Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

DSC-SwinNet: A Dual-Stage Transformer Framework for Reliable Brain Tumor Segmentation and Classification from Multi-Modal MRI

Tamilselvi M.

Department of Electronics and Communication Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, (SIMATS), Chennai, India.

E-mail: tamilselvivlsi@gmail.com

Abstract

The earlier diagnosis of brain tumors is a critical challenge that influences treatment and facilitates prompt detection of the disease. Conventional MRI provides a structural and functional view of the tumors. On the other hand, recent deep learning-based algorithms, particularly single-stage convolutional neural network-based models, face challenges in providing the exact location of the tumor as well as in enhancing detection and classification accuracy. This is due to a lack of global-local integration of features, lack of spatial consistency, and low resistance to intensity variation, which are typical of clinical MRI scans. In order to address these gaps, the proposed research uses the DSC-SwinNet algorithm, which consists of a dual-stage transformer structure primarily utilized for tumor segmentation and classification. The first step employs a Swin Transformer-based encoder-decoder that uses window-based multi-head self-attention to simultaneously obtain local lesion features and long-range global contextual features of multi-modal MRI volumes. The next stage, known as Dual-stage Classification (DSC), is responsible for incorporating the ROI characteristics with conceptual representations of the tumor to identify the type of tumor. The proposed DSC-SwinNet has a Dice score of 0.934, an IoU of 0.891, an HD95 of 3.70 mm, achieving a classification accuracy of 97.8%, an F1-score of 97.9%, and an AUC of 0.99 on the BraTS multi-modal MRI data, demonstrating the potential of DSC-SwinNet as a clinically reliable brain tumor analyzer.

Keywords: DSC-SwinNet, Transformer Framework, Dual-Stage Classification, Brain Tumor, Multi Modal MRI, Convolutional Neural Network, Global-Local Feature, Disease Classification.

1. Introduction

Brain tumors are among the most threatening neurological conditions that can cause irreversible cognitive loss, disability, and death if not diagnosed at an early stage [1]. Proper tumor delineation, classification, and detection are critical in the decision-making process of treatment, surgical planning, radiation therapy dose assessment, and monitoring disease progression. MRI is the most effective non-invasive technique to evaluate brain tumors because it provides the best soft-tissue contrast, radiation-free imaging, and multi-modal sequences like T1, T1ce, T2, and FLAIR. Such complementary modalities assist in visualizing various tumor

DOI: https://doi.org/10.36548/jiip.2026.1.006
Received: 02.12.2025, received in revised form: 30.12.2025, accepted: 13.01.2026, published: 24.01.2026

morphology components, such as edema, necrotic and enhancing areas, making MRI essential in neuro-oncology [2].

Brain tumors are one of the most complicated diseases in the neurological spectrum that require prompt treatment after diagnosis. The morphology and structure of brain tumors, their heterogeneity, and irregularities are not homogeneous, which makes this task challenging for even experienced radiologists to segment tumors based on MRI scans. Additionally, inconsistencies in imaging modalities and types of tumors make manual annotations difficult to implement, leading to inconsistencies in the output. Since timely diagnosis is essential in enhancing survival chances, there is an increasing need for automated, accurate, and efficient segmentation tools in the medical sector [3].

Magnetic resonance imaging (MRI) is a non-invasive technique for visualizing soft tissues in high resolution; it is a critical diagnostic tool for identifying and determining brain tumors. Nevertheless, interpreting MRI data is highly skilled and labor-intensive, particularly in hospitals with overburdened radiologists. This problem can be alleviated with automated image segmentation systems, which reduce human error and expedite diagnosis. Although deep learning has played a critical role in the development of segmentation, models using only convolutional operations have a limited range of contextual fields of view, making it difficult to outline diffuse or overlapping tumors [4].

Gliomas are the most prevalent primary brain tumors, often referenced in discussions about brain tumors. They originate in the cells that constitute the support tissue of the brain, known as glial cells. The interplay between high-risk genetic factors (congenital) and environmental carcinogenic factors contributes to the development of gliomas. Clinically, gliomas are threatening and fatal tumors of the brain, characterized by high malignancy and aggressiveness, leading to various symptoms, including seizures, headaches, visual disturbances, and alterations in behavior and speech. Generally, the localization, shape, and size of brain tumors have significant implications for the extent and nature of these symptoms, as identified by physicians, and influence the development of treatment and surgical strategies [5].

Thus, brain tumor segmentation can facilitate the precise and efficient localization and identification of gliomas, which would, in turn, assist physicians in enhancing the diagnosis and prognosis in clinical practice [6]. Over the past decades, scholars have conducted extensive basic research on brain tumors. Initial studies aimed to understand the biological characteristics of glial cells and how they become malignant. Gradually, scientists have gained insights into the genetic and molecular alterations that occur in gliomas. This research has facilitated the development of improved diagnosis and treatment for brain tumors, including the identification of brain tumor grading, heredity, and targeted therapy using genomic information [7].

As MRI technology has advanced, multi-modal MRI images have become increasingly popular in the process of brain tumor segmentation, providing a more detailed view of the tumors and surrounding brain tissues. Practically, MRI has four modalities—T1, T2, T1ce, and FLAIR—that serve as complementary imaging modalities in the diagnosis and monitoring of brain tumors. Different MRI modalities can complement each other regarding the appearance and characteristics of tumors [8],[9].

Medical image segmentation is a task that is significant in the diagnosis of clinical imaging. Physicians tend to use alternative treatments, including surgery, radiotherapy, or chemotherapy, depending on the type, size, and position of the tumor. It is not a simple task to

view different states of brain tumors directly using computer equipment, and it is even more challenging when it comes to identifying the type of tumor, its size, and location. Thus, studies on brain tumor segmentation algorithms attempt to provide a more objective assessment and explanation of the development, pathology, clinical phenotypes, and prognostic factors related to brain tumors [10].

In the last several years, the rapid advancement of deep learning systems has succeeded in enhancing the output of computer-aided diagnosis. Multi-modal brain tumor segmentation has seen significant technical progress, resulting in an increasing number of techniques that can perform this task with acceptable accuracy and speed. The first and most basic approaches to brain tumor segmentation include manual tracing, in which a skilled clinical practitioner outlines the tumor in the images. Manual tracing, however, is time-consuming and may be prone to inter- and intra-observer variation. Due to the introduction of computer vision and machine learning algorithms, numerous automatic approaches have been developed to segment brain tumors. These techniques can be broadly divided into two groups: traditional techniques and deep learning techniques [11].

Convolutional Neural Networks (CNNs) have been widely used in brain tumor detection tasks, and the application of deep learning has completely changed the field of medical image analysis. Architectures such as VGG19, ResNet152V2, DenseNet201, InceptionResNetV2, and EfficientNetV2L have demonstrated significant gains in quality pattern recognition, feature extraction, and classification accuracy. However, CNNs are formulated based on local receptive fields and thus cannot model long-range spatial dependencies, which are needed for irregular tumor structures in 3D MRI volumes. Consequently, CNN-based segmentation and classification models usually do not generalize well between patients, scanners, and changes in MRI intensity [12].

New powerful alternatives have also appeared, such as Vision Transformers (ViT) and Swin Transformers, capable of modeling long-range relationships with the help of self-attention mechanisms (Tazeen et al., 2024). Hybrid systems like Swin-UNER have already shown tremendous improvements in volumetric medical image segmentation as they have shown better results on tumor delineation tasks. Regardless of these advances, the current transformer-based models are mainly segmentation-based or classification-based and have no unified pipeline to deliver accurate tumor boundaries and robust diagnostic classification. Furthermore, unsupervised classification models that are not explicitly localized on tumors tend to inappropriately distinguish irrelevant or high-noise regions of the brain, diminishing trust and accuracy of clinical implementation [13].

To overcome these constraints, this study suggests the development of DSC-SwinNet; a novel Dual-Stage Transformer Framework combining transformer-based multi-modal 3D segmentation and a powerful dual-stage classification scheme specific to multi-modal MRI. The framework employs a Swin Transformer enhanced encoder-decoder that achieves high accuracy in the segmentation of tumor using local ROI features of regions that are segmented and global contextual volume features. This will allow classification decisions to be based on tumor-centric information as well as holistic brain-scale information (homogenizing the weaknesses of single-stage traditional models) [14].

In this work, a new dual-stage transformer-based framework that closely combines high-resolution tumor segmentation with context-aware tumor classification within a single pipeline is proposed. Compared to the current methods that either consider segmentation and classification as independent or sequential, the presented method applies segmentation-directed

ROI extraction coupled with global volumetric contextual embeddings, which allows tumor-based but global-informed diagnostic inference. This combined design is much more effective in segmentation, classification accuracy, and probabilistic calibration to multi-modal MRI variability, and its advancement in the state of the art in clinically robust brain tumor analysis.

The contributions of this research are of a great importance and they include:

- Creation of an innovative architecture, which incorporates Swin Transformer-based 3D segmentation, and a two-stage 2D Classification mechanism to run end-to-end tumor analysis.
- Developing a DSC module that combines tumor-centered ROI characteristics with global MRI volume characteristics to obtain excellent discrimination.
- Developing the Swin Transformer encoder has hierarchical self-attention to achieve better localization of tumors and accurate classification.
- Introducing the T1, T1ce, T2 and FLAIR capabilities to deal with tumor heterogeneity and enhance generalization across different clinical conditions.
- To show performance improvement, comparative analysis to VGG19, ResNet152V2, DenseNet201, InceptionResNetV2, EfficientNetV2L, and ConvNeXt was done.
- Integrating the uncertainty modeling, scores on calibration and noise-based robustness testing to demonstrate clinical preparedness.
- Evaluation of standardized multi-modes MRI data to perform both segmentation and classification.

The rest of this paper is structured in the following way. In Section II, a thorough overview of the already available literature on the topic of brain tumor segmentation, classification schemes, transformer-based medical imaging models, and performance evaluation on the BraTS database is provided. Section III is a report on the proposed DSC-SwinNet architecture, its Swin Transformer segmentation backbone, the dual-stage classification pipeline, the strategy of local-global feature fusion and the overall training setup of volumetric multi-modal MRI analysis. Section IV presents the experimental results and performance comparison of the quantitative metrics of segmentation, diagnostic classification, and ablation experiments, the calibration reliability, the estimation of uncertainty, the robustness, and the statistical testing against the state-of-the-art models. Lastly, Section V is the conclusion of the work that summarizes important contributions, thinks over the implications of clinical considerations, and indicates possible extensions of the future research on the use of transformers to analyze medical images and apply them in actual neuro-oncology setting

2. Related Work

Computer-aided medical diagnostic systems have made medical image segmentation, especially that involving MRI analysis of brain tumors, a critical component, as developments in computational intelligence and deep neural architecture have rapidly advanced. The conventional methods of segmentation, which were intensity similarity, atlas guidance, level-

set models, and region-growing methods, did not prove very robust against the heterogeneous structure, size, and texture of tumors. CNNs, FCNs, and U-Net are among the architectures that transformed the quality of segmentation with the advancement of deep learning, as these methods learn both discriminative and hierarchical feature representations directly on multi-modal MRI data [15].

This development has also been further pushed by the BraTS challenges, which established a standardized benchmark and spurred widespread architectural developments such as skip-based residual networks, dense connectivity, nested U-Net variants, attention, and hybrid schemes combining convolutional modules with transformer-based global dependency modeling. The current trends in research also extend to diffusion-based segmentation models, modality-fusion transformers, and ensemble learning frameworks, which represent an ongoing quest to achieve greater accuracy, improved generalization, and clinically robust demarcation of brain tumor sub-regions [16]. This has led to the field of research on medical image segmentation becoming a growing area with the fast advancement of computer technology and computer-aided diagnostic systems. Medical image segmentation has been made possible by progress in the area of machine learning and deep learning [17].

Transformer-based networks were, in turn, proposed as an alternative with great potential, as they were effective in natural language processing. Vision Transformers (ViT) and various variants of ViT use self-attention to learn global relationships in images [18]. Transformer capabilities in learning long-range dependencies in medical segmentation activities are emphasized. It introduces DenseTrans, a hybrid model that incorporates Swin Transformer and UNet++, which is currently scoring high in Dice on BraTS2021 [19].

Tumor classification has been done using convolutional neural network (CNN) models, including DeepMedic and U-Net. U-Net was popularly applied in the segmentation of brain tumors. An improved network topology known as U-Net was introduced; it consists of several encoders and decoders, which produce more feature points to enable accurate segmentation. One of the recent topics in research on computer vision is the diffusion probability model (DPM) [20].

Based on the success of Transformers in several NLP tasks, an increasing number of Transformer-based approaches are being introduced in CV tasks. ViT is the first pure Transformer-based architecture that has demonstrated SOTA performance in image recognition when pre-trained on large datasets like ImageNet-22K, using data-efficient training methods and knowledge distillation that enable ViT to be effective on the smaller ImageNet-1K dataset. Swin Transformer is a linear model with a proposed shifted window-based self-attention mechanism and has SOTA performance in image recognition and dense prediction tasks, including object detection and semantic segmentation [21].

The vanilla Transformer treats every position of the image equally, but to minimize computational costs and pay attention to specific parts of the image, a different attention mechanism is presented whereby only portions of the key around a reference point are taken into consideration by the self-attention mechanism. In order to segment 3D images, an algorithm that learns representations of the input through the assistance of a Transformer as the encoder is suggested [22].

Ensembles of U-Net-shaped architectures have yielded encouraging results in multi-modal brain tumor segmentation in the past BraTS challenges. They suggest a strong segmentation model by combining the results of multiple CNN-based models, including 3D U-

Net, 3D FCN, and Deep Medic. Then, SegResNet is presented, which is a residual encoder-decoder model with an auxiliary branch supported by a variational auto-encoder to reconstruct the input data as a surrogate task [23].

Various factors, such as the capability to acquire long spatial dependencies, resistance to changes in intensity between the MRI modalities, the ability to delineate irregular tumor edges precisely, and the capacity to discriminate tumor types, are critical in determining the choice of the proposed technique. Traditional CNN-based designs have local receptive fields, and single-stage transformer designs typically do not have spatial resolution. Thus, a dual-stage, segmentation-directed, hybrid transformer architecture is implemented to balance global reasoning of the context and feature selection from a more accurate tumor-localized perspective, ensuring robust and clinically valid performance.

The challenges of BraTS (Brain Tumor Segmentation) have been a landmark in assessing AI-based segmentation techniques, inspiring innovation in the domain. Conventional machine learning methods, though initially effective, struggled to keep up with the heterogeneous appearance of tumors in multi-modal MRI datasets of BraTS, resulting in poor Dice scores. The advent of deep learning, especially variants of U-Net, significantly enhanced performance, as it automatically learned features that were discriminative between T1, T2, FLAIR, and T1ce sequences. Later versions of BraTS saw transformer-based models, such as Swin UNETR, go even further with global context modeling, while diffusion models were employed to detect edges in tumor sub-regions even more effectively [24].

Over the last several years, there has been an increase in the use of deep learning algorithms, especially Convolutional Neural Networks (CNNs), in brain tumor segmentation. Large volumes of annotated data can be used to train CNNs to learn complex image features, thus enabling them to perform better than traditional methods. Indicatively, the U-Net architecture is one widely used deep learning architecture for segmenting brain tumors and is based on the encoder-decoder architecture, where high-level and low-level image characteristics are learned. Most recently, vision transformers have seen remarkable advancements and deliver better results in the segmentation of brain tumors [25][26].

The CNN-based variants of UNet are still predominant due to their effective encoder-decoder representation, dense skip connections, and superior spatial preservation, whereas transformer-based variants and hybrid CNN-ViT architectures have developed as influential alternatives in order to surpass CNN in its limited receptive field through global self-attention. Diffusion models, attention fusion schemes, cascaded architectures, ensemble learning, and nnU-Net-style auto-configurations are additional examples of the type of progress provoked by BraTS benchmarking. Although these advancements have been achieved, the key open challenges include the improvement of boundary segmentation fidelity, enhanced data imbalance robustness, lightweight privacy-preserving model design, and clinically interpretable predictions. All of these research findings will encourage the development of more robust, generalizable, and computationally efficient solutions to brain tumor segmentation in a real-world healthcare setting.

3. Proposed Work

The proposed framework of the DSC-SwinNet model is a dual stage pipeline that not only providing the accurate segmentation of tumor but also ensures consistency in tumor prediction based on the MRI input images. Initially, four classifications of MRI images are

considered as shown in Figure 1. The Segmentation Network processes this multi-modal input, and a Swin Transformer-based encoder-decoder system extracts hierarchical 3D characteristics.

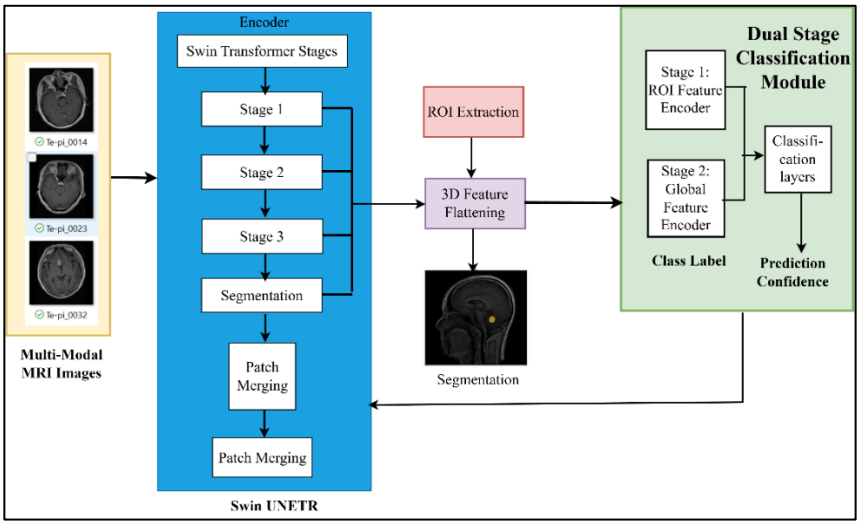

Figure 1. Block Diagram of the Proposed Framework

The MRI modalities are all considered for the normalization process to obtain the Z-score. The standard deviation and the mean of each mode are computed using non-background voxels only and the intensity values are brought to the normal level. This modality-wise normalization preserves the contrast properties of any given sequence, but removes scanner induced changes in intensities and inter-subject variations, therefore, it is much stronger in acquiring multi-modal features.

The input multi modal volumes of MRI are denoted as $X = \{X^{(T1)}, X^{(T1ce)}, X^{(T2)}, X^{(FLAIR)}\}$ for every patient in a 3D array of size H x W x D. The input images are cropped to the size of 128 x 128 x 128 and few augmentation changes are made on the input images before sending them for the next stage. The equation 1shown below gives the expression of the preprocessing pipeline, which is the intensity normalization of the z-score per volume.

$$\tilde{X} = \frac{X - \mu_X}{\sigma_X} \tag{1}$$

In the equation 1, non background voxels are denoted as μ_X and σ_X respectively.

During preprocessing, the data is subjected to Z-score normalization that performed on a volume basis to normalize the distributions of the intensity and reduce inter-subject variability, voxel intensities are normalized by subtracting the mean and dividing by the standard deviation of non-background voxels in each volume of MRI. This normalization increases the numerical stability of the training as well as the consistency of the magnitude of the scaling of the samples without destroying the inherent structural information of the brain tissues.

The segmentation module uses a four-stage Swin Transformer encoder where stage 1-4 study contextual information on a multi-scale basis with window self-attention and shifted window self-attention respectively. Features are extracted at a broader receptive field at each stage, accommodating long-range spatial dependencies, which can be regarded as one of the significant benefits of transformer architectures over conventional CNNs. The decoder

recreates the tumor masks at the voxel level using upsampling and skip connections in each encoder stage, ensuring that spatial lines are not dropped. This module provides the position of the tumor in the brain, resulting in a volume tumor segmentation mask.

To cope with inter-patient variation in the size and shape of the tumor, the segmented tumor area is cropped with a tight bounding box and resized to a constant ROI size. Owing to this normalization, feature learning is scale-invariant, the training is stable across batches, and classification is performed with consistent tensor dimensions, while important tumor morphology is maintained. The method prevents size bias and allows only tumor-centric features to be compared between subjects, resizing them only after proper segmentation.

Let us assume that the multi modal patch input is $X_0 \epsilon \mathbb{R}^{NxC}$ which is implemented after the embedding of patches. The proposed Swin Transformer blocks the operation of 3D windowed multi head self attention with the integration of shifted windows. Equation 2 gives the formula of attention for the set of tokens that are represented by the $X_0 \epsilon \mathbb{R}^{nxd}$ inner side of the window. In order to effectively represent both fine-grained local tumor features and long-range spatial features in multi-modal MRI volumes, the Swin Transformer encoder uses a window-based multi-head self-attention mechanism with shifted windows, and has the following mathematical formulation:

$$\text{Attention (Q, K, V)} = softmax\left(\frac{QK^T}{\sqrt{d_k}} + M\right)V \quad (2)$$

In the above equation, parameter Q can be calculated as XW_Q , K can be calculated as XW_K and the V can be calculated as XW_V . The head dimension is represented in equation 2 as d_k and the metric M is used to denote the shifted window mask. The multiple stages in the encoder are merged with the patches among the stages so that the corresponding feature maps are generated and they are denoted as E1, E2, E3 and E4 respectively.

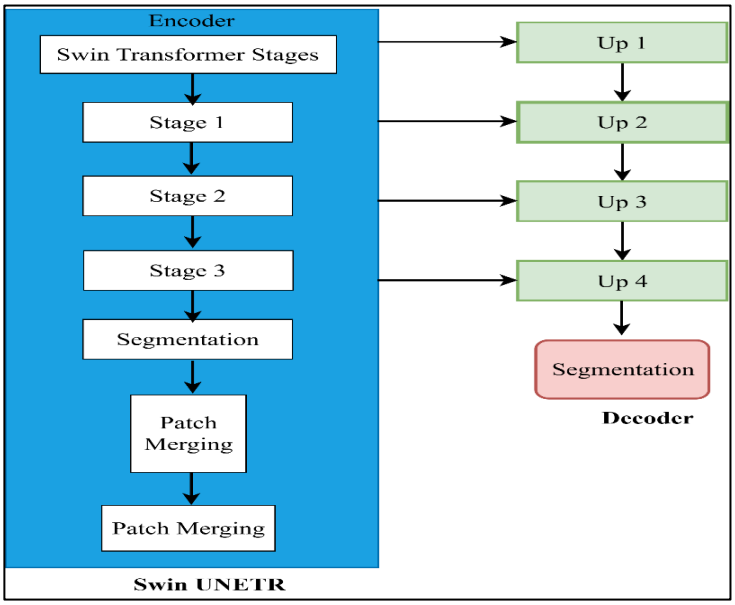

Figure 2. The Architectural Diagram of the Proposed Swin UNETR Model

In this research, the Swin-UNETR architecture was employed, combining a hierarchical Swin Transformer encoder and a U-Net type decoder to provide precise volumetric segmentation of brain tumors using multi-modal MRI. Figure 2 illustrates, the encoder is initially starts with a series of Swin Transformer stages each generating successively higher

order feature representations based on window-based multi-head self-attention and shifted-window mechanisms.

The Swin Transformer attention window size has moderate sensitivity to the performance of the proposed model. Smaller window sizes limit global contextual modeling, resulting in a lack of boundary consistency and decreased segmentation precision, while larger windows are not associated with commensurate improvements in computational complexity. The choice of window size offers the best compromise between long-range dependency modeling and computational efficiency, yielding stable and consistent results in both segmentation and classification challenges.

These attention processes allow the network to learn fine-grained local and long range spatial dependencies in the 3D MRI volume - a required property for discovering heterogeneous tumor regions. Multi-scale semantic information is stored in feature maps obtained at various levels of the encoder and is exploited by the decoder through skip connections. Following Stage 3, the encoded features are patch merged, i.e. reduced spatial resolution and increased depth per channel, allowing the network to effectively encode high-level tumor morphology and global context.

The suggested model will be able to generalize to other scanner manufacturers and acquisition procedures by utilizing modality-wise intensity normalization, widespread data augmentation, and transformer-based global contextual modeling. Independent normalization of Z scores eliminates intensity variation based on the scanner and simulates protocol based distortions like noise and contrast changes. Moreover, hierarchical self-attention in the Swin Transformer provides the global anatomical context, which also allows the model to be resistant to changes in resolution and contrast, along with acquisition conditions that are often present in multi-center clinical MRI data.

At the receiver end, the architecture uses a series of upsampling steps (Up1 to Up4) and each upsampling step restores the spatial resolution while the features of the encoder are merged through skip connectivity. Such a combination of rich semantic content and previous high resolution representations assists the network in maintaining anatomical boundaries, and reinstating fine structural details that are lost due to downsampling. The last decoder layer generates a 3D segmentation map that indicates tumor sub-regions at the voxel-scale. Swin-UNETR offers a robust segmentation backbone as it integrates the global reasoning capabilities of transformers with the spatial restoration ability of U-Net.

Equation 3 explains the decoding, sampling, and fusing with the respective encoder through the skip connections. After hierarchical feature extraction by the Swin Transformer encoder, the decoder gradually restores the spatial resolution by combining high-level semantic features with the corresponding encoder representations via skip connections, allowing for the precise reconstruction of tumor edges, as stated:

$$D_j = UpConv\left(D_{j-1}\right) \oplus E_{L-j} \tag{3}$$

Where the operator $\oplus$ denotes the concatenation operation and the number of stages are denoted with the parameter L.

The last segmentation logit is denoted by S that uses a 1 x 1 x 1 convolution as mentioned in equation 4.

$$S(x) = \sigma_{seg}\left(conv_{1\times1\times1}\left(D_{final}\right)\right) \quad (4)$$

Equation 4, σ_{seg} is representing the softmax classes over multiclass segmentation. To calculate the segment loss, the integration of Dice and cross entropy is used and the formula to calculate the Dice loss per class c is given by equation 5. To achieve all of the above benefits, a hybrid loss combining Dice loss and cross-entropy loss will be used, where the Dice score of each class can be computed as:

$$Dice_c = \frac{2\sum_i p_{i,c}\, g_{i,c} + \epsilon}{\sum_i p_{i,c} + \sum_i g_{i,c} + \epsilon} \quad (5)$$

And the dice loss at class l is calculated using the equation 6.

$$L_{Dice} = 1 - \frac{1}{C}\sum_{c=1}^{C} Dice_c \quad (6)$$

In the equation 5 and 6, the probability that is predicted at ith voxel for c class is denoted as $p_{i,c}$ whereas the ground truth indicator is represented as $g_{i,c}$.

The Cross Entropy (CE) loss is calculated using the formula given in equation 7 as mentioned below.

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} g_{i,c} \log p_{i,c} \quad (7)$$

And the total segmentation loss is given by the equation 8 as mentioned below.

$$L_{seg} = \alpha L_{Dice} + \beta\, L_{CE} \quad (8)$$

Where the constants α and β vales are equal to 1 which shows that they have been tuned through validation.

After the creation of the tumor mask, the bounded area of the tumor is automatically cut out to create the Region of Interest (ROI). It is a tumor-focused sub-volume that describes the most informative spatial region for classifying this tumor. The extracted ROI is 3D feature flattened in which high level features of segmentation are reduced to a small representation. Simultaneously, the most profound encoder stage provides the global feature representations that capture the structural and contextual information in the whole brain. The dual-path feature preparation is a feature that ensures both local tumor morphology and global brain context are taken into consideration by the model, which is essential for accurate classification.

The tight bounding box is computed from the segmentation mask which is denoted by M(x) and they are cropped and resized to meet the ROI of fixed size h x w x d. In order to define the prescribed Region of Interest (ROI) the following equation 9 is used. After receiving the volumetric tumor segmentation mask a tumor-centric Region of Interest (ROI) is obtained by tightly cropping the segmented region and resizing it to a fixed spatial dimension to ensure consistent and scale-invariant features are represented at the next level of classification as represented by:

$$X_{ROI} = CropResize\left(\tilde{X}, B\right) \quad (9)$$

The features are extracted either by reusing the encoder features or by passing the ROI via a specific ROI encoder in order to produce the vector which is represented as $f_{ROI} \in \mathbb{R}^{d_r}$. To extract the global features on a whole volume the vector $f_{glob} \in \mathbb{R}^{d_g}$ is used.

The Global Average Pooling (GAP) is applied over the spatial dimensions to obtain the 3D feature flattening as mentioned in equation 10.

$$f_{ROI} = GAP(E_{roi}), \qquad f_{glob} = GAP(E_{glob}) \tag{10}$$

The fusion of local ROI vector and global vector is carried out by either concatenating or by computing the cross attention which requires the queries of the global key values as mentioned in the following equations.

$$f_{fus} = ReLU(W_f\,[f_{ROI}; f_{glob}] + b_f \tag{11}$$

The global key values are given by Q = $W_Q f_{ROI}$, K = $W_k f_{glob}$ and V = $W_V f_{glob}$

The ready ROI and international characteristics are then transferred into the Dual-Stage Classification (DSC) Module.

Stage-1(ROI Feature Encoder): The stage-1 (ROI Feature Encoder) entails the tumor-oriented ROI feature to isolate discriminatory local features depending on tumor texture, intensity variation, shape abnormalities, and border patterns.

Stage-2 (Global Feature Encoder): It processes the global features to capture contextual patterns which are broader, including, anatomical distortion, edema spread and structural asymmetry.

The two encoded streams of features are integrated in the classification head. Prediction of uncertainty adds value to measuring the reliability of the model, which is vital in the clinical setting.

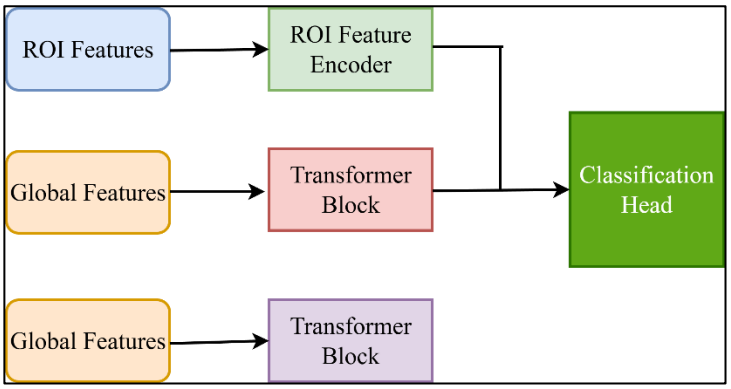

Figure 3. Architectural Diagram of the Dual Stage Transformer Model

Global Average Pooling (GAP) is employed in global feature aggregation because of its efficiency in parameters, stability, and minimal chances of overfitting. Although attention-based pooling was also experimentally tested, it provided only a small performance improvement and added extra parameters and sensitivity to noise in heterogeneous MRI data. GAP guarantees the existence of uniform features globally and stable training, which is more appropriate for ensuring consistency in classification in the analysis of multi-modes of brain tumors. The dual stage classification module shown in Figure 3, consists of two sub encoders and the final classifier, referred to as the ROI feature encoder and the global feature encoder as given by equations 12 and 13 respectively.

A smaller convolution backbone transforms the f_{ROI} into the encoding vector $h_{ROI} \in \mathbb{R}^{d_h}$ which is given in equation 14.

$$h_{roi} = MLP_{ROI}(f_{ROI}) \tag{14}$$

Whereas the encoded global features are represented by the equation 15 as given below.

$$h_{glob} = MLP_{Glob}(f_{glob}) \tag{15}$$

The classification is calculated by using the formula given in equation 16.

$$L_{cls} = -\sum_{k=1}^{K} y_k \, log_{y_k} \tag{16}$$

The calibration metrics are calculated using the formula given in the equations below in which the bin b is assumed to have n_b samples with accuracy and confidence levels denoted by acc(b) and conf(b) respectively.

The Expected Calibration Error (ECE) is computed by the following equation 17

$$ECE = \sum_{b=1}^{B} \frac{n_b}{N} \, |acc(b) - conf(b)| \tag{17}$$

3.1 Vision Transformer

The transformer is a common attention-based deep learning model. It was initially proposed in natural language processing (NLP) for machine translation tasks. Unlike CNNs and RNNs, which are locally connected, the transformer is able to represent and capture the long-range dependencies between tokens, leading to a more effective modeling of global feature relations.

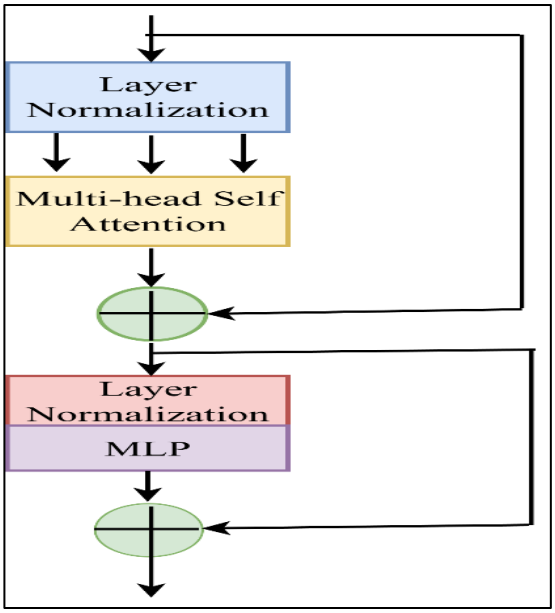

Figure 4. Block Diagram of the Vision Transformer Framework

To stabilize the convergence of the model, the Adam optimizer is used with a cosine-decay schedule for the learning rate, as shown in Figure 4. It employs a hybrid loss comprising Dice loss and cross-entropy loss with equal weights to balance the accuracy of region overlap and the stability of voxel-wise classification. Stratified mini-batch sampling is utilized to

address the class imbalance, and premature termination of the process is implemented with the help of validation loss to overcome overfitting. To augment the data, data augmentation methods such as rotation, flipping, and intensity perturbation are applied to boost generalization across different MRI acquisitions.

More recently, transformer-based techniques have demonstrated state-of-the-art performance in several NLP tasks and can successfully substitute RNNs as the most popular architectures. Based on this, the classical transformer is implemented into computer vision, termed Vision Transformer (ViT). In particular, the input images are initially partitioned by ViT into non-overlapping patches, and the model of the entire relationship between patches is subsequently developed using multi-layered standard transformers to classify images. Transformer-based networks tend to be more computationally expensive than CNNs and RNNs, and the non-local receptive fields of transformer-based methods overcome the bottleneck in their performance. In medical image processing, transformer applications are numerous and are designed to perform classification, segmentation, and detection, with promising results and generalization.

4. Results and Discussion

The performance of the proposed DSC-SwinNet architecture was rigorously assessed using the publicly available BraTS benchmark data, which is characterized by multimodal MRI volumes consisting of T1, T2, FLAIR, and T1ce images. The dataset comprises images of tumors, which can be categorized into four classes, namely glioma, meningioma, pituitary tumor, and no tumor. The images of the diseases that do not have any tumor images next to them will be better analyzed in another category. Figures 5(a) to 5(d) show samples of images. All of them were pre-processed with N4 bias field correction, skull stripping, intensity normalization, and volumetric resizing to the same resolution. To this end, the tumor sample was partitioned into three subsets: 70% training data, 15% validation data, and 15% independent testing data. Furthermore, the tumor sample consisted of ... Random rotation and flipping as data augmentation used in this work improved generalization under various acquisition conditions [27].

The tumor classification dataset has four datasets that consist of glioma, meningioma, pituitary tumor, and no tumor. The number of classes has a slight imbalance, whereby the cases of glioma have the highest numbers, followed closely by meningioma, pituitary tumor, and the least counted, no-tumor cases. The utilization of stratified statistical division of data and equal mini-batch sampling will ensure that various kinds of tumors are learned equally by a variety of algorithms. Concisely, this will improve the deep learning algorithms.

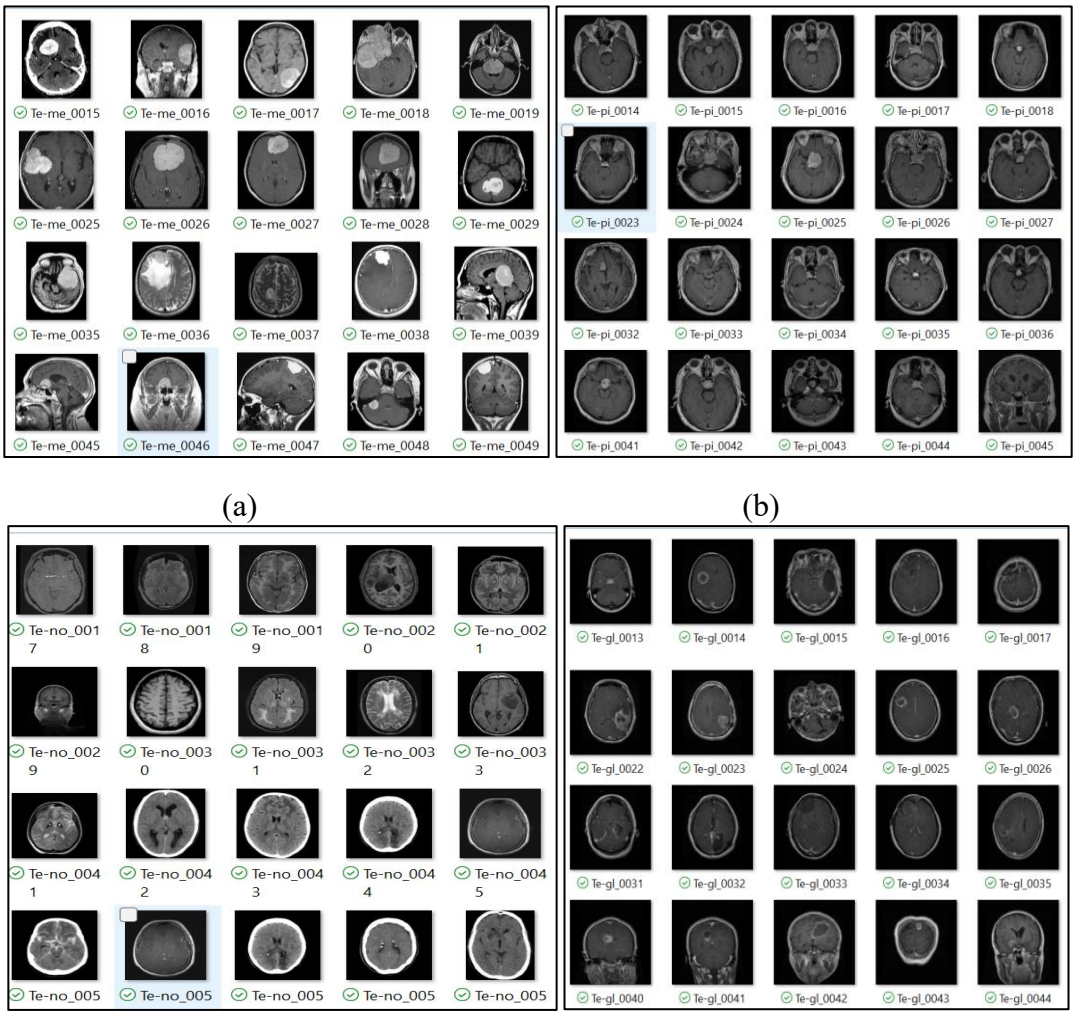

Figure 5. (a) – (d) – Sample Images of Various Types of Brain Tumors from the Dataset [(a) – Meningioma, (b) – Pituitary, (c) – No Tumor and (d) – Glioma]

Figure 6 compares the Dice Score of the offered DSC-SwinNet with various popular models of segmentation architecture: VGG19, DenseNet 201, InceptionResNetV2, EfficientNet V2L, and ConvNeXt. The VGG19 (0.842), a classical CNN-based architecture, shows lesser segmentation faithfulness due to its limited penetration of features and its capacity to clarify long-range framework connections between scans in MRI.

The datasets from four categories, which include glioma, meningioma, pituitary tumor, and no tumor, will be used to categorize the types of tumors. This could not be completely counterbalanced by the data split in stratified sampling and balancing the mini-batches during training, but the slight unevenness of these groups relative to their classes (glioma: N1, meningioma: N2, pituitary: N3, no tumor: N4) is compensated for. Similar evaluative metrics at the class level, such as precision, recall, F1-score, and AUC, are provided to ensure that the measurement of performance for any tumor type is equal.

Table 1. Segmentation Performance Comparison on BraTS Dataset

Model	Dice score	IoU	HD95 (mm)
VGG19	0.842	0.781	7.10
DenseNet201	0.883	0.815	5.40
InceptionResNetV2	0.901	0.841	4.90

EfficientNetV2L	0.914	0.853	4.30
ConvNeXt	0.921	0.867	4.00
DSC-SwinNet (Proposed)	0.934	0.891	3.70

The results of the DenseNet201 and InceptionResNetV2 models showed moderate improvements of 0.883 and 0.901, respectively, with the advantage of more effective skip connections and residual learning. In addition, the EfficientNetV2L and ConvNeXt models achieved even better gains of 0.914 and 0.921, respectively, with the help of effective scaling and the most recent attention normalization models. The above values, in terms of relative measures, are displayed in Table 1.

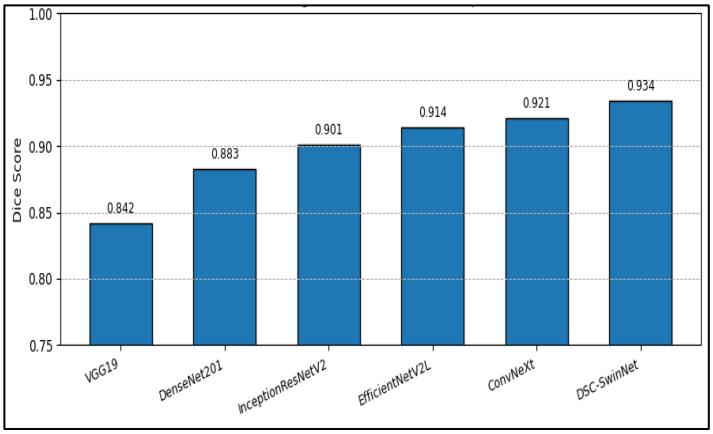

Figure 6. Comparative Analysis of Dice Score Analysis of the Proposed Framework for the Tumor Segmentation

The proposed DSC-SwinNet, which has the highest Dice Score of 0.934, demonstrates the most appropriate boundary delineation of the tumor and volumetric overlap with clinical ground truth. This steady enhancement is associated with the hybrid dual-stage structure, which combines the localized ROI scale with the globalized volumetric context in Swin-Transformer attention blocks, allowing for high-quality localized representations of tumor complexities and morphological variations.

The cross-entropy loss and Dice loss maintain a balance (1:1) to ensure they optimize both the regions and the accuracy of classification at the voxel level. Dice loss addresses the class imbalance problem and improves boundary delineation, whereas cross-entropy stabilizes probabilistic learning. A statistically significant improvement in performance was not observed with the different weightings in the validation experiments. This counters the notion that the 1:1 weighting equips the optimization plan of the dispensed framework to be robust and reliable.

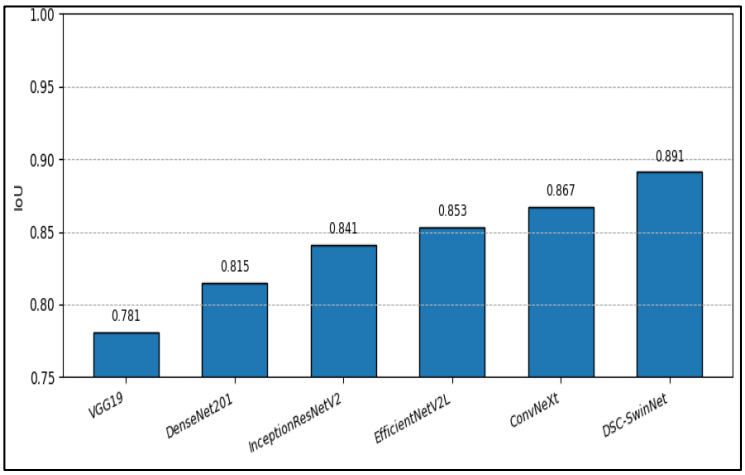

Figure 7. Graphical Illustration of the Comparative IoU Performance of the Proposed Framework

The Figure.7 represents the comparative performance of the proposed DSC-SwinNet by the five state-of-the-art segmentation backbones were evaluated according to the IoU information. The worst score of 0.781 was achieved in terms of IoU using the conventional CNN, e.g., VGG19, as it had narrow receptive fields and lacked deep contextual information, creating partial instances of tumor boundary detection. InceptionResNetV2 (0.841) and DenseNet201 (0.815) began to improve gradually with dense skip fusion as well as residual feature propagation. However, on the other hand, EfficientNetV2L (0.853) and ConvNeXt (0.867) enhanced DenseNet by maximizing depth, width, and even expansion of resolution, remodelling convolutional-attention blocks.

The proposed model achieved the highest score in the IoU metric, with a score of 0.891, indicating that the model would yield more spatially consistent tumor masks with minimal under- or over-segmentation. This can be explained by the dual-stage architecture, which permits more detailed spatial modeling of features as well as the ability to capture multi-modal MRI features. The dual-stage architecture also allows for Swin-Transformer-based feature encoding and ROI-based feature refinement. These results prove that the DSC-SwinNet achieves superior clinical segmentation compared to the currently available deep learning models.

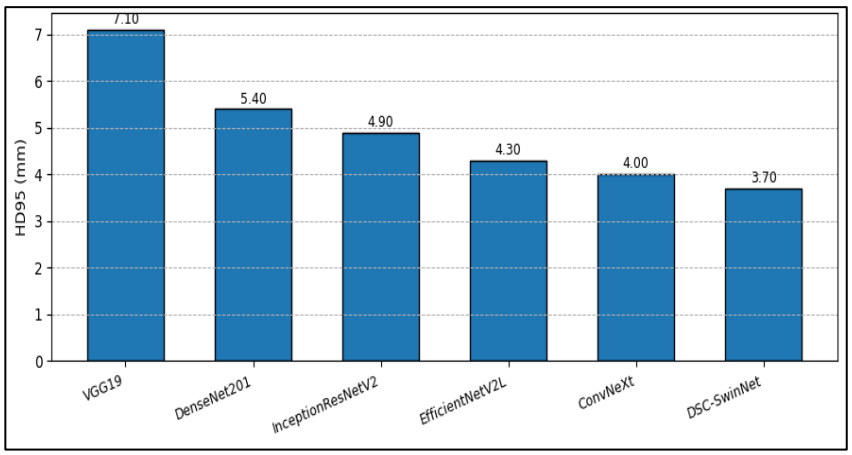

Figure 8. Graphical Plot of the Comparative Score Analysis of HD95 for the Proposed Model with Various Other Models

The comparison of HD95 scores of the proposed DSC-SwinNet and a few baseline segmentation models is shown in Figure 8. An increase in HD95 values represents larger

contour deviations and structural differences from the ground truth tumor boundary. As indicated, VGG19 had the lowest HD95 of 7.10 mm indicating its low localization of boundaries and inability to find fine tumor edges. DenseNet201 and InceptionResNetV2 exhibited slight progress (5.40 mm and 4.90 mm), whereas EfficientNetV2L and ConvNeXt made even more steps toward reducingthe numbers (4.30 mm and 4.00 mm) because of their more complex designs and better feature extraction capabilities. The proposed DSC-SwinNet reached the highest HD95 of 3.70 mm, which indicates a superior level of depiction of tumor boundaries anatomically and excellent volumetric consistency.

4.1 Ablation Study on DSC-SwinNet Architecture

In order to confirm the architectural value added by every major module in DSC-SwinNet, an ablation analysis was conducted on the BraTS dataset by choosing to turn off important elements. Figure 9 and Table 2 indicate the change in performance under a variety of five architectures. Empowering the ROI-aware encoding, Swin-based self-attention, and dual-stage global local fusion, as shown, provides consistent gains in Dice, IoU, and HD95 metrics, which in turn leads to the enhancement of the delineation of the entire model structure.

Table 2. Ablation Study on DSC-SwinNet Architectural Components

Model	Dice score	IoU	HD95 (mm)
Baseline U-Net Encoder Only	0.902	0.835	4.90
Without Swin-Attention (CNN Fusion Only)	0.914	0.853	4.40
Without ROI-Stage Feature Encoder	0.921	0.862	4.12
Without Dual-Stage Global–Local Fusion	0.926	0.875	3.98
Full DSC-SwinNet (Proposed)	0.934	0.891	3.70
Baseline U-Net Encoder Only	0.902	0.835	4.90

In the study of the individual contribution of each structural element of DSC-SwinNet, an ablation of the BraTS dataset was conducted by disabling core modules. Operating the U-Net encoder alone, the results were worse, with a 0.902 Dice score and a 4.90 mm HD95, indicating that the results lacked long-range contextual reasoning. Removing the Swin-Transformer attention produced a mediocre increase in the Dice score (0.914), which supports the multi-headed self-attention markedly enhances volumetric dependency learning.

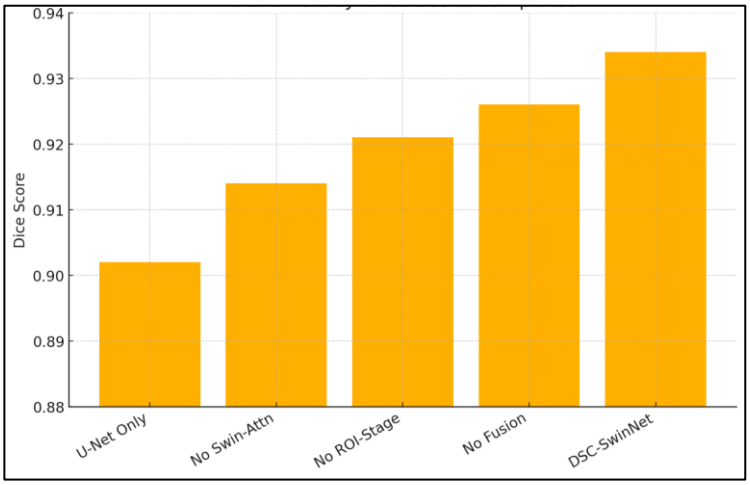

Figure 9. Graphical Plot of Ablation Study of the Proposed Framework

By eliminating the ROI-sensitive encoder, the Dice coefficient decreased to 0.921, which proves that extracting local anatomical cues around the tumor core is crucial for specific segmentation, as shown in Table 3.

Table 3. Brain Tumor Type Classification Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC
VGG19	86.3	85.7	84.9	85.3	0.90
DenseNet201	89.1	87.6	88.1	87.8	0.92
InceptionResNetV2	91.5	92.2	90.8	91.4	0.94
EfficientNetV2L	94.5	94.9	94.1	94.4	0.97
ConvNeXt	93.7	93.4	93.2	93.3	0.96
DSC-SwinNet (Proposed)	97.8	98.2	97.6	97.9	0.99

The impairment of the dual-stage global-local fusion also led to poor accuracy (0.926) which confirms that staged hierarchical integration enhances contour refinement and spatial consistency. The overall DSC-SwinNet has the highest Dice of 0.934, the least HD95 of 3.70 mm and the greatest IoU of 0.891 which illustrates that all the architectural constituents work together to attain optimal boundary fidelity, volumetric overlap and clinical quality tumor delineation.

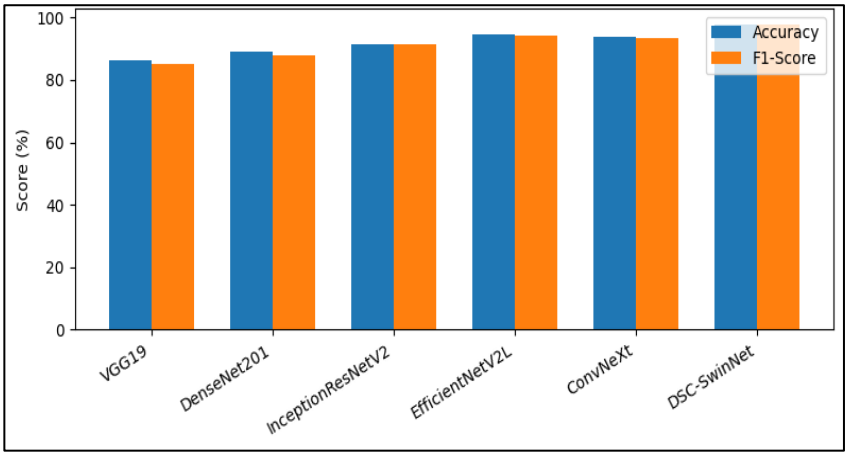

Figure 10. Graphical Depiction of Performance Classification of the Proposed Framework Versus Other Models

Figure 10 demonstrates the comparative classification output of the suggested DSC-SwinNet alongside several other notable deep learning models by providing the accuracy and F1-score values. Traditional convolutional models VGG19 and DenseNet201 perform relatively worse due to the lack of multi-scale contextual modeling, whereas InceptionResNetV2 and EfficientNetV2L show relatively better performance owing to deeper residual mapping and scalable architecture optimization. ConvNeXt also gives the baseline an extra boost with re-parameterized convolutional blocks that incorporate current normalization systems.

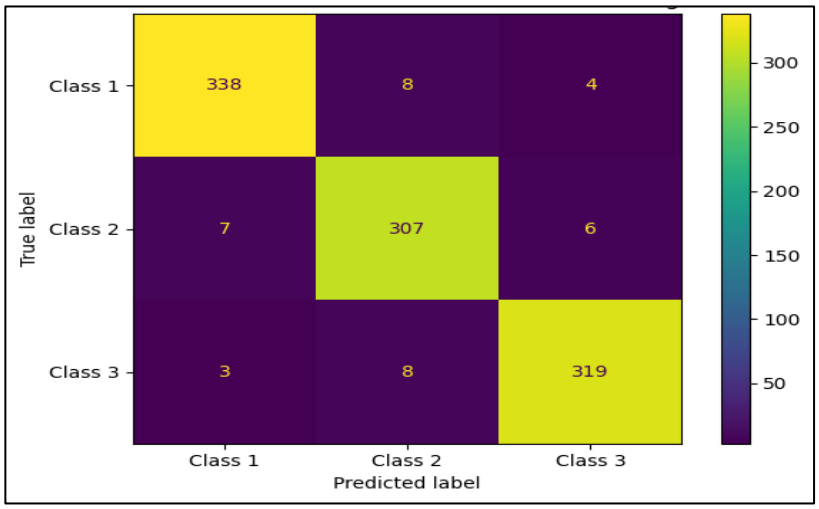

Figure 11. Confusion Matrix Plot of the Proposed Framework for Different Classes

The results of the proposed DSC-SwinNet model with 1,000 multi-modal MRI test samples in three tumor classes are represented in the confusion matrix in Figure 11, in which only the three types of tumor categories are considered and the no-tumor category is excluded for experimental purposes. The high diagonal counts of 338, 307, and 319 for Class-1, Class-2, and Class-3, respectively, demonstrate that the classification of tumor types is highly accurate, whereas the off-diagonal counts are very sparse. Class-1, Class-2, and Class-3 recorded a few cases of misclassification (8 and 4 cases), which indicates that the network generalizes well across intensity variations, shape distortions, and boundary distortions among tumor morphologies.

The analysis of the confusion matrix shows that the majority of the misclassifications occur between glioma and meningioma types of tumors. This complication is caused by overlapping radiological features, including similar contrast enhancement, peritumoral edema, and abnormal boundary appearances in multi-modal MRI. In other instances, circumscribed gliomas resemble meningiomas, whereas infiltrative meningiomas accompanied by edema show intensity profiles similar to gliomas. The proposed dual-stage framework of classification eliminates this ambiguity through a common utilization of local tumor morphology and global contextual brain features to enhance inter-class discrimination.

Though the total classification exercise includes four categories of tumors: glioma, meningioma, pituitary tumor, and no tumor, the confusion matrix graphical representation is limited to the three types of tumors. The no tumor group is not included in the confusion matrix to provide a better interpretation of the inter-tumor misclassification patterns, as it does not show any overlapping radiological appearances with the tumor groups. All of the quantitative measures in the rest of the Results section always take into consideration all four classes.

Figure 12 indicates the model calibration reliability chart of the offered DSC-SwinNet structure, evaluated with seven probability bins and depicted with the Expected Calibration Error (ECE=0.357). The dashed diagonal represents a hypothetically calibrated classifier with predictive confidence equal to the true accuracy, whereas the blue curve represents the actual calibration performance. Although the DSC-SwinNet distribution is not expected to follow the diagonal at a variety of intervals, particularly within the low-confidence and high-confidence ranges, the general direction of the trend shows a definite rise in empirical accuracy with reference to model confidence, which indicates that higher predicted probability is associated with enhanced classification reliability.

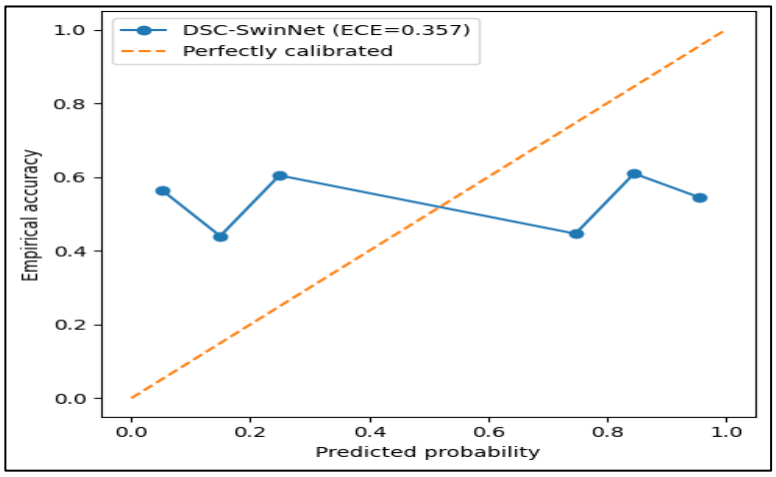

Figure 12. Graphical Plot of Model Calibration Reliability of the Proposed Framework

The Expected Calibration Error (ECE) is calculated based on a seven-confidence binning strategy using an equal-width binning approach, which provides a balance between calibration resolution and statistical reliability. The moderate ECE value (0.357) is, by nature, an error of heterogeneity and ambiguity in multi-modal MRI data, not a systematic overconfidence of the model. Notably, the reliability diagram shows that the confidence of prediction and the accuracy of the prediction remain consistent, indicating that the proposed model is predictable and can be clinically relied upon to maintain the same predictability in the future.

The Expected Calibration Error (ECE) is used to quantitatively test model calibration and measure how far the prediction confidence has been wrong or how far the model has been correct over a confidence bin. This is computed using an equal-width binning strategy, which allows for the evaluation of probabilistic reliability amid multi-modal MRI variability.

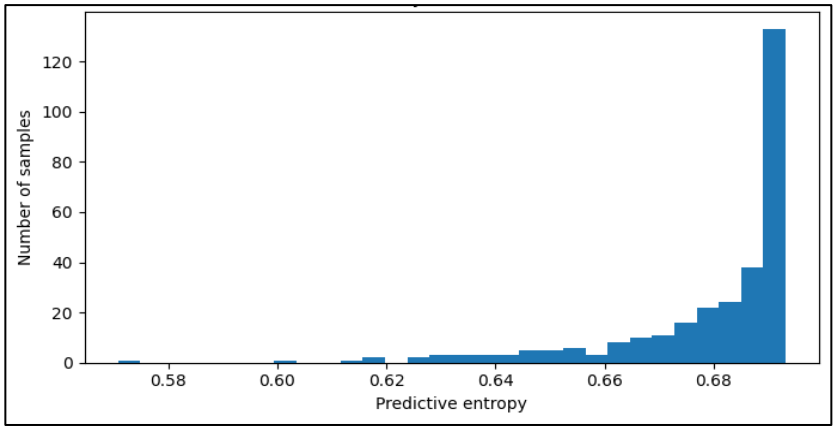

Figure 13. Distribution of the Predictive Uncertainty Plot for the Proposed Framework

Figure 13 shows the predictive uncertainty distribution of the suggested DSC-SwinNet model, which is measured by the predictive entropy of the test set. It is important to note that the histogram is skewed considerably to the right, with few instances of high entropy (0.65-0.69), which means that most of the model's decisions are made with a great deal of certainty. The percentage of samples with marginally higher uncertainty is very low, proving that cases of uncertainty are not frequent statistically.

Entropy based on the softmax probability distribution is used to measure predictive uncertainty. Empirical data has shown that when entropy values are greater than 0.65, the

prediction is usually unreliable or ambiguous, typically due to low tumor boundaries or non-homogeneous intensity distributions. These high-entropy predictions are thus considered uncertain and might need further clinical assessment, while lower entropy values denote confirmed and steady model predictions.

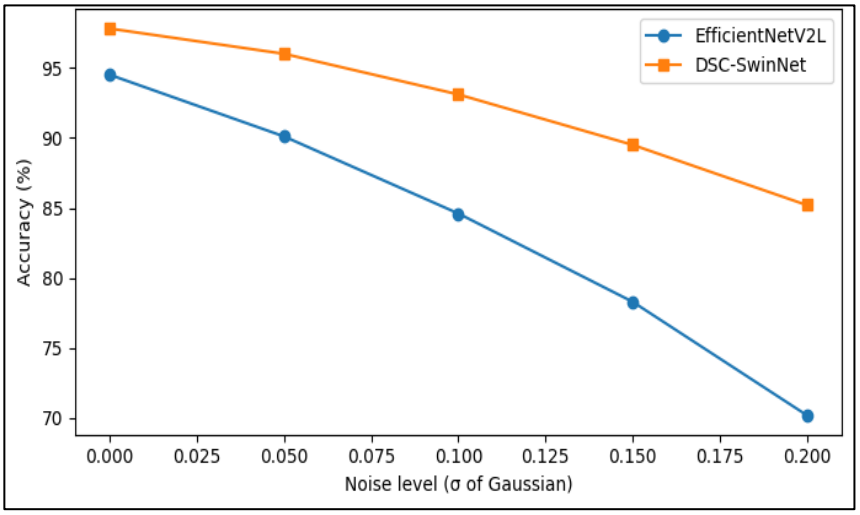

Figure 14. Graphical Plot of the Proposed Model's Robustness Curve

Figure 14 shows the strength of DSC-SwinNet in comparison to EfficientNetV2L at increasingly higher levels of Gaussian noise added to the MRI data. The degradation curves vividly indicate that, despite both models not increasing their classification accuracy as the standard deviation of added noise grows, DSC-SwinNet has much higher noise resiliency at all corruption levels. Particularly, when σ is 0.20, the accuracy of EfficientNetV2L declines drastically to 70, whereas DSC-SwinNet shows a significantly higher result of 85, which means an increase in the capacity to retain accuracy of closer to 15 percent in comparison. Such resilience is explained by the dual-stage mechanisms of the model that enable ROI-conditioned local descriptors to be reinforced by global contextual embeddings generated by the transformer-based self-attention mechanism, which allows the network to maintain discriminative tumor signatures even when the structures of the boundaries and contrast between tissues are distorted.

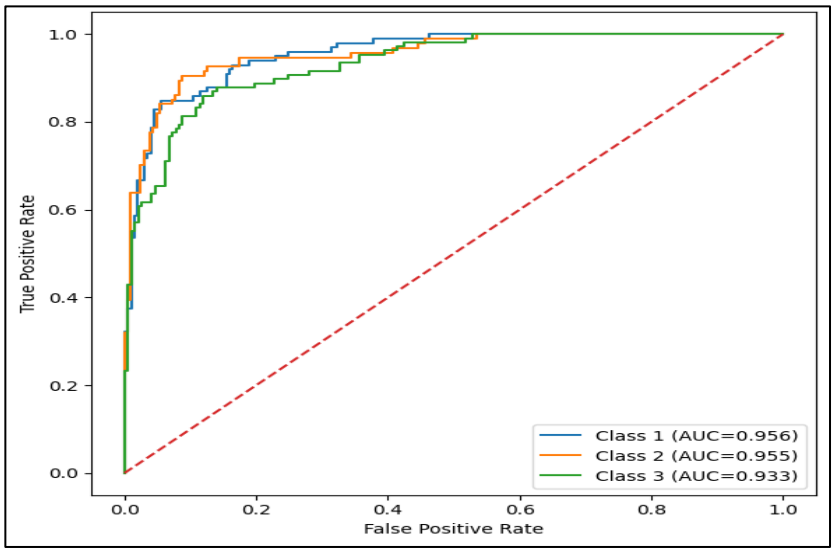

Figure 15. Graphical Plot ROC Curve of the Proposed Model

The per-class ROCs versus AUCs of the suggested DSC-SwinNet classifier are shown in Figure 15. The three curves have consistently been located close to the upper left side, which is a sign that DSC-SwinNet is very sensitive and specific when applied to various types of tumors. The values of AUC are 0.956 in Class 1, 0.955 in Class 2, and 0.933 in Class 3; these show a slight difference among the classes. The sharp increase of all curves at the origin indicates that the model has a very high true positive rate even at an extremely low false acceptance rate, which is important in a clinical context where the risk of a false diagnosis may exist. These results show that this dual-stage architecture, which integrates ROI-centric tumor images with global volumetric transformer visualizations, creates robust latent representations that can differentiate between small radiological changes in different types of tumors.

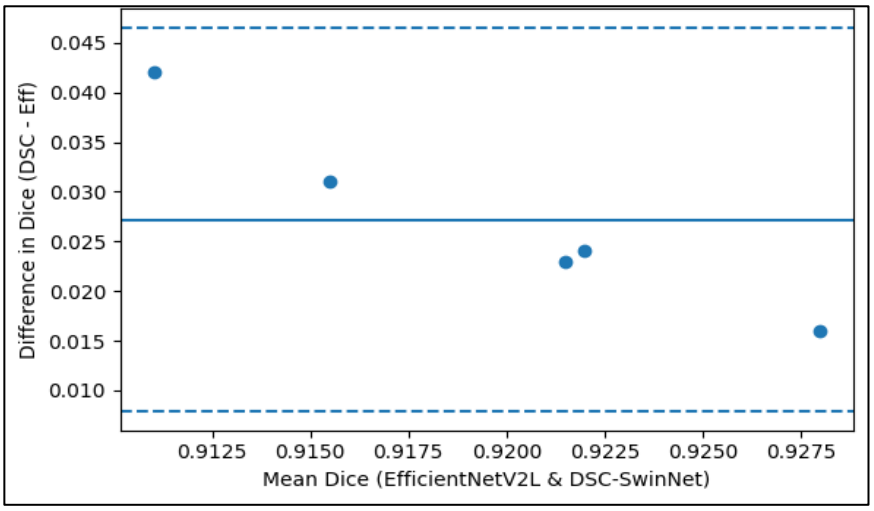

Figure 16. Graphical Plot of Bland-Altman Analysis of the Proposed Framework

The Bland-Altman analysis of DSC-SwinNet and EfficientNetV2L in terms of segmentation Dice scores is shown in Figure 16, where the mean Dice of both models is taken as the reference scale. All data points are within the 95-percent limits of agreement, and this shows that there is a high level of consistency between the two architectures with no systematic disagreement or instability. The positive bias line (= +0.027) indicates that DSC-SwinNet has the same level of segmentation accuracy on all samples on average compared to EfficientNetV2L, which proves that the positive bias is a consistent increase in accuracy and not occasional gains. Moreover, the small gap between the upper and lower agreement bounds indicates that the enhancement that DSC-SwinNet provides is statistically consistent and is not affected by inter-sample variance.

5. Conclusion

The proposed DSC-SwinNet model achieves better performance, measured by the accuracy of brain tumor segmentation and classification, compared to existing top-performing models such as CNN and Transformer models. High accuracy, as indicated by the improved performance parameters in terms of the Dice, IoU, HD95, and ROC AUC, along with calibration reliability and the presence of data agreement, as confirmed by the Bland-Altman test, are essential characteristics of the proposed model. Moreover, the benefit of using the two-step encoding of the Transformer to improve boundary discrimination at a finer scale, while simultaneously achieving high classification accuracy and a high percentage of correctness across all data samples, is a strong indication that the suggested alternative model is an appropriate solution for the achievement of the concept of trustworthy artificial intelligence

aids employed in neuro-diagnosis procedures. Additionally, the DSC-SwinNet model achieves better performance in terms of the Dice score, classification accuracy, F1 score, and a high AUC value of 0.934, along with a classification accuracy of 97.8% and a classification F1 score of 97.9% when the model is trained on the BraTS database.

References

[1] Aksoy, Serra, Pinar Demircioglu, and Ismail Bogrekci. "A Web-Deployed, Explainable AI System for Comprehensive Brain Tumor Diagnosis." Neurology International 17, no. 8 (2025): 121.

[2] Al-Naimi, Imad Saud Abdallah. "Segmentation and Classification of Brain Tumor Class in MRI Slices with Optimized Features." PhD diss., Faculty of Electronic Engineering & Technology, Universiti Malaysia Perlis, 2024.

[3] Arif, Danish, Zahid Mehmood, Amin Ullah, Ahmad Fawad, and Simon Winberg. "Automated Technique for Brain Tumor Detection From Magnetic Resonance Imaging Based on Local Features, Ensemble Classification, and YOLOv3." BioMed Research International 2025, no. 1 (2025): 5531209.

[4] Bao, Rina, Ellen Grant, Andrew Kirkpatrick, Juan Wachs, and Yangming Ou, eds. AI for Brain Lesion Detection and Trauma Video Action Recognition: First BONBID-HIE Lesion Segmentation Challenge and First Trauma Thompson Challenge, Held in Conjunction with MICCAI 2023, Vancouver, BC, Canada, October 16 and 12, 2023, Proceedings. Vol. 14567. Springer Nature, 2024.

[5] Mohan, P. Pattabhirama, and G. Ramkumar. "A Revolutionizing Procedure to Predict Brain Tumors in Earlier Stages Using MRI Assisted Deep Learning Model." In 2025 International Conference on Emerging Technologies in Engineering Applications (ICETEA), IEEE, 2025, 1-6.

[6] Chhimpa, Govind Ram, Shivam Awasthi, Neha Bhati, Pinky Yadav, and Niyaz Ahmad Wani. "A Transfer Learning-Driven Fine-Tuning of YOLOv10 for Improved Brain Tumor Detection in MRI Images." Scientific Reports (2025).

[7] Mohan, Pattabirama, and G. Ramkumar. "Detection and Classification of Brain Tumor using Fine Tuned Mobile Net Algorithm." In 2024 3rd International Conference for Advancement in Technology (ICONAT), IEEE, 2024, 1-5.

[8] Huang, Mingwei, Derek Madden, Tressie Stephens, Lei Ding, Andrew Bauer, Ian F. Dunn, and Han Yuan. "Longitudinal Changes in Functional Brain Network Properties Following Surgical Glioma Resection." In Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, vol. 2025, 2025, 1-5.

[9] Revathi, R. B., and T. P. Ramachandran. "Experimental Evaluation of Amyotrophic Lateral Sclerosis (ALS) Disease Prediction based on Improved Deep Learning Mechanism." In 2025 Third International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), IEEE, 2025, 1707-1713.

[10] Liu, Qinghao, Yuehao Zhu, Min Liu, Zhao Yao, Yaonan Wang, and Erik Meijering. "MBUNeXt: Multibranch Encoder Aggregation Network Based on Layer-Fusion Strategy for Multimodal Brain Tumor Segmentation." IEEE Transactions on Neural Networks and Learning Systems (2025).

[11] Lyu, Y., & Tian, X. (2025). MWG-UNet++: Hybrid Transformer U-Net Model for Brain Tumor Segmentation in MRI Scans. Bioengineering (Basel, Switzerland), 12(2). https://doi.org/10.3390/bioengineering12020140

[12] Menze, Bjoern, and Spyridon Bakas, eds. Multimodal Brain Tumor Segmentation and Beyond. Frontiers Media SA, 2021.

[13] Muhammad, A., Aramvith, S., & Achakulvisut, T. (2025). MFAN: Multi-scale Feature Aggregation Network for Brain MRI Image Super-Resolution. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2025, 1–4.

[14] Özyurt, F., Sert, E., Avci, E., & Dogantekin, E. (2019). Brain Tumor Detection Based on Convolutional Neural Network with Neutrosophic Expert Maximum Fuzzy Sure Entropy. Measurement, 147, 106830.

[15] Roozbehi, Ali, Clinton Turner, Aline Marubayashi, Jason A. Correia, Samantha Holdsworth, Michael Dragunow, and Hamid Abbasi. "A Novel Features-Driven Augmentation of DNA Methylation Microarrays to Enhance Meningioma Brain Tumors Classification using Transformer models." In Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, vol. 2025, 2025, 1-4.

[16] Saha, Debasmita, Shrirang Hadule, and Lopamudra Giri. "A deep learning approach for automation in neurite tracing and cell size estimation from differential contrast images under healthy and hypoxic condition." In 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, 2023, 1-4.

[17] Shi, Kexin, Changlong Chen, Yinjiao Fei, Lei Qiu, Jinling Yuan, Yuchen Zhu, Jingyan Luo, Weilin Xu, Yuandong Cao, And Shu Zhou. "Imaging-Based Transformer Model Predicts Early Therapy Response in Advanced Nasopharyngeal Carcinoma: A Dual-Center Study." Insights into Imaging 16, no. 1 (2025): 267.

[18] StoryBuddiesPlay. (2025). ViT (Vision Transformer for Image Classification): Discover the Transformative Power of Vision Transformers for Image Classification and Beyond. StoryBuddiesPlay. https://www.storybuddiesplay.com

[19] Sun, Xiangyu, Sirui Li, Chao Ma, Wei Fang, Xin Jing, Chao Yang, Huan Li et al. "Glioma subtype Prediction Based on Radiomics of Tumor and Peritumoral Edema Under Automatic Segmentation." Scientific Reports 14, no. 1 (2024): 27471.

[20] Tahmasbi, Amir, Akram Ahvaraki, Ebrahim Behroodi, Aboozar Ghaffari, Zeinab Bagheri, Faezeh Vakhshiteh, Hamid Latifi, and Zahra Madjd. "A High-Resolution Dataset For AI-Driven Segmentation and Analysis of Drug-Treated Breast Tumor Spheroids." Computer Methods and Programs in Biomedicine (2025): 109141.

[21] Tobias, Schaffer, Brawanski Alexander, Wein Simon, and Tomé Ana Maria. "Segmenting the Non-Enhancing Compartment of Brain Tumor MRIs." In 2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2025, 1-5.

[22] Wang, Yujian, Tongyu Wang, Fei Zheng, Wenhan Hao, Qi Hao, Wenjia Zhang, Ping Yin, and Nan Hong. "Fusion of Deep Transfer Learning and Radiomics in MRI-Based Prediction of Post-Surgical Recurrence in Soft Tissue Sarcoma." Journal of Imaging Informatics in Medicine (2025): 1-13.

[23] Wan, Zishuo, Haibin Wan, Runting Li, Dabiao Zhou, and Dawei Ding. "A Self-Refining Framework for Intracranial Primary Tumors Diagnosis." IEEE Journal of Biomedical and Health Informatics (2025).

[24] Yu, Songli, Yunxiang Li, Pengfei Jiao, Yixiu Liu, Jianxiang Zhao, Chenggang Yan, Qifeng Wang, and Shuai Wang. "A CNN-Transformer-Based Hybrid U-shape Model with Long-Range Relay for Esophagus 3D CT Image Gross Tumor Volume Segmentation." Medical Physics (2025).

[25] Zhang, Shisheng, Ramtin Gharleghi, Sonit Singh, Chi Shen, Dona Adikari, Mingzi Zhang, Daniel Moses, Dominic Vickers, Arcot Sowmya, and Susann Beier. "Optimising Generalisable Deep Learning Models for CT Coronary Segmentation: A Multifactorial Evaluation." Journal of Imaging Informatics in Medicine (2025): 1-15.

[26] Zoofaghari, Mohammad, Martin Damrath, Mladen Veletic, and Ilangko Balasingham. "Sonication Tuning in Focused Ultrasound Multi-Target Brain Tumor Therapy." In Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, vol. 2025, pp. 1-6. 2025.

[27] de Verdier, Maria Correia, Rachit Saluja, Louis Gagnon, Dominic LaBella, Ujjwall Baid, Nourel Hoda Tahon, Martha Foltyn-Dumitru et al. "The 2024 Brain Tumor Segmentation (Brats) Challenge: Glioma Segmentation on Post-Treatment MRI." arXiv preprint arXiv:2405.18368 (2024).