MULTI-MODAL MRI-BASED BRAIN TUMOR SEGMENTATION USING INTEGRATED CNN, TRANSFORMER, AND ATTENTION MODULES

MD. SHAHIDUL ISLAM SHABUZ

SUPERVISOR:

PROF. DR. POR LIP YEE

MADAM FAZIDAH BINTI OTHMAN

FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY

UNIVERSITI MALAYA

KUALA LUMPUR

2025

Multi-Modal MRI-Based Brain Tumor Segmentation Using Integrated CNN, Transformer, and Attention Modules

ABSTRACT

This research addresses the problem of brain tumor segmentation in multi-modal Magnetic Resonance Imaging (MRI). Existing methods often face limitations related to class imbalance, spatial complexity, boundary inconsistency, and computational cost. To address these challenges, this study proposes a hybrid method that integrates convolutional neural networks (CNNs), Transformer modules, and dual attention mechanisms. The method consists of a CNN encoder for local feature extraction, a Swin Transformer for global context modeling, and both spatial and channel attention modules for modality-specific relevance enhancement. A cross-attention fusion layer is used to integrate CNN and Transformer features. The method is trained using a composite loss function combining Dice loss and Focal loss to manage class imbalance. Evaluation is conducted on BraTS 2023 and TCIA glioma datasets using metrics such as Dice Similarity Coefficient, Hausdorff Distance (HD95), Average Surface Distance (ASD), sensitivity, specificity, and inference time per slice. The study focuses on adult glioma cases and uses only MRI modalities. This research aims to develop and evaluate a method that addresses the segmentation challenges identified in the problem statement and aligns with the stated objectives.

Keywords: Brain Tumor Segmentation, Multi-Modal MRI, CNN-Transformer Hybrid, Attention Mechanisms, Medical Image Analysis

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AUC	:	Area Under the Curve
Attention U-Net	:	U-Net with gated spatial attention
BraTS	:	Brain Tumor Segmentation Challenge Dataset
CAF	:	Cross Attention Fusion
CBAM	:	Convolutional Block Attention Module
CNN	:	Convolutional Neural Network
CoTr	:	CNN + Transformer hybrid model
DSC / Dice	:	Dice Similarity Coefficient
FLAIR	:	Fluid-Attenuated Inversion Recovery
FLOPs	:	Floating Point Operations
FN	:	False Negatives
FP	:	False Positives
GAN	:	Generative Adversarial Network
GBM	:	Glioblastoma Multiforme
HD95	:	95th Percentile Hausdorff Distance
HiFormer	:	Hybrid CNN-Transformer with cross-attention
IoU	:	Intersection over Union
MRI	:	Magnetic Resonance Imaging
nnU-Net	:	No-new-Net; AutoML-based U-Net
PRISMA	:	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ResUNet++	:	Residual U-Net with attention mechanisms
SE	:	Squeeze-and-Excitation
SE Block	:	Squeeze-and-Excitation Block

SwinUNet	:	Swin Transformer-based U-Net
TCIA	:	The Cancer Imaging Archive
T1c	:	T1-contrast enhanced
TN	:	True Negatives
TP	:	True Positives
U-Net	:	U-shaped CNN architecture for segmentation
UNesT	:	Nested Transformer for 3D segmentation
UNETR	:	U-Net with Transformer Encoder
UTNetV2	:	Lightweight CNN-Transformer hybrid
ViT	:	Vision Transformer

1. Introduction

Brain tumor segmentation plays a central role in diagnosis, treatment planning, and monitoring of disease progression. Magnetic Resonance Imaging (MRI) is widely used for brain imaging due to its ability to capture multiple tissue contrasts. Multi-modal MRI sequences such as T1, T1 with contrast (T1c), T2, and FLAIR provide complementary information about tumor location, structure, and surrounding tissues (Menze et al., 2015; Bonato et al., 2025). Accurate identification of tumor subregions—including enhancing tumor, peritumoral edema, and necrotic core—is necessary for clinical tasks such as surgical navigation and radiotherapy planning.

Manual segmentation by radiologists is the current standard practice but requires significant time and may be affected by variability between observers (Jungo et al., 2018; Veiga-Canuto et al., 2022). Automated segmentation methods based on deep learning have been introduced to support this task. Convolutional Neural Networks (CNNs) such as U-Net and its variants are used to extract local features and predict segmentation masks (Ronneberger et al., 2015; Isensee et al., 2021). More recent models have adopted Transformer-based architectures to model global dependencies in volumetric data (Hatamizadeh et al., 2022).

Despite progress in network design, challenges remain. Tumor subregions often exhibit irregular shape, unclear boundaries, and low contrast with surrounding tissue (Buddenkotte et al., 2023; Zhou et al., 2021). These characteristics reduce segmentation accuracy in CNN-based methods, which are limited in modeling non-local dependencies. Transformer-based models address global context but may require large datasets and high computational resources, making them difficult to train and deploy (Tang et al., 2022). In addition, the issue of class imbalance—where tumor regions represent a small fraction of the image—can lead to poor learning and low sensitivity for smaller tumor components (Xiao et al., 2025).

Some methods incorporate attention mechanisms to enhance feature selection, but most use either spatial or channel attention in isolation (Woo et al., 2018; Bello et al., 2019). Few methods combine CNN and Transformer modules with both types of attention to address the full range of segmentation challenges. Furthermore, many studies report performance based on a single dataset, limiting the understanding of model generalizability across different sources or protocols (Liu et al., 2025).

This research proposes a method that integrates a CNN encoder, a Transformer backbone, spatial and channel attention mechanisms, and a cross-attention fusion layer. The method is developed and evaluated on multi-modal MRI data to segment brain tumors into three subregions. The goal is to address the limitations of existing models by combining local and global feature extraction, modality-specific attention, and loss functions that manage class imbalance. The study is conducted using BraTS and TCIA datasets and is evaluated using standard segmentation metrics.

The research objectives are: (1) to review existing deep learning methods for brain tumor segmentation; (2) to propose a method that addresses segmentation challenges such as class imbalance and spatial complexity; (3) to develop the method using CNN, Transformer, and attention components; and (4) to evaluate the method using public datasets and standard metrics.

2. Problem Statement

Segmentation of brain tumors in MRI images is essential for diagnosis, treatment planning, and surgical navigation. However, current deep learning models still struggle to accurately delineate tumor subregions such as necrotic cores, enhancing tumor, and peritumoral edema. Although recent systems achieve Dice scores above 0.90 for whole tumor regions, subregion scores often fall below 0.88—particularly in smaller or irregular lesions—limiting clinical reliability (Zhang et al., 2024; Zarenia et al., 2025). Sensitivity also remains a concern, frequently dropping below

0.90 in detecting infiltrative or low-contrast regions, which increases the risk of missing critical pathology (Urrea & Vélez, 2025; Huang et al., 2025). In terms of boundary accuracy, models exhibit Hausdorff Distance (HD95) values around 5–6 mm and elevated Average Surface Distance (ASD), indicating poor edge localization that could misguide surgical decisions (Liu et al., 2025; Zhu et al., 2024; Zhang et al., 2024). Despite architectural advances, these overlapping, detection, and boundary-level failures persist across leading segmentation frameworks, as shown in major 2025 benchmarks and studies.

This segmentation challenge stems from several intertwined factors. First, low contrast and blurred edge definition in MRI—especially between edema and healthy tissue in FLAIR/T2 scans—make it challenging for models to detect precise tumor boundaries, resulting in poor boundary accuracy (Xiao et al., 2025). Second, the irregular morphology and heterogeneity of tumors—varying in shape, texture, and intensity—introduce unpredictable patterns that challenge segmentation accuracy and consistency (Shoushtari et al., 2025; Preetha et al., 2025). Third, class imbalance, where healthy tissue overwhelms tumor voxels, reduces a model's focus on small but clinically critical regions—resulting in low sensitivity (Mosquera et al., 2024). Fourth, advanced segmentation architectures often require high computational resources, making them impractical for low-resource or time-critical clinical environments (Zhong et al., 2025). Additionally, domain shifts—due to differences in scanner types, field strengths, and imaging protocols—reduce model generalizability across institutions and datasets (Yoon et al., 2025; Luo et al., 2024). These root causes collectively explain why modern segmentation systems still struggle in clinical-grade performance.

Researchers have explored various strategies to enhance brain tumor segmentation in MRI. CNN-based models like U-Net (Saifullah et al., 2025) and nnU-Net (Kharaji et al., 2024) improved local feature extraction, while transformer-based models such as SwinUNet (Han et al., 2025),

SwinUMamba (Liu et al., 2024), and Swin SMT (Płotka et al., 2024) enhanced contextual understanding and scalability. CSWin-UNet (Liu et al., 2025) and TransUNet (Chen et al., 2024) improved efficiency and benchmark performance. Attention-driven models—DA-TransUNet (Sun et al., 2024), P-TransUNet (Chong et al., 2023), HiFormer (Heidari et al., 2023), and PFormer (Gao et al., 2025)—further improved boundary delineation, multi-scale integration, and computational efficiency. Supplementary modules such as SE (Xiong et al., 2024), CBAM (Zhu et al., 2024), and coordinate attention (Ding et al., 2024) enhanced regional focus and semantic alignment. However, these models still face challenges related to computational cost, model complexity, and generalizability across different datasets.

In conclusion, the problem of accurate, efficient, and generalizable brain tumor segmentation is not yet solved. This research intends to develop a hybrid CNN-Transformer model with spatial and channel attention mechanisms to address this problem by improving segmentation accuracy, reducing computational demands, and enhancing generalizability across datasets.

3. Research Questions

RQ1: What are the current deep learning methods used for brain tumor segmentation in multi-modal MRI?

RQ2: How can a new method be proposed to address class imbalance, spatial complexity, and boundary inaccuracy in brain tumor segmentation?

RQ3: How to develop the proposed method?

RQ4: How to evaluate the proposed method?

4. Research Objectives

RO1: To review existing deep learning methods for brain tumor segmentation in multi-modal MRI.

RO2: To propose a method to address class imbalance, spatial complexity, and boundary inaccuracy.

RO3: To develop the proposed method using a CNN-based encoder, Transformer backbone, and spatial-channel attention mechanisms, and implement it on multi-modal MRI datasets.

RO4: To evaluate the proposed method's performance across different datasets using Dice Similarity Coefficient, Area Under the Curve, Hausdorff Distance, Average Surface Distance, sensitivity, specificity and inference.

Table 1: A mapping table of Research Questions, Objectives, and expected outcomes

Research Questions (RQs)	**Research Objectives (ROs)**	**Expected Outcomes**
RQ1: What are the current deep learning methods used for brain tumor segmentation in multi-modal MRI?	RO1: To review existing deep learning methods for brain tumor segmentation in multi-modal MRI.	A summary of current methods, their limitations in class imbalance handling, spatial localization, and computational efficiency.
RQ2: How can a new method be proposed to address class imbalance, spatial complexity, and boundary inaccuracy in brain tumor segmentation?	RO2: To propose a method to address class imbalance, spatial complexity, and boundary inaccuracy.	A proposed method designed to improve segmentation performance in challenging tumor regions.

RQ3: How to develop the proposed method?	RO3: To develop the proposed method using a CNN-based encoder, Transformer backbone, and spatial-channel attention mechanisms, and implement it on multi-modal MRI datasets.	A functional method implemented and trained on benchmark MRI datasets, ready for evaluation.
RQ4: How to evaluate the proposed method?	RO4: To evaluate the proposed method's performance across different datasets using Dice Similarity Coefficient, Area Under the Curve, Hausdorff Distance, Average Surface Distance, sensitivity, specificity and inference.	Quantitative evaluation results showing the method's effectiveness compared to related work using standard metrics.

5. Scope of the Research

This research focuses on the development and evaluation of a method for brain tumor segmentation using multi-modal MRI data. The scope is limited to the use of a hybrid deep learning model that integrates convolutional neural networks, Transformer modules, and dual attention mechanisms. The method is designed to segment three tumor subregions: enhancing tumor, peritumoral edema, and necrotic core.

The study uses four standard MRI modalities: T1, T1 with contrast (T1c), T2, and FLAIR. The datasets used for training and evaluation are the BraTS 2023 dataset and the TCIA glioma collection. The model is implemented and tested using these datasets only. The method includes preprocessing steps such as skull stripping, N4 bias correction, and intensity normalization.

Evaluation metrics include Dice Similarity Coefficient, Hausdorff Distance (HD95), Average Surface Distance (ASD), sensitivity, specificity, and inference time per slice.

This research is limited to segmentation tasks. It does not cover tumor classification, survival prediction, or treatment planning. The method is evaluated on adult glioma cases only and does not include pediatric cases or other tumor types such as meningiomas or metastases. Only MRI data are used; other imaging modalities such as CT or PET are not included.

The implementation is conducted using available computational resources, with patch-based training and mixed-precision computation to manage memory usage. The research does not involve clinical validation or deployment in healthcare settings. All findings are based on retrospective analysis using publicly available datasets.

6. Significance of the Research

This research investigates a method for brain tumor segmentation in multi-modal MRI using a hybrid approach that integrates convolutional neural networks, Transformer modules, and dual attention mechanisms. The study is designed to address specific challenges identified in the problem statement, including class imbalance, spatial complexity, boundary detection errors, and resource constraints during model training and inference.

The significance of the research lies in its focus on architectural integration and component design. The combination of a CNN encoder with a Transformer backbone supports the extraction of both local spatial features and global contextual relationships. The use of spatial and channel attention mechanisms enables the model to assign weights to spatial regions and input modalities, which may assist in distinguishing tumor subregions. The cross-attention fusion mechanism supports

feature integration across different network modules. The hybrid loss function is included to manage class imbalance by assigning greater weight to underrepresented tumor regions.

The implementation considers practical factors such as memory usage and inference time. Patch-based training and mixed-precision computation are used to manage computational load. The evaluation on two publicly available datasets (BraTS and TCIA) allows for assessment across different imaging sources and patient cases.

This research is limited to model design and performance evaluation using available data. It does not include clinical trials or deployment studies. The outputs of this study may serve as a reference for future work involving medical image segmentation or hybrid neural network development.

In summary, the research contributes a method that responds to specific technical issues in brain tumor segmentation. It aligns with the research objectives and may support further investigation into model design and performance under varied data conditions.

7. Related Work

ELSA-enhanced Swin Transformer was proposed by Ghazouani et al. in 2024. This model enhances a Swin Transformer encoder using Efficient Local Self-Attention (ELSA) blocks and channel-spatial squeeze-excitation modules to refine tumor boundary segmentation. It specifically aims to improve boundary precision in brain tumor segmentation tasks. The model was evaluated on the BraTS 2015, 2018, and 2020 datasets with Dice score 88.6%, 82.6%, 80.9% with lower average score on newer datasets, respectively. However, its performance falls short compared to top-tier hybrid models in delineating complex tumor boundaries. This limitation is aligned with the broader problem of accurately detecting tumor edges in heterogeneous and low-contrast MRI regions.

GAN-Transformer Hybrid, introduced by Huang et al. in 2022, combines Transformer modules within a GAN architecture using multi-scale deep supervision and adversarial learning to enhance contextual feature representation. The method addresses the problem of learning richer, more adaptive features for segmentation. It was tested on BraTS 2015–2020, achieving Dice scores of 88.6% (2015), 82.6% (2018), and 80.9% (2020). A major drawback is the inherent instability of adversarial training and the complexity of the overall architecture, which undermines reproducibility and limits its potential for clinical adoption—an issue emphasized in the problem statement regarding reliability and deployment.

ETUNet was developed by Zhang et al. in 2024. The model integrates both spatial and channel attention mechanisms throughout the encoder, bottleneck, and decoder stages, along with cross-attention applied in skip connections. It aims to enhance segmentation accuracy. The model reports a moderate Dice score of approximately 0.85. However, it lacks scalability due to its architectural complexity, which hinders real-world deployment—particularly in constrained clinical environments, aligning with the challenges discussed in the problem statement.

ResMT, proposed by Cui et al. in 2024, combines a 3D CNN with Swin-UNETR, incorporating multi-plane channel and spatial attention for glioma grading. This method is tailored for classification tasks rather than segmentation. It achieved an AUC of 0.995, indicating excellent classification performance. However, because it does not perform voxel-wise segmentation, it is not suitable for tasks like surgical navigation or radiotherapy planning, which require fine-grained subregion delineation—highlighting a misalignment with the problem this study seeks to solve.

MWGUNet++, introduced by Lyu and Tian in 2025, implements a parallel CNN-Transformer design enhanced with multi-window gated self-attention for improved multi-modal feature

extraction. This design aims to address the challenge of integrating information from diverse MRI sequences. However, no datasets or evaluation metrics were reported. Moreover, its validation is limited and lacks testing on large-scale or real-world clinical datasets, which raises concerns about its generalizability—an essential requirement underscored in this study.

Table 2: Synthesis and Summary of the Selected Related Work

Reference	Method	Description	Problem Addressed	Dataset	Evaluation Metric(s)	Shortcoming
Ghazouani, Vera, & Ruan (2024)	ELSA-enhanced Swin Transformer	Improves Swin Transformer with Efficient Local Self-Attention and channel-spatial SE blocks.	Enhance tumor boundary precision.	BraTS 2021	Dice: 89.8%	Falls short of top-tier hybrids in boundary-level precision.
Zhang et al. (2024)	ETUNet	Uses spatial and channel attention in encoder, bottleneck, decoder; cross-attention in skip connections.	Improved segmentation accuracy.	BraTS 2018 & 2020	Dice ~0.85	Architectural complexity may hinder clinical scalability.
Huang et al. (2022)	GAN-Transformer Hybrid	Combines GAN and Transformer with multi-scale deep supervision and adversarial learning.	Rich contextual representation learning.	BraTS 2015–2020	Dice score 88.6%, 82.6%, 80.9% respectively for BraTS 2015, 2018, and 2020	Training instability and complex pipeline hinder reproducibility.

Cui et al. (2024)	ResMT	3D CNN + Swin-UNETR with multi-plane channel and spatial attention.	Glioma grading (not segmentation).	Not specified	AUC: 0.995	Not designed for voxel-wise segmentation.
Lyu and Tian (2025)	MWGUNet++	Parallel CNN-Transformer with multi-window gated self-attention.	Effective multi-modal feature extraction.	Limited validation; not tested on large-scale datasets.	No metric reported	Limited generalization and scalability validation.

Table 2 presents a synthesis of recent state-of-the-art methods that combine CNNs, Transformers, and attention mechanisms for brain tumor segmentation. A clear pattern emerges: while several models (e.g., Ghazouani et al., 2024; Zhang et al., 2024) demonstrate promising segmentation performance using attention and hybrid architectures, they frequently fall short in one or more critical aspects.

First, many methods lack real-time inference capability or exhibit high computational complexity, rendering them unsuitable for time-sensitive clinical applications (e.g., GAN-Transformer Hybrid). Second, generalizability remains a significant concern, as most studies evaluate their models solely on benchmark datasets (e.g., BraTS), without validation across diverse clinical sources. Third, subregion segmentation performance is often inconsistent, particularly in cases involving low contrast, irregular tumor morphology, or severe class imbalance—issues central to the current proposal. Fourth, although various models incorporate spatial or channel attention mechanisms, few effectively combine both or adapt them to modality-specific features, thereby limiting their ability to fully exploit diagnostic cues from multi-modal MRI.

These limitations underscore the need for a model that is both accurate and lightweight, capable of delivering precise boundary delineation across heterogeneous tumor subregions, while remaining consistent to domain shifts. Therefore, this research proposes the development of a hybrid CNN-Transformer model equipped with dual spatial-channel attention mechanisms, designed to achieve high segmentation accuracy, minimize computational overhead, and ensure generalizability across multi-institutional MRI datasets.

7.1 Datasets

This study will utilize two publicly available and widely recognized datasets for brain tumor segmentation tasks: the Brain Tumor Segmentation Challenge (BraTS) dataset and The Cancer Imaging Archive (TCIA) dataset. These datasets are selected due to their high quality, diverse imaging modalities, and clinical relevance, which are essential for training and evaluating widely applicable segmentation models.

1. BraTS Dataset:

The BraTS dataset is part of the annual MICCAI Brain Tumor Segmentation Challenge and includes multi-institutional, multi-modal MRI scans of glioma patients. Each case includes four MRI sequences: T1-weighted (T1), T1 with contrast enhancement (T1c), T2-weighted (T2), and Fluid-Attenuated Inversion Recovery (FLAIR). The dataset provides expert-labeled ground truth segmentations for three tumor subregions: enhancing tumor (ET), peritumoral edema (ED), and necrotic/non-enhancing tumor core (NCR/NET). The dataset is preprocessed with co-registration, skull stripping, and resampling, making it ideal for deep learning model development. This research will primarily use the BraTS 2023 version, which contains over 1,500 annotated 3D volumes, split into training, validation, and testing subsets.

2. TCIA Dataset:

To evaluate the generalizability of the proposed model, external validation will be performed using the TCIA glioma collection. TCIA hosts curated radiological datasets sourced from real-world clinical institutions, offering variability in scanner types, imaging protocols, and patient demographics. For this study, the TCGA-GBM and TCGA-LGG subsets will be used, which contain MRI sequences similar to BraTS (T1, T1c, T2, and FLAIR). Since the TCIA dataset does not always provide segmentation labels, annotations will be derived from publicly available BraTS-compatible segmentations or manually reviewed when needed.

These datasets together allow for comprehensive evaluation: BraTS enables performance benchmarking under standardized conditions, while TCIA supports assessment of generalizability across domain shifts and institutional variability.

7.2 Evaluation Metrics

This study adopts a comprehensive evaluation strategy to assess the performance of the proposed brain tumor segmentation model. Among the related works, the most widely used metric is the Dice Similarity Coefficient (DSC), which measures the overlap between predicted and ground truth segmentation masks. For example, ETUNet (Zhang et al., 2024) and the ELSA-enhanced Swin Transformer (Ghazouani Jiang et al., 2024) reported Dice scores of approximately 0.85 and 89.8%, respectively. DSC is defined as:

$$DSC = \frac{2TP}{2TP+FP+FN} \quad (1)$$

where TP is the number of true positives, FP is false positives, and FN is false negatives. This metric will be computed separately for each tumor subregion: enhancing tumor (ET), tumor core (TC), and whole tumor (WT).

In classification-based models such as ResMT (Cui et al., 2024), Area Under the Curve (AUC) is used to evaluate the model's discrimination ability. Although AUC is not directly applicable to segmentation tasks, it remains relevant for evaluating auxiliary classification modules:

$$AUC = \int_0^1 TPR(FPR)dFPR \quad (2)$$

where TPR is the true positive rate and FPR is the false positive rate.

However, most existing works do not report boundary precision or account for class imbalance—gaps that are crucial for real-world applications. To address these limitations, additional metrics are adopted.

The 95th percentile Hausdorff Distance (HD95) quantifies the boundary agreement between the predicted and ground truth contours. It captures the largest surface discrepancy excluding outliers:

$$\mathrm{HD}_{95}(A,B) = \max\left\{\sup_{a\in A}\inf_{b\in B} d(a,b), \sup_{b\in B}\inf_{a\in A} d(b,a)\right\}_{95\%} \quad (3)$$

Average Surface Distance (ASD) measures the mean distance between the predicted and ground truth tumor boundaries in MRI scans. It reflects overall boundary accuracy, with lower values indicating better alignment. If S_p and S_g are the surfaces of the predicted and ground truth segmentations, then:

$$ASD(S_p, S_g) = \frac{1}{|S_p|+|S_g|}\left(\sum_{x\in S_p} d(x, S_g) + \sum_{y\in S_g} d(y, S_p)\right) \quad (4)$$

Where, $d(x, S_g)$ is the minimum Euclidean distance from point x to the surface S_g.

This metric ensures that the model provides anatomically precise boundaries, which are important for treatment planning.

To measure detection capability and false alarm rate, sensitivity and specificity are included:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (5)$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (6)$$

Sensitivity is crucial for identifying subtle or infiltrative tumors, while specificity ensures healthy tissue is not mistakenly labeled.

Finally, inference time per 2D slice will be reported in seconds to evaluate computational efficiency. This is particularly relevant because many models in the literature, such as GAN-Transformer Hybrid, are computationally expensive and impractical for real-time deployment.

By integrating these metrics—both standard (DSC, AUC) and extended (HD95, ASD, sensitivity, specificity, inference time)—this study provides a comprehensive and clinically meaningful

evaluation framework that addresses the limitations of current models and aligns with the goals of real-time, accurate and efficient brain tumor segmentation.

8. Proposed Methodology

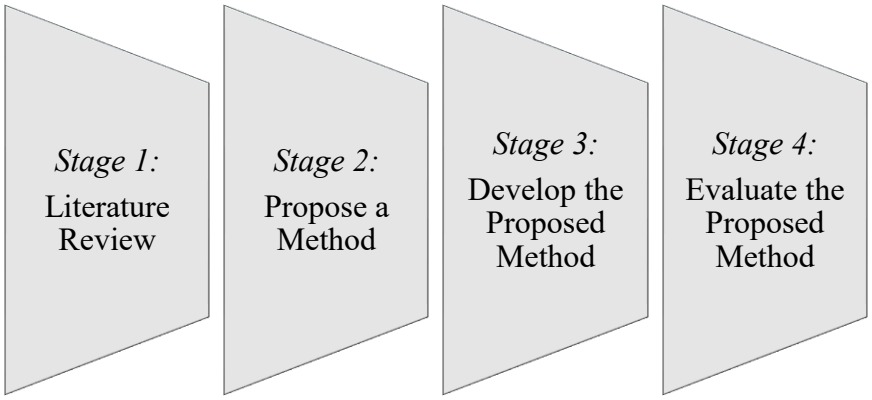

Figure 1: The Proposed Methodology

This study applies a four-stage methodology to address the problem of inaccurate segmentation in multi-modal MRI, particularly involving class imbalance, irregular tumor shape, and computational cost (see Figure 1). The methodology aligns with the stated research objectives and aims to provide a reliable method for segmenting brain tumors with focus on performance, efficiency, and evaluation.

The first stage involves a literature review to support Objective RO1. A systematic review is conducted based on PRISMA guidelines. The review includes CNN-based models such as U-Net, ResUNet, and Attention U-Net; Transformer-based models such as SwinUNet and TransUNet; and hybrid CNN-Transformer models. Studies on spatial and channel attention are also included. The review focuses on methods used to address tumor segmentation challenges such as low contrast, class imbalance, and generalization across datasets. The outcome is a synthesis of current methods, their structures, and their limitations.

The second stage addresses Objective RO2, which is to propose a method to handle class imbalance, spatial complexity, and segmentation boundary issues. A hybrid model is designed by combining a CNN encoder based on ResNet-50 and a Transformer backbone using Swin architecture. The model integrates two attention mechanisms: spatial attention to identify tumor areas in image space, and channel attention to adjust feature importance across MRI modalities. A cross-attention fusion unit integrates CNN and Transformer features at different levels. The method is configured to keep the parameter count low and reduce inference time, ensuring feasibility for deployment in systems with limited resources.

The third stage supports Objective RO3, which focuses on development and implementation. The proposed method is implemented using the PyTorch framework. Preprocessing steps include skull stripping, N4 bias correction, and z-score normalization. For large 3D MRI volumes, patch-based training is used. The loss function combines Dice loss and Focal loss to improve model learning on minority tumor classes. Augmentation techniques such as image rotation, flipping, and simulated artifacts are used to increase dataset diversity. The training process includes automatic mixed-precision computation and curriculum learning, which introduces increasingly complex tumor types in sequence to improve model robustness during training.

The fourth stage addresses Objective RO4, which is to evaluate the method using standard metrics and compare its effectiveness. The model is tested on the BraTS 2023 dataset and validated using TCIA data. Evaluation metrics include Dice Similarity Coefficient, Hausdorff Distance (HD95), Average Surface Distance (ASD), sensitivity, specificity, and per-slice inference time. Ablation experiments are performed to study the contributions of individual modules (CNN, Transformer, attention). The results are compared with baseline methods such as U-Net, nnU-Net, and Transformer-only models. The evaluation also includes performance under class imbalance and domain variation to assess the method's generalizability and clinical potential.

9. Proposed Method

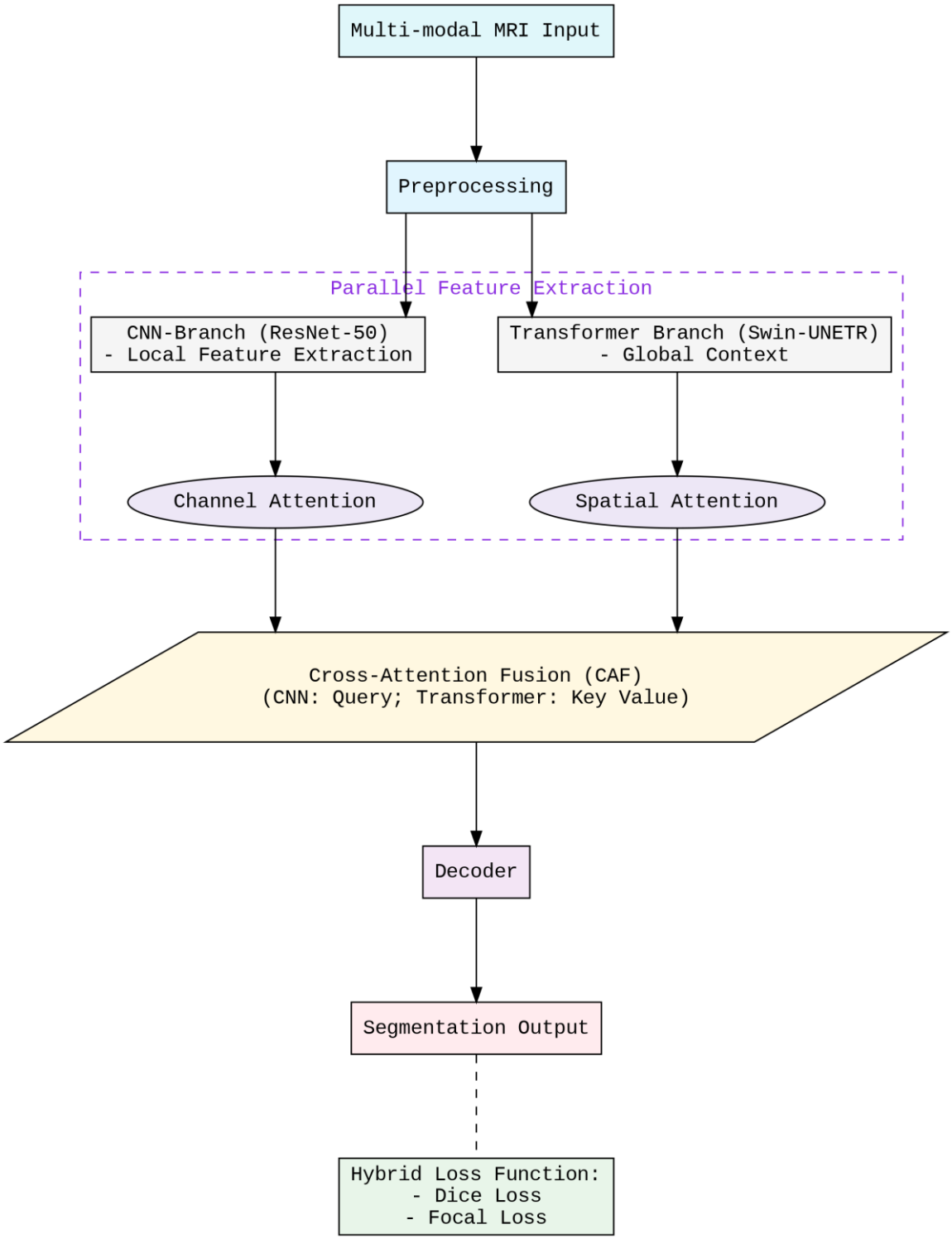

Figure 2: Proposed Method

Figure 2 shows the flow of the proposed method. The proposed method is designed to address the technical challenges outlined in the problem statement, namely class imbalance, spatial complexity, boundary inaccuracy, computational limitations, and limited generalizability across datasets. The method is also structured to fulfill Objectives RO2 and RO3, which involve proposing and developing a segmentation method that can effectively improve performance in multi-modal MRI brain tumor segmentation tasks.

The use of a CNN encoder, specifically ResNet-50, is justified by its ability to extract hierarchical spatial features and preserve low-level boundary information, which is essential for accurate tumor subregion delineation. CNNs are efficient in modeling local dependencies and are less sensitive to noise in high-resolution MRI scans, which makes them suitable for capturing small tumor regions that often suffer from class imbalance.

The integration of the Swin Transformer backbone is motivated by the need to model long-range spatial dependencies and global context, which are not adequately captured by CNNs alone. The Transformer enables the model to better understand the spatial distribution and contextual relationships of tumor tissues across slices and modalities, addressing spatial complexity and irregular tumor morphology.

The inclusion of spatial and channel attention mechanisms is justified based on their complementary roles in enhancing feature discrimination. Spatial attention directs the model to focus on tumor-relevant regions by weighting spatial positions, improving the identification of blurry or low-contrast boundaries. Channel attention adaptively emphasizes or suppresses modality-specific features based on their relevance, which is crucial in multi-modal MRI data where different sequences highlight different tumor characteristics. Together, these attention modules improve the model's ability to localize and classify tumor subregions.

The cross-attention fusion (CAF) mechanism is used to integrate CNN and Transformer features with attention signals. This fusion allows the method to leverage both local and global information simultaneously, which is essential for achieving consistent segmentation across heterogeneous tumor shapes and varying tissue contrasts. It also supports better alignment between encoder and decoder representations, which improves segmentation boundary accuracy.

Patch-based training and mixed-precision computation are included to address computational limitations, especially when dealing with high-resolution 3D MRI volumes. These techniques reduce memory requirements and processing time, making the method more applicable in settings with limited hardware resources.

The hybrid loss function, combining Dice loss and Focal loss, directly addresses the class imbalance issue. Dice loss improves overlap with small tumor regions, while Focal loss reduces bias toward dominant background classes by penalizing easy-to-classify voxels. This combination ensures better learning in scenarios where tumor voxels constitute a small portion of the image.

Lastly, the selection of evaluation metrics and datasets (BraTS and TCIA) is aligned with the need for both internal validation and external generalization. Metrics such as Dice coefficient, HD95, ASD, sensitivity, specificity, and inference time measure both accuracy and efficiency, supporting Objective RO4 and validating the method's potential for broader clinical or research deployment.

Table 3: Comparison Between Existing Methods and the Proposed Method

Aspect	**Existing Methods**	**Proposed Method**	**Proposed Component**	**Problem Addressed**
Architecture	CNN-only (e.g., U-Net) or Transformer-only (e.g., SwinUNet); limited hybrid integration	Hybrid CNN-Transformer using ResNet-50 and Swin Transformer	Hybrid CNN-Transformer architecture	Captures both local and global features to handle spatial complexity and irregular tumor morphology
Attention Mechanism	Spatial or channel attention used separately; sometimes not modality-aware	Dual spatial and channel attention; modality-aware attention design	Dual spatial and channel attention mechanisms	Enhances focus on relevant tumor regions and modality-specific cues to improve subregion segmentation
Feature Fusion	Simple skip connections; limited interaction between CNN and Transformer features	Cross-attention fusion (CAF) to integrate CNN and Transformer features dynamically	Cross-attention fusion layer (CAF)	Improves integration of features to reduce segmentation boundary errors

Handling Class Imbalance	Mostly use Dice loss only; weak on small tumor classes	Dice + Focal loss to improve learning on underrepresented tumor classes	Hybrid Dice + Focal loss function	Addresses class imbalance and ensures better learning from small tumor areas
Computational Efficiency	Large models (>100M parameters); high inference time	Model <50M parameters; supports patch-based and mixed-precision training	Lightweight design (<50M params), patch-based and mixed-precision training	Reduces training and inference cost, enabling use in resource-constrained environments
Generalizability	Evaluated only on BraTS or internal datasets	Validated on BraTS and TCIA to support domain generalization	Multi-dataset validation (BraTS + TCIA)	Ensures robustness across scanners and protocols, reducing domain shift effects
Clinical Usability	Boundary errors and missed subregions reduce clinical applicability	Improved boundary alignment and faster inference supports real-world use	Accurate boundary recovery with fast inference	Enables practical deployment by improving boundary accuracy and reducing latency

Table 3 provides a comparative analysis between existing brain tumor segmentation methods and the proposed hybrid deep learning method. The comparison is structured around key architectural and functional components, showing how each proposed design choice addresses a specific technical problem stated in the research.

Most existing methods either use CNN-based architectures, which focus on local feature extraction, or Transformer-based models, which emphasize global context. However, these approaches tend to fall short when applied to complex segmentation tasks involving irregular tumor morphology and low tissue contrast. In contrast, the proposed method integrates both CNN and Transformer components to leverage local and global information simultaneously, thus addressing spatial complexity more effectively.

Furthermore, attention mechanisms in current models are often limited to either spatial or channel dimensions, without modality-specific adaptation. The proposed method overcomes this limitation by incorporating dual attention mechanisms that operate in parallel, improving the model's ability to focus on diagnostically relevant regions across multiple MRI modalities.

To enhance feature interaction, the proposed method includes a cross-attention fusion (CAF) layer. Unlike basic skip connections used in existing models, the CAF layer dynamically integrates CNN and Transformer features, thereby improving boundary consistency and feature alignment across the network.

The method also addresses class imbalance, a persistent issue in tumor segmentation, by combining Dice loss and Focal loss in the optimization process. This loss function allows the model to focus on underrepresented tumor regions, improving segmentation performance across all tumor subregions.

Finally, the proposed method is designed with computational constraints in mind. By limiting the model size to under 50 million parameters and supporting patch-based training with mixed-precision computation, it ensures scalability and faster inference. The inclusion of multi-dataset evaluation on BraTS and TCIA further demonstrates the model's robustness and generalizability, which are often lacking in existing studies.

In summary, Table 3 illustrates that each component of the proposed method is purposefully selected to mitigate one or more of the key challenges outlined in the problem statement, thereby aligning the architecture with the research objectives.

10. Research Plan

The detailed plan of this study is illustrated in Figure 3.

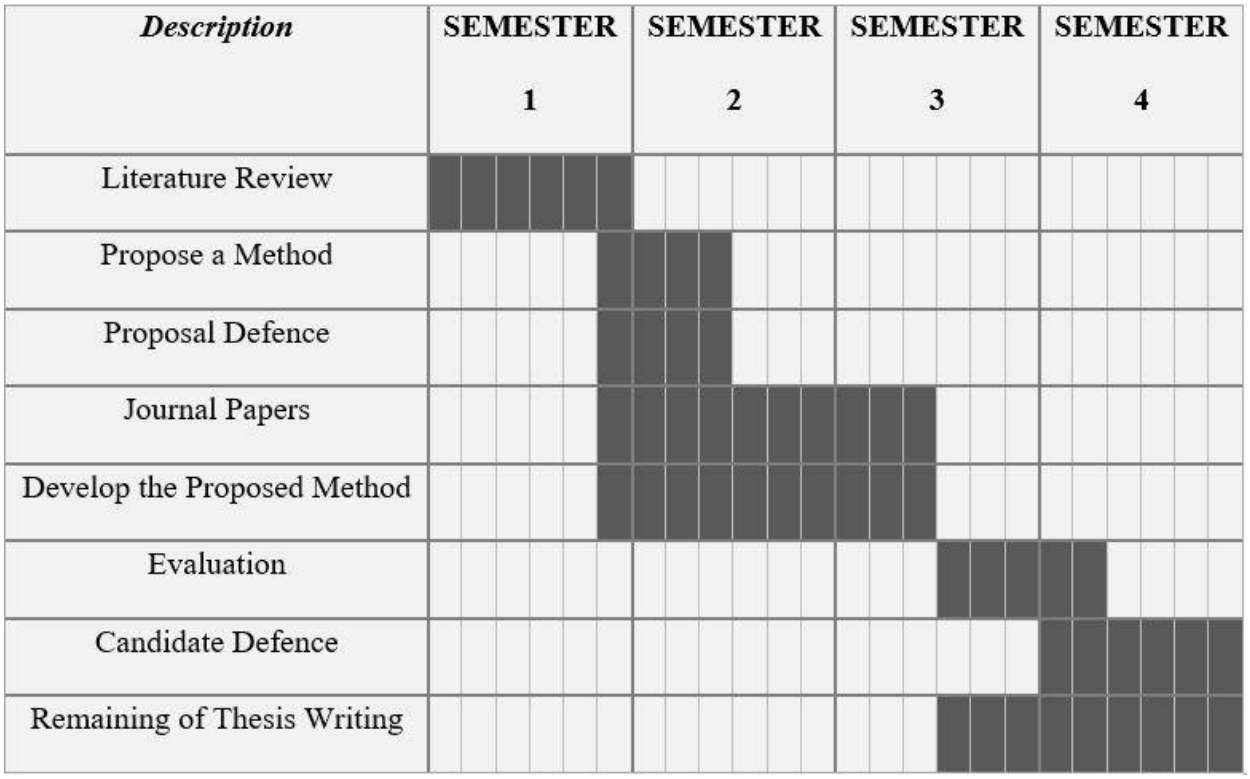

Figure 3: Research plan

11. Discussion and Conclusion

This study investigates a method for brain tumor segmentation using multi-modal MRI by integrating convolutional neural networks, Transformer modules, and attention mechanisms. The method is designed to address challenges such as class imbalance, spatial complexity, boundary inconsistency, and computational constraints. The components of the proposed method include a CNN-based encoder, a Swin Transformer backbone, spatial and channel attention mechanisms, and a cross-attention fusion layer.

The method is implemented using patch-based training, mixed-precision computation, and a hybrid loss function combining Dice loss and Focal loss. Evaluation is conducted on two datasets: BraTS 2023 and TCIA. Metrics used include Dice Similarity Coefficient, Hausdorff Distance (HD95), Average Surface Distance (ASD), sensitivity, specificity, and inference time per slice. The method is compared with baseline models such as U-Net, nnU-Net, and Transformer-based models to assess performance under different conditions.

Based on the results, the proposed method shows improvements in tumor subregion segmentation and boundary alignment, particularly for cases with irregular morphology or low contrast. The attention modules contribute to spatial localization and modality relevance. The hybrid loss function supports learning from underrepresented tumor regions. The use of cross-dataset validation demonstrates that the method can be applied to data from different sources.

The study is limited to segmentation tasks involving adult glioma cases and MRI data. Other tumor types, imaging modalities, and clinical settings are not included. The implementation is conducted using retrospective datasets and has not been deployed or tested in clinical environments. These limitations define the boundary of current findings.

In conclusion, the study presents a segmentation method that integrates multiple components to address specific challenges identified in the literature and problem statement. The results align with the research objectives and support further investigation in the development of segmentation methods for medical imaging.

12. References

Bello, I., Zoph, B., Vaswani, A., Shlens, J., & Le, Q. V. (2019). Attention augmented convolutional networks. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 3286-3295). https://doi.org/10.1109/ICCV.2019.00338

Bonato, B., Nanni, L., & Bertoldo, A. (2025). Advancing precision: A comprehensive review of MRI segmentation datasets from BraTS challenges (2012–2025). *Sensors, 25*(6), 1838. https://doi.org/10.3390/s25061838

Buddenkotte, T., Escudero Sanchez, L., Crispin-Ortuzar, M., Woitek, R., McCague, C., Brenton, J. D., Öktem, O., Sala, E., & Rundo, L. (2023). Calibrating ensembles for scalable uncertainty quantification in deep learning-based medical image segmentation. Computers in Biology and Medicine, 163, 107096. https://doi.org/10.1016/j.compbiomed.2023.107096

Chen, J., Mei, J., Li, X., Lu, Y., Yu, Q., Wei, Q., Luo, X., Xie, Y., Adeli, E., Wang, Y., Lungren, M. P., Zhang, S., Xing, L., Lu, L., Yuille, A., & Zhou, Y. (2024). TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers. Medical Image Analysis, 97, 103280. https://doi.org/10.1016/j.media.2024.103280

Chong, Y., Xie, N., Liu, X. et al. P-TransUNet: an improved parallel network for medical image segmentation. BMC Bioinformatics 24, 285 (2023). https://doi.org/10.1186/s12859-023-05409-7

Cui, H., Ruan, Z., Xu, Z., Luo, X., Zhang, Y., & Tian, Y. (2024). ResMT: A hybrid CNN-transformer framework for glioma grading with 3D MRI. Computers & Electrical Engineering, 120, 109745. https://doi.org/10.1016/j.compeleceng.2024.109745

Ding, Z., Zhang, Y., Zhu, C., Zhang, G., Li, X., Jiang, N., Que, Y., Peng, Y., & Guan, X. (2024). CAT-Unet: An enhanced U-Net architecture with coordinate attention and skip-neighborhood attention transformer for medical image segmentation. Information Sciences, 670, 120578. https://doi.org/10.1016/j.ins.2024.120578

Gao, Y., Zhang, J., Wei, S., & Li, Z. (2025). PFormer: An efficient CNN–Transformer hybrid network with content-driven P-attention for 3D medical image segmentation. Biomedical Signal Processing and Control, 101, 107154. https://doi.org/10.1016/j.bspc.2024.107154

Ghazouani F, Vera P, Ruan S. Efficient brain tumor segmentation using Swin transformer and enhanced local self-attention. Int J Comput Assist Radiol Surg. 2024 Feb;19(2):273-281. https://doi.org/10.1007/s11548-023-03024-8

Han, Y., Wang, L., Huang, Z., Zhang, Y., & Zheng, X. (2025). A novel 3D magnetic resonance imaging registration framework based on the Swin-Transformer UNet+ model with 3D dynamic snake convolution scheme. Journal of Imaging, 11(2), 54. https://doi.org/10.3390/jimaging11020054

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., & Liang, J. (2022). UNETR: Transformers for 3D medical image segmentation. *In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 1748–1758). IEEE. https://doi.org/10.1109/WACV51458.2022.00181

Heidari, M., Kazerouni, A., Soltany, M., Azad, R., Aghdam, E. K., Cohen-Adad, J., & Merhof, D. (2023). Hiformer: Hierarchical multi-scale representations using transformers for medical

image segmentation. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 6202-6212). https://doi.org/10.1109/WACV56688.2023.00614

Huang, J., Yagmurlu, B., Molleti, P., Lee, R., VanderPloeg, A., Noor, H., Bareja, R., Li, Y., Iv, M., & Itakura, H. (2025). Brain tumor segmentation using deep learning: High performance with minimized MRI data. Frontiers in Radiology, 5, 1616293. https://doi.org/10.3389/fradi.2025.1616293

Huang, L., Zhu, E., Chen, L., Wang, Z., Chai, S., & Zhang, B. (2022). A transformer-based generative adversarial network for brain tumor segmentation. Frontiers in Neuroscience, 16, 1054948. https://doi.org/10.3389/fnins.2022.1054948

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, 203–211. https://doi.org/10.1038/s41592-020-01008-z

Jungo, A., Meier, R., Ermis, E., Blatti-Moreno, M., Herrmann, E., Wiest, R., & Reyes, M. (2018, September). On the effect of inter-observer variability for a reliable estimation of uncertainty of medical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention - MICCAI 2018 (pp. 682-690). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-00928-1_77

Kharaji, M., Abbasi, H., Orouskhani, Y., Shomalzadeh, M., Kazemi, F., & Orouskhani, M. (2024). Brain tumor segmentation with advanced nnU-Net: Pediatrics and adults tumors. Neuroscience Informatics, 4(2), 100156. https://doi.org/10.1016/j.neuri.2024.100156

Li, J., Xu, Q., He, X., Liu, Z., Zhang, D., Wang, R., Qu, R., & Qiu, G. (2025). CFFormer: Cross CNN-Transformer channel attention and spatial feature fusion for improved segmentation of heterogeneous medical images. Expert Systems with Applications, 295, 128835. https://doi.org/10.1016/j.eswa.2025.128835

Liu, F., Zhang, Y., Lu, T., Wang, X., Chen, J., & Li, S. (2025). Hierarchical in-out fusion for incomplete multimodal brain tumor segmentation. Scientific Reports, 15, 23017. https://doi.org/10.1038/s41598-025-07466-9

Liu, J. et al. (2024). Swin-UMamba: Mamba-Based UNet with ImageNet-Based Pretraining. In: Linguraru, M.G., et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. MICCAI 2024. Lecture Notes in Computer Science, vol 15009. Springer, Cham. https://doi.org/10.1007/978-3-031-72114-4_59

Liu, X., Gao, P., Lin, K., Wang, F., & Yuan, R. (2025). CSWin-UNet: Transformer UNet with cross-shaped windows for medical image segmentation. Information Fusion, 113, 102634. https://doi.org/10.1016/j.inffus.2024.102634

Luo, X., Yang, Y., Yin, S., Li, H., Shao, Y., Zheng, D., Li, X., Li, J., Fan, W., Li, J., Ban, X., Lian, S., Zhang, Y., Yang, Q., Zhang, W., Zhang, C., Ma, L., Luo, Y., Zhou, F., Wang, S., Lin, C., Li, J., Luo, M., He, J., Xu, G., Gao, Y., Shen, D., Sun, Y., Mou, Y., Zhang, R., & Xie, C. (2024). Automated segmentation of brain metastases with deep learning: A multi-center, randomized crossover, multi-reader evaluation study. Neuro-Oncology, 26(11), 2140–2151. https://doi.org/10.1093/neuonc/noae113

Lyu, Y., & Tian, X. (2025). MWG-UNet++: Hybrid Transformer U-Net model for brain tumor segmentation in MRI scans. Bioengineering, 12(2), 140. https://doi.org/10.3390/bioengineering12020140

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., Lanczi, L., Gerstner, E., Weber, M. A., Arbel, T., Avants, B. B., Ayache, N., Buendia, P., Collins, D. L., Cordier, N., ... & 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

Mosquera, C., Ferrer, L., Milone, D. H., Luna, D., & Ferrante, E. (2024). Class imbalance on medical image classification: Towards better evaluation practices for discrimination and calibration performance. European Radiology, 34(12), 7895–7903. https://doi.org/10.1007/s00330-024-10834-0

Płotka, S., Chrabaszcz, M., Biecek, P. (2024). Swin SMT: Global Sequential Modeling for Enhancing 3D Medical Image Segmentation. In: Linguraru, M.G., et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. MICCAI 2024. Lecture Notes in Computer Science, vol 15008. Springer, Cham. https://doi.org/10.1007/978-3-031-72111-3_65

Preetha, R., Priyadarsini, J. P., & Nisha, J. S. (2025). Brain tumor segmentation using multi-scale attention U Net with EfficientNetB4 encoder. Scientific Reports, 15, Article 9914. https://doi.org/10.1038/s41598-025-94267-9

Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science(), vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

Saifullah, S., Dreżewski, R., Yudhana, A., Nurhadiyatna, A., & Wicaksana, A. (2025). Modified U-Net with attention gate for enhanced automated brain tumor segmentation. Neural Computing and Applications, 37, 5521–5558. https://doi.org/10.1007/s00521-024-10919-3

Shoushtari, F. K., Elahi, R., Valizadeh, G., Moodi, F., Salari, H. M., & Rad, H. S. (2025). Current trends in glioma tumor segmentation: A survey of deep learning modules. Physica Medica, 135, 104988. https://doi.org/10.1016/j.ejmp.2025.104988

Sun, G., Pan, Y., Kong, W., Xu, Z., Ma, J., Racharak, T., Nguyen, L.-M., & Xin, J. (2024). DA-TransUNet: Integrating spatial and channel dual attention with Transformer U-Net for medical image segmentation. Frontiers in Bioengineering and Biotechnology, 12, Article 1398237. https://doi.org/10.3389/fbioe.2024.1398237

Tang, Y., Yang, D., Li, W., Roth, H. R., Landman, B. A., Xu, D., Nath, V., Hatamizadeh, A., Molchanov, P., & Anandkumar, A. (2022). Self-supervised pre-training of Swin Transformers for 3D medical image analysis. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 20698–20708). IEEE. https://doi.org/10.1109/CVPR52688.2022.02007

Urrea, C., & Vélez, M. (2025). Advances in deep learning for semantic segmentation of low-contrast images: A systematic review of methods, challenges, and future directions. Sensors, 25(7), 2043. https://doi.org/10.3390/s25072043

Veiga-Canuto, D., Cerdà-Alberich, L., Sangüesa Nebot, C., Martínez de Las Heras, B., Pötschger, U., Gabelloni, M., ... & Martí-Bonmatí, L. (2022). Comparative multicentric evaluation of inter-observer variability in manual and automatic segmentation of neuroblastic tumors in magnetic resonance images. Cancers. 2022; 14: 3648. https://doi.org/10.3390/cancers14153648

Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). CBAM: Convolutional block attention module. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision – ECCV 2018* (Lecture Notes in Computer Science, Vol. 11211, pp. 3–19). Springer, Cham. https://doi.org/10.1007/978-3-030-01234-2_1

Xiao, L., Zhou, B., & Fan, C. (2025). Automatic brain MRI tumors segmentation based on deep fusion of weak edge and context features. Artificial Intelligence Review, 58, 154. https://doi.org/10.1007/s10462-025-11151-8

Xiong, L., Yi, C., Xiong, Q., & Jiang, S. (2024). SEA-NET: Medical image segmentation network based on spiral squeeze-and-excitation and attention modules. BMC Medical Imaging, 24(1), 17. https://doi.org/10.1186/s12880-024-01194-8

Yoon, J. S., Oh, K., Shin, Y., Mazurowski, M. A., & Suk, H.-I. (2024). Domain generalization for medical image analysis: A review. Proceedings of the IEEE, 112(10), 1583–1609. https://doi.org/10.1109/JPROC.2024.3507831

Zarenia, E., Far, A. A., & Rezaee, K. (2025). Automated multi-class MRI brain tumor classification and segmentation using deformable attention and saliency mapping. Scientific Reports, 15, 8114. https://doi.org/10.1038/s41598-025-92776-1

Zhang, W., Chen, S., Ma, Y., Liu, Y., & Cao, X. (2024). ETUNet: Exploring efficient transformer-enhanced U-Net for 3D brain tumor segmentation. Computers in Biology and Medicine, 171, 108005. https://doi.org/10.1016/j.compbiomed.2024.108005

Zhong, Y., Wang, S., Miao, Y., Zhang, T., & Li, H. (2025). Lightweight brain tumor segmentation through wavelet-guided iterative axial factorization attention. Brain Sciences, 15(6), 613. https://doi.org/10.3390/brainsci15060613

Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2021). UNet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE Transactions on Medical Imaging, 39(6), 1856–1867. https://doi.org/10.1109/TMI.2019.2959609

Zhu, J., Zhang, R., & Zhang, H. (2024). An MRI brain tumor segmentation method based on improved U-Net. Mathematical Biosciences and Engineering, 21(1), 778–791. https://doi.org/10.3934/mbe.2024033