Brain Tumor 3D Image Segmentation: Comparative Analysis

Atharv Arya
Department of Computer Science
University of Southern California
Los Angeles, CA, USA
atharvar@usc.edu

Erana Wan
Department of Computer Science
University of Southern California
Los Angeles, CA, USA
eranawan@usc.edu

Harshita Srinivas
Department of Computer Science
University of Southern California
Los Angeles, CA, USA
hs65685@usc.edu

Mudra Limbasia
Department of Computer Science
University of Southern California
Los Angeles, CA, USA
limbasia@usc.edu

Shreyas Malewar
Department of Computer Science
University of Southern California
Los Angeles, CA, USA
malewar@usc.edu

***Abstract*—Accurate brain tumor segmentation in MRI scans is crucial for clinical diagnosis and treatment planning. This study presents a comparative evaluation of four leading deep learning models—3D U-Net, 3D ResNet, nnU-Net, and SegFormer3D on the BraTS 2023 dataset for 3D brain tumor segmentation. Each model represents a distinct architectural paradigm, ranging from convolutional encoders to transformer-based designs. We performed a detailed analysis of their performance using standard metrics such as Dice coefficient, IoU, and F1-score. Our findings highlight notable trade-offs: SegFormer3D achieved the highest F1-score (0.8694) and precision (0.8914), while nnU-Net delivered the best spatial overlap with a Dice score of 0.8859 and IoU of 0.8216. These results underscore the impact of architecture on segmentation quality and computational efficiency. This work provides practical insights for selecting appropriate models in clinical and research settings and lays the groundwork for future advancements in automated neuro-oncological imaging.**

***Index Terms*—segmentation, deep learning, tumor, volumetric**

I. Introduction

Accurate and efficient brain tumor segmentation in MRI scans is critical for early diagnosis, treatment planning, and patient prognosis [1]. Traditional manual segmentation by radiologists is time-intensive, subject to inter-observer variability, and requires specialized expertise, making automated deep learning-based approaches highly desirable [2]. While numerous deep learning models have demonstrated promising results, their relative performance varies depending on factors such as tumor heterogeneity, dataset characteristics, and computational efficiency.

This study aims to present a comparative analysis of four deep learning models, 3D U-Net, 3D ResNet, 3D nnU-Net, and SegFormer3D on brain tumour segmentation using the BRATS dataset. Studies comparing 2D, 2.5D, and 3D approaches have consistently found that 3D models maintain higher accuracy, especially when training data are limited [3]. Thus we have elected to focus on analysis on 3D models which are most clinically relevant for the given task. Additionally we selected models based on popularity and diversity. Exisiting comparative analysis tend to be foucsed on traditonal ML and basic (non-ML) techniques [4] [5]. Recent research in this area has included reviews of leading models, but only include similar architectures, such as only 3D U-Net-based models [6]. In order to cover a more representative set of model architectures, we have included both 3D ResNet and 3D U-net as our baseline models, as well as U-net based nnU-Net and transformer based SegFormer3D.

II. Literature Review

A. 3D U-Net

Automated medical image segmentation is a crucial step in computer-aided diagnosis and treatment planning. However, obtaining fully annotated 3D datasets is labor-intensive and costly. To address this challenge, Ö. Çiçek et. al. proposed the 3D U-Net, an extension of the widely used 2D U-Net architecture, designed for volumetric medical image segmentation from sparsely annotated data [7]. 3D U-Net extends the original U-Net model to three-dimensional data, incorporating a fully convolutional encoder-decoder architecture with skip connections to capture both local and global contextual information. The encoder extracts hierarchical features through successive convolution and downsampling operations, while the decoder restores spatial resolution through upsampling layers. The skip connections facilitate precise localization by combining feature maps from corresponding encoding and decoding layers. To handle sparse annotations, the model employs a training strategy that samples small patches from volumetric data, enabling efficient learning from partially labeled slices. The loss function is adapted to focus on annotated regions, mitigating the limitations of incomplete ground truth. The network is trained using stochastic gradient descent with data augmentation techniques to improve generalizability.

The 3D U-Net demonstrated superior segmentation performance on biomedical imaging datasets, particularly in scenarios with limited labeled data. The authors evaluated the model on a brain tumor segmentation task and showed its effectiveness in reconstructing complete 3D segmentations from sparsely annotated slices. The model significantly outperformed traditional methods and earlier deep learning approaches. Its ability to leverage sparse annotations makes it a practical choice for applications where manual labeling is resource-intensive. The 3D U-Net model represents a significant advancement in volumetric medical image segmentation. By leveraging 3D convolutions and an encoder-decoder structure with skip connections, the model efficiently reconstructs dense segmentations from sparse labels. Its impact on medical imaging research continues to grow, with ongoing improvements and adaptations extending its applicability to various clinical tasks.

B. 3D ResNet

The 3D ResNet architecture, introduced by Hara et al. in 2018 [8], extends the popular 2D ResNet model by using 3D convolutions to capture spatiotemporal and volumetric features, making it particularly effective for video analysis and medical image interpretation. Unlike traditional 2D CNNs, 3D ResNet processes data across three dimensions, allowing it to learn both spatial structure and temporal or depth-related context, which is critical in applications like brain tumor segmentation from MRI scans or action recognition in video frames [8]. Building upon the residual learning framework proposed by He et al. (2015) [9], 3D ResNet facilitates the training of deep models by using identity shortcuts to address the vanishing gradient problem. Its adoption in medical imaging has been further advanced through transfer learning approaches, such as in the Med3D framework [11], where 3D ResNet backbones pretrained on multiple 3D datasets demonstrated strong generalization across various tasks including segmentation and classification. While the model is computationally intensive, it remains a cornerstone for 3D vision tasks due to its strong performance and modular design, and continues to serve as a baseline for newer architectures in volumetric learning.

C. nnU-Net

The paper "nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation" by Fabian Isensee et al. presents a significant advancement in the field of medical image segmentation. The nnU-Net framework addresses the challenge of diversity in medical imaging datasets by offering a self-configuring method that automatically adapts to dataset properties without manual intervention [12]. Unlike traditional approaches that require custom pipeline designs for specific datasets, nnU-Net analyzes training cases and automatically configures preprocessing, network architecture, training, and post-processing parameters, addressing the inherent diversity in biomedical imaging datasets that vary drastically in modality, image sizes, voxel spacings, and class ratios [12]. This adaptability has enabled nnU-Net to surpass many specialized solutions across 23 public datasets used in international biomedical segmentation competitions, establishing it as a standardized baseline, an out-of-the-box segmentation tool, and a framework for developing new segmentation methods [12].

Recent comparative studies have consistently demonstrated nnU-Net's superior performance in various medical imaging applications, particularly in tumor segmentation tasks. In a recent study on pediatric brain tumor segmentation, nnU-Net significantly outperformed the DeepMedic model with higher mean Dice scores across tumor subregions and better generalization capabilities on external validation datasets [13]. In 2024, the original authors behind nnU-Net published another study comparing nnU-Net to various other segmentation methods, including CNN-based, Transformer-based, and Mamba-based approaches. They concluded that CNN-based U-Net models within the nnU-Net framework, when scaled to modern hardware resources, continue to deliver state-of-the-art performance despite the emergence of newer architectures [14]. We aim to expand on the existing literature in our study and conduct an unbiased analysis of top models across a variety of segmentation approaches.

D. SegFormer-3D

SegFormer3D, an extension of SegFormer, enhances 3D medical image segmentation by incorporating volumetric feature learning. This model introduces 3D patch embedding to capture volumetric contextual information efficiently, hierarchical feature extraction in 3D to enable multi-scale feature learning tailored for 3D medical images, and a lightweight 3D decoder to aggregate multi-level feature representations for high-quality segmentations. Additionally, SegFormer3D is optimized for GPU memory efficiency, making it feasible for large-scale medical datasets like BRATS. Since brain tumor segmentation requires precise localization and delineation of tumor subregions, SegFormer3D's ability to efficiently process volumetric data and capture contextual dependencies enhances segmentation performance, making it a promising choice for BRATS dataset applications.

While SegFormer3D offers significant advantages, challenges remain in adapting it for medical image segmentation. Processing high-resolution 3D volumes requires extensive GPU resources, posing memory constraints. Additionally, Transformer models typically require large datasets, and medical imaging datasets like BRATS are limited in size. Another challenge is the robustness and generalization of SegFormer3D, as it needs extensive fine-tuning to generalize across different MRI modalities and patient variations.

SegFormer3D represents a state-of-the-art approach in 3D medical image segmentation, showing strong potential for brain tumor segmentation tasks. Future research should focus on optimizing the model for better generalization, addressing memory constraints, and leveraging data augmentation techniques to improve segmentation accuracy in medical applications.

III. Exploratory Data Analysis

The BraTS 2023 dataset is a comprehensive collection of brain imaging data designed for the evaluation of tumor segmentation and glioma analysis. It includes MRI scans of 1010 patients with high-grade gliomas and low-grade gliomas, with each subject imaged using four different MRI modalities: T1, T1ce (contrast-enhanced T1), T2, and FLAIR. The dataset features 3D volumes with varying image resolutions and intensities, offering diverse and clinically relevant data for research and development in medical image processing and glioma segmentation.

This dataset is specifically designed to aid in the training and evaluation of algorithms for automated tumor detection, segmentation, and analysis in the context of gliomas, which are among the most aggressive forms of brain tumors.

A. MRI Modalities Description

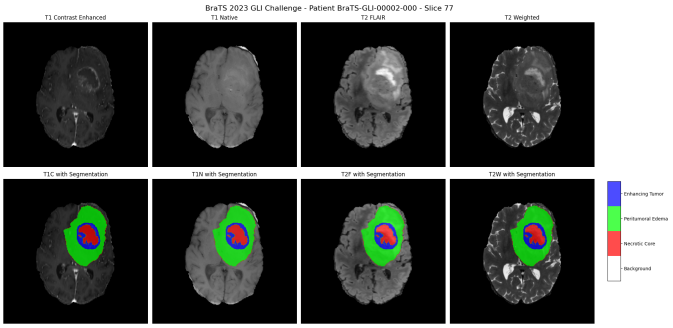

Fig. 1. Tumor Segmentation Classes (Patient 2)

- **T1 (T1-weighted Imaging):** T1-weighted images are commonly used in MRI to highlight anatomical structures and provide high contrast between different tissue types, such as white matter, gray matter, and cerebrospinal fluid. T1 images are useful for assessing brain structure and detecting abnormalities such as tumors or areas of brain atrophy. They provide high spatial resolution and clear differentiation of tissue structures.

- **T1ce (T1-weighted Contrast-Enhanced Imaging):** This modality is similar to T1-weighted imaging but includes the use of a contrast agent (typically gadolinium). The contrast agent enhances the visibility of abnormal tissue, particularly tumors, by highlighting regions with abnormal blood-brain barrier permeability. T1ce images are critical for detecting and delineating tumors and lesions with active blood supply or disrupted blood-brain barrier.

- **T2 (T2-weighted Imaging):** T2-weighted images are used to detect changes in water content within tissues, making them sensitive to fluid-filled structures like cysts or edematous regions. In brain imaging, T2 images are highly effective for identifying lesions, inflammation, and swelling. They provide excellent contrast between gray matter, white matter, and cerebrospinal fluid.

- **FLAIR (Fluid-Attenuated Inversion Recovery Imaging):** FLAIR is a modified T2-weighted image where the cerebrospinal fluid (CSF) signal is suppressed. This suppression allows better visualization of lesions and abnormalities within the brain, especially in the periventricular areas. FLAIR images are particularly useful for detecting white matter changes, such as those seen in multiple sclerosis or ischemic changes, as well as in the identification of tumor margins.

B. Intensity Statistics

TABLE I
Intensity statistics for each MRI modality.

Modality	Min	Max	Mean	Std
T1	0	9234	309.77	374.60
T1ce	0	19713	472.78	319.68
T2	0	3048	99.40	50.69
FLAIR	0	4857	181.94	125.38

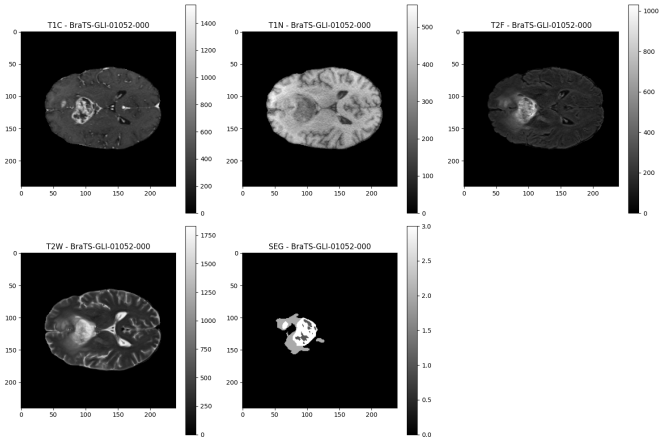

Fig. 2. Sample Visualization for Patient 1052

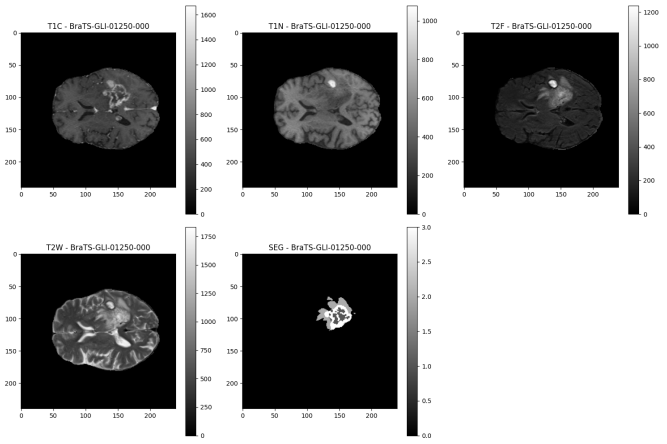

Fig. 3. Sample Visualization for Patient 1250

- **Min:** The minimum intensity value observed in the imaging modality, which reflects the lowest signal detected.

This often corresponds to regions with little or no tissue or structure, such as background or empty spaces in the scan.

- **Max:** The maximum intensity value recorded in the modality, typically indicating areas with the most intense signal. These areas often correspond to dense tissues, such as tumors or abnormal growths, which absorb more signal than surrounding tissue.

- **Mean:** The average intensity value across all pixels in the image for that modality, providing a general sense of the signal strength in the scan. In medical imaging, it reflects the overall tissue characteristics and can indicate tissue health or abnormalities.

- **Std (Standard Deviation):** This measures the variability of intensity values within the image. A higher standard deviation indicates significant variation in tissue types (e.g., contrast between tumors and normal brain matter), while a lower standard deviation suggests more uniform tissue characteristics.

IV. Methodology

Our comparative analysis consists of four main steps: data preprocessing, model training, model evaluation, and final analysis. Data preprocessing is standardized across all models - MRI scans were normalized, cropped to regions of interest, and stacked across four modalities to form multi-channel 3D volumes suitable for training. Model training as well as hyperparameter tuning are done individually for each model. After completion of training, model performances are assessed using precision, recall, F1-scores, and dice scores. Finally, we compare models on overall performance as well as computational efficiency.

A. Preprocessing

Prior to training the models, we applied standardized preprocessing steps to all MRI volumes. We used 1010 of the 1251 BraTS patients with ground truth masks as our training data and the 241 remaining patients as our test data. Preprocessing was applied to both categories.

First we sorted the raw data files by patient and use the nibabel library to extract the data from each .nii file as a numpy array. We then applied intensity normalization (min-max scaling) per modality to standardize voxel intensity distributions. Next we cropped the scans to focus on the region of interest (brain area), reducing computational load. Each modality is cropped from its original 240x240x155 dimensions into a 128x128x128 cube by cropping along indices $[56 : 184, 56 : 184, 13 : 141]$. This also satisfied the cubic input shape requirements for 3D nnU-Net and SegFormer3D. Finally all 4 modalities are stacked into a single image as 4 channels and the resulting 4x128x128x128 arrays are saved as .npy files.

B. Training

All models were trained using the preprocessed BraTS dataset divided into training and validation pytorch datasets on a 80/20 random split. Pytorch dataloaders were also used for batching and multiprocessing. Hyperparameters were independently selected and tuned for each model during training to accommodate the differences in model architecture and allow to best model performance.

Each model employed the optimizer from the original paper proposing its architecture, and most used the original criterion. For architectures which did not originally employ a Dice criterion, we used the closest Dice-Loss variant in order to optimize parameters and handle class imbalances. Batch sizes and learning rates were tuned using grid search independently for each model. Below are the final hyperparameters for each model after tuning.

TABLE II
Final Model Hyperparameters

Model	**3D U-Net**	**3D ResNet**	**3D nnU-Net**	**Segformer3D**
Optimizer	AdamW	Adam	SGD	Adam
Criterion	Dice+CE	Dice+CE	Dice+CE	Dice
LR	3e-4[a]	2e-4	1e-2[a]	1e-3
Batch Size	4	1	2	10
Patch Size	N/A	N/A	128x160x112	N/A
Epochs	100	100	800	800
Patience	10	10	100	100
Best Epoch	95	8	729	221

[a]With variable learning rate.

Training was performed on GPU-accelerated environments using the PyTorch framework. Early stopping and model checkpointing were employed to prevent overfitting.

C. Evaluation and Comparative Analysis

After training, all models were evaluated on the same test set using the standardized metrics. The test set comprised 241 of the 2451 patients in the original BraTS training dataset, set aside before training. For each architecture, the model with the lowest validation loss was loaded from the earlier saved state dict. Evaluation was run with batch size 2. Model performance was assessed using:

- Precision
- Recall
- F1-score
- Dice Score
- Intersection over Union (IoU)
- Computational efficiency (training/inference time, GPU memory usage)

2D visualizations of model predictions were also created during this process, using the middle-slice by depth. Segmentation errors were analyzed using visualization of segmentation masks and ground-truth masks next to each of the MRI modality slices.

Results were compared to identify strengths and limitations of each architecture in terms of segmentation accuracy, robustness, and computational efficiency

V. Results

TABLE III
PERFORMANCE METRICS OF DIFFERENT 3D SEGMENTATION MODELS

Model	Precision	Recall	F1-score	Dice	IoU
Segformer3D	**0.8914**	**0.8492**	**0.8694**	0.8109	0.7276
3D ResNet	0.8164	0.7438	0.7587	0.7659	0.6709
3D U-Net	0.8002	0.8344	0.8582	0.7880	0.7933
3D nnU-Net	0.7914	0.8012	0.8477	**0.8859**	**0.8216**

A. 3D U-Net

3D U-Net performed competitively across all metrics, showing a well-balanced trade-off between precision and recall, which translated into a solid F1-score (0.8582) and Dice (0.7880). This model follows a classic *encoder-decoder architecture with skip connections*, which effectively preserves spatial information across different resolutions. A variable learning rate scheduler was employed to prevent the learning from getting stuck in plateaus, which was an occurrence during training, which adaptively adjusted the learning rate to achieve optimal performance. However, the lack of adaptivity (compared to nnU-Net) and global context modeling (as in transformers) might have limited its ability to match the top performers in edge cases or highly variable anatomical structures.

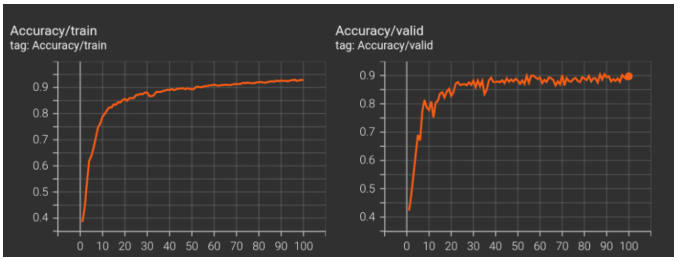

Fig. 4. Train and Validation Accuracy Curve for U-net

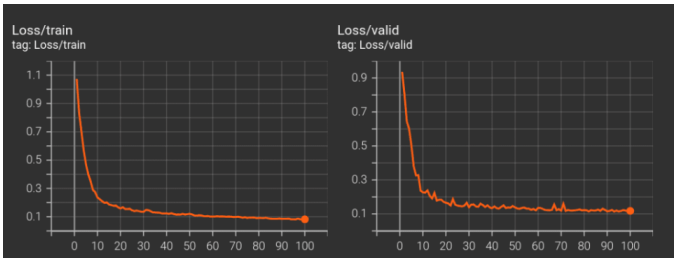

Fig. 5. Train and Validation Loss Curve for U-net

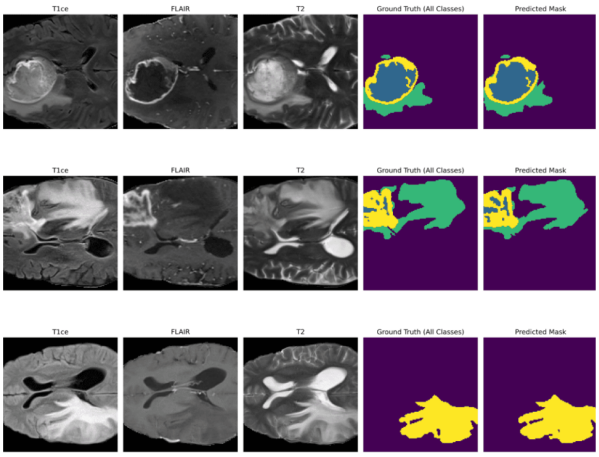

Fig. 6. 3D U-net Results

B. 3D ResNet

3D ResNet underperformed relative to the other models across all metrics. This architecture is primarily designed for *feature extraction rather than segmentation*, using residual connections to allow deep networks to train effectively. However, without the U-Net style decoder and multi-scale aggregation, ResNet struggles with fine-grained localization, as evidenced by its lower Dice (0.7659) and IoU (0.6709). Additionally, it lacks mechanisms for spatial recovery of lost resolution, leading to less precise segmentation boundaries.

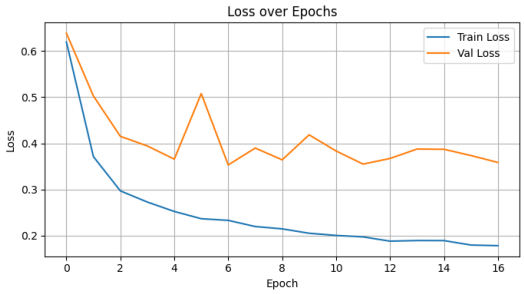

Fig. 7. Train and Validation Loss Curve for 3D ResNet

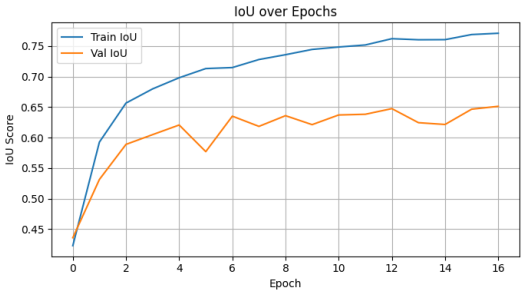

Fig. 8. Train and Validation IoU Curve for 3D ResNet

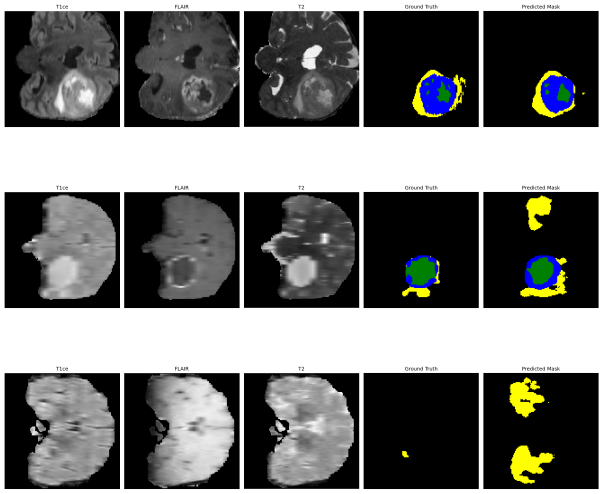

Fig. 9. 3D ResNet Results

C. 3D nnU-Net

While Segformer3D excelled in classification-style metrics, 3D nnU-Net delivered the best overlap-based performance with a Dice coefficient of 0.8859 and IoU of 0.8216, indicating excellent spatial agreement between predictions and ground truth.

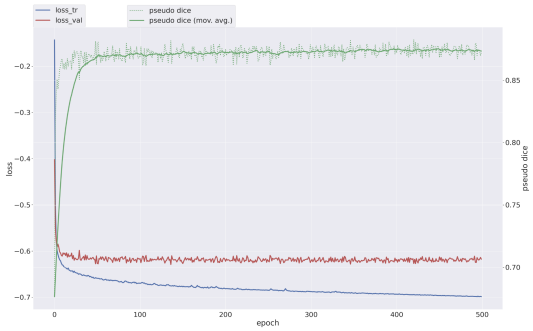

Fig. 10. Train & Validation Loss Curves for nnUnet

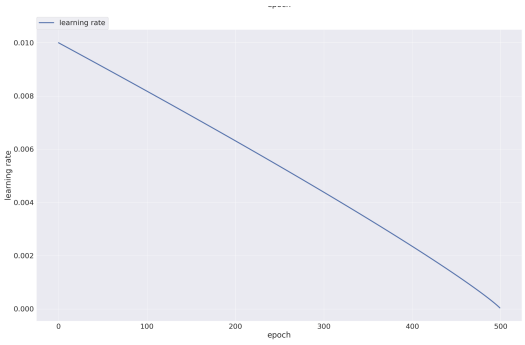

Fig. 11. Adaptive Learning rate Curve for nnUnet

This is expected given that nnU-Net dynamically adapts its architecture (e.g., depth, feature map size, and patch size) based on dataset characteristics. It also includes advanced pre- and post-processing techniques like region-based crop selection and test-time augmentation. These architectural and procedural enhancements likely contributed to the high-quality segmentation performance, especially in complex volumetric regions.

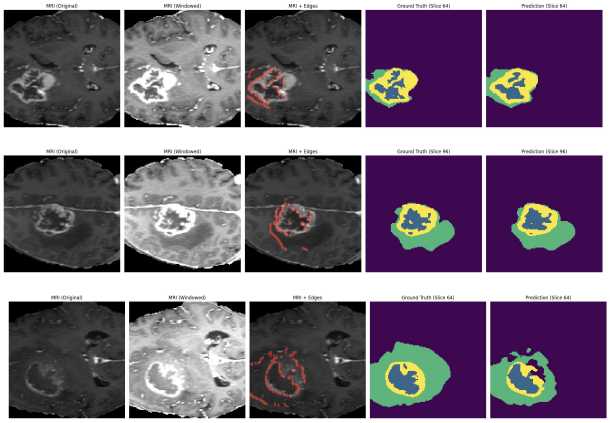

Fig. 12. nnUNet Results

D. SegFormer 3D

Segformer3D achieved the highest precision (0.8914), recall (0.8492), and F1-score (0.8694) among all models. This superior performance can be attributed to its *transformer-based architecture*, which enables effective long-range spatial dependency modeling. Unlike CNNs that rely on local receptive fields, transformers can capture global context more naturally, helping the model distinguish subtle differences in tissue boundaries and reduce both false positives and false negatives. Its relatively strong Dice (0.8109) and IoU (0.7276) scores support its effectiveness, though the slightly lower overlap metrics compared to nnU-Net suggest room for improvement in spatial alignment.

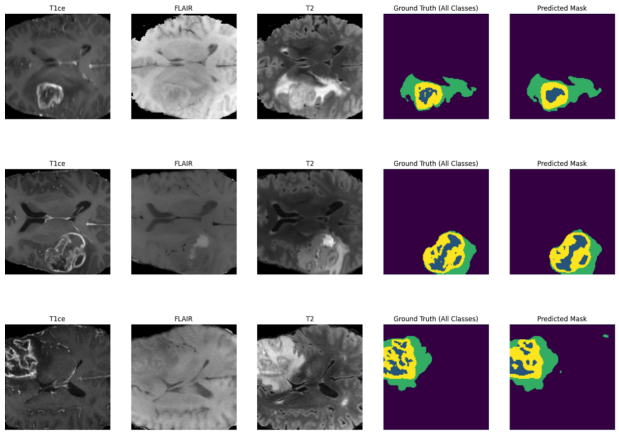

Fig. 13. Segformer3D Results

E. Visual Comparison

A visual comparison of the model predictions for patient 401 (14) aligns with our measurements of model performance. Of our four models, 3D nnU-Net and Segformer3D out peform baseline models 3D U-Net and 3D ResNet, capturing more of the intricacies of the different tumor regions. This is especially evident for the core necrotic regions (blue color), which tends to have irregular sharp angles and smaller disconnected patches.

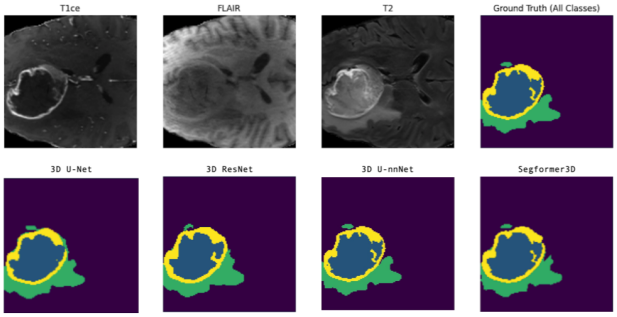

Fig. 14. Model Predictions for Patient 00401-000

VI. Training Efficiency and Resource Utilization

TABLE IV
Training configuration of different 3D segmentation models

Config	Segformer	ResNet	U-Net	nnU-Net
GPU	A40	A100	A40 (x2)	RTX 3090
Batch Size	**10**	1	4	2
#Workers	**4**	2	2	2
Time/Epoch	**41.76s**	41.76s	140s	60s

In addition to segmentation accuracy, practical deployment of deep learning models also requires consideration of computational efficiency, memory consumption, and hardware resource usage. In this context, the training and inference statistics for the evaluated models reveal notable differences tied closely to architectural design and implementation characteristics.

Segformer3D emerged as the most computationally efficient model, completing an epoch in just **41.76 seconds** on a single Nvidia A40 GPU using a batch size of 10 and 4 CPU workers. Despite its transformer-based complexity, Segformer3D exhibited excellent parallelism and GPU utilization, aided by modern attention mechanisms that scale efficiently with batch size and allow effective memory management. Its total GPU memory consumption was modest at **19.41 GB**, which is notable given its performance advantages. Furthermore, Segformer3D demonstrated rapid inference capabilities, processing 241 validation samples in just **9 minutes and 6 seconds**, making it well-suited for real-time or near-real-time clinical applications.

The 3D nnU-Net, while achieving the best overlap-based performance in segmentation (highest Dice and IoU), was comparatively slower and more resource-intensive. Trained on a single Nvidia RTX 3090 GPU with a batch size of 2 and 2 data loader workers, it required **60 seconds per epoch** and consumed up to **21.8 GB** of GPU memory. This is expected, as nnU-Net dynamically adapts its architecture based on dataset properties, often resulting in deeper or wider networks and aggressive patch-based training, which increases memory overhead. These trade-offs reflect its focus on robustness and generalization at the cost of training efficiency.

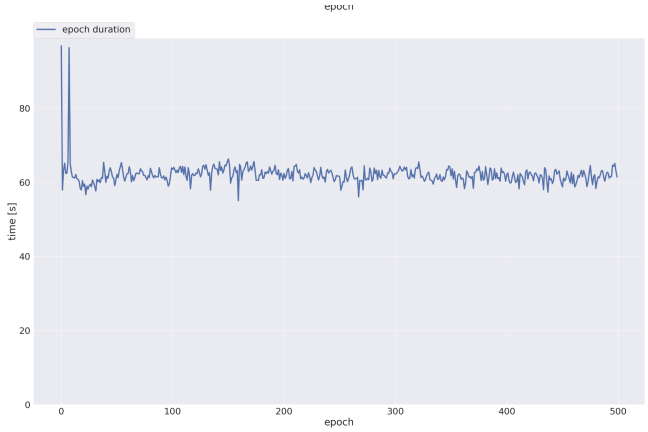

Fig. 15. Epoch Duration Curve for nnUnet

The 3D U-Net model struck a balance between performance and computational cost. Using two Nvidia A40 GPUs, a batch size of 4, and 2 workers, it achieved a per-epoch time of **140 seconds**, slightly better than nnU-Net, while consuming a similar amount of memory (**21.8 GB**). This can be attributed to its simpler encoder-decoder structure with skip connections, which, although less flexible than nnU-Net, remains computationally stable and memory-aware across different dataset scales.

In contrast, 3D ResNet was the least resource-efficient among the models evaluated. Despite being run on a powerful Nvidia A100 GPU, it only utilized a batch size of 1 and 2 workers, yielding the same epoch time as Segformer3D (**41.76 seconds**) but offering significantly lower segmentation performance. The model consumed **19.41 GB** of memory, which is disproportionate given its shallow inference utility. This inefficiency stems from its lack of a decoding pathway or multi-scale feature aggregation, causing less optimal use of both memory and compute resources in segmentation contexts.

Overall, Segformer3D provided the best trade-off between speed, memory consumption, and segmentation performance. While 3D nnU-Net remains the gold standard in overlap metrics, its computational footprint may be limiting for resource-constrained environments. These results suggest that transformer-based architectures like Segformer3D could offer a compelling alternative, especially when fast deployment and scalable training are prioritized.

VII. Limitations

This section aims to look into the limitations of the current study in terms of the dataset, computation resources and experiments conducted.

- **Dataset Bias and Limited Diversity**: The current study relies on the BraTS dataset, which is well-curated but limited in diversity in terms of scanner types, demographics, institutions. Thus, the study may not generalize to real-world clinical data with variable quality, noise, or pathology heterogeneity and may thus require external validation on out-of-distribution datasets for e.g., from a different hospital or country.

- **Limited Compute resources**: Given the data intensive nature of the task (3-D data), alongside large models, with millions of parameters, the task demands a large amount of compute resources. In this study, we had access to Carc, which was still limited in our scope. With increasingly long training times, limited the amount of experiments (for e.g., hyperparameter tuning, ablation study, architecture search) we could perform. We still, however, managed to tune the hyperparameters based on the suggestions made in the original papers in which the models were proposed.

- **Lack of Evaluation Metrics**: In our study, we utilize the Dice coefficient to evaluate the segmentation performance of the models and Hausdorff Distance for boundary accuracy. To further enhance the study, exploring additional metrics to measure the sensitivity to tumor core or enhancing tumor in the segmentation task and studying the volume correlation of the tumor regions, may lead to a more comprehensive study.

- **Limited Use of Advanced Training Techniques**: The current study does not utilize advanced training techniques like data augmentation, mixed precision training, or multi-GPU parallelism that were infeasible or inefficient on constrained hardware. Further, other training techniques such as fine-tuning large pre-trained models (like Vision Transformers or CLIP) were unexplored due to training limitations.

VIII. Future Work

The experimental results across the four architectures—SegFormer 3D, 3D ResNet, 3D U-Net, and nnU-Net—highlight the strengths and limitations of each model. Building on this, we outline several directions for future research to enhance performance, reliability, and clinical applicability:

- **Extension to Multimodal and Multiregional Datasets**: Future work will involve applying the trained models to a broader spectrum of datasets, including CT scans and anatomical regions beyond the brain (e.g., lungs, liver). This will help assess the models' generalizability to diverse medical imaging domains and promote cross-domain transferability.

- **Interpretability and Clinical Relevance Analysis**: To foster trust in clinical deployment, future efforts will focus on integrating explainability techniques such as Grad-CAM, attention rollouts, and SHAP for 3D segmentation. These visualizations will be used to verify whether the model's predictions align with radiological reasoning.

- **Uncertainty Estimation for Reliable Deployment**: To improve clinical safety, future work will incorporate uncertainty estimation using techniques like Monte Carlo dropout and test-time augmentation. Voxel-wise uncertainty maps will help flag low-confidence predictions, especially near tumor boundaries, enabling better error detection and supporting manual review in critical cases.

- **Error Analysis and Robustness Testing**: Detailed post-hoc error analysis using attention heatmaps and pixel-level uncertainty overlays will be performed. This will aid in identifying common failure patterns (e.g., tumor boundary ambiguity) and developing correction strategies through architectural adjustments or data augmentation.

- **Low-Resource Model Optimization for Real-Time Use**: Given the computational demands of 3D models, optimizing architectures for low-latency inference on edge devices (e.g., in operating rooms or remote settings) will be crucial. Techniques such as knowledge distillation, pruning, and quantization will be explored for deployment-ready efficiency.

- **Active Learning Pipeline for Continuous Improvement**: A future iteration of this work will include a human-in-the-loop setup where uncertain or low-confidence predictions trigger manual annotations. This can feed into an active learning loop, improving the model iteratively with minimal labeling effort.

IX. Conclusion

In this study, we conducted a comprehensive comparative analysis of four state-of-the-art deep learning models—3D U-Net, 3D ResNet, nnU-Net, and SegFormer3D—for brain tumor segmentation using the BraTS 2023 dataset. Our results demonstrate that each model has distinct advantages depending on the evaluation criteria. SegFormer3D achieved the highest precision and demonstrated efficient training times, suggesting promise in transformer-based volumetric segmentation. nnU-Net, on the other hand, delivered the highest Dice and IoU scores, validating its robustness and effectiveness in medical segmentation tasks even without manual configuration. 3D ResNet exhibited strong generalization capabilities with balanced performance, while 3D U-Net showed competitive segmentation accuracy and was computationally lightweight.

However, the study also reveals several limitations—namely, reliance on a single curated dataset (BraTS), restricted experimentation due to hardware constraints, and limited use of advanced training techniques such as data augmentation and uncertainty estimation. Additionally, certain clinical nuances such as tumor subregion-specific accuracy and model interpretability remain underexplored in this phase. Despite these constraints, our findings provide valuable insights into the trade-offs between model accuracy, efficiency, and deployment feasibility. This work lays the foundation for future improvements, particularly in enhancing clinical readiness through uncertainty quantification, robustness evaluation, interpretability techniques, and low-resource optimization strategies. Ultimately, the insights gained from this study aim to inform the selection and refinement of automated segmentation models for real-world neuro-oncological applications.

REFERENCES

[1] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., "The multimodal brain tumor image segmentation benchmark (BRATS)," IEEE Trans. Med. Imaging, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[2] U. Baid, S. Ghodasara, S. Mohan, M. Bilello, E. Calabrese, E. Colak, et al., "The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification," arXiv preprint arXiv:2107.02314, 2021.

[3] A. Avesta, S. Hossain, M. Lin, M. Aboian, H. M. Krumholz, and S. Aneja, "Comparing 3D, 2.5D, and 2D approaches to brain image auto-segmentation," Bioengineering, vol. 10, no. 2, p. 181, Feb. 2023.

[4] S. S. Varshney, N. Rajpal, and R. Purwar, "Comparative study of image segmentation techniques and object matching using segmentation," in Proc. Int. Conf. Methods Models Comput. Sci. (ICM2CS), New Delhi, India, 2009, pp. 1–6, doi: 10.1109/ICM2CS.2009.5397985.

[5] W. Wang, A. N. Kravchenko, A. J. M. Smucker, and M. L. Rivers, "Comparison of image segmentation methods in simulated 2D and 3D microtomographic images of soil aggregates," Geoderma, vol. 162, no. 3–4, pp. 231–241, 2011.

[6] M. Saleh, M. Salih, M. Ahmed, and A. Hussein, "From traditional methods to 3D U-Net: A comprehensive review of brain tumour segmentation techniques," J. Biomed. Sci. Eng., vol. 18, pp. 1–32, 2025.

[7] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, Proc., Part II, vol. 19, pp. 424–432.

[8] K. Hara, H. Kataoka, and Y. Satoh, "Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet?" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2018.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016.

[10] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), 2015.

[11] S. Chen, K. Ma, and Y. Zheng, "Med3D: Transfer learning for 3D medical image analysis," arXiv preprint arXiv:1904.00625, 2019.

[12] F. Isensee, P. F. Jaeger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, "nnU-Net: A self-adapting framework for U-Net-based medical image segmentation," Nat. Methods, vol. 18, no. 2, pp. 203–211, 2021.

[13] A. Vossough, N. Khalili, A. M. Familiar, D. Gandhi, K. Viswanathan, W. Tu, D. Haldar, S. Bagheri, H. Anderson, S. Haldar, P. B. Storm, A. Resnick, J. B. Ware, A. Nabavizadeh, and A. Fathi Kazerooni, "Training and comparison of nnU-Net and DeepMedic methods for autosegmentation of pediatric brain tumors," AJNR Am. J. Neuroradiol., vol. 45, no. 8, pp. 1081–1089, 2024.

[14] F. Isensee, T. Wald, C. Ulrich, M. Baumgartner, S. Roy, K. Maier-Hein, and P. F. Jäger, "NnU-Net revisited: A call for rigorous validation in 3D medical image segmentation," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2024: 27th Int. Conf., pp. 488–498.

[15] S. Perera, P. Navard, and A. Yilmaz, "SegFormer3D: An efficient transformer for 3D medical image segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW), pp. 4981–4988, 2024.