3D Brain Tumor Segmentation using U-Net Variants on the BraTS 2020 Dataset

Leon Chen
Student, Department of Computer Science
University of Canterbury
lch250@uclive.ac.nz

Prof Richard Green
Professor, Department of Computer Science
University of Canterbury
richard.green@canterbury.ac.nz

Abstract

Brain tumor segmentation is critical for assisting radiologists in diagnosing and planning treatment for brain tumor patients. With the advancement of deep learning methods, automatic segmentation has become more accurate and efficient. This paper explores 3D brain tumor segmentation using the BraTS 2020 dataset, focusing on convolutional neural network (CNN) architectures such as 3D U-Net and its variants, including Attention Residual U-Net and Depthwise Separable U-Net. The proposed methods aim to segment tumor subregions, including enhancing tumor (ET), tumor core (TC), and whole tumor (WT), with high accuracy. We discuss the trade-offs between model complexity and segmentation performance, along with challenges encountered during the implementation process.

1 Introduction

Brain tumors are among the most life-threatening forms of cancer, requiring accurate diagnosis and treatment planning. Magnetic Resonance Imaging (MRI) is the primary imaging modality used for brain tumor diagnosis, providing multi-modal scans such as T1, T1-contrast, T2, and FLAIR. Manual segmentation of brain tumors from MRI images is labor-intensive, time-consuming, and prone to inter-observer variability. Automated brain tumor segmentation can provide more consistent, rapid, and accurate analysis, aiding radiologists in clinical decision-making.

This paper explores the performance of different U-Net variants for 3D brain tumor segmentation using the BraTS 2020 dataset. Specifically, we investigate three architectures: the standard 3D U-Net, Attention Residual U-Net, and a lightweight Depthwise Separable U-Net. Our primary contributions include:

- Implementation and evaluation of three U-Net variants for brain tumor segmentation
- Comparison of model complexity and segmentation performance across architectures
- Analysis of segmentation performance for different tumor subregions (ET, TC, WT)
- Insights into the trade-offs between model size and segmentation accuracy

The remainder of this paper is organized as follows: Section 2 provides background on brain tumor segmentation and deep learning approaches; Section 3 details our methodology; Section 4 presents experimental results; Section 5 discusses findings and limitations; and Section 6 concludes with implications and future directions.

2 Background

2.1 Brain Tumor Segmentation

Brain tumors, particularly gliomas, present significant challenges in clinical oncology due to their heterogeneous morphology and aggressive behavior. Accurate segmentation of tumor subregions—enhancing tumor (ET), tumor core (TC), and whole tumor (WT)—from MRI is crucial for diagnosis, treatment planning, and outcome prediction. Manual segmentation, although common in practice, is time-consuming and subject to inter-observer variability. As such, automated segmentation methods have become a major research focus.

The Brain Tumor Segmentation (BraTS) challenge has provided standardized datasets to promote the development of automated segmentation tools. These datasets contain multi-modal MRI

images (T1, T1c, T2, FLAIR) along with expert-annotated labels for the three subregions [6–10]. While highly valuable, BraTS datasets have limitations such as a lack of variation in imaging protocols and scanner types, which may affect generalizability to real-world clinical settings. An example of the different MRI modalities used in BraTS is shown in Figure 1.

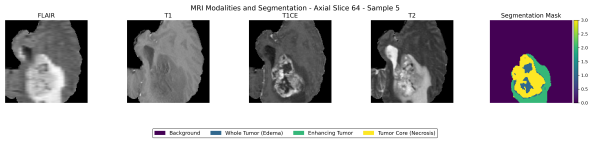

Figure 1: Examples of different MRI modalities provided in the BraTS dataset: T1, T1c, T2, and FLAIR.

2.2 Deep Learning for Medical Image Segmentation

Convolutional neural networks (CNNs) have become the standard for medical image segmentation. The U-Net architecture, introduced by Ronneberger et al. [1], uses an encoder-decoder structure with skip connections, enabling precise localization while maintaining contextual understanding. Çiçek et al. [2] extended this design to 3D U-Nets to accommodate volumetric imaging data like MRI. These 3D U-Nets have shown strong performance in medical applications but are computationally demanding and memory-intensive.

2.3 U-Net Variants and Their Performance

To enhance segmentation accuracy, several variants of U-Net have been proposed. Attention mechanisms have been introduced to help models focus on salient tumor features, while residual connections improve training stability and allow deeper architectures.

For example, Zhang et al. introduced AResU-Net, which combines attention gates and residual units. Their model achieved Dice scores of 0.897 (WT), 0.859 (TC), and 0.838 (ET) on the BraTS dataset [12]. Similarly, Agarwala et al.'s A-UNet model reported a mean Dice score of approximately 0.85 across tumor subregions, demonstrating the efficacy of attention mechanisms in 3D medical imaging [11]. More recently, Butt and Jabbar proposed a hybrid multi-head attention U-Net that achieved Dice scores of 0.902 (WT), 0.868 (TC), and 0.872 (ET), outperforming traditional CNN-based architectures [13]. Gao et al.'s dilated multi-scale residual attention U-Net further improved segmentation accuracy, reporting Dice scores of 0.913 (WT), 0.894 (TC), and 0.869 (ET) [15].

In contrast to the high-performing architectures reported in the academic literature, results from Kaggle implementations on the BraTS 2020 dataset vary significantly depending on model architecture and training strategy. For example, a baseline 2D U-Net implementation on Kaggle achieved Dice scores hovering around 65% [17], highlighting the limitations of 2D models in capturing volumetric context across MRI slices. In comparison, a more advanced solution incorporating a 3D U-Net with a 3D autoencoder reported improved Dice scores of 88.3% for whole tumor (WT), 79% for tumor core (TC), and 75.2% for enhancing tumor (ET) [16]. These findings emphasize the impact of architectural choices and computational resources on segmentation performance.

While these attention and residual-enhanced architectures deliver strong results, they tend to increase model complexity, requiring more parameters and memory—making them less ideal for deployment in resource-limited environments.

2.4 Lightweight U-Nets

To address these limitations, recent work has focused on designing lightweight segmentation networks. One such effort is the use of depthwise separable convolutions, inspired by the MobileNet architecture [5]. These convolutions significantly reduce the number of trainable parameters by factorizing standard convolutions into depthwise and pointwise operations.

Avazov et al. proposed a lightweight U-Net with dynamic focus on tumor boundaries, which employed attention modules alongside depthwise separable convolutions. Remarkably, their model achieved Dice scores above 0.93 for WT while using significantly fewer parameters compared to conventional 3D U-Nets [14]. This indicates that it is possible to maintain high accuracy even with efficient architectures, provided the model design effectively captures key spatial and contextual features.

2.5 Proposed Approach

Motivated by these findings, our work investigates the trade-off between model complexity and segmentation performance by implementing three architectures: a baseline 3D U-Net, an Attention Residual 3D U-Net, and a Depthwise Separable 3D U-Net. While prior studies have often explored either attention or residual enhancements independently, our Attention Residual 3D U-Net integrates both. The Depthwise Separable 3D U-Net, in contrast, serves as a lightweight counterpart designed to test how much model simplification can be tolerated before accuracy significantly drops. This

comparative analysis aims to inform future development of efficient and effective segmentation models suitable for both high-performance systems and resource-constrained environments.

In a practical sense, this would mean that the Attention Residual U-Net is best suited for deployment in scenarios demanding high precision, such as pre-surgical planning, where accurate delineation of tumor boundaries and aggressive subregions is critical for treatment decisions. In contrast, the Depthwise Separable U-Net would be ideal for use in resource-constrained or real-time environments—such as mobile health applications, remote clinics, or edge devices—due to its low parameter count and efficient computation, enabling deployment without the need for powerful hardware.

3 Methodology

3.1 Dataset

We utilized the BraTS 2020 dataset, which contains multi-modal MRI scans (T1, T1ce, T2, and FLAIR) from 369 patients for train data. Validation data was provided but not labeled and hence not used. We further split the training data into training (70%), internal validation (20%), and testing (10%) sets. Each patient's data consists of four MRI modalities, each with dimensions of $240 \times 240 \times 155$ voxels. For each patient, the four modalities are stacked along the channel dimension to form a 4-channel input volume, which is fed into the model. The ground truth segmentation masks include four labels: background (label 0), necrotic and non-enhancing tumor core (label 1), peritumoral edema (label 2), and enhancing tumor (label 4).

3.2 Data Preprocessing

We applied the following preprocessing steps to prepare the data for model training:

- Intensity normalization of each MRI modality to zero mean and unit variance
- Cropping to a region of interest ($128 \times 128 \times 128$) centered around the brain to reduce computational requirements
- Data augmentation techniques including random 90-degree rotations around random axes (50% probability), random flipping along spatial axes (50% probability per axis), random Gaussian noise (mean 0, std dev 0.05, 30% probability), and random intensity shifts (± 0.1, 30% probability) to improve model generalization.

3.3 Model Architectures

We implemented and evaluated three U-Net variants:

3.3.1 3D U-Net

The standard 3D U-Net architecture consists of an encoder and decoder path with skip connections. The encoder path uses repeated applications of $3 \times 3 \times 3$ convolutions, ReLU activations, and max pooling to extract features at different scales. The decoder path uses up-convolutions and concatenations with corresponding encoder features via skip connections, followed by regular convolutions.

3.3.2 Attention Residual U-Net

The Attention Residual U-Net extends the standard U-Net by incorporating both attention gates and residual connections. The attention gates are placed at each skip connection, using the higher-level features from the decoder to emphasize relevant activations from the encoder, effectively suppressing irrelevant regions in the feature maps. Additionally, residual connections are introduced within each encoder and decoder block, allowing for better gradient flow during training and enabling the network to learn more complex features. This dual enhancement helps the model focus on regions of interest while maintaining stable training dynamics, particularly beneficial for the complex task of brain tumor segmentation.

3.3.3 Depthwise Separable U-Net

We proposed a lightweight variant by replacing standard 3D convolutions with depthwise separable convolutions. This architecture factorizes a regular convolution into a depthwise convolution (applying a single filter per input channel) followed by a pointwise convolution ($1 \times 1 \times 1$ convolution) to combine the outputs. This modification significantly reduces the number of parameters while maintaining the overall network structure.

3.4 Training Strategy

All models were trained using the following configuration:

- Loss function: Combination of Dice loss and cross-entropy loss to handle class imbalance
- Optimizer: Adam with an initial learning rate of 1×10^{-4}
- Batch size: 2 due to memory constraints
- Training duration: 80 epochs with early stopping based on validation performance (patience 20 epochs)

- Learning rate scheduler: Reduce on plateau with a factor of 0.5 and patience of 10 epochs

3.5 Evaluation Metrics

We evaluated our models using the following metrics:

- **Dice Similarity Coefficient (DSC):** Measures spatial overlap between predicted and ground truth segmentations
- **Precision:** Ratio of true positive predictions to total positive predictions
- **Recall:** Ratio of true positive predictions to total actual positives

These metrics were calculated for each tumor subregion: ET, TC, and WT, as well as for the background class (no tumor).

4 Results

4.1 System Configuration

The experiments were conducted on a local machine with the following hardware and software specifications:
Device: Laptop
CPU: AMD Ryzen 9 5800 (8C/16T)
RAM: 16 GB
GPU: NVIDIA GeForce RTX 3070
OS: Windows 10 (64-bit)
Python: 3.11
IDE: Visual Studio Code
Framework: PyTorch 2.0.1

4.2 Model Performance Comparison

Table 1 presents the performance metrics for the three implemented U-Net variants across different tumor subregions. The Dice score, precision, and recall were calculated for each tumor component and the background class.

4.3 Analysis of Results

The quantitative results presented in Table 1 are further illustrated by qualitative examples in Figure 2. This figure showcases the segmentation performance of the Attention Residual U-Net on representative slices, comparing the predicted segmentation masks against the ground truth labels.

4.3.1 Performance by Tumor Subregion

All three models achieved excellent segmentation accuracy for the background (no tumor) class, with Dice scores exceeding 0.99, indicating high reliability in identifying healthy brain tissue. The most challenging segmentation task proved to be the enhancing tumor (ET) component, with all models exhibiting their lowest Dice scores in this category (0.5091–0.6900). This is consistent with findings in the literature, as the enhancing tumor regions are often smaller and more heterogeneous in appearance.

Table 1: Performance comparison of U-Net variants on the BraTS 2020 dataset

Metric	Region	U-Net	AttnRes	DepthSep
Dice	No Tumor	0.9957	**0.9973**	0.9964
	WT	0.7796	**0.8705**	0.8006
	TC	0.6986	**0.7595**	0.5945
	ET	0.6458	**0.6900**	0.5091
Prec.	No Tumor	0.9967	0.9966	0.9967
	WT	0.8136	**0.9096**	0.8450
	TC	0.7123	**0.8182**	0.7042
	ET	0.6647	**0.7793**	0.6323
Recall	No Tumor	0.9948	**0.9981**	0.9961
	WT	0.7956	**0.8524**	0.7909
	TC	0.7142	**0.7600**	0.6135
	ET	0.6758	**0.6872**	0.5089
Params		5.65M	5.76M	0.59M
Val. Loss		0.1688	**0.1647**	0.1910
Val. Dice		**0.8221**	0.7734	0.7151

4.3.2 Attention Residual U-Net

The Attention Residual U-Net (AttnRes) achieved the best overall performance, with the highest Dice scores across all tumor subregions (WT: 0.8705, TC: 0.7595, ET: 0.6900). Its attention mechanism improved focus on relevant features, while residual connections enabled deeper learning with only a slight increase in parameters (5.76M vs. 5.65M for standard U-Net). Precision was notably higher, especially for WT (0.9096 vs. 0.8136) and TC (0.8182 vs. 0.7123), indicating fewer false positives—clinically valuable for reducing misclassification.

4.3.3 Standard 3D U-Net

The standard U-Net showed moderate performance (Dice: WT 0.7796, TC 0.6986, ET 0.6458) with 5.65M parameters. Though it had the highest validation Dice score (0.8221), it underperformed compared to AttnRes, suggesting less efficient use of its parameter budget and possible overfitting.

4.3.4 Depthwise Separable U-Net

The DepthSep U-Net had the fewest parameters (0.59M) and the lowest performance overall, but still achieved a competitive Dice score for WT (0.8006). Performance dropped significantly for TC (0.5945) and ET (0.5091), highlighting limitations

in fine detail segmentation due to aggressive parameter reduction.

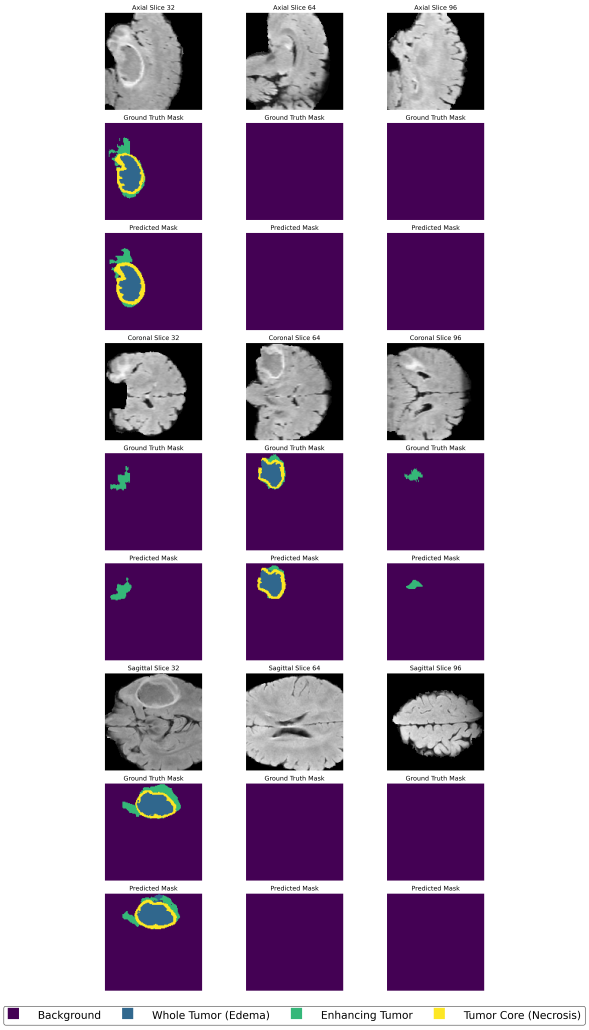

Figure 2: Example segmentation results comparing Ground Truth (top box) with predictions from the Attention Residual U-Net model (bottom box) on different slices from the BraTS 2020 dataset. Different colors represent distinct tumor subregions (ET, TC, WT)

5 Discussion

5.1 Model Complexity vs Performance Trade-off

Our results highlight a clear trade-off between model size and segmentation accuracy. The Attention Residual U-Net achieved the best performance with only a minor increase in parameters over the standard U-Net. In contrast, the Depthwise Separable U-Net reduced parameters by 90% while maintaining decent whole tumor performance.

5.2 Segmentation Challenges

Several patterns emerged across all models regarding segmentation challenges:

- Enhancing tumor regions were consistently the most difficult to segment accurately, likely due to their smaller size and variable enhancement patterns.
- Models generally exhibited higher precision than recall for tumor subregions, indicating a tendency to underestimate tumor extent rather than overestimate it.
- The heterogeneity of brain tumors in terms of size, shape, and appearance likely contributed to variability in segmentation performance.

One notable observation is that all models achieved relatively high Dice scores for whole tumor segmentation compared to tumor core and enhancing tumor, suggesting that distinguishing the overall tumor boundary from healthy brain tissue is easier than differentiating between tumor subregions. This aligns with the hierarchical nature of the segmentation task, where identifying the tumor boundary is a prerequisite for subregion classification.

5.3 Limitations

Our study has several limitations worth noting:

- The evaluation was performed on a subset of the BraTS 2020 dataset, which may not fully represent the diversity of brain tumors encountered in clinical practice.
- While our data augmentation strategies aimed to improve model generalization, domain shift remains a concern when applying these models to data from different institutions or imaging protocols.
- The class imbalance inherent in brain tumor segmentation (where tumor voxels are significantly outnumbered by healthy tissue) continues to pose challenges, despite our use of Dice loss to mitigate this issue.
- The models were trained and evaluated on preprocessed data with fixed dimensions ($128 \times 128 \times 128$), which may not capture the full extent of larger tumors or subtle details in smaller ones.
- Batch size limitation of 1 constrained the stability and generalization ability of training.
- Epoch limitation of 80 due to compute time may have prevented full model convergence.

- The training set (369 cases) is relatively small. Limited labeled data and severe class imbalance (tumor vs. background and among tumor subregions) make models prone to overfitting and poor generalization.
- Limited receptive field: U-Net variants (even with attention gates) can miss very small tumor regions. In practice, some subpopulations (e.g. pediatric tumors or metastases) are underrepresented or absent.
- Attention U-Net and depthwise U-Net architectures, while efficient, may struggle with multi-modal feature fusion and multi-class segmentation. Depthwise separable convolutions process channels independently, which can limit cross-modality learning. Additionally, attention mechanisms designed for binary segmentation may not capture fine-grained distinctions between tumor subregions.

Additionally, at the time training was halted, all models were still showing improvement, indicating that the performance could further enhance.

6 Conclusion

6.1 Summary

This study investigated three U-Net variants for 3D brain tumor segmentation on the BraTS 2020 dataset. The Attention Residual U-Net achieved the best overall performance, particularly for tumor subregion differentiation, while the Depthwise Separable U-Net offered a lightweight alternative with competitive performance for whole tumor segmentation despite having only 10% of the parameters.

Our findings highlight the importance of architectural choices in balancing segmentation accuracy and computational efficiency. The attention mechanism and residual connections proved beneficial for capturing complex tumor features, while depthwise separable convolutions demonstrated their utility in creating lightweight models suitable for resource-constrained environments.

However, the overall segmentation performance, particularly for the tumor subregions, did not reach the high Dice scores reported in recent academic studies using BraTS2020, such as 0.8514 for tumor core (TC) and 0.8175 for enhancing tumor (ET) in [15], or up to 0.9339 for whole tumor (WT) in [14]. Nonetheless, our models achieved performance comparable to several top-performing implementations from the Kaggle community. For instance, while basic 2D U-Net solutions on Kaggle typically yield Dice scores around 65%, more advanced 3D architectures on the platform—such as those incorporating autoencoders—achieve scores of 88.3% (WT), 79% (TC), and 75.2% (ET) [16]. Our results fall within a competitive range of these more sophisticated Kaggle solutions. While promising, these results should be interpreted in light of several limitations.

6.2 Future Research

Several promising directions emerge from this work that warrant further investigation:

- **Loss Function Optimization:** Implement Focal Loss to improve segmentation accuracy for challenging tumor subregions. Explore combinations of losses (e.g., Tversky, boundary loss) for enhanced performance.
- **Multi-scale attention mechanisms:** Incorporating multi-scale or hybrid attention (e.g., combining spatial and channel-wise attention) could improve sensitivity to small tumor structures and subtle boundaries.
- **Advanced Network Architectures:** Integrate transformer-based modules to leverage long-range dependencies while maintaining localization strengths.
- **Model Ensemble Strategies:** Develop sophisticated ensembles to leverage strengths of different architectures, improving overall performance.
- **Multi-scale Feature Fusion:** Explore multi-scale feature fusion to improve segmentation at different tumor scales.
- **Increased Data, Computational Resources, and Training:** Increase training data, computational resources, and extend training schedules to improve model performance and enable more complex models.
- **External Validation:** Test models on independent datasets to assess real-world generalization.

By pursuing these research directions, we aim to further advance the state-of-the-art in brain tumor segmentation, ultimately contributing to improved diagnostic accuracy and treatment planning for brain tumor patients.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[2] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2016, pp. 424–432.

[3] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, Apr. 2018.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[5] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, Apr. 2017.

[6] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[7] S. Bakas *et al.*, "Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features," *Scientific Data*, vol. 4, p. 170117, Jul. 2017.

[8] S. Bakas *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," *arXiv preprint arXiv:1811.02629*, Nov. 2018.

[9] S. Bakas *et al.*, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-GBM collection," *The Cancer Imaging Archive*, 2017. [Online]. Available: `https://doi.org/10.7937/K9/TCIA.2017.KLXWJJ1Q`.

[10] S. Bakas *et al.*, "Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection," *The Cancer Imaging Archive*, 2017. [Online]. Available: `https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF`.

[11] S. Agarwala, S. Sharma, and B. U. Shankar, "A-UNet: Attention 3D UNet architecture for multiclass segmentation of brain tumor," in *Proc. IEEE Region 10 Symposium (TENSYMP)*, Mumbai, India, 2022, pp. 1–5, doi: 10.1109/TENSYMP54529.2022.9864546.

[12] J. Zhang, X. Lv, H. Zhang, and B. Liu, "AResU-Net: Attention residual U-Net for brain tumor segmentation," *Symmetry*, vol. 12, no. 5, p. 721, May 2020.

[13] M. A. Butt and A. U. Jabbar, "Hybrid multihead attentive Unet-3D for brain tumor segmentation," *arXiv preprint arXiv:2405.13304*, May 2024.

[14] K. Avazov, S. Mirzakhalilov, S. Umirzakova, A. Abdusalomov, and Y. I. Cho, "Dynamic focus on tumor boundaries: A lightweight U-Net for MRI brain tumor segmentation," *Bioengineering*, vol. 11, no. 12, p. 1302, Dec. 2024.

[15] Y. Gao, Y. Zhang, and J. Li, "Three-dimensional (3D) dilated multi-scale residual attention U-Net for brain tumor segmentation," *Quantitative Imaging in Medicine and Surgery*, vol. 13, no. 4, pp. 2345–2357, Apr. 2023.

[16] M. Marco Polo, "BraTS 2020 Dataset (Training + Validation)," Kaggle, 2021. [Online]. Available: `https://www.kaggle.com/datasets/awsaf49/brats20-dataset-training-validation`. [Accessed: May 2025].

[17] R. Reza *et al.*, "BraTS 2020 Dataset (Training + Validation)," Kaggle, 2021. [Online]. Available: `https://www.kaggle.com/datasets/awsaf49/brats20-dataset-training-validation`. [Accessed: May 2025].