

BRAIN TUMOR SEGMENTATION USING MULTI-ATTRIBUTE AGGREGATION AND ADAPTIVE GRADIENT DICE LOSS MECHANISM

SUBIA SALMA^{1*}, S C LINGAREDDY², VINEET KUMAR³,
SHILPASHREE S⁴

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, RESEARCH SCHOLAR, VVIT, BANGALORE, INDIA¹

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, PRINCIPAL, VVIT, BANGALORE, INDIA²

TECHNICAL DIRECTOR, PLANET I TECHNOLOGIES, BANGALORE, INDIA³

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, NAGARJUNA COLLEGE OF ENGINEERING AND TECHNOLOGY, ASSOCIATE PROFESSOR, BANGALORE, INDIA⁴

¹Subiasalma89@gmail.com, ²sclingareddy@gmail.com ³vineetkumarvk97@gmail.com, ⁴shilpatani@gmail.com

Abstract

The human brain is a remarkable entity, central to all bodily functions. Given its critical role, prompt reporting of any irregularities mainly tumors that further lead to cancer in its structure is vital for reducing mortality rates. Segmenting abnormal regions is key for effective treatment planning and monitoring. The most crucial task in this process is the differentiation of healthy tissue from abnormal areas. To date, various imaging techniques have been employed for the early detection of these anomalies, with Magnetic Resonance Imaging (MRI) standing out as a notable and noninvasive diagnostic tool. This research presents an integrated architecture Multi-Attribute Aggregation (MAA) and an Adaptive Gradient Dice Loss (AGDL) (MAA+AGDL) mechanism to significantly improve the accuracy of segmenting brain tumors from MRI scans. The novel framework employs a dual-path architecture that processes foreground and background regions separately, incorporating enhanced feature extraction through improved encoders, a multi-attribute aggregation module for comprehensive feature representation, and mapping prediction modules for precise segmentation. To overcome traditional challenges associated with gradient vanishing and feature misalignment, the framework integrates an AGDL mechanism for stable training and a Deformable Convolution-based Feature Alignment (DCFA) for adaptive feature alignment. Evaluated on the BRATS 2019 and 2020 datasets, our model demonstrates superior performance over existing approaches, achieving higher Dice Similarity Coefficient scores, improved sensitivity, and reduced Hausdorff Distance. These advancements signify a substantial step forward in the use of deep learning for medical imaging, promising enhanced clinical diagnostics and treatment planning for brain tumor patients.

Keywords: Multi-Attribute Aggregation (MAA); Adaptive Gradient Dice Loss (AGDL); Brain Tumor Segmentation; Deep Learning Framework; Magnetic Resonance Imaging (MRI).

1 INTRODUCTION

The central nervous system (CNS) of humans encompasses the brain, the primary component of the human nervous system, along with the spinal cord. It is the brain that governs the body's myriad functions, encompassing processing, integration, coordination, decision-making, and dispatching directives to the body. Featuring a highly intricate anatomical architecture, the human brain is central to our neurological functions. Disorders affecting the CNS range from brain tumors, injuries sustained through trauma, and developmental irregularities, to conditions such as multiple sclerosis, strokes, dementia, infections, and even migraines. The incidence of brain tumors has been on an upward trajectory worldwide, attributed to the proliferation of cancerous cells that migrate to the brain via the vascular system. Between 2004 and 2022, there has been a notable increase in brain tumor cases, estimated between 10% to 15%. [1]-[3] Imaging techniques such as Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) are

critical in diagnosing brain irregularities. MRI, in particular, is a noninvasive tool that excels in investigating CNS disorders, offering detailed views of the brain in slices without invasive procedures. This modality stands out for its sensitivity in detecting CNS conditions, providing invaluable data for diagnosis. MRI's superiority over CT includes enhanced image quality and the ability to observe brain structures in real-time without exposing patients to ionizing radiation [4]. Its ability to generate high-contrast images across three dimensions—axial, sagittal, and coronal—enables comprehensive analysis of the brain, spinal cord, and vascular structures. Unlike CT scans, MRI uses a combination of magnetic fields and radio waves to produce detailed images, eliminating radiation exposure risks. MRI's high contrast resolution is particularly beneficial for examining the brain's soft tissues, offering clear views in three-dimensional planes to aid in understanding the brain and spinal cord's complex anatomy. The technology is adept at distinguishing between different types of tissue, making it essential for diagnosing various CNS disorders. MRI sequences like T1-weighted, T2-weighted, and Fluid-Attenuated Inversion Recovery (FLAIR) scans provide diverse insights into brain tissue properties, aiding in the differentiation of fluids like cerebrospinal fluid (CSF) from other abnormalities, these modalities are presented in figure 1[5]-[7].

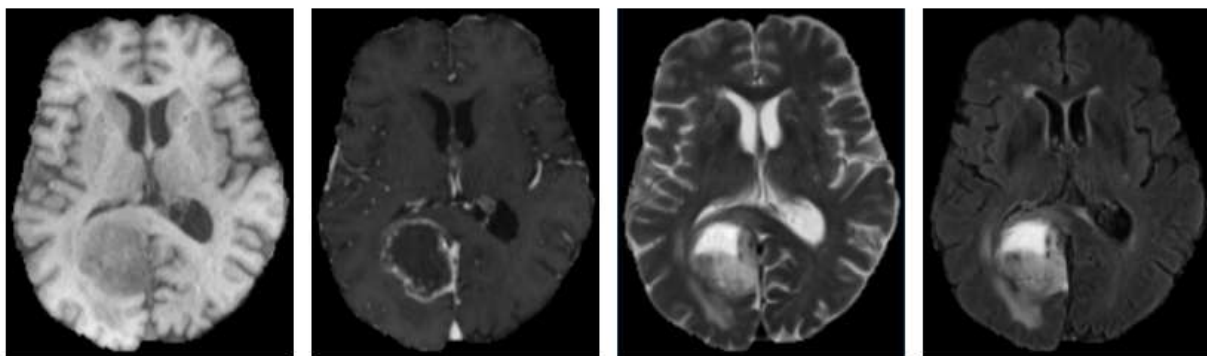


Figure 1 Different MRI imaging modalities, from left to right T1, T1ce, T2, and FLAIR

The field of brain tumor segmentation has seen significant advancements with the adoption of machine learning and deep learning techniques [5]-[11]. Traditional methods relied on extracting specific features from MRI data but faced challenges due to the diverse nature of brain tumors and the limitations of certain MRI modalities. Deep learning approaches, particularly Convolutional Neural Networks (CNNs) and advanced structures like the U-Net [8], have demonstrated a strong capability for automatic brain tumor segmentation by learning complex features directly from the data. These techniques have evolved to address the heterogeneity of tumors and incorporate multimodal MRI data for more precise segmentation. Innovations such as multi-task networks, deep feature fusion frameworks, and 3D CNNs [9] have significantly improved the accuracy and efficiency of brain tumor segmentation. Recent strategies have also focused on overcoming the challenges posed by incomplete MRI datasets in clinical settings, which can result from varied scanning protocols and patient conditions. Solutions such as Generative Adversarial Networks (GANs) [10] and Variational Autoencoders (VAEs) [11] have been explored to reconstruct missing modalities, enhancing segmentation performance even when complete multimodal data isn't available. This progress reflects a broader move towards leveraging deep learning for nuanced understanding and management of CNS disorders, underscoring the importance of continuous innovation in medical imaging technology. Moreover, deep learning has a heavy reliance on large annotated datasets, which are often scarce and expensive to obtain in the medical field. While deep learning models excel in extracting complex features and patterns from extensive data, they struggle to generalize from limited or highly specific datasets, limiting their applicability in varied medical scenarios. The incorporation of meta-learning introduces the ability to learn and adapt quickly from a few examples, overcoming the data dependency hurdle. This synergy not only boosts segmentation accuracy in scenarios with limited data but also mitigates the issue of overfitting, a common challenge in deep learning models, by enabling models to generalize better to new, and unseen medical images.

1.1 Motivation and contribution

The issue of brain tumors and their impact on mortality rates worldwide has become a growing concern. To combat this, researchers have developed an innovative approach to improve the accuracy of diagnosis. The automated brain tumor segmentation model that has been created allows experts to identify brain tumors more efficiently, ultimately leading to a more timely diagnosis. Traditional methods, while useful, often fall short of capturing the complex features of brain tumors, leading to a pressing need for more advanced, automated, and precise segmentation techniques. The motivation behind this research stems from the necessity to enhance the accuracy and efficiency of brain tumor diagnosis, which is crucial for timely intervention and effective treatment planning.

This research introduces a novel integrated architecture combining Multi-Attribute Aggregation (MAA) with Adaptive Gradient Dice Loss (AGDL) to significantly improve brain tumor segmentation from MRI scans. The proposed framework stands out for its dual-path approach for foreground and background segmentation, innovative use of deformable convolution-based feature alignment, and the application of an adaptive gradient dice loss mechanism to enhance segmentation accuracy. Specifically, the study contributes to the field in several key ways:

- **Innovative Integration of MAA and AGDL:** By integrating MAA with AGDL, the framework improves upon existing models in terms of segmentation accuracy, demonstrating superior performance on the BRATS 2019 and 2020 datasets with enhanced Dice Similarity Coefficient scores, sensitivity, and reduced Hausdorff Distance.
- **Dual-Path Segmentation Approach:** The dual-path architecture, designed for precise segmentation of both foreground and background regions, incorporates advanced encoders, multi-attribute aggregation modules, and mapping prediction modules. This approach facilitates a more nuanced and accurate segmentation process.
- **Adaptive Gradient Dice Loss Mechanism:** The introduction of an adaptive gradient optimization to the Dice loss function addresses the challenge of gradient vanishing, ensuring a more stable and effective training process.
- **Deformable Convolution-based Feature Alignment:** The utilization of deformable convolution layers for feature alignment across different image scales represents a significant advancement in capturing and aligning critical features for accurate tumor segmentation.
- **Extensive Evaluation and Validation:** The comprehensive performance evaluation on the BRATS datasets, using metrics such as Dice Similarity Coefficient, sensitivity, specificity, and Hausdorff Distance, validates the effectiveness and efficiency of the proposed framework in real-world clinical settings.

2 RELATED WORK

The rapid detection of brain tumors presents a formidable challenge, yet the advancement of deep learning technologies has offered promising avenues for their accurate identification using digital imagery. Notably, the U-Net framework [12], as introduced in one study, marks a significant stride in brain tumor image segmentation, specifically within the Brain Tumor Segmentation Challenge (BraTS). This study proposed enhancements to the U-Net architecture, focusing on post-processing, region-specific training, data augmentation, and further analysis to refine segmentation performance. Despite these improvements, the model faced limitations due to the absence of explicit spatial constraints and scalability issues. Further research introduced a 3-Dimensional U-Net design [13], integrating channel and spatial attention within the decoder network to forge an attention-based Convolutional Neural Network (CNN) capable of brain tumor detection from MRI scans. The complexity of this model, however, raised concerns regarding its computational demand and interpretability. Subsequently, the UNet++ [14] architecture was proposed, featuring a fully trained encoder-decoder network where nested dense skip connections bridge the encoder and decoder networks, reducing the information gap. Yet, selecting the optimal set of hyperparameters for this model posed a challenge. Another advancement, the 2-dimensional attention recurrent residual U-Net (AttR2U-Net) [15], employed multitask deep supervision (MTDS) to draw semantic insights from images, enhancing segmentation accuracy but at the risk of overfitting. Following this, the LKAU-Net [16-17] model emerged, building on the U-Net structure with the integration of large-kernel attention mechanisms.

This approach improved brain tumor segmentation by capturing long-range dependencies through large-kernel convolutions and focusing on critical regions with attention mechanisms. Lastly, the PraNet model introduced a parallel reverse attention network utilizing a parallel partial decoder (PPD) to refine the segmentation of polyps from

colonoscopy images. Despite its success in enhancing segmentation precision by leveraging high-level feature aggregation, the model's effectiveness was contingent on access to extensive training datasets and was marked by significant computational demands. In a recent advancement, the Swin-Unet model [18] was introduced, incorporating a transformer architecture akin to the traditional U-Net for medical image segmentation. This innovative approach features a symmetric Swin transformer-based decoder with a patch-expanding layer to restore the spatial resolution of feature maps. Simultaneously, a hierarchical Swin transformer with shifted windows serves as the encoder, adeptly extracting context features. This model signifies a pivotal shift towards incorporating transformer models in medical imaging. Another noteworthy development is the Multi-Inception-U-Net [19], designed to enhance scalability. Tested against the Brain Tumor Segmentation Challenge (BraTS) datasets from 2015, 2017, and 2019, this model demonstrated commendable performance, leveraging the foundational strengths of the U-Net. Specifically, it achieved an accuracy of 83% on the training set and 82% on the validation set using the 3D-UNet architecture for MRI brain tumor segmentation on the BraTS 2020 dataset. Further exploration in the field led to the introduction of the DCSAU-Net [20], a novel, deeper, and more compact split-attention UNet model tailored for medical image analysis. This model excels by integrating high and low-level information through a primary features conservation framework and a compact attention block, showcasing superior performance in medical image segmentation by outperforming existing methods in key metrics. The study also presented the SPP-Net [21], a network designed for brain tumor segmentation that combines spatial pyramid pooling (SPP) and attention blocks, eschewing traditional residual connections. This design choice enables the network to synthesize information across varying down-sampling blocks, thus enhancing the comprehensiveness of the reconstruction process. The attention blocks further enrich the model by integrating global dependencies with local features, culminating in a notable accuracy score on the Brats 2021 dataset. Moreover, the BiTr-Unet [22] model emerged, blending CNN and transformer architectures for segmenting brain tumors using multi-modal MRI data. This model demonstrated exceptional performance on the BraTS 2021 validation dataset, achieving impressive mean Dice scores and Hausdorff distances across different tumor regions. An additional innovative approach combined the U-Net model, CNN, and Self Organizing Feature Map (SOFM) [23] into an ensemble technique for precise brain tumor segmentation. This method applied to the BRATS 2020 dataset, underscores the ongoing exploration and integration of diverse computational techniques to enhance the accuracy and efficiency of medical image segmentation.

In a comprehensive analysis, the study [24] contrasts two advanced object detection frameworks, YOLOv5 and YOLOv7, applied to brain cancer classification and detection within MRI scans. This investigation covers three types of brain tumors: meningiomas, gliomas, and pituitary tumors, using sophisticated mask alignment techniques for accurate tumor segmentation. Research [25] introduces an enhanced Res-UNet model, integrating attention-guided and scale-aware strategies for precise brain tumor segmentation. The model incorporates a mechanism for capturing essential contextual information through attention and feature fusion, alongside a strategy for unearthing and dynamically integrating multi-scale features, boosting feature detection and enhancement across various scales. Study [26] unveils the M2FTrans framework, designed to merge cross-modality features in multi-modality settings, even when some modalities are missing. It uses learnable fusion tokens and masked self-attention to maintain long-range modality dependencies flexibly and introduces spatial and channel-wise fusion transformers for balancing and reducing redundancy across modal features. Finally, [27] proposes a brain tumor segmentation method employing a multimodal transformer network adapted to handle incomplete MRI data sets. This method is built upon a U-Net framework featuring modality-specific encoders, a multimodal transformer for correlating features across modalities, and a shared-weight decoder. This approach aims to seamlessly integrate multimodal and multi-level features for enhanced brain tumor segmentation.

3 PROPOSED METHODOLOGY

Meta-learning introduces a powerful paradigm for brain tumor segmentation by equipping models with the ability to rapidly adapt to new, specific tasks of identifying and segmenting brain tumors from minimal examples. By drawing on previously acquired knowledge and optimizing the learning process itself, meta-learning enables the creation of highly flexible and efficient segmentation models. This approach is particularly valuable in the medical imaging field, where obtaining large, annotated datasets is challenging, and the ability to quickly adjust to different tumor types or rare conditions can significantly enhance diagnostic accuracy and patient care. This research work adopts the framework of meta-learning as a base and develops a novel architecture MAA+AGDL to enhance the performance of brain tumor segmentation; the MAA+AGDL model includes designing the MAA architecture, feature

reshaping module, optimal features for each class module, Generation of multiple feature relations maps, a novel AGDL mechanism with DFCM.

3.1 Pre-Processing

In the proposed methodology for enhancing brain tumor detection in MRI images, a comprehensive preprocessing strategy plays a pivotal role in ensuring the accuracy and reliability of subsequent analyses. This strategy encompasses several key steps designed to refine and standardize the MRI data, preparing it for effective tumor segmentation and identification. The steps include:

- **Image Rescaling:** Initially, all MRI scans are rescaled to a uniform resolution of 512 by 512 pixels. This standardization is crucial for consistent analysis across various scans, facilitating the comparison and evaluation of tumor features within a unified framework.
- **Intensity Normalization:** Due to the inherent variability in image intensities and brightness among MRI scans, a normalization process is applied to standardize the intensity levels across all images. This step ensures that the images can be accurately analyzed, with consistent brightness and contrast levels that are conducive to identifying tumor characteristics.
- **Noise Reduction:** Recognizing the sensitivity of brain tumor images to noise, a targeted noise reduction technique is employed to enhance image quality. The use of a median filter is selected for its effectiveness in reducing noise without compromising the integrity of the tumor features. This approach significantly improves the visibility of tumors and aids in their precise segmentation.

3.2 Problem Definition

The Brain Tumor Segmentation model, employing meta-learning, focuses on training the model with a large dataset $\mathcal{F}_b$ comprising numerous labeled samples whose classes $\mathcal{E}_b$ are known. This enables the model to quickly implement segmentation for previously unknown classes, denoted as $\mathcal{E}_n$, in which case $\mathcal{E}_b \cap \mathcal{E}_n$ is equal to Null. To train the proposed model, a meta-learning-based segmentation technique is utilized. Specifically, for each instance of limited data meta-learning, we employ a random sampling method to divide the data into novel classes $\mathcal{F}_n = \{(U, S)\}$ for the support image dataset $U = \{(K_u^k, \mathcal{O}_u^k)\}_{k=1}^M$ and the query dataset $S = \{(K_s^k, \mathcal{O}_s^k)\}_{k=1}^{P_s}$, where the novel class dataset comprises P classes and includes two masked images as pairs $(K_u^k, \mathcal{O}_u^k)$ for a given period. Here, K_u^k as well as $\mathcal{O}_u^k$ represent the k -th image and its corresponding segmentation, respectively. However, the $\mathcal{O}_s^k$ is primarily used for training purposes. M denotes the number of masked image pairs input for each class within the support images, and P_s indicates the number of queries (where $M = 5$ or 1). Furthermore, u and s represent the pairs of masked images related to the support data U as well as the query data S . It is noted that the input for the proposed model consists of a masked image pair from the support data and a query image, while the output is the predicted binary masked segmentation $\widehat{\mathcal{O}}_s$ for the query image. In this study, each instance of limited data meta-learning is applied to prior studies and experimental setups.

3.3 Architecture

The architecture illustrated in Figure 1 demonstrates the functionality of the proposed study, which is divided into two parallel paths leading to analogous operations. One path is designated for the prediction of foreground regions, while the other is dedicated to predicting background regions. Each path comprises the following components:

- A feature encoder for attribute sharing is utilized to extract attributes from both query and support images.
- The generation of various descriptors to introduce a specific descriptor that fully represents the class's attribute distribution.
- The prediction is based on different Feature Relations maps for the foreground and background areas in the query images.

Specifically, the support and query images are fed into the attribute-sharing encoder, which extracts feature vectors from the query and support data. This encoder, pre-trained and based on a ResNet architecture, facilitates attribute retrieval. The proposed method incorporates Structure Refinement and Boundary Extension operations on masked support images to generate images with structure refinement-supported masks and Boundary extension-supported masks, respectively. The structure refined support masked images are subtracted from the original image, and the original support masked image is subtracted from the dilated masked image, yielding inner and outer boundary foreground masks. For background area prediction, average pooling is applied to the support masked images,

encompassing both background and exterior boundary areas, to generate respective prototypes. Subsequently, element-wise multiplication is performed between the support attributes and background masking to obtain a background attribute with foreground values set to zero. These zero values are then removed from the background attribute, and a constant number of background attribute points, equivalent to the number of foreground points, are sampled. These points are assigned to areas with zero value in the background areas. The attribute vector is reshaped, with each point associated with the background passing through a generic multi-layer perceptron (MLP) to generate various descriptors. Finally, Feature Relations maps computation represents background and out-boundary prototypes, descriptors specific to the background, and query attributes. This data is input into lightweight decoders to predict background areas. The operations for foreground path prediction follow a similar process. Figure 2 shows the proposed MAA + ADLC architecture.

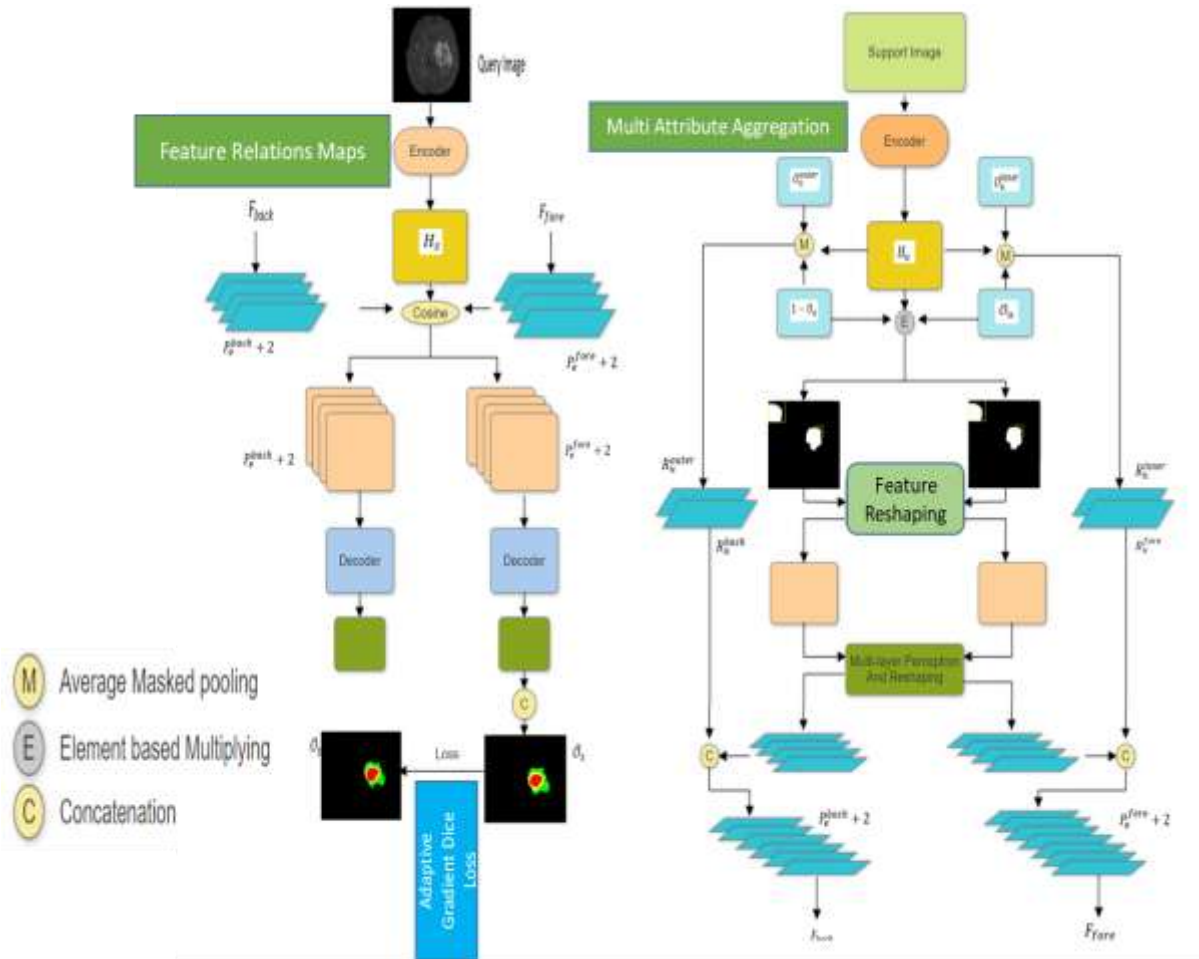


Figure 2 proposed MAA+ AGDL architecture

3.4 Multi-Attribute Aggregation

Initially, the network structure is frozen in the attribute-sharing encoder, denoted as $h_2(\cdot)$ which is used to retrieve attribute vectors for both input support and query images, $\mathfrak{A}$ represents the network attribute. The attributes for the support and query images are represented as $H_u = h_2(K_u)$ and $H_s = h_2(K_s)$ respectively, where H_u and H_s belong to $\mathbb{T}^{J \times Y \times e}$, here J and Y represent the height and width of the attributes, respectively, and e denotes the channel depth of the attribute.

To address common issues related to segmentation and blurry boundaries, a technique is proposed for generating approximate boundary masks to improve the segmentation accuracy of the foreground class boundary. Initially, image processing methods, including structure refinement and Boundary extension, are applied to the support

mask $\mathcal{O}_u$. This results in the SRF (structure refinement) of the support mask $\mathcal{O}_u^{erosion}$ and the BE (Boundary extension) support mask denoted as $\mathcal{O}_u^{dilation}$. The structure refinement support mask is then subtracted from the initial support mask, and similarly, the initial support mask is subtracted from the dilated mask. These operations generate masks for the inner and outer boundaries in the foreground, denoted as $\mathcal{O}_u^{inner}$ and $\mathcal{O}_u^{outer}$, respectively. The process is formulated using the following equations (1).

$$\mathcal{O}_u^{SRF}(j, y) = \underset{(z, a) \text{ belongs to } D}{\text{minimum}} \mathcal{O}_u(j + z, y + a) \quad (1)$$

$$\mathcal{O}_u^{BE}(j, y) = \underset{(z, a) \text{ belongs to } D}{\text{maximum}} \mathcal{O}_u(j + z, y + a)$$

Given the equations mentioned earlier (j, y) represents the initial mask's position, and (z, a) denotes the pixel placement within the element D , which is a matrix kernel with dimensions 3 by 3.

$$\mathcal{O}_u^{inner} = \mathcal{O}_u - \mathcal{O}_u^{SRF} \quad (2)$$

$$\mathcal{O}_u^{outer} = \mathcal{O}_u - \mathcal{O}_u^{BE}$$

It is observed that the inner and outer boundary masks, denoted as $\mathcal{O}_u^{inner}$ and $\mathcal{O}_u^{outer}$ are retrieved using the specified formulation. This process is crucial for predicting the paths in the maps of the foreground and background, respectively. In line with traditional practices for the proposed method, retrieving the main data from the foreground and background areas of the support images plays a vital role. An average pooling operation is performed on the support attribute using the masks of both the background and foreground. This operation yields the prototypes of the foreground and background, denoted as r_u^{fore} and r_u^{back} , respectively. This is expressed as given below in equation (3).

$$r_u^k = \left(\sum_{j, y} H_u(j, y) \cdot 1[\mathcal{O}_u(j, y) = l] \right) \left(\sum_{j, y} 1[\mathcal{O}_u(j, y) = l] \right)^{-1} \quad (3)$$

In the context described, the notations, $l = 1$ while k is *fore*, serve to differentiate the evaluations applied to the foreground and background. While $l = 0$ while k is *back*, implying that the evaluation is applied to foreground and $1[\cdot]$ is used to indicate that it is set to 1 if the condition inside the bracket is true, else equal is 0. The evaluation of prototypes for the foreground and background is elaborated in equation (3), where average pooling operations are conducted on the support attribute. This operation leverages the previously obtained foreground support masks, both outer (r_u^{inner}) and (r_u^{outer}), to generate prototypes for the inner supporting boundary. These prototypes are instrumental in distinctly differentiating the data for the inner and outer boundaries of the targeted object. The formulation can be expressed as:

$$r_u^v = \left(\sum_{j, y} H_u(j, y) \odot \mathcal{O}_u^v(j, y) \right) \left(\sum_{j, y} \mathcal{O}_u^v(j, y) \right)^{-1} \quad (4)$$

Considering equation (4), as a reference point, the symbol $\odot$ represents the Hadamard product, a term used to denote element-wise multiplication between two matrices of the same dimensions. In this context $v = inner$, r_u^{inner} denotes the prototype for the inner boundary associated with the supporting image, evaluated using the corresponding mask $\mathcal{O}_u^{inner}$ and $v = outer$, r_u^{outer} denotes the prototype for the outer boundary associated with the supporting image, evaluated using the corresponding mask $\mathcal{O}_u^{outer}$. The assumption is made that in the worst-case scenario, one of the prototypes loses a part of the comprehensive data, leading to an inadequate representation of the complete feature for class distribution. To address this, the generation of various descriptors for representation is

proposed, which is distinct from traditional practices. This ensures that both the foreground and background areas of the support attribute are sampled during the relative operation at random. Consequently, the foreground and background are masked on the support attribute to result in the support attribute for the foreground denoted as H_u^{fore} and background H_u^{back} , which is given below in equation (5).

$$H_u^k = H_u \odot O_u^k \quad (5)$$

In this context, where the element-wise multiplication is denoted by $\odot$. While $k = fore$, we use $O_u^{fore} = O_u$ and the evaluated output is the support attribute for foreground and when $k = back$, we use $O_u^{back} = 1 - O_u$ and the evaluated output is the support attribute for the foreground. We observe that the values for the points of the background for the support attribute in the foreground are = 0, and vice-versa.

3.5 Feature Reshaping Module

To address the task of balancing the number of feature points between the foreground and background after having filtered out zero-value features and reshaping the support attributes, we proceed with the formulation of the feature reshaping as described in algorithm 1. This technique aims to ensure a consistent number of feature points. $R_{fore} = [r_{fore}^1, r_{fore}^2, \dots, r_{fore}^{P_{fore}}]$ belongs to $\mathbb{T}^{P_{fore} \times e}$ and $R_{back} = [r_{back}^1, r_{back}^2, \dots, r_{back}^{P_{back}}]$ belongs to $\mathbb{T}^{P_{back} \times e}$, P_{fore} and P_{back} denote the complete count of points in the fore as well as background respectively. To ensure a consistent number of feature points between the foreground and background attributes, a random sampling method is applied to select P_{fore} and P_{back} from their respective sets of feature points, where $\widetilde{R}_{fore}$ belongs to $\mathbb{T}^{P_{back} \times e}$ are selected directly from the attribute points of the foreground at random. Contrarily, the attribute points of the foreground are repeated p times before choosing t samples at random from the feature points of the foreground. These datasets of feature points for the foreground are combined resulting in the P_{back} samples utilized for occupying the areas having zero values for foreground attributes. This technique is formulated mathematically as given below in equations (6) and (7).

$$\widetilde{R}_{fore} = \mathcal{H}_{sample}(R_{fore}, P_{back}) \quad (6)$$

While P_{fore} is lesser than P_{back} :

$$o = \lfloor (P_{back})(P_{fore})^{-1} \rfloor$$

$$t = P_{back} - oP_{fore} \quad (7)$$

$$R'_{fore} = \mathcal{H}_{sample}(R_{fore}, t)$$

$$\widetilde{R}_{fore} = \widetilde{R}_{fore} \oplus \widetilde{R}_{fore} \oplus \dots \oplus R'_{fore} \oplus \widetilde{R}_{fore}$$

In this case, $P_{fore} + P_{back}$ is equivalent to JY , the function that chooses p at random is denoted as $\mathcal{H}_{sample}(Z, p)$ from the point Z , the operation is denoted by $\lfloor \cdot \rfloor$, the concatenation for the dimension elements is represented as $\oplus$. The sampling technique at random for the attribute points for the background are same as given above and also observed in the algorithm explained below. However, we attain P_{fore} samples where $\widetilde{R}_{fore}$ belongs to $\mathbb{T}^{P_{fore} \times e}$, to occupy the background attribute.

Algorithm 1: Feature reshaping algorithm

Algorithm	Feature Reshaping algorithm
Input	Features or attributes in the foreground H_u^{fore}
Step 1:	The count of points for foreground as well as background is given as P_{fore} and P_{back}
Step 2:	Eliminate the feature or attribute points that have the value zero in H_S^{fore} and Retrieve the feature points of the foreground R_{fore} from H_S^{fore}
Step 4	Choosing P_{back} samples at random $\widetilde{R_{fore}}$ from R_{fore}
Step 5	If P_{fore} is greater than or equal to P_{back} then Choose P_{back} samples at random $\widetilde{R_{fore}}$ from R_{fore} , Else:
Step 8	Obtain o by dividing P_{back} by P_{fore} using equation (7)
Step 9	The remainder t of P_{back} over P_{fore} using equation (7)
Step 10	Choose t samples at random $\widetilde{R'_{fore}}$ from R_{fore} from equation (7)
Step 11	R_{fore} is repeated o times and combined with $\widetilde{R'_{fore}}$ to result in $\widetilde{R_{fore}}$ End if
Step 13	Attain the index position Z_1 for the background in H_u^{fore} from equation (8)
Step 14	Occupy the positions that are in the background in H_u^{fore} with $\widetilde{R_{fore}}$ relating to Z_1
Output	Reshaped foreground H_u^{fore} feature

The foreground attributes H_u^{fore} (H_u^{back}) along with the zero numbered back(fore) ground area is occupied with samples of fore(back) ground attribute points $\widetilde{R_{fore}}$ ($\widetilde{R_{back}}$). This process assures that the points of the feature for the foreground H_u^{fore} are relative to the foreground, while the entire attribute points for the background H_u^{back} is related to the background. This procedure is expressed as given below in equation (8) and (9).

$$\begin{aligned} Z_0 &= \mathcal{V}[H_u^{back} = 0] \\ Z_1 &= \mathcal{V}[H_u^{fore} = 0] \end{aligned} \quad (8)$$

$$\begin{aligned} H_u^{back}[Z_0] &= R_{back}^{sp} \\ H_u^{fore}[Z_1] &= R_{fore}^{sp} \end{aligned} \quad (9)$$

Considering the equations given above, the index of spatial location for the back, as well as foreground, is denoted as Z_0 and Z_1 , respectively. The logic function is given as $\mathcal{V}[\cdot]$ and is used for the determination if the particular condition inside the bracket is true or not.

3.6 Dual Network for optimal features in each class

For generating various descriptors that fully capture the feature data for each class, the features of the occupied foreground and background are passed through two lightweight multi-layer perceptrons (MLPs) which are denoted as

$\mathbb{M}$ that have a straightforward structure. In this way, the representation descriptors for the foreground and background classes are created. These are viewed as types of various prototypes, as detailed below in equation (10).

$$\begin{aligned} f_{fore} &= \mathbb{M}(H_u^{fore}; \theta_1) \\ f_{back} &= \mathbb{M}(H_u^{back}; \theta_0) \end{aligned} \quad (10)$$

Here, f_{fore} belongs to $\mathbb{T}^{P_e^{fore} \times e}$ and f_{back} belongs to $\mathbb{T}^{P_e^{back} \times e}$ expresses P_f^{fore} generated descriptors for foreground and P_f^{back} generated descriptors for background. The occupied attributes for the foreground as well as background are given as H_u^{fore} and H_u^{back} which are shaped again according to the size of $JY \times e$ and the $\mathbb{M}(\cdot; \theta_*)$ which is a function for the multi-layer perceptron that has attributes θ_* . We observe, the multi-layer perceptron has a structure that consists of a completed linked layer, an activation layer (ReLU) as well as another completely linked layer, that maps a constant count of attribute points for various represented descriptors. In conclusion, the combination of the normally represented descriptors for the foreground along with its supporting prototype r_u^{fore} as well as the prototype for the inner boundaries r_u^{inner} from the equation (3) and equation (4) that results in F_{fore} . In the same manner, we combine the normally represented descriptors for background along with its supporting prototype r_u^{back} as well as the prototype for the outer boundaries r_u^{outer} from the equation (3) and equation (4) that results in F_{back} . This procedure is explained using the equations that are given below:

$$\begin{aligned} F_{fore} &= f_{fore} \oplus r_u^{fore} \oplus r_u^{inner} \\ F_{back} &= f_{back} \oplus r_u^{back} \oplus r_u^{outer} \end{aligned} \quad (11)$$

In this case, F_{fore} belongs to $\mathbb{T}^{(P_e^{fore}+2) \times e}$ and F_{back} belongs to $\mathbb{T}^{(P_e^{back}+2) \times e}$ that will be utilized for the path predictions in the fore as well as background maps related to query pictures.

3.7 Multiple Feature Relations Maps

A two-path architecture as well as combines the Feature Relations maps produced by various descriptors for the prediction queries of back as well as foreground maps. Therefore, this efficiently decreases the miscategorization of the model. Firstly, the affinities are evaluated individually for the various generated fore as well as background descriptors as given below in equation (12).

$$C_{fore}(j, y) = (F_{fore}^V \cdot H_s(j, y))(\|F_{fore}\| \cdot \|H_s(j, y)\|)^{-1} \quad (12)$$

$$C_{back}(j, y) = (F_{back}^V \cdot H_s(j, y))(\|F_{back}\| \cdot \|H_s(j, y)\|)^{-1}$$

In this case, C_{fore} belongs to $\mathbb{T}^{(P_e^{fore}+2) \times J \times Y}$ and C_{back} belongs to $\mathbb{T}^{(P_e^{back}+2) \times J \times Y}$ are used to denote Feature Relations maps of the foreground as well as the background, respectively. Further, a decoder is developed for the concatenation of the maps C_{fore} and C_{back} . The concluding map of the foreground $\widehat{\mathcal{O}}_{fore}$ is generated using this decoder and the map of the background $\widehat{\mathcal{O}}_{back}$ is generated using the following equation (13).

$$\begin{aligned} \widehat{\mathcal{O}}_{fore} &= \text{Decoding}(C_{fore}; \rho_1) \\ \widehat{\mathcal{O}}_{back} &= \text{Decoding}(C_{back}; \rho_0) \end{aligned} \quad (13)$$

Here, $Decoding(\cdot; \rho_*)$ is used to express the function for a convolutional neural network having a lightweight along with the attribute ρ_* , and the result is an individual map for prediction $\widehat{\mathcal{O}}_*$ belongs to $\mathbb{T}^{I \times Y}$. In conclusion, we concatenate the maps that are predicted finally using the operation of concatenation as well as the function of softmax such as given below in equation (14).

$$\widehat{\mathcal{O}} = \alpha(\widehat{\mathcal{O}}_{back} \oplus \widehat{\mathcal{O}}_{fore}) \quad (14)$$

For the equation (14) given above, $\alpha(\cdot)$ is used to denote the softmax that is used, and the final map that is predicted is given as $\widehat{\mathcal{O}}$.

3.8 Adaptive Gradient Dice Loss (AGDL)

Our proposed methodology introduces a novel approach to enhance the segmentation accuracy of medical images through the implementation of Adaptive Gradient Dice Loss and Deformable Convolution-based Feature Alignment (DCFA). These advancements address key challenges in conventional segmentation methods, particularly in handling gradient vanishing issues and improving feature alignment across different image scales.

Adaptive Gradient Dice Loss is designed the GODL function is a pivotal innovation designed to mitigate the instability associated with the Dice loss during the model training phase. The Dice loss is formulated as given in equation (15).

$$Dice_{Loss} = 1 - \frac{2|A \cap B|}{|A| + |B|} \quad (15)$$

Where A and B represent the predicted and ground truth segmentation masks, respectively. To address the gradient vanishing problem inherent in the Dice loss, we introduce the GODL, which modifies the loss function based on the gradient's magnitude during backpropagation. This ensures a consistent and robust optimization path. The AGDL is mathematically represented as given in equation (16).

$$AGDL = (Dice_{Loss})X(1 + \eta(\nabla \mathcal{O})) \quad (16)$$

Where ∇ is a tuning parameter that controls the degree of gradient optimization, and $\nabla \mathcal{O}$ signifies the magnitude of the gradient of the loss function, facilitating enhanced learning dynamics.

3.8.1 Deformable Convolution-based Feature Alignment (DCFA)

Incorporates deformable convolution layers to enable adaptive feature alignment across different scales. This module effectively captures and aligns disparate features, crucial for accurately segmenting intricate details in medical images. The alignment process is governed by the deformable convolution operation, which adjusts the kernel's spatial sampling locations according to the learned offsets, thereby optimizing the feature fusion across layers for improved segmentation performance.

4 PERFORMANCE EVALUATION

The methodology for brain tumor detection through MRI scans employs PyTorch for model development and training, utilizing the computational power of an Nvidia GeForce RTX 3090 graphics card. This setup provides a robust platform for handling the complexities of neural network training and image processing tasks. The methodology encompasses both training and testing phases, meticulously designed to optimize performance:

A. Training Stage:

- **Batch Size:** Set to 2, this small batch size allows for efficient memory usage and more frequent updates of the model weights, which can be beneficial for learning fine-grained details in the MRI images.
- **Optimizer:** The Adam optimizer is chosen for its effectiveness in handling sparse gradients and adapting the learning rate for different parameters, with an initial learning rate of $1e-5$. This choice is aimed at achieving a balanced optimization process that is both robust and sensitive to the nuances of the training data.
- **Epochs and Learning Rate Decay:** The model is trained over 200 epochs to ensure sufficient exposure to the training data. The learning rate is designed to decay by a factor of 0.5 every 10 epochs, which helps in fine-tuning the model's parameters as the training progresses, ensuring steady convergence.

- **Early Stopping:** To mitigate the risk of overfitting, an early stopping mechanism is employed. Training is halted if the validation loss does not improve for 50 consecutive epochs, ensuring that the model retains its generalization ability without memorizing the training data.

B. Testing Stage:

The efficiency of the model is highlighted by its quick inference time, where testing a single sample takes approximately 0.4 seconds. This rapid testing capability indicates the model's practical applicability in real-world settings, where timely diagnosis can significantly impact treatment outcomes.

C. Dataset Details

The evaluation of the proposed model for brain tumor detection utilizes data from the Brain Tumor Segmentation Challenge (BRATS) 2019 and 2020. These datasets are pivotal for benchmarking the performance of segmentation models due to their comprehensive collection of MRI sequences and annotated cases.

- **BRATS2019 Dataset [28]:** Contains 335 cases, providing a substantial volume of data for training and validation. The dataset includes four types of $240 \times 240 \times 155$ MRI sequences: FLAIR, T1, T2, and T1ce. The annotations available cover various tumor regions: necrotic and non-enhancing tumor (NET), enhancing tumor (ET), peritumoral edema (ED), and the non-tumor background (BG).
- **BRATS2020 Dataset [29]:** Comprises 369 cases, each featuring the same four MRI sequence types as the 2019 dataset, further enriching the diversity and volume of data available for model evaluation. Similar to the 2019 dataset, annotations are provided for distinct tumor subregions, facilitating detailed analysis and segmentation performance assessment.

The model's performance is meticulously assessed across three critical tumor subregions: the whole tumor (WT), the tumor core (TC), and the enhancing tumor (ET). This approach ensures a comprehensive evaluation of the model's ability to detect and segment various components of brain tumors, reflecting its practical utility in clinical settings. A five-fold cross-validation method is employed across the two datasets to ensure the robustness and reliability of the evaluation results. For the BRATS2019 dataset, 268 cases are designated for training, with the remaining 67 cases allocated for testing. Similarly, for the BRATS2020 dataset, 295 cases are used for training, and 74 cases are set aside for testing. This division allows for a balanced distribution of data across training and testing sets, ensuring that the model is exposed to a wide variety of cases during training and evaluated on a representative sample of the dataset.

4.1 Evaluation Metrics

To evaluate the proposed brain tumor segmentation model, four key metrics are utilized:

- **Dice Similarity Coefficient (DSC):** Measures the overlap between the model's predictions and the ground truth. Higher values indicate better accuracy.
- **Sensitivity:** Assesses the model's ability to correctly identify tumor regions, with higher values indicating fewer missed tumor areas.
- **Specificity:** Evaluates the accuracy in identifying non-tumor regions, with higher values showing precise exclusion of non-tumor areas.
- **Hausdorff Distance (HD95):** Quantifies the accuracy of the tumor boundary predictions, with lower values indicating more precise boundary segmentation.

These metrics provide a comprehensive assessment of the model's performance in segmenting brain tumors, comparing its accuracy and precision against existing models.

4.2 Experiment Results

The verification of the Proposed Multi-Attribute Aggregation (MAA) and an Adaptive Gradient Dice Loss (AGDL) (MAA+AGDL) involves a comprehensive comparison with leading segmentation methods on the BRATS2020 dataset. The models compared included U-Net [30], SwinBTS [31], nnU-Net [32], SA-Net [33], H2NF-Net [34], TransBTS [35], BiTr-Unet [36], SegTransVAE [37], mmFormer [38], UNETR [39], Attention U-Net [40], SA-LuT-Net [41], the studies by Wang et al. [42], Li et al. [43], and both CorrDiff and the enhanced CorrDiff+ [44]. This evaluation showcased the MAA+AGDL's performance against a wide array of state-of-the-art segmentation technologies, highlighting its efficacy and innovation in brain tumor segmentation on a well-recognized dataset.

The visual comparison provided in Table 1 from BRATS 2019 displays various brain tumor segmentation models alongside the ground truth. The segmentation accuracy of each model can be assessed by the proximity of its highlighted tumor regions to the ground truth. The MAA+AGDL appears to closely match the ground truth, indicating a high degree of accuracy in tumor detection and delineation. It demonstrates precise segmentation of the tumor's core and periphery, aligning well with the ground truth and showing clear improvements over other methods like Segmentation U-Net, TransBTS, mmFormer, BiTr-Unet, and CorrDiff+. The MAA+AGDL's segmentation is notably more consistent and contiguous, suggesting superior performance in identifying and outlining the tumor area.

Table 1 Brats 2019

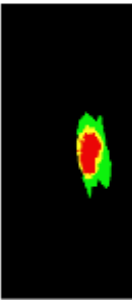
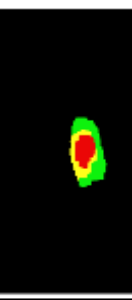
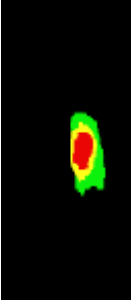
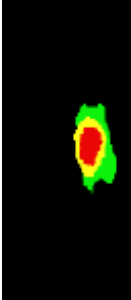
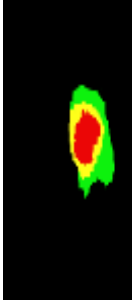
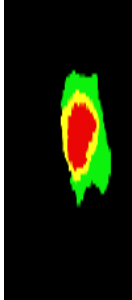
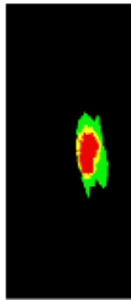
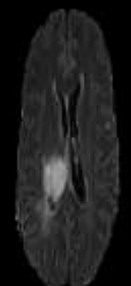
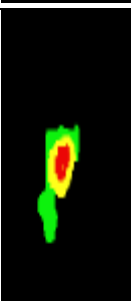
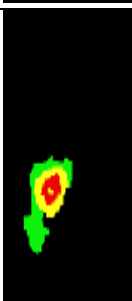
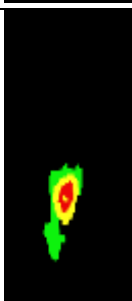
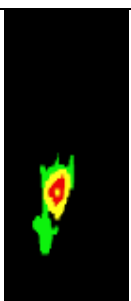
Image	Ground Truth	Segmentation U-Net	TransBTS	mmFormer	BiTr-Unet	COrrDiff+	Proposed(MAA+AGDL)
							
							

Table 2 displays segmentation results on the BRATS 2020 dataset across multiple models in comparison to the ground truth. Observations indicate that the proposed model achieves highly accurate segmentation, with tumor boundaries that closely resemble those in the ground truth. The consistency and continuity in the tumor regions segmented by the proposed model suggest that it effectively captures both the core and peripheral tumor areas. Other models, such as the Segmentation U-Net, TransBTS, mmFormer, BiTr-Unet, and CorrDiff+, show varying degrees of congruence with the ground truth. However, the proposed model seems to provide the best match, offering fewer false positives and negatives, which implies improved sensitivity and specificity. The precision in boundary delineation also appears to be superior in the proposed model, likely resulting in higher Dice scores and lower Hausdorff distances, although these would need to be confirmed with quantitative metrics. Overall, the proposed model demonstrates a strong ability to segment brain tumors accurately, which is essential for effective diagnosis and treatment planning.

Table 2 Brats 2020 dataset

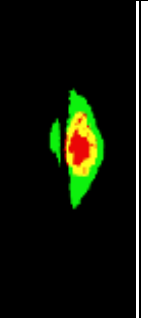
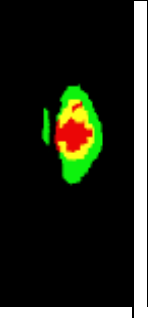
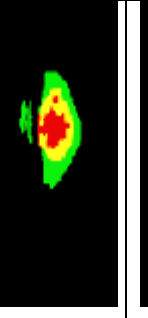
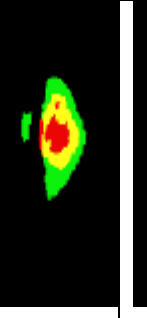
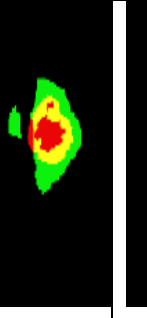
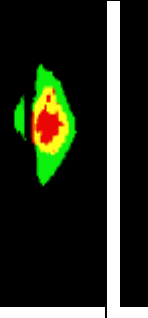
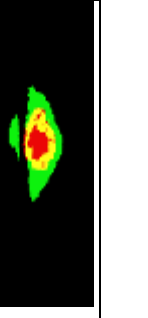
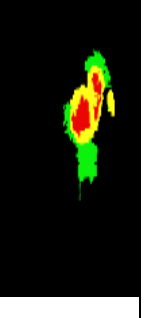
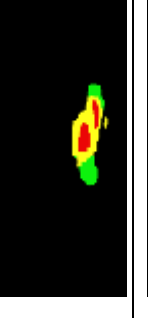
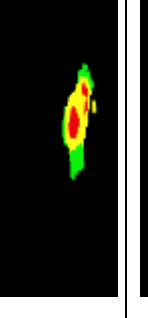
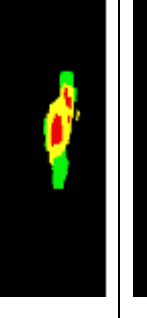
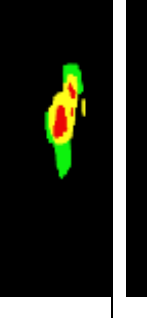
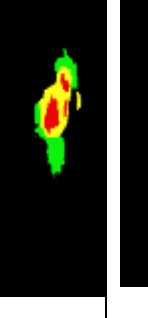
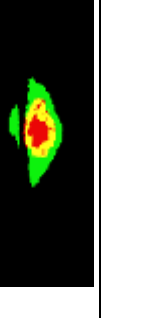
Image	Ground Truth	Segmentation on U-Net	TransBTS	mmFormer	BiTr-Unet	COrrDiff+	Proposed(MAA+AG DL)
							
							

Table 3 shows the comparison of Dice Similarity Coefficient (DSC) scores across various segmentation models on the BRATS 2019 dataset and highlights the superior performance of the proposed model, especially in segmenting enhancing tumor (ET), whole tumor (WT), and tumor core (TC) regions. With DSC scores of 93.64 for ET, 96.54 for WT, and 93.54 for TC, the proposed model significantly outperforms other state-of-the-art methods. Notably, BiTr-Unet and CorrDiff also show commendable results, but the proposed model's scores are markedly higher, demonstrating its enhanced capability in accurately segmenting brain tumors. This superiority is particularly evident in its comprehensive detection across all tumor subregions, indicating a significant advancement in brain tumor segmentation technology; a graphical comparison is presented in Figure 3.

Table 3 BRATS 2019 DSC comparison

Methods	ET (DSC)	WT (DSC)	TC (DSC)
U-Net	73.34	88.26	75.98
CANet	75.9	88.5	85.1
Attention U-Net	75.96	88.81	77.2
SA-LuT-Net	78.21	90.79	84.82
Wang et al.	73.7	89.4	80.7

Li et al.	77.1	88.6	81.3
TransBTS	78.9	90.01	81.9
BiTr-Unet	81.09	89.46	81.98
SegTransVAE	80.85	90.15	85.25
mmFormer	77.62	89.43	83.56
UNETR	80.28	90.23	84.91
CorrDiff	81.23	90.47	85.87
Proposed(MAA+AGDL)	93.64	96.54	93.54

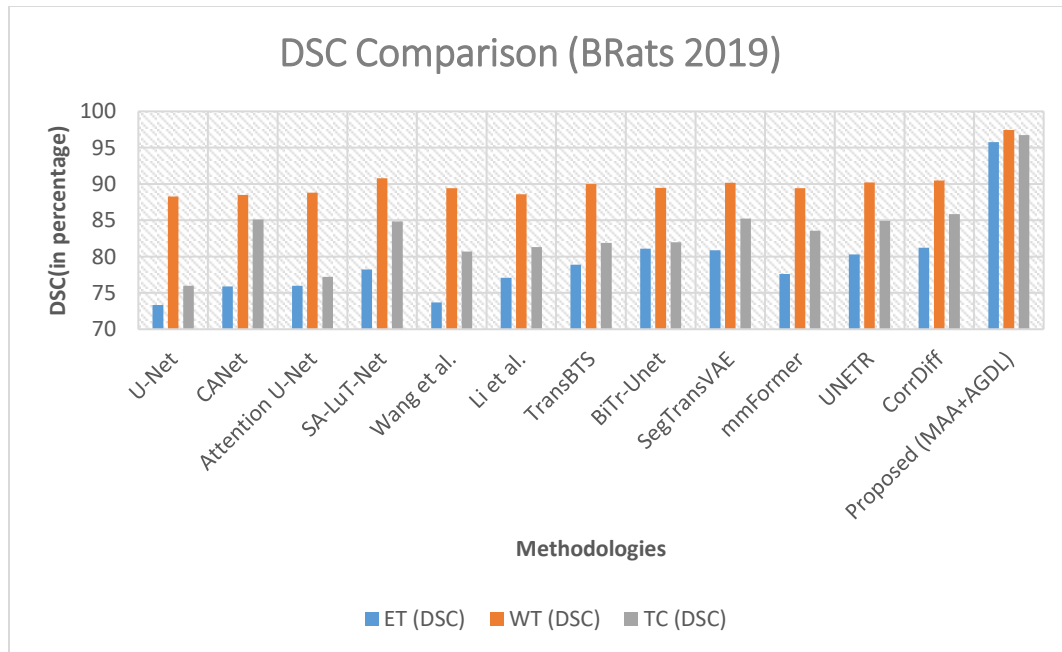


Figure 3 DSC comparison on BRATS 2019

Table 4 presents BRATS 2019 dataset's Hausdorff Distance (HD95) comparison and underscores the proposed model's exceptional precision in delineating tumor boundaries across enhancing tumor (ET), whole tumor (WT), and tumor core (TC) regions. Achieving HD95 scores of 3.25 for ET, 2.64 for WT, and 3.8 for TC, the proposed model demonstrates unparalleled accuracy in capturing the intricate details of tumor margins when compared to other leading methods. This is notably superior to the next best performances, such as those from SA-LuT-Net and SegTransVAE, indicating a significant advancement in ensuring precise tumor segmentation. The proposed model's ability to minimize the maximum distance of the segmented boundary from the ground truth, even in complex tumor structures, sets a new benchmark in the field. A graphical comparison is given in Figure 4.

Table 4 BRATS 2019 HD95 comparison

Methods	ET (HD95)	WT (HD95)	TC (HD95)
U-Net	6.21	6.17	8.68
CANet	4.81	7.09	8.41
Attention U-Net	5.2	7.76	8.26
SA-LuT-Net	3.69	4.46	5.26
Wang et al.	5.99	5.68	7.36
Li et al.	6.03	6.23	7.41
TransBTS	4.73	5.64	6.04
BiTr-Unet	5.76	4.62	7.82
SegTransVAE	5.14	4.47	5.67
UNETR	5.23	4.53	6.87
CorrDiff	5.49	4.18	6.52
Proposed(MAA+AGDL)	3.25	2.64	3.8

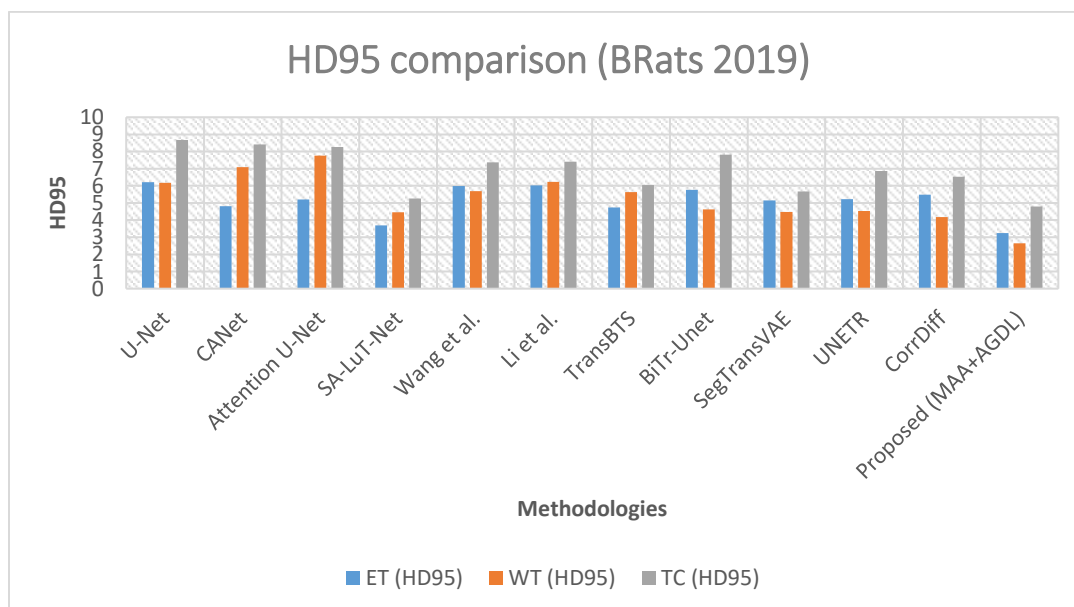


Figure 4 BRATS 2019 HD95 comparison

Table 5 outlines the Dice Similarity Coefficient (DSC) values for various deep learning models applied to brain tumor segmentation on Brats 2020, evaluated across three metrics: Enhancing Tumor (ET), Whole Tumor (WT), and Tumor Core (TC). The DSC value, a statistical measure ranging from 0 to 1, quantifies the similarity between the

predicted segmentation and the ground truth, with 1 indicating perfect agreement. The U-Net model, a foundational architecture in medical image segmentation, shows consistent performance with DSC values of 76.36 for ET, 83.27 for WT, and 81.78 for TC. However, it is outperformed by more advanced models such as nnU-Net and SA-Net, indicating the evolution and improvement in segmentation accuracy. Notably, the CorrDiff+ model demonstrates significant improvement with scores of 82.17 for ET, 92.28 for WT, and 86.21 for TC, suggesting the effectiveness of incorporating correlation and difference strategies in segmentation tasks. The proposed model significantly outshines all others, with DSC values of 92.45 for ET, 97.82 for WT, and 94.28 for TC, indicating a substantial leap in segmentation accuracy. This model's performance underscores the potential of advanced deep learning techniques in accurately delineating tumor boundaries, which is crucial for effective diagnosis and treatment planning. Furthermore, the graphical comparison is given in Figure 5.

Table 5 BRATS 2020 DSC comparison

Methods	ET	WT	TC
U-Net	76.36	83.27	81.78
SwinBTS	77.36	89.06	80.3
nnU-Net	79.45	91.19	85.24
SA-Net	79.27	91.08	85.29
H2NF-Net	78.75	91.29	85.46
TransBTS	78.73	90.09	81.73
BiTr-Unet	82.05	90.16	82.89
SegTransVAE	81.32	90.18	85.57
mmFormer	77.57	89.59	85.69
UNETR	79.81	90.19	83.97
CorrDiff	80.92	91.13	85.06
CorrDiff+	82.17	92.28	86.21
Proposed(MAA+AGDL)	92.45	97.82	94.28

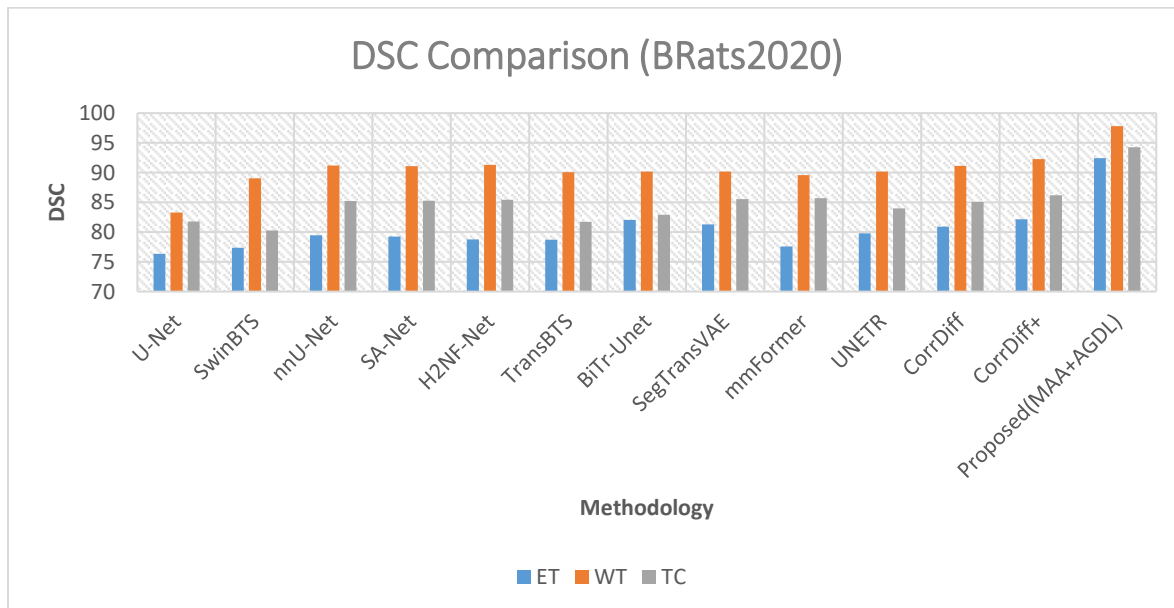


Figure 5 BRATS DSC comparison

Table 6 presents the sensitivity performance of various deep learning models for brain tumor segmentation, focusing on three key areas: Enhancing Tumor (ET), Whole Tumor (WT), and Tumor Core (TC). Sensitivity, a crucial metric in medical imaging, measures the true positive rate, indicating the model's ability to correctly identify tumor regions. The U-Net architecture shows solid baseline performance with sensitivity scores of 76.65 for ET, 88.34 for WT, and 78.65 for TC, demonstrating its reliability in tumor detection. Advanced models such as nnU-Net and SA-Net offer improvements, with nnU-Net achieving higher sensitivity across all categories, particularly 80.97 for ET, 91.17 for WT, and 84.38 for TC, underscoring its efficiency in capturing tumor features. The CorrDiff+ model, with sensitivity values of 83.32 for ET, 92.73 for WT, and 85.14 for TC, represents a significant advancement, suggesting that incorporating correlation and difference strategies can enhance tumor detection accuracy. Remarkably, the Proposed_Model sets a new standard with exceptional sensitivity scores of 94.35 for ET, 96.12 for WT, and 93.84 for TC. This performance highlights the model's superior ability to accurately identify tumor regions, making it a potential game-changer in the field of medical image segmentation. A graphical comparison is presented in Figure 6.

Table 6 BRATS Sensitivity

Methods	ET (Sensitivity)	WT (Sensitivity)	TC (Sensitivity)
U-Net	76.65	88.34	78.65
nnU-Net	80.97	91.17	84.38
SA-Net	77.81	91.08	83.86
H2NF-Net	79.16	91.75	84.27
BiTr-Unet	78.12	90.89	83.48
SegTransVAE	80.27	91.32	83.89
UNETR	78.73	91.21	84.02
CorrDiff	81.78	91.36	84.09
CorrDiff+	83.32	92.73	85.14
Proposed(MAA+AGDL)	94.35	96.12	93.84

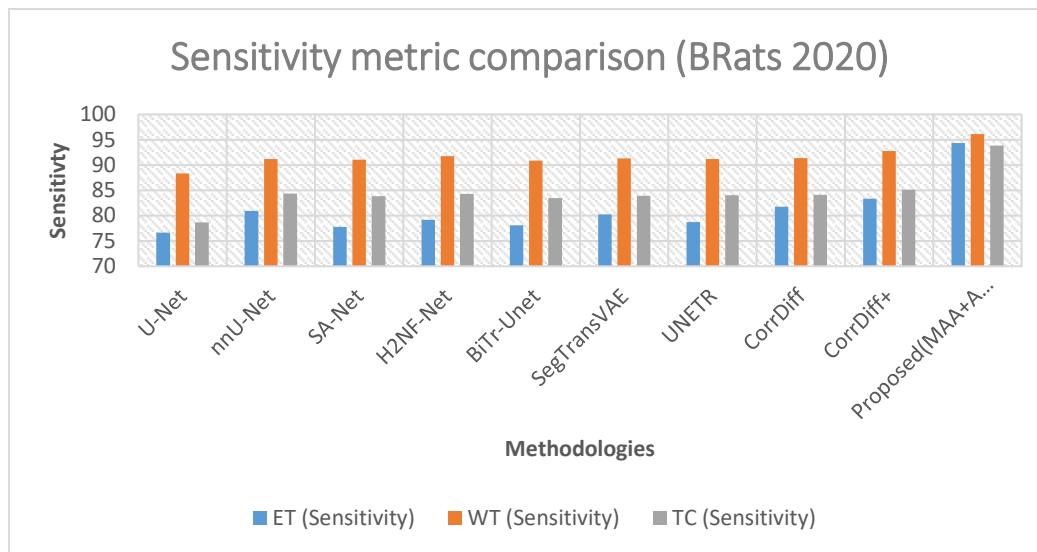


Figure 6 Sensitivity comparison on Brats 2020 dataset

Table 7 presents Hausdorff Distance (HD95) scores for brain tumor segmentation models on the BRATS 2020 dataset and showcases the superior performance of the proposed model across all tumor regions: Enhancing Tumor (ET), Whole Tumor (WT), and Tumor Core (TC). With HD95 scores of 12.89 for ET, 2.65 for WT, and 3.57 for TC, the proposed model demonstrates the highest accuracy in outlining tumor boundaries, significantly outperforming other advanced models including CorrDiff+, BiTr-Unet, and SA-Net. This indicates the effectiveness of the proposed model's novel architectural features and methodologies in capturing the intricate details of tumor margins, thereby offering a significant advancement in the precision of medical image segmentation for brain tumors. Graphical representation is given in Figure 7.

Table 7 HD95 comparison on Brats 2020 dataset

Methods	ET	WT	TC
U-Net	30.21	8.37	16.78
SwinBTS	26.84	8.56	15.78
nnU-Net	29.23	3.79	7.77
SA-Net	18.19	4.09	5.88
H2NF-Net	26.57	4.18	4.97
TransBTS	17.95	4.96	9.76
BiTr-Unet	15.76	4.58	13.78
SegTransVAE	16.97	4.57	5.73
UNETR	17.99	4.75	9.37
CorrDiff	17.89	4.67	8.49
CorrDiff+	15.64	3.68	6.93
Proposed(MAA+AGDL)	12.89	2.65	3.57

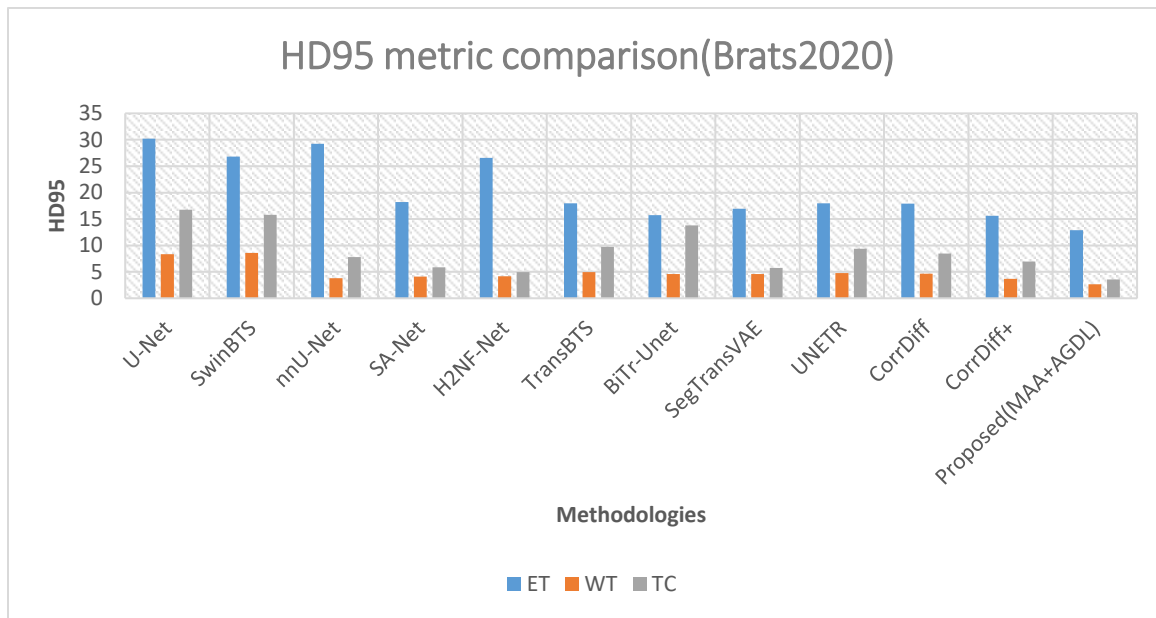


Figure 7 HD95 comarison on BRATs 2020 dataset

4.3 Comparative Analysis

The analysis of the performance improvement in Table 8 reveals significant enhancements by the proposed model over CorrDiff+ across various metrics and datasets. For BRATS 2019 and 2020:

- DSC Improvements are notable, especially for Enhancing Tumor (ET) regions, indicating a more accurate tumor segmentation.
- HD95 Reductions suggest the model's increased precision in outlining tumor boundaries.
- Increased Sensitivity in the 2020 dataset underscores the model's efficiency in detecting true tumor areas.

Overall, these results highlight the model's potential to improve brain tumor diagnosis and treatment planning through superior segmentation accuracy and precision.

Table 8 Performance improvisation of proposed (MAA+AGDL)

Metric / Dataset	ET Improvement	WT Improvement	TC Improvement
DSC (BRATS 2019)	12.41%	4.26%	7.67%
HD95 (BRATS 2019)	-1.94 (Improvement in distance)	-1.54 (Improvement in distance)	-2.07 (Improvement in distance)
DSC (BRATS 2020)	10.28%	5.54%	8.07%
Sensitivity (BRATS 2020)	11.03%	3.39%	8.70%

CONCLUSION

This study introduces a novel deep learning framework for brain tumor segmentation, leveraging Multi-Attribute Aggregation (MAA) and Adaptive Gradient Dice Loss (AGDL) mechanisms. The proposed dual-path architecture, incorporating advanced feature extraction and mapping prediction modules, addresses critical challenges in the segmentation process, such as the differentiation of tumor regions from healthy brain tissue and the accurate delineation of tumor boundaries. The implementation of Adaptive Gradient Dice Loss and Deformable Convolution-based Feature Alignment within this framework significantly improves segmentation accuracy. These innovations offer a robust solution to the gradient vanishing problem and enhance the alignment of features across different image scales, respectively. As a result, the proposed model achieves superior performance on the BRATS 2019 and 2020 datasets, as evidenced by enhanced Dice Similarity Coefficient scores, sensitivity, and reduced Hausdorff Distance. These improvements not only underscore the technical advancements of the proposed model but also highlight its potential to significantly impact clinical diagnostics and treatment planning. Future research directions include exploring the integration of additional imaging modalities to enhance segmentation accuracy.

Conflict of interest: The authors declare no conflict of interest.

Data Availability: Datasets utilized in this research are mentioned in reference [28], [29].

REFERENCES

1. M. Xu, L. Guo and H. -C. Wu, "Novel Robust Automatic Brain-Tumor Detection and Segmentation Using Magnetic Resonance Imaging," in *IEEE Sensors Journal*, doi: 10.1109/JSEN.2024.3367123.
2. S. Karim *et al.*, "Developments in Brain Tumor Segmentation Using MRI: Deep Learning Insights and Future Perspectives," in *IEEE Access*, vol. 12, pp. 26875-26896, 2024, doi: 10.1109/ACCESS.2024.3365048.
3. R. Preetha, M. J. P. Priyadarsini and J. S. Nisha, "Comparative Study on Architecture of Deep Neural Networks for Segmentation of Brain Tumor using Magnetic Resonance Images," in *IEEE Access*, vol. 11, pp. 138549-138567, 2023, doi: 10.1109/ACCESS.2023.3340443.
4. A. Jabbar, S. Naseem, T. Mahmood, T. Saba, F. S. Alamri and A. Rehman, "Brain Tumor Detection and Multi-Grade Segmentation Through Hybrid Caps-VGGNet Model," in *IEEE Access*, vol. 11, pp. 72518-72536, 2023, doi: 10.1109/ACCESS.2023.3289224.
5. F. Isensee *et al.*, "nnU-Net: A self-configuring method for deep learningbased biomedical image segmentation," *Nature Methods*, vol. 18, no. 2, pp. 203–211, 2021, DOI: 10.1038/s41592-020-01008-z.
6. Ahmad, P., Qamar, S., Shen, L., Saeed, A. (2021). Context Aware 3D UNet for Brain Tumor Segmentation. In: Crimi, A., Bakas, S. (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2020. Lecture Notes in Computer Science()*, vol 12658. Springer, Cham. https://doi.org/10.1007/978-3-030-72084-1_19.
7. D. Zhang *et al.*, "Exploring Task Structure for Brain Tumor Segmentation From Multi-Modality MR Images," in *IEEE Transactions on Image Processing*, vol. 29, pp. 9032-9043, 2020, doi: 10.1109/TIP.2020.3023609.
8. B. Mallampati, A. Ishaq, F. Rustam, V. Kuthala, S. Alfarhood and I. Ashraf, "Brain Tumor Detection Using 3D-UNet Segmentation Features and Hybrid Machine Learning Model," in *IEEE Access*, vol. 11, pp. 135020-135034, 2023, doi: 10.1109/ACCESS.2023.3337363.
9. J. Zhao *et al.*, "Uncertainty-Aware Multi-Dimensional Mutual Learning for Brain and Brain Tumor Segmentation," in *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 9, pp. 4362-4372, Sept. 2023, doi: 10.1109/JBHI.2023.3274255.
10. B. Zhan, D. Li, X. Wu, J. Zhou and Y. Wang, "Multi-Modal MRI Image Synthesis via GAN With Multi-Scale Gate Mergence," in *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 17-26, Jan. 2022, doi: 10.1109/JBHI.2021.
11. J. Cheng *et al.*, "Multimodal Disentangled Variational Autoencoder With Game Theoretic Interpretability for Glioma Grading," in *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 2, pp. 673-684, Feb. 2022, doi: 10.1109/JBHI.2021.3095476.
12. Islam, M., Vibashan, V.S., Jose, V.J.M., Wijethilake, N., Utkarsh, U., Ren, H. (2020). Brain Tumor Segmentation and Survival Prediction Using 3D Attention UNet. In: Crimi, A., Bakas, S. (eds) *Brainlesion:*

- Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2019. Lecture Notes in Computer Science(), vol 11992. Springer, Cham. https://doi.org/10.1007/978-3-030-46640-4_25.
13. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov, D., et al. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018 2018. Lecture Notes in Computer Science(), vol 11045. Springer, Cham. https://doi.org/10.1007/978-3-030-00889-5_1.
14. Ma, S., Tang, J., & Guo, F. (2021). Multi-Task Deep Supervision on Attention R2U-Net for Brain Tumor Segmentation. *Frontiers in oncology*, 11, 704850. <https://doi.org/10.3389/fonc.2021.704850>.
15. Li, H., Nan, Y., Yang, G. (2022). LKAU-Net: 3D Large-Kernel Attention-Based U-Net for Automatic MRI Brain Tumor Segmentation. In: Yang, G., Aviles-Rivero, A., Roberts, M., Schönlieb, CB. (eds) Medical Image Understanding and Analysis. MIUA 2022. Lecture Notes in Computer Science, vol 13413. Springer, Cham. https://doi.org/10.1007/978-3-031-12053-4_24.
16. Fan, Deng-Ping & Ji, Ge-Peng & Zhou, Tao & Chen, Geng & Fu, Huazhu & Shen, Jianbing & Shao, Ling. (2020). PraNet: Parallel Reverse Attention Network for Polyp Segmentation. 10.1007/978-3-030-59725-2_26.
17. Cao, H. et al. (2023). Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. In: Karlinsky, L., Michaeli, T., Nishino, K. (eds) Computer Vision – ECCV 2022 Workshops. ECCV 2022. Lecture Notes in Computer Science, vol 13803. Springer, Cham. https://doi.org/10.1007/978-3-031-25066-8_9.
18. Ballestar, L.M., Vilaplana, V. (2021). MRI Brain Tumor Segmentation and Uncertainty Estimation Using 3D-UNet Architectures. In: Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2020. Lecture Notes in Computer Science(), vol 12658. Springer, Cham. https://doi.org/10.1007/978-3-030-72084-1_34.
19. R. El Jurdi, C. Petitjean, P. Honeine and F. Abdallah, "BB-UNet: U-Net With Bounding Box Prior," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1189-1198, Oct. 2020, doi: 10.1109/JSTSP.2020.3001502.
20. Xu, Q., Ma, Z., He, N., & Duan, W. (2023). DCSAU-Net: A deeper and more compact split-attention U-Net for medical image segmentation. *Computers in biology and medicine*, 154, 106626. <https://doi.org/10.1016/j.compbimed.2023.106626>.
21. Vijay, S., Guhan, T., Srinivasan, K., Vincent, P. M. D. R., & Chang, C. Y. (2023). MRI brain tumor segmentation using residual Spatial Pyramid Pooling-powered 3D U-Net. *Frontiers in public health*, 11, 1091850. <https://doi.org/10.3389/fpubh.2023.1091850>
22. Jia, Q., Shu, H. (2022). BiTr-Unet: A CNN-Transformer Combined Network for MRI Brain Tumor Segmentation. In: Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Lecture Notes in Computer Science, vol 12963. Springer, Cham. https://doi.org/10.1007/978-3-031-09002-8_1.
23. D. S. Vinod, S. P. S. Prakash, H. AlSalman, A. Y. Muaad and M. B. B. Heyat, "Ensemble Technique for Brain Tumour Patient Survival Prediction," in *IEEE Access*, doi: 10.1109/ACCESS.2024.3360086
24. M. F. Almufareh, M. Imran, A. Khan, M. Humayun and M. Asim, "Automated Brain Tumor Segmentation and Classification in MRI Using YOLO-Based Deep Learning," in *IEEE Access*, vol. 12, pp. 16189-16207, 2024, doi: 10.1109/ACCESS.2024.3359418.
25. X. Li, Z. Fang, R. Zhao and H. Mo, "Brain Tumor MRI Segmentation Method Based on Improved Res-UNet," in *IEEE Journal of Radio Frequency Identification*, doi: 10.1109/JRFID.2023.3349193.
26. J. Shi, L. Yu, Q. Cheng, X. Yang, K. -T. Cheng and Z. Yan, "MFTrans: Modality-Masked Fusion Transformer for Incomplete Multi-Modality Brain Tumor Segmentation," in *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 1, pp. 379-390, Jan. 2024, doi: 10.1109/JBHI.2023.3326151.
27. H. Ting and M. Liu, "Multimodal Transformer of Incomplete MRI Data for Brain Tumor Segmentation," in *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 1, pp. 89-99, Jan. 2024, doi: 10.1109/JBHI.2023.3286689.
28. <https://www.med.upenn.edu/cbica/brats2019/data.html>
29. <https://www.med.upenn.edu/cbica/brats2020/data.html>
30. 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.
31. Jiang Y, Zhang Y, Lin X, Dong J, Cheng T, Liang J. SwinBTS: A Method for 3D Multimodal Brain Tumor Segmentation Using Swin Transformer. *Brain Sci.* 2022 Jun 17;12(6):797. doi: 10.3390/brainsci12060797. PMID: 35741682; PMCID: PMC9221215.
32. F. Isensee et al., “nnu-net: a self-configuring method for deep learning based biomedical image segmentation,” *Nature Methods*, vol. 18, pp. 203–211, 2020, DOI: 10.1038/s41592-020-01008-z.
33. Yuan, Y. (2021). Automatic Head and Neck Tumor Segmentation in PET/CT with Scale Attention Network. In: Andrearczyk, V., Oreiller, V., Depeursinge, A. (eds) *Head and Neck Tumor Segmentation. HECKTOR 2020. Lecture Notes in Computer Science()*, vol 12603. Springer, Cham. https://doi.org/10.1007/978-3-030-67194-5_5
34. H. Jia, W. T. Cai, H. Huang, and Y. Xia, “H2nf-net for brain tumor segmentation using multimodal mr imaging: 2nd place solution to brats challenge 2020 segmentation task,” *ArXiv*, vol. abs/2011.03188, 2020, doi.org/10.48550/arXiv.2012.15318.
35. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., Li, J. (2021). TransBTS: Multimodal Brain Tumor Segmentation Using Transformer. In: de Bruijne, M., et al. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. MICCAI 2021. Lecture Notes in Computer Science()*, vol 12901. Springer, Cham. https://doi.org/10.1007/978-3-030-87193-2_11.
36. Jia, Q., Shu, H. (2022). BiTr-Unet: A CNN-Transformer Combined Network for MRI Brain Tumor Segmentation. In: Crimi, A., Bakas, S. (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2021. Lecture Notes in Computer Science*, vol 12963. Springer, Cham. https://doi.org/10.1007/978-3-031-09002-8_1.
37. Q. -D. Pham et al., "Segtransvae: Hybrid Cnn - Transformer with Regularization for Medical Image Segmentation," 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI), Kolkata, India, 2022, pp. 1-5, doi: 10.1109/ISBI52829.2022.9761417.
38. Zhang, Y. et al. (2022). mmFormer: Multimodal Medical Transformer for Incomplete Multimodal Learning of Brain Tumor Segmentation. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. MICCAI 2022. Lecture Notes in Computer Science*, vol 13435. Springer, Cham. https://doi.org/10.1007/978-3-031-16443-9_11.
39. A. Hatamizadeh et al., "UNETR: Transformers for 3D Medical Image Segmentation," 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2022, pp. 1748-1758, doi: 10.1109/WACV51458.2022.00181.
40. O. Oktay et al., “Attention u-net: Learning where to look for the pancreas,” *ArXiv*, vol. abs/1804.03999, 2018, doi.org/10.48550/arXiv.1804.03999.
41. B. Yu et al., "SA-LuT-Nets: Learning Sample-Adaptive Intensity Lookup Tables for Brain Tumor Segmentation," in *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1417-1427, May 2021, doi: 10.1109/TMI.2021.3056678.
42. Wang, F., Jiang, R., Zheng, L., Meng, C., Biswal, B. (2020). 3D U-Net Based Brain Tumor Segmentation and Survival Days Prediction. In: Crimi, A., Bakas, S. (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2019. Lecture Notes in Computer Science()*, vol 11992. Springer, Cham. https://doi.org/10.1007/978-3-030-46640-4_13
43. Li, X., Luo, G., Wang, K. (2020). Multi-step Cascaded Networks for Brain Tumor Segmentation. In: Crimi, A., Bakas, S. (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2019. Lecture Notes in Computer Science()*, vol 11992. Springer, Cham. https://doi.org/10.1007/978-3-030-46640-4_16.
44. W. Li, W. Huang and Y. Zheng, "CorrDiff: Corrective Diffusion Model for Accurate MRI Brain Tumor Segmentation," in *IEEE Journal of Biomedical and Health Informatics*, doi: 10.1109/JBHI.2024.3353272