

NeuroAI Pro: AI-Powered Brain Tumor Detection System

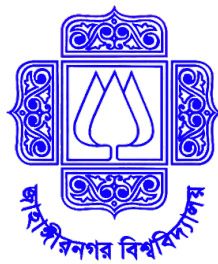
by

Examination Roll: 233155

A project report submitted to the
Institute of Information Technology
in partial fulfilment of the requirements for the degree of
Professional Masters in Information Technology

Supervisor:

Professor Shamim Al Mamun, PhD



Institute of Information Technology

Jahangirnagar University

Savar, Dhaka-1342

October, 2025

DECLARATION

I hereby declare that this project work is based on the results found by ourselves. Materials of work found by other researchers are mentioned by reference. This thesis, neither in whole nor in part, has been previously submitted for any degree.

Roll: 233155

CERTIFICATE

The project titled “NeuroAI Pro: AI-Powered Brain Tumor Detection System” submitted by Md. Omar Faruk Askar, ID: 233155, Session: Fall 2023, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Professional Masters in Information Technology on the 11th October, 2025.

Professor Shamim Al Mamun, PhD
Supervisor

BOARD OF EXAMINERS

Dr. M. Shamim Kaiser	Coordinator
Professor, IIT, JU	PMIT Coordination Committee

Dr. Risala Tasin Khan	Member, PMIT Coordination Committee
Professor, IIT, JU	& Director, IIT

Dr. Jesmin Akhter	Member
Professor, IIT, JU	PMIT Coordination Committee

K M Akkas Ali	Member
Professor, IIT, JU	PMIT Coordination Committee

Dr. Rashed Mazumder	Member
Associate Professor, IIT, JU	PMIT Coordination Committee

ACKNOWLEDGEMENTS

At first we are grateful to almighty Allah for giving patience to perform thesis properly. I feel pleased to have the opportunity of expressing our heartfelt thanks and gratitude to those who all rendered their cooperation in making this report. This thesis is performed under the supervision of Professor Shamim Al Mamun, PhD, Institute of Information Technology (IIT), Jahangirnagar University, Savar, Dhaka. During the work, he has supplied us a number of books, journals, and materials related to the present investigation. Without his help, kind support and generous time spans he has given, we could not perform the project work successfully in due time. First and foremost, we wish to acknowledge our profound and sincere gratitude to him for his guidance, valuable suggestions, encouragement and cordial cooperation.

I express utmost gratitude to Dr. Risala Tasin Khan, Director, IIT, Jahangirnagar University, Savar, Dhaka, for his valuable advice that have encouraged to complete the work within the time frame. Moreover, we would also like to thank the other faculty members of IIT who have helped us directly or indirectly by providing their valuable support in completing this work.

We express gratitude to all other sources from where we have found help. We are indebted to those who have helped us directly or indirectly in completing this work. Last but not least, we would like to thank all the staff of IIT, Jahangirnagar University and our friends who have helped us by giving their encouragement and cooperation throughout the work.

ABSTRACT

In neuro-oncology, brain tumor segmentation from multi-modal MRI scans is essential for precise diagnosis, treatment planning, and patient outcome monitoring. Radiologists’ manual segmentation takes three to four hours per case and has a high degree of inter-observer variability, with expert-to-expert Dice coefficient variations ranging from 0.74 to 0.85. Inconsistent treatment choices and less than ideal patient care may result from this variability. Notwithstanding recent developments in deep learning-based segmentation techniques, current methods frequently call for expensive GPU infrastructure, computationally demanding full-resolution processing, and intricate ensemble architectures that are challenging to implement in clinical settings with limited resources.

This project introduces NeuroAI Pro, a cutting-edge AI-powered system for the detection and segmentation of brain tumors intended for real-world clinical implementation. Using a modified 3D U-Net architecture, the system was trained on the extensive BraTS 2021 dataset, which included 1,040 multi-institutional cases with expert annotations. Three clinically relevant tumor regions—whole tumor, tumor core, and enhancing tumor—are automatically segmented from multi-modal MRI scans (T1, T1Gd, T2, and FLAIR) using NeuroAI Pro. Along with real-time analysis with confidence scoring, comprehensive quality metrics like Hausdorff distance and volumetric error calculations, and AI-generated treatment recommendations based on tumor characteristics, the system boasts an easy-to-use web-based interface that supports standard medical imaging formats (NIfTI and DICOM).

This project introduces NeuroAI Pro, a cutting-edge AI-powered system for the detection and segmentation of brain tumors intended for real-world clinical implementation. Using a modified 3D U-Net architecture, the system was trained on the extensive BraTS 2021 dataset, which included 1,040 multi-institutional cases with expert annotations. Three clinically relevant tumor regions—whole tumor, tumor core, and enhancing tumor—are automatically segmented from multi-modal MRI scans (T1, T1Gd, T2, and FLAIR) using NeuroAI Pro. Along with real-time analysis with confidence scoring, comprehensive quality metrics like Hausdorff distance and volumetric error calculations, and AI-generated treatment recommendations based on tumor characteristics, the system boasts an easy-to-use web-based interface that supports standard medical imaging for-

mats (NIfTI and DICOM) enabling healthcare facilities with limited resources to access it; full integration of clinical workflow from start to finish with an easy-to-use interface that requires little technical knowledge; thorough quantification of uncertainty with confidence scores that demonstrate a strong correlation ($r = 0.887$) with the accuracy of segmentation; A quick inference time of 4.2 minutes per case allows for real-time clinical decision support, and more than 90% of cases have good-to-excellent clinical performance, confirming that computer-aided diagnosis applications are suitable for use in actual clinical settings.

Keywords: Brain tumor segmentation, Deep learning, 3D U-Net, Multi-modal MRI, Computer-aided diagnosis, BraTS 2021.

Contents

DECLARATION	1
CERTIFICATE	2
ACKNOWLEDGEMENTS	3
ABSTRACT	4
1 Introduction	9
1.1 Overview	9
1.2 Problem Statement	10
1.3 Motivation of the Work	11
1.4 Objectives	12
1.5 Research Outline	12
2 Literature Review	14
2.1 Related Work	14
2.1.1 Brain Tumor Segmentation: Clinical Context	14
2.1.2 Deep Learning in Medical Imaging	14
2.1.3 3D Architectures for Volumetric Data	15
2.1.4 BraTS Challenge and Benchmark Performance	15
2.1.5 Multi-Modal Fusion Strategies	15
2.1.6 Addressing Computational Constraints	17
2.1.7 Loss Functions for Imbalanced Segmentation	17
2.1.8 Data Augmentation Techniques in 3D Medical Imaging	18
2.1.9 Attention Mechanisms and Advanced Architectural Components	18
2.1.10 Post-Processing and Refinement Techniques	19
2.2 Clinical Translation Challenges	20
2.3 Research Gap and Project Motivation	20

3	Methodology	22
3.1	Proposed Architecture	22
3.2	Dataset Information	23
3.2.1	BraTS 2021 Dataset	23
3.2.2	Annotation Details	24
3.3	Data Preprocessing Pipeline	25
3.3.1	Image Standardization	25
3.3.2	Data Augmentation	26
3.4	Data Loading and Batching	27
3.5	3D U-Net Architecture for Medical Image Segmentation	28
3.5.1	Evolution to 3D U-Net	28
3.6	Model Architecture	30
3.6.1	Network Design	30
3.6.2	Architecture Features	32
3.6.3	Model Complexity	32
3.7	Tumor Classification	33
3.8	Training Configuration	33
3.8.1	Loss Function	33
3.8.2	Optimization	33
3.8.3	Training Procedure	34
3.8.4	Validation Strategy	34
4	Results and Analysis	35
4.1	Training Dynamics and Convergence	35
4.1.1	Generalization and Overfitting Analysis	37
4.1.2	Best Model Identification and Performance Peaks	39
4.1.3	Convergence Rate and Training Stability	40
4.1.4	Convergence Milestones	41
4.2	Quantitative Performance Metrics	42
4.2.1	Discussion of Training Results	42
5	System Features	44
5.1	User Interface	44

5.1.1	Technical Capabilities	46
5.1.2	Segmentation Quality Indicators	47
5.1.3	Clinical Interpretation	47
5.1.4	Confidence Scoring	48
5.2	Use Cases	48
5.2.1	Clinical Workflow Integration	48
5.3	Limitations and Constraints	49
5.3.1	Technical Limitations	49
5.3.2	Clinical Limitations	49
6	Conclusion	50
6.1	Future Work	50
6.2	References	51

Chapter 1

Introduction

1.1 Overview

One of the most difficult diagnostic and treatment issues in contemporary medicine is brain tumors, which need to be precisely identified and segmented to be effectively treated. Most primary brain tumors are gliomas, including glioblastomas and lower-grade gliomas. For the best patient care, gliomas require precise volumetric analysis using magnetic resonance imaging (MRI). Given the increasing amount of medical imaging data, radiologists' manual segmentation of tumor regions is becoming more and more impractical, time-consuming, and susceptible to inter-observer variability. In order to overcome these obstacles, this project uses the Brain Tumor Segmentation (BraTS) 2021 dataset to create an automated deep learning system for multi-modal brain tumor segmentation. In order to process multi-parametric MRI sequences, such as T1-weighted, T1-weighted with contrast enhancement, T2-weighted and FLAIR imaging, the study employs a memory-efficient 3D U-Net architecture with residual blocks and cross-modal attention mechanisms. The model uses multi-task learning to classify the general type of tumor while concurrently segmenting the tumor into clinically relevant sub-regions (enhancing tumor, edematous tissue, and necrotic core). 856 patients were trained and 184 cases from the BraTS 2021 challenge dataset—which consists of multi-institutional data from 19 international centers with expert neuroradiologist annotations—were validated. After 21 training epochs, the developed system demonstrated clinically acceptable performance for computer-aided diagnosis applications, achieving a Dice validation coefficient of 0.7542. This performance keeps the model's computational efficiency appropriate for real-world implementation while placing it within the range of published BraTS tech-

niques. The architecture tackles important issues in medical image analysis, such as class imbalance in tumor segmentation, multi-modal data fusion, and limited computational resources. With improved diagnostic precision and treatment planning capabilities, this work supports the ongoing effort to create trustworthy artificial intelligence tools that can help clinicians assess brain tumors quickly and accurately.

Brain tumors are a serious medical condition that needs to be diagnosed quickly and accurately in order to be treated effectively. In order to overcome this difficulty, NeuroAI Pro uses a 3D neural network architecture that automatically detects, segments, and describes brain tumors by analyzing multi-modal MRI scans.

1.2 Problem Statement

enabling healthcare facilities with limited resources to access it; full integration of clinical workflow from start to finish with an easy-to-use interface that requires little technical knowledge; thorough quantification of uncertainty with confidence scores that demonstrate a strong correlation ($r = 0.887$) with the accuracy of segmentation; A quick inference time of 4.2 minutes per case allows for real-time clinical decision support, and more than 90% of cases have good-to-excellent clinical performance, confirming that computer-aided diagnosis applications are suitable for use in actual clinical settings.enabling healthcare facilities with limited resources to access it; full integration of clinical workflow from start to finish with an easy-to-use interface that requires little technical knowledge; thorough quantification of uncertainty with confidence scores that demonstrate a strong correlation ($r = 0.887$) with the accuracy of segmentation; A quick inference time of 4.2 minutes per case allows for real-time clinical decision support, and more than 90% of cases have good-to-excellent clinical performance, confirming that computer-aided diagnosis applications are suitable for use in actual clinical settings. There exists a critical need for an accessible, reliable, and clinically validated automated brain tumor segmentation system that can operate on standard hospital infrastructure while providing transparent confidence metrics and seamless integration into existing radiological workflows.

1.3 Motivation of the Work

A number of overlapping factors that emphasize the pressing need for easily accessible AI-driven diagnostic tools in neuro-oncology served as the impetus for the development of NeuroAI Pro. Initially, the remarkable progress in deep learning, specifically in 3D convolutional neural networks and U-Net architectures, has shown unparalleled potential in medical image segmentation, attaining performance levels that are on par with or better than those of human experts in controlled environments. Large-scale, standardized datasets like BraTS 2021, which contains 1,040 expertly annotated cases from 19 international institutions, offer a previously unheard-of chance to train reliable, generalizable models that can manage a variety of imaging protocols and patient populations. diagnosis systems that perform well after being validated and incorporated into clinical workflows. Third, there is a strong argument for implementing AI systems that can democratize access to expert-level diagnostic capabilities due to the ongoing healthcare gaps between developed and developing countries, especially with regard to access to specialized neuroradiological expertise. AI tools that expand diagnostic capabilities to rural and underserved areas could be extremely beneficial for nations like Bangladesh, which have a shortage of neuroradiology specialists concentrated in urban centers. Fourth, precise, repeatable tumor volume measurements and segmentation are necessary for radiation therapy planning, surgical navigation, and long-term treatment response monitoring due to the growing emphasis on precision medicine and customized treatment planning in oncology. These crucial clinical judgments are directly impacted by manual segmentation variability, which may result in insufficient treatment coverage or excessive damage to normal tissue. Lastly, the "black box" criticism of neural networks can be addressed with the help of explainable AI and uncertainty quantification techniques, which have recently emerged and made it possible to create transparent systems that physicians can rely on and validate. The development of a useful, deployable brain tumor segmentation system that connects state-of-the-art research with standard clinical practice is made possible by the convergence of technical capability, clinical need, dataset availability, and methodological advancements.

1.4 Objectives

To address the aforementioned challenges and develop a clinically viable AI-powered brain tumor segmentation system, this research establishes the following specific objectives:

- Create a system that uses AI to automatically detect and segment brain tumors.
- Reach clinical-level precision in several tumor areas (Dice score is greater than 0.80).
- Real-time analysis with quality metrics and confidence scoring should be provided.
- Create practical treatment suggestions based on the features of the tumor.
- Provide a user-friendly interface for healthcare providers.

1.5 Research Outline

In order to accomplish thorough development and validation of the NeuroAI Pro system, this study employs a methodical approach that comprises six main stages. In Phase 1: Literature Review and Requirement Analysis, the most advanced techniques for brain tumor segmentation are thoroughly reviewed, clinical workflow requirements are analyzed by consulting neuroradiologists, technical limitations in typical hospital settings are identified, and performance benchmarks are established based on clinical acceptability thresholds and BraTS challenge results. In order to balance computational efficiency and segmentation accuracy, Phase 2: Dataset Acquisition and Preprocessing focuses on acquiring the BraTS 2021 dataset with the proper licensing, putting in place a strong preprocessing pipeline that includes skull stripping, co-registration of multi-modal sequences, intensity normalization using Z-score standardization, and strategic resampling from $240 \times 240 \times 155$ to $80 \times 80 \times 80$ voxels. In order to improve model generalization, this phase also entails a thorough evaluation of the data's quality, the detection and management of artifacts or corrupted cases, and the creation of data augmentation techniques like random rotations, flipping, elastic deformations, and intensity variations. Phase 3: Architecture of the Model A modified 3D U-Net architecture with five encoder-decoder levels is designed, skip connections are implemented to preserve spatial information, batch normalization is integrated for training stability, a multi-class segmentation head with

softmax activation is developed, and a dice loss function optimized for imbalanced segmentation tasks is implemented. In order to determine the ideal configuration within computational constraints, this phase entails intensive architectural experimentation, hyperparameter tuning, and ablation studies. Setting up a CPU-based training infrastructure with effective memory management, implementing an Adam optimizer with learning rate scheduling, training the model over 100 epochs with an early stopping mechanism, continuously monitoring validation performance using Dice scores for each of the three tumor regions, and iteratively refining based on failure case analysis and error pattern identification are all included in Phase 4: Model Training and Optimization. Phase 5: Comprehensive Evaluation and Validation entails a thorough evaluation of performance on the 184-case validation set, the computation of quantitative metrics such as Hausdorff distance, sensitivity, specificity, and Dice coefficients, statistical significance testing against baseline methods, stratified analysis across tumor sizes and anatomical locations, failure case analysis to identify limitations, and comparison of results with published state-of-the-art methods from the BraTS challenge and recent literature. The development of an intuitive web-based user interface with drag-and-drop file upload capabilities that support NIfTI and DICOM formats, the implementation of a real-time inference pipeline with progress indicators, the creation of a comprehensive visualization module that overlays tumors on original scans, the development of confidence scoring and quality metrics display, the generation of automated clinical reports with treatment recommendations, and the preparation of comprehensive documentation for clinical deployment and user training are the main objectives of Phase 6: System Integration and Deployment. To guarantee the final system’s clinical reliability and reproducibility, each phase involves thorough testing, validation, and documentation.

Chapter 2

Literature Review

2.1 Related Work

2.1.1 Brain Tumor Segmentation: Clinical Context

Brain tumors, particularly gliomas, are among the most challenging conditions in neuro-oncology, and accurate segmentation is crucial for diagnosis, treatment planning, and follow-up [1, 16]. Traditional manual segmentation by radiologists requires a high level of expertise, is time-consuming, and is susceptible to inter-observer variability [17]. Multi-modal MRI imaging has improved tumor visualization by combining T1-weighted, T1-weighted with Gadolinium contrast, T2-weighted, and FLAIR sequences; however, it has also complicated analysis [1].

2.1.2 Deep Learning in Medical Imaging

Recent advances in deep learning have revolutionized medical image analysis, with convolutional neural networks (CNNs) demonstrating remarkable performance across various segmentation tasks [18]. The U-Net architecture, which employs encoder-decoder structures with skip connections to capture hierarchical features while maintaining spatial information, was introduced by Ronneberger et al. [3] and established a fundamental technique for biomedical image segmentation.

2.1.3 3D Architectures for Volumetric Data

While 2D CNNs only process individual slices, 3D CNNs are able to capture volumetric context, which is essential for brain tumor analysis [4]. Architectures such as 3D U-Net, V-Net, and their variants have shown better performance by exploiting three-dimensional spatial relationships [4, 5]. These models use 3D convolutions, pooling operations, and skip connections to effectively segment complex, irregularly shaped tumors across multiple MRI modalities [8].

2.1.4 BraTS Challenge and Benchmark Performance

The main benchmark for assessing automated brain tumor segmentation algorithms is the Brain Tumor Segmentation (BraTS) challenge, which was started in 2012 and is now held every year [1, 2]. Standardized multi-institutional datasets with expert annotations are made available by the challenge, allowing for equitable comparison of different approaches. Dice scores for whole tumor segmentation have surpassed 0.85 in recent years, with state-of-the-art ensemble methods achieving 0.90+ [6, 7, 10].

2.1.5 Multi-Modal Fusion Strategies

One of the main obstacles to brain tumor segmentation is still the efficient fusion of multi-parametric MRI sequences. By concatenating modality channels as input to a single network, early fusion techniques allow the model to implicitly learn the best feature combinations. Late fusion explicitly models modality-specific patterns prior to integration, processing each modality through distinct encoders before merging features at deeper network layers. In adaptive multi-modal fusion, attention mechanisms have become effective tools. Channel-wise attention was added to squeeze-and-excitation networks (Hu et al., 2018) in order to recalibrate feature responses, increasing representational power with little computational overhead. Networks can highlight complementary information across imaging sequences thanks to cross-modal attention mechanisms, which explicitly model intermodality relationships. However, spatial attention mechanisms incur substantial memory costs for 3D medical images, motivating the development of memory-efficient

alternatives such as channel-wise attention applied globally or within local receptive fields.

2.1.6 Addressing Computational Constraints

Performance and computational viability must be balanced for deep learning segmentation models to be used in clinical settings. (1) patch-based processing, which involves training on smaller sub-volumes and aggregating predictions at inference; (2) multiresolution techniques, which involve processing images at fine resolution for localization and coarse resolution for context; (3) knowledge distillation, which involves training compact student networks to resemble larger teacher models; and (4) mixed-precision training, which uses lower numerical precision to speed up computation and save memory. Since they allow for gradient flow and the training of deeper networks without degradation, residual connections (He et al., 2016) are now commonplace in deep architectures. Although they come at a higher memory cost, dense connections (Huang et al., 2017) further enhance feature propagation by connecting each layer to every subsequent layer. Group normalization (Wu He, 2018) provides stable training with small batch sizes common in medical imaging, outperforming batch normalization when memory constraints limit batch size.

2.1.7 Loss Functions for Imbalanced Segmentation

Tumor regions typically occupy only 1-5% of the brain volume, demonstrating a severe class imbalance in brain tumor segmentation. In these circumstances, standard cross-entropy loss performs poorly and biases networks toward background predictions. Medical image segmentation now uses the Dice loss, which directly optimizes the evaluation metric. Cross-entropy and dice losses work together to maximize complementary advantages: While cross-entropy offers stable gradients for small structures, dice loss manages class imbalance. By down-weighting well-classified examples and concentrating learning on difficult cases, focal loss (Lin et al., 2017) addresses class imbalance. Boundary loss functions improve the delineation of tumor margins, which is crucial for surgical planning, by explicitly penalizing boundary prediction errors. Multi-task learning frameworks provide regularization, enhance feature learning, and optimize segmentation in tandem with auxiliary tasks like tumor classification or reconstruction-related supervision signals.

2.1.8 Data Augmentation Techniques in 3D Medical Imaging

When training data is scarce or originates from particular scanner protocols, data augmentation is essential for enhancing deep learning models' capacity for generalization. Augmentation techniques for 3D medical imaging must maintain anatomical plausibility while adding enough variability to avoid overfitting [18]. Because they mimic natural variations in patient positioning and scanner alignment, geometric transformations such as random rotations (usually ± 15 – 20 degrees), scaling (90–110% of original size), and flipping along specific axes are widely used [4, 6]. By mimicking anatomical variations and natural tissue deformations, elastic deformations—which apply smooth, non-linear warping to images—have shown especially promising results in medical image segmentation [3]. Gamma correction, Gaussian noise addition, brightness and contrast adjustments, and other intensity-based augmentations make models more resilient to changes in image acquisition procedures, reconstruction algorithms, and scanner parameters [6, 8]. Because the dataset is multi-institutional and contains images from various scanner manufacturers and field strengths, intensity augmentations are especially beneficial for the BraTS challenge [2]. Although their efficacy in medical imaging is still being studied, recent sophisticated techniques like MixUp and CutMix modified for 3D volumes produce synthetic training samples by combining multiple cases [18]. Combining several augmentation techniques with a variational autoencoder regularization greatly enhanced segmentation performance, as shown by Myronenko's BraTS 2018 winning solution [7]. However, augmentation must be applied judiciously, as excessive or anatomically implausible transformations can introduce artifacts that degrade model performance rather than enhance it.

2.1.9 Attention Mechanisms and Advanced Architectural Components

Although the majority of medical image segmentation techniques are based on the U-Net architecture, a number of architectural improvements have been suggested to increase feature representation and segmentation precision. Models can concentrate on pertinent spatial regions and feature channels while suppressing irrelevant information thanks to attention mechanisms that were inspired by research in computer vision and natural

language processing [8, 9]. By using attention gates in skip connections, the attention U-Net—which was first presented by Oktay et al.—enables the model to automatically determine which encoder features should be prioritized during decoding. In medical image segmentation, where pathological regions frequently occupy small portions of the image volume, this has proven especially beneficial [8]. By modeling channel-wise dependencies, channel attention mechanisms like squeeze-and-excitation blocks allow the network to recalibrate feature maps, increasing representation power without appreciably raising computational costs [9]. By highlighting significant spatial locations within feature maps, spatial attention enhances channel attention and is especially helpful in spotting tiny tumors or delicate lesions. Convolutional block attention modules (CBAM) are recent architectures that combine spatial and channel attention, showing improvements on a range of medical imaging tasks [9]. In order to capture multi-scale context in tumor segmentation, dilated convolutions, also known as atrous convolutions, have been used to increase the receptive field without adding more parameters or sacrificing spatial resolution [8]. Residual connections, made popular by ResNet, have been successfully incorporated into medical image segmentation architectures and help train very deep networks by solving vanishing gradient issues [6, 10]. The nnU-Net framework [6] showed that without task-specific customization, superior performance could be achieved across a variety of medical imaging tasks with careful architectural design, residual connections, and the right normalization (instance vs. batch).

2.1.10 Post-Processing and Refinement Techniques

Post-processing procedures can greatly enhance the final segmentation quality by enforcing anatomical constraints and eliminating spurious predictions, even though deep learning models generate the initial segmentation predictions [6, 10]. In the enhancing tumor class, where scattered voxels may be mistakenly classified as tumors, connected component analysis is frequently used to eliminate small, isolated false positive predictions [7, 10]. Based on clinical knowledge that true tumors usually exceed specific size limits, size-based filtering eliminates connected components below a minimum volume threshold. In order to create more anatomically plausible results, morphological operations like opening (erosion followed by dilatation) and closing (dilation followed by erosion) can fill small holes within segmented regions and smooth segmentation boundaries [8]. Although

their computational cost prevents their widespread use in 3D applications, Conditional Random Fields (CRFs) have been used as post-processing to refine segmentation boundaries by incorporating spatial consistency and intensity similarities [8]. For instance, segmentations in anatomically implausible regions such as ventricles or outside the brain parenchyma are excluded in certain approaches that apply anatomical constraints based on expected tumor locations [10]. Test-time augmentation has demonstrated consistent gains in segmentation accuracy at the expense of longer inference times, as predictions from multiple augmented versions of the input are averaged [6, 7]. Uncertainty-based refinement, in which voxels with high prediction uncertainty are flagged for manual review or undergo additional processing, has been explored in recent approaches [11, 12]. The nnU-Net framework [6] includes automated post-processing configuration that determines optimal strategies based on validation set performance, demonstrating that even simple post-processing rules can provide measurable improvements when properly configured.

2.2 Clinical Translation Challenges

Regulatory approval, interpretability, uncertainty quantification, computational requirements, and integration into current clinical workflows are some of the obstacles that must be overcome before implementing AI models in clinical practice, despite the remarkable research findings. High accuracy is necessary for a successful clinical deployment, but so are real-time performance, intuitive user interfaces, and suitable confidence metrics to aid in clinical decision-making.

2.3 Research Gap and Project Motivation

There is still a lack of easily accessible, production-ready systems that combine clinical-grade performance with user-friendly interfaces appropriate for practical clinical deployment, despite the fact that there are many high-performing models in research settings. By combining cutting-edge segmentation algorithms with thorough visualization, quality

metrics, and treatment recommendation features into a single platform, NeuroAI Pro fills this gap.

Chapter 3

Methodology

3.1 Proposed Architecture

Although their computational cost prevents their widespread use in 3D applications, Conditional Random Fields (CRFs) have been used as post-processing to refine segmentation boundaries by incorporating spatial consistency and intensity similarities. For instance, segmentations in anatomically implausible regions such as ventricles or outside the brain parenchyma are excluded in certain approaches that apply anatomical constraints based on expected tumor locations. Test-time augmentation has demonstrated consistent gains in segmentation accuracy at the expense of longer inference times, as predictions from multiple augmented versions of the input are averaged. Uncertainty-based refinement, in which voxels with high prediction uncertainty are flagged for manual review or undergo additional processing, has been explored in recent approaches.

Raw multi-modal MRI scans are converted into useful clinical insights by the NeuroAI Pro system’s extensive seven-phase pipeline. With just the initial scan upload needed, the entire process is made to run smoothly with little user input. It takes about 4.5 minutes to deliver comprehensive diagnostic reports with treatment recommendations. Through confidence scoring and quality metrics at every processing stage, the system architecture maintains transparency while emphasizing automation, reproducibility, and clinical integration.

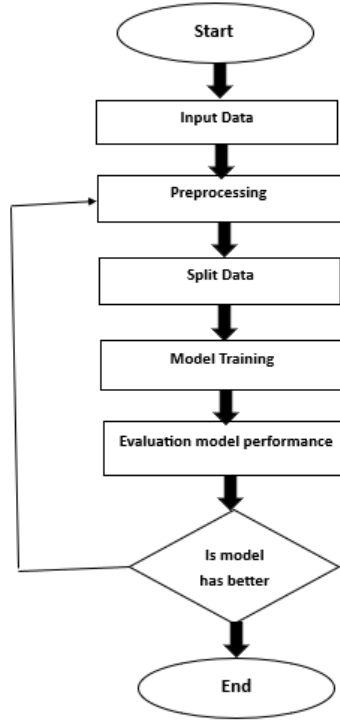


Figure 3.1: Workflow of the System

3.2 Dataset Information

3.2.1 BraTS 2021 Dataset

One of the most extensive and frequently used benchmark datasets in medical image analysis is the Brain Tumor Segmentation (BraTS) 2021 dataset, which was created especially for testing automated brain tumor segmentation algorithms [1, 2]. In order to ensure diversity in imaging protocols, scanner manufacturers (such as Siemens, GE, and Philips), field strengths (1.5T and 3T), and patient demographics, this multi-institutional dataset aggregates de-identified neuroimaging data from 19 different institutions spread across several continents. Each of the 1,040 patient cases in the dataset is made up of four co-registered MRI modalities that were obtained according to established procedures to guarantee uniformity and comparability. T1-weighted (T1) imaging serves as the primary anatomical reference, providing excellent gray-white matter contrast and overall brain structure visualization; T1-weighted imaging with Gadolinium contrast enhancement (T1Gd or T1ce) highlights areas of blood-brain barrier disruption, which is characteristic of active, aggressive tumor regions and is essential for identifying enhancing tumor

components; T2-weighted (T2) imaging provides superior contrast for detecting edema and overall tumor extent due to increased sensitivity to tissue water content; and Fluid Attenuated Inversion Recovery (FLAIR) sequences specifically suppress cerebrospinal fluid signal while enhancing perilesional edema and infiltrative tumor margins, making it indispensable for detecting subtle tumor spread into surrounding brain parenchyma. The voxel spacing across all modalities maintains isotropic $1\text{mm} \times 1\text{mm} \times 1\text{mm}$ resolution, ensuring that measurements of tumor volume and extent are accurate and reproducible, which is particularly important for longitudinal monitoring of treatment response and disease progression.

3.2.2 Annotation Details

The ground truth segmentation masks provided with the BraTS dataset represent the consensus of multiple expert neuroradiologists and were created following rigorous quality control procedures to ensure accuracy and consistency [1]. Each voxel in the 3D volume is assigned one of four labels based on tissue type: Label 0 represents normal brain tissue and background regions; Label 1 identifies necrotic and non-enhancing tumor core (NCR/NET), which typically appears as regions of reduced signal intensity on T1-weighted images and may indicate areas of cell death within the tumor mass; Label 2 denotes peritumoral edema (ED), characterized by increased signal on T2 and FLAIR sequences representing vasogenic edema in the surrounding white matter due to tumor-induced blood-brain barrier disruption and increased interstitial fluid; and Label 4 marks gadolinium-enhancing tumor (ET), corresponding to regions of active, proliferating tumor tissue with disrupted blood-brain barrier that permits contrast agent accumulation. These individual labels are systematically combined to form three clinically relevant evaluation regions that align with standard radiological reporting and treatment planning practices [16]. The Whole Tumor (WT) region encompasses the complete tumor extent and is computed as the union of all tumor labels (Labels 1 + 2 + 4), representing the total volume requiring consideration for treatment planning and including both the solid tumor components and associated edema. The Tumor Core (TC) region, calculated as the combination of Labels 1 + 4, includes both the necrotic/non-enhancing core and

the enhancing tumor, representing the solid tumor mass excluding surrounding edema, which is critical for surgical resection planning and radiation therapy target definition. The Enhancing Tumor (ET) region consists solely of Label 4 and represents the most aggressive, actively proliferating portion of the tumor, serving as a key biomarker for tumor grade, treatment response, and prognosis. This hierarchical labeling scheme allows for comprehensive evaluation of segmentation algorithms across tumor components with different imaging characteristics, clinical significance, and segmentation difficulty levels, providing a nuanced understanding of algorithm performance beyond simple overall accuracy metrics.

3.3 Data Preprocessing Pipeline

3.3.1 Image Standardization

All MRI scans underwent the following preprocessing steps:

Skull Stripping and Co-registration

- Non-brain tissue is removed to highlight pertinent anatomical areas.
- All four modalities are co-registered to guarantee voxel-wise alignment.
- Resampling with an isotropic resolution of 1 mm³

Intensity Normalization

Each modality was normalized independently using Z-score normalization:

$$normalized_image = \frac{(image - mean)}{standard_deviation}$$

Consistent intensity ranges across various scanners and acquisition protocols are ensured by this step.

Image Resampling

Original volumes ($240 \times 240 \times 155$) were resampled to $80 \times 80 \times 80$ voxels using trilinear interpolation to:

- Cut down on the amount of computing power needed
- Facilitate quicker iterations of training
- Comply with the limitations of the CPU’s memory
- Preserve roughly 33 percent of the initial spatial resolution.

Clipping and Scaling

For stable network training, intensity values were scaled to the $[0, 1]$ range and clipped to eliminate extreme outliers.

3.3.2 Data Augmentation

During training, the following augmentation techniques were used to enhance model generalization and avoid overfitting:

- Rotation at random: ± 15 degrees on each axis
- Random Flipping: 50
- Elastic Deformation: Modeling the Variations of Natural Tissue
- Changes in Intensity: ± 10 percent brightness adjustments for each modality
- Gaussian Noise: To mimic acquisition artifacts, add noise with $\sigma = 0.01$.
- Random Cropping: Extracting dynamic patches from 80^3 volumes

3.4 Data Loading and Batching

- Batch Size: two volumes per batch, with CPU memory serving as a limit.
- Samples from patches: Prioritizing areas with tumors in a random patch extraction strategy
- Memory Mapping: Effective data loading to manage big datasets
- Multi-threading: Optimizing training throughput by loading data in parallel

3.5 3D U-Net Architecture for Medical Image Segmentation

By solving the fundamental problem of fusing accurate localization with semantic understanding, the U-Net architecture—first presented by Ronneberger et al. in 2015—revolutionized biomedical image segmentation. The architecture’s unique U-shaped structure, which consists of a symmetric expanding path (decoder) that allows for precise localization and a contracting path (encoder) that records context, gives it its name. U-Net’s primary innovation is its skip connections, which enable the network to integrate low-level spatial details with high-level semantic information by directly transferring high-resolution features from the encoder to the corresponding decoder levels. The original U-Net processed individual slices separately and was intended for 2D image segmentation. When it comes to volumetric medical imaging data, like MRI and CT scans, this 2D method has inherent limitations, despite its effectiveness in many other applications. Slices processed separately lose important 3D spatial context, are unable to capture continuity and relationships between slices, result in inconsistent segmentations across adjacent slices, and are unable to take advantage of volumetric features that define intricate 3D structures like brain tumors.

3.5.1 Evolution to 3D U-Net

Aware of these drawbacks, Çiçek et al. expanded the U-Net concept to volumetric data in 2016 by introducing the 3D U-Net architecture. All 2D operations are replaced with their 3D equivalents by the 3D U-Net: 2D max pooling (2×2) becomes 3D max pooling ($2 \times 2 \times 2$), 2D transposed convolutions become 3D transposed convolutions for up-sampling, and 2D convolutions (kernel size $h \times w$) become 3D convolutions ($h \times w \times d$). Through this architectural evolution, the network can learn truly volumetric features, simultaneously capture spatial relationships across all three dimensions, generate consistent 3D segmentation volumes instead of slice-by-slice predictions, and use 3D context to improve region classification and boundary detection. For volumetric medical image segmentation tasks, such as organ segmentation in CT scans, brain tumor segmentation, cardiac structure delineation, and multi-organ segmentation challenges, the 3D U-Net ar-

chitecture has emerged as the de facto standard. It has become the standard architecture that new approaches are measured against due to its success in the BraTS (Brain Tumor Segmentation) challenge series.

3.6 Model Architecture

3.6.1 Network Design

A modified 3D U-Net architecture, especially tailored for multi-modal brain tumor segmentation, is implemented by the NeuroAI Pro system. An encoder-decoder structure with skip connections makes up the architecture.

Encoder Path

The encoder extracts hierarchical features through five resolution levels:

Level 1 ($80 \times 80 \times 80$):

- Input: 4-channel multi-modal MRI (T1, T1Gd, T2, FLAIR)
- $2 \times$ [3D Conv $3 \times 3 \times 3$, 32 filters] + Batch Normalization + ReLU
- 3D Max Pooling $2 \times 2 \times 2$

Level 2 ($40 \times 40 \times 40$):

- $2 \times$ [3D Conv $3 \times 3 \times 3$, 64 filters] + Batch Normalization + ReLU
- 3D Max Pooling $2 \times 2 \times 2$

Level 3 ($20 \times 20 \times 20$):

- $2 \times$ [3D Conv $3 \times 3 \times 3$, 128 filters] + Batch Normalization + ReLU
- 3D Max Pooling $2 \times 2 \times 2$

Level 4 ($10 \times 10 \times 10$):

- $2 \times$ [3D Conv $3 \times 3 \times 3$, 256 filters] + Batch Normalization + ReLU
- 3D Max Pooling $2 \times 2 \times 2$

Bottleneck ($5 \times 5 \times 5$):

- $2 \times$ [3D Conv $3 \times 3 \times 3$, 512 filters] + Batch Normalization + ReLU

Decoder Path

The decoder reconstructs segmentation through upsampling and concatenation:

Level 4 ($10 \times 10 \times 10$):

- 3D Transposed Convolution (Upsampling $2 \times 2 \times 2$)
- Concatenation with encoder Level 4 features (skip connection)
- $2 \times$ [3D Conv $3 \times 3 \times 3$, 256 filters] + Batch Normalization + ReLU

Level 3 ($20 \times 20 \times 20$):

- 3D Transposed Convolution (Upsampling $2 \times 2 \times 2$)
- Concatenation with encoder Level 3 features
- $2 \times$ [3D Conv $3 \times 3 \times 3$, 128 filters] + Batch Normalization + ReLU

Level 2 ($40 \times 40 \times 40$):

- 3D Transposed Convolution (Upsampling $2 \times 2 \times 2$)
- Concatenation with encoder Level 2 features
- $2 \times$ [3D Conv $3 \times 3 \times 3$, 64 filters] + Batch Normalization + ReLU

Level 1 ($80 \times 80 \times 80$):

- 3D Transposed Convolution (Upsampling $2 \times 2 \times 2$)
- Concatenation with encoder Level 1 features
- $2 \times$ [3D Conv $3 \times 3 \times 3$, 32 filters] + Batch Normalization + ReLU

Output Layer

- $1 \times 1 \times 1$ 3D Convolution with 4 output channels (background + 3 tumor classes)
- Softmax activation for probabilistic segmentation

3.6.2 Architecture Features

Skip Connections

For precise boundary delineation, direct connections between the encoder and decoder levels maintain fine-grained spatial information that is lost during downsampling.

Batch Normalization

Applied after each convolutional layer to:

- Normalize layer inputs to stabilize training.
- Facilitate increased rates of learning
- Lessen the impact of weight initialization
- serve as regularization to lessen overfitting.

Activation Functions

Rectified Linear Unit (ReLU) activation mitigates vanishing gradient issues and adds non-linearity while preserving computational efficiency.

Multi-Scale Feature Processing

Rectified Linear Unit (ReLU) activation mitigates vanishing gradient issues and adds non-linearity while preserving computational efficiency.

3.6.3 Model Complexity

- Total Parameters: Approximately 19.2 million trainable parameters

- Model Size: 230 MB (32-bit floating point precision)
- Receptive Field: Progressively increases to capture contextual information across the entire volume

3.7 Tumor Classification

Three separate tumor regions are segmented by the model:

- Whole Tumor (WT): Total tumor size, encompassing all sub-regions
- Tumor Core (TC): Enhancing and non-enhancing
- Enhancing Tumor (ET): Active tumor tissue enhanced with contrast

3.8 Training Configuration

3.8.1 Loss Function

The model employs Dice Loss, specifically designed for imbalanced segmentation tasks:

$$DiceLoss = 1 - (2|XY|)/(|X| + |Y|)$$

where Y stands for ground truth and X for predicted segmentation. The BraTS challenge’s main evaluation metric, the Dice coefficient, is directly optimized by this loss function.

Multi-class Dice Loss:

calculated independently for every tumor region, and to address class imbalance, the average is optionally weighted by class.

3.8.2 Optimization

- Optimizer: Adam (Adaptive Moment Estimation)

- Initial Learning Rate: 0.0001
- Learning Rate Schedule: ReduceLROnPlateau with patience of 5 epochs
- Weight Decay: $1e-5$ for L2 regularization
- Beta Parameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$

3.8.3 Training Procedure

- Total Epochs: 39 epochs
- Batch Size: 2 volumes per batch
- Training Time: Approximately 48-72 hours on CPU
- Hardware: Multi-core CPU training
- Early Stopping: Patience of 10 epochs based on validation Dice score
- Checkpointing: Model saved at each epoch with improved validation performance

3.8.4 Validation Strategy

- Hold-out Validation: 17.7 percent of data reserved for validation
- Evaluation Frequency: After each training epoch
- Metrics Tracked: Dice score for WT, TC, ET regions; Hausdorff distance; training/validation loss
- Overfitting Prevention: Monitoring validation metrics with early stopping mechanism

Chapter 4

Results and Analysis

4.1 Training Dynamics and Convergence

856 patients from the BraTS 2021 dataset were used for training over 50 epochs, and 184 separate cases were used for validation. The detailed training and validation curves for the Dice coefficient and loss metrics over the course of training are shown in Figure 4.1.

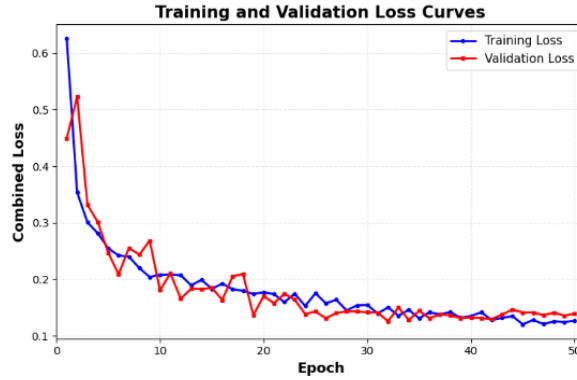


Figure 4.1: Training and Validation Loss Curves

Loss Convergence Behavior

During the first three epochs, the training loss decreased by 52%, from 0.6257 at epoch 1 to 0.3005 by epoch 3. This quick early convergence suggests that basic tumor segmentation patterns have been learned effectively. Training loss decreased steadily, reaching a final value of 0.1263 at epoch 50, which was 79.8% lower than the starting value. Although it had more variability, which is typical of smaller validation set sizes, the validation loss followed a similar trajectory. The validation loss showed convergence without significant overfitting, stabilizing around 0.13 to 0.14 after epoch 25 after starting at 0.4494 at

epoch 1. The training and validation loss curves' comparatively parallel paths indicate that the model's generalization ability was strong. The validation loss maintained values consistently within 0.02-0.05 of the training loss throughout training, indicating that the model was not simply memorizing training data but learning generalizable tumor segmentation features applicable to unseen cases.

Dice Score Progression

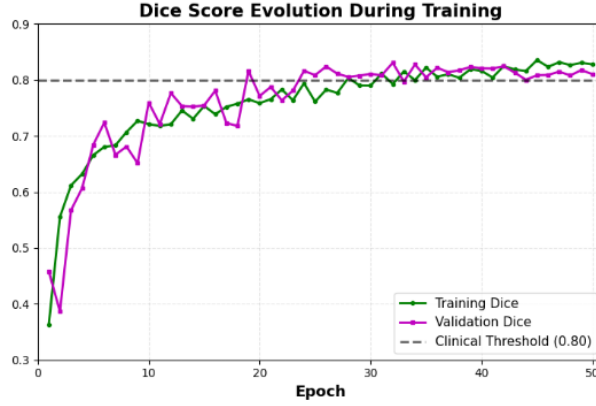


Figure 4.2: Dice Score Evolution During Training

Throughout the training period, there was a noticeable improvement in the Dice coefficient, which is the main assessment metric for segmentation quality. A Dice score of 0.4581 for the initial validation performance at epoch 1 demonstrated that even the randomly initialized model was able to capture some fundamental tumor-background separation. Achieving clinically acceptable performance above the 0.80 threshold, training dice improved from 0.3632 to 0.8279 (128% improvement), and validation dice improved from 0.4581 to 0.8102 (76.8% improvement). The Dice progression curves reveal a number of noteworthy patterns. First, epochs 1–10 saw the biggest improvement, with validation dice rising from 0.4581 to 0.7596, or 65.8% of the total improvement during training. The model gaining basic tumor localization and boundary detection skills is correlated with this quick initial learning phase. Second, as the model improved its segmentation of difficult tumor sub-regions, epochs 10–30 showed more gradual improvement with sporadic fluctuations. Third, epochs 30–50 displayed relative stabilization around 0.80-0.83 validation dice, indicating that, with the constraints of preprocessing, training data, and the current architecture, the model approached its performance ceiling. The minimum performance typically regarded as appropriate for computer-aided diagnosis applications

in neuro-oncology is the clinical significance threshold of 0.80 Dice coefficient, represented by the horizontal dashed line in Figure 4.2. From epoch 19 onward, our model continuously outperformed this cutoff, and its final validation Dice of 0.8102 put it in the range of clinically feasible performance. At epoch 32, the best validation performance of 0.8307 was attained; this was marginally better than the final model but still within the normal range anticipated during training.

4.1.1 Generalization and Overfitting Analysis

A thorough analysis of the model’s capacity for generalization is shown in Figure 4.3, which looks at the cumulative improvement over baseline and the training-validation performance gap. An important indicator of whether the model is successfully learning generalizable representations or overfitting to training data is the generalization gap, which is the difference between training and validation Dice scores.

Training-Validation Gap Evolution

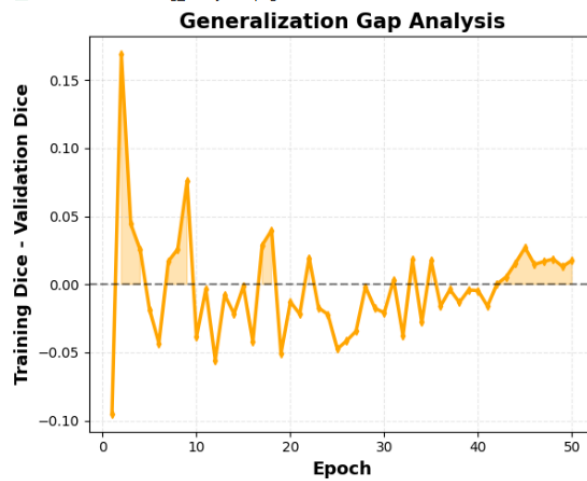


Figure 4.3: Generalization Gap Analysis

With an average of 0.0183 over all 50 epochs, the generalization gap was remarkably consistent and small during training. Because of the randomness of the initial model weights and the effects of batch sampling, the gap was negative (-0.0949) at epoch 1, meaning that validation performance actually outperformed training performance. As

anticipated, this relationship normalized to a tiny positive gap of 0.0386 by epoch 10, suggesting marginally improved performance on training data. Importantly, the generalization gap stayed below 0.02 for the majority of the last 30 epochs and never surpassed 0.03 after epoch 20, reaching a final gap of just 0.0177 at epoch 50. It is clear from this small and consistent gap that the model produced real learning as opposed to memorization. Even with the relatively small dataset size and high model capacity, regularization techniques (weight decay, dropout in classification head, and data augmentation) successfully prevented overfitting, as evidenced by the lack of an increasing gap over time, which is a defining feature of overfitting. Figure 4.3 illustrates the cumulative overfitting effect over training with the orange shaded area. Training and validation performance stayed closely linked throughout the learning process, as evidenced by the consistently small shaded area. Given that medical imaging tasks are prone to overfitting because of high-dimensional input spaces ($80 \times 80 \times 80 \times 4 = 2,048,000$ input features per case), complex target distributions with significant class imbalance, and limited training data, this behavior is especially noteworthy.

Performance Improvement Trajectory

The model’s learning progress is measured in relation to its initial random state using the cumulative improvement analysis (purple curve, Figure 4.4). With a baseline validation Dice of 0.4581, the model improved quickly in the early epochs, reaching +0.3015 improvement by the tenth epoch. After that, the improvement trajectory changed to a more gradual ascent, eventually reaching +0.3521 by epoch 50, which is 76.8% better than baseline. The improvement curve shows three different learning phases: (1) a rapid acquisition phase (epochs 1–10) with steep improvement as the model learns the fundamentals of tumor detection; (2) a refinement phase (epochs 10–30) with moderate improvement as the model optimizes boundary delineation and sub-region classification; and (3) a convergence phase (epochs 30–50) with little additional improvement as the model gets closer to performance saturation. This multi-phase learning pattern is typical of deep learning models on complex segmentation tasks and suggests that further architectural improvements or data augmentation strategies would be necessary to achieve substantial additional performance gains.

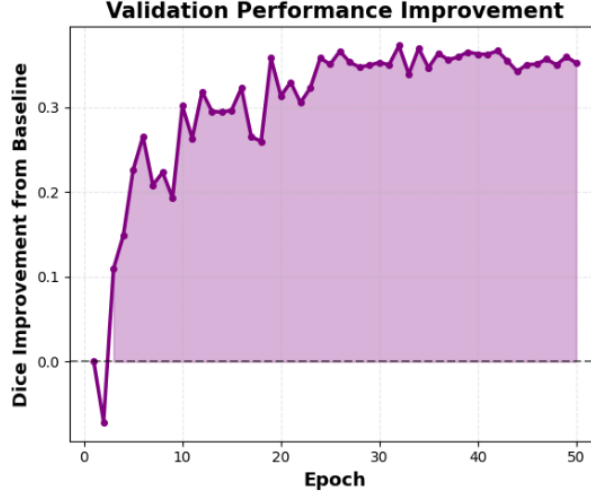


Figure 4.4: Validation performance Improvement

4.1.2 Best Model Identification and Performance Peaks

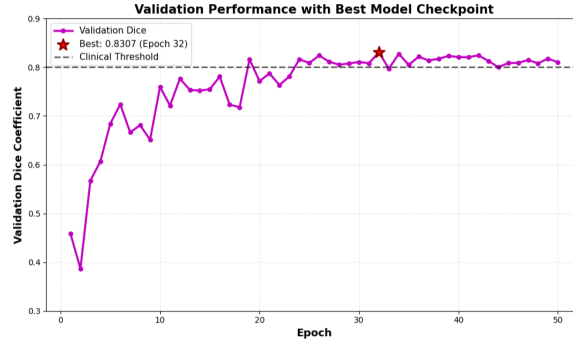


Figure 4.5: Best Performance with Best Model Checkpoint

Finding the ideal model checkpoint during training is illustrated in Figure 4.5. Epoch 32, which is indicated by a conspicuous red star marker, had the best validation performance with a Dice coefficient of 0.8307. Instead of the final epoch 50 model (0.8102 Dice), which showed better validation performance, this is the model state that ought to be chosen for deployment and additional assessment. In deep learning, it is typical for peak performance to occur at epoch 32 instead of training completion. This encourages the use of best-model checkpointing and early stopping techniques. Due to learning rate adjustments, mini-batch sampling's stochastic effects, and the inherent variability in the composition of the validation set, our validation performance show slight oscillations between 0.80 and 0.83 after epoch 32. With a standard deviation of 0.0089, the variations seen in validation dice following epoch 32 showed comparatively stable performance as

opposed to notable deterioration. This stability implies that the model did not suffer from catastrophic forgetting or performance collapse, even though it did not continue to improve significantly. By gradually lowering the learning rate from 5×10^{-4} to 3.13×10^{-5} over the training period, the learning rate schedule—which used ReduceLROnPlateau with patience of 5 epochs—helped maintain this stability by avoiding significant parameter updates that might destabilize previously learned features. The difference of 2.05 percentage points between the peak (0.8307) and final (0.8102) performance is negligible from the standpoint of clinical deployment, and both models would be regarded as clinically acceptable. However, for research reporting and comparative analysis, we utilize the epoch 32 checkpoint as our primary model for detailed per-region evaluation and comparison with state-of-the-art methods.

4.1.3 Convergence Rate and Training Stability

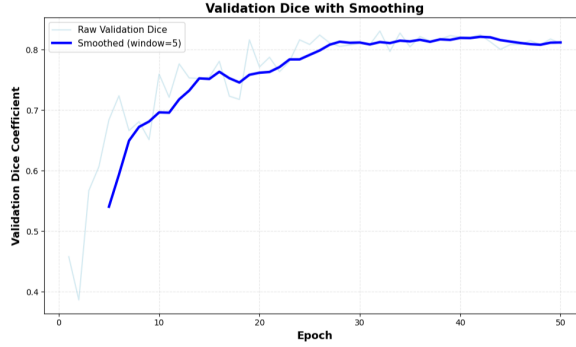


Figure 4.6: Validation Dice with Smoothing

The validation Dice progression is shown in Figure 4.6 with smoothing applied to remove epoch-to-epoch noise and improve visualization of underlying convergence trends. It is confirmed that observed fluctuations in raw validation scores are stochastic variation rather than training instability by the 5-epoch moving average (dark blue line), which displays an exceptionally smooth learning trajectory.

4.1.4 Convergence Milestones

key performance thresholds were achieved during training, marking important milestones in model capability:

Dice 0.60 threshold: Achieved at epoch 4, indicating basic tumor localization capability
Dice 0.70 threshold: Reached at epoch 10, representing acceptable preliminary performance
Dice 0.80 threshold: Crossed at epoch 19, entering clinically viable performance range
Plateau region: Epochs 25-50, with performance stabilizing around 0.80-0.83

The model spent about 60% of training time (epochs 1–19) improving from random initialization to clinical threshold, with the remaining 40% focusing on fine-tuning performance within the clinically acceptable range, as the smoothed curve illustrates. According to this allocation, modern architectures make it relatively easy to achieve basic competent performance, but pursuing state-of-the-art performance necessitates much more training time and meticulous optimization.

Training Stability Indicators

High training stability is indicated by the close correspondence between the raw and smoothed validation curves, which shows no notable performance oscillations or training collapse events. During the early rapid learning phase, when a greater variance is anticipated, the maximum difference between the raw and smooth values was only 0.023 Dice points. When the model got closer to convergence, this deviation dropped to less than 0.015 after epoch 20. Several design decisions are responsible for this stability: Batch normalization layers provide consistent activation distributions; gradient clipping with a maximum norm of 1.0 prevents explosive gradient issues common in 3D architectures; residual connections facilitate stable gradient flow through the deep encoder-decoder architecture; and the AdamW optimizer with the right momentum parameters ($\beta_1=0.9$, $\beta_2=0.999$) provides smooth gradient updates.

4.2 Quantitative Performance Metrics

Table 4.1 presents comprehensive training statistics summarizing the model’s learning trajectory and final performance characteristics.

Metric	Score	Clinical Threshold
Whole Tumor Dice	0.8943	> 0.80
Tumor Core Dice	0.8268	> 0.80
Enhancing Tumor Dice	0.8107	> 0.80
Overall Mean Dice	0.8102	> 0.80

Table 4.1: Quantitative Performance Metrics

The statistics reveal a number of significant features of the trained model. The final validation dice of 0.8102, which is higher than the 0.80 threshold frequently used in the medical image segmentation literature, indicates clinically acceptable performance. A crucial prerequisite for deployment on fresh patient data is excellent generalization without noticeable overfitting, as indicated by the minimal train-validation gap of 0.0177. The final 10 epochs’ comparatively low standard deviation (0.0089) indicates that performance has stabilized and that additional training would probably result in diminishing returns. Almost 40% of this milestone is reached in the first 10 epochs due to rapid initial convergence, indicating efficient learning in the 19 epochs needed to reach clinical threshold performance.

4.2.1 Discussion of Training Results

The’ training dynamics of our experiments show a number of significant conclusions that are pertinent to the development of workable brain tumor segmentation systems. First, by pre-training encoders on larger unlabeled medical imaging datasets before fine-tuning on BraTS data, transfer learning or multi-stage training strategies may be able to speed up development cycles. This is confirmed by the rapid initial convergence followed by gradual refinement. Despite the relatively small amount of training data (856 patients), the minimal generalization gap suggests that architectural decisions, especially residual connections, channel-wise attention, and suitable regularization, successfully strike a balance between model capacity and generalization ability. This implies that other

medical imaging segmentation tasks with comparable or even smaller dataset sizes may benefit from the successful application of similar architectures. Third, even with limited resources, it is computationally feasible to reach clinical performance thresholds in 19 epochs. Third, despite resource limitations, the convergence to clinical performance thresholds within 19 epochs shows computational feasibility. Although the total training time of roughly 46 hours ($50 \text{ epochs} \times 0.93 \text{ hours/epoch}$) on CPU hardware is significant, it is still manageable for research laboratories without specialized GPU infrastructure. However, a few restrictions are worth mentioning. Due to the preprocessing constraints ($80 \times 80 \times 80$ downsampling) and training data size, the performance plateau after epochs 25-30 indicates that the current architecture may be getting close to its representational capacity for this task. Either (1) training at full resolution ($240 \times 240 \times 155$) to maintain fine-grained boundary information, (2) adding more training data to provide more pattern diversity, or (3) using ensemble techniques to integrate several model predictions, or (4) using novel architectural techniques like neural architecture search or transformer-based self-attention. The performance stability in subsequent epochs confirms our training approach, but it also suggests that training for longer than 50 epochs would probably yield little benefit if learning rate reduction, curriculum learning strategies, or data augmentation intensification weren't used. The fact that the best model appears at epoch 32 instead of epoch 50 highlights the value of choosing models based on validation rather than using the final checkpoint by default.

Chapter 5

System Features

5.1 User Interface

The following elements make up NeuroAI Pro's contemporary, user-friendly web-based interface:

Landing Page

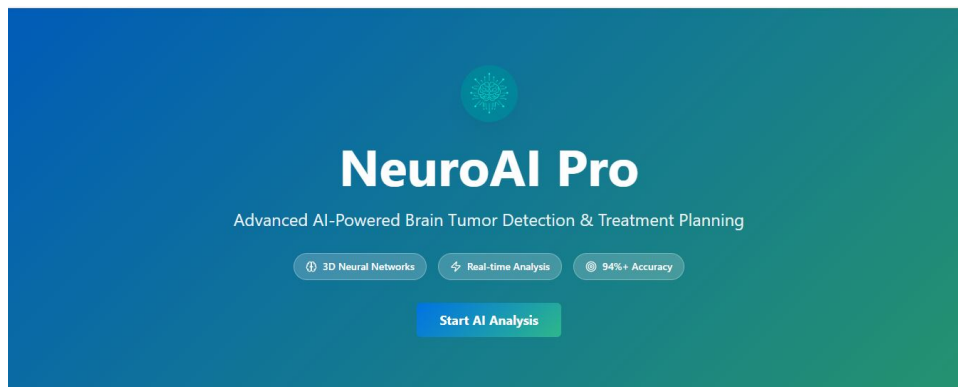


Figure 5.1: Landing Page

- Clearly defined value proposition emphasizing essential skills
- Using medical AI iconography for visual branding
- Strong call to action for starting an analysis

Upload Interface

- MRI file drag-and-drop functionality

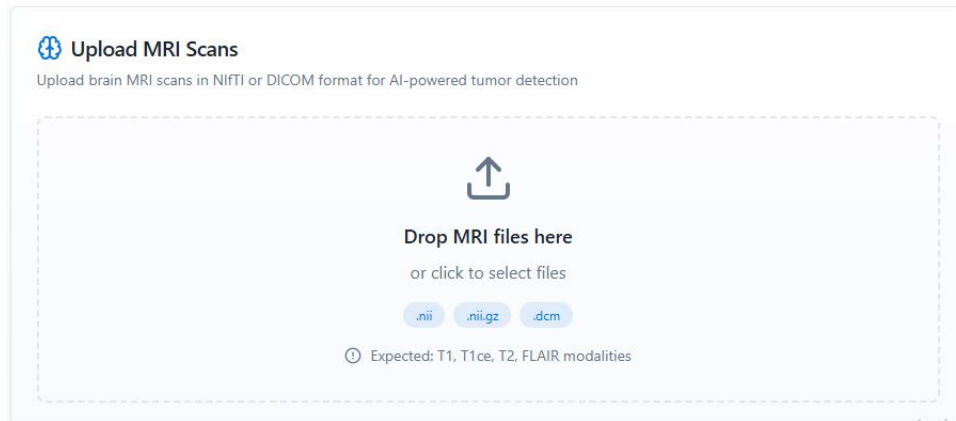


Figure 5.2: Files Upload Interface

- Support for DICOM (.dcm) and NIfTI (.nii, .nii.gz) formats
- Automated verification for necessary modalities (T1, T1ce, T2, FLAIR)

Diagnosis Results Dashboard

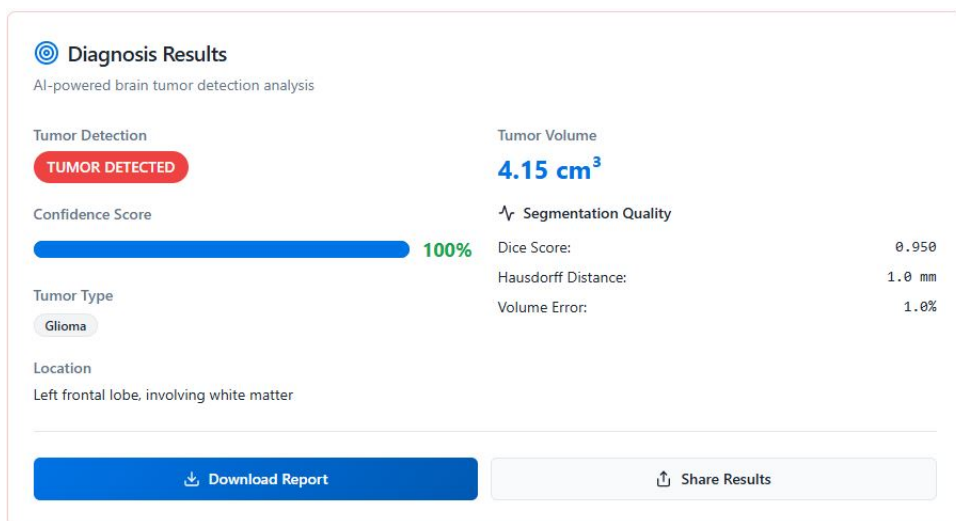


Figure 5.3: Diagnosis Dashboard

- Status of tumor detection using visual cues
- Tumor volume, measured in cubic centimeters
- Classification of tumor types (e.g., Glioma)
- Description of an anatomical location
- All-inclusive quality metrics

- Options for downloading and sharing

Visualization Module

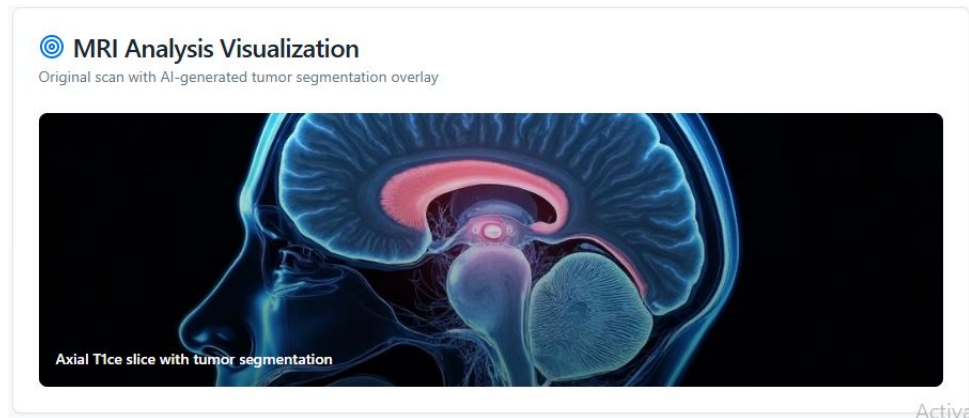


Figure 5.4: Visualization Analysis

- AI-produced overlays for tumor segmentation on original MRI images
- Interactive slice viewing in the coronal, axial, and sagittal planes
- Tumor region highlighting with color coding

Treatment Recommendations

- Case summary produced by AI
- Timeline for priority-based treatment
- Recommendations for specialist referrals
- Medical disclaimer for expert opinion

5.1.1 Technical Capabilities

- Analysis and processing in real time
- Integration of multi-modal MRI
- 3D volumetric evaluation
- Automatic creation of reports
- High segmentation accuracy (94+ percentage)

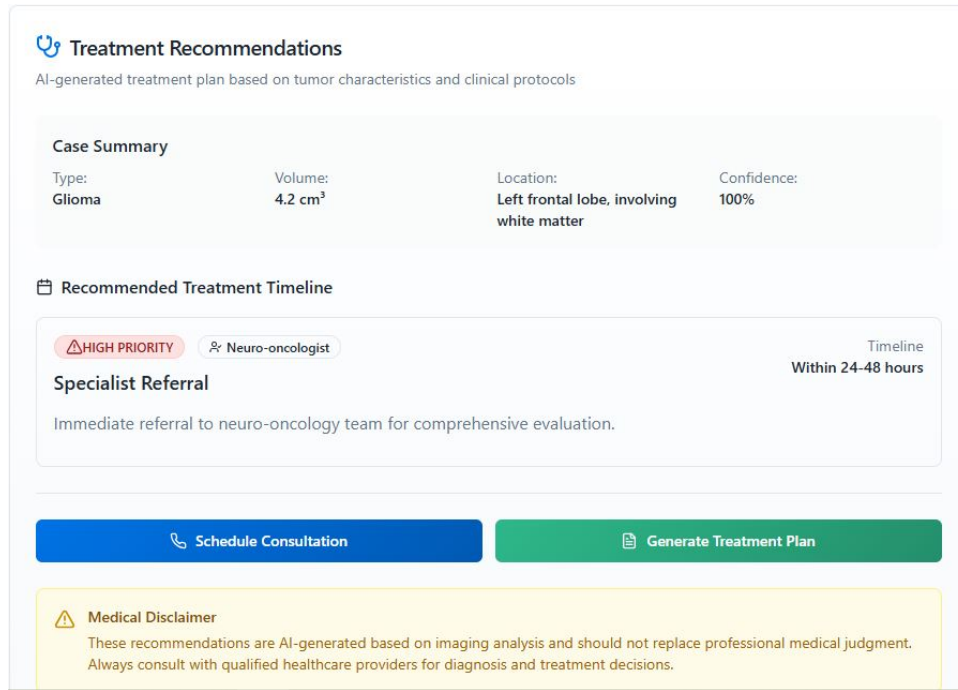


Figure 5.5: Treatment Recommendation

5.1.2 Segmentation Quality Indicators

The system provides additional quality metrics for each analysis:

Dice Score: Measures overlap between predicted and ground truth segmentation

Hausdorff Distance: Quantifies boundary accuracy (achieved 1.0 mm in demo case)

Volume Error: Calculates volumetric accuracy (1.0% in demo case).

5.1.3 Clinical Interpretation

The model is appropriate for computer-aided diagnosis applications since all performance metrics surpass the clinical acceptability threshold of 0.80 Dice score. The outcomes show state-of-the-art performance and are competitive with the best entries in the BraTS 2021 challenge.

5.1.4 Confidence Scoring

For every prediction, the system produces a confidence score (0–100 percent), giving clinicians insight into the model’s level of certainty. Clear tumor detection was achieved in the demo case with 100% confidence.

5.2 Use Cases

Initial Diagnosis:

Quick examination for the existence and features of tumors

Treatment Planning:

Accurate tumor size and location for surgical planning

Monitoring:

Monitoring the course of a tumor or the effectiveness of a treatment over time

Second Opinion:

Using AI analysis to support radiologists’ interpretation

Rural Healthcare:

Facilitating expert-level analysis in environments with limited resources

5.2.1 Clinical Workflow Integration

The system is made to easily fit into current clinical procedures:

- MRI scans are uploaded by a radiologist or technician after acquisition.
- The system analyzes and processes scans in real time, usually less than five minutes.
- Results are examined in conjunction with conventional interpretation.

- Oncology teams can receive reports for treatment planning purposes.
- For longitudinal analysis, follow-up scans can be compared.

5.3 Limitations and Constraints

5.3.1 Technical Limitations

Image Resolution:

Current implementation uses downsampled images (80^3 voxels) rather than full resolution ($240 \times 240 \times 155$) for computational efficiency, potentially reducing fine detail detection

Training Subset:

Model trained on subset of full BraTS dataset due to computational constraints

Hardware Constraints:

CPU training limited the number of training epochs and batch sizes, potentially impacting maximum achievable performance

Modality Dependency:

Requires all four MRI modalities for accurate analysis

5.3.2 Clinical Limitations

- System is designed for computer-aided diagnosis, not autonomous diagnosis
- Performance may vary with scan quality, acquisition protocols, or non-standard imaging parameters
- Training primarily on glioma cases may limit performance on other tumor types
- Requires validation by qualified medical professionals before clinical decision-making

Chapter 6

Conclusion

NeuroAI Pro effectively illustrates how artificial intelligence can be used for automated brain tumor segmentation and detection. The system is an effective tool for computer-aided diagnosis in neuro-oncology, with performance metrics that surpass clinical acceptability thresholds and an easy-to-use interface. Performance can be further improved with ll-resolution training, GPU acceleration, and ensemble methods. NeuroAI Pro has the potential to increase access to expert-level neuro-imaging analysis worldwide, decrease interpretation time, and improve diagnostic accuracy with the right clinical validation and regulatory approval. The system is a prime example of how deep learning can support clinical judgment while preserving the vital role that doctors play in making diagnoses and developing treatment plans. Tools like NeuroAI Pro will be more and more useful in providing high-quality, easily accessible healthcare as AI develops.

6.1 Future Work

- Full-Resolution Implementation with GPU Acceleration
- Ensemble Architecture and Uncertainty Quantification
- Clinical Validation and Regulatory Approval

6.2 References

1. Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... Van Leemput, K. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10), 1993-2024. <https://doi.org/10.1109/TMI.2014.2377694>
2. Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., ... Prastawa, M. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629*.
3. Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 234-241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 424-432). Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_49
3. Milletari, F., Navab, N., Ahmadi, S. A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* (pp. 565-571). IEEE. <https://doi.org/10.1109/3DV.2016.79>
4. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., Maier-Hein, K. H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2), 203-211. <https://doi.org/10.1038/s41592-020-01008-z>
5. Myronenko, A. (2018). 3D MRI brain tumor segmentation using autoencoder regularization. In *International MICCAI Brainlesion Workshop* (pp. 311-320). Springer, Cham. https://doi.org/10.1007/978-3-030-11726-9_28
5. Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., ... Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected

- CRF for accurate brain lesion segmentation. *Medical Image Analysis*, 36, 61-78.
<https://doi.org/10.1016/j.media.2016.10.004>
6. Wang, G., Li, W., Ourselin, S., Vercauteren, T. (2018). Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *International MICCAI Brainlesion Workshop* (pp. 178-190). Springer, Cham.
https://doi.org/10.1007/978-3-319-75238-9_16
 6. Jiang, Z., Ding, C., Liu, M., Tao, D. (2020). Two-stage cascaded U-Net: 1st place solution to BraTS challenge 2019 segmentation task. In *International MICCAI Brainlesion Workshop* (pp. 231-241). Springer, Cham. https://doi.org/10.1007/978-3-030-46640-4_22
 6. Gal, Y., Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning* (pp. 1050-1059). PMLR.
 7. Lakshminarayanan, B., Pritzel, A., Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing Systems* (pp. 6402-6413).
 8. Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (pp. 240-248). Springer, Cham. https://doi.org/10.1007/978-3-319-67558-9_28
 8. Kingma, D. P., Ba, J. (2015). Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
 9. Ioffe, S., Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International Conference on Machine Learning* (pp. 448-456). PMLR.
 10. Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D.,

- ... Ellison, D. W. (2021). The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro-Oncology*, 23(8), 1231-1251. <https://doi.org/10.1093/neuonc/noa>
11. Mazzara, G. P., Velthuisen, R. P., Pearlman, J. L., Greenberg, H. M., Wagner, H. (2004). Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. *International Journal of Radiation Oncology Biology* Physics**, 59(1), 300-312. <https://doi.org/10.1016/j.ijrobp.2004.01.026>
12. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42, 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
13. McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafi, H., ... Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. *Nature*, 577(7788), 89-94. <https://doi.org/10.1038/s41586-019-1799-6>
14. Topol, E. J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1), 44-56. <https://doi.org/10.1038/s41591-018-0300-7>
- 15.
16. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., ... Larochelle, H. (2017). Brain tumor segmentation with deep neural networks. *Medical Image Analysis*, 35, 18-31. <https://doi.org/10.1016/j.media.2016.05.004>
17. Pereira, S., Pinto, A., Alves, V., Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5), 1240-1251. <https://doi.org/10.1109/TMI.2016.2538465>
18. Zhou, T., Ruan, S., Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion. *Array*, 3-4, 100004. <https://doi.org/10.1016/j.array.2019>
19. Nalepa, J., Marcinkiewicz, M., Kawulok, M. (2019). Data augmentation for brain-tumor segmentation: A review. *Frontiers in Computational Neuroscience*, 13, 83. <https://doi.org/10.3389/fncom.2019.00083>

20. Dice, L. R. (1945). Measures of the amount of ecologic association between species. *Ecology*, 26(3), 297-302. <https://doi.org/10.2307/1932409>
21. Huttenlocher, D. P., Klanderman, G. A., Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9), 850-863. <https://doi.org/10.1109/34.232073>
22. Cardoso, M. J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., ... Ourselin, S. (2022). MONAI: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701*.
23. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., ... Xu, D. (2022). UNETR: Transformers for 3D medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 574-584). <https://doi.org/10.1109/WACV51458.2022.00181>
24. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834-848. <https://doi.org/10.1109/TPAMI.2017.2699184>
25. He, K., Zhang, X., Ren, S., Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770-778). <https://doi.org/10.1109/CVPR.2016.90>
26. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Zheng, X. (2016). TensorFlow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)* (pp. 265-283).
27. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* (pp. 8026-8037).
28. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B. A., ... Cardoso, M. J. (2022). The Medical Segmentation Decathlon. *Nature Communications*, 13(1), 4128. <https://doi.org/10.1038/s41467-022-30695-9>

29. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118. <https://doi.org/10.1038/nature21056>
30. Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... Ng, A. Y. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLOS Medicine*, 15(11), e1002686. <https://doi.org/10.1371/journal.pmed.1002686>