Biomedical Signal Processing and Control 71 (2022) 103077

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors

Dhiraj Maji, Prarthana Sigedar, Munendra Singh *

Department of Mechatronics Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, Karnataka 576104, India

ARTICLE INFO

Keywords:
MRI
Brain tumor segmentation
Res-UNet
Attention gates
Guided decoder

ABSTRACT

The automatic segmentation of brain tumors in Magnetic Resonance Imaging (MRI) plays a major role in accurate diagnosis and treatment planning. The present study proposes a new deep learning generator architecture called Attention Res-UNet with Guided Decoder (ARU-GD) for the segmentation of brain tumors. The proposed generator architecture have the capability to explicitly guide the learning process of each decoder layer. The individual loss function to each decoder layer helps to supervise the learning process of each layer in the decoder and thereby enables them to generate better feature maps. The attention gates in the generator focuses on the activation of relevant information instead of allowing all information to pass through the skip connections in the Res-UNet. Our model performed well in comparison to the baseline models i.e. UNet, Res-UNet, and Res-UNet with attention gates. The proposed ARU-GD is compared with popular deep learning models VGG-Net, MobileNet, QuickNAT, DenseNet and XceptionNet, and BraTS 2019 leaderboard models. The proposed ARU-GD has achieved Dice Scores of 0.911, 0.876 and 0.801 and mean IoU of 0.838, 0.781 and 0.668 on the whole tumor, tumor core and enhancing tumor respectively on unseen High-Grade Glioma test data. The implementation code is available on the following Github link.

1. Introduction

Growth of abnormal cells in the brain may cause brain tumor. The gliomas are the type of tumors, which originate from glial cells. They are broadly categorized as Higher Grade Gliomas (HGG) and Lower Grade Gliomas (LGG). The HGG grows rapidly with a high degree of cell infiltration in comparison to the LGG. Several technologies are used for non-invasive diagnosis, and to monitor, and analyze the tumors such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), and Computed Tomography (CT). However, MRI is an effective diagnosis technique for detecting soft tissue tumors and hence popular for brain imaging.

The radiologists need to diagnose brain tumors accurately to decide the best course of treatment for the patients. Manual segmentation requires a lot of expertise and skills, still it can sometimes be erroneous. Automatic segmentation of tumors in MRI images helps in analyzing pathological conditions in a better way by measuring the size, location, structure and grade of the tumor. Gliomas can diffuse out, have poor contrasts and are difficult to detect due to the intensity variations in MRI scans. This makes the task of segmenting the brain tumors quite arduous. Previously, a wide range of techniques were developed for the segmentation of brain tumors in MRI images [30,2]. Some of the common image segmentation methods include, region growing methods [8], clustering based methods [27,26], and watershed algorithms [28,22]. However, the effectiveness of these techniques are limited because of their sensitivity to noise and artifacts.

Recently, brain tumor segmentation using Deep Learning techniques has shown promising results due to their ability to learn complex features from the data [19]. Convolutional Neural Networks (CNNs) and Fully Convolutional Networks (FCNs) based Deep Learning models like SegNet [3], Deep Neural Networks [10], U-Net [23], QuickNAT [24], DenseNet [13] and their variants [1,15,20,31] have shown noteworthy performance in segmentation tasks.

The present study proposes a new deep learning architecture, that introduces guided decoder and inculcates the properties of Res-UNet [32] and attention gates [21], which we call as the Attention Res-UNet with Guided Decoder (ARU-GD). The main contributions of this proposed work are as follows:

* Corresponding author.
E-mail addresses: munendra107@gmail.com, munendra.singh@manipal.edu (M. Singh).

https://doi.org/10.1016/j.bspc.2021.103077
Received 6 November 2020; Received in revised form 1 July 2021; Accepted 20 August 2021
Available online 3 September 2021

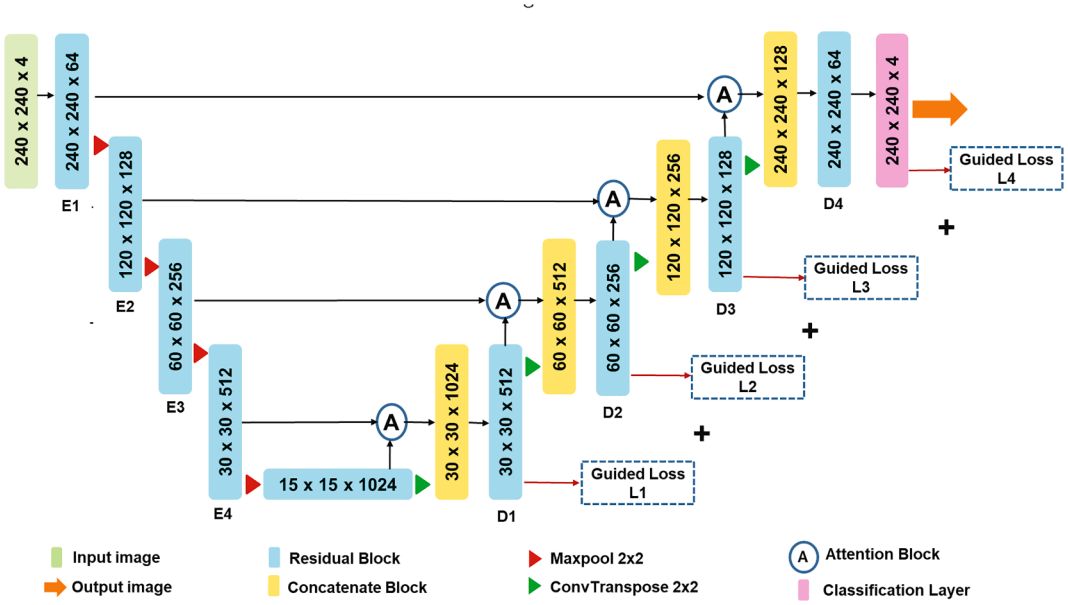

Fig. 1. Attention Res-UNet with Guided Decoder architecture.

1. The network architecture consists of a new guided decoder, which supervises the learning process of the decoder and helps in producing improved features.
2. The proposed weighted guided loss improves prediction capabilities of each layer of the decoder, thereby increasing the prediction accuracy of the final output layer.
3. Hybrid network architecture inculcating Attention Gates with a backbone of Res-UNet architecture, which focuses on the activation of relevant regions.
4. Evaluation is done on the HGG data of BraTS 2019 to show that proposed ARU-GD outperforms its baseline models and current state of the art segmentation techniques.

The organization of the paper is as follows. Section 2 discusses the related deep learning models for the segmentation of brain tumors. Section 3 formulates methodology, where we have discussed the highlights of the network architecture, and its components. Section 4 presents the results obtained by proposed model in terms of ablation study and comparison study. Finally, Section 5 draws the conclusion of the present study.

2. Related work

Nowadays, the image segmentation techniques are overshadowed by wide use of deep learning models, which provide better segmentation results [19]. Badrinarayanan et al. presented an architecture, named SegNet based on Fully Convolutional Networks (FCNs), which uses a novel technique of unpooling to perform non-linear upsampling [3]. Alqazzaz et al. implemented the SegNet for brain tumor segmentations [1]. This work has combined four SegNet models, which are trained on the four different modalities individually. Roy et al. proposed QuickNAT, which is one of the quick and accurate segmentation method [24]. This model used an encoder decoder architecture for segmenting the tumors in MRI scans in 20 s. Huang et al. proposed DenseNet, which has used the output of each layer as inputs for all the subsequent layers [13]. Multimodal Brain Tumor segmentation (BraTS) is an internationally recognized medical image segmentation challenge [18,5,6]. They focus on evaluating the cutting edge models for segmenting brain tumors from MRI images. In the 2019 BraTS challenge, Jiang et al. [14], Zhao et al. [34], and McKinley et al. [17] were the top performers. These state of the art models have gained noteworthy results in brain tumor segmentation. However, there is a huge possibility for improvement in the accuracy of semantic segmentation of brain tumors in MRI images.

Previously, Havaei et al. has proposed a Convolutional Neural Network (CNN) based Deep Neural Network, which captures both local features and global contextual features by using a two-pathway architecture [10]. Long et al. has shown that FCNs are highly efficient and produce praiseworthy results [16]. Since then CNNs and FCNs have become very commonly used neural networks for semantic segmentation tasks. Theoretically, deeper networks should perform better in segmentation tasks. However, as the model gets deeper, training becomes arduous and complex. This problem is overcome by the introduction of residual networks [11]. Lei Bi et al. used residual networks in liver lesions segmentation task and proposed a novel architecture by cascading ResNet to detect the fine boundaries of the liver and liver lesions [7]. Further, Ronneberger et al. proposed the U-Net [23], which is based on FCNs and is the most popular architecture in biomedical imaging. It gained popularity due its skip connections, which is used to pass important information from encoder to decoder stage. A lot of modifications have been made to this well-known architecture in order to further improve its accuracy. Yang et al. have modified the U-Net by adding residual blocks [32], which help U-Net in extracting more features at every layer. Due to its greater performance and effectiveness in feature extraction, Res-UNet is widely used as a base model for many architectures in Deep Learning [31]. Kermi et al. used a slightly modified architecture of U-Net with a combined loss function of weighted cross entropy and generalized dice loss to procure huge improvements in segmentation accuracy [15]. However, these methods pass all the extracted features through the skip connection to the decoder stage.

Attention gates were introduced to suppress the irrelevant activations and to primarily focus on the features relevant to the specific task [21]. This technique reduces the possibility of false positives and helps direct focus of the model to the task thereby increasing the performance of the model. Attention gate was first used in the area of natural language processing for machine translation [4]. Recently, Oktay et al. used the attention gates with U-Net for pancreas segmentation from abdominal CT and has shown that the attention gates remove the requirement of a separate object localization module in CNN [21]. Attention gates with U-Net have also shown promising results in brain tumor segmentation [20] and other tasks [29]. Zhang et al. implemented the attention gates on Res-UNet for the segmentation of brain tumors and have shown its effectiveness over the Res-UNet [33]. We investigate the similar

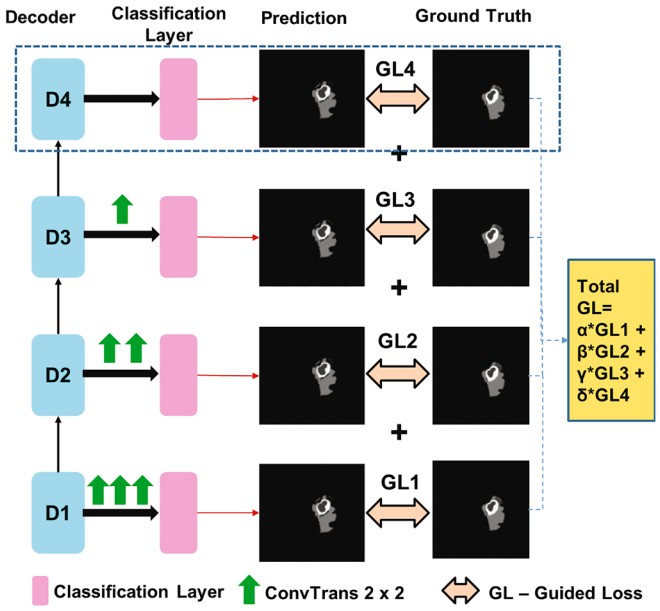

Fig. 2. Overview of Guided Loss.

techniques and inculcate Attention Res-UNet along with Guided Decoder to achieve better segmentation results than the current state-of-the-art models.

3. Method

This section discusses the network architecture details of our proposed model Attention Res-UNet with Guided Decoder (ARU-GD) and the loss functions used in the training process. Further, this section discusses the dataset, preprocessing steps and implementation details of Attention Res-UNet with Guided Decoder.

3.1. Network architecture

The overview of proposed architecture is demonstrated in Fig. 1. The different parts of the model namely attention gate and guided decoder have been delineated as follows.

3.1.1. Proposed guided decoder

The present study proposes to train each decoder layer explicitly with its individual loss function and hence, predicts output at each layer instead of only at the final layer. This arrangement guides network to generate better feature representation at each layer. Further, the prediction at individual decoder layer and transfer of weighted loss of each intermediate layer to the final layer contributes to enhance the segmentation accuracy of the model. A representation of proposed Guided Decoder is shown in Fig. 2. The output of decoder layers D1, D2 and D3 is upsampled to the input size 240 × 240 and compared with the ground truth. This process enables us to generate improved features at these layers. The different weights α, β, γ, and δ of loss functions is passed for the prediction at layer D4 as per their importance. These weights determine the amount of contribution the loss of each layer has on the overall loss of the model. The loss at final layer is given more weightage, whereas the lower layers are given relatively less weightage. It should be noted that only image generated at final layer D4 is considered as the segmented output image of the model.

3.1.2. Attention gates

Attention mechanism focuses on a specific region of the image while ignoring the others. It is similar to the concept on how human visual attention works, where human visuals can focus on a specific point or area while suppressing the surrounding regions. Attention gates can suppress feature activations of irrelevant areas in the image thereby reducing false positives. In present study, they promote updation of parameters in a model in spatial regions, which are relevant to segmentation of tumors. The skip connection connects the encoder to its corresponding decoder through an attention gate as shown in Fig. 1. The attention gate receives two inputs – one from the corresponding encoder containing all contextual and spatial information in that layer and the second input as the gating signal from the decoder layer below it. The output from the attention gate is further passed to the decoder for concatenation. A representation of the attention gate used in our proposed model is shown in Fig. 3.

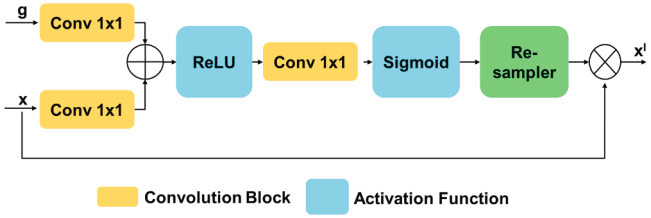

Fig. 3. Depiction of attention gate.

3.2. Attention Res-UNet with Guided Decoder

The segmentation of different sub-regions of tumor is important to estimate its growth and treatment planning. However, the task of multi-class segmentation of tumor is quite challenging. Hence, deeper networks are required to learn more complex features from the data. However, as the network goes deeper, it loses more spatial information thereby reducing the segmentation accuracy. UNet uses skip connections to pass contextual and spatial information between the encoder and decoder. This helps in retrieving important spatial information lost during downsampling. However, the UNet passes all the information through its skip connections. The attention gates are used to merge only relevant feature activations in the decoder. The Residual blocks help the UNet in extracting more features at every layer. Further, to improve the segmentation accuracy, we have proposed and added guided decoder into the attention Res-UNet to produce a new hybrid architecture – "Attention Res-UNet with Guided Decoder" as shown in Fig. 1. The loss in guided decoder helps each layer of the decoder to generate better features, and subsequently increases the segmentation accuracy at the final output layer.

3.2.1. Encoder path

The encoder path consists of four layers, each containing a Residual block and then a 2 × 2 Max-pooling layer. Our model takes an input of size 240 × 240 × 4, where the image resolution is 240 × 240, and we have four channels corresponding to four modalities of the MRI dataset (Flair, T1, contrast enhanced T1 and T2). The residual blocks consist of two 3 × 3 dilated convolutional layers, each followed by a Batch Normalization (BN) layer and Leaky Rectified Linear Unit (Leaky ReLU) activation function. The first residual block has 64 feature maps of size 240 × 240. During down-sampling, the number of feature maps is doubled and the feature map size is halved at each layer. Therefore, the residual block at the fourth layer has 512 feature maps of size 30 × 30. The convolutional blocks in the first and second encoder layers have a dilation rate of 1, whereas the third and fourth encoder layers have dilation rates of 2 and 4 respectively.

3.2.2. Bottleneck

The bottleneck layer connects the encoder and decoder of the proposed model. It contains a residual block with 1024 feature maps of size 15 × 15. The convolution layers in the residual block have a dilation rate of 8. The output of the bottleneck layer moves in two different paths.

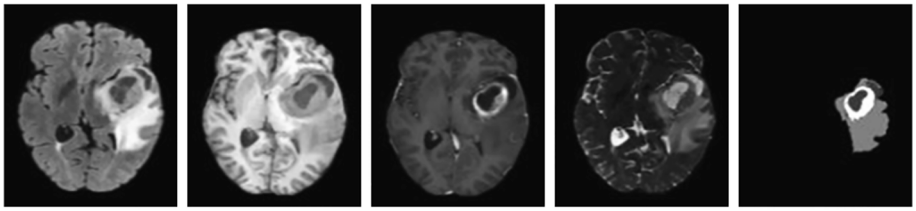

Fig. 4. Different modalities of MRI used in the study (a) FLAIR, (b) T1, (c) contrast enhanced T1 (d) T2 and (e) ground truth of brain tumour and its sub-regions.

One enters the decoder path into the convolutional transpose layer and the other becomes the gating signal for the attention gate.

3.2.3. Decoder path

The decoder path also consists of four layers, each of which has a residual block preceded by a 2 × 2 convolutional transpose layer. Each decoder layer is linked with the corresponding encoder layer using a skip connection and attention gate. Upsampling is done using 2 × 2 convolutional transpose. The output from attention gate concatenates with the upsampled output from the previous decoder layer. After concatenation, the output is passed to the residual blocks. Same residual blocks are used as those in the encoder path. In each decoder layer, the number of feature maps is halved while their size is doubled. The output of the residual block at each layer is upsampled to the size of 240 × 240 and passed through a classification layer and softmax activation function. The classification layer is used to decrease the number of feature maps to N channels (N = 4). A softmax activation converts the feature maps into probabilities. As a result, each channel in the final output of the model becomes probability maps for a particular class.These outputs are then compared with the ground truth and the loss is calculated. Only the output of the last decoder layer D4 is considered as the main output of the model.

3.3. Loss function

Present study utilizes combined weighted dice loss and weighted cross-entropy loss for the segmentation of MRI data.

3.3.1. Weighted dice loss

The Dice Scores of each class are calculated separately and is given as follows:

$$DS_n = \frac{2\sum p_n g_n}{\sum p_n + \sum g_n} \tag{1}$$

where p_n is a predicted output of our model for class n and g_n is the ground truth for class n. Here class denotes different regions of tumour. Segmentation of necrosis and enhancing tumor region is relatively more difficult than the other regions, hence, we have assigned different weights to the Dice Scores of each class. The weights are assigned as 1, 5, 2, 4 for non-tumour region along with background, necrosis, edema and enhancing tumor respectively. The weighted Dice Score is given as follows:

$$DS_{weighted} = \frac{\sum_n w_n DS_n}{\sum_n w_n} \tag{2}$$

where w_n is the weight factor for class n and DS_n is the dice score of class n. The Weighted Dice Loss (WDL) is calculated with the help of above mentioned weighted dice score and is defined as follows:

$$WDL = 1 - DS_{weighted} \tag{3}$$

3.3.2. Weighted cross-entropy loss

In combination of WDL, the Weighted Cross-Entropy (WCE) loss is used to train our model, which helps the model to converge faster. It is given as follows:

$$WCE = -\sum_{n\epsilon N} w_n g_n log(p_n) \tag{4}$$

where w_n is the weight for class n, g_n is the ground truth of class n and p is the predicted output corresponding to class n. The weights for each class are kept in similar way as in WDL i.e. 1, 5, 2, 4 for background, necrosis, edema and enhancing tumour region respectively.

3.3.3. Combined Loss

Finally, a Combined Loss (CL) function is obtained by adding the Weighted Cross-Entropy loss and the Weighted Dice Loss to train the proposed model.

$$CL = WDL + WCE \tag{5}$$

3.3.4. Guided Loss

The combined loss CL is used to train every decoder layers individually by generating predictions from each of them and comparing with the ground truth. The loss function of each layer is referred as Guided Loss (GL_x) for the x^{th} layer. The total Guided Loss function is calculated as follows:

$$GL_{overall} = \alpha GL_1 + \beta GL_2 + \gamma GL_3 + \delta GL_4 \tag{6}$$

where $\alpha, \beta, \gamma$and $\delta$are the weights for the guided loss for $1^{st}, 2^{nd}, 3^{rd}$ and 4^{th} layer respectively. As the 4th layer predicts the main output of the model, we have assigned higher weights to it and lesser weight to the losses of other layers. We carried out a series of experiments with different values and selected $\alpha = \beta = \gamma = 0.125$and $\delta = 0.5$. The experiment details are provided in the Tables 5 and 6 (Appendix).

3.4. Dataset

BraTS 2019 training dataset obtained from the BraTS official website is used to evaluate proposed architecture. The BraTS dataset is divided into 259 scans of HGG and 76 scans of LGG patients. The MRI scans of each patient consist of four sequences namely T1, contrast enhanced T1, T2 and FLAIR. The ground truth segmented images are divided into 4 classes with labels - 0: background and non-tumor region, 1: Necrosis, 2: Edema, and 4: enhancing tumor (ET). The dimensions of images are (240, 240, 155). The slices in four modalities along with ground truth label are shown in Fig. 4. Present study focuses on HGG patients only. By taking 25 slices from each modality of each of the 259 HGG patients, we had a total of 6475 images of each modality.

3.5. Data preprocessing

Each modality of the MRI scans is a 3D scan of the brain having size (240, 240, 155). From each volume (or 155 slices), the study selects 25 slices, and extracts slices with relatively more distinct information. In this view, we have chosen alternate slices from the first 50 slices near the center, as slices away from the center have less region of interest. The

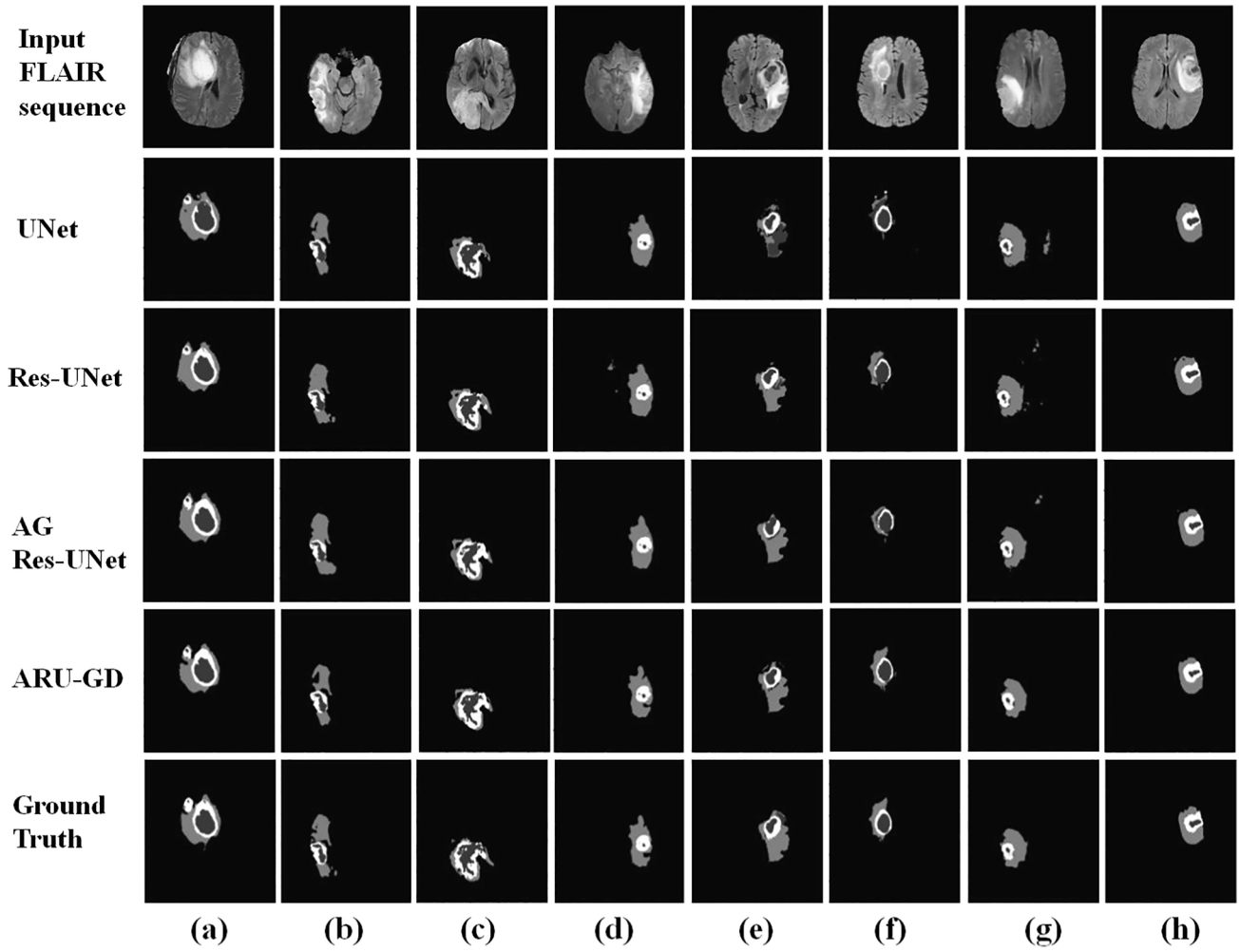

Fig. 5. Predictions of our model and the baseline models on randomly chosen eight unseen images from the test set; (row-wise) 1: flair images, 2: U-Net, 3: Res-UNet, 4: Res-UNet with AG, 5:Attention Res-UNet with Guided Decoder, 6: ground truth. White color in segmented image corresponds to enhancing tumor, dark grey corresponds to necrosis, and light grey corresponds to edema.

selection of alternate slices provides variations within the data taken from each patient and avoids over burdening of closely related data. For training purposes, we have taken a total of 6475 images for each modality from HGG patients. We have normalized the input data to ensures all inputs have a similar distribution. For this, Z-score is used, which transforms the data in such a way that the mean becomes 0 and standard deviation as 1. The Z-score equation implemented on each image slice is given as $I_n = \frac{I-\mu}{\sigma}$, where I_n is the normalized slice, I is the original slice, and μ and σ are the mean and standard deviation of input slice I respectively.

As the four modalities of the MRI scan (Flair, T1, T1CE and T2) are used together to detect different regions of the tumor, we have combined the slices from the four sequences in the form of channels. The size of input data becomes [6475, 240, 240, 4], where 6475 is the number of images, 240×240 is the size of each image and 4 represents the number of sequences. The data is then divided into a training set - 4700 images, validation set - 775 images, and testing set - 1000 images. In order to prevent overfitting, we have randomly performed data augmentations of horizontal and vertical flip, height and width shifts, rotations, zoom, and shear on our training set during training of our models.

3.6. Implementation details

The proposed model is trained on a batch size of 8 for 100 epochs. We have used the Adam optimizer with a learning rate of 0.0001, batch normalization to increase the stability of the network and to normalize our model at every layer. For model regularization, we have used L2 regularization of 0.00001. The model has been trained on the NVIDIA Tesla K80 GPU provided by Google Colab. We have used Keras with Tensorflow backend to implement our project.

4. Experimental results and discussion

This section elaborates on the experiments carried out to test and evaluates proposed model. The ablation study delineates the performance of our model against its individual counterparts. Finally, comparative study shows performance of proposed model against current state-of-the-art architectures.

4.1. Evaluation

The evaluation of the proposed model has been done on 1000 unseen test images, which are not used for training or validation. The brain tumor segmentation suffers hugely from class imbalance issues. The Dice Score (DS) coefficient takes care of the class imbalance problem and hence it has been used as the evaluation metric along with Intersection over Union (IoU) for all the experiments. For practical applications, the tumors are grouped into the following three regions: (i) Whole Tumor region, which includes necrosis, edema and enhancing tumor (label 1, 2, and 4), (ii) Tumor Core region includes necrosis and enhancing tumor

Table 1
Comparison of our proposed model and the baseline models on the BraTS 2019 dataset.

Method	Dice Score			Mean IoU		
	WT	TC	ET	WT	TC	ET
UNet	0.858	0.822	0.700	0.752	0.699	0.539
Res-UNet	0.871	0.803	0.725	0.772	0.672	0.570
AG Res-UNet	0.893	0.857	0.761	0.807	0.750	0.614
AG Res-UNet with Guided Decoder	**0.911**	**0.876**	**0.801**	**0.838**	**0.781**	**0.668**

(label 1 and 4), and (iii) Enhancing Tumor (ET) region (label 4). The proposed model has been evaluated against its individual counterparts and current state of the art architectures based on their dice scores and mean IoU for the whole tumor, tumor core and enhancing tumor regions.

4.2. Ablation study

The ablation study shows the utility of each module present in the proposed model. In this view, we have tested and compared proposed Attention Res-UNet with Guided Decoder (ARU-GD) model against its baseline models i.e. UNet, Res-UNet, and Res-UNet with attention gates (AG Res-UNet).

In ground truths and predicted segmentation images, the white region represents the enhancing tumors (ET), dark grey represents the necrosis and the light grey region corresponds to edema. Fig. 5 shows the comparison of predictions for Attention Res-UNet with Guided Decoder (ARU-GD) with respect to its baseline models i.e. UNet, Res-UNet and Res-UNet with attention gates (AG Res-UNet) on randomly chosen eight unseen images from the test set.

UNet has skip connections, which function as paths to share information between corresponding encoder and decoder layers. The U-Net architecture acts as the backbone for the proposed model and does not have specific highlights in the network hence producing a relatively less accurate segmented image. Further, the main concept behind residual blocks is adding an "identity shortcut connection" that jumps over multiple layers. This connection is used to add the input of the previous layer to the output of the current layer. The summation of these outputs is used as the input for the next layer. Residual blocks have been implemented over the UNet architecture to overcome the problem of degradation and saturation, and to extract more features at each layer. By adding the residual blocks to the U-Net, the Res-UNet is formed, which has improved the segmentation quality over the UNet model. The attention gate allows only activations from relevant regions thereby increasing the spatial accuracy at each layer. Hence, addition of attention gates over the Res-UNet improved segmentation accuracy in terms of dice score and mean IoU. The guided decoder is the main architectural highlight of the network. Its inculcation at the individual layers for better feature generation and utilization of the same in the overall loss function improves the learning process, which further improved the segmentation results. Hence, proposed ARU-GD yielded best segmentation of individual brain tumor regions i.e. whole tumor (WT), tumor core (TC) and enhancing tumor (ET) regions. Table 1 demonstrates the quantitative comparison of the proposed model with the baseline models. It is very much evident from the results that there is incremental improvement in the segmentation results as we are adding modules one by one (i.e. residual learning, attention gates and guided decoder) to the base U-Net model. The results show that proposed ARU-GD has achieved better segmentation accuracy than its baseline models in terms of dice coefficient and mean IoU for the WT, TC and ET regions. For statistical analysis of the results obtained in ablation study, we have considered 0.05 as a threshold for the p-value. The results show that proposed model has performed significantly better with respect to Unet and Res-Unet in the segmentation of all the three regions i.e. WT, TC and ET.

Table 2
Comparison of the proposed model on different loss functions.

Loss Variants	Dice Score			Mean IoU		
	WT	TC	ET	WT	TC	ET
ARU-GD + DL	0.897	0.835	0.795	0.814	0.717	0.661
ARU-GD + WCE	**0.915**	0.843	0.790	**0.845**	0.730	0.654
ARU-GD + (WCE + DL)	0.911	0.857	0.800	0.815	0.718	0.626
ARU-GD + (WCE + WDL)	0.911	**0.876**	**0.801**	0.838	**0.781**	**0.668**

Moreover, the proposed model ARU-GD performs significantly better in ET segmentation as compared to AG-ResUnet. However, even though our model yielded better Dice score than AG-ResUnet in WT and TC, the increase in performance is not that significant.

We have also trained our proposed model on multiple loss functions that include dice loss, weighted cross-entropy loss, combination of weighted cross entropy and dice loss and combination of weighted cross-entropy and weighted dice loss. It can be observed in Table 2 that the best results are obtained by using a combination of the weighted cross-entropy and weighted dice loss with highest dice scores and mean IoU for TC and ET. The values of dice score and mean IoU for combination of the weighted cross-entropy and weighted dice loss is very close to weighted cross-entropy loss for WT. These results suggested us to choose combination of weighted cross-entropy and weighted dice loss for the present study. Ablation study for loss was performed on proposed ARU-GD model.

4.3. Comparative study

The proposed architecture is compared with the state-of-the-art models used for semantic segmentation purposes. The qualitative comparison of predictions of proposed ARU-GD with the state-of-the-art models VGG-Net [25], MobileNet [12], QuickNAT [24], DenseNet [13] and Xception-Net [9] on randomly chosen unseen images from the test set is shown in Fig. 6. It is very much evident from the figure that the proposed model is able to segment these individual tumour regions most closely to the ground truth in comparison to other segmentation methods. The most difficult region to segment was the enhancing tumour (i.e. white region) and its segregation with necrosis (i.e. dark grey region), hence all segmentation models performed poorly in segmenting enhancing tumour and core tumour. On the other hand proposed ARU-GD model segmented these regions satisfactorily. Further, Table 3 shows the quantitative performance of proposed ARU-GD, which outperforms the state-of-the-art models for the segmentation of TC and ET groups in the MRI data. In case of WT, the proposed model performed at par with Xception-Net. Overall, it clearly indicates that the proposed model is more competitive to generate segmented images and results that are closest to the ground truth. The performance of Xception-Net is close to proposed model, which is followed by DenseNet, QuickNAT, MobileNet, and VGG-Net for all the individual regions i.e. WT, TC and ET.

The present study determines the difference in performance between proposed model and the other models. In this view, we have used Box-Whisker plot method of analysis as shown in Fig. 7, which indicates the distribution of the outputs of each model. It is evident from the plots that the median, quartile 1 and quartile 2 values of proposed method have an edge over the other methods.

Further, we have performed statistical analysis of comparative study. The proposed model has shown significantly better performance ($p < 0.05$) in TC region compared to the state-of-the-art models. In the segmentation of ET region, our model has shown p-values less than 0.05 compared to all the given models except the Xception-Net. Lastly, in spite of getting better Dice score in WT segmentation, the difference in performance is not significant except as compared to MobileNet.

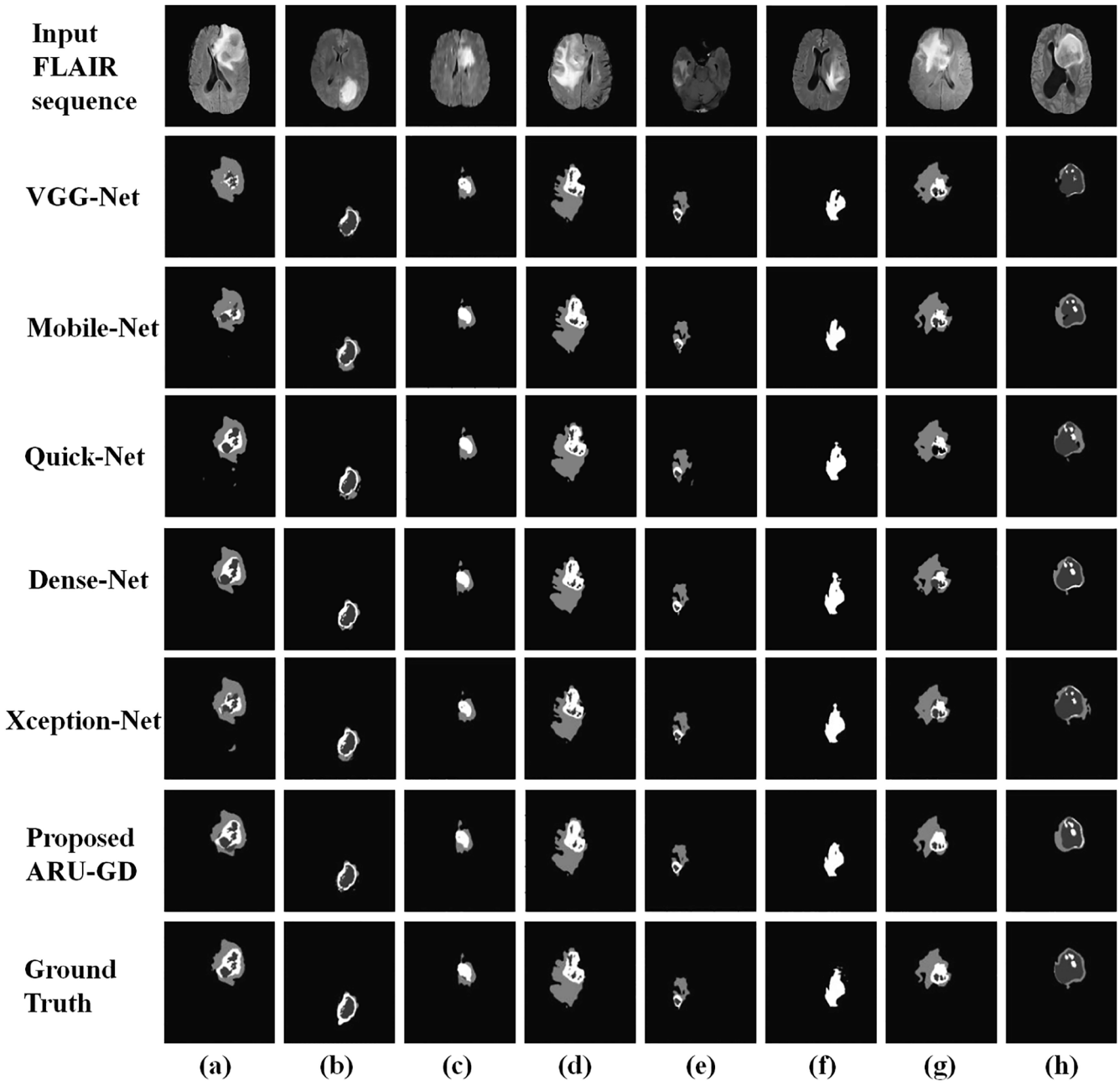

Fig. 6. Predictions of our proposed model and the state-of-the-art models on eight unseen images from test data; (row-wise) 1: FLAIR images, 2: VGG-Net, 3: MobileNet 4: QuickNAT, 5: DenseNet, 6: Xception-Net, 7:Attention Res-UNet with Guided Decoder. 8: ground truth. White color in segmented image corresponds to enhancing tumor, dark grey corresponds to necrosis, and light grey corresponds to edema.

Table 3
Comparison of our proposed model and the existing state-of-the-art models on the test images obtained from BraTS 2019 dataset.

Method	Dice Score			Mean IoU		
	WT	TC	ET	WT	TC	ET
VGG-Net [25]	0.879	0.742	0.658	0.784	0.590	0.490
MobileNet [12]	0.855	0.774	0.642	0.747	0.632	0.473
QuickNAT [24]	0.877	0.793	0.722	0.781	0.658	0.566
DenseNet [13]	0.907	0.841	0.725	0.831	0.727	0.570
Xception-Net [9]	**0.915**	0.837	0.779	**0.844**	0.720	0.638
AG Res-UNet with Guided Decoder	0.911	**0.876**	**0.801**	0.838	**0.781**	**0.668**

4.3.1. Comparison on HGG and LGG data

The comparative analysis in this section compares the performance of our proposed model against the current leaderboard models of BraTS 2019. Since the leaderboard models utilize both HGG and LGG data to train and test their performance, we have also performed a set of experiments to train and test our model on combined data of HGG and LGG patients. For this study, we used 20 slices per scan of each of the 335 patients (259 HGG and 76 LGG) which gave a total of 6700 images. We then divided it into Training set – 4380 images, Validation set – 1000 images and Testing set – 1320 images.

The proposed model obtained Dice Scores of 0.923, 0.847 and 0.834 on WT, TC and ET region respectively on testing set. Here, we have directly cited the results of the leaderboard models as mentioned in their respective research papers. As it is evident from Table 4, the proposed

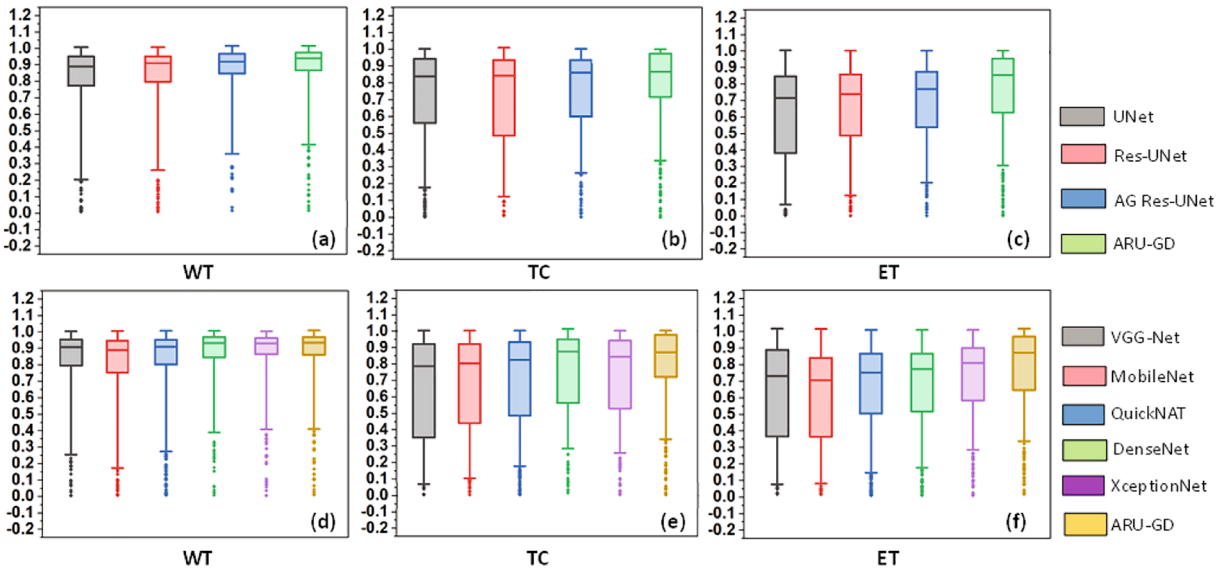

Fig. 7. Box-Whiskers plot for Ablation (a, b, c) and Comparative study (d, e, f) demonstrating the difference between the Dice scores.

Table 4
Comparison of proposed model and existing leaderboard models of BraTS 2019.

Method	Dice Score		
	WT	TC	ET
Jiang et al. [14]	0.887	0.836	0.832
Zhao et al. [34]	0.883	**0.861**	0.810
McKinley et al. [17]	0.890	0.830	0.810
Proposed ARU-GD	**0.923**	0.847	**0.834**

model performs better in comparison to the models of Jiang et al. [14], Zhao et al. [34] and McKinley et al. [17]. Our model has an edge in terms of segmentation of Whole Tumour region as compared to the others and provides at par performance in the segmentation of Tumor Core and Enhancing Tumour regions.

5. Conclusions

In this study, we have proposed deep learning architecture Attention Res-Unet with Guided Decoder (ARU-GD) for the semantic segmentation of brain tumors. The architecture has a new guided decoder, and inculcated attention gates. These changes in the base network Res-UNet improved the learning process by generating superior feature maps at the decoder and allowed only activations from relevant regions at the encoder side, and they worked together to increase the performance of segmentation. The proposed ARU-GD network is trained and tested on the BRATS 2019 dataset. The proposed ARU-GD has achieved Dice Scores of 0.911, 0.876 and 0.801 and mean IoU of 0.838, 0.781 and 0.668 on the whole tumor, tumor core and enhancing tumor segmentations respectively on unseen HGG test data. The present study has shown that our model outperforms its baseline models i.e. UNet, Res-UNet, and Res-UNet with attention gates. The experimental results have shown that the proposed architecture outperformed the current state of the art models such as VGG-Net, MobileNet, QuickNAT, DenseNet and Xception-Net. In future, 3D input of MRI data can be used for training as it contains more contextual and local information among nearby slices. The training of the proposed model is computationally heavy, hence work can be extended to improve memory requirements and computational time.

CRediT authorship contribution statement

Dhiraj Maji: Investigation, Formal Analysis, Methodology, Software, Data curation. **Prarthana Sigedar:** Writing - Original draft, Formal analysis, Methodology, Resources, Software. **Munendra Singh:** Conceptualization, Investigation, Validation, 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.

Appendix A

A.1. Study on selection of weight values of loss function

We have carried out a series of experiments with different values for α, β, γ and δ and based on those results, we chose the best combination. The experiment details are provided in the Tables 5 and 6.

Table 5
Dice scores obtained with different values for α, β, γ and δ.

$\alpha = \beta = \gamma$	δ	WT	TC	ET
0.2	0.4	0.892	0.810	0.771
0.167	0.5	0.916	0.824	0.789
0.125	0.5	0.911	0.876	0.801
0.13	0.6	0.907	0.839	0.783

Table 6
Mean IOU obtained with different values for α, β, γ and δ.

$\alpha = \beta = \gamma$	δ	WT	TC	ET
0.2	0.4	0.805	0.681	0.627
0.167	0.5	0.845	0.702	0.651
0.125	0.5	0.838	0.781	0.668
0.13	0.6	0.830	0.723	0.644

References

[1] S. Alqazzaz, X. Sun, X. Yang, L. Nokes, Automated brain tumor segmentation on multi-modal mr image using segnet, Computational Visual Media 5 (2019) 209–219.

[2] M. Angulakshmi, G. Lakshmi Priya, Automated brain tumour segmentation techniques–a review, International Journal of Imaging Systems and Technology 27 (2017) 66–77.

[3] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 2481–2495.

[4] D. Bahdanau, K. Cho, Y. Bengio, Neural machine translation by jointly learning to align and translate, 2014. arXiv preprint arXiv:1409.0473.

[5] 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, Scientific Data 4 (2017), 170117.

[6] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R.T. Shinohara, C. Berger, S.M. Ha, M. Rozycki, et al., 2018. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. arXiv preprint arXiv:1811.02629.

[7] L. Bi, J. Kim, A. Kumar, D. Feng, Automatic liver lesion detection using cascaded deep residual networks, 2017. arXiv preprint arXiv:1704.02703.

[8] S. Chakraborty, S. Chatterjee, A. Das, K. Mali, Penalized fuzzy c-means enabled hybrid region growing in segmenting medical images, in: Hybrid Machine Intelligence for Medical Image Analysis. Springer, 2020, pp. 41–65.

[9] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.

[10] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P. M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks, Medical Image Analysis 35 (2017) 18–31.

[11] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[12] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017. arXiv preprint arXiv:1704.04861.

[13] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[14] Z. Jiang, C. Ding, M. Liu, D. Tao, Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task, International MICCAI Brainlesion Workshop, Springer (2019) 231–241.

[15] A. Kermi, I. Mahmoudi, M.T. Khadir, Deep convolutional neural networks using u-net for automatic brain tumor segmentation in multimodal mri volumes, International MICCAI Brainlesion Workshop, Springer (2018) 37–48.

[16] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 3431–3440.

[17] R. McKinley, M. Rebsamen, R. Meier, R. Wiest, Triplanar ensemble of 3d-to-2d cnns with label-uncertainty for brain tumor segmentation, International MICCAI Brainlesion Workshop, Springer (2019) 379–387.

[18] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, et al., The multimodal brain tumor image segmentation benchmark (brats), IEEE Transactions on Medical Imaging 34 (2014) 1993–2024.

[19] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, D. Terzopoulos, Image segmentation using deep learning: A survey. arXiv preprint arXiv:2001.05566, 2020.

[20] M. Noori, A. Bahri, K. Mohammadi, Attention-guided version of 2d unet for automatic brain tumor segmentation, in: 2019 9th International Conference on Computer and Knowledge Engineering (ICCKE), IEEE, 2019, pp. 269–275.

[21] O. Oktay, J. Schlemper, L.L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, et al., 2018. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999.

[22] V. Rajinikanth, K.P. Thanaraj, S.C. Satapathy, S.L. Fernandes, N. Dey, Shannon's entropy and watershed algorithm based technique to inspect ischemic stroke wound, Smart Intelligent Computing and Applications, Springer (2019) 23–31.

[23] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer (2015) 234–241.

[24] A.G. Roy, S. Conjeti, N. Navab, C. Wachinger, A.D.N. Initiative, et al., Quicknat: A fully convolutional network for quick and accurate segmentation of neuroanatomy, NeuroImage 186 (2019) 713–727.

[25] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014. arXiv preprint arXiv:1409.1556.

[26] C. Singh, A. Bala, A transform-based fast fuzzy c-means approach for high brain mri segmentation accuracy, Applied Soft Computing 76 (2019) 156–173.

[27] M. Singh, V. Venkatesh, A. Verma, N. Sharma, Segmentation of mri data using multi-objective antlion based improved fuzzy c-means, Biocybernetics and Biomedical Engineering 40 (2020) 1250–1266.

[28] V. Sivakumar, N. Janakiraman, A novel method for segmenting brain tumor using modified watershed algorithm in mri image with fpga, Biosystems 198 (2020), 104226.

[29] V. Venkatesh, N. Sharma, M. Singh, Intensity inhomogeneity correction of mri images using inhomonet, Computerized Medical Imaging and Graphics 101748 (2020).

[30] A. Wadhwa, A. Bhardwaj, V.S. Verma, A review on brain tumor segmentation of mri images, Magnetic Resonance Imaging 61 (2019) 247–259.

[31] X. Xiao, S. Lian, Z. Luo, S. Li, Weighted res-unet for high-quality retina vessel segmentation, in: 2018 9th International Conference on Information Technology in Medicine and Education (ITME), IEEE, 2018, pp. 327–331.

[32] C. Yang, X. Guo, T. Wang, Y. Yang, N. Ji, D. Li, H. Lv, T. Ma, Automatic brain tumor segmentation method based on modified convolutional neural network, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2019, pp. 998–1001.

[33] J. Zhang, Z. Jiang, J. Dong, Y. Hou, B. Liu, Attention gate resu-net for automatic mri brain tumor segmentation, IEEE Access 8 (2020) 58533–58545.

[34] Y.X. Zhao, Y.M. Zhang, C.L. Liu, Bag of tricks for 3d mri brain tumor segmentation, International MICCAI Brainlesion Workshop, Springer (2019) 210–220.