DOI: http://dx.doi.org/10.26483/ijarcs.v17i1.7408

ISSN No. 0976-5697

Volume 17, No. 1, January-February 2026

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

AN ATTENTION-ENHANCED NESTED 3D U-NET WITH COMPOUND LOSS FOR TUMOR CORE SEGMENTATION IN MULTISEQUENCE VOLUMETRIC BRAIN MRI

Ceena Mathews
Department of Computer Science
Prajyoti Niketan College
Thrissur, India

Anuj Mohamed
School of Computer Sciences
Mahatma Gandhi University
Kottayam, India

Abstract: Maximal removal of the tumor core tissues while preserving healthy brain tissues is essential to decrease tumor recurrence and improve patient outcomes. Therefore, a comprehensive pixel-wise understanding of brain MR images becomes imperative for the precise and automated identification of tumor core regions, such as enhancing tumor, non-enhancing tumor, and necrosis. Attention mechanisms allow deep learning models to focus on relevant regions, thereby improving the accuracy of tumor delineation, especially in the presence of class imbalance. In this work, we extend our previously proposed attention-enhanced nested U-Net and compound loss framework to a fully volumetric formulation for multisequence MRI segmentation. We trained and tested the proposed 3D model using 190 multisequence volumetric brain MR images in the BraTS 2019 benchmark dataset. The proposed model achieved Dice scores of 0.88, 0.78, and 0.84 on the BraTS 2019 dataset, for the whole tumor, enhancing tumor, and tumor core, respectively. Experimental results demonstrate competitive and consistent improvements in tumor core segmentation compared with recent 3D deep learning models.

Keywords: Brain tumor; Segmentation; Attention gate; 3D nested U-Net; Compound loss; Multisequence 3D MRI

\

I. INTRODUCTION

Malignant brain tumors such as gliomas are heterogeneous and infiltrative, with distinct subregions such as enhancing tumor (ET), non-enhancing tumor (NET), necrosis (NCR), and edema (ED). To treat tumor patients, several procedures such as microsurgical resection, radiotherapy, and chemotherapy are used. The surgical goal is to remove only the tumor core (ET, NET, NCR) tissues for a better prognosis. The ED surrounding the tumor is not resected, but is managed with medications like steroids to improve the patient's quality of life. Therefore, effective automated segmentation of tumor core regions from magnetic resonance images (MRI) before surgery is crucial.

However, precise automatic segmentation of tumor subregions from MRI is challenging, mainly due to class imbalance and the heterogeneous nature of gliomas. Class imbalance in brain MR images refers to an uneven distribution of certain tissue types or structures. In the brain tumor images, the distribution of ED, NET, and ET regions within MR images may vary significantly. One of these regions may dominate the image, while the other regions are relatively sparse. If there are more ED regions than ET regions, the model may become biased toward ED segmentation, resulting in a higher Dice similarity coefficient (DSC) for ED. Moreover, uniform tissue characteristics exhibited by the ED region make it relatively easier for segmentation algorithms to identify and differentiate it from surrounding normal brain tissues. It is apparent from the literature [1] - [7] that ED has a better DSC than ET and TC, and hence whole tumor (WT), which consists of ED and TC, has a large DSC score. Such a diagnosis may mislead the physician during tumor core resection, thereby increasing the risk of tumor recurrence. Tumor core structures must therefore be precisely segmented from brain MRI to reduce the risk of tumor recurrence.

Furthermore, enhancing tumor regions may have more heterogeneous appearances, and the boundaries between the tumor and the surrounding tissues can be less clear, making it more challenging to segment ET areas accurately. The motivation of our work is to predict with greater accuracy all tumor subregions, especially tumor core tissues, from brain MR images for the effective prognosis of the disease.

Despite significant advances in deep learning-based brain tumor segmentation, accurately delineating tumor core subregions remains challenging due to severe class imbalance and weak boundary contrast in volumetric MRI. Existing encoder–decoder architectures often achieve high performance for whole tumor segmentation, while underperforming on clinically critical tumor core regions. In such cases, complete removal of the tumor through surgery may not occur, thus increasing the risk of tumor recurrence.

The underlying problem of the encoder-decoder architectures, such as nested U-Net, is that the skip connections in the encoder-decoder model concatenate 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 the poor performance of such networks [8].

This limitation motivates the need for architectures that explicitly enhance feature selection along skip connections and loss functions that emphasise hard-to-classify tumor structures. By integrating attention gates proposed by Oktay et al. [9] with a nested U-Net architecture, the model can better emphasise important regions of the image while reducing the influence of irrelevant regions. This can lead to improved segmentation accuracy.

In our earlier work [10], an enhanced attention gate and a compound loss function were introduced for slice-wise (2D) multimodal MRI segmentation. In this work, we extend the previously proposed framework to a fully volumetric (3D) formulation. These components are retained without structural modification, and systematically evaluate its effectiveness on multisequence 3D MRI volumes from the BraTS 2019 dataset, where inter-slice contextual information is critical for accurate tumor core segmentation. The 3D model was evaluated using DSC score, sensitivity, and specificity.

The main contributions of this work are as follows:

1. Reformulation of a previously proposed attention-based nested U-Net and compound loss framework into a fully volumetric (3D) architecture for brain tumor segmentation.
2. Comprehensive evaluation of the 3D model on multisequence volumetric MRI, analysing its effectiveness in segmenting clinically critical tumor core subregions.
3. An empirical comparison with recent 3D brain tumor segmentation models on the BraTS 2019 benchmark.

II. RECENT STUDIES

Deep learning algorithms are gaining popularity with state-of-the-art results in brain tumor segmentation. This section discusses studies based on 3D CNN-based brain tumor segmentation models evaluated using the BraTS 2019 dataset.

Islam et al. [1] adopted a 3D U-Net architecture and integrated channel and spatial attention with the decoder network to perform segmentation. Cheng et al. [11] proposed a novel memory-efficient cascade 3D U-Net (MECU-Net) for brain tumor segmentation. The model was trained and evaluated using edge loss and weighted Dice loss.

Weninger et al. [2] proposed a multi-tasking method where the network is trained for tasks such as tumor segmentation, image reconstruction, and enhancing tumor detection. These three tasks share an encoder but have different decoder architectures. For the tumor segmentation task, they used 3D U-Net architecture, whereas the reconstruction branch is based on a fully convolutional variational auto-encoder, and to enhance the tumor detection network added a classification branch to the segmentation network. The model was trained and evaluated using different weighted loss functions.

The aforementioned studies on the volumetric MRI dataset in the literature demonstrate that the DSC scores for the tumor subregions ET and TC are lower compared to WT. This would adversely affect the prognosis of the disease. In this study, to increase the predictive accuracy of tumor core structures such as ET, NET, and NCR in volumetric images, we adopted the nested 3D U-Net model integrated with the enhanced structure of attention gates and compound loss function proposed in our previous work.

III. MATERIALS AND METHODS

A. *BraTS Dataset*

The volumetric MRI dataset used in this study is provided by the CBICA (Center for Biomedical Image Computing and Analytics) as part of the multimodal Brain Tumor Segmentation (BraTS) 2019 challenge jointly organised by the International Conference on Medical Image Computing and Computer Assisted Interventions (MICCAI). The BraTS 2019 training dataset includes pre-operative multimodal MRI scans of 335 patients, of which 259 are HGG, and 76 are LGG cases. These MRI scans were acquired under different clinical protocols and using various scanners at multiple institutions (n=19).

Each MRI volume has a dimension of 240x240x155 (width x height x depth). Each volume has 155 slices, and each slice is a two-dimensional (240x240) image. There are three different planes in a volume, such as the axial, sagittal, and coronal planes. These planes view the brain from three directions and contain different spatial information. The images are co-registered to the same anatomical template and skull stripped. Since the MR images are acquired using different scanners, they are of different resolutions. Therefore, the images are interpolated to the same resolution (1 mm^3) [12]-[14]. The MR images are stored in the Neuroimaging Informatics Technology Initiative (NIfTI) format with a '.nii.gz' extension. NIfTI files are used very commonly in imaging informatics for neuroscience and neuroradiology research.

Each patient case in the dataset has four MRI sequences (multi-parametric), such as T1-weighted (T1), T1-weighted with gadolinium contrast (T1Gd), T2-weighted (T2), fluid-attenuated inversion recovery (FLAIR), and has a manually segmented ground truth annotated by expert neuroradiologists. Tumor regions in the ground truth are annotated as background (label 0), NCR and NET (label 1), ED (label 2), and ET (label 4). Label 3 is not assigned to any region. Each MRI sequence is significant in identifying different tumor subregions. In T1Gd, ET appears brighter, whereas NET and NCR appear darker. In FLAIR images, ET, NET, and ED appear brighter.

Additionally, for evaluating the performance of the segmentation of brain tumors, three subregions are suggested by the dataset providers:

1) Tumor core (TC), which includes NCR, NET, and ET;
2) ET area
3) Whole tumor (WT), where WT comprises TC and ED.

B. *Nested 3D U-Net with Enhanced Attention Gate*

In this study, we used the network architecture proposed in our previous work, a nested U-Net with an enhanced attention gate shown in Fig. 1.

The nested U-Net model consists of an encoder and decoder. The encoder captures the context information and passes it to the decoder of the corresponding convolutional block through the attention gate so that more precise and relevant location features of the foreground objects are extracted. The input to each convolutional block in the decoder part is the concatenated output of two feature maps, which are as follows:

(1) the output from the attention gate before the decoder

(2) upsampled feature map from the lower deconvolution block.

The enhanced structure of the attention gate used is shown in Fig. 2. The gate accepts feature maps from the previous nodes in the same skip connection pathway along the same depth as the input, in addition to the inputs used with the original attention gate. The hypothesis is that incorporating the skip-connection feature maps into the attention gate improves gradient flow and reduces the semantic gap between encoder and decoder feature maps.

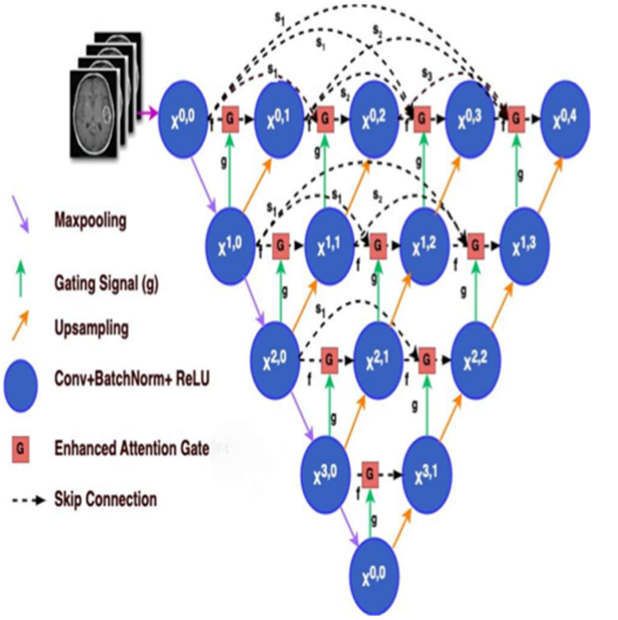

Figure.1. Nested U-Net with enhanced attention gate [10]

employed to select more crucial features from the encoded feature maps and pass them on to the subsequent decoder. On each of these inputs, the convolution operation (*W*), and batch normalisation (*b*) are applied before being added pixel by pixel. The combined output is then activated using ReLU (σ_1). This is followed by a convolution operation (W_0), batch normalisation (b_0), and sigmoid activation (σ_2).

A sigmoid function calculates the attention weights for each input feature vector. These attention weights assign a level of importance to each feature vector, resulting in a weighted sum of features that is then passed to the next layer of the neural network. The sigmoid function is a nonlinear activation that captures complex relationships between input features and their relative importance. This allows the attention module to identify non-linear patterns in the input data that may be relevant to the target task. It has a smooth gradient that facilitates backpropagation and enables efficient training of neural networks. It maps the input values to a range between 0 and 1, which is the attention coefficient ϕ. This allows the attention layer to prioritise features most relevant to the target task while minimising the impact of irrelevant features. This attention map guides the model focus on relevant areas and helps identify and segment tumor regions accurately.

The weights assigned to the attention scores are used to calculate how much each component contributes to the final output representation. Higher attention scores are given greater weight, while lower attention levels have less impact on the output. The attention mechanism can scale or weight each component of the input sequence separately with element-wise multiplication. Element-wise multiplication allows the attention mechanism to pay attention to relevant information while ignoring less crucial components. The attention coefficient ϕ, feature maps of the encoder, and those of the skip connections are multiplied element-wise to generate the output as shown in (1).

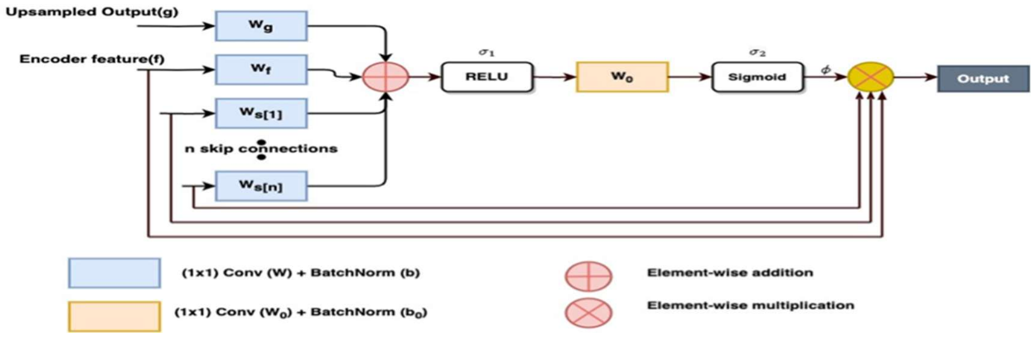

Figure.2. Enhanced structure of Attention Gate [10]

The inputs to the enhanced attention gate are as follows:

(1) up-sampled gating signal (*g*)
(2) encoder feature map (*f*) and
(3) feature maps *[i]* of the skip connections at the 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. 1 accepts as input the feature maps of $X^{0,0}$ and $X^{0,1}$, up-sampled output from $X^{1,2}$ and the feature map of the encoder $X^{0,2}$. The feature maps generated by nodes such as $X^{0,1}$ represent the low-level features extracted from the input image. The up-sampled output captures the high-level features learned from the previous stage of the network. The gating signal is

$$output = \prod_{i=1}^{n} s[i] * f * \phi \tag{1}$$

where $\phi = \sigma_2\left(W_0 \times \left(\sigma_1(\sum_{i=1}^{n}(W_{s[i]} \times s[i] + b_{s[i]}) + (W_f \times f + b_f) + (W_g \times g + b_g))\right) + b_0\right)$ (2)

where *s[i]* is the feature map of the skip connections at the same level where (*i* = 1…*n*), *n* indicates the number of nodes before the encoder, *f* indicates the feature map of the encoder, ϕ indicates the attention coefficient, σ_1 indicates RELU activation, and σ_2 indicates sigmoid activation, W indicates convolution, b indicates batch normalisation. The model has 13.4 million trainable parameters.

C. *Compound loss*

To address class imbalance, we employed the compound loss function as in (3), proposed in our earlier work [10], which combines weighted binary cross-entropy (WBCE) in (4) and focal Tversky loss in (5). This loss formulation is retained in the present study to ensure consistency while evaluating the impact of volumetric learning on tumor core segmentation.

$$Loss = Loss_{ft} + Loss_{wbce} \tag{3}$$

The classes defined for the tumor subregions are smaller in volume than the background classes representing healthy brain tissues. This results in the average prediction of tumor subregions. From the literature, it is observed that the Tversky loss, Focal Tversky loss, and WBCE loss, among others, work effectively on class-imbalanced datasets.

The $Loss_{wbce}$ defined in (4) assigns different weights to different classes, enabling us to distinguish regions of different classes and learn significant patterns in the image. In the proposed loss, we modified the decay parameter of the Focal Tversky loss function, denoted by γ, which effectively captures small classes, as given in (5).

$$Loss_{wbce} = \frac{-(T \times log(P) \times w + (1 - T) \times log(1 - P))}{N} \tag{4}$$

where T indicates ground truth values, P indicates predicted values, N indicates the number of samples, and w is a hyperparameter that enables a tradeoff between false positives and false negatives. To reduce the number of false negatives, set w>1; to decrease the number of false positives, set w<1.

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

where $Loss_{ft}$ represents Focal Tversky loss, and TI represents Tversky index as defined in (6).

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

where p_{ic} is the predicted value of the pixel i of the 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. Hyperparameters α and β can be tuned to shift the emphasis to improve recall in the case of large class imbalance.

From the literature [15] - [19], it is observed that compound loss function-based models gain better segmentation results. The advantage of using multiple loss functions is that it can lead to better generalization and robustness as the model is trained on multiple criteria at the same time. Therefore, for better optimisation of the segmentation model, we used the compound loss function proposed in our previous work.

The resultant loss function assigns more weight to difficult instances through the focal loss component and assigns more importance to specific classes through WBCE loss. This can lead to a more accurate model that can handle imbalanced datasets and difficult predictions.

D. *Evaluation Metrics*

We tested the performance of the model using evaluation metrics such as DSC, sensitivity, and specificity. DSC measures the amount of overlap between the predicted mask and ground truth labels as indicated in (7). Sensitivity measures the rate of true positives considering the positives in both ground truth and predicted mask and is calculated using (8). Specificity measures the true negative rate considering the negatives in both ground truth mask and the predicted mask as mentioned in (9).

$$DSC = \frac{2 \times |T \cap P|}{|T| + |P|} \tag{7}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{8}$$

$$Specificity = \frac{TN}{TN + FP} \tag{9}$$

where T represents the ground truth values, P indicates the predicted values, *TP, TN, FP*, and *FN* represents number of true positives, true negatives, false positives, and false negatives, respectively.

IV. RESULTS AND DISCUSSION

Volumetric MRI images provide a clearer view of the contextual information surrounding tumor structures. In volumetric MRI, each voxel captures distinct physical properties of the tissue type. These aspects are critical as even minor pathological variations are considered valuable. Hence, we used 3D multisequence MRI in the BraTS 2019 dataset to test the effectiveness of the proposed model.

For training and testing, 190 multisequence MR images of HGG patients from the dataset are used. Each patient case contains four different MRI sequences, including T1, T2, T1Gd, and FLAIR. Each image has the following dimensions: 240x240x155, where 240x240 denotes the height and breadth of a slice, and 155 is the total number of slices.

Each volume is reduced to 160x192x128 dimensions due to computational and memory constraints. Since each MRI sequence captures different tumor characteristics, all four sequences were merged to form a 4D volume, enabling more precise detection of tumor subregions. Of the 190 high-grade glioma MR images, 60% (114 multisequence MR images) are used for training, 20% (38 volumetric images) are used for validation, and 20% (38 volumetric images) for testing. Data were preprocessed using a simple normalisation technique to scale the voxel values in the range of 0 to 1.

The 3D implementation of the proposed model uses 3D convolutional layers instead of the 2D layers of the nested U-Net model. The number of filters used in this framework is 8,16,32, and 64, and the kernel size is 3 x 3 x 3. Since the number of filters used in the convolutional layers is less, the number of trainable parameters is only 244 thousand.

The proposed model was implemented using NVIDIA Tesla V100-SXM2 with 16 GB GPU memory and 54 GB RAM. Keras with Tensorflow (v2.4.3) is used as the backend. Since we have only 114 volumetric images for training, the model was trained using the stochastic gradient descent (SGD) optimiser with a learning rate of $3e^{-2}$ and momentum of 0.95. SGD is used because it is popular for its faster convergence, especially on smaller datasets, where it can swiftly adapt to the limited amount of data. Because stochastic sampling introduces noise, SGD can be useful in reducing overfitting on small datasets. To optimise the performance of the model, we have used the proposed compound loss function. The model is trained for nearly 650 epochs, since the number of samples used for training is less. Each epoch took nearly 97 seconds to train 244 thousand parameters. The model was trained only with a batch size of 2 due to model complexity and the volumetric dataset.

In this study, we substituted the hyperparameter *w* with a value greater than 1 to decrease the number of false negatives. The proposed model's performance was evaluated using metrics such as DSC, sensitivity, and specificity for different values of α, β, and γ.

From the literature and based on our 2D implementation, it has been discovered that the network models trained using Focal Tversky loss perform better with hyperparameter values α = 0.7, β = 0.3, and **γ** = 0.75. Henceforth, due to the memory and computational constraints, we conducted experiments using the same Focal Tversky parameter values with two values of weight *w* ie., 2 and 4.

Tables I and II show the mean DSC, sensitivity, and specificity scores for WT, ET, and TC subregions of the tumor using HGG cases in the BraTS 2019 dataset for some combinations of the hyperparameters α, β, **γ**, and *w*. It is observed that the model gives comparable performance for the tumor subregions ET and TC (ET + NCR + NET).

Table I. Mean DSC score of the 3D model for different values of w, α, β, and γ with BraTS 2019 dataset

w	**α**	**β**	**γ**	**Mean DSC**		
				WT	**TC**	**ET**
2	0.7	0.3	0.75	**0.8771**	0.8355	**0.7808**
2	0.7	0.3	0.9	0.8720	0.8556	0.7730
4	0.7	0.3	0.75	0.8631	0.8387	0.7637
4	0.7	0.3	0.9	0.8538	**0.8656**	0.7234

Since the model gives better results for the values w = 2, α = 0.7, β = 0.3, and **γ** = 0.75, we used the segmentation results using these values for comparison against other state-of-the-art 3D segmentation models and the results of the comparison are shown in Table III.

The proposed model demonstrates a favourable trade-off between enhancing tumor and tumor core segmentation accuracy. Although the model shows consistent improvements, performance may vary across datasets due to differences in acquisition protocols and tumor heterogeneity. With increased computational and memory capacity, the performance of our model can be further improved by training with more patient data and by using a larger batch size while training. Our model achieves DSC scores of 0.88, 0.78, and 0.84 for WT, ET, and TC respectively.

Fig. 3 depicts the box plot for DSC metric of the proposed model for the tumor subregions WT, ET, and TC. It shows that 50% of the test samples achieve a score of 0.9 for WT and TC, whereas 0.8 for ET. A sample of segmentation results of 5 patients in the BraTS 2019 dataset using the proposed model is illustrated in Fig. 4.

Superior performance has been achieved because of the integration of enhanced attention gates into nested 3D U-Net and the utilisation of the compound loss function, which strongly focuses on the hard-to-classify classes in the image. Despite the higher accuracy achieved in the tumor areas ET and TC, the DSC metrics for WT show a lower score because it includes the DSC score of the ED subregion and the DSC score of TC. The ED region exhibits a low score of 0.74. Training with more datasets can increase the prediction accuracy of all the tumor subregions.

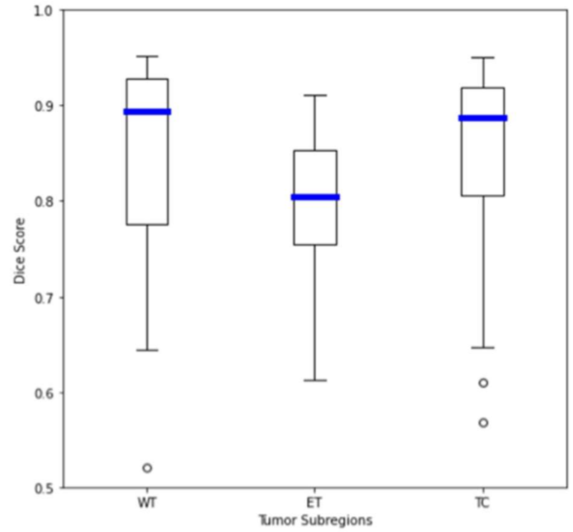

Figure 3. Boxplot showing the DSC metric of the proposed 3D model for the WT, ET and TC subregions

Table II. Sensitivity and specificity values of the proposed 3D model for different values of w, α, β, and γ with BraTS 2019 dataset

w	α	β	γ	Sensitivity			Specificity		
				WT	TC	ET	WT	TC	ET
2	0.7	0.3	0.75	**0.8681**	0.7619	0.7471	0.9976	**0.9995**	0.9992
2	0.7	0.3	0.9	0.8300	0.7769	0.7129	**0.9984**	**0.9995**	**0.9994**
4	0.7	0.3	0.75	0.8559	**0.8488**	**0.8206**	0.9973	0.9985	0.9984
4	0.7	0.3	0.9	0.8577	0.8189	0.7303	0.9968	0.9994	0.9990

Since more images are used to train the 2D model (8100 samples) than the 3D model, the 2D implementation of the model performs better than the 3D version. This is because only 114 volumetric images are utilised for training in the 3D implementation. A high-end GPU is required for training more 3D samples. With appropriate computing power, optimisation algorithms like gridsearchCV, random search, and bayesian optimisation methods may be used to choose the best subset of hyperparameters to employ with the proposed loss function.

Table III. Comparison of the model with state of the art 3D segmentation models on the BraTS 2019 dataset.

Model	Mean DSC		
	WT	ET	TC
Islam et al.[1]	0.87	**0.778**	0.78
Weninger et al.[2]	0.82	0.65	0.76
Cheng et al.[11]	**0.89**	0.756	0.81
Proposed 3D	0.88	**0.781**	**0.836**

Despite the promising results, the proposed model is limited by the relatively small number of volumetric training samples due to computational constraints. Additionally, hyperparameter optimisation was restricted to a limited search space. Future work will focus on large-scale training and automated hyperparameter tuning to further improve segmentation robustness.

Compared to the previously reported 2D implementation [10], the volumetric model demonstrates improved consistency in tumor core segmentation by leveraging inter-slice contextual information. Although the 3D model is trained with fewer samples due to memory constraints, it achieves competitive performance, highlighting the advantage of volumetric feature representation over slice-wise learning.

Extending a 2D segmentation framework to a volumetric setting is non-trivial due to the increased computational complexity, memory constraints, and the need to effectively capture inter-slice contextual dependencies. Unlike 2D slice-wise segmentation, volumetric learning enables spatial continuity across axial, sagittal, and coronal planes, which is essential for accurately segmenting heterogeneous tumor core regions. Therefore, validating the proposed framework in a 3D setting is necessary to assess its clinical applicability.

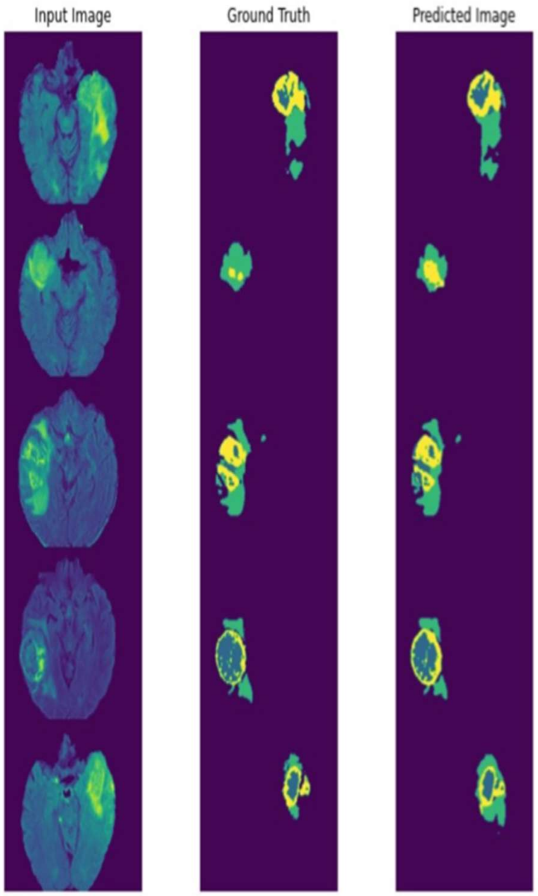

Figure 4. Segmentation of brain tumors in 5 patients using the proposed 3D model. In both the ground truth and predicted mask, green indicates ED, blue indicates ET, and yellow indicates the NETs and NCR

V. CONCLUSION

The sagittal, axial, and coronal planes of an MRI volume provide three distinct perspectives of the brain and different spatial data. Therefore, we used multisequence MR volumetric images in the BraTS 2019 dataset for training and testing to attain better segmentation results. The model was trained with an SGD optimiser using 190 multisequence MR images in the BraTS 2019 dataset. We achieved DSC scores of 0.88, 0.78, and 0.84 for the WT, ET, and TC, respectively. The segmentation results of the proposed 3D model are compared against the state of the art 3D models and the proposed model achieves competitive performance with recent state of the art 3D segmentation models, particularly for tumor core subregions. The segmentation accuracy of the 3D model can be improved by increasing the number of filters in the 3D convolutional layers which need high-end computing devices.

This study is limited by the number of volumetric training samples due to computational constraints, and experiments were conducted only on BraTS 2019. Hyperparameter optimisation was restricted to a limited search space. Future work will extend the evaluation to newer BraTS datasets and explore automated hyperparameter tuning under higher computational capacity.

The improved performance of the model can be attributed to the following factors:

1) The deep selection of important information from the hierarchically extracted features along the skip connection pathway using an improved attention gate.

2) The compound loss function concentrates more on the important foreground and difficult-to-classify pixels.

VI. REFERENCES

[1] M. Islam, V. Vibashan, V. J. M. Jose, N. Wijethilake, U. Utkarsh, and H. Ren, “Brain tumor segmentation and survival prediction using 3D attention U-Net,” in International MICCAI Brainlesion Workshop. Springer, 2019, pp. 262–272.

[2] Weninger, Q. Liu, and D. Merhof, “Multi-task learning for brain tumor segmentation,” in International MICCAI brainlesion workshop. Springer, 2019, pp. 327–337.

[3] W. Shi, E. Pang, Q. Wu, and F. Lin, “Brain tumor segmentation using dense channels 2D U-Net and multiple feature extraction network,” in International MICCAI Brainlesion Workshop. Springer, pp. 273–283, 2019.

[4] M. Hamghalam, B. Lei, and T. Wang, “Convolutional 3D to 2D patch conversion for pixel-wise glioma segmentation in MRI Scans,” in International MICCAI Brainlesion Workshop. Springer, pp. 3–12, 2019.

[5] R. Agravat and M. S. Raval, “Brain tumor segmentation and survival prediction,” in International MICCAI Brainlesion Workshop. Springer, pp. 338–348, 2019.

[6] Z. Dai, N. Wen, and E. Carver, “Brain tumor segmentation using non-local mask r-cnn and single model ensemble,” in International MICCAI Brainlesion Workshop. Springer, 2021, pp. 239–248.

[7] 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 International MICCAI Brainlesion Workshop. Springer, 2020, pp. 230–240.

[8] 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.

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

[10] C. Mathews and A. Mohamed, “Nested U-Net with enhanced attention gate and compound loss for Semantic Segmentation of brain tumor from multimodal MRI,” International Journal of Intelligent Engineering & Systems, vol. 15, no. 4, 2022.

[11] X. Cheng, Z. Jiang, Q. Sun, and J. Zhang, “Memory-efficient cascade 3D U-Net for brain tumor segmentation”, in International MICCAI Brainlesion Workshop. Springer, 2019, pp. 242–253.

[12] S. Bakas et al., “Advancing the Cancer Genome Atlas Glioma MRI Collections with Expert segmentation labels and radiomic features,” Nature Sci. Data 4, 170117, 2017.

[13] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, “Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection,” The Cancer Imaging Archive, Vol. 286, 2017.

[14] 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, vol. 34, no. 10, pp. 1993–2024, 2014.

[15] M. Yeung, E. Sala, C.-B. Sch¨onlieb, and L. Rundo, “Unified focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation,” arXiv preprint arXiv:2102.04525, 2021.

[16] 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.

[17] Y. Zhang, S. Liu, C. Li, and J. Wang, “Rethinking the Dice loss for deep learning lesion segmentation in medical images,” Journal of Shanghai Jiaotong University (Science), vol. 26, pp. 93–102, 2021.

[18] W. Zhu, Y. Huang, L. Zeng, X. Chen, Y. Liu, Z. Qian, N. Du, W. Fan, and X. Xie, “Anatomynet: deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy,” Medical physics, vol. 46, no. 2, pp. 576–589, 2019.

[19] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, “Combo loss: handling input and output imbalance in multi-organ segmentation,” Computerized Medical Imaging and Graphics, vol. 75, pp. 24–33, 2019.