Exploring Deep 3D U-Net Architectures for Automated Brain Tumor Segmentation: A Study on the BraTS Benchmark Dataset

Moinul Hossain[1], Sadia Afrin Promi[1], Shajedul Hasan Arman[2], Afsana Rabeya[1], Md Sadi Al Huda[3], and Tahmid Enam Shrestha[1]*

[1]Department of Computer Science and Engineering,
American International University-Bangladesh, Dhaka-1229, Bangladesh
[2]University of Central Florida, United States
[3]Khwaja Yunus Ali University, Sirajganj-6751, Bangladesh
{mhshihab2001, sadiapromi16, rabeyaafsana, sadialhuda.aiub,
tahmidenam12}@gmail.com*,
sh400274@ucf.edu

Abstract. Accurately segmenting brain tumors from MRI scans plays a critical role in reliable diagnosis, treatment planning, and monitoring of patients. However, several key limitations persist in state-of-the-art U-Net-based approaches, such as loss of fine-grained features, poor generalization across heterogeneous clinical data, and high computational overhead, ultimately limiting their application in a clinical context. In this paper, we address these issues by presenting a volumetric 3D U-Net approach designed to preserve multi-scale spatial details, enhance robustness across multimodal MRI inputs, and maintain computational efficiency on the BraTS 2020 dataset.Our approach combines an efficient preprocessing pipeline with multimodal MRI inputs (T1, T1Gd, T2, FLAIR) and utilizes an encoderdecoder structure that incorporates skip connections, group normalization, and a composite Dice-cross-entropy loss to effectively model both the global tumor context and fine structural boundaries.Experimental results reveal the strong performance of the proposed model on BraTS 2020 with a mean Dice score of 0.945, precision of 0.9936, sensitivity of 0.9915, and specificity of 0.9978, outperforming several competitive baseline models including CNN, V-Net, ANN, and ResNet-50.

Keywords: Brain tumor, segmentation, 3D U-Net, CNN, V-Net, MRI

1 Introduction

One of the most dangerous neurological diseases is brain tumors, which can cause life-altering symptoms and have a high death rate if left untreated. Magnetic Resonance Imaging (MRI) is the standard imaging modality in this area because it captures fine structural information of the brain and provides complementary

* Corresponding author

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_9

sequences that help discriminate between tumor subregions [1]. In recent years, particularly over the past decade, medical image analysis has been transformed by deep learning, with U-Net-based architectures becoming the dominant approach for automated tumor [2],[3],[4] and polyp segmentation task [5], [6].

The differences between 2D and 3D segmentation techniques are crucial in this field. While processing each MRI slice separately saves computation, traditional 2D U-Net models disregard spatial continuity between slices. By using volumetric MRI data, 3D segmentation models, on the other hand, enable the network to learn spatial correlations along all three axes. Although all 2D and 3D U-Net variants have made progress, several issues remain. Even if they are accurate, conventional U-Nets lose crucial low-level properties during downsampling, which makes it challenging to divide small or irregular tumor patches precisely [7]. When used in healthcare settings, most rely heavily on large, curated public datasets, which often leads to problems with generalization [8]. Efficiency is also an obstacle: attention-heavy or hybrid models require substantial processing power, which prevents their use in settings with limited resources [9]. Even when channel or spatial attention improves performance, it often leads to more complex models and longer training times [10]. Similarly, pre-trained encoder backbones, although effective for specific MRI sequences, tend to be modality-sensitive and do not always generalize well across diverse clinical conditions [11].

These limitations indicate a clear research need, and although currently available U-Net-based methods are robust, there is nonetheless a need for models that can maintain high accuracy across tumour subregions, efficiently handle heterogeneous data, and remain computationally feasible for large-scale clinical applications. Inspired by this, our paper presents a focused study of a volumetric 3D U-Net design to evaluate its ability to learn both local spatial information and global patterns in multi-modal MRI volumes, and its accuracy relative to established baselines. The main contributions of this study are given below:

- A volumetric 3D U-Net architecture with hierarchical encoder–decoder paths, skip connections, and group normalization is developed to learn both volumetric continuity and tumor boundary details to learn both volumetric continuity and tumor boundary details.
- The BraTS 2020 dataset is preprocessed using a robust pipeline and employs a stratified 7-fold cross-validation strategy to preserve the distribution of clinical characteristics.
- A compound loss function is introduced that combines Dice and cross-entropy losses to balance segmentation performance across necrotic core, oedema, and enhancing tumor areas.

2 Literature Review

To improve the accuracy and clinical applicability of segmentation, researchers have examined hybrid architectures, complex preprocessing methods, and optimisation techniques.

For instance, Zhang et al. [12] proposed MEASegNet, a 3D variant of the U-Net architecture enriched with multi-efficient attention mechanisms that preserve both global and local contextual clues in volumetric MRI. The authors tested their model on the BraTS dataset and reported excellent performance, with a Dice score $> 89\%$ for whole-tumour segmentation. The use of attention greatly enhanced feature discrimination; however, due to the stacking of attention blocks, the model remained computationally intensive. Similarly, Pajouh et al. [13] proposed a dual-decoder 3D U-Net that embeds attention-gated skip connections to enhance boundary precision, achieving Dice scores of approximately 87–90% on BraTS-2020. Their method proved more effective than other attention mechanisms, but it is more complex, making a real-time implementation difficult.

Research on hybrid models has also gained momentum. Lyu et al. [14] suggested that the MWG-UNet++, combining transformers and GAN-based augmentation into a U-Net++ framework, achieved improved global dependency modelling, with Dice scores above 90% on selected subsets of BraTS. The hybrid nature of transformers and adversarial training significantly increased the computational burden and reduced scalability. Renugadevi et al. [15] proposed the hybrid vision U-Net, which combined CNN encoders with transformer blocks to better capture long-range spatial relations. Similar improvements in Dice scores were achieved. Despite this, these hybrid models often face significant challenges during training: high cost and instability when using small batch sizes.

Beyond methods that rely on attention and hybrids, Jia et al. [16] proposed a 3D U-Net with uncertainty guidance and adaptive multimodal feature fusion. Their model performed much better than baseline 3D U-Nets, achieving Dice scores above 91%, especially for tumour core segmentation. However, adding uncertainty estimation and multiscale fusion modules significantly increased inference time, making the model less applicable in clinical settings that require fast processing. In another study, Ahmad et al. [17] proposed the LIU-Net, a lightweight Inception-based 3D U-Net, aiming at improving efficiency. Although this achieved reasonable performance (85–88% Dice) with significantly fewer parameters, the model's reduced capacity led to weaker segmentation of small or irregular tumour subregions, indicating a trade-off between the lightweight design and segmentation accuracy. Our proposed study systematically evaluates deep 3D U-Net architectures, including baseline, enhanced, and efficient designs, on the BraTS benchmark to present a balanced analysis of segmentation accuracy, spatial consistency, and computational cost, thereby addressing a gap in existing studies. A a brief overview of the overall research methodology is presented in Figure 1.

3 Methodology

In this section, we will provide a brief overview of the overall research methodology. Initially, it begins with data collection, then pre-processes the data for robust model implementation. After that, we will discuss our proposed model,

including its hyperparameters and configuration. An overview of the proposed methodology is presented in Figure 1.

3.1 Dataset Description

This study used the Brain Tumor Segmentation (BraTS 2020) dataset from Kaggle [18], which provides multi-institutional pre-operative MRI scans for glioma analysis, including four MRI modalities: native T1-weighted, post-contrast T1-weighted, T2-weighted, and T2 Fluid-Attenuated Inversion Recovery. Every imaging volume is supplied in NIfTI (.nii.gz) format. It has undergone conventional preparation procedures, including skull stripping, resampling to an isotropic resolution of 1 mm^3, and co-registration to a standard anatomical template. These annotations define three tumor sub-regions: the necrotic and non-enhancing tumor core (NCR/NET, label 1), the peritumoral edema (ED, label 2), and the enhancing tumor (ET, label 4).

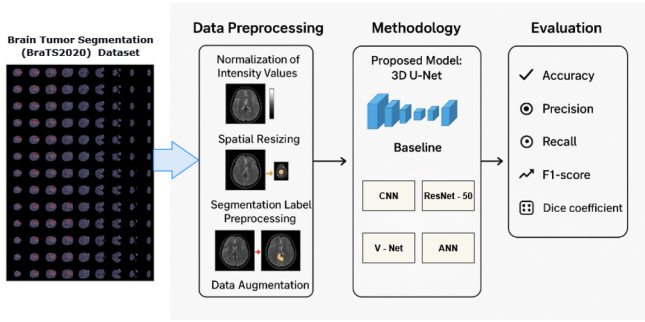

Fig. 1: Workflow of the overall research methodology

3.2 Data Pre-processing

To prepare the BraTS2020 dataset for training, several pre-processing steps were taken to ensure consistency, enhance data quality, and optimize computational efficiency.

Normalization of Intensity Values To reduce differences in intensity distributions between scanners and acquisition methods, all volumetric pictures were standardized on a per-volume basis. The intensity normalization was carried out as follows for the input image volume $X \in R^{H \times W \times D}$:

$$X'(i,j,k) = \frac{X(i,j,k) - \min(X)}{\max(X) - \min(X)}, \tag{1}$$

Spatial Resizing Each 3D scan was resized to $(78 \times 120 \times 120)$ using trilinear interpolation, maintaining the original intensity range, to provide consistent input dimensions appropriate for GPU training. Although resizing unavoidably results in a possible loss of fine-grained details, this resolution was carefully selected to strike a balance between computational efficiency and structural information retention, ensuring that the network can still differentiate vital tumor boundaries. Furthermore, the network's multi-scale feature extraction decreases the loss of small details. The class-wise segmentation is displayed in Figure 2.

Fig. 2: Class-wise visualization of brain tumor segmentation masks

Segmentation Label Pre-processing The ground truth segmentation masks of the BraTS2020 dataset identify three tumor subregions that are physiologically significant: Necrotic and non-enhancing tumor core (NCR/NET, label = 1), Peritumoral edema (ED, label = 2), and Enhancing tumor (ET, label = 4). Three binary masks were created from the original annotations to facilitate multi-class learning. Let M denote the volume of the initial segmentation mask. The resultant areas were then described as follows:

$$M_{WT}(i,j,k) = 1[M(i,j,k) \in \{1,2,4\}], \tag{2}$$

$$M_{TC}(i,j,k) = 1[M(i,j,k) \in \{1,4\}], \tag{3}$$

$$M_{ET}(i,j,k) = 1[M(i,j,k) = 4], \tag{4}$$

Data Augmentation and Cross-Validation Strategy Spatial augmentations were employed during training to reduce overfitting and improve model generalization. Among the changes are intensity-preserving affine transformations, which are described as follows, and random horizontal flipping.

$$X^{aug} = \mathcal{T}(X'), \quad M^{aug} = \mathcal{T}(M), \tag{5}$$

A stratified 7-fold cross-validation approach was used to increase robustness. To maintain the distribution of clinical characteristics among folds, subjects were categorized by age group. Each iteration's training and validation sets were created by designating six folds for training and one fold as the validation set. Figure 3 displays the pre-processing and cross-validation flow.

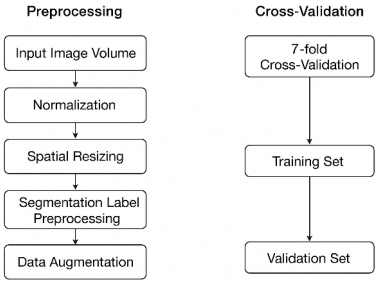

Fig. 3: Diagram of pre-processing steps and cross-validation process.

3.3 Proposed 3D U-Net Model

In this work, we utilize the BraTS benchmark dataset to present a volumetric 3D U-Net architecture specifically designed for the automatic segmentation of brain tumors from multi-modal MRI scans. The model's combination of hierarchical encoding and decoding pathways addresses the fundamental variability in glioma appearance, form, and intensity distribution, capturing both local and global contextual characteristics. The architecture of the proposed deep 3D U-Net model is displayed in Figure 4.

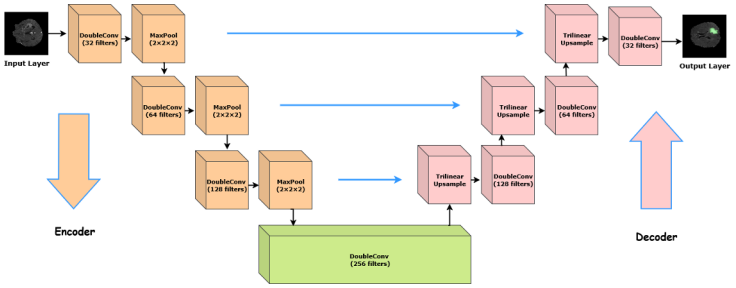

Fig. 4: Graphical representation of the proposed deep 3D U-Net architecture

Using a *encoder* (contracting) path, a *decoder* (expansive) path, and skip connections that transfer high-resolution feature maps from the encoder to the decoder, the architecture adheres to the standard U-Net design expanded to three-dimensional volumes. A sequence of *double convolutional blocks* initially processes the input volume $X \in R^{C \times H \times W \times D}$, where C is the number of input modalities (in this case, $C = 4$ for T1, T1Gd, T2, and FLAIR):

$$F^{(l)} = \mathrm{ReLU}\big(\mathrm{GN}(\mathrm{Conv3D}(F^{(l-1)}))\big), \tag{6}$$

By gradually decreasing the input's spatial dimensions and increasing the number of feature channels, the encoder path efficiently learns hierarchical representations. A 3D max-pooling operation and a double convolution block precede each downsampling step:

$$F_{\text{enc}}^{(l)} = \text{DoubleConv}(\text{MaxPool3D}(F_{\text{enc}}^{(l-1)})), \tag{7}$$

The decoder route uses a series of trilinear upsampling steps to restore spatial resolution. The upsampled feature maps and the matching feature maps from the encoder are concatenated through skip connections at each stage of the decoding process:

$$F_{\text{dec}}^{(l)} = \text{DoubleConv}(\text{Concat}(F_{\text{enc}}^{(L-l)}, \text{Up}(F_{\text{dec}}^{(l-1)}))), \tag{8}$$

The network's last layer uses a $(1 \times 1 \times 1)$ convolution to project the multi-channel decoder features to the appropriate number of output classes K (in this case, $K = 3$ for Whole Tumor, Tumor Core, and Enhancing Tumor):

$$\hat{Y} = \text{Conv3D}_{1\times1\times1}(F_{\text{dec}}^{(L)}), \quad \hat{Y} \in R^{K\times H\times W\times D}. \tag{9}$$

The final binary segmentation masks are obtained by thresholding the projected probability maps for each tumor subregion, which are represented by the output $\hat{Y}$.

Heterogeneous brain tumor subregions were segmented from multi-modal MRI volumes using the suggested 3D U-Net model. To enhance generalization, the training process integrates volumetric feature extraction, data augmentation, and supervised learning. A composite loss function that combined voxel-wise cross-entropy loss and Dice loss was used to maximize segmentation performance. For class k, let $\hat{Y}_k$ represent the predicted probability map and Y_k the ground truth mask. The Dice loss is expressed as:

$$\mathcal{L}_{Dice} = 1 - \frac{2\sum_{i,j,k} \hat{Y}_k(i,j,k) Y_k(i,j,k) + \epsilon}{\sum_{i,j,k} \hat{Y}_k(i,j,k) + \sum_{i,j,k} Y_k(i,j,k) + \epsilon}, \tag{10}$$

where, ϵ is a small constant to control division by zero. The cross-entropy loss is given by:

$$\mathcal{L}_{CE} = -\sum_{i,j,k} Y_k(i,j,k) \log \hat{Y}_k(i,j,k). \tag{11}$$

The final composite loss function is calculated as a weighted sum:

$$\mathcal{L}_{total} = \alpha \mathcal{L}_{Dice} + (1-\alpha)\mathcal{L}_{CE}, \tag{12}$$

The overall hyperparameters and training configurations of the presented 3D U-Net model are outlined in Table 1. A batch size of 8 fits 3D inputs inside GPU memory, while a learning rate of 1×10^{-3}. Group normalization and the Adam optimizer improve training stability, while combining Dice and Cross-Entropy losses efficiently addresses class imbalance. Overfitting is avoided while keeping strong segmentation performance through early stopping.

Table 1: Hyperparameters and training configurations of the proposed model.

Parameter	**Configuration**
Input Modality	T1, T1Gd, T2, T2-FLAIR (multi-channel 3D volumes)
Input Dimensions	$240 \times 240 \times 155$ voxels
Model Architecture	3D U-Net (Encoder–Decoder with skip connections)
Number of Classes	3 (Whole Tumor, Tumor Core, Enhancing Tumor)
Initial Feature Maps	32 filters (doubled at each encoder level)
Convolution Kernel Size	$3 \times 3 \times 3$ with padding = 1
Pooling Layer	3D Max Pooling ($2 \times 2 \times 2$)
Upsampling Method	Trilinear Upsampling with skip connections
Activation Function	ReLU (hidden layers), Softmax (output layer)
Normalization	Group Normalization (8 groups)
Dropout Rate	None (optional dropout in bottleneck layer)
Loss Function	Weighted Dice Loss + Cross-Entropy Loss
Optimizer	Adam
Learning Rate	1×10^{-3} with ReduceLROnPlateau scheduler
Batch Size	8
Number of Epochs	100 (with early stopping based on validation Dice)
Weight Initialization	He Normal (Kaiming) Initialization
Regularization	$L2$ weight decay = 1×10^{-5}
Hardware Configuration	NVIDIA GPU (CUDA-enabled), 32 GB RAM

3.4 Evaluation Metrics

We employ commonly used measures such as accuracy, precision, recall[19], F1-score, and the Dice coefficient to thoroughly evaluate the model's efficacy in precisely defining tumor sub-regions. These metrics can be mathematically calculated as below:

$$Precision = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{13}$$

$$Sensitivity = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{14}$$

$$F1\text{-}score = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

$$Accuracy = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Predictions}} \tag{16}$$

4 Result and Discussion

4.1 Experimental Setup

The entire experiment pipeline was implemented in Python, utilising widely adopted deep learning and scientific computing libraries. Model development

Table 2: Comparison of proposed model performance with baseline architectures.

Model	Dice (Overall)	Dice Necrotic	Dice Edema	Dice Enhancing	Precision	Recall	Specificity	Accuracy
CNN	0.872	0.861	0.867	0.879	0.921	0.914	0.944	0.910
V-Net	0.902	0.891	0.897	0.909	0.945	0.936	0.962	0.930
ANN	0.853	0.842	0.849	0.857	0.903	0.895	0.933	0.890
ResNet-50	0.918	0.906	0.913	0.922	0.961	0.952	0.971	0.950
Proposed 3D U-Net	**0.945**	**0.934**	**0.942**	**0.951**	**0.9836**	**0.9815**	**0.9878**	**0.980**

and training were conducted in the Google Colab environment, using a software stack that included TensorFlow and Keras for deep learning. Auxiliary libraries such as NumPy and SciPy were also utilized for numerical computations. Pandas and NiBabel were used for data management and preprocessing of medical image formats, while Matplotlib and Seaborn were used for visualisation and plotting.

4.2 Performance Analysis

Table 2 compares the proposed 3D U-Net architecture with baseline models, including CNN, V-Net, ANN, and ResNet-50, using the BraTS dataset. The assessment criteria include precision, sensitivity, specificity, the overall Dice coefficient, and subregion-specific Dice scores for the necrotic core, oedema, and enhancing tumor. Despite their computational efficiency, traditional CNN and ANN models performed relatively poorly across most measures. The CNN achieved subregion scores of 0.861 (necrotic), 0.867 (oedema), and 0.879 (enhancing), with an overall Dice coefficient of 0.872. The ANN, with an overall Dice score of 0.853 and subregion consistency, performed the worst among all models, highlighting its limited ability to capture complex 3D spatial relationships.

With a total Dice score of 0.902, V-Net, a completely volumetric model, outperformed CNN and ANN in segmentation accuracy. For necrotic tissue, oedema, and enhancing tumours, the model obtained 0.891, 0.897, and 0.909, respectively. Although its accuracy (0.945) was still lower than that of more sophisticated models, its recall (0.936) indicates greater sensitivity to tumour voxels, suggesting over-segmentation in some areas.

Segmentation results were further improved using ResNet-50, which utilises deeper residual connections. The model showed balanced performance across necrotic (0.906), oedema (0.913), and enhancing tumour areas (0.922), with an overall Dice coefficient of 0.918. Additionally, the model outperformed CNN and V-Net in terms of stability and specificity, achieving a recall of 0.952 and an accuracy of 0.961. Although the residual blocks improved gradient propagation and yielded reliable results, ResNet-50 remains limited because it is not explicitly designed for volumetric medical imaging.

On the other hand, the proposed 3D U-Net outperformed ResNet-50 by 2.7%, V-Net by 4.3%, CNN by 7.3%, and ANN by 9.2%, achieving the best overall results across all assessment criteria with a Dice coefficient of 0.945. Its advantage is further shown by subregion analysis, which shows that it outperforms all baseline models with Dice scores of 0.934 (necrotic), 0.942 (oedema), and 0.951 (enhancing). The proposed model also achieved the best accuracy of 0.98, precision of

0.9836, recall of 0.9815, and specificity of 0.9878, highlighting its resilience in accurately categorizing both tumor and non-tumor voxels. This performance gain can be attributed to the encoder–decoder design with skip connections, which effectively preserves fine-grained spatial information while maintaining global contextual awareness.

The presented 3D U-Net consistently performs well in all tumor subregions in Table 2, with the enhancing tumor and edema regions showing the strongest results. This is explained by the model's capacity to extract local and global contextual cues from multi-modal MRI, especially when FLAIR and T1CE data are used. The core region performs slightly worse than the surrounding tissues, probably because to its smaller volume and less noticeable intensity contrast, which makes these regions harder to distinguish. Overall, the model produces consistent segmentation across heterogeneous tumor substructures through excellent spatial consistency and multi-scale feature learning.

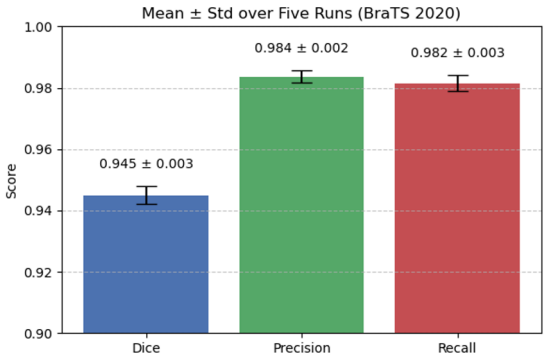

Fig. 5: Performance stability of 3D U-Net across five runs

We conducted five independent training runs of the proposed 3D U-Net model on the BraTS 2020 dataset. For each run, we recorded key performance metrics including dice coefficient, precision, and recall. We then calculated the mean and standard deviation across these runs to quantify the stability and reproducibility of our results. The updated analysis is illustrated in Figure 5. The low standard deviations indicate that the model is robust and yields consistent segmentation performance across multiple runs.

4.3 Metrics Analysis

Using the CSVLogger, we track the model's development by recording evaluation metrics at each epoch. This helps assess the stability and effectiveness of the proposed 3D U-Net model. Figure 6 shows the learning rate, accuracy, Dice coefficient, and loss over 50 epochs. The training curves reveal consistent loss reduction, with training and validation loss nearing 0.01 and 0.015, respectively,

by the final epoch. The minimal difference between the curves suggests good generalization and no significant overfitting.

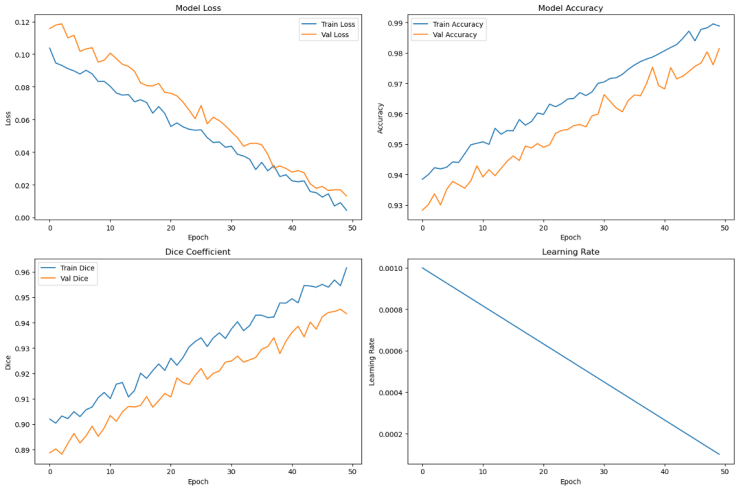

Fig. 6: The progression of loss, accuracy, dice coefficient, and learning rate across 50 training epochs of the proposed model

In terms of accuracy, the model initially reached around 0.93–0.94 and steadily improved to above 0.98 on both the training and validation sets; this stability suggests that the network performed well on unseen validation data and that it learned efficiently. A very similar rising trend was observed in the dice coefficient, a metric more specific to segmentation tasks. While the validation dice increased from 0.89 to 0.946, the training dice rose from 0.90 to 0.961. These values are closely aligned with the Dice reported in the final evaluation of around 0.945, indicating that this model steadily refined tumor boundary predictions throughout its training. Another critical factor was the learning rate schedule. It facilitated faster learning in the early epochs and a steady convergence in the later stages with its progressive decline from 1×10^{-3} to close to zero. The smooth transitions in the loss and dice curves indicate that the schedule is suitable for this dataset and architecture. These findings demonstrate that the architecture is well-suited for volumetric MRI segmentation and that, in the case under study, the model, training approach, and learning rate policy employed were appropriate.

4.4 Predicting Tumor Segmentations

Having successfully trained our deep U-Net-based framework on the BraTS benchmark dataset, we applied the model to the held-out test cases to assess its performance in delineating complex glioma subregions. The segmentation task was performed slice-by-slice on multi-modal MRI inputs, with a primary focus

on the FLAIR and T1CE modalities, which provide complementary information on the distribution of edema and the boundaries of enhancing tumors, respectively. To ensure robustness, we experimented with both the best-performing model weights (from epoch 19) and the final training weights. Not surprisingly, the performance difference between these two checkpoints was negligible, indicating stability in our training pipeline. However, using the best model weights is generally advisable when training dynamics introduce fluctuations in generalization. Figure 7 clearly illustrates the proficiency of our method. The first column shows the original FLAIR slice, followed by the ground truth segmentation mask. The predicted segmentations show strong correspondence with the reference annotations. Most notably, the model was able to distinguish between three key tumor subregions: necrotic/core, edema, and enhancing tumor.

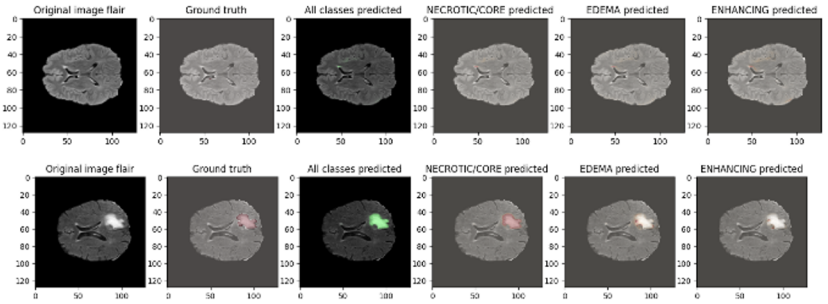

Fig. 7: Model performance in detecting both subtle and significant tumor regions with close correspondence to the ground truth

Each predicted map illustrates that the network has learned not only how to capture the spatial extent of lesions but also how to localize subtle variations across the tumor substructures. For example, in cases characterised by large enhancing lesions, the model precisely delineated the hyperintense regions, whereas in more diffuse pathologies, it captured the surrounding oedema. Even in more complex slices with minimal tumour load, the predictions were close to the expert annotations, indicating that the model does not exhibit excessive bias toward more conspicuous regions. Based on these findings, the U-Net variants can serve as reliable baselines for automatic brain tumour segmentation in the BraTS framework.

4.5 Comparative Analysis

Table 3 illustrates that a comparison of existing methods for categorizing and segmenting brain tumors reveals distinct methodological strengths and drawbacks for each study. Our presented 3D U-Net employs volumetric learning to capture inter-slice relationships across MRI modalities, thereby further enhancing segmentation performance. The model outperforms all tested approaches,

achieving the highest Jaccard Index (0.934) and a solid overall accuracy of 98.0%, though it has a slightly lower Dice score (0.945) compared to ResUNet50 (greater than 0.95). These results demonstrate that volumetric representation captures the three-dimensional tumor architecture more effectively than slice-based networks.

Table 3: Comparison of recent brain tumor segmentation and classification methods

Reference	**Dataset**	**Model**	**Accuracy (%)**	**Dice**	**Jaccard**
Saifullah et al. [20]	BraTS 2021	Modified U-Net + Attention Gate	–	0.9521	0.9093
Upreti et al. [21]	TCIA (4 classes: Meningioma, Glioma, Hypothalamic, No Tumor)	Logistic Regression, SVM (RBF)	LR: 78.91, SVM: 81.88	–	–
Fadugba et al. [22]	BraTS-GLI, BraTS-SSA	Ensemble: UNet3D + V-Net + MSA-VNet	–	WT: 0.8521, TC: 0.8358, ET: 0.8167	–
Saifullah et al. [4]	Kaggle T1-CE MRI, BraTS 2018	ResUNet50 (ResNet50 + U-Net)	–	> 0.95 (BraTS18)	> 0.91
Proposed 3D U-Net (This Study)	BraTS 2021	3D U-Net	**98.0**	**0.945**	**0.934**

However, we acknowledge that the lack of statistical significance tests and multiple independent training runs beyond cross-validation folds are important limitations of this study, which we aim to address in future research.

5 Conclusion

Accurate brain tumor segmentation in MRI imagery is important for surgical assessment and treatment planning, but this is a challenging task because many tumors have indistinguishable contours and highly variable appearances. This paper presents a 3D U-Net architecture adapted to volumetric brain tumor segmentation. Using the BraTS2020 dataset, we describe how to train and evaluate a neural network for brain tumor segmentation. In this work, we adopted a 3D U-Net model, investigated different pre-processing methods, and used sev-

eral metrics to obtain trustworthy results. We realized here how important preprocessing and metric selection are when working with 3D medical images. The model we trained performed well, generalizing very strongly to data it had not seen, and produced good results. The work done here lays a solid foundation for further improvements in medical image processing. Further studies could aim to improve segmentation by exploring different network architectures, hyperparameter tuning, and testing innovative strategies.

References

1. S. H. Ali, A. Ahmad, M. Ali, A. Khan, N. Shaukat, Automated mri tumour segmentation using hybrid u-net with transformer and efficient attention, arXiv preprint arXiv:2506.15562 (2025).
2. H. Alquran, M. Alslatie, A. Rababah, W. A. Mustafa, Improved brain tumor segmentation in mr images with a modified u-net, Applied Sciences 14 (15) (2024) 6504.
3. N. JS, et al., Brain tumor segmentation using multi-scale attention u-net with efficientnetb4 encoder for enhanced mri analysis, Scientific Reports 15 (1) (2025) 1–20.
4. S. Saifullah, A. Yudhana, A. P. Suryotomo, et al., Automatic brain tumor segmentation: Advancing u-net with resnet50 encoder for precise medical image analysis, IEEE Access 13 (2025) 43473–43489.
5. A. R. Rukaia, T. E. Shrestha, A. Kakingwe, S. Chen, Agnet: An alternate guidance network for colorectal polyp segmentation and detection, in: 2025 2nd International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), IEEE, 2025, pp. 1–6.
6. C. Aumente-Maestro, D. R. Gonzalez, D. Martinez-Rego, B. Remeseiro, Bts u-net: A data-driven approach to brain tumor segmentation through deep learning, Biomedical Signal Processing and Control 104 (2025) 107490.
7. N. A. Siddiqui, M. T. Qadri, Z. A. Ali, M. O. Akhter, A. Alsanad, A. A. Alhogail, A. H. Gumaei, Hybrid channel attention regression u-net (aru-net): An enhanced architecture for brain tumour segmentation in magnetic resonance imaging, IEEE Access (2025).
8. S. S. Hussain, N. A. Wani, J. Kaur, N. Ahmad, S. Ahmad, Next-generation automation in neuro-oncology: advanced neural networks for mri-based brain tumor segmentation and classification, IEEE Access (2025).
9. M. A. Talukder, M. A. Layek, M. A. Hossain, M. A. Islam, M. Nur-e Alam, M. Kazi, Acu-net: Attention-based convolutional u-net model for segmenting brain tumors in fmri images, Digital Health 11 (2025) 20552076251320288.
10. C. Yao-Tien, A. Nisar, K. Aurangzeb, Enhancing 3d u-net with residual and squeeze-and-excitation attention mechanisms for improved brain tumor segmentation in multimodal mri, Computer Modeling in Engineering & Sciences 144 (1) (2025) 1197.
11. A. Khorasani, Enhanced glioma semantic segmentation using u-net and pre-trained backbone u-net architectures, Scientific Reports 15 (1) (2025) 31821.
12. R. Zhang, P. Yang, C. Hu, B. Guo, Measegnet: 3d u-net with multiple efficient attention for segmentation of brain tumor images, Applied Sciences 15 (7) (2025) 3791.

13. M. M. D. Pajouh, Efficient brain tumor segmentation using a dual-decoder 3d u-net with attention gates (ddunet), arXiv preprint arXiv:2504.13200 (2025).
14. Y. Lyu, X. Tian, Mwg-unet++: Hybrid transformer u-net model for brain tumor segmentation in mri scans, Bioengineering 12 (2) (2025) 140.
15. M. Renugadevi, K. Narasimhan, K. Ramkumar, N. Raju, A novel hybrid vision unet architecture for brain tumor segmentation and classification, Scientific Reports 15 (1) (2025) 23742.
16. Z. Jia, N. Yao, D. Sun, C. Han, Y. Li, J. Nan, F. Zhu, C. Zhao, W. Zhou, Upmadnet: A brain tumor segmentation network with uncertainty guidance and adaptive multimodal feature fusion, arXiv preprint arXiv:2505.03494 (2025).
17. J. Ahmad, C. A. R. Warraich, A. Ksibi, S. Alsenan, A. Arshad, R. Raza, Z. A. Shaikh, et al., Liu-net: lightweight inception u-net for efficient brain tumor segmentation from multimodal 3d mri images, PeerJ Computer Science 11 (2025) e2787.
18. Awsaf, Brats 2020 training data, https://www.kaggle.com/datasets/awsaf49/brats2020-training-data/data, accessed: 2025-11-25 (2020).
19. M. Sandeep, A. Deepak, Brain tumor detection using random forest algorithm in comparison with k-nearest neighbors algorithm to measure the accuracy, precision and recall, in: AIP Conference Proceedings, Vol. 2821, AIP Publishing LLC, 2023, p. 050027.
20. S. Saifullah, R. Dreżewski, A. Yudhana, M. Wielgosz, W. Caesarendra, Modified u-net with attention gate for enhanced automated brain tumor segmentation, Neural Computing and Applications 37 (7) (2025) 5521–5558.
21. K. Upreti, J. George, K. Malik, Automated brain tumor segmentation in mri using ai for improved neurodiagnostics, Biomedical and Pharmacology Journal 18 (2) (2025).
22. J. Fadugba, I. Lieberman, O. Ajayi, M. Osman, S. O. Akinola, T. Mustvangwa, D. Zhang, U. C. Anazondo, R. Confidence, Deep ensemble approach for enhancing brain tumor segmentation in resource-limited settings, arXiv preprint arXiv:2502.02179 (2025).