansion increases the model's capacity to capture global and complex spatial features. A ReLU 6 activation follows this step to ensure non-linearity and numerical stability.
2. Depthwise 3×3 Convolution: A 3×3 depthwise convolution is applied independently to each channel, focusing on local patterns such as edges and textures. It avoids inter-channel computations, which reduces complexity. Another ReLU6 activation is applied to enhance the feature representations.
3. Channel Reduction (1×1 Convolution): A second 1×1 convolution reduces the channel dimensions from C_{exp} back to C_{out}. This step limits the number of parameters, keeps the model efficient, and preserves key information.
4. Residual Connection (if applicable): When $C_{in} = C_{out}$ and the stride is 1, a residual connection is used (Figure 3a), combining input and output. This helps retain useful features and improves gradient flow during training. If $C_{in} \neq C_{out}$ or stride equals 2, the residual connection is omitted (Figure 3b).
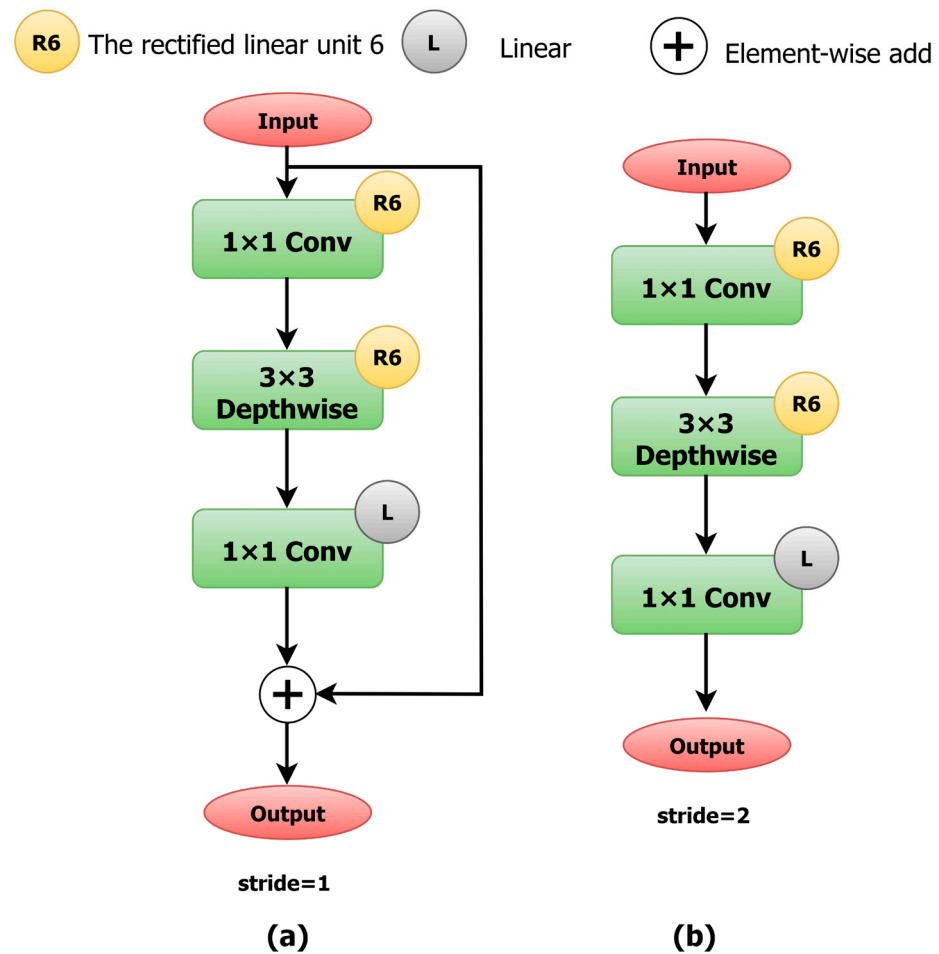
Figure 3. Structure of the Inverted Residual Block. (**a**) Stride-2 block without a residual connection (**b**) Stride-1 block with a residual connection.

The Inverted Residual Block reduces computational complexity by using depthwise convolutions. A standard convolution has a complexity of:

$$O\left(H \times W \times C_{in} \times C_{out} \times K^2\right),\tag{1}$$

where H and W denote the height and width of the feature map, respectively, C_{in} is the number of input channels, C_{out} is the number of output channels, C_{exp} is the expanded number of channels in the inverted residual block, and K is the kernel size. Depthwise convolution reduces this to:

$$O\left(H \times W \times C_{exp} \times K^2\right),\tag{2}$$

This reduction in operations enables PRA-UNet to perform real-time segmentation while maintaining high accuracy. Figure 3 illustrates the structure of the Inverted Residual Block and its use of residual connections based on stride and channel compatibility.

3.2. Bridge

In our brain tumor segmentation approach, we integrate the CBAM [35] as a bridge between the encoder and decoder to enhance the representation of extracted features before the spatial reconstruction phase. CBAM applies a sequential attention mechanism on the channel and spatial dimensions, allowing it to emphasize regions of interest and filter out irrelevant information. This approach is particularly beneficial for distinguishing tumor

tissues from healthy brain structures. The feature refinement process through CBAM follows a sequential update defined by:

$$F' = M_c(F) \otimes F, \tag{3}$$

$$F'' = M_s(F') \otimes F', \tag{4}$$

where F denotes the original feature map produced by the encoder, F' corresponds to the feature representation refined through channel attention, and F'' designates the final spatially enhanced feature map forwarded to the decoder. The terms $M_c(F)$ and $M_s(F)$ are the weighting maps generated by the channel and spatial attention modules, respectively, and $\otimes$ denotes element-wise multiplication.

Figure 4 illustrates the role of CBAM in improving the representation of relevant features for tumor segmentation. The generated attention maps help reduce the influence of irrelevant structures and enhance segmentation accuracy, which is crucial for improving tumor delineation in clinical applications.
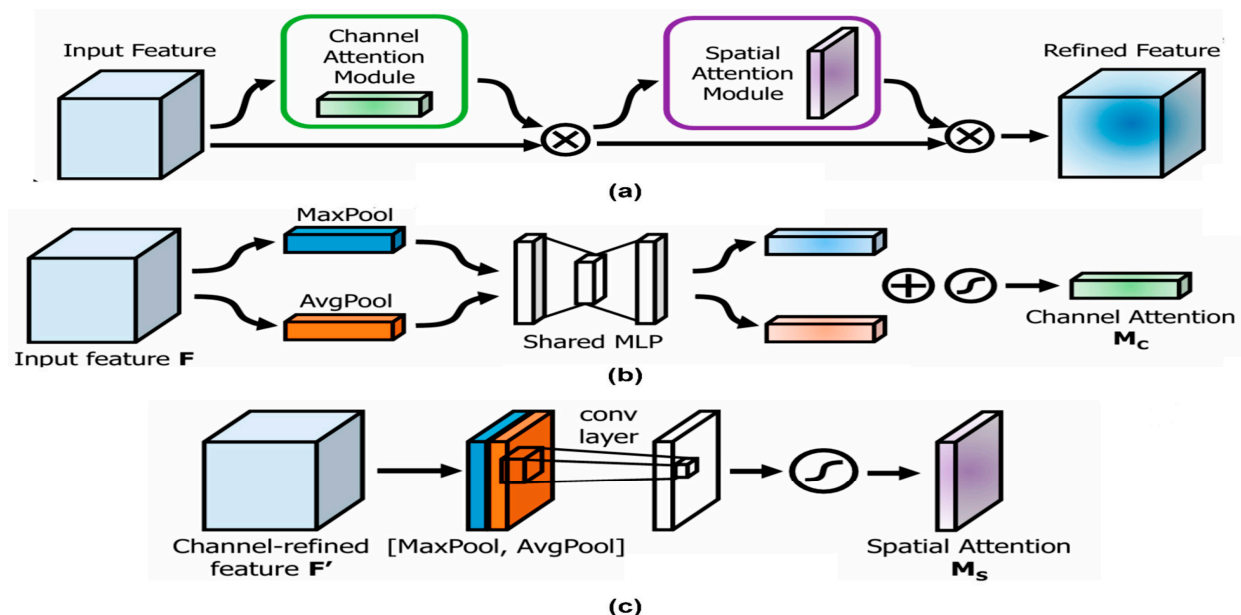

Figure 4. (**a**) General CBAM Framework, (**b**) Channel Attention Module, (**c**) Spatial Attention Module [35].

To elucidate this mechanism in greater detail, channel attention assesses the importance of each feature map by combining information from Global Average Pooling and Global Max Pooling [36]. These operations summarize the information in each channel as global statistics, which are then processed by a Multi-Layer Perceptron (MLP). The output is normalized through a sigmoid activation function σ, allowing an adaptive weighting of each channel and amplifying those most relevant for tumor segmentation. This process is defined by:

$$M_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))), \tag{5}$$

where F is the input feature map, $AvgPool(F)$ and $MaxPool(F)$ represent the global average and maximum pooling operations, respectively, and $MLP(\cdot)$ is a multi-layer perceptron.

Spatial attention then refines tumor localization by leveraging the feature maps adjusted by channel attention. It is based on a 7×7 convolution applied to the concatenation

of the maps obtained through Global Average Pooling and Global Max Pooling, generating a spatial weighting map defined as [37]:

$$M_s(F) = \sigma(Conv_{7\times7}([AvgPool(F); MaxPool(F)])), \tag{6}$$

where $Conv_{7\times7}$ represents a convolution with a 7×7 kernel applied to the concatenated maps $AvgPool(F)$ and $MaxPool(F)$, capturing essential spatial relationships for accurate tumor segmentation.

3.3. Decoder

The decoder reconstructs a high-resolution segmentation map from the encoded features. It adopts a symmetric structure aligned with the encoder and includes three core components: DSC blocks, skip connections (SC), and a final classification layer. DSC blocks refine feature maps while keeping computational cost low [38]. Skip connections transmit detailed spatial information from the encoder to the decoder. These features are refined using a Spatial Attention Module (SAM), which enhances relevant regions and suppresses non-informative areas [39]. Finally, a 1×1 convolution followed by a sigmoid activation produces a probabilistic segmentation map that quantifies tumor likelihood at the pixel level; a thresholding step converts this representation into a binary mask. This decoder design ensures accurate reconstruction while preserving computational efficiency and compactness for deployment on edge devices.

3.3.1. DSC Block Architecture

Each DSC block consists of two operations. A Depthwise Convolution (DWConv) with a 3×3 kernel applies independently to each input channel to extract spatial information. A Pointwise Convolution (PConv) with a 1×1 kernel follows this to integrate channel-wise information. Each step follows Batch Normalization (BN) and ReLU activation (see Figure 5). Each DSC block is followed by an up-sampling operation ($\times 2$) to restore spatial resolution. The Floating-Point Operations (FLOPs) required for DSC versus standard convolution are given by:

$$FLOPs_{DSC} = C_{in} \times K^2 \times H \times W + C_{in} \times C_{out} \times H \times W, \tag{7}$$

$$FLOPs_{conv} = C_{in} \times C_{out} \times K^2 \times H \times W, \tag{8}$$

where C_{in} and C_{out} are the input and output channels, K is the kernel size, and H and W are the spatial dimensions.

Figure 5. Depthwise Separable Convolution Block Architecture.

3.3.2. Integration of SAM

Skip connections transfer feature maps from the encoder to decoder layers at the same spatial resolution. To enhance their informativeness, SAM is applied before concatenation. SAM combines average and max pooling across the channel axis, followed by a 7×7 convolution and sigmoid activation, generating an attention map $M_s(F)$. This modulates SC features by:

$$F_{output} = F_{skip} \cdot M_s(F), \tag{9}$$

This enhances relevant spatial regions and reduces noise. The architecture of SAM is shown in block (c) of Figure 4. It improves spatial selectivity and supports precise segmentation in heterogeneous tumor areas [39].

3.3.3. Final Segmentation Map

The decoder output passes through a 1×1 convolution to reduce the number of channels, followed by a sigmoid activation function to generate the binary segmentation map. The resulting pixel values are normalized in the range [0, 1], using a fixed cutoff of 0.5, where pixels with values ≥ 0.5 are classified as tumor and those below this threshold are considered background. This probabilistic output enables pixel-level interpretation, supporting accurate tumor delineation while maintaining low computational cost.

3.4. Datasets

We used a well-organized and publicly available dataset to evaluate the performance of the PRA-UNet model for brain tumor segmentation. The following sections describe the dataset, preprocessing steps, data augmentation procedures, and validation strategy.

3.4.1. BraTS2020

We chose the BraTS2020 dataset [10,40,41] because it is known for being consistent and useful in clinical settings. There are 369 cases, and each case includes four types of brain MRIs: FLAIR, T1, T1c (T1 with contrast), and T2. The resolution of all volumes is $240 \times 240 \times 155$ voxels. We obtained the dataset from Kaggle and used it for our experiments [42].

3.4.2. Preprocessing

We selected one axial slice from each case to adapt the 3D MRI volumes into the 2D structure of PRA-UNet. The slice selection relies on the segmentation mask by identifying the slice containing the largest number of tumor pixels, ensuring that training focuses on the region with the highest informative content. After slice selection, each image is resized to 256×256 pixels to maintain consistent input dimensions.

The four MRI modalities (T1, T1c, T2, and FLAIR) were then concatenated into a single multi-channel image of shape (256, 256, 4). Combining these sequences allows the model to benefit from their complementary contrast mechanisms and capture a richer spectrum of anatomical and pathological information for whole-tumor segmentation. Specifically, T1 provides structural detail, T1c highlights contrast-enhancing components, T2 emphasizes fluid-related regions, and FLAIR accentuates peritumoral edema. Integrating these modalities within a unified representation strengthens the network's ability to delineate the tumor as a single region of interest.

Figure 6 presents examples of the selected axial slices from the four modalities and their integrated representation with the tumor mask. To support open science and reproducibility, the preprocessed dataset is publicly available on Kaggle [43].

3.4.3. Data Augmentation

Before augmentation, the dataset was divided at the patient level, with 300 patients assigned to the training set and 69 to the test set. This strict patient-wise separation ensured that no augmented instance from a given subject appeared in both subsets, preventing any form of data leakage. After this split, data augmentation was applied independently to the training and test subsets. While the augmentation of the training data follows standard practice to improve generalization, controlled augmentation of the test subset was introduced to compensate for the limited size of the raw test set and to assess model stability and robustness under a wider range of transformation scenarios. The augmentation

pipeline included centered and random cropping, rotations of 90°, 180°, and 270°, and horizontal and vertical flips. These transformations increased visual variability and enabled more rigorous stress-testing of the model's behavior. After augmentation and preprocessing, the final dataset consisted of 3500 images, with 3000 reserved for training and validation and 500 forming the fixed, independent test set used for evaluation.
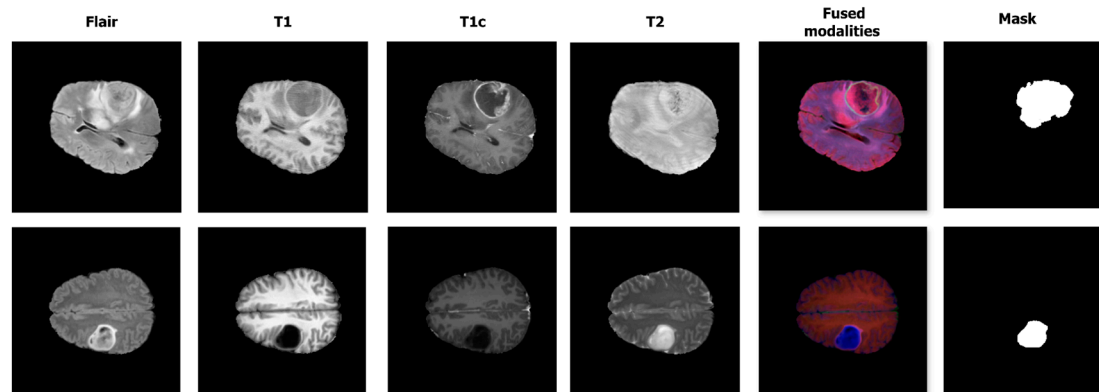

Figure 6. BRATS 2020: Axial View of Multi-Modal MRI Scans (FLAIR, T1, T1c, T2) and Fused Image with Brain Tumor Segmentation Mask.

3.4.4. Cross-Validation

We used a five-fold cross-validation scheme to evaluate model performance, but we applied it only to the training subset derived from the 300 patients. We partitioned the 3000 augmented training images into five folds, and in each iteration, we used four folds for training and the remaining fold for validation. We repeated this procedure five times to obtain a balanced and reliable estimate of performance across the training distribution. Throughout the entire process, the test set, composed of 500 images from 69 independent patients, remained fixed and was never included in any training or validation split. All evaluations focused on whole-tumor segmentation to measure the model's accuracy and consistency.

3.5. Experimental Configuration

We implemented all experiments using Python 3.8 and TensorFlow 2.11.0. We used additional libraries, including NumPy 1.23.5, OpenCV 4.7.0, scikit-learn 1.2.2, and Matplotlib 3.7.1. We executed the code in a Windows 10 environment with CUDA 11.2 and cuDNN 8.1, using an AMD Ryzen 7 5800X processor, an NVIDIA GeForce RTX 3060 GPU, 16 GB of DDR4 RAM, and a 512 GB SSD. To address class imbalance and improve segmentation accuracy, we adopted a composite loss function combining Dice Loss and Binary Cross-Entropy (BCE) Loss. The loss function is formulated as follows:

$$L = \alpha \cdot Dice\ Loss + \beta \cdot Binary\ Cross - Entropy\ Loss, \tag{10}$$

In all experiments, the coefficients were fixed at $\alpha = 0.7$ and $\beta = 0.3$. This choice is supported by the ablation study in Section 4.2.2, which shows that the Dice-dominated composite loss achieves better Dice and IoU scores than both single-loss baselines and the balanced configuration.

The Adam optimization algorithm was employed with an initial learning rate of 0.0001, which was progressively reduced using a cosine annealing schedule to refine convergence in later epochs. The training process comprised 50 epochs with a batch size of 16, ensuring stable optimization and model convergence.

3.6. Evaluation Metrics

The performance evaluation of the proposed approach is based on three fundamental metrics: Accuracy, Dice, and IoU. These metrics are widely used in segmentation to quantify the similarity between model predictions and reference data (ground truth). Mathematically, the Accuracy is expressed as:

$$Accuracy \ = \ \frac{TP \ + TN}{TP \ + TN + FP \ + FN}, \tag{11}$$

While the Dice and IoU are defined as:

$$Dice \ Score = \frac{2TP}{2TP \ + FN \ + FP}, \tag{12}$$

$$IoU = \frac{TP}{TP + FP + FN}, \tag{13}$$

In these equations:

- True Positive (TP): The number of brain tumor pixels correctly identified.
- False Positive (FP): The number of normal brain tissue pixels incorrectly classified as brain tumor pixels.
- True Negative (TN): The number of normal brain tissue pixels correctly classified.
- False Negative (FN): The number of brain tumor pixels incorrectly classified as normal brain tissue pixels.

The Dice is mainly used to assess the similarity between detected and reference regions, providing a measure of relative agreement. On the other hand, the IoU gives an exact measure of how predictions and annotations overlap. Accuracy is also a key measure of how many pixels in the whole image are correctly classified. This makes the evaluation of segmentation performance more complete. In addition to these core evaluation metrics, several statistical measures were incorporated to provide a more comprehensive and reliable performance analysis. These include the average (Avg), standard deviation (σ), standard error of the mean (SEM), and confidence intervals at 95% (CI95) and 99% (CI99). The average of a given metric is computed as:

$$avg = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{14}$$

where x_i denotes the score obtained for sample i, and N is the total number of samples. The standard deviation, which quantifies the spread of the values around the average, is given by:

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - Avg)^2}, \tag{15}$$

The SEM, which measures the precision of the estimated average, is computed as:

$$SEM = \frac{\sigma}{\sqrt{N}}, \tag{16}$$

Based on the SEM, the 95% and 99% confidence intervals are calculated, respectively, as:

$$CI95 = Avg \pm 1.96 \times SEM, \tag{17}$$

$$CI99 = Avg \pm 2.576 \times SEM, \tag{18}$$

These intervals indicate the range within which the actual average performance of the segmentation metrics is expected to lie with 95% or 99% confidence. This integration ensures a more robust and scientifically meaningful assessment of the model's segmentation

performance, reflecting not only its accuracy but also the consistency and stability of its predictions across the dataset.

4. Results and Discussion

This section assesses the performance of PRA-UNet using both quantitative and qualitative methods. It begins by presenting the training configurations and validation results, including validation on the BraTS 2020 dataset to ensure robustness and generalization. Then, the overall testing results and detailed performance evaluation are discussed. A subsequent comparative analysis highlights the superiority of PRA-UNet over existing state-of-the-art methods. Finally, the deployment prospects and clinical integration potential of PRA-UNet are discussed.

4.1. Training and Test Phase Analysis

This section provides a detailed review of the PRA-UNet setup used to segment 2D brain tumors from MRI data. We created five configurations to systematically evaluate the role of each architectural component and its impact on the performance of the final model. Each configuration omits specific modules from the architecture, allowing their effect on segmentation performance to be measured. The evaluation uses both quantitative results and visual examples to support design choices and clarify performance differences. Figure 7 shows an example of the four configurations.

The configurations are defined as follows:

- Configuration A: This configuration removed inverted residual blocks in the encoding phase. The objective is to examine whether using only bottleneck residual blocks improves segmentation accuracy through enhanced feature extraction.
- Configuration B: This configuration removed bottleneck residual blocks in the encoding phase. It is designed to assess the effect of depthwise separable convolutions, applied via inverted residual blocks, on model performance.
- Configuration C: In this configuration, SAMs are removed from the skip connections. This isolates the effect of spatial attention and allows us to evaluate whether direct encoder–decoder connections suffice for accurate segmentation.
- Configuration D: In this configuration, the CBAM is removed from the bridge, and a direct connection is established between the encoder and decoder. This setup tests the impact of removing channel attention in favor of uninterrupted information.
- PRA-UNet: This is the original configuration. It integrates both bottleneck and inverted residual blocks, includes SAMs in the skip connections, and retains the CBAM in the bridge. This full-featured setup serves as the reference configuration for comparison.

Table 3 presents the performance evaluation of the models during the training phase, using a 5-fold cross-validation methodology. This analysis highlights the strengths and limitations of each configuration in brain tumor segmentation.

The results clearly demonstrate the superior segmentation performance of PRA-UNet compared to configurations A, B, C, and D. Specifically, PRA-UNet achieves a maximum Dice of 98.15%, coupled with a peak IoU of 96.36% and a remarkable accuracy of 99.83% during the second training iteration. Throughout the training iterations, PRA-UNet maintains high metrics consistently, with Dice ranging from 97.09% to 98.15% and IoU values between 94.34% and 96.36%. These results confirm the robustness and effectiveness of PRA-UNet in accurately segmenting medical images, notably outperforming other comparative architectures.
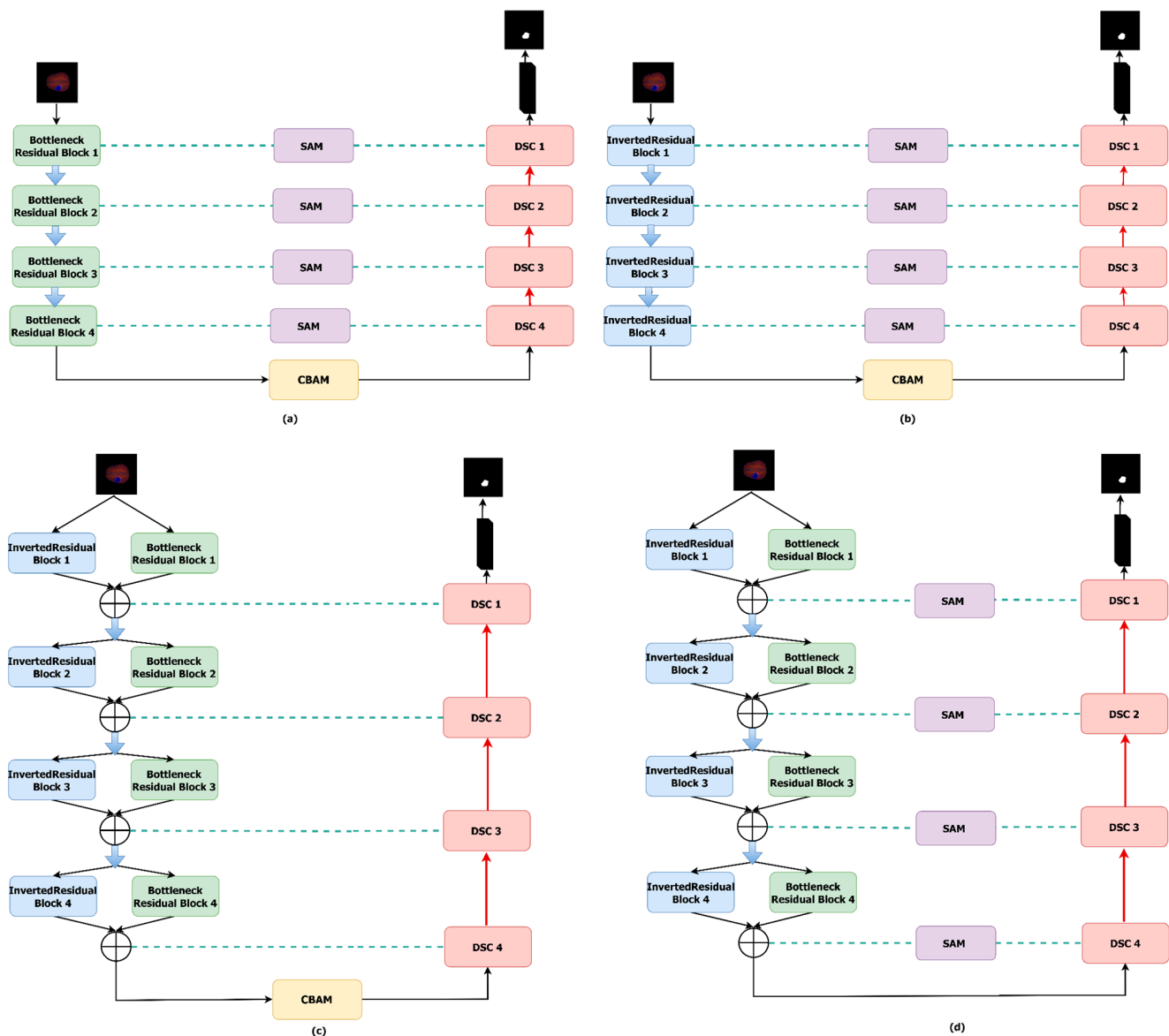
Figure 7. Architectural Variants of the PRA-UNet Model for Evaluating Brain Tumor Segmentation Performance: (**a**) Configuration A, (**b**) Configuration B, (**c**) Configuration C, (**d**) Configuration D.

In contrast, configurations A and B, which exclusively use bottleneck residual blocks or inverted residual blocks, exhibit lower performance, with Dice scores below 95.86%. Configuration C, excluding SAM from skip connections, achieves a modest improvement yet remains inferior compared to configuration D. Configuration D, which lacks the CBAM from its bridge, performs better but does not reach the high Dice and IoU scores achieved by PRA-UNet.

The narrow gap between accuracy and Dice in PRA-UNet indicates a well-calibrated model that distinguishes tumor structures. Accuracy mainly focuses on how many pixels were correctly classified, while Dice is more sensitive to FN, making it a stricter test of missed segmentation areas. The slight difference between the Dice and IoU scores (1.79–2.75%) shows that PRA-UNet is even more accurate. This small gap indicates that the predicted and actual segmentation masks are very similar and have perfect overlap.

PRA-UNet also has much lower loss values during training, between 0.0114 and 0.0179. This shows that it is better at optimizing and converging than other models. This lower loss matches up perfectly with the great DSC and IoU metrics, which show that the model can extract features quickly and segment images accurately. The learning curves in Figure 8

further substantiate these findings, as they exhibit a stable validation loss without any indication of overfitting. This means that PRA-UNet can reliably generalize new data.

Table 3. Comparison of Performance Metrics for the Different Models During the Training Phase with 5-Fold Cross-Validation.

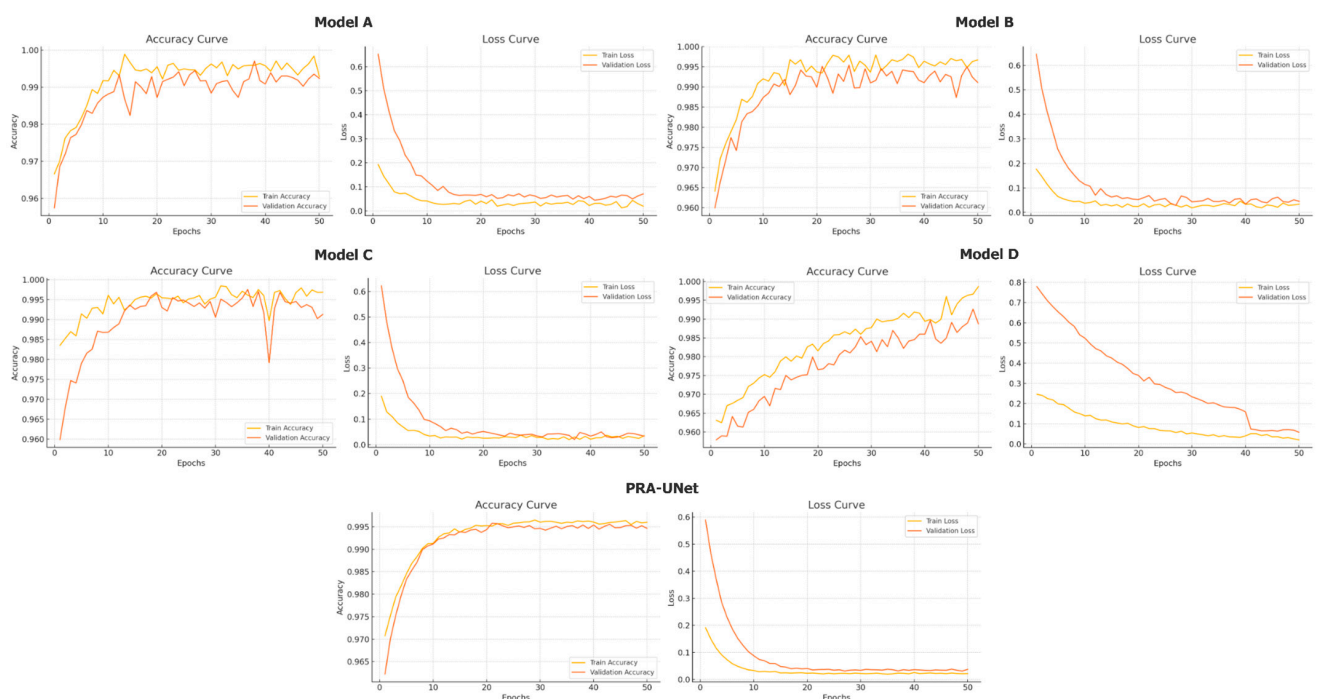| Configuration	Iteration	Accuracy (%)	Dice (%)	IoU (%)	Loss
Configuration A	1	99.62	95.86	92.05	0.0258
	2	99.54	95.15	90.77	0.0304
	3	99.61	95.72	91.79	0.0269
	4	99.44	93.95	88.59	0.0383
	5	99.61	95.77	91.88	0.0264
Configuration B	1	99.62	95.84	92.03	0.0263
	2	99.58	95.32	91.08	0.0290
	3	99.55	94.98	90.46	0.0313
	4	99.61	95.76	91.89	0.0266
	5	99.58	95.31	91.07	0.0292
Configuration C	1	99.62	95.62	91.71	0.0267
	2	99.57	95.30	91.03	0.0292
	3	99.66	96.24	92.75	0.0230
	4	99.61	95.58	91.55	0.0269
	5	99.61	95.83	92.00	0.0258
Configuration D	1	99.74	97.21	94.58	0.0171
	2	99.62	96.02	92.34	0.0246
	3	99.65	96.12	92.54	0.0240
	4	99.65	96.15	92.59	0.0235
	5	99.64	96.04	92.40	0.0240
PRA-UNet	1	99.77	97.52	95.16	0.0152
	2	99.83	98.15	96.36	0.0114
	3	99.76	97.40	94.93	0.0160
	4	99.81	97.90	95.89	0.0129
	5	99.73	97.09	94.34	0.0179

Figure 8. Training and Validation Accuracy and Loss Curves for Different Configurations.

The results in Table 4, which show the average performance metrics for all training iterations, back up these claims. PRA-UNet has the best overall average performance of all the configurations tested. It has an average Dice of 97.61%, an Accuracy of 99.78%, and an IoU of 95.34%, as well as the lowest average loss of 0.0147.

Table 4. Average Performance Metrics for Different Configurations During the Training Phase.

Configuration	Avg Accuracy (%)	Avg Dice (%)	Avg IoU (%)	Avg LOSS
Configuration A	99.56	95.29	91.02	0.0296
Configuration B	99.59	95.44	91.31	0.0285
Configuration C	99.61	95.71	91.81	0.0263
Configuration D	99.66	96.31	92.89	0.0226
PRA-UNet	99.78	97.61	95.34	0.0147

Table 5 looks at how well the configurations can work with unseen data. It displays the average results from the test phase. PRA-UNet shows that it works well most of the time, with an average Dice of 95.71%, an Accuracy of 99.61%, and an IoU of 91.78%. Even though the training metrics show a small drop, this drop is not significant, which shows that PRA-UNet can effectively generalize real-world data. On the other hand, some configurations show much bigger drops, which shows that they are not as good at generalizing. This difference in performance shows how strong and useful PRA-UNet is in the clinic for accurately separating brain tumors.

Table 5. Average Performance Metrics for Different Configurations During the Test Phase.

Configuration	Avg Accuracy (%)	Avg Dice (%)	Avg IoU (%)	Avg LOSS
Configuration A	99.17	91.34	84.14	0.0575
Configuration B	99.30	92.45	86.02	0.0498
Configuration C	99.34	93.15	88.92	0.0378
Configuration D	99.08	90.16	82.19	0.0662
PRA-UNet	99.61	95.71	91.78	0.0272

Table 6 provides a deeper statistical exploration of the segmentation performance of all configurations by incorporating not only the Avg and σ, but also the SEM and CI95 and CI99. These metrics offer a more rigorous evaluation of result stability and the reliability of performance differences between models.

Table 6. Descriptive Statistics and Confidence Intervals (95% and 99%) of Dice and IoU Metrics for Different Configurations During the Test Phase.

Configuration	Metric	Avg (%) $\pm$ σ	SEM	CI95 (Avg(%) $\pm$ Error)	CI99 (Avg(%) $\pm$ Error)
A	Dice	91.34 $\pm$ 6.66	2.98	91.34 $\pm$ 5.84	91.34 $\pm$ 7.68
	IoU	84.14 $\pm$ 1.09	0.488	84.14 $\pm$ 0.956	84.14 $\pm$ 1.257
B	Dice	92.45 $\pm$ 0.30	0.134	92.45 $\pm$ 0.263	92.45 $\pm$ 0.345
	IoU	86.02 $\pm$ 0.51	0.228	86.02 $\pm$ 0.447	86.02 $\pm$ 0.587
C	Dice	93.15 $\pm$ 0.41	0.183	93.15 $\pm$ 0.358	93.15 $\pm$ 0.472
	IoU	88.92 $\pm$ 0.72	0.322	88.92 $\pm$ 0.631	88.92 $\pm$ 0.830
D	Dice	90.16 $\pm$ 0.83	0.371	90.16 $\pm$ 0.728	90.16 $\pm$ 0.956
	IoU	82.19 $\pm$ 1.37	0.612	82.19 $\pm$ 1.199	82.19 $\pm$ 1.578
PRA-UNet	Dice	95.71 $\pm$ 0.14	0.0626	95.71 $\pm$ 0.122	95.71 $\pm$ 0.161
	IoU	91.78 $\pm$ 0.26	0.116	91.78 $\pm$ 0.227	91.78 $\pm$ 0.299

Overall, PRA-UNet shows the most minor variance, the lowest SEM, and the tightest confidence intervals among all configurations. These results indicate superior consistency across the five test runs, confirming the model's robustness and its ability to maintain stable predictions even when exposed to variation in test data.

For the Dice, PRA-UNet achieves an average of 95.71 ± 0.14, with an extremely small SEM (0.0626) and very narrow confidence intervals (CI95 = 95.71 ± 0.122, CI99 = 95.71 ± 0.161). These narrow intervals demonstrate that the differences between test folds are minimal and that the actual Dice performance of PRA-UNet almost certainly lies very close to the reported Avg.

In comparison, the other configurations (A–D) exhibit larger σ and wider CI ranges, indicating greater variability across test folds and thus lower statistical reliability. This variability shows that although some configurations achieve acceptable mean performance, they do not consistently maintain this behavior across different test splits.

A similar pattern appears for the IoU metric. PRA-UNet again displays the best stability, with an Avg IoU of 91.78 ± 0.26, an SEM of 0.116, and very compact confidence intervals (CI95 = 91.78 ± 0.227, CI99 = 91.78 ± 0.299). This statistical precision indicates a high level of repeatability in segmentation predictions, an essential requirement for real clinical deployment. Conversely, the IoU values for configurations A and D exhibit wider confidence intervals, indicating less predictable performance and greater sensitivity to data variation.

Taken together, the metrics in Table 6 confirm that PRA-UNet is not only superior in terms of average performance but also statistically more reliable and stable. Its consistently narrow CI ranges underline its strong generalization capability and reinforce its suitability for medical image segmentation tasks where precision and reliability are crucial.

Beyond quantitative metrics, qualitative assessment remains crucial for evaluating segmentation quality in clinical applications. Figure 9 presents a visual comparison of segmentation results, demonstrating that PRA-UNet achieves more accurate, refined tumor segmentation than other configurations. However, as shown in sample 5, all configurations, including our proposed approach, struggle with the segmentation of small tumors, highlighting an ongoing challenge in brain tumor segmentation.

Although PRA-UNet exhibits exceptional performance in brain tumor segmentation, its real-time efficiency depends on its computational complexity. To assess its feasibility in a clinical setting, several parameters are considered, including the number of parameters, memory size, inference latency, and FLOPS. Here, we compare PRA-UNet with configurations A, B, C, and D, which were previously evaluated for their segmentation performance.

The results in Table 7 show that configuration A is the most lightweight, with only 0.33 million parameters, a memory footprint of 4.18 MB, and the lowest latency (54.10 ms). Despite its high efficiency and suitability for real-time applications, its simple architecture ultimately limits its segmentation performance. Configuration B, slightly more complex with 0.70 million parameters and a memory size of 8.36 MB, shows a significant increase in latency (96.44 ms) without delivering a meaningful improvement in segmentation accuracy, reducing its practical relevance compared to PRA-UNet.

Configurations C and D exhibit increased complexity without proportional gains in segmentation quality. Configuration C, with 1.68 million parameters and a memory footprint of 19.65 MB, achieves moderate latency (67.72 ms) but incurs a high computational cost (9.52 GFLOPs), offering only a modest improvement over previous models. Configuration D shows a similar trend with 1.55 million parameters and 18.15 MB in memory, achieving lower latency (56.97 ms) but still maintaining a high computational cost (9.50 GFLOPs), with no notable performance benefit. In contrast, PRA-UNet, with 1.69 million

parameters and 19.71 MB of memory, effectively balances complexity and accuracy. It achieves a competitive latency of 60 ms and a computational cost of 9.56 GFLOPs, while ensuring robust precision suitable for critical medical applications such as brain tumor segmentation.
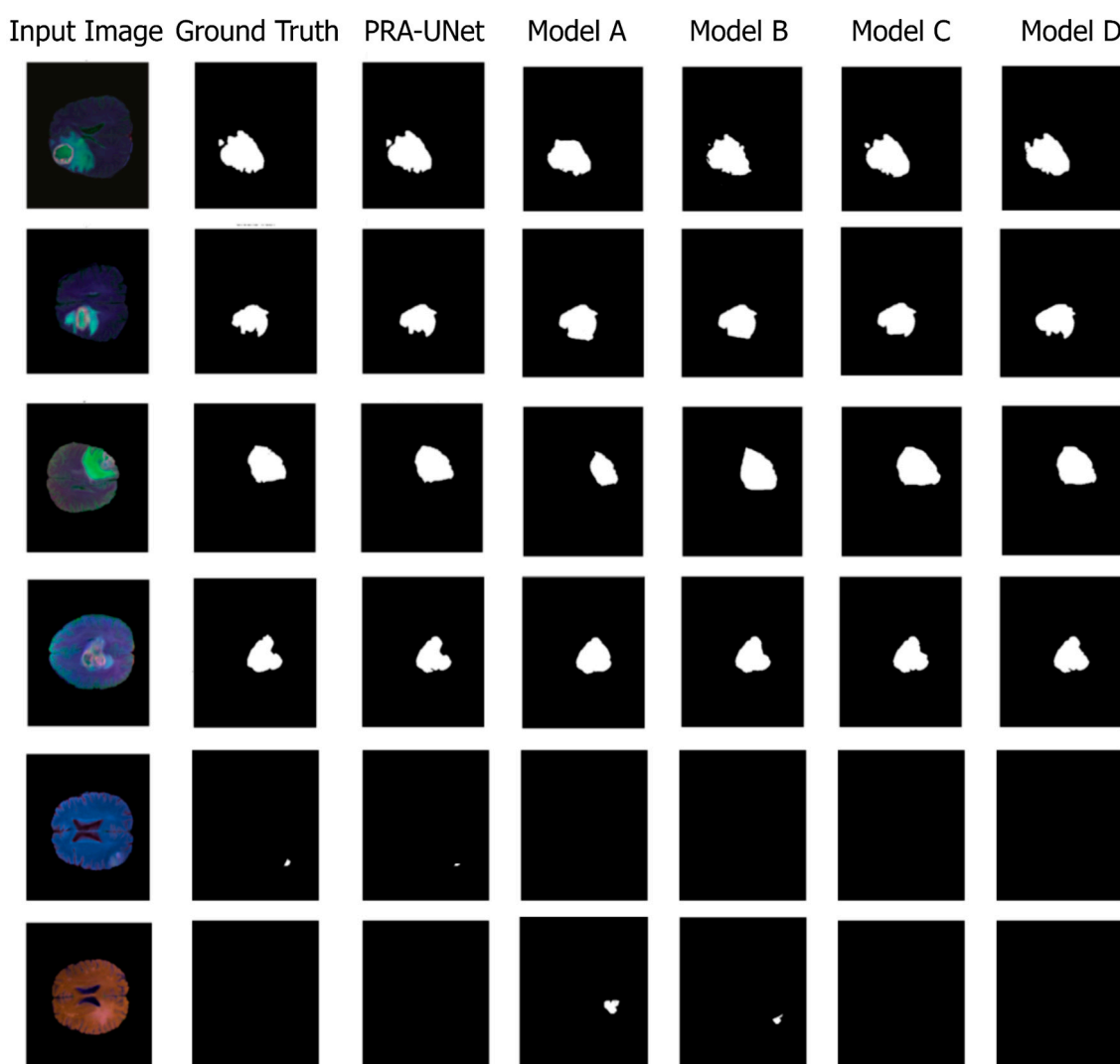

Figure 9. Sample Segmentation Results on Test Data.

Table 7. Comparison of Computational Complexity Across Configurations.

Configuration	Parameters (M)	Size (MB)	Latency (ms)	GFLOPs
Configuration A	0.33	4.18	54.10	2.82
Configuration B	0.70	8.36	96.44	4.7
Configuration C	1.68	19.65	67.72	9.52
Configuration D	1.55	18.15	56.97	9.50
PRA-UNet	1.69	19.71	60.00	9.56

4.2. Ablation Study

To better understand the design choices underlying PRA-UNet, we conducted a series of ablation experiments. These experiments focus on two key aspects of the architecture and training strategy: the depth of the encoder–decoder network and the design of the loss function used to handle the strong foreground–background imbalance. In all cases, only

one factor was modified at a time, while the remaining components and training settings were kept fixed to isolate the contribution of each factor.

4.2.1. Depth Analysis

We first analyze the effect of network depth on the performance and computational cost of PRA-UNet. To this end, we compare three variants with three, four, and five encoder–decoder levels, respectively, while keeping all other architectural components and training settings identical.

The results of this study, summarized in Table 8 using five-fold cross-validation, show that increasing the depth from three to four layers leads to a significant improvement in segmentation scores. Specifically, the Dice increases from 90.4% to 95.71%, representing a gain of 5.31%, while the IOU rises from 83.6% to 91.78%, marking a gain of 8.18%. This notable improvement is accompanied by moderate increases in parameters from 0.98 to 1.69 million, operations from 6.21 to 9.56 GFLOPs, and training time from 17.2 to 19.2 s per epoch. This suggests that adding a layer allows the model to better capture relevant image features, thereby improving segmentation accuracy. In contrast, extending to five layers results in only a marginal gain in Dice (+0.09%) to 95.80% and a slight decrease in IOU (−0.01%) to 91.77%. This evolution requires significantly more resources, with the number of parameters rising to 6.63 million, the number of operations to 12.29 GFLOPs, and the time per epoch to 25.4 s. These results indicate that four layers provide the optimal depth for PRA-UNet, balancing accuracy, complexity, and computational efficiency. This setup maximizes performance at a low cost, which is especially important when resources are limited.

Table 8. Average results of PRA-UNet variants by depth during the testing phase.

Model Depth	Avg Dice (%)	Avg IOU (%)	Parameters (M)	GFLOPs	Training Time (s/Epoch)
3 layers	90.40	83.60	0.98	6.21	17.2
4 layers (PRA-UNet)	95.71	91.78	1.69	9.56	19.2
5 layers	95.80	91.77	6.63	12.29	25.4

4.2.2. Loss Function Ablation

To further evaluate the training behavior of PRA-UNet, we performed an ablation study of the weighting strategy in the composite loss. Five configurations were examined under identical conditions, allowing the impact of each weighting choice to be clearly isolated.

The results in Table 9 reveal distinct performance patterns. The pixel-wise–only configuration ($\alpha = 0$, $\beta = 1$) produces the lowest scores (Dice = 89.96%, IoU = 86.02%), as it favors background pixels and struggles to detect small tumor regions. Increasing the emphasis on the region-overlap component improves the results: the setting $\alpha = 1$, $\beta = 0$ reaches a Dice of 91.68% and an IoU of 88.10%, reflecting enhanced sensitivity to tumor structures, although boundary predictions remain less stable.

A balanced combination ($\alpha = 0.5$, $\beta = 0.5$) offers a more robust compromise, achieving 92.00% Dice and 88.96% IoU by leveraging both region-based and pixel-wise information. In contrast, over-weighting the pixel-wise component ($\alpha = 0.3$, $\beta = 0.7$) slightly degrades performance (90.36% Dice, 86.99% IoU), reintroducing bias toward the background.

The best results are obtained with a region-dominated configuration ($\alpha = 0.7$, $\beta = 0.3$), which achieves the highest Dice (95.71%) and IoU (91.78%). This setting provides the most effective balance for imbalanced medical images, enhancing tumor sensitivity while maintaining stable predictions. Overall, these findings confirm that PRA-UNet benefits

significantly from a loss dominated by the region-overlap component, which is therefore adopted as the optimal configuration.

Table 9. Average results of loss function ablation during the testing phase.

Model	α	β	Avg Dice (%)	Avg IOU (%)
	0	1	89.96	86.02
	1	0	91.68	88.10
PRA-UNet	0.5	0.5	92.00	88.96
	0.3	0.7	90.36	86.99
	0.7	0.3	95.71	91.78

4.2.3. Evaluation on the Raw Test Set

To complement the previous ablation experiments and provide full transparency regarding the influence of test-set augmentation, we also report an evaluation performed exclusively on the raw, non-augmented test images. This analysis quantifies the behavior of PRA-UNet under unmodified acquisition conditions, reflecting the exact data distribution encountered in real-world clinical scenarios.

Table 10 summarizes the performance obtained on the original test set before any augmentation. The model achieves an average Dice of 97.12%, an IoU of 93.64%, and an overall accuracy of 99.71%, indicating a noticeable improvement compared to the augmented-test results. This performance gain shows that PRA-UNet leverages the natural consistency of the raw images, while the synthetic transformations introduced during augmentation make the evaluation significantly more challenging.

Table 10. PRA-UNet performance on the raw and augmented test sets.

Model	Dataset	Avg Accuracy (%)	Avg Dice (%)	Avg IoU (%)	Avg LOSS
PRA-UNet	Raw	99.71	97.12	93.64	0.0199
PRA-UNet	Augmented	99.61	95.71	91.78	0.0272

It is important to note that the raw test set remains relatively small, which limits the statistical robustness of the evaluation. A reduced number of samples increases sensitivity to case-specific variations and may artificially inflate performance metrics. For this reason, relying exclusively on the raw test set could result in an optimistic estimation of the model's generalization capabilities.

To address this limitation, we present both evaluations side by side—the conservative augmented-test results and the optimistic raw-test results—to provide a balanced understanding of the model's behavior. The conservative scenario was retained for comparisons against state-of-the-art methods to ensure fairness and scientific rigor.

4.3. Validation on BraTS 2021

To assess the generalizability of PRA-UNet, we evaluated the model on 340 multimodal MRI images from the BraTS 2021 [44] dataset. The preprocessing steps were consistent with those used for the BraTS 2020 dataset, ensuring comparability. In addition, the same training procedures and computational environment were maintained to ensure that no external variations influenced the performance differences between the datasets. PRA-UNet achieved a Dice of 94.03%, an IoU of 90.53%, a loss of 0.0327, and an accuracy of 99.45% on BraTS 2021, results that are comparable to those obtained on BraTS 2020. The minor decrease in performance is attributed to several factors. BraTS 2021 presents greater tumor heterogeneity, featuring a wider variety of tumor types and more diverse

imaging conditions due to a larger number of cases from multiple institutions and equipment. Additionally, MRI acquisition protocols in BraTS 2021 are more varied, with greater differences in scanners and clinical practices, contributing to the complexity of segmentation. Despite these challenges, PRA-UNet demonstrated robust performance across both datasets, confirming its strong generalizability and potential for clinical application across diverse imaging conditions. The results are presented in Table 11.

Table 11. PRA-UNet Performance on BraTS 2020 and 2021 Datasets During the Test Phase.

Model	Dataset	Avg Accuracy (%)	Avg Dice (%)	Avg IoU (%)	Avg LOSS
PRA-UNet	BraTS 2021	99.45	94.03	90.53	0.0327
PRA-UNet	BraTS 2020	99.61	95.71	91.78	0.0272

4.4. Analysis of PRA-UNet Performance

The results confirm that PRA-UNet provides an optimal configuration for brain tumor segmentation. Several key factors contribute to this superior performance:

1. Hybrid Architecture: PRA-UNet integrates inverted residual blocks and bottleneck residual blocks within the encoder, enabling multi-scale feature extraction. This design enhances the model's ability to capture both fine details and global contextual information, ensuring more precise and robust tumor segmentation.
2. Attention Mechanisms: To improve feature refinement, the model incorporates the CBAM, which enhances spatial and channel-wise feature selection. This allows the network to focus on critical tumor regions while filtering out irrelevant information. Additionally, SAM in skip connections refines feature propagation, reducing background noise and further improving segmentation accuracy.
3. Optimized Loss Function: The segmentation of small and imbalanced tumor regions is enhanced through a combination of Dice Loss and Binary Cross-Entropy Loss. This optimization reduces false negatives, increases reliability, and improves overall detection performance, especially for difficult-to-segment tumor structures.
4. Robustness and Generalization: PRA-UNet maintains an optimal architectural depth that balances feature extraction capacity and computational efficiency, avoiding unnecessary complexity while preserving accuracy. The reduction in model parameters helps minimize overfitting, leading to more reliable performance. Additionally, the five-fold cross-validation strategy validates the model's robustness and confirms its strong generalization ability across different data distributions, ensuring consistent and reliable segmentation in real-world applications.

All these factors demonstrate why PRA-UNet is one of the most effective models for brain tumor segmentation.

4.5. Comparative Analysis

The primary objective of this section is to compare the performance of the PRA-UNet model to that of other open-source architectures that are often used for medical image segmentation. We chose the U-Net, U-Net++, Attention U-Net, ResU-Net, and U-Net++ with Dense models to ensure that the comparison was fair and objective. These architectures are well known in the literature for their efficiency and reproducibility, ensuring that our comparison is transparent and replicable. We used the BraTS 2020 dataset [45] as a reference to train and test all of the models in the same experimental setting. To reduce external variations that might affect the results, we kept the same experimental conditions as in our previous study.

The evaluation is based on two main criteria: segmentation accuracy and computational efficiency. The DSC and IoU are standard metrics in medical image segmentation that are used to measure accuracy. Inference latency, model parameter count, and memory size are among the metrics used to assess computational efficiency. To assess the robustness of the models, we tested each model using 500 test images that were not part of the training set. We split the images into 20 groups of 25 images each to enable batch-wise evaluation with uniform memory usage. The final reported results represent the overall average computed across all test images.

Table 12 shows that PRA-UNet is highly efficient and competitive, making it a strong candidate among existing medical image segmentation models. PRA-UNet differs from many other cutting-edge architectures that focus solely on accuracy. It strikes a good balance between segmentation performance and computational cost, which is vital for real-world deployment.

Table 12. Comparison of PRA-UNet and State-of-the-Art Models During Test Phase (Average Over Five-Fold Cross-Validation).

Model	Avg Dice (%)	Avg IoU (%)	Latency (ms)	Parameters (M)	Size (MB)
U-Net	85.80	75.13	94	7.80	120
U-Net++	90.46	82.58	120	9.20	148
Attention U-Net	93.97	88.62	111	8.50	123.9
ResU-Net	92.00	85.18	102	8.10	123
U-Net++ with Dense	96.32	92.91	235	10.60	150
PRA-UNet (Proposed)	95.71	91.78	60	1.69	19.71

PRA-UNet is the best model in terms of computational usage. It has the lowest latency (60 ms), the fewest parameters (1.69 million), and the smallest memory size (19.71 MB). These features suggest that the model could be usable in systems with limited hardware resources, such as portable devices or diagnostic tools operating on the edge. On the other hand, models like U-Net++ and Attention U-Net need a lot more resources, with more than 8 million parameters and memory sizes of more than 120 MB. UNet++ with Dense has the highest Dice score (96.32%) and IoU (92.91%), but it is also the most computationally intensive model, with 10.6 million parameters, a 150 MB size, and a latency of 235 ms. This makes it less suitable for time-sensitive environments.

PRA-UNet is one of the best models in terms of accuracy, with a Dice score of 95.71% and an IoU of 91.78%. This level of accuracy is kept up without slowing down or expanding the model size, even though it is slightly lower than U-Net++ with Dense. Attention U-Net (93.97% Dice, 88.62% IoU) and ResU-Net (92.00% Dice, 85.18% IoU) also perform well, but they are less precise and less efficient than PRA-UNet. U-Net is lightweight in terms of computation, but it fails to segment effectively (85.8% Dice, 75.13% IoU), making it unsuitable for clinical settings where high accuracy is required.

Finally, U-Net++ (90.46% Dice, 82.58% IoU) is more accurate than the original U-Net, but its higher latency (120 ms) and memory requirements (148 MB) do not offer substantial benefits over PRA-UNet. These observations support the conclusion that PRA-UNet provides a superior balance among accuracy, inference time, and model complexity compared to alternative architectures.

PRA-UNet is the most balanced and ready-to-use model of the ones we evaluated. It is a practical and scalable solution for real-time medical imaging applications because it can keep high segmentation accuracy while lowering latency and resource use.

While Table 12 emphasizes the balance between accuracy and computational cost, Table 13 complements this analysis by examining the statistical behavior of the models

across the five folds. This additional layer of evaluation reveals how consistently each architecture performs, beyond its average Dice and IoU values.

Table 13. Comparison of PRA-UNet and State-of-the-Art Models During Test Phase with Confidence Interval (CI95/CI99).

Model	Metric	Avg(%) $\pm$ σ	SEM	CI95 (Avg(%) $\pm$ Error)	CI99 (Avg(%) $\pm$ Error)
U-Net	Dice	85.80 ± 0.50	0.224	85.80 ± 0.44	85.80 ± 0.58
	IoU	75.13 ± 0.60	0.268	75.13 ± 0.52	75.13 ± 0.69
U-Net++	Dice	90.46 ± 0.40	0.179	90.46 ± 0.35	90.46 ± 0.046
	IoU	82.58 ± 0.50	0.224	82.58 ± 0.44	82.58 ± 0.58
Attention U-Net	Dice	93.97 ± 0.35	0.156	93.97 ± 0.31	93.97 ± 0.40
	IoU	88.62 ± 0.40	0.179	88.62 ± 0.35	88.62 ± 0.46
ResU-Net	Dice	92.00 ± 0.45	0.201	92.00 ± 0.39	92.00 ± 0.52
	IoU	85.18 ± 0.55	0.246	85.18 ± 0.48	85.18 ± 0.63
U-Net++ with Dense	Dice	96.32 ± 0.17	0.076	96.32 ± 0.150	96.32 ± 0.196
	IoU	92.91 ± 0.30	0.134	92.91 ± 0.263	92.91 ± 0.345
PRA-UNet (Proposed)	Dice	95.71 ± 0.14	0.0626	95.71 ± 0.122	95.71 ± 0.161
	IoU	91.78 ± 0.26	0.116	91.78 ± 0.227	91.78 ± 0.299

The reference models, including U-Net, U-Net++, Attention U-Net, and ResU-Net, show confidence intervals that remain relatively wide, indicating noticeable fluctuations from one fold to another. These variations suggest that their segmentation performance, although acceptable on average, is more sensitive to data partitioning effects. Attention U-Net and ResU-Net, for instance, present improved mean scores compared to U-Net, yet their dispersion across folds remains significant, revealing a level of instability that is not visible in the mean figures alone.

Among the architectures, U-Net++ with Dense achieves the highest average accuracy, but its confidence bounds still reflect a measurable degree of variability. This nuance, when combined with its high computational demand reported in Table 11, reduces its practicality in scenarios where both performance and operational constraints must be considered simultaneously.

In contrast, the confidence intervals of PRA-UNet are notably narrower than those of all other models. This pattern indicates that its predictions are more tightly clustered across folds, reflecting a stable and reproducible behavior. These statistical characteristics reinforce the observations from Table 12: PRA-UNet not only maintains competitive segmentation accuracy with low computational cost but also exhibits a more reliable performance profile, making it particularly suitable for deployment in real-world environments where consistency is essential.

The confusion matrices in Figure 10 further support the performance advantages of PRA-UNet. The model records 1,426,130 true positives and 31,214,195 true negatives, while keeping false positives at 71,258 and false negatives at 56,417. This reflects accurate detection with limited misclassification.

Compared to other models, PRA-UNet shows a better balance. U-Net++ with Dense has almost similar FN (56,751) and FP (70,924) but requires far more resources. Attention U-Net, although achieving more TP, produces over 100,000 false negatives and more than 111,000 false positives, which affects its reliability.

U-Net and U-Net++ both show higher error rates, especially in false negatives. U-Net, in particular, reaches 232,232 FN, the highest among all models. These results confirm that PRA-UNet achieves fewer errors in both directions—missing fewer tumors and producing

fewer false detections—while remaining efficient. This supports its use in clinical settings where both precision and speed are essential.
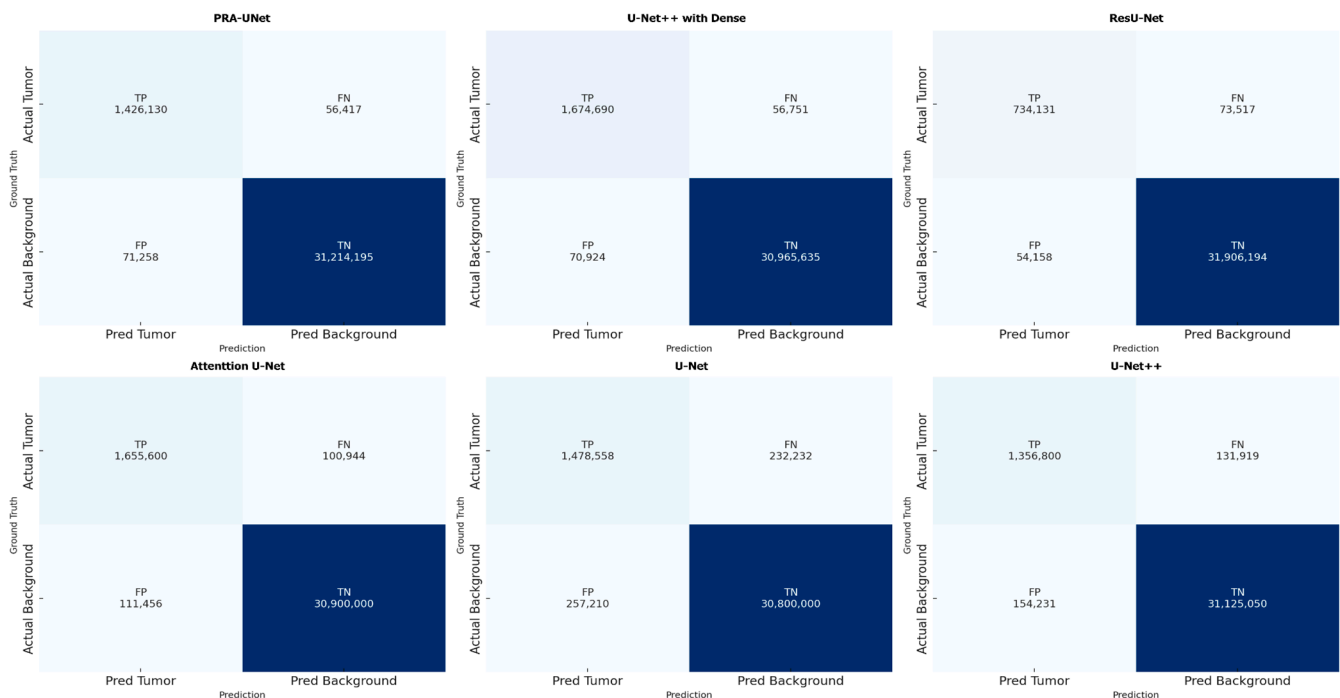

Figure 10. Comparison of Confusion Matrices During the Test Phase (Average Over 5-Fold Cross-Validation) for Different U-Net Variants.

4.6. Deployment Prospects and Clinical Integration

PRA-UNet's architecture is designed to minimize computational complexity while maintaining high segmentation accuracy. This characteristic is particularly relevant for healthcare systems with constrained infrastructure [45], where many hospitals—especially in rural areas—lack advanced imaging hardware, computing resources, and specialized radiologists.

From a clinical workflow perspective, PRA-UNet can be incorporated into DICOM-based imaging infrastructures through an edge/cloud architecture designed to support resource-constrained healthcare systems. Rather than relying exclusively on centralized computing infrastructures, segmentation can be performed directly at peripheral nodes, thereby reducing data transfer requirements and enabling quasi real-time processing even where connectivity is intermittent. Within this distributed workflow, PRA-UNet would function as an assistive segmentation component embedded into the standard DICOM pipeline, automatically generating structured delineations that support radiologists during routine interpretation. Such local preprocessing not only contributes to consistent contouring and reduced variability but also helps prioritize urgent cases by accelerating the triage process in busy or understaffed environments. Additionally, the possibility of deploying the model on affordable hardware opens the door to reliable telemedicine workflows, where clinicians in remote facilities can obtain immediate preliminary analyses before forwarding studies to referral centers.

Clinically, this may translate into reduced diagnostic delays, improved access to imaging-based evaluations, and a tangible reduction in workload for healthcare professionals in resource-limited settings. It could also serve as a decision-support mechanism, enhancing consistency in diagnosis and enabling earlier intervention in time-critical cases.

4.7. Limitations and Future Directions

Although PRA-UNet exhibits strong segmentation accuracy and computational efficiency, it is essential to contextualize these findings within the broader methodological constraints of the current study. This reflective perspective not only situates the reported performance within its scientific boundaries but also highlights structural factors that limit the model's generalizability. First, the inherent reliance on 2D axial slices, while crucial for achieving real-time inference on lightweight hardware, inevitably restricts the capacity to capture complete volumetric continuity and long-range inter-slice dependencies. Such constraints may limit the model's ability to represent tumor morphology when spatial heterogeneity extends across multiple planes. Second, despite its robustness on public benchmark datasets, the model has not yet undergone large-scale validation on locally acquired clinical MRIs, which typically exhibit greater variations in acquisition protocols, scanner characteristics, and noise patterns. This is a critical step toward confirming the model's clinical reliability under real-world conditions. Third, segmentation of extremely small or low-contrast lesions remains challenging, reflecting a broader limitation across many lightweight architectures, where fine-grained feature representation competes with computational constraints. Fourth, although PRA-UNet is explicitly designed for use in resource-limited environments, its actual deployment on real edge hardware has not yet been experimentally validated.

Taken together, these factors naturally motivate several avenues for future research. A promising direction involves the exploration of hybrid 2D–3D or pseudo-3D variants that preserve computational efficiency while enhancing volumetric feature continuity. Further improvements may also arise from integrating adaptive attention mechanisms explicitly designed for micro-lesion detection or from employing self-supervised and domain-adaptation techniques to strengthen robustness across heterogeneous clinical centers. Additionally, evaluating the model directly on embedded clinical devices would provide concrete evidence of its feasibility for deployment in low-resource healthcare infrastructures. Finally, expanding validation to multi-institution cohorts and real-world clinical workflows will be essential to ensure that PRA-UNet transitions from a high-performing research architecture to a reliable clinical decision-support tool.

5. Conclusions

In this study, we presented PRA-UNet, a lightweight and efficient deep learning architecture designed for brain tumor segmentation from multimodal MRI scans. By integrating bottleneck residual blocks, inverted residual blocks, and attention mechanisms, the proposed model achieves a strong balance between segmentation accuracy and computational efficiency. On the BraTS2020 test set, PRA-UNet obtained an average Dice score of 95.71%, an IoU of 91.78%, and an accuracy of 99.61%, with narrow confidence intervals confirming the stability of the predictions across the five test folds. The model also maintains a compact computational profile, requiring only 1.69 million parameters, 19.71 MB of memory, and a latency of 60 ms per image, making it suitable for real-time applications and deployment on resource-constrained systems.

The qualitative evaluation further highlights the model's strengths. Visual analyses show that PRA-UNet generates clear and anatomically coherent tumor boundaries, reduces segmentation noise, and preserves fine structural details, particularly in cases with irregular margins or heterogeneous tissue textures. Confusion matrix analysis corroborates these observations by demonstrating high true-positive and true-negative rates with limited misclassifications.

External validation on the BraTS2021 dataset reinforces the robustness of the proposed architecture. PRA-UNet achieved a Dice score of 94.03%, an IoU of 90.53%, and an accuracy

of 99.45%, indicating stable generalization across variable acquisition protocols and multi-institution data. The consistency observed between the two datasets confirms the model's adaptability and reliability under diverse imaging conditions.

Overall, PRA-UNet provides a practical and resource-efficient solution for automated brain tumor segmentation. Its combination of accurate predictions, qualitative consistency, and low computational cost positions it as a promising tool for integration into clinical workflows, especially in scenarios where rapid inference and limited hardware resources are critical considerations.

Author Contributions: Conceptualization, A.Z.L.; methodology, A.Z.L., M.M. and S.M.; software, A.Z.L.; validation, A.Z.L., M.M. and S.M.; formal analysis, A.Z.L.; investigation, M.M. and S.M.; resources, A.Z.L.; data curation, A.Z.L.; writing—original draft preparation, A.Z.L.; writing—review and editing, A.Z.L., M.M. and S.M.; visualization, A.Z.L.; supervision, M.M. and S.M.; project administration, M.M. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets analyzed in this study are the publicly available BraTS2020 dataset. The data can be accessed at BraTS2020 Dataset: https://www.kaggle.com/datasets/sanglequang/brats2020 (accessed on 21 June 2025), Preprocessed Dataset: https://www.kaggle.com/datasets/alilebanizakaria/2d-multichannel-brats2020 (accessed on 23 June 2025).

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

References

1. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*; Springer: Munich, Germany, 2015; pp. 234–241.
2. Gonzalez, R.C.; Woods, R.E. *Digital Image Processing*, 3rd ed.; Prentice Hall: Upper Saddle River, NJ, USA, 2008.
3. Tiwari, A.; Srivastava, S.; Pant, M. Brain tumor segmentation and classification from magnetic resonance images: Review of selected methods from 2014 to 2019. *Pattern Recognit. Lett.* **2020**, *131*, 244–260. [CrossRef]
4. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]
5. Pereira, S.; Pinto, A.; Alves, V.; Silva, C.A. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1240–1251. [CrossRef] [PubMed]
6. Dorfner, F.J.; Patel, J.B.; Kalpathy-Cramer, J.; Gerstner, E.R.; Bridge, C.P. A review of deep learning for brain tumor analysis in MRI. *NPJ Precis. Oncol.* **2025**, *9*, 2. [CrossRef] [PubMed]
7. Wang, J.; Yu, Z.; Luan, Z.; Ren, J.; Zhao, Y.; Yu, G. Rdau-net: Based on a residual convolutional neural network with DFP and CBAM for brain tumor segmentation. *Front. Oncol.* **2022**, *12*, 805263. [CrossRef]
8. Saeed, M.U.; Ali, G.; Bin, W.; Almotiri, S.H.; AlGhamdi, M.A.; Nagra, A.A.; Masood, K.; Amin, R.U. Rmu-net: A novel residual mobile U-Net model for brain tumor segmentation from MR images. *Electronics* **2021**, *10*, 1962. [CrossRef]
9. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [CrossRef] [PubMed]
10. Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Rueckert, D. The multimodal brain tumor image segmentation benchmark (BraTS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [CrossRef]
11. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. *arXiv* **2018**, arXiv:1802.02611. [CrossRef]
12. Strudel, R.; Garcia, R.; Laptev, I.; Schmid, C. Segmenter: Transformer for semantic segmentation. *arXiv* **2021**, arXiv:2105.05633. [CrossRef]
13. Cheng, B.; Schwing, A.G.; Kirillov, A. Masked-attention mask transformer for universal image segmentation. *arXiv* **2021**, arXiv:2112.01527.
14. Wang, J.; Sun, K.; Cheng, T.; Jiang, B.; Deng, C.; Zhao, Y.; Liu, D.; Mu, Y.; Tan, M.; Wang, X.; et al. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 3349–3364. [CrossRef]
15. Kirillov, A.; Wu, Y.; He, K.; Girshick, R. Pointrend: Image segmentation as rendering. *arXiv* **2020**, arXiv:1912.08193. [CrossRef]

16. Caron, M.; Touvron, H.; Misra, I.; Joulin, A.; Bojanowski, P.; Douze, M.; Cord, M.; Jégou, H. Emerging properties in self-supervised vision transformers. *arXiv* **2021**, arXiv:2104.14294. [CrossRef]

17. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Granada, Spain, 2018; pp. 3–11. Available online: https://arxiv.org/abs/1807.10165 (accessed on 21 June 2025).

18. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; King, T.; Novikov, M.; Khanal, B.; McCague, C.; Castro, E.; Pinto, A.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999. [CrossRef]

19. Diakogiannis, F.I.; Waldner, F.; Caccetta, P.; Wu, C. ResUNet-A: A deep learning framework for semantic segmentation of remotely sensed data. *arXiv* **2019**, arXiv:1904.00592. [CrossRef]

20. Naqvi, N.Z.; Chhikara, M.; Garg, A.; Agrawal, M.; Kumar, P. A modified dense-UNet for pulmonary nodule segmentation. *INFOCOMP J. Comput. Sci.* **2023**, *22*, 2.

21. Alom, M.Z.; Hasan, M.; Yakopcic, C.; Taha, T.M.; Asari, V.K. Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation. *arXiv* **2018**, arXiv:1802.06955.

22. Chen, J.; Lu, Y.; Yu, Q.; Luo, X.; Zhou, B.; Burdick, J.; Hillel, J. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. *arXiv* **2021**, arXiv:2102.04306. [CrossRef]

23. Cao, H.; Wang, Y.; Chen, J.; Jiang, D.; Zhang, X.; Tian, B.; Wang, J.; Qian, C.; Yan, Y. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. *arXiv* **2021**, arXiv:2105.05537.

24. Hatamizadeh, A.; Nath, V.; Yang, D.; Roth, H.R.; Xu, D. UNETR: Transformers for 3D Medical Image Segmentation. *arXiv* **2022**, arXiv:2103.10504.

25. Chahbar, F.; Merati, M.; Mahmoudi, S. MPB-UNet: Multi-Parallel Blocks UNet for MRI Automated Brain Tumor Segmentation. *Electronics* **2024**, *14*, 40. [CrossRef]

26. Zhang, Y.; Han, Y.; Zhang, J. MAU-Net: Mixed attention U-Net for MRI brain tumor segmentation. *Math Biosci. Eng.* **2023**, *20*, 20510–20527. [CrossRef] [PubMed]

27. Vatanpour, M.; Haddadnia, J. TransDoubleU-Net: Dual scale swin transformer with dual level decoder for 3D multimodal brain tumor segmentation. *IEEE Access* **2023**, *11*, 125511–125518. [CrossRef]

28. Naser, M.; Deen, M. Brain tumor segmentation and grading of lower-grade glioma using deep learning in MRI images. *Comput. Biol. Med.* **2020**, *121*, 103758. [CrossRef]

29. Kunjumon, A.; Jacob, C.; Resmi, R. An efficient U-Net based model for low grade glioma segmentation in MRI images. In Proceedings of the 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), Vellore, India, 22–23 February 2024; pp. 1–5.

30. Makwana, C.; Nikunjkumar, N.; Maheshwari, S. Brain tumor segmentation using U-Net++ with dense connection. *Multidiscip. Sci. J.* **2024**, *6*, 2024264. [CrossRef]

31. Talukder, M.A.; Layek, M.A.; Hossain, M.A.; Islam, M.A.; Nur-e Alam, M.; Kazi, M. Acu-net: Attention-based convolutional U-Net model for segmenting brain tumors in fMRI images. *Digit. Health* **2025**, *11*, 20552076251320288. [CrossRef]

32. Wen, P.Y.; van den Bent, M.; Youssef, G.; Cloughesy, T.F.; Ellingson, B.M.; Weller, M.; Galanis, E.; Barboriak, D.P.; de Groot, J.; Gilbert, M.R.; et al. RANO 2.0: Update to the response assessment in neuro-oncology criteria for high-and low-grade gliomas in adults. *J. Clin. Oncol.* **2023**, *41*, 5187–5199. [CrossRef]

33. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

34. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.

35. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. CBAM: Convolutional Block Attention Module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

36. Lin, M.; Chen, Q.; Yan, S. Network in Network. *arXiv* **2013**, arXiv:1312.4400.

37. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet Classification with Deep Convolutional Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS), Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.

38. Chollet, F. Xception: Deep Learning with Depthwise Separable Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

39. Fang, W.; Han, X.-H. Spatial and channel attention modulated network for medical image segmentation. In Proceedings of the Asian Conference on Computer Vision (ACCV) Workshops, Kyoto, Japan, 30 November–4 December 2020.

40. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, B.M.; Farahani, K.; Davatzikos, C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Nat. Sci. Data* **2017**, *4*, 170117. [CrossRef]

41. Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.T.; Berger, C.; Ha, S.M.; Rozycki, M.; et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BraTS Challenge. *arXiv* **2018**, arXiv:1811.02629. [CrossRef]

42. Quang, S.L. Brats2020—Brain Tumor Segmentation Dataset. Available online: https://www.kaggle.com/datasets/sanglequang/brats2020 (accessed on 21 June 2025).

43. Lebanizakaria, A. 2D-MultiChannel-Brats2020. Kaggle, 2021. Available online: https://www.kaggle.com/datasets/alilebanizakaria/2d-multichannel-brats2020 (accessed on 23 June 2025).

44. SyedSajid. Brain Tumor Segmentation (BraTS 2021) Dataset; Kaggle, 2021. Available online: https://www.kaggle.com/datasets/syedsajid/brats2021 (accessed on 19 November 2025).

45. Marey, A.; Ambrozaite, O.; Afifi, A.; Agarwal, R.; Chellappa, R.; Adeleke, S.; Umair, M. A perspective on AI implementation in medical imaging in LMICs: Challenges, priorities, and strategies. *Eur. Radiol.* **2025**, 1–12. Available online: https://link.springer.com/article/10.1007/s00330-025-12031-z (accessed on 15 November 2025).