Nested U-Net with Enhanced Attention Gate and Compound Loss for Semantic Segmentation of Brain Tumor from Multimodal MRI

Ceena Mathews[1*]   Anuj Mohamed[2]

[1]*Prajyoti Niketan College, India*
[2]*School of Computer Sciences, Mahatma Gandhi University, India*
Corresponding author's Email: ceenamathews@prajyotiniketan.edu.in

Abstract: The maximum resection of small tumor structures such as enhancing and non-enhancing tumor increases the overall survival rate of patients in all age groups. Therefore, it is necessary to precisely detect and delineate the tumor sub-regions such as necrosis, enhancing tumor and non-enhancing tumor in brain MR images for effective prognosis of tumor. To achieve this objective, in this study, we implement semantic segmentation using nested U-Net with an enhanced attention gate to classify, localize and segment the tumor sub-regions effectively. Additionally, we propose a compound loss function to handle the problem of class imbalance seen dominantly in brain MR images. We trained and tested the proposed model using two multimodal MRI benchmark datasets, BraTS 2019 and BraTS 2020. We achieved dice scores of 0.91, 0.92 and 0.93 with BraTS 2019 and dice scores of 0.91, 0.92 and 0.92 with BraTS 2020 dataset for the whole tumor, enhancing tumor and tumor core respectively. Experimental results on both datasets showed that our proposed model outperforms state-of-the-art 2D brain tumor segmentation frameworks using deep learning in the prediction of small tumor structures such as enhancing tumor, necrosis and non-enhancing tumor.

Keywords: Attention gate, Compound loss, Focal tversky loss, Nested U-Net, Weighted binary cross-entropy.

1. Introduction

A brain tumor is one of the lethal and devastating diseases. Patients with highly malignant tumors have only median survival despite the best available treatment procedures such as surgery, radiation therapy and chemotherapy [1]. According to the international agency for research on cancer (IARC), an agency of the world health organization (WHO), the worldwide incidence rate and mortality rate of brain and central nervous system (CNS) cancer, in 2018 for both males and females of all ages is 2,96,851 and 2,41,037 respectively. It also estimates that the number of incident cases of brain and CNS cancer for both sexes of all ages worldwide, from 2018 to 2040, is 4,35,554 [2].

The type of brain tumor that originates in the brain's glial cells is called gliomas. Based on the intensity of its spread, gliomas are classified as high-grade glioma (HGG) and low-grade glioma (LGG). Glioma cells are divided into different regions such as peritumoral edema (ED), necrotic core (NCR), enhancing tumor (ET) and non-enhancing tumor (NET). The life expectancy of patients with HGG is very less with a median survival rate of 12 to 14 months since diagnosis, despite advances in treatment procedures such as microsurgical resection, radiation, and chemotherapy [3]. The overall survival rate of patients in all age groups can be increased through maximum resection of ET and NET sub-regions of a brain tumor [4]. Therefore, it would be desirable to non-invasively diagnose the extent of the brain tumor meticulously before doing any invasive procedure.

Magnetic resonance imaging (MRI) is the best non-invasive method used for the differential diagnosis of brain tumors. Normally, tumors are manually segmented from brain MRI by radiologists. The efficacy of manual segmentation depends on the perception of the expertise. Such subjective errors can be reduced by using automatic segmentation. However, precise automatic segmentation of tumor sub-regions from MRI becomes a complicated task

mainly due to two reasons: (1) gliomas are heterogeneous and infiltrative (2) In brain MR images, there are more non-tumor regions pixels than the tumor pixels, resulting in greater class imbalance.

Semantic segmentation is a computer vision task where similar parts of the image with the same class are grouped. It involves both localization and segmentation. Presently, deep learning architectures with encoder-decoder structures are effective in the process of semantic segmentation. Encoder-decoder architectures such as U-Net [5] and its variants such as 3D U-Net [6], Attention U-Net [7], V-Net [8], conditional generative adversarial networks (cGAN) [9], nested U-Net (U-Net++) [10] are more prevalently used in the medical image segmentation.

However, the potential issue of the encoder-decoder architecture is that the skip connections in the encoder-decoder architecture concatenate the inadequate contextual information in the high-level features extracted at the beginning of the encoder path, with the corresponding feature map at the end of the decoder path. This results in poor performance of such networks [11]. To address this issue, attention gates proposed by Schlemper et al. are integrated with encoder-decoder architectures, since they automatically learn to focus on the structures at the region of interest without the need for additional supervision.

The attention gate was first integrated into the U-net model by Oktay et al. to realize spatial localization and subsequent segmentation. It was integrated into nested U-Net (ANU-Net) by Cheng Li et al. to improve the segmentation results [12].

Class imbalance is a major issue in medical image segmentation. Various studies have been performed to compare optimization achieved by different loss functions to deal with the class imbalance. Sanchez – Peralta et al. compared seven loss functions on the CVC-EndoScenestill dataset and observed that the region-based losses give better performance than the cross-entropy loss [13]. Jadon conducted a comparison of fifteen losses using the NBFS skull-stripped dataset and found that the region-based losses such as focal Tversky loss and Tversky loss outperformed the other loss functions [14]. Ma et al. compared twenty loss functions using four datasets for liver, liver tumor, pancreas and multi-organ segmentation and concluded that the compound loss functions perform better than the region-based and distribution-based loss functions [15]. The literature shows that the region-based or compound loss functions perform more consistently than the distribution-based loss functions.

The motivation of our work is to predict with greater accuracy all the tumor sub-regions from brain MR images for the effective prognosis of the disease. To achieve this objective, in this study, we implement semantic segmentation using encoder-decoder architecture, to classify, localize and segment the tumor sub-regions effectively. Further, we enhanced the structure of the attention gate so that it aids in the deep selection of the hierarchically extracted relevant features. We adopted a nested U-Net without deep supervision as the base model for evaluating the performance of the proposed attention gate. The proposed attention gate is integrated between all the convolution blocks in the nested U-Net model. Additionally, we propose a compound loss function by combining focal Tversky loss and weighted binary cross-entropy loss functions to handle the class imbalance issue common in the brain MRI dataset.

The proposed model was trained and evaluated on brain tumor segmentation (BraTS) 2019 and 2020 challenge datasets using metrics dice similarity coefficient (DSC), hausdorff distance (HD), sensitivity and specificity. The dataset providers conventionally use three sub-regions for evaluating the brain tumor segmentation such as

1) tumor core (TC) where TC contains ET, NET, and NCR
2) whole tumor (WT) where WT comprises TC and ED
3) ET

The paper focuses on the semantic segmentation of small tumor structures such as NCR, enhancing and non-enhancing tumors using Nested U-Net with enhanced attention gate and compound loss function. Related works are described in section 2 of the paper. Datasets, proposed network architecture, and proposed compound loss function used for the effective segmentation of the tumor sub-regions are discussed in section 3. Results and discussion of the experiments are given in section 4 and the conclusion in section 5.

2. Related works

Deep learning algorithms are gaining popularity with state-of-the-art results in brain tumor segmentation. Convolutional neural network (CNN) is the popular deep learning method. This section discusses some of the recent studies based on 2D convolutional neural networks using BraTS 2019 and BraTS 2020 datasets for brain tumor segmentation.

Shi et al. [16] developed a multiple feature extraction network based on dense channels 2D U-Net. The network was trained and evaluated on the BraTS 2019 dataset. Hamghalam, Baiying and Wang [17] proposed a framework to convert 3D patches
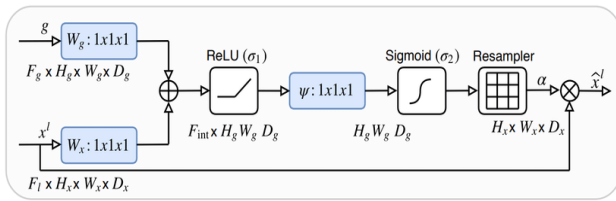
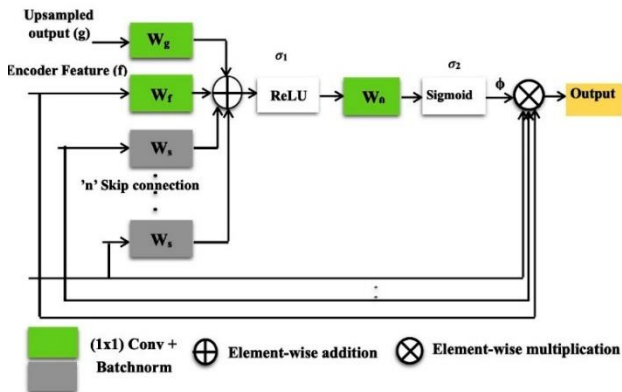Figure. 1 Structure of attention gate proposed by Schlemper et al. [22]

Figure. 2 Enhanced attention gate

into 2D images which are then fed into classifier blocks with 2D kernels for the prediction. Agravat and Raval [18] use three-layer deep U-Net based encoder-decoder architecture for semantic segmentation. Each layer of the encoding side includes dense modules and the decoding side uses convolution modules. The network was trained on the BraTS 2019 dataset using soft dice loss and focal loss function. Wang et al. [19] proposed a modality pairing network to segment brain tumor substructures. The proposed network consists of paralleled branches using different modalities as input. The model was trained on the BraTS 2020 dataset using modality pairing loss.

The above studies in the literature show that the DSC scores for the regions ET and TC are less, reducing the effect of segmentation. This would adversely affect the prognosis of the disease. In the proposed study, in order to increase the predictive accuracy of small tumor structures such as ET, NET and NCR, enhanced attention gates are used with nested 2D U-Net to focus more on the hierarchically extracted relevant features of the structures at the region of interest. Further, a novel compound loss function is used to deal with the class imbalance issue seen in brain MRI.

3. Materials and methods

This section describes the dataset used for experimentation, the architecture of the nested U-Net

model with the enhanced attention gate and the compound loss function.

3.1 BraTS dataset

This work uses 3D MR images of 150 HGG patients from the BraTS 2019 training dataset and 150 MR images from the BraTS 2020 training dataset. Each patient case has four MRI sequences such as T1-weighted (T1), T1-weighted with gadolinium contrast (T1Gd), T2-weighted (T2), fluid-attenuated inversion recovery (FLAIR) and ground truth. The ground truths in these datasets are manually segmented and annotated by the experts as background (label 0), NET (NCR/NET) (label 1), ED (label 2) and ET (label 4). Label 3 is not used by the experts. The images are co-registered, skull stripped and re-sampled to 1 mm^3 [20, 21].

3.2 Network architecture

The network architecture proposed in our work is based on the attention gate and nested U-Net model. The structure of the attention gate was enhanced and incorporated between the convolution blocks of the nested U-Net model. The model is trained and tested using a novel compound loss function.

Enhanced attention gate: In the original attention gate shown in Fig. 1 proposed by Schlemper et al., only two inputs are used (1) the up-sampling feature in the decoder also known as gating signal (g) and (2) the corresponding depth feature of the encoder.

In this work, we redesigned the gate to accept feature maps from the previous nodes in the same skip connection pathway along the same depth as input, in addition to the inputs used with the original attention gate. The hypothesis is that the addition of the feature maps of the skip connections to the attention gate greatly optimizes the gradient flow and helps to reduce the semantic gap between the encoder and decoder feature map. The structure of the proposed attention gate is shown in Fig. 2.

The inputs to the attention gate are as follows :
  (1) up-sampled gating signal (g)
  (2) encoder feature map (f) and
  (3) feature map s[i] of the skip connections at same level, where (i = 1…n) and n indicates the number of nodes before the encoder.

The attention gate between nodes $X^{0,2}$ and $X^{0,3}$ in Fig. 3 accepts as input the feature maps of $X^{0,0}$ and $X^{0,1}$, up-sampled output from $X^{1,2}$ and feature map of the encoder $X^{0,2}$. The gating signal is used to select more essential features from the encoded feature
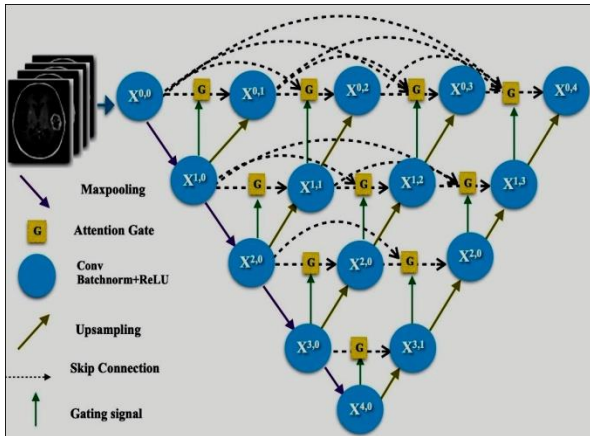
Figure. 3 Nested U-Net with enhanced attention gate

maps and forwards them to the next decoder. Convolution operation (W) and batch normalization (b) are performed on each of these inputs and are then added together pixel by pixel. The combined output is then activated using rectified linear unit (ReLU) σ_1 as defined in Eq. (1).

$$\sigma_1 = \max(0, x) \qquad (1)$$

This is followed by convolution operation (W_0), batch normalization (b_0) and sigmoid activation σ_2 as defined in Eq. (2).

$$\sigma_2 = \frac{1}{1 + e^{-x}} \qquad (2)$$

This enables the network to learn to predict values in the range [0,1] and it is used as the attention coefficient ϕ. The attention coefficient ϕ, feature maps of the encoder and that of the skip connections are multiplied element-wise to generate the output as shown in Eq. (3).

$$F = \sigma_1(a + (W_f \times f + b_f) + (W_g \times g + b_g))$$
$$\text{where } a = \sum_{i=1}^{n}(W_{s[i]} \times s[i] + b_{s[i]})$$
$$\phi = \sigma_2[W_0 \times F + b_0]$$
$$Output = \prod_{i=1}^{n} s[i] \times f \times \phi \qquad (3)$$

Nested U-Net with enhanced attention gate: In this study, we adopted nested U-Net without deep supervision as the backbone for our network architecture. We inserted the redesigned attention gate between each of the convolution blocks in the nested U-Net model as shown in Fig. 3.

The nested U-Net model consists of encoder and decoder parts. The encoder captures the context information and passes it to the decoder of the corresponding convolution block through the attention gate so that more precise and relevant

location features of the foreground objects are extracted. The input to each convolJition block in the decoder part is the concatenated output of two feature maps (1) output from the attention gate before the decoder and (2) the up-sampled feature map from the lower deconvolution block.

3.3 Compound loss function

The classes defined for the tumor sub-regions are less in volume than the background classes representing the healthy brain tissues. This results in the average prediction of tumor sub-regions. From the literature, it is realized that the region-based loss functions and compound loss functions work effectively with class imbalance dataset. Hence in our work, we propose a compound loss function that combines focal Tversky (region-based loss function) and weighted binary cross-entropy (distribution-based loss function) loss functions.

Weighted binary cross-entropy defined in Eq. (4), assigns different weights to different classes, enabling us to distinguish regions of different classes and learn significant patterns in the image.

$$Loss_{wbce} = \frac{-(T \times \log(P) \times w + (1 - T) \times \log(1 - P))}{N}$$

$$(4)$$

where T indicates ground truth values, P indicates predicted values, N indicates the number of samples and w is a hyper parameter that enables a tradeoff between false positives and false negatives. In order to reduce the number of false negatives, w > 1 is used and to reduce the number of false positives, w < 1 is used.

Focal Tversky loss is an extension of Tversky loss with the hyper parameter γ. It helps to focus well on the hard classes within the region of interest and predict those classes with greater accuracy. The value of the hyper parameter is chosen in such a way that there is a balance between precision and recall. The focus on hard classes can be increased with γ < 1. The focal Tversky loss is given in Eq. (5).

$$Loss_{ft} = (1 - TI)^{\gamma} \qquad (5)$$

where $Loss_{ft}$ represents focal Tversky loss, γ is a hyper parameter whose value may range between 0 and 1 and TI represents Tversky index as defined in Eq. (6) and (1-TI) represents Tversky loss.

$$TI =$$
$$\frac{\sum_{i=1}^{N} p_{ic}g_{ic} + \varepsilon}{\sum_{i=1}^{N} p_{ic}g_{ic} + \alpha\sum_{i=1}^{N} p_{ic}g_{i\bar{c}} + \beta\sum_{i=1}^{N} p_{i\bar{c}}g_{ic} + \varepsilon} \tag{6}$$

where p_{ic} is the predicted value of the pixel i, of tumor class c and $p_{i\bar{c}}$ is the predicted value of pixel i, of the non-tumor class $\bar{c}$, g_{ic} and $g_{i\bar{c}}$ represent the ground-truth value of the pixel i of the tumor class c and non-tumor class $\bar{c}$, respectively. Hyper parameters α and β enable a trade-off between false positives and false negatives to improve recall in the case of large class imbalance. The values of α and β may range between 0 and 1, such that α + β = 1. ε is a small constant to avoid division by zero error [23].

The proposed compound loss defined in Eq. (7) is the combination of weighted binary cross-entropy and focal Tversky loss functions.

$$Loss = Loss_{wbce} + Loss_{ft} \tag{7}$$

3.4 Evaluation metrics

We tested the performance of the model using evaluation metrics such as DSC, sensitivity, specificity and HD [24, 25]. DSC measures the amount of overlap between the predicted mask and ground truth labels as indicated in Eq. (8). Sensitivity measures the rate of true positives considering the positives in both ground truth and predicted mask and is calculated using Eq. (9). Specificity measures the true negative rate considering the negatives in both the ground truth mask and the predicted mask as mentioned in Eq. (10). HD computes the distance between the set of non-zero pixels of two images according to Eq. (11). HD is one of the most informative and useful criteria because it is an indicator of the largest segmentation error [26].

$$DSC = \frac{2 \times |T \cap P|}{|T| + |P|} \tag{8}$$

where T represents ground truth pixels and P defines predicted pixels.

$$Sensitivity = \frac{TP}{TP + FN} \tag{9}$$

$$Specificity = \frac{TN}{TN + FP} \tag{10}$$

where TP indicates the number of true positives, FN indicates the number of false negatives, TN indicates the number of true negatives and FP represents false positives.

$$HD(T, P) = \max\big(h(T, P), h(P, T)\big) \tag{11}$$

where

$$h(T, P) = \max_{t \in T} \min_{p \in P} \|t - p\|$$

and $\|t - p\|$ is the Euclidean distance on the points t and p, T represents the ground truth values and P represents the predicted values.

4. Results and discussion

In this study, we investigated the performance of the proposed model using two 3D multimodal MRI datasets of the BraTS challenge, BraTS 2019 and BraTS 2020. 150 MR images of HGG patients from the BraTS 2019 and 150 MR images from the BraTS 2020 dataset were used for training and testing. In both datasets, each patient case has four MRI modalities such as T1, T2, T1Gd and FLAIR. Each image has a dimension of 240x240x155 where 240x240 indicates the height and width of a slice and 155 specifies the number of slices (depth). Each modality is significant in identifying different tumor regions. In T1Gd, the area that has ET appears brighter and TC appears darker, whereas WT appears brighter in FLAIR images. So, in this work, we combine all four sequences of MRI for the effective segmentation of the various sub-regions of the tumor.

Contextual information about the tumor structures can be better observed in 3D volumetric MR images. However, due to the computational and memory limitations, we extracted only the middle ninety 2D slices from each modality since the remaining slices do not convey much information. These extracted slices from each modality are then merged to form 2D images. Additionally, each image is cropped to a size of 192x192. Hence, the dimension of the input image becomes 192x192x4. The dataset is split in the ratio 6:2:2 for training, validation and testing. The data is then preprocessed using a simple normalization technique to scale the pixel values in the range of [0,1].

The proposed model was trained on both BraTS 2019 and BraTS 2020 datasets separately using an Adam optimizer with a learning rate of 3e-5. The network is trained for 80 epochs with a batch size of 16.

253

Table 1. Performance metrics of the proposed model using different values of the hyper parameters α, β, γ and w, on HGG cases of BraTS 2019 dataset

Proposed Model				DSC			Sensitivity			Specificity		
w	α	β	γ	WT	ET	TC	WT	ET	TC	WT	ET	TC
2	0.3	0.7	0.75	.8985	.9131	.9303	.9261	.8802	.9318	**.9993**	.9995	.9997
2	0.3	0.7	0.9	.8974	.9137	.9274	.9263	.8801	.9280	**.9993**	**.9996**	**.9998**
2	0.7	0.3	0.75	.9065	.9165	.9335	**.9650**	.9162	.9602	.9985	.9994	.9996
2	0.7	0.3	0.9	.8989	.9111	.9299	.9619	.9118	**.9617**	.9984	.9994	.9995
4	0.3	0.7	0.75	.8975	**.9170**	.9306	.9318	.8997	.9331	.9992	.9995	**.9998**
4	0.3	0.7	0.9	.8996	.9141	.9284	.9288	.8924	.9364	**.9993**	**.9996**	.9997
4	**0.7**	**0.3**	**0.75**	**.9125**	**.9170**	**.9341**	.9594	**.9517**	.9582	.9988	.9994	.9996
4	0.7	0.3	0.9	.8841	.9108	.9324	.9565	.9015	.9549	.9988	.9995	.9996

Table 2. Performance comparison of the proposed model with state-of-the-art 2D frameworks on BraTS 2019 dataset

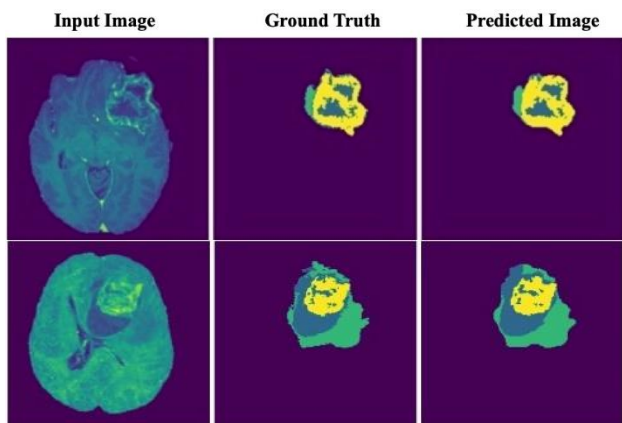| Method	DSC		
	WT	ET	TC
W.Shi et al. [16]	0.9019	0.8266	0.8394
M. Hamghalam et al. [17]	0.89	0.74	0.80
R.R Agravat and M. S Raval [18]	0.89	0.74	0.85
Proposed	0.9125	0.9170	0.9341

Figure. 4 Example of segmentation results on HGG cases of BraTS 2019 dataset. In the ground truth and predicted images, green area shows edema, blue shows ET and yellow shows NET and NCR sub-regions

4.1 Performance on BraTS 2019 dataset

We experimented with different values for the hyper parameters α, β, γ and w, on the BraTS 2019 dataset. In the experiment, since the classes representing tumor structures are fewer than the healthy tissues, we use w > 1 which catalyzes the decrease in the number of false negatives. Table 1 shows the DSC, sensitivity and specificity scores for WT, ET and TC sub-regions of the tumor using HGG cases in the BraTS 2019 dataset for the different values of the hyper parameters α, β, γ and w. It is observed that the model gives comparable performance measures for the sub-regions such as ET and TC (ET+NCR+NET) with all values of hyper parameters. Since the model gives optimum results for the values w = 4, α = 0.7, β = 0.3 and γ = 0.75, we used the evaluation metrics and segmentation results using these values for comparison against other models.

Fig. 4 shows examples of the segmentation results on two patient cases in the BraTS 2019 dataset. Fig. 5 (a) and (b) show the box plots using DSC and HD metrics respectively for the tumor sub-regions WT, TC and ET. The performance of the proposed model is compared against the state-of-the-art 2D segmentation frameworks based on BraTS training dataset since our study utilizes only 2D slices from 3D MRI BraTS dataset. Table 2 compares the DSC scores of the proposed model with the state-of-the-art 2D segmentation frameworks on the BraTS 2019 dataset.

4.2 Performance on BraTS 2020 dataset

We also experimented with different values for the hyper parameters α, β, γ and w, on 150 cases in the BraTS 2020 dataset. Table 3 shows the DSC, sensitivity and specificity scores for WT, ET and TC using the BraTS 2020 dataset for the different values of the hyper parameters α, β, γ and w. It is observed that the model gives comparable results for ET and TC with all hyper parameter values.

Although the predictive accuracy of the smaller tumor structures is high, the accuracy of WT is low since the DSC score of ED is less. It is evident from the results that the model performs better than the state-of-the-art 2D deep learning-based brain tumor segmentation methods for the tumor sub-regions ET,

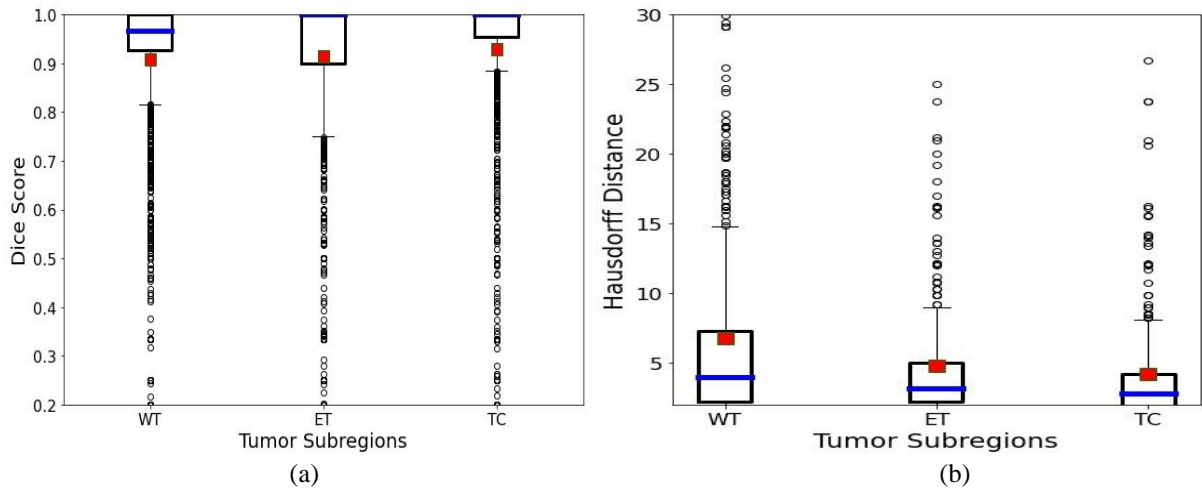(a)   (b)

Figure. 5 Boxplots of the proposed model for WT, ET and TC sub-regions of brain tumors on the BraTS 2019 dataset
for: (a) DSC and (b) HD metrics

Table 3. Evaluation metrics for the proposed model using different values of the hyper parameters α, β, γ and w, on
BraTS 2020

| Proposed Model				DSC			Sensitivity			Specificity		
w	α	β	γ	WT	ET	TC	WT	ET	TC	WT	ET	TC
2	0.3	0.7	0.75	.8881	.9220	.9211	.9244	.8814	.9211	.9992	**.9997**	**.9996**
2	0.3	0.7	0.9	.9121	**.9349**	**.9334**	.9353	.9121	.9318	.9994	**.9997**	**.9996**
2	0.7	0.3	0.75	.9004	.9254	.9271	.9512	.9162	.9502	.9985	.9995	.9993
2	0.7	0.3	0.9	.9137	.9276	.9327	**.9656**	.9104	**.9638**	.9986	.9996	.9992
4	0.3	0.7	0.75	.8944	.9179	.9193	.9188	.8656	.9162	.9993	**.9997**	**.9996**
4	0.3	0.7	0.9	.9004	.9275	.9261	.9271	.9008	.9349	**.9995**	**.9997**	**.9996**
4	0.7	0.3	0.75	**.9137**	.9329	.9330	.9578	**.9385**	.9575	.9989	.9996	.9993
4	0.7	0.3	0.9	.9039	.9213	.9246	.9548	.9221	.9555	.9984	.9992	.9994
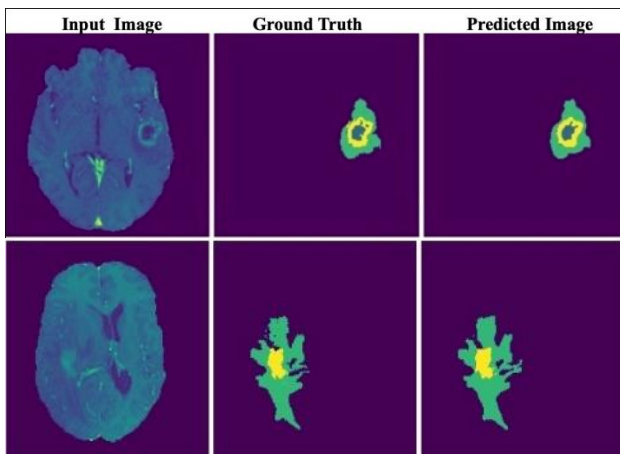
Figure. 6 Segmentation results on BraTS 2020 dataset

Table 4. Performance comparison of the proposed model
with state-of-the-art 2D segmentation framework on
BraTS 2020 dataset

Method	DSC		
	WT	ET	TC
Y.Wang et al. [28]	**.924**	.863	.898
Proposed	.9137	**.9213**	**.9246**

Table 5. HD metrics on BraTS 2019 and BraTS 2020
dataset for w=4, α=0.7,β=0.3 and γ=0.75.

Dataset	Hausdorff Distance		
	WT	ET	TC
BraTS 2019	6.4	4.6	4
BraTS 2020	6.5	4.9	4.6

NET and NCR. The segmentation results of two
patients in the BraTS 2020 dataset using the model is
shown in Fig. 6. Table 4 compares the DSC scores of
the proposed model with the state-of-the-art 2D
segmentation frameworks on the BraTS 2020 dataset.
Table 5 reports the HD metrics on both datasets using
values w = 4, α = 0.7, β = 0.3 and γ = 0.75.

From the Tables 2 and 4, it is evident that the
proposed model applied on both datasets gives better
results for the tumor sub-regions ET and TC, since
the enhanced attention gates increases the learning of
the small tumor sub-regions effectively. Attention
gates also accelerate the efficacy of the dense skip
connections. The addition of the feature maps of the
skip connections at the same depth, as input to the

attention gate greatly optimizes the gradient flow and helps to reduce the semantic gap between the encoder and decoder feature map.

5. Conclusion

In this work, we enhanced the structure of the attention gate and proposed a novel compound loss function for effective segmentation of all tumor sub-regions such as WT, ET and TC. Further, the model helps to work out class imbalance problems seen in brain MR images. The proposed compound loss function is the combination of weighted binary cross-entropy and focal Tversky loss functions. The enhanced gate is incorporated into nested U-Net architecture. The proposed model was trained and evaluated using two multimodal MRI BraTS datasets, BraTS 2019 and BraTS 2020. We achieved dice scores of 0.91, 0.92 and 0.93 with BraTS 2019 and dice scores of 0.91, 0.92 and 0.92 with the BraTS 2020 dataset for the whole tumor, enhancing tumor and tumor core respectively. Compared to state-of-the-art 2D segmentation methods using deep learning, our proposed method gave superior results for sub-regions such as ET and TC. The model accomplished better performance due to the deep selection of relevant information from the hierarchically extracted features along the skip connection pathway using an enhanced attention gate and since the compound loss function focuses more on the relevant foreground pixels. Utilizing increased computational power and high memory, the accuracy of the model can be improved further by training more samples. The prediction accuracy can be further improved by using multimodal 3D volumetric images, instead of the 2D slices extracted from the 3D images, since volumetric images convey more spatial information about the tumor structures.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

The conceptualization, methodology, resources, implementation, result analysis, original draft preparation, draft editing and visualization has been done by first author. The supervision, review of work and project administration has been done by second author.

References

[1] C. Horbinski, K. L. Ligon, P. Brastianos, J. T. Huse, M. Venere, S. Chang, J. Buckner, T. Cloughesy, R. B. Jenkins, C. Giannini, L. B. Nabors, P. Y. Wen, K. J. Aldape, R. V. Lukas, E. Galanis, C. G. Eberhart, D. J. Brat, J. N. Sarkaria, "The medical necessity of advanced molecular testing in the diagnosis and treatment of brain tumor patients", *Neuro-oncology*, Vol. 21, No. 12, pp. 1498-1508, 2019.

[2] J. Ferlay, M. Ervik, F. Lam, M. Colombet, L. Mery, M. Pineros, A. Znaor, I. Soerjomataram, and F. Bray, "Global cancer observatory: cancer today", Lyon, France: *International Agency for Research on Cancer*, pp. 1-6, 2018.

[3] Q. T. Ostrom, H. Gittleman, J. Fulop, M. Liu, R. Blanda, C. Kromer, Y. Wolinsky, C. Kruchko, and J. S. B. Sloan, "CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012", *Neuro-Oncology*, Vol. 17, Suppl. 4, pp. iv1- 62, 2015.

[4] A. M. Molinaro, S. H. Jumper, R. A. Morshed, J. Young, S. J. Han, P. Chunduru, Y. Zhang, J. J. Phillips, A. Shai, M. Lafontaine, and J. Crane, "Association of maximal extent of resection of contrast-enhanced and non–contrast-enhanced tumor with survival within molecular subgroups of patients with newly diagnosed glioblastoma", *JAMA Oncology*, Vol. 6, No. 4, pp. 495-503, 2020.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation", In: *Proc. of International Conf. on Medical Image Computing and Computer-Assisted Intervention,* pp. 234-241, 2015.

[6] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation", In: *Proc. of International Conference on Medical Image Computing and Computer-Assisted Intervention,* pp. 424-432, 2016.

[7] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, and B. Glocker, "Attention u-net: Learning where to look for the pancreas", *arXiv Preprint* arXiv:1804.03999, 2018.

[8] F. Milletari, N. Navab, and S. A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation", In: *Proc. of Fourth International Conf. on 3D Vision (3DV)*, pp. 565-571, 2016.

[9] B. Yu, L. Zhou, L. Wang, J. Fripp, and P. Bourgeat, "3D cGAN based cross-modality MR image synthesis for brain tumor segmentation", In: *Proc. of IEEE 15th International Symposium*

on Biomedical Imaging (ISBI 2018), pp. 626-630, 2018.

[10] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation", *Deep Learning in medical image analysis and multimodal learning for clinical decision support. Lecture Notes in Computer Science,* Vol.11045, pp. 3-11, 2018.

[11] T. L. B. Khanh, D. P. Dao, N. H. Ho, H. J. Yang, E. T. Baek, G. Lee, S. H. Kim, and S. B. Yoo, "Enhancing U-Net with spatial-channel attention gate for abnormal tissue segmentation in medical imaging", *Applied Sciences*, Vol. 10, No. 17, p. 5729, 2020.

[12] C. Li, Y. Tan, W. Chen, X. Luo, Y. He, Y. Gao, and F. Li, "ANU-Net: Attention-based Nested U-Net to exploit full resolution features for medical image segmentation", *Computers & Graphics*, Vol. 90, pp. 11-20, 2020.

[13] L. F. S. Peralta, A. Picon, J. A. A. Barroso, J. F. O. Moran, F. M. S. Margallo, and J. B. Pagador, "Eigenloss: combined PCA-based loss function for polyp segmentation", *Mathematics*,Vol. 8, No. 8, p. 1316, 2020.

[14] S. Jadon, "A survey of loss functions for semantic segmentation", In: *Proc. of 2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE*, pp. 1-7, 2020.

[15] J. Ma, J. Chen, M. Ng, R. Huang, Y. Li, C. Li, X. Yang, and A. L. Martel, "Loss odyssey in medical image segmentation", *Medical Image Analysis*, Vol. 71, p. 102035, 2021.

[16] W. Shi, E. Pang, Q. Wu, and F. Lin, "Brain tumor segmentation using dense channels 2D U-net and multiple feature extraction network", In: *Proc. of International MICCAI Brainlesion Workshop, Springer, Cham*, pp. 273-283, 2019.

[17] M. Hamghalam, B. Lei, and T. Wang, "Convolutional 3D to 2D patch conversion for pixel-wise glioma segmentation in MRI scans", In: *Proc. of International MICCAI Brainlesion Workshop*, pp. 3-12, 2019.

[18] R. R. Agravat, and M. S. Raval, "Brain tumor segmentation and survival prediction", In: *Proc. of International MICCAI Brainlesion Workshop*, pp. 338-348, 2019.

[19] Y. Wang, Y. Zhang, F. Hou, Y. Liu, J. Tian, C. Zhong, Y. Zhang, and Z. He, "Modality-pairing learning for brain tumor segmentation", In: *Proc. of International MICCAI Brainlesion Workshop*, pp. 230-240, 2020.

[20] B. H. Menze, A. Jakab, S. Bauer, J. K. Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, and L. Lanczi, "The multimodal brain tumor image segmentation benchmark (BRATS)", *IEEE Transactions on Medical Imaging*, Vol. 34, No. 10, pp. 1993-2024, 2014.

[21] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features", *Scientific Data*, Vol. 4, No. 1, pp. 1-13, 2017.

[22] J. Schlemper, O. Oktay, M.Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images", *Medical Image Analysis*, Vol. 53, pp. 197-207, 2019.

[23] M. Yeung, E. Sala, C. B. Schonlieb, and L. Rundo, "Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation", *Computerized Medical Imaging and Graphics*, Vol. 95, p. 102026, 2022.

[24] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, pp. 850-863,1993.

[25] W. R. Crum, O. Camara, and D. L. Hill, "Generalized overlap measures for evaluation and validation in medical image analysis", *IEEE Transactions on Medical Imaging*, Vol. 25, No. 11, pp. 1451-1461, 2006.

[26] D. Karimi, and S. E. Salcudean, "Reducing the hausdorff distance in medical image segmentation with convolutional neural networks", *IEEE Transactions on Medical Imaging*, Vol. 39, No. 2, pp. 499-513, 2019.