# Multimodal attention-gated cascaded U-Net model for automatic brain tumor detection and segmentation

Siva Koteswara Rao Chinnam , Venkatramaphanikumar Sistla , Venkata Krishna Kishore Kolli [*]

*Department of Computer Science & Engineering, VFSTR Deemed to be University, Vadlamudi, Guntur, Andhra Pradesh, India*

## ARTICLE INFO

## ABSTRACT

During the last decade, several studies have been conducted to improve efficiency and robustness in the detection and segmentation of brain tumors based on different parameters like size, shape, location, and contrasts. This study proposes Multimodal Attention-gated Cascaded U-Net (MAC U-Net) model to address the performance issues observed in the detection and segmentation of low-grade tumors. The effectiveness of group normalization with attention gate is also explored with skip connections to segment small-scale brain tumors using several highlighted salient features. The model is evaluated on the brain tumor benchmark dataset BraTS2018 over various performance metrics such as Dice, IoU, Sensitivity, Specificity, and Accuracy. Experimental results illustrate that the proposed MAC U-net on BraTS 2018 dataset outperforms baseline U-nets with 94.47, 84.12, and 82.72 dice similarity coefficient values on HGG and 85.71, 78.85 and 74.16 on LGG subjects with Ground Truth values of Complete Tumor, Tumor Core, and Enhancing tumor, respectively. The proposed model is also evaluated on BraTS 2019 and BraTS 2020 datasets. Moreover, MAC U-net achieves superior performance over typical conventional brain tumor segmentation methods especially in terms of low-grade gliomas.

## 1. Introduction

A brain tumor is a collection or mass of abnormal cells in the brain (central nervous system tumor), which, in turn, disrupts the functions of the brain. Brain tumors can be mainly classified into malignant (cancerous) and benign (non-cancerous) tumors. Brain tumors occur mainly due to excessive exposure to radiation or cancer treatments or due to hereditary transmission. As per WHO (World Health Organization) [1–3], 30% of people diagnosed with a brain tumor do not recover from the disease. Gliomas, the most common type of brain tumor, develop from glial cells and meningiomas, which originate in the meninges. WHO classified gliomas as high-grade glioma (HGG) and low-grade glioma (LGG) based on their characteristics. They are further divided into Grade I (non-infiltrative, slow in growth, curable via surgery and long-term survival of victim), Grade II (infiltrative and relatively slow in growth and may turn into high grade), Grade III (malignant, infiltrative, moderate in growth and high turn into a high grade), and Grade IV (most malignant, widely infiltrative, and rapid in growth). These tumors are treated by medical practitioners using surgery, radiation therapy, or chemotherapy alone or in combination.

Clinical expertise is extremely essential in the detection (estimating the location and type) and segmentation of brain tumors using magnetic resonance imaging (MRI). The task is cumbersome because of lesion localization, and slice-wise decision-making. The effectiveness of the manual approach in identification and segmentation largely depends on the expertise of the clinician. Automated brain tumor segmentation is an alternative approach that can lead to reliable decision-making with high accuracy. However, developing an automated system that detects brain tumors requires expertise in both computer engineering as well as medical science. Creating an automated system to detect brain tumors is a challenging task because tumor intensity varies in surrounding healthy cells. Furthermore, these tissues often show variations in size, shape, and localization. Recently, several studies were conducted on deep learning models to improve tumor segmentation, thereby increasing diagnostic and treatment accuracy and precision. MRI imaging typically allows different types of sequences to best visualize data aspects of the brain by adjusting repetition time (TR) and echo time (TE) parameters. In T1 sequences, good contrasts observed between tissues, whereas the brightness of lipids is relatively higher than that of cerebrospinal fluid (CSF); dense bones also appear. Contrast-enhanced T1-weighted (TICE) sequences are used to detect tumor borders, facilitating an easier distinction between necrotic and active tumors. T2 sequences determine

whether the edema is associated with a tumor. The fast Fluid-Attenuated Inversion Recovery (FLAIR) sequence is effective at distinguishing edema from CSF. Generally, tumors like enhancing tumors and tumor core are very small in size. The performance of the existing automated segmentation models is limited in segmenting tumor core or enhancing tumor. Hence, deep learning models need to be employed with more "attention" to ensure effective documentation in all such cases.

In general, U-Net is limited in accuracy in segmenting small scale tumors due to reduction of dimensions during down-sampling. Brain tumors are generally in complex shapes with diverse sizes. The task of tumor segmentation is to enhance accuracy in segmentation using attention mechanism, which enhance the local feature expression. In this work, authors aimed to explore the effectiveness of the attention gate in segmentation of small-scale tumors.

Majority models in the literature are suffering with network convolution across different modalities. The computation of feature difference between glioma and normal tissue is evolved as a major challenge for categorization of different grades of tumor. In this connection, discrimination of tissue contours needs multi-scale semantic information to reduce the information loss during convolution. Single modal U-Net could not be able to segment the grade and contour of the brain accurately.

Basically, model accuracy is highly correlated with batch size in batch normalization, but due to memory limitations in model building, authors intended to take the advantage of group normalization. Group normalization addresses channel dependencies and memory intensive issues.

Majority of the literature models are more successful in handling HGG when compared to LGG. Segmentation of LGG is complex due to its simple internal structure with lower contrast and compactness. Furthermore, LGG also does not exhibit any necrosis, perifocal edema and hemorrhagic foci. To overcome the above limitations, in this work gradient levels are well controlled and convergence process was speed up with reduced effect of internal covariate shift through the usage of Group Normalization and Self Attention module. In general, LGG tumors have uniform intensity distribution across boundaries, and their lesion area is more distinguishable from T2 and flair modalities, are intended for proposed model building.

To address all the above limitations, authors propose the Multimodal Attention-gated Cascaded U-Net Model for Automatic Brain Tumor Detection and Segmentation with following Objectives:

- Design and development of 9-layer attention-gated U-Net model with Group Normalization (GN) that for detection of full tumor from Flair and T2 MRI modalities in the first phase.
- The combination of Average and Maxpooling aimed to minimize the feature loss during downsmapling.
- Design and development of 7-layer attention gated U-Net model with GN to segment small-size tumors from T1 CE in the second phase.
- Decisions level fusion is performed over predicted full tumor in phase I, Enhanced Tumor and Tumor Core in phase -II to detect and segment low-grade tumors. The outcomes are compared with the Ground Truth and clinician (domain expert) opinion.

The rest of the paper is organized as follows: Section 2 presents a thorough review of the literature published on unimodal and multimodal brain tumor segmentation. Section 3 presents the proposed MAC U-Net architecture for multimodal brain segmentation. Section 4 presents experimental results and discussions. Finally, the conclusion and future scope are presented in Section 5.

## 2. Related work

Brain tumor segmentation involves the separation of tumor tissues such as an active tumor, edema, and necrosis from normal tissues. Brain tumor segmentation approaches are primarily categorized into three approaches such as manual, semi-automatic, and fully automatic. These algorithms include Atlas, image analysis, machine learning, and hybrid algorithms [4–7].This segmentation process focuses chiefly on identifying the whole tumor, tumor core, and enhanced tumor. Manual segmentation approaches are tedious and have low accuracy and precision [8].There is a need fora robust and efficient automatic technique in segmentation. Convolution Neural Network (CNN), a class of deep learning neural networks, surpassed the results of traditional machine learning models like support vector machines (SVMs), decision trees, and random forests [9]. An automatic CNN called Fully Convolutional Networks (FCNs) [10],has been shown to minimize computational cost in generating tumor label maps. Inspired by FCNs, an efficient FCN named U-Net [11] was designed with a contracting path to capture context and a symmetric expanding path that enables precise localization to do efficient segmentation of tumor core.

Deep learning has a critical role in biomedical imaging and facilitates radiologists in better characterization and supports decision making. Also, this improves the diagnosis and treatment of patients. The deep learning-based U-Net [12] has remained seminal in the great transformation of biomedical image segmentation efforts ever since. Radiologists segment the tumor manually using anatomical and physiological information [13,14]. In semi-automatic methods [15–17], radiologists define the region of interest, adjust parameters to suit input images, use automatic algorithms to process, adjust the response depending on the feedback, evaluate the results, and repeat the process until the best results are achieved. In contrast, fully automatic methods are fast and perform better without the need for any manual intervention [18–21]. An automatic 3D brain tumor segmentation [22,23] is proposed using a sequential U-Net architecture. The Brain Tumor Segmentation model [24] performs with better accuracy when compared to traditional brain tumor segmentation algorithms. The Nearest-Neighbor resampling-based Elastic-Transformed (NNRET) CNN framework [25] was put forward to perform better segmentation in terms of accuracy over the classic U-Net Model. The Performance of classical U-Net is poor in the processing of complex images; therefore, the Half-Dense U-Net model [26] was proposed to locate the boundary of a brain tumor; this model outperformed when compared to Dense U-Net Model and ResNet.

Recent trends show that researchers are mainly employing stacking networks to have higher accuracy in tumor detection despite having higher computational complexity [27–30].When the depth of the network increases, the excessive loss of information over the layers affects the performance of the model. Another study [31] addressed the increase in the number of parameters and loss of information due to the increase in depth of the network, using the Stack Multi-Connection Simple Reducing Net (SMCSRNet) framework. This model overcomes the loss of information by including a series of bridge connections among the stacked cascade networks. SMCSRNet has outperformed stacked U-Net and achieved significant efficiency over other Deep Convolutional Neural Networks such as DenseNet or ResNet. Attention is required to notable values and for segmentation to get semantics, along with the focus on boundaries that require details. In order to resolve this predicament, one study [32] adopted 2D U-Net with denser skip connections based on stride dilated convolution (SDC) to reduce resolution that preserves more details over samples of the BraTS2017 dataset. Authors [33] performed well over classical U-Net and included the extraction of global and local feature extraction paths, thereby improving the accuracy of segmentation on five tumor regions over a large BraTS2018 dataset [34] verified the performance of Group Normalization (GN), Instance Normalization (IN)in U-Net and observed that the accuracy of the model with GN is higher and yields better generalization. GN does not depend on batch size, it divides the sequences into various subgroups, and calculates the mean and variance. In the U-net model [35,36] employed dense blocks in place of skip connections at the encoder part and applied deep supervision at the decoder part, and achieved higher segmentation accuracy on the BraTS2018 data set [37] extracted dense feature information using

**Table 1**
Systematic Survey on Models used in Brain Tumor Segmentation.

| Ref | Modal Type | Methods Used: | Data Set | Performance Metrics |
|---|---|---|---|---|
| [18] | Multi-Channel T1, T1CE, T2, FLAIR | First order Features Extracted using Texture Analysis. Classification Using SVM CRF based Regularization | Contra Cancrum | Dice– 0.84 |
| [19] | Multi-Channel T1, T1CE, T2, FLAIR | After the multi-sequence image has been pre-processed, voxel-wise features are extracted, followed by classification and sub sequent spatial regularization. | Brats2013 | Dice– 0.83 Jaccard −0.72 PPV – 0.86 Sensitivity – 0.897 |
| [20] | Single Channel - FLAIR | Deep medic - Automatic detection and segmentation of brain metastases on multimodal MR images with a DCNN | Customized data set | Dice– 0.79 Sensitivity –0.98 False Positivity −4.4 |
| [21] | Single Channel - Independently | Automatic brain tumor detection and segmentation using U-net based fully convolutional networks | Brats2015 | Dice– 0.86 Sensitivity – 0.65 |
| [23] | Multi-Channel T1, T1CE, T2, FLAIR | Sequential 3D U-net with BN | Brats2017 | Dice– 0.883 |
| [24] | Multi-Channel T1, T1CE, T2, FLAIR | Feature Recombination BN | Brats2015 | Dice– 0.90 PPV –0.88 Sensitivity −0.93 |
| [25] | Single Channel - Independently | Nearest-neighbor Re-sampling-based Elastic-Transformation BN | Brats2017 | Dice Similarity Coefficient – 0.8976 Jaccard −0.8869 |
| [31] | Multi-Channel T1, T1CE, T2, FLAIR | A Stacked Multi-Connection Simple Reducing Net BN | Brats2017 | Dice Similarity Coefficient – 0.83 PPV – 0.789 SCS – 0.90 |
| [32] | Single Channel - Independently | Enhanced U-Net | Brats2017 | Dice Similarity Coefficient – 0.89 Sensitivity – 0.8592 Specificity – 0.9981 |
| [33] | Multi-Channel T1, T1CE, T2, FLAIR | Modified U-Net that incorporates both global and local feature extraction paths | Brats2018 | Dice Similarity Coefficient – 0.94 Jaccard – 0.69 |
| [35] | Multi-Channel T1, T1CE, T2, FLAIR | Multimodal Brain Tumour Segmentation using Densely Connected 3D CNN | Brats2018 | Dice Similarity Coefficient – 0.87 |
| [37] | Multi-Channel T1, T1CE, T2, FLAIR | Attention Gate ResU-Net BN | Brats2017 Brats2018 Brats2019 | Dice– 0.872 Haussdroff95 – 6.87 Dice– 0.872 Haussdroff95 – 5.62 Dice– 0.87 |

downsampling and propagated the spatial and location information of low-level features through up-sampling. Furthermore, integration of attention gates, AGResUNet, with residual modules in U- Net outperformed both U-Net and ResNet. Inception residual modules were integrated with Dense U-net (DENSE-Inception U Net) [38] and evaluated over different data sets. The model outperformed as compared with the state-of-the-art models. List of models used for Brain tumor segmentation is tabulated in Table 1.

As, spatial contextual information of 3D MR images is not fully considered by CNN, K. Hu et al., [39] proposed an efficient framework consists of fine segmentation that uses the MCCNN with fully connected CRFs for smoothing tumor edges and eliminating false positives. An efficient 3D residual neural network (ERV-Net) proposed in [40], to avoid degradation with fusion loss function to address the issues of network convergence, and data imbalance, has shown high efficiency compared to state of art methods.

## 3. Proposed work

From the literature, it is clear evident that the majority segmentation approaches were successful in handling HGG when compared to LGG. The internal structure of the LGG is simpler than that of the HGG, the segmentation of the LGG is considered more difficult because of its lower contrast and smaller size [14]. In general, LGG tumors have uniform intensity distribution across boundaries, and their lesion area is more distinguishable from T2 and flair modalities. Keeping in the view of above advantages, T2 and flair modal MRI images are intended for proposed model building.

Fig. 1 shows the proposed MAC U-Net architecture in which the encoder extracts texture features and background information to detect tumor location. A structurally symmetric decoder is used to reconstruct or segment the tumor.

A nonhomogeneous magnetic field or sensitive movement of the patient during acquisition results in non-uniform intensities, contrast differences, and noise in MRI scans. BraTS2018 images were acquired using multiple scanners from different sites/subjects with different clinical protocols, and they suffered from non-standardized intensity distributions. To remove these limitations, the N4ITK bias field correction algorithm was applied as a first step of preprocessing. To eliminate scanner variation, we adopted the Fuzzy C-means intensity normalization technique. The objective function [41–45] of Fuzzy C-means is given by,

$$J(U, V) = \sum_{k=1}^{c} \sum_{i=1}^{n} u_{ik}{}^{m} \|X_i - V_k\|^2 \tag{1}$$

$$\sum_{k=1}^{c} \sum_{i=1}^{n} u_{ik} = 1 \tag{2}$$

where $X$ denotes the set of data points, $V$ denotes the set of centers, $c$ denotes cluster count, $N$ denotes the number of data points, $V_k$ is a fuzzy cluster centroid of the $k^{th}$ cluster and $m$ denote the weighting exponent. Accuracy can be improved by removing noise, adding spatial information, and modifying the objective function of standard fuzzy c-means.

### 3.1. Detection of full tumor using 9 layer attention gated U-Net

Generally, radiologists use the FLAIR modality for segmenting edema, discriminate edema against ventricles/ fluid-filled structures, and T2 for edema extension. Furthermore, we applied FLAIR and T2 modality samples as input to proposed architecture given in Fig. 2.

The encoder block consists of two 3*3 convolutions subject to group normalization (GN) [46], followed by a ReLu and one 2*2 max-pooling layer with stride 2. For tumor recognition, an optimized feature map was formed by adjusting hyper-parameters [47], using a convolution layer with an input $Row(R_1) \times Column(C_1) \times Dimension(D_1)$. The hyper-
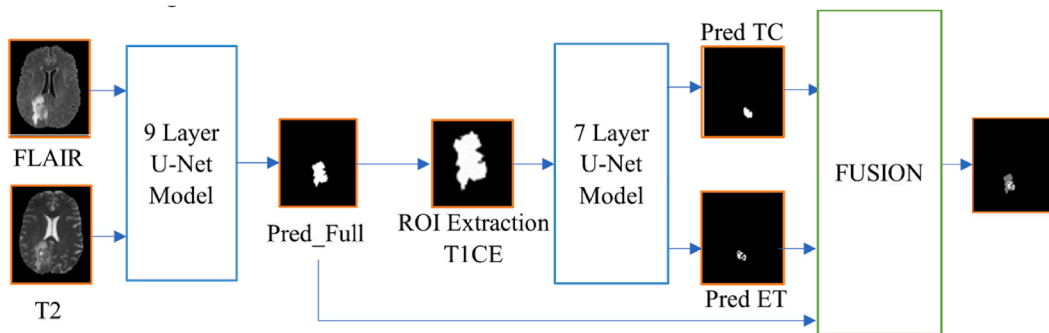
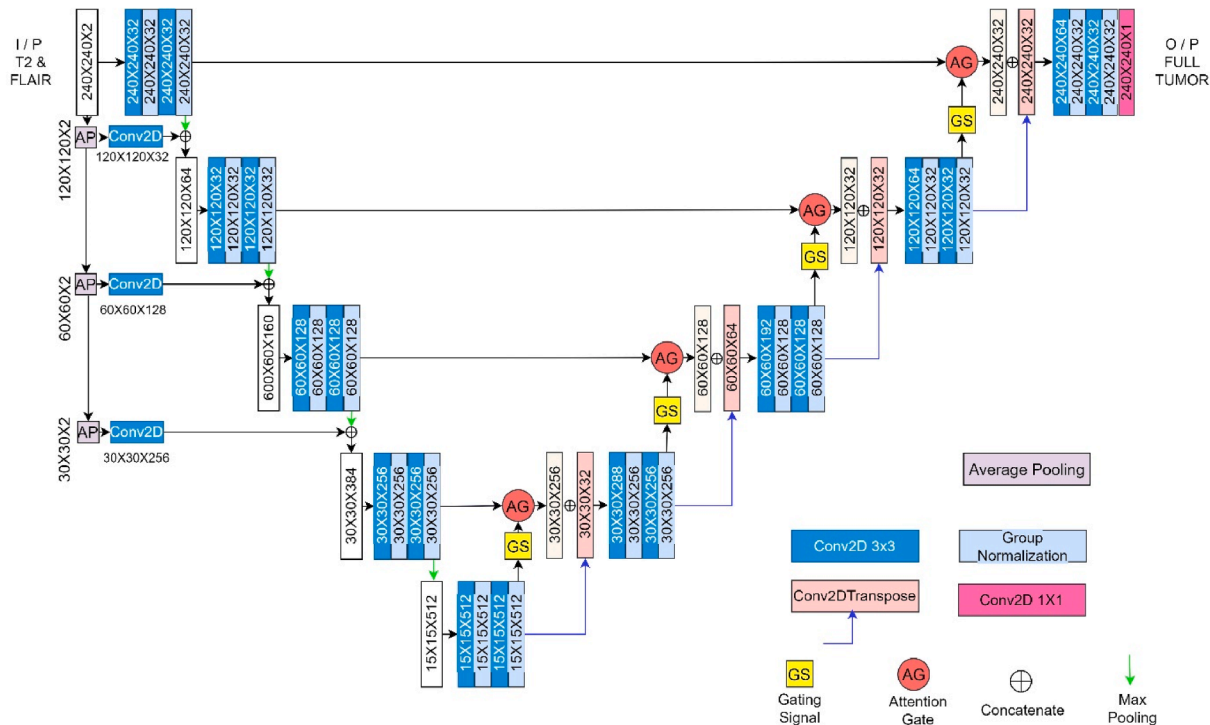**Fig. 1.** Proposed MAC U-Net architecture for brain tumor segmentation.



**Fig. 2.** 9-Layer Attention gated U-Net architecture for full tumor segmentation.
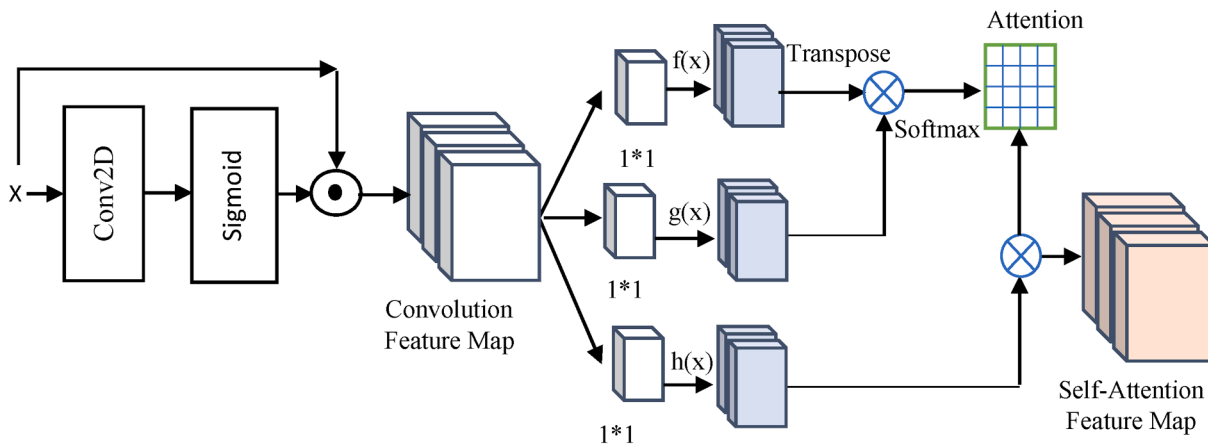


**Fig. 3.** Architecture of the self-attention module.

parameter of feature maps, like several filters (K), size of the filter (F), stride (S) with zero paddings (P) are calculated using the following equations [48]:

$$Width = \frac{R_1 - F + 2P}{S} + 1 \tag{3}$$

$$Height = \frac{C_1 - F + 2P}{S} + 1 \tag{4}$$

$$Dim = F \tag{5}$$

The filters are convolved with an input image that results in feature maps:

$$Out = \sum_{k=0}^{n} (FilterWeight*Input) + bias \tag{6}$$

In general, average pooling smoothens the image, but failed to extract the sharp features. Max pooling well suits for the extraction of the bright features from the dark background. But both these pooling approaches could not be able to hold down sampling's unique characteristics. Segmentation of small-scale tumor need the minimum feature loss during the down-sampling. In this work, authors supplemented the model with reduced feature loss by concatenating the feature maps of both "maximum" and "average" pooling. The weights in the pooling kernel are adaptively computed based on the input MR images or feature maps to extract more useful features during down-sampling and enhance the segmentation (See Fig. 3).

### 3.1.1. Exploiting channel dependencies using group normalization

Keeping in the view of memory constraints and to realize optimized performance, the proposed model was built with minimal batch size. But, batch normalization with lower batch size tends to incorrect estimation of batch statistics [10,49] and presents detrimental impact on performance of DNN [46,50,51]. To realize the limitations such as lower training error, easy optimization, authors adopted Group Normalization [50,52,53] to normalize the features within each group and which is independent of batch sizes and presented the stable performance. To apprehend the regularization ability of the group normalization, L2 regularizer is used in the model building. Pre-processing plays a vital role in deep learning, and which address the effect of imbalanced gradients leads to slow learning. In the model building, the weights are adjusted using stochastic gradient descent to avoid imbalance. The output of the activation function is normalized after each layer. The activation function's normalized output is sent to the next layer. The feature map of a layer $x$ and index $i = (i_N, i_C, i_H, i_W)$, where N is the batch axis, C is the channel axis, H is spatial height, and W is the spatial width.

$$\widehat{x_i} = \frac{x_i - \mu_i}{\sigma_i} \tag{7}$$

$$S_i = \{k | k_C = i_C\} \tag{8}$$

$$\mu_i = \frac{1}{m} \sum_{k \in S_i} x_k \tag{9}$$

$$\sigma_i = \sqrt{\frac{1}{m} \sum_{k \in S_i} (x_k - \mu_i)^2 + \epsilon} \tag{10}$$

$$y_i = \gamma \widehat{x_i} + \beta \tag{11}$$

In GN, with the help of $G$ (hyper-parameter), the number of layers (set to 32 in this work), $C$ denotes the number of channels per group, and computes $\mu$ and $\sigma$ in a set $S_i$ defined by Eqs. (7)–(10):

$$S_i = \{k | (k_N = i_N, \lfloor k_C*G/C \rfloor = \lfloor i_C*G/C \rfloor)\} \tag{12}$$

For faster convergence, Rectified Linear Unit (ReLU) is employed as an element-wise activation function:

$$R(z) = Max(0, z), \text{for} z > 0 \tag{13}$$

To reduce complexity and to control overfitting, we used a pooling layer with stride 2, in order to minimize the number of training parameters [53]. The input volume having a width ($W_1$), height ($H_1$), and dimension ($D_1$) produce a volume of size with width ($W_2$), height ($H_2$), and dimension ($D_2$):

$$W_2 = \frac{W_1 - F}{S} + 1 \tag{14}$$

$$H_2 = \frac{H_1 - F}{S} + 1 \tag{15}$$

$$D_2 = D_1 \tag{16}$$

Softmax was used as an activation function for classification:

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}} \tag{17}$$

### 3.1.2. Attention based salient feature extraction in tumor segmentation:

The attention gate (AG) [55–58] was implemented in the decoder block for extracting salient features, to minimize usage of computational resources, to reduce the impact of noise, ambiguous, and irrelevant features. In this model, skip connections of the up-sampling path are combined with downsampling path spatial information using soft attention that reduces redundant features [58]. The major component of a transformer is the unit of the multi-head self-attention mechanism. The encoded input is a set of key-value pairs ($K$, V), and preceded output is a query ($Q$) in the view of a transformer. The next output is produced by the mapping query and the set of key-value pairs.

The scaled dot product is adopted by the transformer during the down-sampling of the neural network. After $'l'$ layers, the convolution feature map is $x^l$. The feature map is then branched into $K$, $Q$, and $V$, were.

$$Key : f(x) = W_f x \tag{18}$$

$$Query : g(x) = W_g x \tag{19}$$

$$Value : h(x) = W_h x \tag{20}$$

Then, the dot-product attention leads to the self-attention feature maps:

$$\propto_{i,j} = softmax\left(f(x_i)^T g(x_j)\right) \tag{21}$$

$$O_j = \sum_{i=1}^{N} \propto_{i,j} h(x_i) \tag{22}$$

$\propto_{i,j}$: entry in attention map, denotes the attention of $i$ th position synthesized at *the* $j$ th location. $W_f, W_g$, and $W_h$ are 1*1 convolution filters. The output $O_j$ is a column vector of final output and is multiplied with parameter $\gamma$ and added to the original input feature map $x_i$.

$$y = x_i + \gamma O_i \tag{23}$$

For the better segmentation, residual block features of small tumors are combined with feature maps of the succeeding layer. The doubled spatial dimension with bilinear interpolation is concatenated with an output feature map of the attention gate and a 3*3 convolution. At each level in the decoder, the features are minimized by a factor of two and double the dimension of images. At the end of the decoder, output has the same spatial image dimension as the input image and followed by 1*1 convolution producing a binary prediction of the full tumor segmented image.

The encoder block is applied on 240*240 images and downgraded to 15*15; the feature maps are increased from 64 to 1024. As the tumor core is inside of edema and the enhancing tumor is part of the tumor
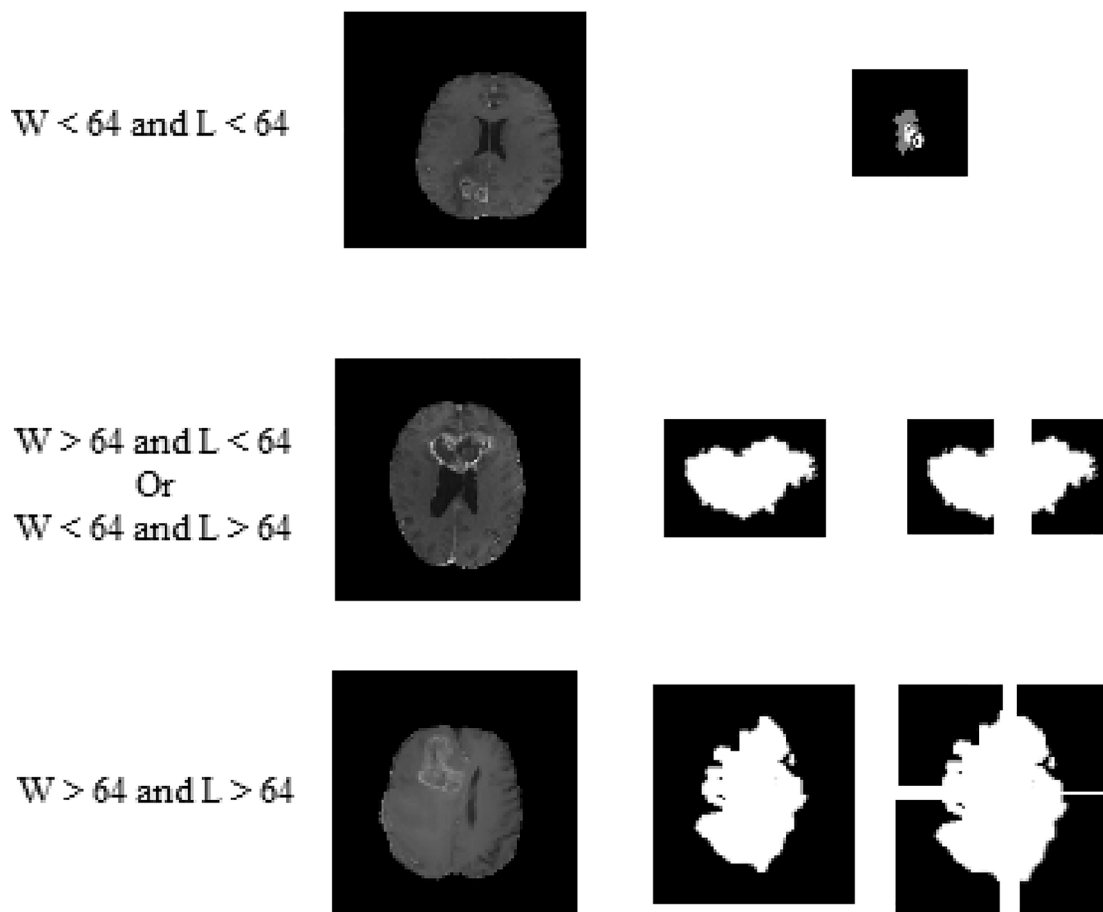
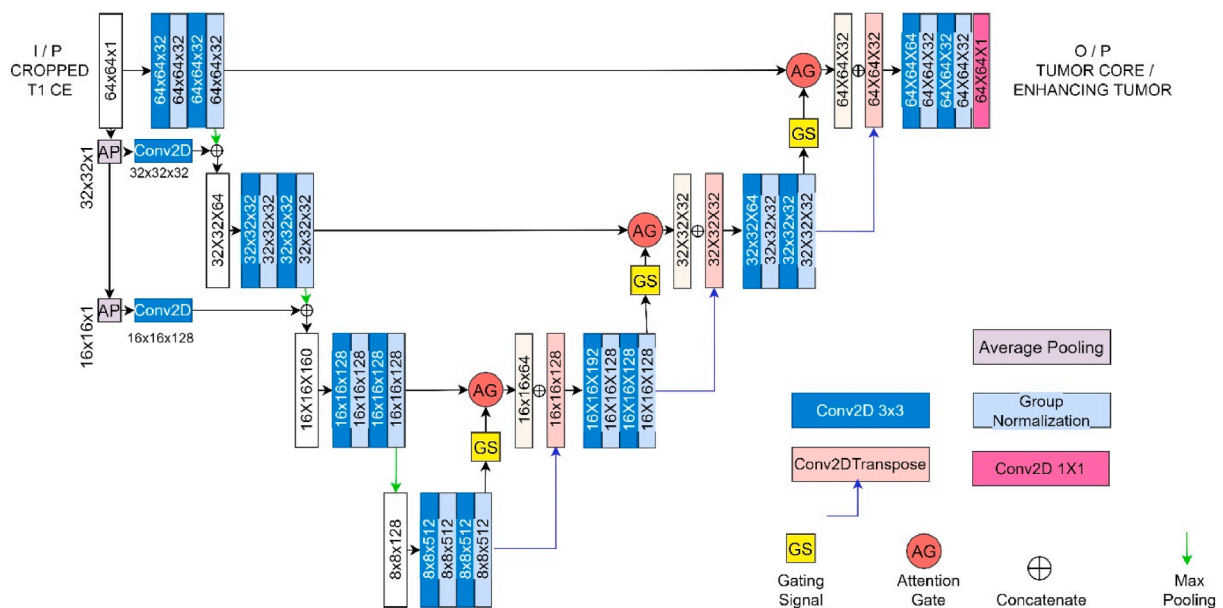**Fig. 4.** Size-based full tumor segmentation scenarios on T1CE slices.



**Fig. 5.** 7-Layer Attention gated U-Net Architecture for segmentation of Tumor Core and Enhanced Tumor.
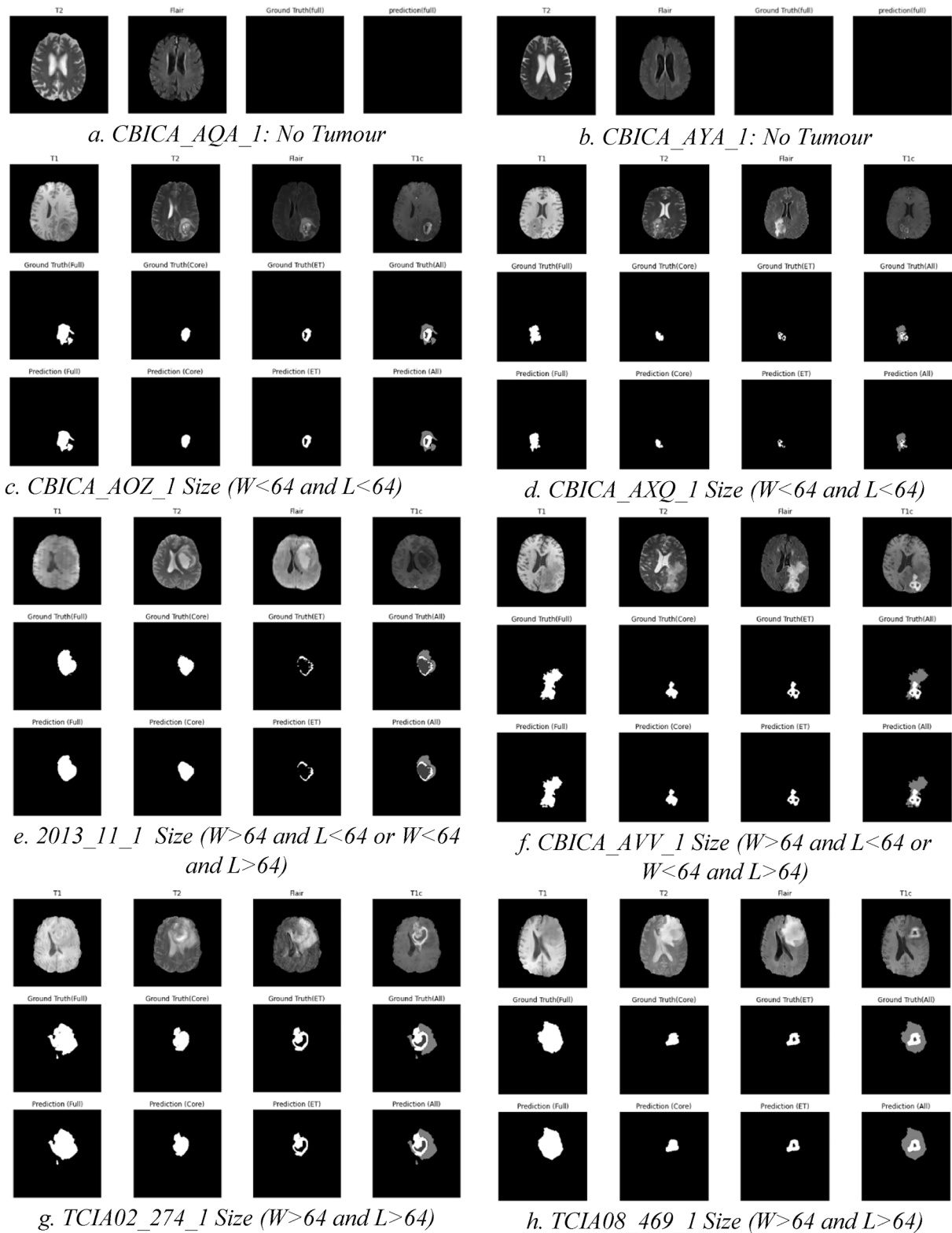
Fig. 6. Segmentation results of proposed method on BraTS18 MRI scans for full, core, & enhanced. a & b: No Tumor slices; c & d Segmentation of small size tumor with w < 64 & l < 64. e, f, g & h: Segmentation of varied length & width tumor slices.

core, it is very small in terms of the number of pixels. Further, center point of full tumor will help in prediction of tumor core and enhanced tumor. The cropping of T1CE will be done based on the size of full tumor and even crop the overlap part to do data-augmentation by fixing the cropping size to 64x64. T1CE will be cropped into one or two or four 64*64 patches using the center of the tumor as shown in Fig. 4.

### 3.2. 7-Layer attention gated U-Net for segmentation of tumor core and enhanced tumor

Fig. 5 shows a 7-Layer Attention gated U-Net Architecture for segmentation of Tumor Core and Enhanced Tumor.

If the size of a tumor is of width or W < 64 and length or L < 64

**Table 2**
List of Hyper parameters.

| Models | 9 Layer U-Net, 7 Layer U-Net |
|---|---|
| Data Set | BraTS 2018, BraTS 2019, BraTS 2020 |
| image size | 240 * 240 |
| Normalization Techniques | N4 Bias Field Correction, Fuzzy C-Means, Batch, and Group Normalization |
| Batch Size | 8 for Batch Normalization, 1 for Group Normalization |
| No of epochs | 30 |
| Optimizers | Adam, SGD and RMSProp |
| Regularizer | L2 |
| Learning Rate | 0.0001 |
| loss function | Dice Coefficient |
| Metrics | Dice, IoU, Sensitivity, Specificity, and Accuracy |

(Fig. 6c & d), using the centroid of the full tumor, T1CE is cropped into one patch. If $W > 64$ and $L < 64$ or $W < 64$ and $L > 64$ then T1CE will be cropped into two 64X64 patches (Fig. 6e & f), else it will be cropped into four patches of each size 64X64 (Fig. 6g & h). These cropped sections are used as input to the 7-layer AG U Net architecture to predict the tumor type. Furthermore, predicted full tumor, predicted tumor core and predicted enhanced tumor core are fused at the center point of the full tumor to estimate the final segmented tumor, as depicted in Fig. 6.

## 4. Results and discussions

The proposed multi-class segmentation approach was evaluated on the BraTS2018 dataset using Dice coefficient and IoU metrics to determine non-enhancing and enhancing tumor regions. The experiment was conducted using the Keras framework using TensorFlow-GPU on a 64 GB Machine: Intel® Xenon® Platinum 8276 CPU with 2.20 GHz processor equipped with NVIDIA GRID V100D-4Q Graphics card. The performance of the proposed model was compared with four different models to ascertain its accuracy.

### 4.1. About dataset

The BraTS2018 dataset [30,44,45] provided by the Medical Image Computing and Computer-Assisted Intervention Society (MICCAI) is a collection of skull-stripped co-registered MRI volumes with an isotropic resolution. BraTS2018 dataset contains skull stripped 210 HGG, 75 LGG MR Scanned subjects for training, and 66 MR scanned subjects for performance evaluation. BraTS2019 dataset comprises of 259 HGG, 76 LGG MR scanned subjects for training, and 125 subjects for testing. BraTS2020 dataset [30,44,45] comprises of 369 subjects for Training and 125 subjects for Testing. Subjects are manually labeled ground truths (L0: Healthy Tissues, L1: Necrosis and Non-Enhancing Tumor, L2: Edema, L4: Enhancing Tumor) of the training dataset. Each MR Scan has four modalities: T1, T1CE, T2, and FLAIR. All modalities are registered together with T1CE to homogenize the data and are resampled to 1 mm isotropic pixel resolution on a normalized axis using a linear interpolator [30–37]. The BraTS2018 dataset also suffers from a class imbalance of MRI volumes, since most of the pixels/ slices within the dataset pertain to a healthy area and very few slices belong to tumors. Hyper parameters of proposed model are presented in Table 2.

### 4.2. Performance evaluation metrics

The Dice Similarity Coefficient (DSC) and Jaccard Similarity Index (JSI / IoU) were employed to validate the different labels of tumor region, where TP corresponds to pixels corresponding to the Total Tumor Region, TN: All healthy tissue pixels, FN: Abnormal tissue pixels i.e., pixels that are not classified by defined model and FP: pixels that are incorrectly classified as a tumor. JSI (i.e., Intersection over Union) measures similarity. JSI / IoU always lies between 1 and 0. The larger value represents more accuracy of segmentation. Modified Hausdorff

**Table 3**
Performance evaluation of proposed model with various attention mechanisms on Brats2018.

| attention type | DSC | | |
|---|---|---|---|
| | Whole | Core | ET |
| Additive Attention [57] | 87.2 | 76.2 | 76.6 |
| Multiplicative Attention [58] | 87.23 | 78.3 | 77.2 |
| Multi Headed Attention [59] | 92.6 | 81.7 | 78.2 |
| Channel Attention [63] | 88.9 | 79.9 | 78 |
| Spatial Attention [64,65] | 89.6 | 80.3 | 79 |
| Self-Attention [66] | **94.47** | **84.12** | **82.72** |

Distance represents adjacency between pixels [44–62].

$$Dice\ Similarity\ Coefficient\ (DSC) = \frac{2TP}{FN + FP + 2TP} \qquad (24)$$

$$Jaccard\ Similarity\ Index\ (JSI) = \frac{TP}{TP + FN + FP} \qquad (25)$$

$$Sensitivity = \frac{TP}{TP + FN} \qquad (26)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (27)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (28)$$

where TP denotes True Positives, FP demotes False Positives, TN denotes True Negatives, and FN indicates False Negatives.

In this study, the authors experimented on BraTS 2018 training dataset with the random split of 80% of data for model building and the remaining 20% for model validation. As per the clinicians' suggestion, authors have considered the slices from 50 to 130 out of 155 for each subject. From BraTS 2018 dataset, 3360 HGG and 1200 LGG samples are used for the model building. Experimentation was performed in four sections with Multi channel cascaded U-Net with Batch Normalization, Multi-Channel cascaded U-Net with Group normalization, Multi-channel Attention-gated cascaded U-Net with Batch Normalization, and Multi-Channel Attention-gated cascaded U-Net with Group normalization. Multi-channel Attention-gated cascaded U-Net with group normalization showed better results when compared with other models for Full, Core, and Enhanced Tumor in terms of performance evaluation metrics such as accuracy, sensitivity, specificity, DSC and IoU. Authors explored the experimentation with various loss functions such as Adam, RMSProp and SGD, from the experimentation it is clearly notified that the performance of Adam is good when compared to other at 0.0001 learning rate. To realize the regularization ability model was built with L2 Regularizer.

### 4.3. The performance evaluation of various models:

Performance evaluation of proposed Multi channel cascaded U-Net with various attention mechanisms on Brats 2018 data set is presented in Table 3. The proposed model with Self-attention performed better in comparison over other attention mechanisms in segmenting Full tumour and Tumour core, and Enhanced tumour (See Table 4).

In the segmentation of High-grade tumors, basic U-Net (9 Layer model) yielded dice similarity coefficient as 86.34, 75.65 and 72.82 and IoU as 76.87, 66.46, 61.18 for Full tumor, Tumor Core and Enhancing Tumor respectively. Basic U-Net suffers more, in segmenting low grade gliomas due to its small in size.

To realize better performance, channel dependencies are preserved by replacing Batch Normalization with Group Normalization. Further U-Net 9-layer architecture is stacked with U-Net 7-layer architecture to emphasize more on segmenting varied tumor sizes as Multi-Channel

**Table 4**
Performance evaluation of proposed architectures with various Normalization Techniques.

| Data Set | Grade | Model | DICE | | | IOU | | | SENSITIVITY | | | SPECIFICITY | | | ACCURACY | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | CT | TC | ET | CT | TC | ET | CT | TC | ET | CT | TC | ET | CT | TC | ET |
| *2018* | HGG | *U Net* | *86.34* | *75.65* | *72.82* | *76.87* | *66.46* | *61.18* | *89.34* | *82.24* | *81.11* | *99.2* | *91.87* | *89.93* | *78.6* | *75* | *73.18* |
| | LGG | | *64.87* | *62.49* | *62.05* | *60.85* | *59.49* | *58.22* | *85.42* | *76.35* | *75.45* | *98.5* | *88.67* | *86.43* | *78.5* | *73.5* | *71.08* |
| | HGG | *U Net + GN* | *88.1* | *80.28* | *76.25* | *78.28* | *67.58* | *61.87* | *89.41* | *83.14* | *82.28* | *99.4* | *92.45* | *89.89* | *82.5* | *77.5* | *74.71* |
| | LGG | | *65.11* | *63.76* | *63.57* | *62.08* | *60.57* | *59.96* | *88.05* | *80.82* | *81.47* | *99.5* | *88.9* | *87.27* | *80.41* | *73.75* | *72.91* |
| | HGG | *MCC U Net + BN* | *88.4* | *81.39* | *76.57* | *79.21* | *68.62* | *62.04* | *89.72* | *83.22* | *83.12* | *99.64* | *92.76* | *90.14* | *90.24* | *80.15* | *76.25* |
| | LGG | | *67.43* | *64.26* | *64.32* | *62.87* | *61.54* | *61.03* | *88.69* | *81.53* | *81.87* | *99.01* | *89.13* | *88.23* | *89.5* | *79.25* | *74.92* |
| | HGG | *MCC U Net + GN* | *88.64* | *82.21* | *76.88* | *79.6* | *69.8* | *62.44* | *92.29* | *88.42* | *82.61* | *99.47* | *93.49* | *91.3* | *91.64* | *83.72* | *77.95* |
| | LGG | | *78.12* | *75.98* | *73.31* | *77.94* | *68.13* | *61.95* | *90.19* | *82.62* | *82.71* | *99.63* | *90.39* | *89.94* | *89.92* | *81.34* | *75* |
| | HGG | MAC U Net + BN | 92.41 | 83.49 | 80.15 | 84.17 | 71.32 | 65.14 | 92.45 | 89.84 | 82.84 | 99.02 | 93.52 | 91.43 | 91.89 | 84.73 | 78.01 |
| | LGG | | 83.18 | 76.4 | 73.33 | 83.36 | 70.96 | 62.66 | 90.68 | 84.43 | 80.92 | 98.56 | 91.8 | 90.23 | 89.8 | 82.69 | 76.83 |
| | **HGG** | **MAC U-Net + GN** | **94.47** | **84.12** | **82.72** | **89.54** | **77.83** | **67.68** | **93.05** | **89.9** | **83.1** | **99.2** | **93.44** | **92.34** | **92.14** | **87.63** | **77.64** |
| | **LGG** | | **85.71** | **78.85** | **74.16** | **84.85** | **71.32** | **66.33** | **91.16** | **84.12** | **80.86** | **98.83** | **92.56** | **90.87** | **90.1** | **84.31** | **75.53** |
| *2019* | HGG | *U Net* | *85.56* | *75.02* | *70.18* | *74.18* | *65.6* | *60.83* | *88.4* | *81.46* | *80.65* | *99.1* | *90.7* | *87.31* | *78.16* | *75.15* | *74.81* |
| | LGG | | *64.07* | *61.91* | *60.15* | *60.5* | *59.35* | *58.02* | *84.22* | *75.15* | *75.5* | *97.13* | *88.16* | *86.3* | *77.92* | *73.43* | *72.69* |
| | HGG | *U Net + GN* | *88.39* | *80.59* | *76.5* | *77.37* | *68.2* | *61.4* | *89.12* | *82.28* | *81.69* | *99.11* | *92.16* | *88.47* | *81.19* | *76.93* | *75.7* |
| | LGG | | *66.63* | *64.42* | *64.73* | *63.6* | *61.15* | *60.09* | *88.31* | *81.2* | *81.54* | *97.77* | *89.19* | *86.91* | *81.39* | *74.77* | *73.49* |
| | HGG | *MCC U Net + BN* | *88.45* | *80.69* | *76.72* | *79.43* | *68.7* | *62.3* | *89.81* | *82.1* | *81.63* | *99.27* | *92.48* | *89.31* | *85.19* | *80.38* | *76.09* |
| | LGG | | *69.31* | *65.52* | *66.02* | *63.66* | *62.04* | *61.91* | *88.09* | *81.63* | *81.87* | *98.2* | *89.63* | *87.29* | *84.22* | *77.86* | *74.89* |
| | HGG | *MCC U Net + GN* | *88.56* | *80.84* | *77* | *81.12* | *69.04* | *62.43* | *89.89* | *86.2* | *81.76* | *99.33* | *92.64* | *89.41* | *87.27* | *82.19* | *77.13* |
| | LGG | | *76.62* | *72.18* | *70.86* | *78.38* | *66.72* | *62.44* | *88.1* | *81.76* | *81.91* | *98.88* | *90.19* | *88.28* | *86.12* | *79.31* | *75.13* |
| | HGG | MAC U Net + BN | 92.86 | 83.09 | 81.85 | 85.69 | 72.56 | 66.46 | 94.15 | 89.9 | 83.04 | 99.2 | 93.81 | 91.19 | 92.07 | 86.13 | 78.1 |
| | LGG | | 85.18 | 77.74 | 74.93 | 83.8 | 71.49 | 63.94 | 91.77 | 84.93 | 81.92 | 99.1 | 92.34 | 90.9 | 90.48 | 83.09 | 77.3 |
| | **HGG** | **MAC U-Net + GN** | **94.75** | **84.23** | **82.84** | **86.49** | **74.42** | **68.81** | **95.13** | **89.87** | **84.72** | **99.6** | **94.57** | **93.09** | **92.75** | **88.33** | **78.47** |
| | **LGG** | | **86.56** | **80.85** | **75.83** | **85.22** | **72.41** | **66.02** | **93.86** | **86.42** | **83.71** | **99.18** | **93.65** | **91.94** | **91.84** | **85.81** | **77.93** |
| *2020* | ALL | *U Net* | *82.62* | *72.71* | *68.37* | *72.47* | *64.26* | *61.23* | *84.53* | *82.62* | *78.36* | *97.6* | *90.17* | *89.63* | *80.51* | *77.76* | *73.58* |
| | ALL | *U Net + GN* | *83.14* | *73.28* | *69.48* | *74.39* | *64.79* | *62.38* | *85.73* | *83.05* | *78.86* | *97.68* | *91.42* | *90.16* | *81.43* | *78.36* | *74.08* |
| | ALL | *MCC U Net + BN* | *83.47* | *73.91* | *70.72* | *75.83* | *65.18* | *62.43* | *86.22* | *82.43* | *79.17* | *98.28* | *92.67* | *90.93* | *82.68* | *79.05* | *74.46* |
| | ALL | *MCC U Net + GN* | *83.9* | *76.25* | *74.63* | *77.41* | *67.69* | *62.58* | *88.96* | *83.77* | *80.58* | *98.8* | *93.28* | *91.59* | *86.5* | *81.27* | *74.83* |
| | ALL | MAC U Net + BN | 89.28 | 82.7 | 80.18 | 82.26 | 72.7 | 66.54 | 92.82 | 88.64 | 83.13 | 99.39 | 94.45 | 92.88 | 90.91 | 85.27 | 77.19 |
| | **ALL** | **MAC U-Net + GN** | **90.45** | **84.3** | **82.16** | **86.59** | **74.38** | **68.45** | **94.73** | **89.4** | **84.68** | **99.3** | **94.4** | **92.7** | **91.49** | **86.29** | **78.58** |
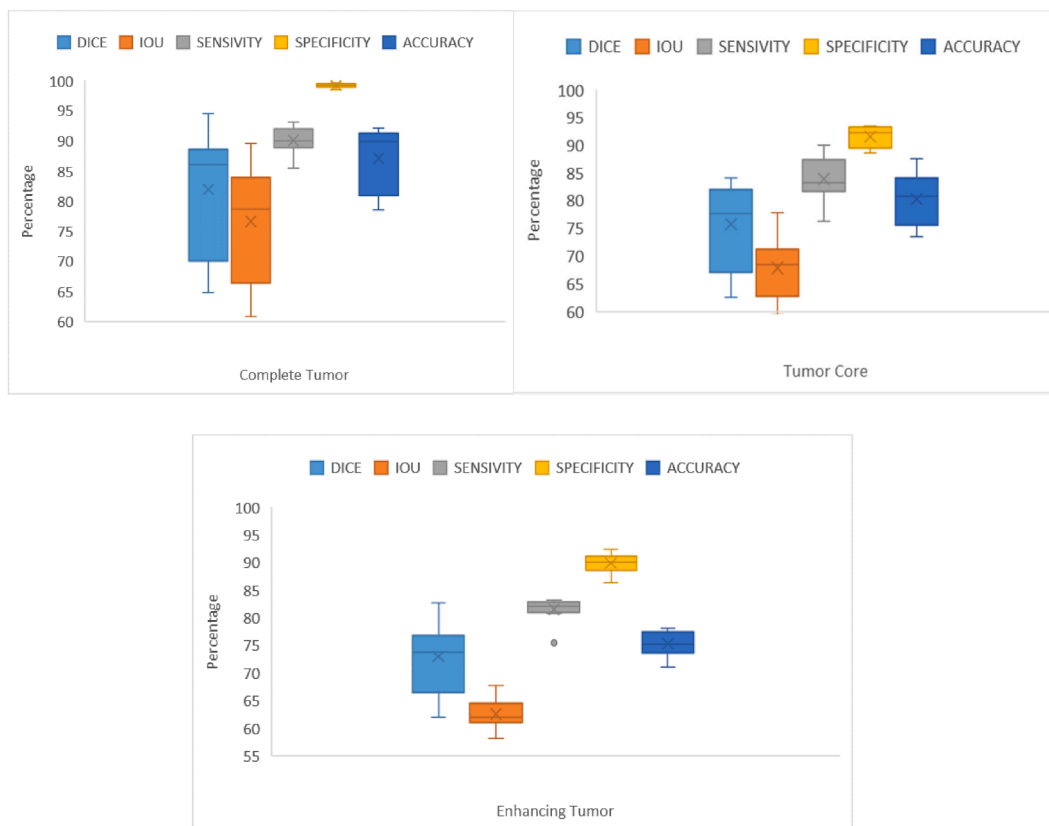
**Fig. 7.** Performance analysis of proposed method on various tumors.

Cascaded U Net Architecture. This multi-channel cascaded U-Net with BN improved the performance by 0.24, 0.82,0.31 and 0.39,1.18,0.4 for both Dice and IoU for segmentation of CT, TC and ET respectively. To reduce the computational complexity, attention mechanism is integrated with cascaded U-Net, that improved the dice performance by 3.77, 1.28, 3.27, and IoU performance by 4.57, 1.52, 2.7 respectively. Further, to take more advantage, for the segmentation of various tumors is boosted with the proposed MAC U Net and its DSC and IoU values are 94.47, 84.12, 82.72, and 89.54, 77.83, 67.68 respectively.

U-Net 9 Layer model has given 85.56, 75.02 and 70.18 of dice score on HGG and 64.07, 61.9, and 60.15 of dice score on LGG of BraTs 2019 dataset in the segmentation of Complete Tumor, Tumor Core and Enhancing Tumors respectively. Further, proposed method with Batch Normalization is evaluated on 2019 dataset and achieved 76.62, 72.18 and 70.86 in the segmentation of CT, TC and ET respectively on Low Grade Gliomas. The proposed work is evaluated on BraTS 2019 dataset and yielded dice scores of 94.75, 84.23, 82.84 and 86.56, 80.85, 75.83 on the segmentation of Full tumor, Tumor Core and Enhancing Tumor of HGG and LGG respectively. IoU coefficients of Full tumor, Tumor Core and Enhancing tumor of the proposed work on HGG and LGG as 86.49, 74.42, 68.81 and 85.22, 72.41, 66.02 respectively. From the results, the model has presented consistent results in the segmentation of small-scale gliomas also. The proposed model is also evaluated on 127 subjects of BraTS 2020 test dataset and achieved dice coefficient values as 90.45, 84.3, and 82.16 on segmentation of Complete Tumor, Tumor Core and Enhancing Tumors respectively. The proposed model has evaluated without Attention gate, and yielded the dice score as 83.9, 76.25 and 74.63 in the segmentation of Complete Tumor, Tumor Core and Enhancing Tumors respectively and it was presenting 6–7% better performance in the segmentation of Tumor Core and Enhancing Tumor with U-Net 9 Layer Architecture. The proposed model with Batch Normalization has realized 89.28, 82.7 and 80.18 of dice score in the segmentation of Complete Tumor, Tumor Core and Enhancing Tumors

**Table 5**
Performance Comparison of the proposed framework with existing frameworks.

| Dataset | Framework | DSC | | |
|---|---|---|---|---|
| | | WT | TC | ET |
| BraTs 18 | Res U-Net [37] | 87 | 80.2 | 76 |
| | AG Res U-Net [37] | 87.2 | 80.8 | 77.2 |
| | MCCN + CRF's[39] | 88.24 | 74.81 | 71.78 |
| | OM-Net[70] | 88.42 | 79.6 | 77.75 |
| | Multi Model 3D U-Net [35] | 87 | 66 | 79 |
| | Proposed MAC U-Net + GN | **94.47** | **84.12** | **82.72** |
| BraTs 20 | 3D-U-Net [67] | 81 | 82 | 77 |
| | Two Stage VAE using AG [68] | 90.41 | 83.5 | 79.58 |
| | Self-ensemble deeply supervised 3D-U Net [69] | 88.59 | 84.27 | 78.5 |
| | ME-Net [71] | 88.3 | 73.9 | 70.2 |
| | Proposed MAC U-Net + GN | **90.45** | **84.3** | **82.16** |

respectively. The comparative analysis clearly points out the need for Attention mechanism and Group Normalization in the segmentation of Tumor Core and Enhancing Tumor. From Fig. 7 it is clearly evident that the proposed model has presented higher accuracy in the segmentation of Complete Tumor, Tumor core and enhanced tumor.

The Res U-Net with BN [37] achieved 87, 80.23, and 76, dice coefficient values for the segmentation of Full, Core, and Enhancing Tumor on BraTs 2018 dataset respectively. K. Hu et al., [39] proposed Multi-Channel Cascaded Neural Network and conditional random fields with all four modalities and achieved a dice score of 88.24, 74.81 and 71.78 in the segmentation of CT, TC and ET respectively. C.Zhou et al., proposed a One Pass Multitask network and achieved the dice score of 88.42, 79.6, and 77.75 in the segmentation of CT, TC and ET respectively. M. Ghaffari et al., proposed Densely connected 3D CNN and achieved a dice score of 87, 66, and 79 in the segmentation of CT, TC and

ET respectively.

Attention gated Res U-Net also presented a similar performance but outperformed Res U Net in the segmentation of Enhancing Tumor. 3D U Net [67] shown higher accuracy in the segmentation of Tumor core. Two stage variation auto encoders with Attention Gate [68] has yielded the dice values of 90.41, 83.5 and 79.58 on Full, Core, and Enhancing Tumor respectively. In addition, the proposed model achieved 94.47, 84.12, and 82.72 as segmentation of Full, Core, and Enhancing Tumor respectively. Further Self Ensemble Deeply Supervised 3D U Net [69] yielded the dice values of 89, 84 and 79 on Full, Tumor core and Enhancing Tumor respectively. Our proposed approach showed consistent performance on Tumor Core and Enhanced Tumor and outperformed whole tumor with 94.47 when compared to [69]. Z. Wenbo et.al., proposed ME Net for the segmentation of CT, TC and ET types and achieved the dice score as 88.3, 73.9, and 70.2 respectively. The performance of the proposed method is compared with other state of art methods is tabulated in Table 5.

This study explored the usage of GN over BN in the case of small batches due to memory constraints. To address the irrelevant and noisy features and to give attention to relevant features, we employed a series of attention gates in skip connections of U-Net architecture at up-sampling and replaced BN with GN to take care of gradient imbalance at each epoch. Our model outperforms basic U-Net, AG U-Net with BN and ResU-Net, AG ResU-Net, 3D U-Net, Two Stage VAE using Attention gate and Self ensemble deeply supervised 3D U-Net in the detection of various tumors.

## 5. Conclusion

The existing deep learning models have limited ability to automatically detect and segment low-grade tumors from edema using MRI imaging sequences. In this work, a novel Attention Gate-based cascaded U-Net ensemble framework was proposed to segment early-stage low-grade brain tumor substructures. The effectiveness of group normalization with attention gate was also explored with skip connections to segment small-scale brain tumors using highlighted salient feature information. Experimental outcomes of the proposed work outperform baseline U-nets on the BraTS 2018, BraTS 2019 and BraTS 2020 datasets. Overall, the proposed model of MAC U-Net achieves superior performance over typical brain tumor segmentation methods. In the future scope, we aimed to propose a 3D model with an appropriate combination of attention mechanisms in brain tumor segmentation to realize better segmentation accuracy.

## CRediT authorship contribution statement

**Siva Koteswara Rao Chinnam:** Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation. **Venkatramaphanikumar Sistla:** Methodology, Validation, Writing – review & editing, Visualization, Investigation. **Venkata Krishna Kishore Kolli:** Methodology, Resources, Supervision, Writing – review & editing.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References:

[1] D. Shen, E.H. Herskovits, C. Davatzikos, An adaptive-focus statistical shape model for segmentation and shape modeling of 3-D brain structures, IEEE Trans. Med. Imaging 20 (4) (2001) 257–270, https://doi.org/10.1109/42.921475.

[2] S. Li and M. Tan, "Gene selection and tissue classification based on support vector machine and genetic algorithm," *1st Int. Conf. Bioinforma. Biomed. Eng. ICBBE*, pp. 192–195, 2007, doi: 10.1109/ICBBE.2007.52.

[3] R.G. Steen, Edema and tumor perfusion: characterization by quantitative 1H MR imaging, Am. J. Roentgenol. 158 (2) (1992) 259–264, https://doi.org/10.2214/ajr.158.2.1729777.

[4] Y.P. Jin Liu, Min Li, Jianxin Wang, Wu Fangxiang, Tianming Liu, A survey of MRI-based brain tumor segmentation methods, Tinshhua Sci. Technol. 19 (6) (2014) 578–595.

[5] A. Wadhwa, A. Bhardwaj, V. Singh Verma, A review on brain tumor segmentation of MRI images, Magn. Reson. Imaging 61 (2019) 247–259, https://doi.org/10.1016/j.mri.2019.05.043.

[6] J. Nalepa, et al., Fully-automated deep learning-powered system for DCE-MRI analysis of brain tumors, Artif. Intell. Med. 102 (2020), https://doi.org/10.1016/j.artmed.2019.101769.

[7] S.K.R. Chinnam, V. Sistla, V.K.K. Kolli, SVM-PUK kernel based MRI-brain tumor identification using texture and Gabor wavelets, Trait. du Signal 36 (2) (Apr. 2019) 185–191, https://doi.org/10.18280/ts.360209.

[8] A. Gooya, K.M. Pohl, M. Bilello, L. Cirillo, G. Biros, E.R. Melhem, C. Davatzikos, GLISTR: glioma image segmentation and registration, IEEE Trans. Med. Imaging 31 (10) (2012) 1941–1954.

[9] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W. M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[10] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 07–12 (2015) 3431–3440, https://doi.org/10.1109/CVPR.2015.7298965.

[11] O. Ronneberger, P. Fischer, T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.

[12] B.J. Erickson, P. Korfiatis, T.L. Kline, Z. Akkus, K. Philbrick, A.D. Weston, Deep learning in radiology: does one size fit all? J. Am. Coll. Radiol. 15 (3) (2018) 521–526, https://doi.org/10.1016/j.jacr.2017.12.027.

[13] A. Işin, C. Direkoğlu, M. Şah, Review of MRI-based brain tumor image segmentation using deep learning methods, Procedia Comput. Sci. 102 (August) (2016) 317–324, https://doi.org/10.1016/j.procs.2016.09.407.

[14] G. Mohan, M.M. Subashini, MRI based medical image analysis: Survey on brain tumor grade classification, Biomed. Signal Process. Control 39 (2018) 139–161, https://doi.org/10.1016/j.bspc.2017.07.007.

[15] A. Hamamci, N. Kucuk, K. Karaman, K. Engin, G. Unal, Tumor-Cut: Segmentation of brain tumors on contrast enhanced mr images for radiosurgery applications, IEEE Trans. Med. Imaging 31 (3) (2012) 790–804, https://doi.org/10.1109/TMI.2011.2181857.

[16] R. Zaouche, A. Belaid, S. Aloui, B. Solaiman, L. Lecornu, D. Ben Salem, S. Tliba, Semi-automatic method for low-grade gliomas segmentation in magnetic resonance imaging, Irbm 39 (2) (2018) 116–128.

[17] S. Saha, A.K. Alok, A. Ekbal, Brain image segmentation using semi-supervised clustering, Expert Syst. Appl. 52 (2016) 50–63, https://doi.org/10.1016/j.eswa.2016.01.005.

[18] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination with Hierarchical Conditional Random Field Regularization," *Int. Conf. Med. Image Comput. Comput. Interv. Springer Berlin Heidelb.*, pp. 354–361, 2011, doi: 10.1007/978-3-642-23626-6_44.

[19] H. Al-Dmour, A. Al-Ani, A clustering fusion technique for MR brain tissue segmentation, Neurocomputing 275 (2018) 546–559, https://doi.org/10.1016/j.neucom.2017.08.051.

[20] R. Meier, S. Bauer, J. Slotboom, R. Wiest, and M. Reyes, "Appearance-and Context-sensitive Features for Brain Tumor Segmentation," *MICCAI BRATS Chall.*, no. October, 2014, doi: 10.13140/2.1.3766.7846.

[21] O. Charron, A. Lallement, D. Jarnet, V. Noblet, J.B. Clavier, P. Meyer, Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network, Comput. Biol. Med. 95 (2018) 43–54, https://doi.org/10.1016/j.compbiomed.2018.02.004.

[22] H. Dong, G. Yang, F. Liu, Y. Mo, Y. Guo, Automatic brain tumor detection and segmentation using U-net based fully convolutional networks, Commun. Comput. Inf. Sci. 723 (2017) 506–517, https://doi.org/10.1007/978-3-319-60964-5_44.

[23] A. Beers et al., "Sequential 3D U-Nets for Biologically-Informed Brain Tumor Segmentation," pp. 1–7, 2017, [Online]. Available: http://arxiv.org/abs/1709.02967.

[24] T. Yang and J. Song, "An automatic brain tumor image segmentation method based on the u-net," *2018 IEEE 4th Int. Conf. Comput. Commun. ICCC 2018*, pp. 1600–1604, 2018, doi: 10.1109/CompComm.2018.8780595.

[25] S. M. Kamrul Hasan and C. A. Linte, "A Modified U-Net Convolutional Network Featuring a Nearest-neighbor Re-sampling-based Elastic-Transformation for Brain Tissue Characterization and Segmentation," *2018 IEEE West. New York Image Signal Process. Work. WNYISPW 2018*, pp. 1–5, 2018, doi: 10.1109/WNYIPW.2018.8576421.

[26] Z. Wu, F. Chen, and D. Wu, "A Novel Framework Called HDU for Segmentation of Brain Tumor," *2018 15th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2018*, pp. 81–84, 2019, doi: 10.1109/ICCWAMTIP.2018.8632590.

[27] J. Bernal, K. Kushibar, D.S. Asfaw, S. Valverde, A. Oliver, R. Martí, X. Lladó, Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: a review, Artif. Intell. Med. 95 (2019) 64–81.

[28] M. Angulakshmi, G.G. Lakshmi Priya, Automated brain tumour segmentation techniques— A review, Int. J. Imaging Syst. Technol. 27 (1) (2017) 66–77, https://doi.org/10.1002/ima.22211.

[29] J.D. Nielsen, K.H. Madsen, O. Puonti, H.R. Siebner, C. Bauer, C.G. Madsen, G. B. Saturnino, A. Thielscher, Automatic skull segmentation from MR images for realistic volume conductor models of the head: assessment of the state-of-the-art, Neuroimage 174 (2018) 587–598.

[30] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B.B. Avants, N. Ayache, P. Buendia, D.L. Collins, N. Cordier, J.J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C.R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N.M. John, E. Konukoglu, D. Lashkari, J.A. Mariz, R. Meier, S. Pereira, D. Precup, S.J. Price, T.R. Raviv, S.M.S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C.A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T.J. Taylor, O.M. Thomas, N.J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D.H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, K. Van Leemput, The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2015) 1993–2024.

[31] Y. Ding, F. Chen, Y. Zhao, Z. Wu, C. Zhang, D. Wu, A stacked multi-connection simple reducing net for brain tumor segmentation, IEEE Access 7 (2019) 104011–104024, https://doi.org/10.1109/access.2019.2926448.

[32] Z. Zhao and Z. Zhao, "An enhanced U-Net for brain tumor segmentation," IEEE Int. Conf. Robot. Biomimetics, ROBIO 2019, no. December, pp. 3054–3058, 2019, doi: 10.1109/ROBIO49542.2019.8961564.

[33] K. Kaewrak, J. Soraghan, G. Di Caterina, D. Grose, Modified U-Net for automatic brain tumor regions segmentation, Eur. Signal Process. Conf. 2019 (2019) 1–5, https://doi.org/10.23919/EUSIPCO.2019.8902767.

[34] X.-Y. Zhou, G.-Z. Yang, Normalization in training U-Net for 2-D biomedical semantic segmentation, IEEE Robot. Autom. Lett. 4 (2) (2019) 1792–1799.

[35] M. Ghaffari, A. Sowmya, R. Oliver, and L. Hamey, "Multimodal Brain Tumour Segmentation using Densely Connected 3D Convolutional Neural Network," 2019 Digit. Image Comput. Tech. Appl. DICTA 2019, pp. 1–5, 2019, doi: 10.1109/DICTA47822.2019.8946023.

[36] M. Ghaffari, A. Sowmya, R. Oliver, Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the BraTS 2012–2018 Challenges, IEEE Rev. Biomed. Eng. 13 (2020) 156–168, https://doi.org/10.1109/RBME.2019.2946868.

[37] J. Zhang, Z. Jiang, J. Dong, Y. Hou, and B. Liu, "Attention Gate ResU-Net for automatic MRI brain tumor segmentation," IEEE Access, vol. 8, pp. 1–1, 2020, doi: 10.1109/access.2020.2983075.

[38] Z. Zhang, C. Wu, S. Coleman, D. Kerr, DENSE-INception U-net for medical image segmentation, Comput. Methods Programs Biomed. 192 (2020), 105395, https://doi.org/10.1016/j.cmpb.2020.105395.

[39] K. Hu, Q. Gan, Y. Zhang, S. Deng, F. Xiao, W. Huang, C. Cao, X. Gao, Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field, IEEE Access 7 (2019) 92615–92629.

[40] X. Zhou, X. Li, K. Hu, Y. Zhang, Z. Chen, X. Gao, ERV-Net: An efficient 3D residual neural network for brain tumor segmentation, Expert Syst. Appl. 170 (2021) 114566.

[41] L. Szilágyi, Z. Benyó, S. M. Szilágyi, and H. S. Adam, "MR Brain Image Segmentation Using an Enhanced Fuzzy C-Means Algorithm," in Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings, 2003, vol. 1, pp. 724–726.

[42] X. Bai, Y. Zhang, H. Liu, Y. Wang, Intuitionistic center-free FCM clustering for MR brain image segmentation, IEEE J. Biomed. Heal. Inform. 23 (5) (2019) 2039–2051, https://doi.org/10.1109/JBHI.2018.2884208.

[43] A. Kouhi, H. Seyedarabi, and A. Aghagolzadeh, "A modified FCM algorithm for MRI brain image segmentation," in 2011 7th Iranian Conference on Machine Vision and Image Processing, MVIP 2011 - Proceedings, 2011, pp. 1–5, doi: 10.1109/IranianMVIP.2011.6121551.

[44] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J.S. Kirby, J.B. Freymann, K. Farahani, C. Davatzikos, Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features, Sci. data 4 (1) (2017), 170117, https://doi.org/10.1038/sdata.2017.117.

[45] S. Bakas et al., "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge," arXiv, 2018, [Online]. Available: http://arxiv.org/abs/1811.02629.

[46] Y. Wu, K. He, Group normalization, Int. J. Comput. Vis. 128 (3) (2019) 742–755, https://doi.org/10.1007/s11263-019-01198-w.

[47] P. Kurian and V. Jeyakumar, Multimodality medical image retrieval using convolutional neural network. Elsevier Inc., 2020.

[48] N.K. Manaswi (Ed.), Deep Learning with Applications Using Python, Apress, Berkeley, CA, 2018.

[49] K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, IEEE Trans. Pattern Anal. Mach. Intell. 42 (2) (2018) 386–397, https://doi.org/10.1109/TPAMI.2018.2844175.

[50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[51] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017, vol. 2017-Janua, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[52] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in 32nd International Conference on Machine Learning, ICML 2015, 2015, vol. 1, pp. 448–456.

[53] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016 (2016) 770–778, https://doi.org/10.1109/CVPR.2016.90.

[54] S.R. Ammarah Farooq, S.M. Anwar, M. Awais, A Deep CNN based Multi-class classification of Alzheimers Disease using MRI, IEEE Instrum. Meas. Soc. (2017) 3–8.

[55] S. Jetley, N. A. Lord, N. Lee, and P. H. S. Torr, "Learn to pay attention," in 6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings, 2018, pp. 1–14.

[56] A. Graves, G. Wayne, and I. Danihelka, "Neural Turing Machines," pp. 1–26, 2014, [Online]. Available: http://arxiv.org/abs/1410.5401.

[57] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–15, 2015.

[58] M. T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing, 2015, pp. 1412–1421, doi: 10.18653/v1/d15-1166.

[59] A. Vaswani, et al., Attention is all you need, Adv. Neural Inf. Process. Syst. vol. 2017-Decem, no. Nips (2017) 5999–6009.

[60] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, J. Garcia-Rodriguez, A survey on deep learning techniques for image and video semantic segmentation, Appl. Soft Comput. J. 70 (2018) 41–65, https://doi.org/10.1016/j.asoc.2018.05.018.

[61] H. Costa, G. M. Foody, and D. S. Boyd, "Supervised methods of image segmentation accuracy assessment in land cover mapping," Remote Sens. Environ., vol. 205, no. December 2016, pp. 338–351, 2018, doi: 10.1016/j.rse.2017.11.024.

[62] M. Dubuisson, A. K. Jain, E. Lansing, and A. B. B, "A modified Hausdorff distance for object matching," in Proceedings of 12th International Conference on Pattern Recognition, 1994, pp. 566–568, doi: 10.1109/ICPR.1994.576361.

[63] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, Comput. Vis. Pattern Recognit. (2019) 1–13.

[64] J.E. Hoffman, Spatial attention in vision, Psychol. Res. 48 (4) (1986) 221–229, https://doi.org/10.1007/BF00309086.

[65] T. Zhou, S. Ruan, Y. Guo, I. Rouen, L. Apprentissage, and I. Rouen, "A MULTI-MODALITY FUSION NETWORK BASED ON ATTENTION MECHANISM FOR BRAIN TUMOR SEGMENTATION," pp. 2020–2023, 2020.

[66] H. Zhao, J. Jia, and V. Koltun, "Exploring Self-attention for Image Recognition," in IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10076–10085, doi: 10.1109/CVPR42600.2020.01009.

[67] L. M. Ballestar and V. Vilaplana, "MRI Brain Tumor Segmentation and Uncertainty Estimation using 3D-UNet Architectures," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12658 LNCS, no. 1, pp. 376–390, 2021, doi: 10.1007/978-3-030-72084-1_34.

[68] C. Lyu and H. Shu, "A Two-Stage Cascade Model with Variational Autoencoders and Attention Gates for MRI Brain Tumor Segmentation," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12658 LNCS, pp. 435–447, 2021, doi: 10.1007/978-3-030-72084-1_39.

[69] T. Henry, et al., Brain Tumor Segmentation with Self-ensembled, Deeply-Supervised 3D U-Net Neural Networks: A BraTS 2020 Challenge Solution, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) vol. 12658 LNCS (2021) 327–339, https://doi.org/10.1007/978-3-030-72084-1_30.

[70] C. Zhou, S. Chen, C. Ding, D. Tao, Learning Contextual and Attentive Information for Brain Tumor Segmentation, vol. 2, Springer International Publishing, 2019.

[71] Z. Wenbo, et al., ME-Net: Multi-encoder net framework for brain tumor segmentation, Int. J. Imaging Syst. Technol. 31 (2021) 1834–1848, https://doi.org/10.1002/ima.22571.