IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, March 2026
Impact Factor: 8.2

Automated 3D Brain Tumor Segmentation Using a Cascaded 2.5D–3D Deep Learning Framework on BraTS 2020

Mr. Parth Prashant Jadhav1, Mr. Ayush Arjun Kadam[2], Mr. Sumit Tatyasaheb Yele[3]
Mr. Mohan Kashinath Mali[4]
Students, Department of Computer Technology[1-3]
Guide, Department of Computer Technology[4]
Bharati Vidyapeeth Institute of Technology Kharghar, Navi Mumbai, Maharashtra, India

Abstract: *Brain tumor segmentation from multi-modal magnetic resonance imaging (MRI) is a critical task for clinical diagnosis, treatment planning, and disease monitoring. Existing full 3D deep learning models often achieve high segmentation accuracy but impose substantial GPU memory requirements, limiting applicability on standard clinical hardware. To address this challenge, this paper presents a two-stage cascaded deep learning framework for accurate and memory-efficient 3D brain tumor segmentation.*

The proposed method combines a coarse 2.5D U-Net for global tumor localization with an attention-based 3D U-Net for fine-grained volumetric refinement. The first stage processes multi-view 2.5D contextual slices (5 adjacent slices × 4 MRI modalities = 20 input channels) to efficiently capture global tumor structure. The second stage refines boundaries using overlapping 96×96×96 voxel 3D patches, guided by probability maps from the first stage. A novel learnable residual mixing mechanism (LearnableResidualMix) adaptively balances coarse and refined predictions via a trainable scalar α.

The framework is trained and evaluated on the BraTS 2020 dataset using four MRI modalities (T1, T1ce, T2, FLAIR) across 369 patients. The final model (32 base filters, depth 5) achieves a mean Dice Similarity Coefficient of 0.865 on the validation set, with Dice scores of 0.932, 0.937, and 0.888 for Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET) respectively — outperforming nnU-Net (single model) and standard Attention U-Net baselines.. .

Keywords: Brain tumor segmentation, deep learning, cascaded neural networks, 2.5D U-Net, 3D U-Net, attention gates, BraTS 2020, medical image analysis, MRI, learnable residual mixing

I. INTRODUCTION

Brain tumors represent one of the most significant challenges in modern neuro-oncology, standing as one of the most life-threatening neurological disorders globally. The complexity of the human brain, coupled with the aggressive and heterogeneous nature of tumorous growths, necessitates high-precision diagnostic tools and meticulous treatment planning to enhance patient survival rates and long-term outcomes. Traditionally, Magnetic Resonance Imaging (MRI) has served as the gold standard for brain tumor analysis due to its superior soft-tissue contrast and non-invasive ability to capture detailed structural and physiological information.[6,7]

Brain tumor segmentation — the process of isolating tumorous tissue from healthy brain matter and identifying sub-regions such as the whole tumor (WT), tumor core (TC), and enhancing tumor (ET) — is the foundational step for tumor volume estimation, radiotherapy planning, and monitoring treatment response. However, clinical practice still relies heavily on manual slice-by-slice delineation by trained radiologists. This approach is time-consuming, labor-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, March 2026
Impact Factor: 8.2

intensive, and subject to significant inter-observer and intra-observer variability. As imaging volumes grow, the demand for automated, reliable, and scalable segmentation solutions has never been higher.[6]

A central technical challenge is the trade-off between dimensionality and computational resources. Full 3D deep learning models capture volumetric spatial context effectively but demand substantial GPU memory that often exceeds standard clinical hardware. Conversely, 2D models are computationally efficient but discard critical spatial context across the depth axis, making them unsuitable for clinical 3D tumor characterization.[1,2]

This work addresses this gap by presenting a hybrid cascaded framework. A lightweight 2.5D U-Net first performs fast global localization, and its output probability maps then guide a targeted 3D Attention U-Net for precise boundary refinement. Central to this design is the LearnableResidualMix layer — a trainable mechanism that adaptively blends Stage 1 and Stage 2 probability outputs, ensuring training stability and continuity between cascade stages. The resulting system delivers competitive 3D segmentation performance while remaining feasible on dual NVIDIA T4 GPUs available in standard cloud environments.

The main contributions of this work are:

- A two-stage cascaded 2.5D–3D segmentation framework that enables memory-efficient processing of full MRI volumes, eliminating the need for high-end volumetric GPU hardware.
- A multi-view 2.5D coarse segmentation network that processes axial, sagittal, and coronal views simultaneously and produces ensemble probability maps as spatial priors.
- An attention-based 3D refinement network guided by Stage 1 probability maps to suppress false positives and improve tumor boundary delineation.
- A novel LearnableResidualMix mechanism that adaptively blends Stage 1 and Stage 2 predictions in probability space via a trained scalar weight α.
- Comprehensive evaluation on BraTS 2020 demonstrating a mean Dice of 0.865, surpassing nnU-Net (single model) and Attention U-Net baselines.

II. MOTIVATION

The motivation behind this project is driven by the critical limitations of existing brain tumor segmentation approaches, which force a difficult compromise between segmentation accuracy and computational feasibility — a compromise that directly impacts the quality and accessibility of neuro-oncological care. Current full 3D deep learning models, while capable of capturing rich volumetric context, impose prohibitive GPU memory demands that restrict their deployment to high-end research hardware. At the same time, purely 2D approaches sacrifice the spatial continuity across MRI slices that is essential for accurate sub-region delineation, particularly for clinically critical structures such as the enhancing tumor core. This project is motivated by the need to resolve this fundamental tension through a principled architectural design that retains the spatial awareness of 3D processing while remaining executable on hardware available in standard clinical and cloud computing environments.

Beyond hardware constraints, this project is motivated by the clinical reality that brain tumor sub-region segmentation — distinguishing the necrotic core, peritumoral edema, and enhancing tumor across four MRI modalities — demands a level of precision and consistency that existing automated systems struggle to deliver reliably. Inter-observer variability among radiologists, combined with the heterogeneous and irregular morphology of high-grade gliomas, introduces diagnostic uncertainty that cascades downstream into treatment planning errors and inconsistent monitoring of therapy response. By developing a cascaded 2.5D–3D framework with attention-guided refinement and a novel learnable blending mechanism, this project seeks to establish a segmentation pipeline that is not only accurate but also architecturally transparent in how it progressively refines its predictions — from coarse global localization to precise boundary delineation — mirroring the structured reasoning process of an experienced radiologist.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 5, March 2026
Impact Factor: 8.2

III. LITERATURE SURVEY

The evolution of brain tumor segmentation has transitioned from manual clinical practice to automated deep learning systems. This section reviews the major developments that motivate the proposed approach.

Manual and Classical Approaches

Traditionally, brain tumor segmentation is performed manually by trained radiologists through slice-by-slice delineation across 3D MRI volumes. While this remains the clinical reference standard, it is exceptionally time-consuming and subject to significant inter-observer variability. Early automated attempts used classical image processing techniques such as intensity thresholding, K-means clustering, Fuzzy C-Means, and watershed segmentation. These methods require heavy manual parameter tuning and lack robustness to the irregular shapes, blurred boundaries, and intensity inhomogeneities characteristic of high-grade gliomas.

Convolutional Neural Network Approaches

The introduction of Convolutional Neural Networks (CNNs) transformed medical image segmentation. The U-Net architecture [1], with its encoder-decoder structure and skip connections, became the benchmark for medical segmentation tasks by capturing both local texture and global spatial features simultaneously. To exploit volumetric context, 3D extensions such as 3D U-Net [2] and V-Net [3] were developed, processing 3D patches to understand spatial continuity across slices. However, these models impose substantial GPU memory requirements, often exceeding the capacity of standard hardware when processing full brain volumes.

The Attention U-Net [4] extended the U-Net by incorporating attention gates at skip connections, allowing the decoder to selectively focus on task-relevant regions and suppress irrelevant background activations. This is particularly valuable in brain tumor segmentation, where the tumor occupies only a small fraction of the total brain volume.

Cascaded and Automated Frameworks

nnU-Net [5] established a self-configuring framework that automatically adapts preprocessing, architecture, and training strategies to any medical segmentation dataset, achieving state-of-the-art results on numerous benchmarks. However, nnU-Net requires substantial compute resources for full 3D operation.

Cascaded architectures have emerged as a practical compromise: a low-resolution or 2D/2.5D model first identifies the region of interest, and a high-resolution 3D model then focuses on that region for detailed segmentation. This coarse-to-fine strategy has been shown to reduce memory requirements significantly while maintaining high accuracy, as the 3D model operates only on small, targeted patches rather than the full volume. The proposed method builds on this paradigm while introducing a novel trainable blending mechanism between cascade stages.[5]

IV. PROPOSED SYSTEM

The proposed system is a cascaded deep learning framework for automated brain tumor segmentation, designed as a two-stage pipeline that processes multi-modal MRI volumes from raw NIfTI input to a final voxel-level segmentation mask. It is built to run on standard dual-GPU cloud hardware, meaning hospitals and research institutions do not require expensive high-memory workstations or institutional-scale computing infrastructure to deploy it. The core of the system is a 2.5D U-Net in the first stage, which processes multi-view contextual slices across axial, sagittal, and coronal planes simultaneously, capturing global tumor structure efficiently without the memory overhead of full volumetric convolutions. This stage produces ensemble probability maps that serve as spatial priors, effectively localizing the tumor region before any detailed refinement is attempted.

Building on this coarse localization, the second stage introduces a 3D Attention U-Net that operates on targeted 96×96×96 voxel patches guided by the Stage 1 probability maps. Attention gates at every skip connection allow the network to selectively focus on tumor boundary voxels and suppress irrelevant background activations, producing precise sub-region delineations across the three clinically defined tumor regions — Whole Tumor, Tumor Core, and

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, March 2026 Impact Factor: 8.2

Enhancing Tumor. A novel LearnableResidualMix mechanism adaptively blends the Stage 1 and Stage 2 probability outputs through a single trainable scalar, ensuring training stability and a graceful transition from coarse to refined predictions across epochs.

Additionally, the system incorporates a robust preprocessing pipeline and carefully designed training strategy tailored to the BraTS 2020 dataset. MAD-based intensity normalization, tumor-biased slice sampling, and multi-view ensemble inference ensure that the model generalizes well despite the limited patient count of 369 cases. By combining memory-efficient 2.5D global localization with attention-guided 3D volumetric refinement and a learnable cascade blending mechanism, the proposed system delivers a mean Dice score of 0.865 on the BraTS 2020 validation set — outperforming nnU-Net single model and Attention U-Net baselines — while remaining fully executable on hardware accessible to standard clinical and cloud environments.

V. PROPOSED FRAMEWORK

Coarse Tumor Localization Framework: The first framework provides a lightweight 2.5D U-Net that processes multi-modal MRI volumes as input. The user supplies four co-registered MRI modalities — T1, T1ce, T2, and FLAIR — which are preprocessed and sampled into multi-view 2.5D slices across axial, sagittal, and coronal planes. The framework outputs ensemble probability maps that coarsely localize the tumor region, serving as spatial priors for the refinement stage.

Volumetric Refinement Framework: The second framework encompasses the 3D Attention U-Net that operates on targeted 96×96×96 voxel patches. It takes as input the four normalized MRI modalities concatenated with the Stage 1 probability maps, forming an 8-channel tensor. The framework applies attention gates at every decoder skip connection to focus on tumor boundaries, producing refined voxel-level segmentation masks across all three tumor sub-regions.

Learnable Cascade Blending Framework: The third framework provides the LearnableResidualMix mechanism that fuses Stage 1 and Stage 2 outputs. It takes the softmax probability maps from both stages as input and blends them using a single trainable scalar α initialized at 0.5. The framework adaptively shifts weight toward the refined Stage 2 prediction during training, outputting a final probability distribution that converges to full reliance on the 3D refinement network.

Preprocessing and Data Pipeline Framework: The fourth framework provides the standardized preprocessing pipeline applied to all input volumes before either network stage. It accepts raw NIfTI MRI files as input and applies spatial resizing, percentile-based intensity clipping, and MAD-based robust z-score normalization. Because this framework operates independently of the network architecture, it ensures consistent, scanner-agnostic input representation and reproducible training across the full 369-patient BraTS 2020 dataset.

VI. RESULTS AND ANALYSIS

The proposed cascaded 2.5D–3D framework was evaluated on the 57-patient held-out validation split of the BraTS 2020 dataset, following the standard BraTS challenge evaluation protocol. The final model configuration, comprising 32 base filters and encoder depth 5, achieved a mean Dice Similarity Coefficient of 0.865 across the three clinically defined tumor sub-regions. Individual sub-region scores were 0.932 for Whole Tumor, 0.937 for Tumor Core, and 0.888 for Enhancing Tumor — all computed volumetrically in 3D rather than slice-by-slice, ensuring that the reported metrics reflect true spatial overlap consistent with clinical measurement standards.

Comparative evaluation against established baselines demonstrates the effectiveness of the proposed approach. The final model outperforms the initial configuration of 24 base filters and depth 4, which achieved a mean Dice of only 0.815, confirming that increased model capacity in the 3D refinement stage translates directly into measurable segmentation gains. More significantly, the proposed cascade surpasses both the nnU-Net single model and the standard Attention U-Net baseline across all three sub-regions, despite being trained on standard dual NVIDIA T4 GPUs rather than the high-memory hardware typically required by full 3D architectures. This demonstrates that the cascaded design successfully closes the performance gap between memory-efficient and resource-intensive approaches.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 5, March 2026 Impact Factor: 8.2

The behavior of the LearnableResidualMix mechanism across training epochs provides additional insight into how the cascade functions. The scalar α began at 0.465 at the start of Stage 2 training, indicating a near-equal blend of Stage 1 and Stage 2 predictions, and progressively rose to 0.907 by epoch 6 before converging to 1.0 from epoch 7 onward. This convergence pattern confirms that the 3D Attention U-Net successfully learned to supersede the coarse 2.5D prior for final delineation, while the blending mechanism provided training stability during the early epochs when Stage 2 was still adapting to the patch-level refinement task. The graceful transition from coarse to refined predictions validates the architectural motivation behind introducing a learnable rather than fixed blending weight.

The Enhancing Tumor sub-region, despite being the smallest and most clinically critical region, achieved a Dice score of 0.888 — notably higher than many comparable single-stage models reported in the literature. This result is largely attributable to the class-weighted Dice loss used in Stage 2 training, which assigned the highest weight of 2.5 to the Enhancing Tumor class, and to the attention gate mechanism that directed the network's focus toward precise boundary voxels in this challenging sub-region. Taken together, the quantitative results and training dynamics demonstrate that the proposed framework achieves a strong balance between segmentation accuracy, architectural efficiency, and clinical relevance across all three tumor sub-regions.

VII. CONCLUSION

This paper presented a two-stage cascaded deep learning framework for automated brain tumor segmentation on the BraTS 2020 dataset, addressing the fundamental trade-off between segmentation accuracy and computational feasibility that limits the clinical deployment of existing full 3D models. By combining a lightweight 2.5D U-Net for global tumor localization with an attention-guided 3D U-Net for volumetric boundary refinement, the proposed system delivers competitive segmentation performance while remaining entirely executable on standard dual NVIDIA T4 GPU hardware. The multi-view ensemble strategy in Stage 1 and the attention gate mechanism in Stage 2 together ensure that the framework captures both global spatial context and fine-grained sub-region boundaries without requiring the prohibitive memory overhead of end-to-end 3D processing.

REFERENCES

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, 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 MICCAI, Springer, 2016, pp. 424–432.

[3] F. Milletari, N. Navab, and S. A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," in 3D Vision (3DV), IEEE, 2016, pp. 565–571.

[4] J. Schlemper, O. Oktay, M. Schaap, et al., "Attention Gating for Improving Completeness of Encoder Representations," in Medical Image Analysis, vol. 53, 2019, pp. 197–207.

[5] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation," Nature Methods, vol. 18, pp. 203–211, 2021.