Three-way Decision for Segmentation of Brain Tumor in MRI

Submitted by

Noman Arif
21i-2385

Supervised by

Dr. Zeshan Khan

Masters of Science (Software Engineering)

A thesis submitted in partial fulfillment of the requirements for the degree of
Masters of Science (Software Engineering)
at National University of Computer & Emerging Sciences

Department of Software Engineering
National University of Computer & Emerging Sciences

Islamabad, Pakistan.

August 2025

Plagiarism Undertaking

I take full responsibility of the research work conducted during the Masters Thesis titled *Three-way Decision for Segmentation of Brain Tumor in MRI* . I solemnly declare that the research work presented in the thesis is done solely by me with no significant help from any other person; however, small help wherever taken is duly acknowledged. I have also written the complete thesis by myself. Moreover, I have not presented this thesis (or substantially similar research work) or any part of the thesis previously to any other degree awarding institution within Pakistan or abroad.

I understand that the management of National University of Computer and Emerging Sciences has a zero tolerance policy towards plagiarism. Therefore, I as an author of the above-mentioned thesis, solemnly declare that no portion of my thesis has been plagiarized and any material used in the thesis from other sources is properly referenced. Moreover, the thesis does not contain any literal citing of more than 70 words (total) even by giving a reference unless I have the written permission of the publisher to do so. Furthermore, the work presented in the thesis is my own original work and I have positively cited the related work of the other researchers by clearly differentiating my work from their relevant work.

I further understand that if I am found guilty of any form of plagiarism in my thesis work even after my graduation, the University reserves the right to revoke my Masters degree. Moreover, the University will also have the right to publish my name on its website that keeps a record of the students who plagiarized in their thesis work.

____

Noman Arif

Date: ____________________

Author's Declaration

I, Noman Arif, hereby state that my Masters thesis titled *Three-way Decision for Segmentation of Brain Tumor in MRI* is my own work and it has not been previously submitted by me for taking partial or full credit for the award of any degree at this University or anywhere else in the world. If my statement is found to be incorrect, at any time even after my graduation, the University has the right to revoke my Masters degree.

I, Noman Arif, hereby declare that my Masters thesis submitted is my own, original work, completed independently without any unauthorized assistance. Only the sources and resources explicitly cited have been utilized. All excerpts, whether quoted directly or paraphrased, have been properly acknowledged.

I confirm that:

■ I have not used any generative AI tools

OR

□ I have used generative AI tools. I have clearly indicated the product name, manufacturer, software version, and the specific purpose (e.g., grammar checking, language enhancement, systematic research). I take full responsibility for the selection, application, and interpretation of any AI-generated content incorporated into this work. List of prompts along with corresponding page number where they are used is listed in the appendix."

____

Noman Arif

Date: ____________________

Certificate of Approval

It is certified that the research work presented in this thesis, entitled "Three-way Decision for Segmentation of Brain Tumor in MRI " was conducted by Noman Arif under the supervision of Dr. Zeshan Khan.

No part of this thesis has been submitted anywhere else for any other degree.

This thesis is submitted to the Department of Software Engineering in partial fulfillment of the requirements for the degree of Masters of Science in Software Engineering

at the

National University of Computer and Emerging Sciences, Islamabad, Pakistan

August' 2025

Candidate Name: Noman Arif Signature: ______________________

Examination Committee:

1. Name: Dr. M. Ehtisham Hassan Signature: ______________________
   Assistant Professor, CUST Islamabad.

2. Name: Dr. Basharat Hussain Signature: ______________________
   Assistant Professor, FAST-NU Islamabad

Dr. ______________________________________
Graduate Program Coordinator, National University of Computer and Emerging Sciences, Islamabad, Pakistan.

Dr. Usman Habib ______________________________________
Head of the Department of Software Engineering, National University of Computer and Emerging Sciences, Islamabad, Pakistan.

Abstract

Brain tumors arise when abnormal cell growth occurs within brain tissue, often leading to poor patient outcomes, with survival typically ranging from several months to just over a year under intensive treatment [1]. Timely detection remains difficult, as accurate identification of tumor boundaries and characteristics demands sophisticated imaging methods and expert interpretation. The BraTS challenge provides standardized MRI datasets to support the development of automated segmentation tools. Among existing approaches, U-Net has become a widely adopted architecture due to its encoder–decoder design with skip connections. Recent advances in deep learning and vision transformers have further expanded possibilities, enabling models to capture both local and global context. In this work, we propose a hybrid ensemble framework that integrates supervised and unsupervised learning with transformer-based modules, aiming to surpass current state-of-the-art results on the BraTS 2024 dataset for post-treatment adult glioma segmentation.

Acknowledgements

I express my deepest gratitude to Almighty Allah, the Most Beneficial and the Most Merciful, for His blessings, and the strength bestowed upon me throughout my academic journey. His divine grace has been the source of inspiration and perseverance, allowing me to overcome challenges and achieve my goals.

I extend my heartfelt appreciation to my supervisor, Dr. Zeshan Khan, whose invaluable guidance, mentorship, and encouragement have played a pivotal role in shaping my research and academic endeavors. His expertise, patience, and dedication have been a guiding light, and I am truly fortunate to have had the opportunity to work under his mentorship.

I am profoundly grateful to my dear friend, Ijaz Hussain, for his indispensable support in this journey. His moral support and appreciation have always motivated me. Whenever I felt down and demotivated, he supported me and made me feel that I could do it.

My gratitude extends to my family for their unending support, love, and understanding. Their encouragement has been a driving force, providing me with the motivation to persevere in the face of challenges.

Dedication

Dedicated to my Abbu and Ammi, who gave it all they had to make me who I am today, I am forever grateful to your love and support in every decision of my life.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	**Full Form**
MRI	Magnetic Resonance Imaging
CNN	Convolutional Neural Network
DL	Deep Learning
DSC	Dice Similarity Coefficient
IoU	Intersection over Union
HD95	95th Percentile Hausdorff Distance
ViT	Vision Transformer
3WC	Three-Way Clustering
BraTS	Brain Tumor Segmentation
AI	Artificial Intelligence
ML	Machine Learning
GPU	Graphics Processing Unit
CPU	Central Processing Unit
DICOM	Digital Imaging and Communications in Medicine
FLAIR	Fluid Attenuated Inversion Recovery
PET	Positron Emission Tomography
ROI	Region of Interest
U-Net	U-shaped Convolutional Neural Network architecture

Chapter 1

Introduction

Segmenting medical images plays a vital role in computer-assisted diagnosis and treatment planning and a critical part in the computer-aided diagnosis paradigm. Among segmentation models, U-Net has become widely adopted due to its encoder-decoder design, flexibility, optimized modular design, and success in all medical image modalities. Over the years, the U-Net model achieved tremendous attention from industrial and academic researchers [2]. Image segmentation is a task in the field of medical imaging and specifically image processing, defined as the division of the entire image into a set of regions, helps in analyzing an image in different ways in a wide range of applications. Medical image segmentation is a crucial example of this domain and offers various benefits for clinical uses.

In Medical Imaging, MRI's, CT scans and X-rays are the most commonly used data to detect and diagnose various diseases like tumor in brain, lungs, kidney etc. BraTS (Brain Tumor Segmentation) is a Segmentation challenge came up every year with new, improved and diverse dataset. Numerous models with competing accuracies and efficiency were proposed in the state-of-the-art with the most popular architecture of U-net that outperformed for the tasks of semantic segmentation.

Semantic segmentation transforms raw medical images into structured, clinically relevant information by delineating regions of interest (ROI) such as organs, lesions, or tissues [3]. This process is indispensable for numerous clinical applications, including radiotherapy planning. Indeed, segmentation has become the most extensively studied task in biomedical image analysis, accounting for nearly 70% of related challenges [4]. Automated segmentation not only accelerates data processing but also provides clinicians with task-specific visualizations and quantitative measurements. These outputs assist practitioners in monitoring disease progression, particularly in cancer treatment, and enhance visualization quality by modeling specific regions relevant to the task at hand (e.g., cardiac or brain segmentation) [2, 5].

Beyond semantic segmentation, image segmentation techniques in digital image processing often involve partitioning an image into multiple regions based on pixel characteristics such as intensity, color, or shape. Applications in medical imaging include identifying and labeling tumor regions in 2D images or 3D volumes. Clustering methods are frequently employed to group similar tissue types, enabling detection and prediction of diseases such as brain tumors, cardiac disorders, bone fractures, and tissue abnormalities using MRI, CT, or X-ray data.

A more recent advancement is Three-Way Clustering (3WC), which addresses uncertainty and incomplete information in medical data [6]. This approach defines two sets for each cluster—the core and the support—corresponding to lower and upper bounds. These sets divide the data into three regions: inside, outside, and partial. Objects in the inside region clearly belong to the cluster, representing its dense core. Those in the outside region are excluded, while objects in the partial region may or may not belong, reflecting vague or sparse boundaries [7]. Such a framework offers a promising way to manage uncertainty in medical imaging, making it particularly relevant for complex tasks like brain tumor segmentation.

1.1 Motivation

The main motivation for this research is that the Three-Way is a modern approach under the umbrella of Three-way Decision theory and Granular Computing, which is more efficient and effective because it can handle the uncertainty in the data and there is still a gap to implement this approach in real-life applications and data to get the desired results with more precision because the Three-Way clustering approach has gained a significant result in the state-of-the-art [8]. As it falls in the domain of image processing, so I decide to implement it in Medical Imaging dataset. In medical imaging, magnetic resonance imaging, CT scans, and X-rays are the most widely used data to detect and diagnose various diseases such as brain tumor. BraTS is a Segmentation challenge that comes up every year with a new, improved, and diverse dataset. Numerous models with competing accuracies and efficiency were proposed in the state of the art with the most popular U-net architecture that outperformed for semantic segmentation tasks

Three-Way Clustering (3WC) has emerged as a contemporary approach for handling data uncertainty more effectively. Rooted in the principles of three-way decision theory, this method introduces two defining sets — the core and the support — to characterize each cluster. These sets partition the data into three distinct regions: inside, outside, and partial. The inside region represents the dense core where membership is certain, the outside region contains elements that clearly do not belong, and the partial region captures ambiguous cases with vague boundaries. This structure enables more nuanced clustering outcomes, particularly in scenarios where clusters exhibit both well-defined centers and uncertain edges. A key challenge, however, lies in the meaningful construction of these sets and the accurate delineation of the three regions to ensure reliable results.

Chapter 2

Methodology

The general framework of three-way clustering was proposed by Hong Yu in 2017 [7] and is represented as follows:

Three-way cluster C as a pair of sets:

$$C = Co(C), Fr(C) \tag{2.1}$$

Here, $Co(C) \subseteq U$ and $\mathrm{Fr}(C) \subseteq U$. Let $\mathrm{Tr}(C) = U \setminus \mathrm{Co}(C) \setminus \mathrm{Fr}(C)$. Then, $Co(C), Fr(C)$ and $Tr(C)$ naturally form the three regions of a cluster as Core, Fringe and Trivial Regions respectively. That is:

$$CoreRegion(C) = Co(C) \tag{2.2}$$

$$FringeRegion(C) = Fr(C) \tag{2.3}$$

$$TrivialRegion(C) = U - Co(C) - Fr(C) \tag{2.4}$$

If $x \in CoreRegion$, the object x belongs to the cluster C definitely;
if $x \in FringeRegion(C)$, the object x might belong to C; if $x \in TrivialRegion(C)$, then the object x does not belong to C.
These subsets have the following properties.

$$U = Co(C) \cup Fr(C) \cup Tr(C) \tag{2.5}$$

$$Co(C) \cap Fr(C) = \emptyset \tag{2.6}$$

$$Fr(C) \cap Tr(C) = \emptyset \tag{2.7}$$

$$Tr(C) \cap Co(C) = \emptyset \tag{2.8}$$

Wang et al. emphasized that the researchers can apply the concepts of three-way clustering in the application areas of machine learning in medical diagnosis . [8]

Chapter 3

Literature Review

Three-Way Clustering (3WC) specifically addresses the challenge of uncertain relationships between objects and clusters. Various approaches have been explored under this framework, including fuzzy clustering, rough clustering, interval clustering, soft clustering, and overlapping clustering. Collectively, these methods aim to manage ambiguity in data assignment. The analysis of 3WC contributes to solving two fundamental issues: first, how a cluster should be formally represented, and second, how clusters can be effectively derived from complex datasets [6].

Tang et al. [9] adapted the Kolmogorov-Arnold Network (KAN) layers into a U-Net framework, producing the UKAN-SE model. This variant incorporates Squeeze-and-Excitation blocks to emphasize informative features through channel-wise attention. Compared to standard U-Net, UKAN-SE achieved superior segmentation accuracy and reduced training time, particularly in enhancing tumor regions. The study highlights the effectiveness of lightweight models augmented with global attention mechanisms..

To addresses the issue of delineation of regions of interest in segmentation of medical imaging I-e; automated segmentation tools for adult glioma post treatment in MRI data [10]. Proposed two approaches to enhance the segmentation performances by incorporation of an additional input based on an artificial input sequence generation (T1Gd-T1) which highlights enhancing tumors and employed different ensemble methods to weigh the contribution of a battery of models. Model is based on three network architectures: 1) nnU-Net 2) nnU-Net ResEnc 3) SegResNet. It employed the default nnU-NET test-time augmentations during inference. It includes Gaussian weighting and mirroring along axis. Explored two different ways of ensemble model predictions: STAPLE and weighted average. STAPLE involves combining predictions of different models. These ensemble models effectively improved the segmentation performance. The label-wise weighted average technique outperformed the STAPLE. Experimentation shows us that T1Gd-T1 subtraction images significantly improved segmentation of enhancing tumors (ET) specifically.

The authors of [11] developed the Medical Imaging Segmentation Toolkit (MIST), designed to streamline the workflow for training and evaluating deep learning models in medical imaging.

MIST standardizes preprocessing and evaluation, ensuring consistency across different architectures and datasets. By supporting multi-GPU scalability and providing a unified pipeline, the framework enables reproducible experiments and fair comparisons between methods. Although MIST itself is not a novel segmentation model, its contribution lies in infrastructure improvements that enhance reliability and accelerate research progress.

In [12], the nnFormer architecture was introduced as a transformer-based solution for 3D volumetric segmentation. Unlike conventional CNNs, nnFormer leverages both local and global self-attention mechanisms to capture contextual information across entire volumes. The model replaces traditional skip connections with attention-based modules, improving segmentation accuracy, particularly in challenging tumor regions. This approach demonstrates the growing relevance of transformer designs in medical imaging, complementing CNN-based methods such as nnU-Net.
par Chen et al. [13]. proposed TransUNet, a hybrid model that combines CNN feature extraction with transformer-based self-attention. The encoder processes image patches into tokens, while the decoder refines segmentation through cross-attention with U-Net features. This design captures both local and global dependencies, improving performance on small and complex tumor regions. Their results demonstrate that integrating transformers into U-Net significantly enhances segmentation accuracy compared to CNN-only baselines.

Ronneberger et al. [14] first proposed the U-Net architecture in 2015, which has since become one of the most influential models in biomedical image segmentation. U-Net's encoder–decoder structure, combined with skip connections, allows precise localization even when training data is limited. By employing data augmentation strategies such as elastic deformation, the model extracts detailed features without requiring extensive annotation. Its efficiency and accuracy established U-Net as a baseline for subsequent segmentation research.

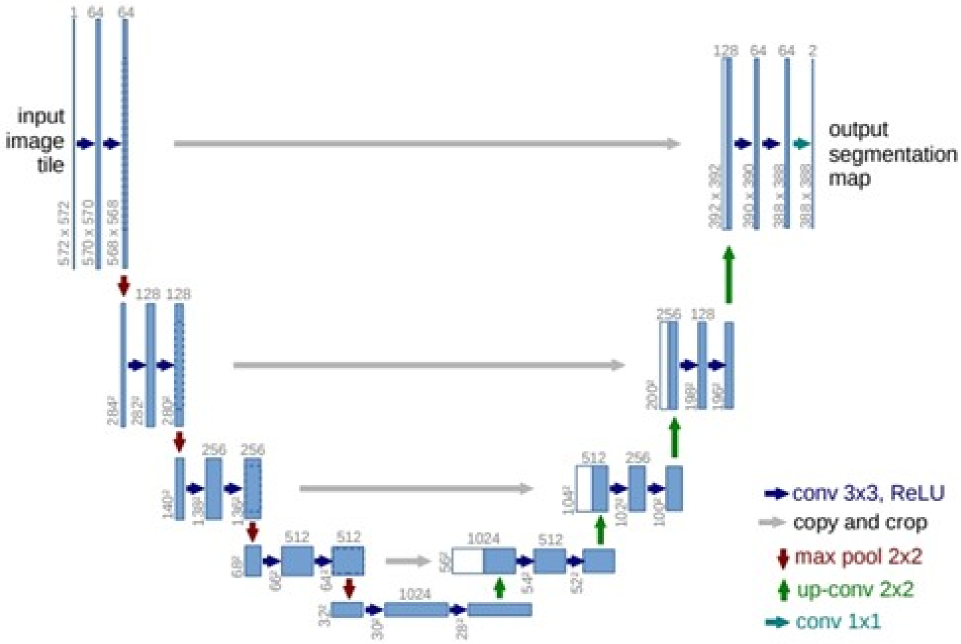

Figure 3.1: 2D U-Net Architecture

Cicek et al. [15] introduced a three-dimensional volumetric U-Net in 2016, motivated by the fact that most medical imaging modalities inherently produce 3D data. Their approach was designed to address the inefficiency of slice-by-slice annotation, which is both time-consuming and redundant since adjacent slices often contain overlapping information. By extending U-Net into a volumetric framework, the model leverages spatial continuity across slices, reducing redundancy and improving segmentation performance. The architecture diagram can be seen in

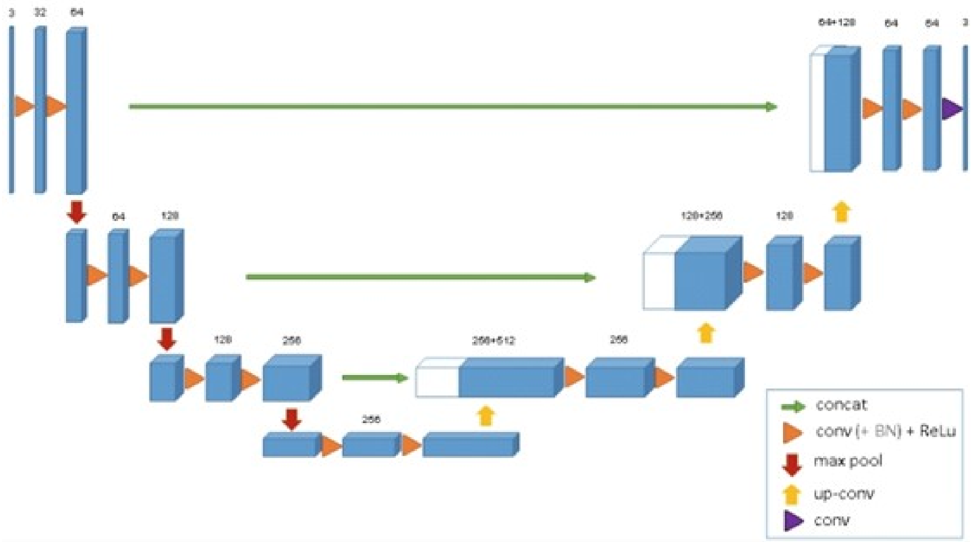

Figure 3.2: 3D U-Net Architecture

A unified 3D Swin Transformer-based segmentation framework that Introduced two variants one with CNN-based decoder and other is with Transformer-based decoder. Integrates a self-supervised pre-training framework using masked image modeling (MIM) to improve downstream segmentation performance [16]. The model surpasses in small structure segmentation (e.g. ET Enhancing Tumor). UNetFormer+ variant is better for large structures (e.g. Whole Tumor), likely due to its global attention mechanism.

Table 3.1: Summary of Related Work in Brain Tumor Segmentation

Authors (Year)	Title	Objective / Purpose	Methodology	Key Findings / Results	Gaps / Limitations	Relevance to Your Research
Ronneberger et al. (2015) [14] [15]	U-Net: Convolutional Networks for Biomedical Image Segmentation	Introduced a novel lightweight segmentation network of U-shaped especially for biomedical images, enabling spatial feature extraction with limited training data.	Proposed a U-shaped CNN: an encoder (contracting path) captures spatial context via convolution and pooling, and a symmetric decoder (expanding path) recovers that spatial resolution by upsampling.	Achieved state-of-the-art performance on cell semantic segmentation benchmarks. Outperformed sliding-window CNN methods, winning ISBI cell tracking challenge with large margin. Fast inference ≈1s per 512•512 image. Demonstrated that such architecture can be trained end-to-end on very limited data.	Not captured global context or uncertainty; heavily depends on training data quality. Does not capture small structures.	Serves as the baseline architecture: our proposed model builds on the U-Net's encoder–decoder design. Understanding U-Net's strengths (data efficiency, precise localization)

Authors (Year)	Title	Objective / Purpose	Methodology	Key Findings / Results	Gaps / Limitations	Relevance to Your Research
Chen et al. (2024) [13]	TransUNet: Rethinking the U-Net Architecture through Transformers	Integrate vision transformers into U-Net to capture global context and refine segmentation for complex tasks	Developed a hybrid CNN–Transformer U-Net. The encoder first extracts CNN feature maps, then a Transformer encoder tokenizes patches for global context. A Transformer decoder uses cross-attention between region proposals and U-Net features. They explore three configurations (encoder-only, decoder-only, encoder+decoder). Both 2D and 3D versions provided.	Significant gains over strong CNN baselines: e.g., +1.06% Dice on multi-organ segmentation and +4.30% on pancreatic tumor segmentation compared to nnU-Net. Outperforms top solution on BraTS 2021. Highlights that Transformer-augmented U-Net captures long-range dependencies and small targets better.	Increased model complexity and training cost. The best performance requires careful architecture tuning (which combination of encoder/decoder). May need large data or pretraining to realize full benefit.	Demonstrates the benefit of incorporating self-attention/transformers into a U-Net-like network. Our model similarly uses attention modules (in skip paths) and could draw on Transformer ideas. The improvements over nnU-Net motivate using global-context modules to better segment complex, post-treatment gliomas in BraTS 2024, which our approach also aims to address.
Wang et al. (2024) [17]	A4-U-Net: Deformable Multi-Scale Attention Network for Brain Tumor Segmentation	Improve multi-scale tumor segmentation by combining several attention mechanisms (spatial, channel, deformable) in a U-Net.	Proposed a U-Net variant with multiple attention modules: (1) Deformable Large Kernel Attention (DLKA) in the encoder for multi-scale feature capture; (2) Swin Spatial Pyramid Pooling (SSPP) with cross-channel attention in the bottleneck for long-range context; (3) Combined Attention Module (CAM) using DCT-based channel weighting and spatial gating in the decoder; (4) Attention Gates on skip connections.	Achieved 94.4% Dice on BraTS 2020. Each attention component incrementally improves boundary accuracy and context understanding. The model sets multiple new records on standard MRI tumor datasets.	Very high architectural complexity and many parameters (risk of overfitting). Relies on large datasets/augmentation to train all modules. Real-time inference may be slower due to extensive attention layers.	Closely related to our approach: they use attention in skip connections and multi-scale attention in encoder/decoder. Their success on BraTS shows the value of such modules by incorporating a three-way classification branch to guide feature focus, addressing post-treatment tumor heterogeneity.
Tang et al. (2024) [9]	3D U-KAN for Multi-modal MRI Brain Tumor Segmentation	Evaluate U-KAN (a U-Net with Kolmogorov–Arnold Network layers) for 3D brain tumor segmentation and efficiency.	Adapted the 2D U-KAN model (which has KAN layers for richer representation) to 3D MRI. Introduced a variant U-KAN-SE by adding Squeeze-and-Excitation blocks for global channel attention. Trained and compared U-KAN and U-KAN-SE against U-Net, Attention U-Net, and Swin UNETR on BraTS 2024. Models are relatively lightweight (10.6M parameters).	U-KAN and UKAN-SE are highly efficient: ¼ training time of U-Net and Attn U-Net, 1/6 of Swin UNETR, while generally surpassing them on most metrics. UKAN-SE slightly outperforms U-KAN. Shows that adding global channel attention (SE) yields better accuracy. Good trade-off between speed and accuracy on BraTS-PT data.	The use of KAN layers may be harder to extend. It is unclear how performance scales with much more data or other tasks.	Illustrates that lightweight models with global attention can be effective on BraTS post-treatment data. Can incorporate SE or similar modules as in UKAN-SE.

Authors (Year)	Title	Objective / Purpose	Methodology	Key Findings / Results	Gaps / Limitations	Relevance to Your Research
Celaya et al. (2024) [11]	MIST: A Simple and Scalable End-to-End 3D Medical Imaging Segmentation Framework	Provide a plug-and-play segmentation pipeline that standardizes preprocessing, augmentation, and training across tasks.	Developed MIST, an open-source framework that automates preprocessing (bias, skull-stripping), data augmentation, and multi-GPU training for 3D medical segmentation. Designed to easily integrate with BraTS post-treatment data. Evaluated MIST on the BraTS adult glioma post-treatment set and other MRI segmentation tasks.	Demonstrated that MIST can produce accurate segmentation masks on BraTS 2024 PT data and other tasks, matching or exceeding manual baselines. The standardized pipeline improved reproducibility and allowed rapid iteration. Showed it can serve as a drop-in tool for BraTS participants	MIST is a framework, not a novel model relies on existing architectures. Its gains come from infrastructure rather than new algorithmic insights. Might not handle novel modalities without adaptation	MIST underlines the importance of standardized training pipelines and robust pre-/post-processing for BraTS segmentation. Our work focuses on model architecture, but we will leverage insights like careful input normalization and augmentation (as in MIST) to maximize the performance of our attention-enhanced U-Net. It also shows that even "simple" ensembles of models can achieve strong results, motivating the use of ensembling design in our approach
Ferreira et al. (2024) [18]	Improved Multi-Task Brain Tumor Segmentation with Synthetic Data Augmentation	Enhance BraTS adult glioma segmentation by augmenting training data with synthetic MRI sequences	Proposed an adversarial MRI sequence synthesis pipeline: generate realistic synthetic MRI modality to supplement real dataset. Used these along with original data scans to train CNN-based segmentation model for BraTS task (post-treatment glioma). Standard architectures (e.g. SegResNet) were trained on combined real+synthetic data.	Using synthetic data increased robustness: achieved winning scores on BraTS task1 (post-treatment gliomas) and 3rd place on task3 (meningioma). Achieved mean Dice of 0.79–0.89 for various sub-regions in task 1. Concluded synthetic images help models generalize better under data scarcity	Gains depend on quality of synthesis. Added complexity in training.	Highlights that data scarcity in BraTS PT can be mitigated via synthetic augmentation. Though our focus is model design, we can incorporate similar ideas (e.g. generate "artificial" input highlighting enhancement).
Kim et al. (2024) [10]	Effective Segmentation of Post-Treatment Gliomas Using Simple Approaches: Artificial Sequence Generation and Ensemble Models	Improve BraTS PT glioma segmentation by simple input transformations and ensembling	Identified that post-surgery tumor enhancement is hard to detect. Created an artificial MRI sequence by linearly combining available modalities to highlight enhancing tumor (ET). Trained multiple deep models (e.g. U-Net) using this artificial sequence as extra input. Also used model ensembling to aggregate predictions.	These "straightforward approaches" significantly boosted segmentation vs baselines. The synthetic enhancing channel helped models better locate ET regions. Ensembling further improved Dice scores over any single model. Showed that even simple tricks can yield large gains in practice.	The methods are task-specific (ET emphasis). Gains depend on how well the artificial sequence correlates with pathology.	Demonstrates the value of task-specific augmentation: crafting inputs that emphasize tumor features. This complements our three-way classification idea: if the encoder is augmented to classify voxels as "clearly tumor", "uncertain", or "background," it similarly provides additional contextual cues. Kim et al. show that auxiliary channels help attention focus; we likewise integrate classification signals in the encoder to improve segmentation under challenging post-treatment appearance.

Authors (Year)	Title	Objective / Purpose	Methodology	Key Findings / Results	Gaps / Limitations	Relevance to Your Research
Maani et al. (2024) [19]	Advanced Tumor Segmentation in Medical Imaging: An Ensemble Approach for BraTS 2023 Adult Glioma and Pediatric Tumor Tasks	Describe a top-performing BraTS 2023 solution using ensemble CNNs and preprocessing.	Built an ensemble of two CNN architectures (SegResNet and MedNeXt) with tailored postprocessing. Preprocessing included bias and motion corrections (ANTS tools). Trained models separately for multi-subregion segmentation. Employed simple morphological refinements to cleanup predictions	Achieved 3rd place in BraTS 2023 ADULT ($Dice \approx 0.8313$, $HD95 \approx 36.4$). Ensemble improved robustness over individual models. The approach was simple (no transformers) but effective due to careful pre/postprocessing. Performance was competitive, albeit a bit below the winning score.	Relied on classical CNN's and hand-tuned postprocessing; lacked advanced attention or transformer modules. May not generalize beyond BraTS specifics.	Emphasizes that strong segmentation can come from well-engineered CNN ensembles and preprocessing. Our model will similarly ensure robust training (e.g. preprocessing pipelines) alongside novel architectural features.
Xing et al. 2025 [20]	Semantic Segmentation of Brain Tumors Using a Local–Global Attention Model	To enhance tumor segmentation combining local and global feature extraction.	LG UNETR: a hybrid model combining 3D Swin UNETR backbone, semantic-oriented masked attention (SMA) in decoder, and Network-in-Network (NiN) blocks in encoder.	Achieved best Dice score (82.51%) and lowest HD95 (8.02mm) on BraTS 2024; showed strong performance with fewer parameters than previous transformer-based models	Did not use full segmentation label set from BraTS 2024 (e.g., excluded Resection Cavity), No external validation or clinical evaluation	Highly relevant: uses BraTS 2024 data and targets post-treatment glioma; model innovations (SMA, NiN) and Swin Transformer base are compatible with your use of attention and 3WC modules in U-Net variants

3.1 Research Gap

While DL based segmentation models, particularly U-Net and its variants, have achieved remarkable success in brain tumor segmentation using pre-treatment datasets (e.g., BraTS 2020–2021). Their performances drop significantly when applied to post-treatment MRI's. This is primarily due to the treatment-induced changes (e.g., surgical cavities), non-standard tumor appearance, and reduced tumor size, all of which introduce ambiguity and visual similarity between tumor and healthy tissues. Although vision transformers ViT's and attention mechanisms improved the focus on tumor regions, and others have applied three-way decision theories in medical image classification tasks. There is limited research ensemble these concepts within a single segmentation pipeline especially for voxel-wise segmentation under uncertainty in post-treatment gliomas.

This gap highlights the need for a novel approach that not only enhances feature discrimination through attention mechanisms but also explicitly handles uncertain regions via three-way decision modeling, tailored for the complex and variable imaging characteristics found in the BraTS 2024 post-treatment glioma dataset.

3.2 Problem Statement

In Medical Imaging, Brain Tumors are one of the complex diseases to detect and predict. BraTS 2024 dataset presents a new challenge by focusing on post-treatment adult glioma cases, which exhibit heterogeneous appearance and treatment-induced artifacts. These complexities significantly degrade the performance of conventional segmentation models. Standard architectures like U-Net struggle to differentiate between treatment effects, tumor recurrence, and healthy tissues due to ambiguous nature of post-treatment imaging. Therefore, this research aims to develop a novel deep learning framework that enhances the U-Net architecture by: Integrating attention mechanisms within the skip connections to better focus only on relevant tumor features, and embedding a three-way classification module in the encoder to explicitly distinguish between tumor, non-tumor, and uncertain regions (boundary voxels).

3.3 Research Questions

The primary objective of this study is to design a robust ensemble deep learning framework capable of accurately segmenting brain tumor sub-regions.
*RQ*1. *Which architecture is more efficient in state-of-the-art for the segmentation of brain tumors in MRI of BraTS 2024?*

The U-Net architecture remains the most widely adopted model for medical image segmentation [3, 14, 15, 17, 21–28]. With over 113,000 citations, it has become the most popular network due to its encoder–decoder design, modular flexibility, and consistent performance across imaging modalities. Over time, numerous U-Net variants have been introduced to address scalability and complexity challenges in medical imaging. Azad et al. [2] categorized these modifications into six groups: (1) skip connection enhancements, (2) backbone design improvements, (3) bottleneck refinements, (4) transformer integration, (5) rich representation enhancements, and (6) probabilistic modeling.

In parallel, transformer-based architectures have gained prominence, with more than 185,000 citations, reflecting their unprecedented success not only in natural language processing but also in computer vision and medical imaging tasks. Researchers have extensively explored the integration of attention mechanisms into CNN-inspired designs [13, 16, 29–36]. These Vision Transformer (ViT) models have proven highly effective, offering the ability to capture long-range dependencies and generate rich feature representations, making them an attractive solution for complex segmentation problems.
*RQ*2. *Can we get better results by applying three-way approach for brain tumor segmentation?*

By applying Three-way concept, I-e; splitting into three regions instead of binary decision, using three-way decision to handle the uncertainty at the boundary values of core and fringe regions. In this context, the tumor boundaries are vague and uncertain. To delineate the boundary

using the three-way concept didn't gave us the promising results by experimenting and tuning of hyper-parameter's as compared to state-of-the-art due to the limitation of resources to handle the large dataset and model's computational complexity.
RQ3. Can implementing Vision Transformer (ViT) in U-Net architecture gives better results as compared to state-of-the-art?

The Implementing Vision Transformers (ViTs) in the U-Net architecture as seen in many recent hybrid models like TransUNet, Swin-Unet, and others can give better results than traditional convolution-based U-Nets in certain scenarios, especially in medical image segmentation. However, whether it outperforms state-of-the-art (SOTA) depends on the specific task, dataset, and architectural choices.

· Benefits of Attention Blocks

- $ViTs$ capture long-range dependencies and global relationships, which standard U-Nets struggle with.

- Especially effective for irregular shapes, heterogeneous textures, or diffuse boundaries (like post-treatment gliomas in BraTS datasets).

- $ViTs$ are good at modeling the overall contextual structure, making them ideal for incorporating high-level semantics in segmentation.

Chapter 4

Proposed Solution

We are proposing a unique hybrid ensemble model for the segmentation of tumor based on 3D U-Net Architecture and the attention mechanism by implementing the 3WC three-way clustering approach integrated with the U-Net pipeline. Vaswani et al. [37] proposed the Transformers as a new attention-driven building block. Due to the ability to model long-range dependencies, self-attention (SA) mechanism attributes in the success of transformers in medical imaging. This enables the model to capture spatial hierarchies, long-range dependencies, and global context, ensuring accurate tumor boundary segmentation. While 3WC method will handle uncertainty and vagueness in medical images more effectively, allowing us to make more granular decisions at each voxel level specially at the boundary region of the tumor.

The ensemble of these two state-of-the-art methods offers an all-encompassing approach for the Segmentation task. Below is an architecture diagram of the proposed model 4.1.

4.1 Evaluation Metrics

The model's performance is assessed using the standard evaluation metrics adopted in the BraTS 2024 challenge, specifically the Dice Similarity Coefficient (DSC) and Hausdorff Distance ($HD95$).

The Dice Similarity Coefficient (DSC), quantifies the voxel-wise overlap between the predicted segmentation mask and the ground truth annotation, while excluding true negative voxels from the calculation. This measure provides an effective way to assess how accurately a model delineates regions of interest, such as tumors, by focusing on the agreement between positive predictions and actual labeled regions.

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \tag{4.1}$$

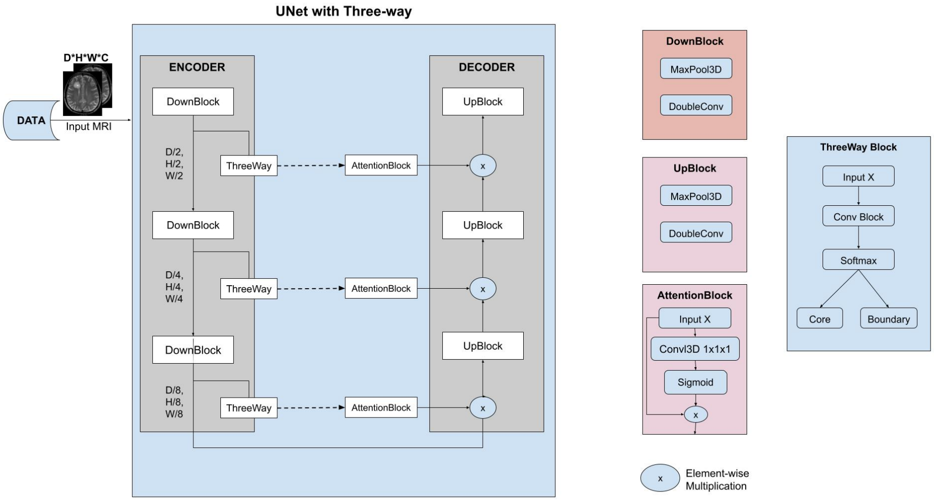

Figure 4.1: Hybrid $3W$-Att-UNet (Three-way Attention UNet)

Hausdorff distance ($HD95$), which measures maximum distance between corresponding points on the boundary of a ground truth and predicted segmentation masks, is another primary metric for the evaluation of segmentation tasks like BraTS.

$$d_H(X,Y) := \max \left\{ \sup_{x \in X} d(x,Y), \sup_{y \in Y} d(X,y) \right\} \tag{4.2}$$

These evaluation metrics equation (4.1) & (4.2) are the standard metrics for the BraTS challenge to evaluate the performance of models. Other than these two metrics, we are also using Intersection over Union IoU, to evaluate the model.

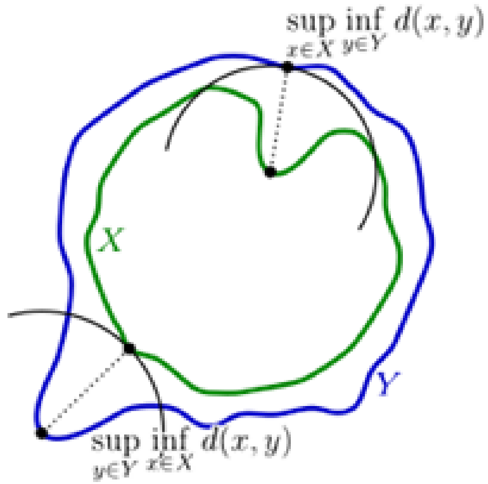

Figure 4.2: Calculation of Hausdorff Distance

Chapter 5

Experimentation & Results

5.1 Dataset

International Brain Tumor Segmentation (BraTS) challenge, has created a benchmarking environment and dataset for the segmentation of adult gliomas tumors. There are 10 different challenges but our focus will be on addressing (Task 1) post-treatment glioma with a new dataset. The challenge is the first to focus on post-treatment scans (contains changes after surgery), which is the reason of selecting this dataset for our experimentation. This challenge introduces a newly curated dataset composed exclusively of post-treatment MRI scans of patients diagnosed with both low-grade and high-grade diffuse gliomas. [1] The data were acquired through clinical imaging protocols that employ multi-parametric MRI sequences, ensuring comprehensive coverage of tumor characteristics. The modalities included are:

1. Pre-contrast T1 weighted (T1n)
2. Contrast-enhanced T1 weighted (T1w)
3. T2 weighted (T2w)
4. T2 fluid-attenuated inversion recovery (FLAIR) T2f

Seven different academic medical centers contributed in Data collection. Post-treatment clinically-acquired MRI scans of glioma, are used as the training, validation, and testing. Ground truth annotations of the tumor sub-regions are created and approved by expert neuroradiologists for every subject included in the training, validation, and testing datasets to quantitatively evaluate the predicted tumor segmentations. All scans are available as 3D nifti files (.nii.gz). The dataset contains 1350 folders, each have 5 ($.nii.gz$) files according to the above mentioned MRI protocol sequence. One file is for the Mask of the tumor. We have requested the authorized body for the

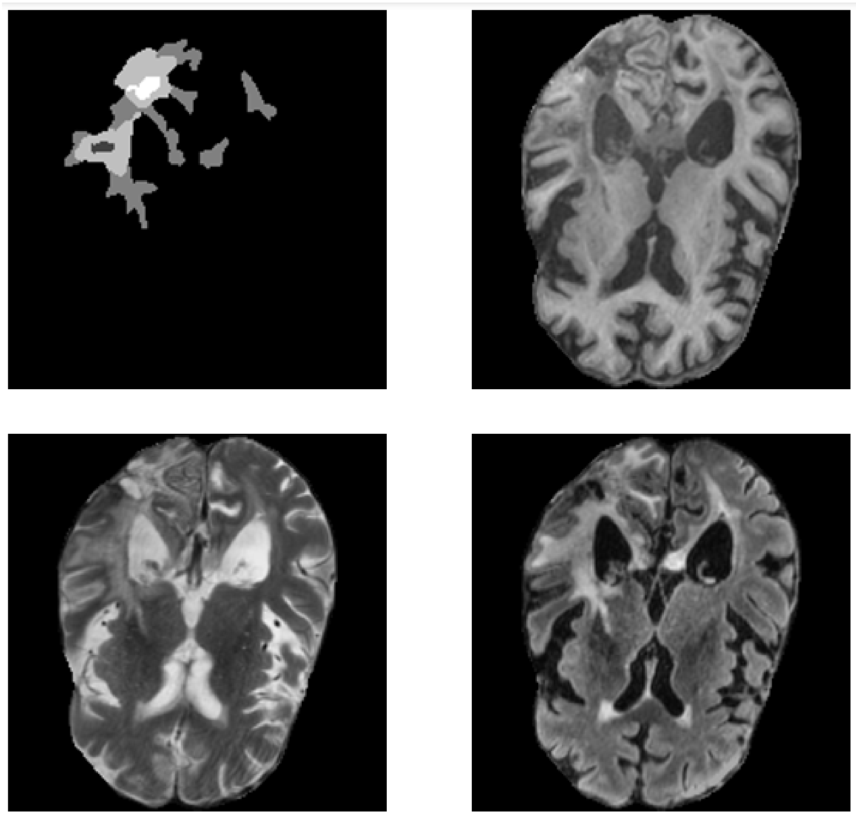

Figure 5.1: Random Sample from BraTS 2024 Adult Glioma Post Treatment Dataset

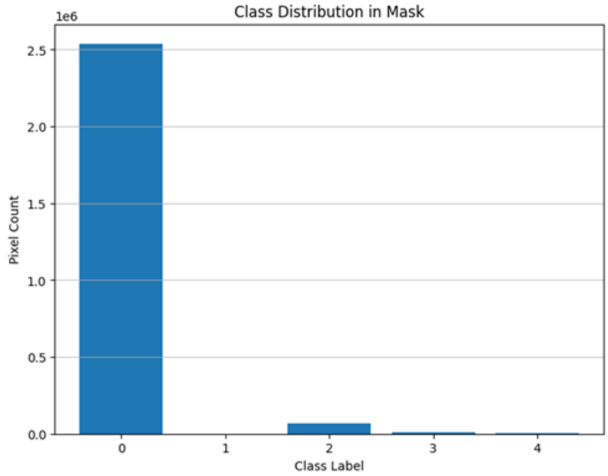

Figure 5.2: Pixel counts against Class labels in a single Mask

dataset after registering of their platform i.e.; Synapse.org. Now we have the access to download and use the dataset for our experimentation after fulfilling their licensing and citation terms and conditions. Out of 1350, we have uploaded 495 on Google Drive due to the storage constraint and 700 on Kaggle. The total size of the provided dataset by BraTS is $35GB$. Each 3D niftii file has a dimension of $182 * 218 * 182(x * y * z)$ with voxel size of $1\,\mathrm{mm}^3$. The link for the Google Drive having dataset and Source code notebooks is Google Drive's Repository. The link for Kaggle is Kaggle's Repository.

5.1.1 Class Imbalance

The dataset contains four classes including background. The following are the representation of classes with their unique values in the mask file: 0: Background 1: Non-enhancing Tumor Core (NETC) 2: Surrounding Non-enhancing FLAIR hyperintensity (SNFH) 3: Enhancing Tumor (ET) Label 1,2 and 3 are the tumor sub-regions. Whole tumor (WT) contains ET + SNFH + NETC, defines the whole extent of the tumor. The labels that contain the ground truth about the tumor has the above values for each class. The background class contains more of the area than the tumor that means our dataset has a class imbalanced as shown in figure below. The above figure displays the class labels in a mask. It contains the background pixels more than 99% of the total pixels. Which means the label is dominant by the background pixels that is black contains no information. The tumor region pixels are very low in number except the label 2 (SNFH). This refers to the class imbalance as number of pixels belonging to different classes are significantly unequal. There are consequences of class imbalance; Firstly, models trained on imbalanced data tend to become biased towards the majority class, as predicting this class more frequently can lead to high overall accuracy. Secondly, this bias often comes at the cost of poor performance on the

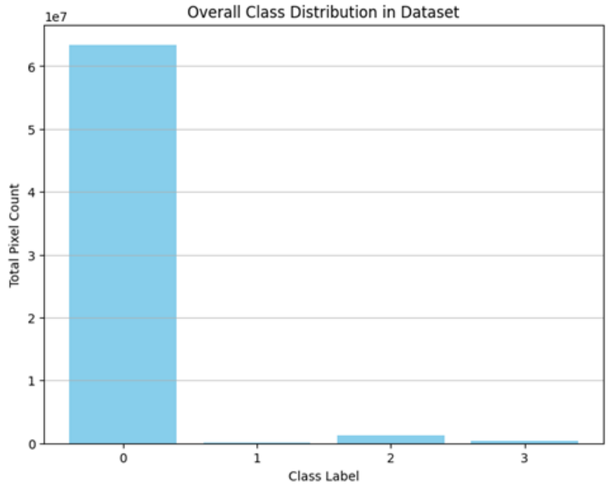

Figure 5.3: Classes Distribution in whole Dataset

minority classes, which are the most critical for accurate segmentation. Lastly, traditional metrics like accuracy can be misleading, as a high accuracy might simply reflect the model's ability to correctly classify the dominant background pixels, while failing to segment the crucial minority objects. The figure above shows the information about the class distribution across the whole dataset labels. To overcome this, we use weights to be assigned to each class. These weights are calculated based on the dataset. By assigning higher weights to the minority class, the model pays more attention to it during training, improving its ability to correctly classify instances from that class. Benefits of Using Class Weights:

· **Improved performance on minority classes:** By giving more weight to the minority class, the model is encouraged to learn its patterns and reduce bias towards the majority class. · **Reduced misclassifications errors:** The higher penalty for misclassify minority class instances helps to minimize errors on these crucial instances. The calculated weights are: Class 0: 0.967580 Class 1: 0.000814 Class 2: 0.019525 Class 3: 0.005481

Loss Function

The weighted cross entropy is used as a loss function. It may slow convergence slightly, but leads to better generalization on validation/test sets. It also reduces the model's bias toward dominant classes (e.g., background).

5.2 Mathematical Representation of Three-Way Module for MRI Voxel Data

Mathematical formulation of the three-way module process, specifically tailored for MRI voxel data and tumor segmentation.
Let $F \in \mathbb{R}^{B\times C\times D\times H\times W}$ represent a $3D$ feature map extracted from the encoder at a specific level, where:
B: batch size
C: number of feature channels
D, H, W: spatial dimensions
R is the set of real numbers.
The Three-Way Clustering module applies a $1 * 1 * 1$ convolution to project features into cluster space:

$$Z = \mathrm{Conv}_{1\times 1\times 1} F \in \mathbb{R}^{B\times K\times C\times D\times H\times W} \tag{5.1}$$

where $K = 3$ is the number of clusters/regions: Core, Fringe, and Boundary.

1. Cluster Probability Estimation A Softmax is applied over the cluster dimension:

$$P_{ijk} = \frac{\exp(Z_{ijk})}{\sum_{l=1}^{K} \exp(Z_{ljk})},\ \forall\ i, j, k \tag{5.2}$$

   where $P \in \mathbb{R}^{B\times K\times C\times D\times H\times W}$ denotes the cluster probability tensor that gives the confidence of each voxel belonging to each cluster.

2. Cluster Assignment & Boundary Estimation Let:

$$P_{\max} = \max_k P_k \tag{5.3}$$

$$C = \arg\max_k P_k \tag{5.4}$$

   Define boundary voxels where model uncertainty is high:

$$B = I(P_{\max} < \tau_b) \tag{5.5}$$

   This creates a binary mask identifying voxels where the model is uncertain (i.e., maximum probability below a boundary threshold $\tau_b = 0.65$.

   Here I is the indicator function that returns 1 if the condition is true, else 0.

3. Semantic Region Definitions

$k \in \{0, 1, 2\}$ and indexes the clusters predicted.

For the dominant cluster e.g., Core assumed at $k = 0$:

- Core Region:

$R_{\text{core}} = I(P_0 > 0.8)$, Mask for Core voxels

- Fringe Region:

$R_{\text{fringe}} = I(0.4 < P_0 \leq 0.8)$, Mask for Fringe voxels

- Boundary Region:

$R_{\text{boundary}} = B$, Mask for Boundary voxels These masks $R_{\text{core}},\ R_{\text{fringe}},\ R_{\text{boundary}} \in \{0, 1\}^{B\times1\times D\times H\times W}$ define semantic subdivisions within the feature space for more robust context-aware fusion.

5.3 Experiment Environment

The Google Colab environment is used to develop and train the model and the data is stored in Google Drive. The Paid version of Colab I-e; Colab PRO is used to train the models because we have 3D spatial data to process with the dimensions of $182 * 218 * 182$ with voxel size of 1 mm^3 which means 1 scan file is having $7, 204, 448$ voxels. Four modalities stack of $float32$ type takes about $122MB's$ of space. One mask file of type $int64$ take around $55MB's$ of space. So, a high GPU memory is required to process this spatial data to train on 3D U-Net. For this, Colab PRO with extended GPU Ram and GPU T4 or A100 will be used for this task.

5.4 Experiments

For experiment # 1,2,5 and 6, the number of epochs was 20, learning rate was 1e−4 and Adam optimizer for all the experiments. The accum_step=2 is used for gradient accumulation; this will simulate the batch size to 2 instead of 1 and the optimizer updates model weights every 2 batches instead of one.

For experiment # 3 and 4, number of epochs are 5, 394 training scans and 101 validation scans. In pre-processing, the original 3D volume of size $182x218x182$ is down sampled to $96x96x96$. The data loaded during the training is in the patches of $64x64x64$ voxels.

Experiment No.	Model	Mean DSC	Mean IoU	Mean HD95
1	3D U-Net	0.23	0.34	49.70
2	Attention U-Net	0.24	0.92	22.58
3	Three-Way Attention U-Net	0.31	0.97	13.49
4	Three-Way Attention U-Net	0.26	0.97	37.26
5	Three-Way Attention U-Net	0.13	0.30	16.50
6	Three-Way Attention U-Net	0.25	0.93	23.35

Table 5.1: Comparison of results of different models

- **3D U-Net (Exp # 01)**

Firstly, custom 3D U-Net model architecture is defined to train on our BraTS dataset with limited number of scans and 20 epochs. The base model with no additional layer (Attention Block) is implemented. Then the model is evaluated using evaluation metrics of DSC, HD95 and IoU. The input of the model is the whole volume being cropped in the pre-processing from the dimensions of $182 * 218 * 182$ to $128 * 160 * 128$ from a center to crop the background. The training curves

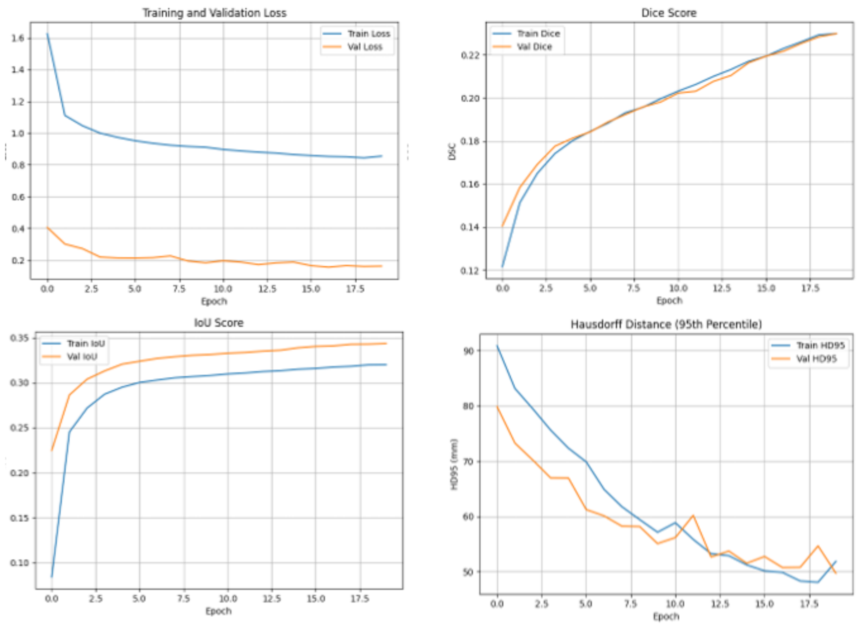

Figure 5.4: 3D U-Net Training Curves

in 5.6 shows us that the Dice score is on the increasing side, this indicates that continuous training of this model can gave us the increased dice score. In 5.5, we can see the model's prediction on test data sample. Even less trained model predicts a much better overall shape of the tumor. If we train this model further, it can give us higher DSC.

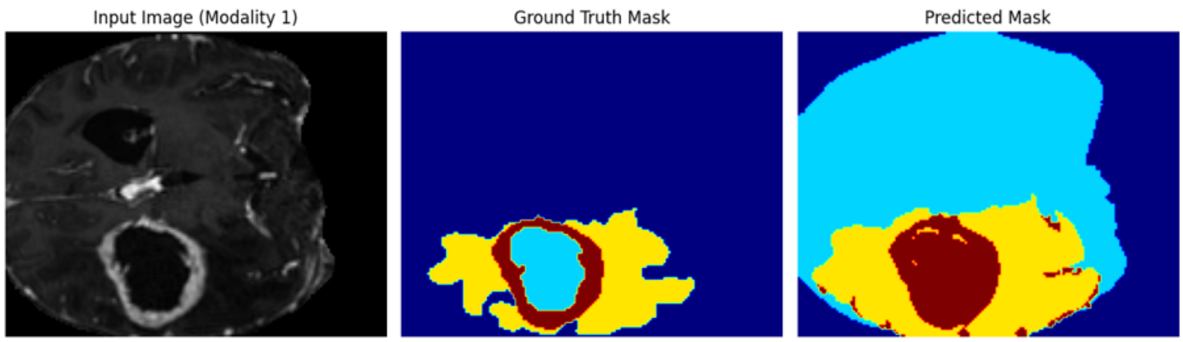

Figure 5.5: 3D U-Net Prediction Results

- **Attention U-Net (Exp # 02)**

The U-Net with the Attention mechanism in the encoder path is defined to train on our dataset. The attention block will focus on the important features only, tells the model to focus on those features. This block is computationally heavy on the GPU due to its nature and computation. This model produces better prediction by trained only on limited data and less epochs. It gives IoU=0.9242, HD=22.5823 and DSC=0.2479. The model prediction in the validation state is shown below. The predicted mask and ground truth mask has a better overlap between them. The testing data also shows a better overlap between predicted and ground truth masks as shown below. The inference on the trained model gave us an Average Dice Score: 0.2473, Average IoU Score: 0.9371 and an Average HD95: 16.6809. The attention block not only predict the bigger tumor but also focuses on smaller regions. This indicates that the attention block is focusing on region of interest ignoring the irrelevant data. The DSC curves also indicates that it is increasing and can be increased with more number of epochs.

- **Our Model (Three-way Attention U-Net) (Exp # 03)**

Below are the results generated by our model with the three-way module integrated in the encoder block of the U-Net architecture after each of the Down Block. The results of this module pass it through the skip connection to the corresponding decoder layer with the concatenation of Attention Block to focus on the tumor regions more while Up Sampling in the decoder block. The experiment run with 5 number of epochs. The data contains 394 scans of the train data and 101 validation scans. In pre-processing, the original 3D volume of size $182x218x182$ is down sampled to $96x96x96$. The data loaded during the training is in the patches of $64x64x64$ voxels. This is due to the limited resources available especially the GPU Ram.

5.4.1 Boundary Region Ratio:

- Model Uncertainty/Confidence:

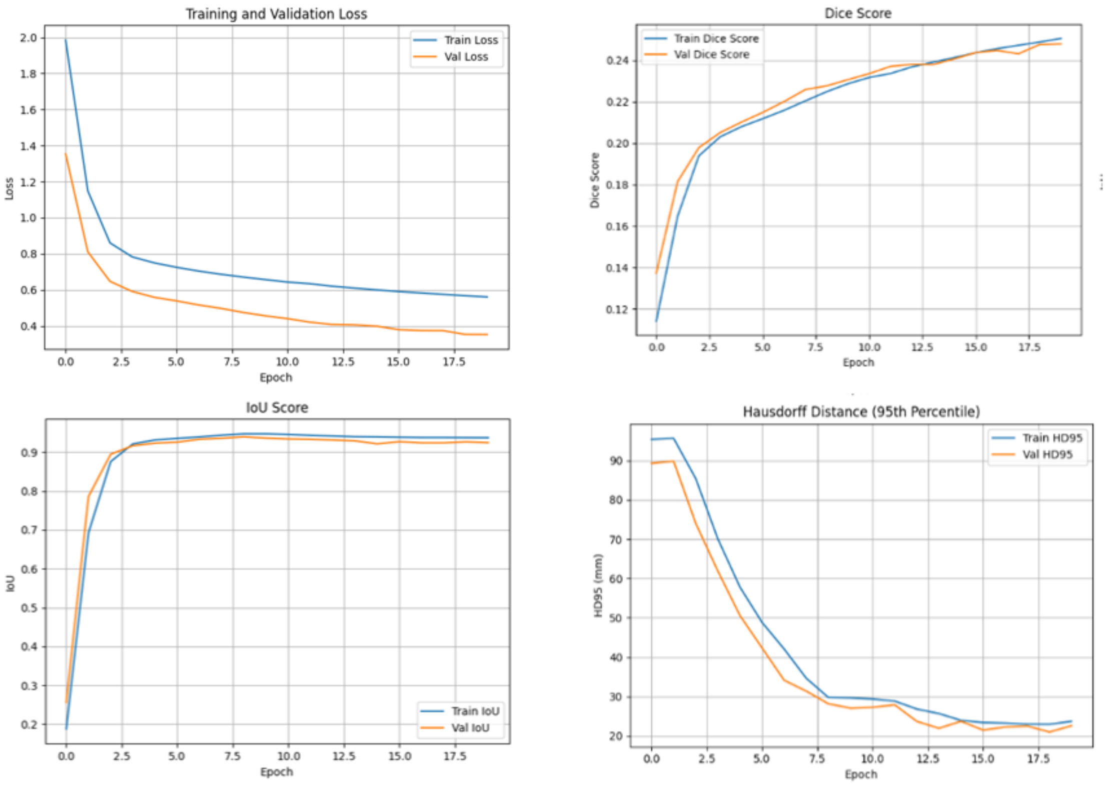

Figure 5.6: Attention U-Net Training Curves over 20 epochs

Val Loss	Val DSC	Val IoU	Val HD95
0.08	0.69	0.97	13.49

Table 5.2: Validation Results

A higher boundary ratio might suggest that the model is more uncertain or less confident in clearly separating features into distinct clusters at that level.

- Feature Discrimination:

Conversely, a lower boundary ratio indicate that the model is learning more discriminative features at that level, leading to clearer separation between clusters.

- Learning Progression:

Tracking the boundary ratio during training could reveal whether the model is becoming more confident in its clustering as training progresses. As seen in the plot, this ratio decrease over epochs as the model learns better representations.

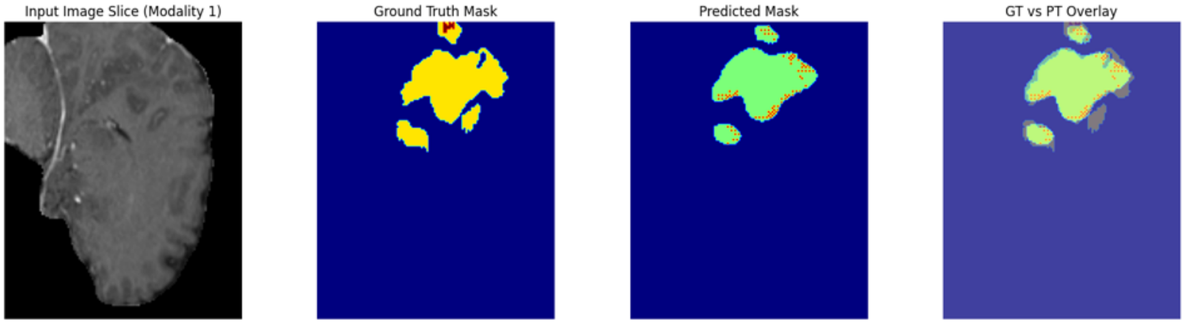

Figure 5.7: Attention U-Net Prediction Results

- Impact of the Boundary Loss:

The boundary loss in the training of the model (weighted by 0.2) specifically penalizes errors in the segmentation within these detected boundary regions. The boundary ratio acts as a measure of how prevalent these difficult-to-segment boundary regions are during training.
Below is the visualization of a specific slice of a 3D volume, shows a patch representing original image, ground truth mask, predicted mask and GT vs Pred Mask overlay.

- Conclusion:

According to the results and visualizations, the use of patches and down sampling of the original image dimensions does not give us the prediction near to ground truth as the model is struggling in segmenting the curves or cuts (tumor structure) of the tumor region due to the loss in spatial information in the pre-processing of the data. Although model predicts the tumor region well even in low epochs. There's a need to alter the pre-processing step I-e; to use the whole 3D volume as an input to capture the spatial information of the 3D space across different modalities.

- **Our Model (Three-way Attention U-Net)(Exp # 04)**

The total number of training samples are 100, validation 20. The patch size used is $80 * 80 * 80$ and the input scan size is 160*160*160 for each modality. Learning rate used is 1e−4 and batch size is 1.

Val Loss	Val DSC	Val IoU	Val HD95
0.21	0.28	0.97	37.26

Table 5.3: Validation Results

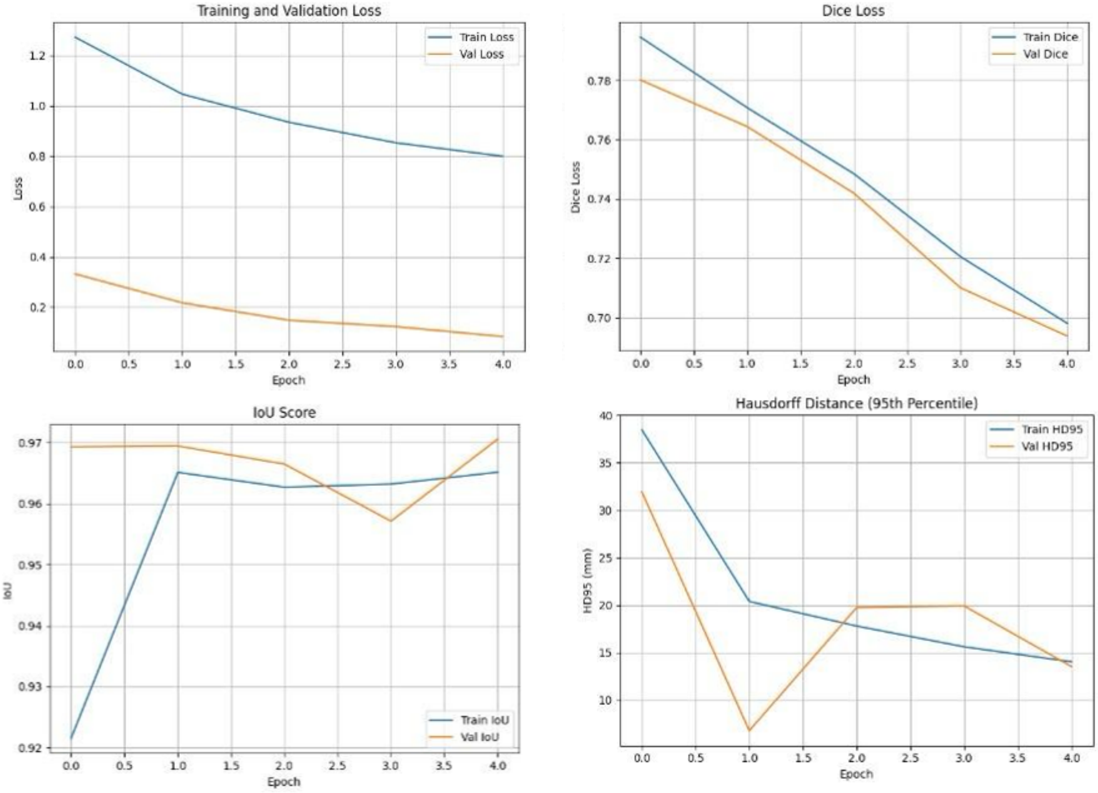

Figure 5.8: 3W-Attention U-Net Training Curves

- **Our Model (Three-way Attention U-Net)(Exp # 05) - Kaggle**

The total number of samples are 600. A patch size used is $32 * 32 * 32$ and the input scan cropped to a size of $128 * 160 * 128$ from $182 * 218 * 182$ for each modality. Learning rate used is 1e-4 and batch size is 1. The training takes about 11.7 hours on Kaggle $2xT4$ GPU. The number of epochs for training was 20.

	Val Loss	Val DSC	Val IoU	Val HD95
Epoch 20	0.7	0.18	0.95	0.32

Table 5.4: Validation Results

- Conclusion

Even the 20 epochs does not gives us the stable results showing that model is struggling to train on the given image and mask patches. The curves are constantly fluctuating which shows that the model is under-fit.

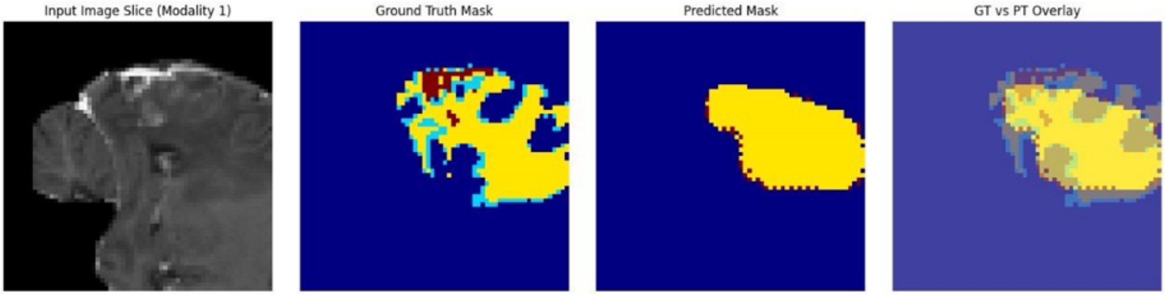

Figure 5.9: 3W-Attention U-Net Prediction Results

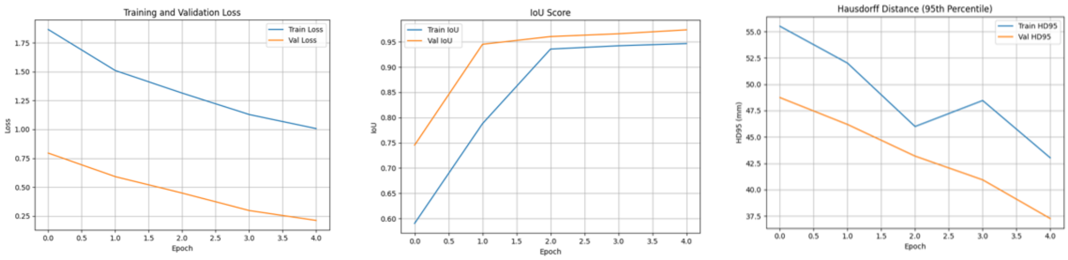

Figure 5.10: Training Curves of Our Model

- **Our Model (Three-way Attention U-Net)(Exp # 06) - Kaggle**

The total number of samples are 300. Instead of patching, the input scan cropped to a size of 128*160*128 from 182*218*182 and used as a whole volume as an input to train our model. The above figure shows a good overlap between predicted mask and ground truth mask in a validation of our model. During inference, our model got a DSC: 0.249, IoU: 0.93 and HD95: 23.35. This gave us a slightly better DSC as compared to previous experiment that uses patching.

More-To-Be-Added

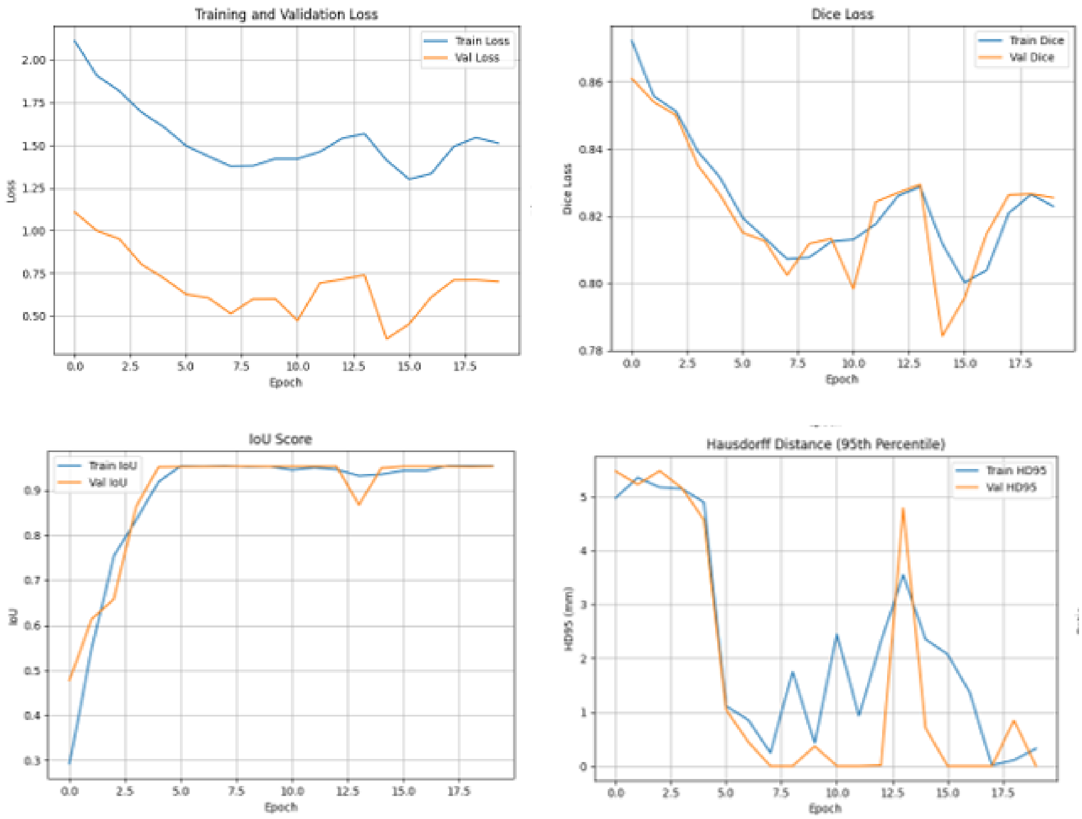

Figure 5.11: Training Curves of Our Model

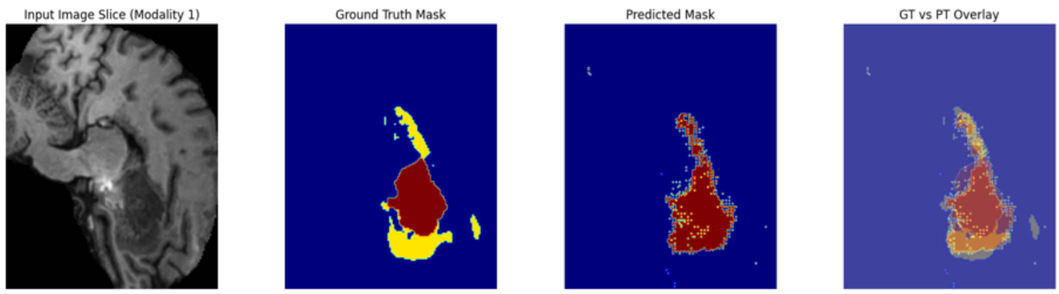

Figure 5.12: Predicted Results of Validation Data

Chapter 6

Conclusion

The experimental results show the clear difference in results of evaluation metrics that our model is giving better results. Although the experimentation is limited in terms of data scans used and epochs. This is due to the available resources such as GPU memory and runtime constraint. Runtime of 12hrs is available on both Colab and Kaggle. It means we can train our model continuously for 12hrs only. As we have 3D spatial data with multiple modalities so it will consume more CPU and GPU memory for computation.

We will do the following in our next step:

· Use all the available data for training.

· Do training for more than 500 epochs by setting up a custom hardware.

· Using dataset other than BraTS for inference.

· Implementing Multi-Head Self Attention for parallel processing.

References

[1] M. C. de Verdier, R. Saluja, L. Gagnon, D. LaBella, U. Baid, N. H. Tahon, M. Foltyn-Dumitru, J. Zhang, M. Alafif, S. Baig, *et al.*, “The 2024 brain tumor segmentation (brats) challenge: Glioma segmentation on post-treatment mri,” *arXiv preprint arXiv:2405.18368*, 2024.

[2] R. Azad, E. Aghdam, A. Rauland, Y. Jia, A. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. Cohen, E. Adeli, and D. Merhof, “Medical image segmentation review: The success of u-net. arxiv 2022,” *arXiv preprint arXiv:2211.14830*, 2024.

[3] F. Isensee, P. F. Jaeger, S. 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, no. 2, pp. 203–211, 2021.

[4] L. Maier-Hein, M. Eisenmann, A. Reinke, S. Onogur, M. Stankovic, P. Scholz, T. Arbel, H. Bogunovic, A. P. Bradley, A. Carass, *et al.*, “Why rankings of biomedical image analysis competitions should be interpreted with care,” *Nature communications*, vol. 9, no. 1, p. 5217, 2018.

[5] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, *et al.*, “The medical segmentation decathlon,” *Nature communications*, vol. 13, no. 1, p. 4128, 2022.

[6] H. Yu, “Three-way decisions and three-way clustering,” in *International joint conference on rough sets*, pp. 13–28, Springer, 2018.

[7] H. Yu, “A framework of three-way cluster analysis,” in *International Joint Conference on Rough Sets*, pp. 300–312, Springer, 2017.

[8] P. Wang, X. Yang, W. Ding, J. Zhan, and Y. Yao, “Three-way clustering: Foundations, survey and challenges,” *Applied Soft Computing*, vol. 151, p. 111131, 2024.

[9] T. Tang, Y. Chen, and H. Shu, “3d u-kan implementation for multi-modal mri brain tumor segmentation,” *arXiv preprint arXiv:2408.00273*, 2024.

[10] H. Kim, L. Milecki, M. C. Moghadam, F. Liu, M. Nguyen, E. Qiu, A. Thanki, and M. R. Sabuncu, "Effective segmentation of post-treatment gliomas using simple approaches: Artificial sequence generation and ensemble models," *arXiv preprint arXiv:2409.08143*, 2024.

[11] A. Celaya, E. Lim, R. Glenn, B. Mi, A. Balsells, D. Schellingerhout, T. Netherton, C. Chung, B. Riviere, and D. Fuentes, "Mist: A simple and scalable end-to-end 3d medical imaging segmentation framework," *arXiv preprint arXiv:2407.21343*, 2024.

[12] H.-Y. Zhou, J. Guo, Y. Zhang, X. Han, L. Yu, L. Wang, and Y. Yu, "nnformer: Volumetric medical image segmentation via a 3d transformer," *IEEE transactions on image processing*, vol. 32, pp. 4036–4045, 2023.

[13] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang, *et al.*, "Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers," *Medical Image Analysis*, vol. 97, p. 103280, 2024.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.

[15] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "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, 2016.

[16] A. Hatamizadeh, Z. Xu, D. Yang, W. Li, H. Roth, and D. Xu, "Unetformer: A unified vision transformer model and pre-training framework for 3d medical image segmentation," *arXiv preprint arXiv:2204.00631*, 2022.

[17] R. Wang, T. Tang, H. Du, Y. Cheng, Y. Wang, L. Yang, X. Duan, Y. Yu, Y. Zhou, and D. Chen, "A4-unet: Deformable multi-scale attention network for brain tumor segmentation," in *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2583–2590, IEEE, 2024.

[18] A. Ferreira, T. Jesus, B. Puladi, J. Kleesiek, V. Alves, and J. Egger, "Improved multi-task brain tumour segmentation with synthetic data augmentation," *arXiv preprint arXiv:2411.04632*, 2024.

[19] F. Maani, A. U. R. Hashmi, M. Aljuboory, N. Saeed, I. Sobirov, and M. Yaqub, "Advanced tumor segmentation in medical imaging: An ensemble approach for brats 2023 adult glioma and pediatric tumor tasks," in *International Challenge on Cross-Modality Domain Adaptation for Medical Image Segmentation*, pp. 264–277, Springer, 2023.

[20] S. Xing, Z. Lai, J. Zhu, W. He, and G. Mao, "Semantic segmentation of brain tumors using a local–global attention model," *Applied Sciences*, vol. 15, no. 11, p. 5981, 2025.

[21] L. Liu and K. Xia, “Btis-net: Efficient 3d u-net for brain tumor image segmentation,” *IEEE Access*, 2024.

[22] R. Raza, U. I. Bajwa, Y. Mehmood, M. W. Anwar, and M. H. Jamal, “dresu-net: 3d deep residual u-net based brain tumor segmentation from multimodal mri,” *Biomedical Signal Processing and Control*, vol. 79, p. 103861, 2023.

[23] J. M. J. Valanarasu, V. A. Sindagi, I. Hacihaliloglu, and V. M. Patel, “Kiu-net: Overcomplete convolutional architectures for biomedical image and volumetric segmentation,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 4, pp. 965–976, 2021.

[24] M. Kharaji, H. Abbasi, Y. Orouskhani, M. Shomalzadeh, F. Kazemi, and M. Orouskhani, “Brain tumor segmentation with advanced nnu-net: pediatrics and adults tumors,” *Neuroscience Informatics*, vol. 4, no. 2, p. 100156, 2024.

[25] M. Futrega, A. Milesi, M. Marcinkiewicz, and P. Ribalta, “Optimized u-net for brain tumor segmentation,” in *International MICCAI brainlesion workshop*, pp. 15–29, Springer, 2021.

[26] M. Bhalerao and S. Thakur, “Brain tumor segmentation based on 3d residual u-net,” in *International MICCAI brainlesion workshop*, pp. 218–225, Springer, 2019.

[27] A. Rawat and R. Kumar, “3d u-net-norm architecture for improving generalization of brats images,” *Multimedia Tools and Applications*, vol. 84, no. 20, pp. 22139–22159, 2025.

[28] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, “S3d-unet: separable 3d u-net for brain tumor segmentation,” in *International MICCAI Brainlesion Workshop*, pp. 358–368, Springer, 2018.

[29] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, “Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images,” in *International MICCAI brainlesion workshop*, pp. 272–284, Springer, 2021.

[30] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, and J. Li, “Transbts: Multimodal brain tumor segmentation using transformer,” in *International conference on medical image computing and computer-assisted intervention*, pp. 109–119, Springer, 2021.

[31] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang, *et al.*, “3d transunet: Advancing medical image segmentation through vision transformers,” *arXiv preprint arXiv:2310.07781*, 2023.

[32] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, “Transunet: Transformers make strong encoders for medical image segmentation,” *arXiv preprint arXiv:2102.04306*, 2021.

[33] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, and D. Xu, “Unetr: Transformers for 3d medical image segmentation,” in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 574–584, 2022.

[34] Y. Xie, C. Zhou, J. Mei, X. Li, C. Xie, and Y. Zhou, “Brain tumor segmentation through supervoxel transformer,” in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, IEEE, 2024.

[35] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, and Y. Yu, “nnformer: Interleaved transformer for volumetric segmentation,” *arXiv preprint arXiv:2109.03201*, 2021.

[36] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, “Swin-unet: Unet-like pure transformer for medical image segmentation,” in *European conference on computer vision*, pp. 205–218, Springer, 2022.

[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[38] F. Isensee, P. F. Jäger, P. M. Full, P. Vollmuth, and K. H. Maier-Hein, “nnu-net for brain tumor segmentation,” in *International MICCAI Brainlesion Workshop*, pp. 118–132, Springer, 2020.