

An Innovative Approach to Multimodal Brain Tumor Segmentation: The Residual Convolution Gated Neural Network and 3D UNet Integration



Islem Gammoudi^{1*}, Raja Ghozi², Mohamed Ali Mahjoub³

¹ National Engineering School of Sousse, LATIS Laboratory of Advanced Technology and Intelligent Systems, University of Sousse, Sousse 4023, Tunisia

² Department of Computer Engineering, American University of Bahrain, Riffa 38884, Bahrain

³ Department of Computer Science, National Engineering School of Sousse, LATIS-Laboratory of Advanced Technology and Intelligent Systems, University of Sousse, Sousse 4023, Tunisia

Corresponding Author Email: islemislem65@gmail.com

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410111>

ABSTRACT

Received: 23 February 2023

Revised: 30 July 2023

Accepted: 10 August 2023

Available online: 29 February 2024

Keywords:

brain tumor segmentation, 3D UNet, ResNet, Res-Gated-3DUNet, BraTS2020 validation dataset

Background: The segmentation of brain tumor images remains a challenging yet vital aspect of medical image analysis. In particular, Gliomas, being the most prevalent type of malignant brain tumors, necessitate early and accurate diagnosis for effective monitoring, analysis, and planning of radiotherapy. However, the segmentation of Glioma images presents two major obstacles: the imbalance of segmented classes and a dearth of training data. Thus, the conception of a dependable and suitable model for such medical image segmentation is paramount. **Methods:** In recent years, the utilization of deep neural networks for the automatic segmentation of Gliomas proved to be very promising. Drawing inspiration from this success, a three-dimensional brain tumor analysis approach, termed as the Residual Convolution Gated Neural Network, which incorporates residual units and signal gating into UNet, thereby enhancing the performance of brain tumor segmentation. By combining the advantages of UNet, ResNet and signal gating, we obtain a novel 3D UNet, featuring block modifications using the newly formulated ResNet 'M' blocks. As a result, our developed Res-Gated-3DUNet network offered improved accuracy in Gliomas image segmentation results. The model was trained and validated using the 2020 Multi-modal Brain Tumor Segmentation Challenge (BraTS2020) datasets. The validation stage using the BraTS2020 dataset resulted in relatively high dice values. The segmentation accuracy was also increased through the incorporation of an innovative combined loss function, proposed in this work. The experimental results reveal that the Res-Gated-3DUNet outperforms conventional brain tumor segmentation algorithms. **Conclusion:** when compared with current literature, the proposed methodology offers more accurate and efficient 3D brain tumor segmentation. Moreover, the Res-Gated-3DUNet architecture proved to be not only an effective, but also a promising technique for Glioma fine segmentation in multimodal images.

1. INTRODUCTION

The field of medical image segmentation has earned growing attention over the past decades due to its critical role in offering automatic identification of regions of interest to medical professionals in general and neurologist in particular for brain tumor identification. In fact, the development of an accurate, automated segmentation methodology for tumors is of critical importance to medical practitioners for it offers unbiased and precise information at the diagnostic and treatment stages. It is important to recall, among various brain tumors, Gliomas remain some of the most aggressive adult brain tumors that typically originate from glial cells and possess high risks to spread to brain tissue [1].

The role of medical imaging has proved over the years to be an invaluable tool in the medical diagnosis, with a wide range of imaging modalities being developed and used for early diagnosis, detection, and treatment stages of medical health

conditions. Among all imaging modalities, the Magnetic Resonance Imaging (MRI) stands out as a high-precision technique that is highly insightful and thus required in clinical health diagnosis. The use of MRI imaging spans a wide range of health conditions including brain strokes, multiple sclerosis, cerebral aneurysms, traumatic brain injuries, tumors, and spinal cord disorders [1]. The high-resolution images produced by the MRI technique provide an insight on various tissue parameters, providing, as a result, a wealth of information on brain tissue. This unique advantage makes MRI an ideal modality for studying brain tumors, thereby rendering the detection of Gliomas using MRI images a major focus in the field of medical image processing and analysis.

Given the profound impact of early and reliable brain tumor detection on patient outcomes, several competitions have been established to foster the development of novel diagnostic methods for such brain diseases. The Multi-modal Brain Tumor Segmentation (BraTS) challenge, held annually since

2012, is especially noteworthy as it encourages the use of innovative Deep Learning (DL) models in providing better to segmentation of brain tumor MRI images [1].

Being an integral part of any Artificial Intelligence (AI) solution, Deep learning, currently plays a critical role in a variety of competitive medical imaging challenges, including that of the BraTS challenge for Glioma segmentation. Amongst the various DL models, Convolutional Neural Networks (CNNs) have shown great efficacy in brain tumor segmentation [2]. The Full-Convolution Neural Network (FCN) [3], proposed by Long et al., represents an important contribution in the field to the field of image segmentation using neural networks [3]. A fully-symmetric Convolutional Neural Network, known as Unet, was built on the FCN has been recently used in biomedical image segmentation [4]. The UNet model was later augmented with an SE-Res module in the contracting path and an SK module in the expanding path for cardiac segmentation tasks [5]. Presently, the 3D UNet model has established itself as a prominent choice for accurate and automated brain tumor segmentation [6]. Enhanced UNet variants, including the dense convolutional network (DenseNet) [7] and the Residual Network (ResU-Net) [8], have been embraced for the segmentation of brain tumor images.

Despite these advancements, accurate Multi-modal three-dimensional (3D) the tumor segmentation of image brain remains a challenge, given the class imbalance problem often encountered in medical imaging datasets. The number of pixels or voxels representing healthy tissue significantly exceeds those representing tumor regions. This class imbalance can lead to biased models that prioritize the majority class, resulting in suboptimal segmentation performance for the minority class (tumor regions).

Segmenting brain tumors is a delicate and challenging task, particularly for localizing various Glioma sub-regions (such as necrotic core, peritumoral edema, enhancing tumor, and non-enhancing tumor core). This process presents additional challenges due to significant variations in image intensities, sizes, and shapes of these sub-tumors [9]. A focal loss function and a model cascade have been proposed in order to address the class imbalance issue, leading to remarkable improvements [10]. Due to their inherent system complexity, these model cascade methods struggle to deliver higher performances.

In light of these challenges, the proposed work presents a novel DL model, denoted as Res-Gated-3DUNet, designed to address the class imbalance problem while enhancing the segmentation performance. Specifically, a novel three-image dimensional 3D UNet architecture is developed with residual encoder-decoder blocks that efficiently learn the hierarchical representations of the input data, thus enabling a precise separation of tumorous regions in these three-dimensional images. The proposed approach seeks to address the challenges posed by class imbalance and aims to deliver more accurate and robust brain tumor segmentation results compared to existing methods. The benefits of skip connections are examined and new skip connection based on signal gate is proposed. Inspired by ResNet [8], 3D UNet [11], and short skip connections using signal gating, a hybrid architecture is therefore introduced in this study for improved accuracy in Glioma image segmentation. The proposed Res-Gated-3DUNet architecture is extensively tested on the BraTS2020 dataset as indicated in the studies [1, 12-15].

The contributions of this study are summarized as follows:

- Proposition of an efficient architecture with block modifications, denoted by ResNet'M, for MRI 3D glioma imaging.
- Introduction of a novel signal gating mechanism within the encoder to allow for efficient information flow in the network. This gating mechanism, combined with short skip connections, enhances model performance by facilitating the network to learn more complicated representations of features.
- Implementation of a multi-class focal loss function to handle the class imbalance problem in MRI glioma imaging.
- Comprehensive evaluation and comparison of the proposed model against state-of-the-art models on the BraTS2020 dataset. The results illustrate the superior performance of our approach in terms of Dice score, Sensitivity, Specificity, and Hausdorff Distance (HD).

The rest of the paper is organized as follows. Section 2 presents a review of related works. Section 3 details the proposed methodology, including data preprocessing, the proposed Res-Gated-3DUNet model, and the loss function. Section 4 details the experimental results and discussion, and Section 5 concludes the research work and suggests directions for future work.

2. RELATED WORK

In surveying the literature on Brain tumor image segmentation, one identifies a wide range of methods have been used [16]. Image segmentation in different homogenous regions has for long relied on edge detection methods, which were used for brain tumor segmentation as well [17]. Another model for different textured regions has often been successfully modeled using Markov random fields [18]. Graph-based approaches using estimate initial boundary delineation for ultimate image segmentation, have also been used for the same purpose with limited success [19-21].

A texture-based model proved to distinguish and thus segment low-grade Gliomas from high-grade Gliomas regions in brain MRI images. Although the aforementioned semi-automatic methods can provide accurate results, they often fail to achieve high performance on large image datasets.

Recently surveyed literature reflects the use of discriminative and generative models for brain tumor segmentation. Among the latter models, those using Deep Learning (DL) are emerging as the most dominant. In fact, a rising number of studies using DL methods in medical image analysis, have been continuously reported. In his review paper, Suzuki [22] gave an overview of the field of DL and its applications in medical image analysis. He assessed what has changed before and after the introduction of DL in the field of ML and identified the main reasons that make DL powerful and its applications in medical image analysis in general.

It is worth noting based on countless recent studies reveal clearly that the DL model presents a powerful tool in brain tumor segmentation. Moreover, the approach based on Convolutional Neural Network (CNN) is one of the most commonly used algorithms of DL, which brings several advantages into the problem of medical image segmentation. One can realize that a CNN-based DL model has provided a reliable architecture for brain tumor segmentation [23], while an FCN was one of the first deep learning networks for image segmentation [3]. Despite its long history, the reported success of CNN models has been limited. More recently, the concept of UNet network was introduced by Olaf Ranneberger who

applied it in training for analysis of medical image [4]. Ever since, UNet has emerged as a leading semantic segmentation framework, playing a pivotal role in the segmentation of brain tumor images. To integrate fine-level information, the UNet architecture utilizes skip-level connections, incorporating higher-resolution feature maps into the encoder path.

When first launched in 2012, the BraTS competition has already proposed the exploration of DL algorithms for tumor segmentation. Ever since, DL has been the algorithm of choice for most entries of the challenge. The use of a combined segmentation framework that includes conditional random fields and Functional Convolution Neural Networks (FCNN), offered a new segmentation framework [24]. Such unified approach was based on image patches to train the FCNNs and their segmentation outcome using the BraTS2016 dataset.

It is worth noting that the use of a 3D UNet [25] model with residual connections, in the encoder as well as the decoder paths, was the focus of a recent study on brain tumor segmentation [26]. To distinguish between cancerous cells and normal ones, an initial investigation focused on a basic DeepSeg network. This network incorporated CNN models, including residual neural networks (ResNet) and NASNet, within a modified UNet architecture and dense convolutional networks (DenseNet) [27]. The performance of this proposed architecture was assessed on the BraTS2019 challenge dataset, resulting in Dice scores ranging from approximately 0.81 to 0.84. In a separate study [28], the authors proposed an

automatic segmentation model based on CNNs, utilizing a 3×3 kernel operator to facilitate deeper architectures. The developed algorithm's performance was evaluated using the BraTS2015 dataset.

Various 3D-CNN architectures, trained on BRATS2015 brain tumor datasets were compared in a reported study [29]. The considered data is based on the 2018 BRATS challenge and the results show that the developed model outperforms both randomly initialized U-Nets and pre-trained ones for all training variables. A summary of research on DL methods for brain tumor segmentation is shown in Table 1. This latter reveals that the 3D-CNN used method, rooted in the DeepMedic CNN, performed best and thus advocated the use of smaller receptive fields with a multiscale architecture [29]. The study [30] proposed a new Deep Learning architecture which relied on both localized and global brain image datasets for accurate segmentation.

The proposed method, called Residual and Pyramid Pool Network (WRN-PPNet) of study [31], proposed to segment brain tumors by first obtaining 2D slices from 3D MRI brain tumor images and then normalizing the 2D slices and inserting them into the model.

Inspired by the aforementioned works and especially the success of the UNet architecture, we sought in our work to further develop the Res-Gated-3DUNet as a novel DL method for brain tumor segmentation, which will be next detailed.

Table 1. Summary of selected studies on brain tumor segmentation based on the Brats challenge

Authors	Method Learning	Dataset	Publication Year	Accuracy
[24]	FCNN	Brats2013	2018	Overall accuracy 0.81
[28]	CNN	Brats2013	2016	DSC 0.88 (WT), 0.83 (TC), 0.77 (ET)
[30]	RDM-Net	Brats2015	2019	Overall Dice 0.86
[26]	CNN	Brats2015	2017	DSC 0.9 (WT), 0.75 (TC), 0.73 (ET)
[27]	DeepSeg	Brats2019	2020	Overall Dice 0.81

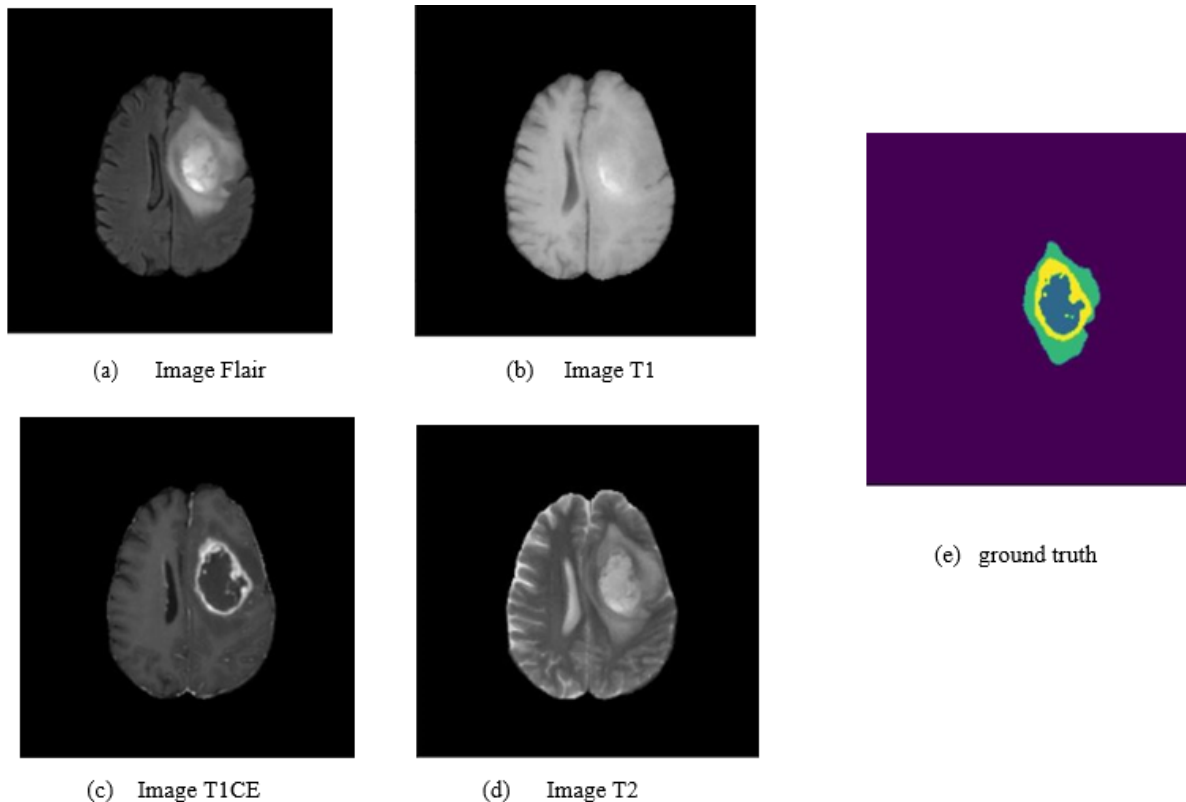


Figure 1. Four MRI modalities of the same patient from the BraTS2020 dataset

3. MATERIALS AND METHODS

In this section, we discuss the adopted data pre-processing and model architecture in our work. Given the critical importance of the database in any study, we have opted for that of BraTS, a world-wide challenge held since 2012 to assess emerging Machine Learning (ML) methods for brain tumor segmentation using multimodal MRI images. It is the largest and most publicly available dataset focused on the segmentation of brain tumors, particularly gliomas, from multi-MRI modalities.

3.1 Data pre-processing

The proposed model in this work, Res-Gated-3DUNet, was trained on the BraTS 2020 dataset which contains 369 Glioma patients and 125 cases, which are unlabeled and used as validation sets [11]. In fact, the BraTS2020 database contains four 3G-MRI sequence modalities, denoted by: T1: Native, T1-weighted: Post-contrast (T1CE), T2: T2-weighted, and Fluid Attenuated Inversion Recovery (FLAIR) volumes for each patient. The set can be divided into subsets for manual segmentation including tumor sub-regions. A summary of the characteristics of the BraTS data set are illustrated in Table 2.

The tumor sub-regions considered for evaluation are:

The whole tumor (WT): Segmenting the whole tumor extent (present in T2-FLAIR: Union of all labels).

The Tumor core (TC): Segmenting the core tumor outline (visible in T2: Union of labels 1 and 4).

The active tumor (AT): Segmenting the active tumor

regions.

Table 2. Summary of the characteristics of adopted the BraTS data set

Acronym	MRI Sequence	Acquisition Method	Property
T1	T1-weighted	Sagittal or Axial	Native image
T1Gd	T1-weighted	Axial 3D acquisition	post-contrast enhancement
T2	T2-weighted	Axial 2D	Native image
T2-FLAIR	T2-weighted	Axial or Coronal or Sagittal 2D	Native image

Figure 1 displays four distinct MRI modalities along with the corresponding ground truth segmentation of patient data from the BraTS2020 dataset. The data is partitioned following the following pre-processing steps:

-All images are stacked into a 4-dimensional array.

-In order to ensure that all MRI images share the same scale, the min-max method is used.

-Cropped images with the smallest bounding box containing the complete brain images were used.

-Four modal MRI images of the same contrast (original size: $240 \times 240 \times 155 \times 1$) are merged to form three-dimensional images with four channels (combined image: $240 \times 240 \times 155 \times 4$).

-Four-dimensional tensor is saved and numbered by array.

We note that the (FLAIR, T1CE, and T2 images are used in order to extract robust features. Figure 2 illustrates sample results of this pre-processing stage.

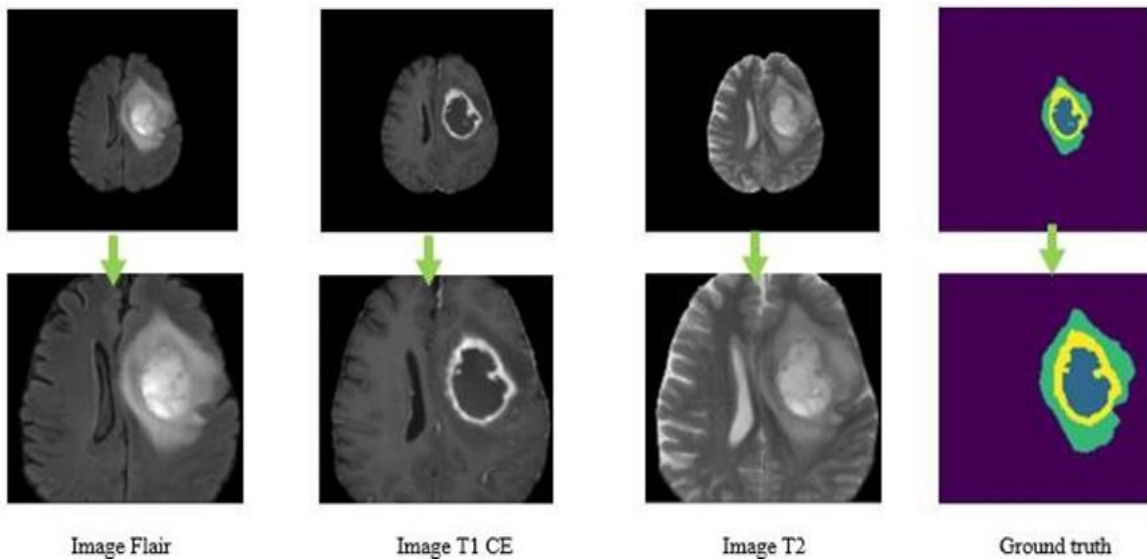


Figure 2. Sample of image pre-processing results

3.2 Proposed model architecture: Deep network

Due to the continuous improvements in imaging technologies and the large amount of applications, new algorithmic solutions for image segmentation are needed. Segmenting an image remains a key step that eventually enable quantitative measurements, diagnoses, and treatment of anatomical structures. Therefore, selecting the appropriate segmentation model is a critical stage in medical imaging in general for its outcome serves physicians to quickly and accurately identify potential regions of interest.

We recall that the image segmentation approach is even

more critical when the subject at hand is the detection of brain tumors. It is well known in the medical field that it is particularly challenging to segment brain tumors due to their irregular, uneven, and unstructured shapes and sizes. In the segmentation of Gliomas (most vicious and common brain tumor) generally, one faces at least two key challenges: Imbalance of classes and lack of training data. Therefore, this task has preoccupied researchers for several years. This has led to the appearance of various methods aiming to efficiently develop suitable algorithms, rooted in advanced ML approaches. The aim of this section is to describe the proposed novel segmentation model for medical image segmentation.

which will provide fast and robust image segmentation using the BraTS dataset for Brain tumor segmentation.

Specifically, Deep Learning approaches are exploited in brain tumor segmentation with multi-modal strategy. The approach suggests building new extensions of the most famous deep network “Unet” used for image segmentation purposes. Inspired by UNet [4], ResNet [8], and signal gated approach, we propose a combined architecture, called Res-Gated-3D UNet, as shown in Figure 3.

First, we note that instead of using the usual convolutional layers in the encoding-decoding path, we introduce ResNet^M blocks to make the learning stage easier. Second, to avoid the vanishing gradient during backpropagation step and improve the segmentation accuracy of the unbalanced classes in a given dataset, we propose a hybrid loss function. Lastly, and in order to show how this deep learning technique can be achieve advanced performances, we used Brats2020 challenging datasets.

The proposed Res-Gated-3DUNet model architecture parameters include the number of layers, filters in each layer, and input/output tensor dimensions. They are described as follows:

Number of Layers: The Res-Gated-3DUNet typically consists of multiple layers, organized into an encoder-decoder

structure. Each layer is responsible for extracting and encoding features at different levels of spatial resolution.

Number of Filters: In each layer, the number of filters determines the depth of feature representation captured by the model. As we progress deeper into the network, the number of filters typically increases to capture more complex patterns.

Input Tensor Dimensions: The input tensor is a 3D medical image, represented as a volume with spatial dimensions and channels corresponding to different modalities (e.g., T1-weighted, T2-weighted, FLAIR). In our case, an input tensor has dimensions like (128, 128, 128, 3), where the spatial dimensions are 128×128×128, and there are 3 channels representing different image modalities.

Output Tensor Dimensions: The output tensor is the segmented mask representing the predicted tumor regions. It has the same spatial dimensions as the input tensor but with a single channel. In our case, the output tensor has dimensions like (128, 128, 128, 1).

Figure 4 provides a visual representation of the encoder-decoder structure, the connections between layers, and the flow of information within the model. The efficiency of the proposed architecture is expected to be high, given its block modifications flexibility, called ResNet^M blocks.

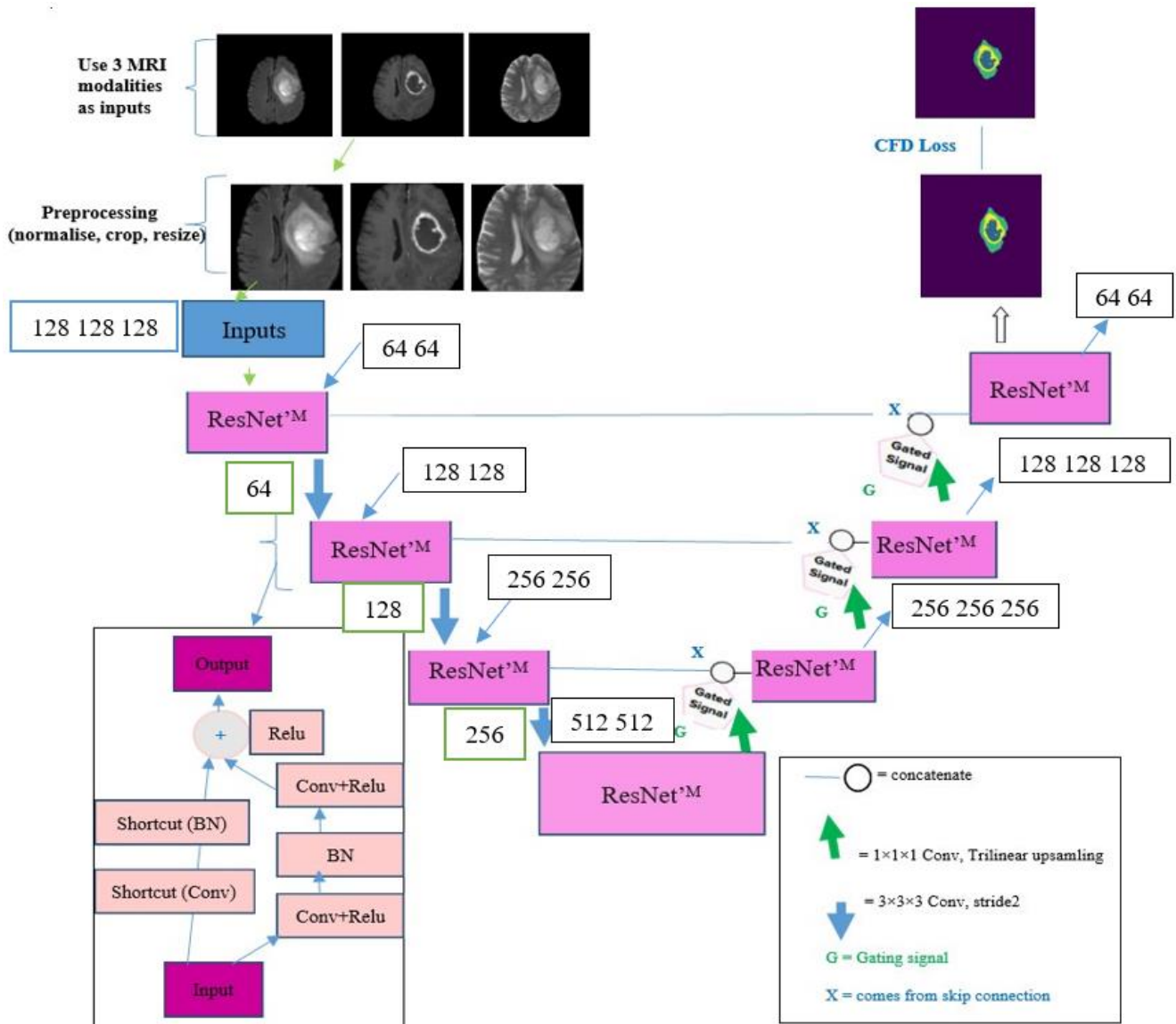


Figure 3. Proposed image segmentation algorithm architecture

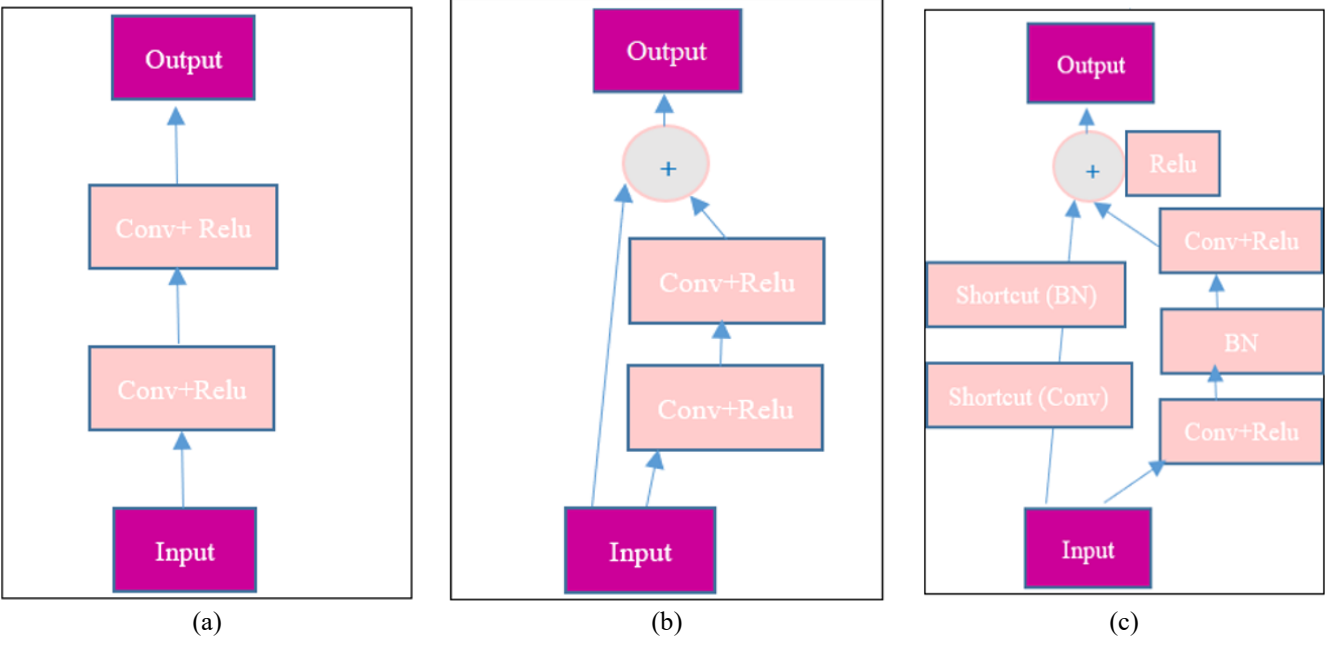


Figure 4. Different variants of convolution units and residual convolutional units: (a) forward convolutional, (b) residual convolutional block, and (c) the proposed residual convolutional block modified (ResNet^M)

It is important to recall that MRI images provide rich information thanks to its multiple acquisition parameters. Therefore, when compared with single modality images, MRI images provide significant improvement to the outcome of the feature extraction process. In fact, by offering multiple views. MRI images bring complementary information, thus contributing to better data representation and ultimately an increased discriminative power of the image analysis process. For instance, T2 and FLAIR images are more suitable in detecting a tumor with peritumoral edema, while T1 and T1c are better suited to detect a tumor core without peritumoral edema. Hence, the use of multi-modal MRI images has proved to be a powerful tool in improving the accuracy of brain tumor image segmentation, and as a result better diagnosis and potential treatment. Therefore, the fusion of multiple image modalities is an important task of multi-modal medical image segmentation. For that, we used the FLAIR, T1CE, and T2 images, in order to extract more robust features.

In this research study, we propose a novel model based on the Res-Gated-3DUNet architecture, incorporating 3D ResNet^M blocks in both the encoder and decoder to enhance the segmentation of brain tumors.

The encoder path of the model is structured such that each stage features a ResNet^M block (refer to Figure 4c) instead of the conventional residual convolutional block (Figure 4b). These ResNet^M blocks consist of two 3D convolutions with Batch Normalization (BN) and Rectified Linear Unit (ReLU), followed by a skip connection (Shortcut) for the addition of identity information. The layers are therefore used in each ResNet^M blocks according to following the pipeline flow:

Input → Convolution layer → ReLU layer → Dropout layer → BN layer → Convolution layer → ReLU layer → Shortcut layer → BN layer → Shortcut+BN layer → ReLU layer → Output.

A Max pooling operation between every two ResNet^M blocks is performed in order to reduce the resolution of the feature maps. We use the encoder to get the feature representation from the three modalities and to accelerate the training process. We note that, different MRI modalities can

highlight different sub-regions, which in turn can provide the complimentary information to analyze the brain tumor. Therefore, ResNet^M block is proposed to produce the most important features from different modalities and to highlight regions that are most relevant to brain tumor segmentation.

Decoder Path: The decoder path has the same architecture as that of the encoder and consists of four levels. This path is used to recover the image details. On the decoder side, and in order to enhance the spatial content, an up-sampling layer with tri-linear interpolation is introduced. The UNet skip connections combine the decoder and encoder paths to maintain accurate spatial information. In addition, this process leads to an appropriate representation of features that were present in the initial layer. To address this problem, we introduce a signal gated in our Res-Gated-3DUNet architecture.

A signal gated implementation in skip connections seeks to suppress activations in irrelevant regions. In fact, bridge between stages using residual connections coupled with gated signals in the encoder part Gated signal takes 2 inputs, denoted by G and X, described below:

G: Gating signal; comes from the next lower network layer. Since it comes from a deeper network layer it has better features representation.

X: Comes from the skip connection. Since it comes from early layers, it has better spatial information.

By combining the spatial information from the skip connection with the feature information from the gating signal, we improve feature reuse and propagation while reducing redundant connections. The integration of the gating signal to the multi-modal brain tumor segmentation process, is therefore adopted. Each residual unit with gated signal can be denoted then using the following expressions:

$$yl = h(xl) + F(xl + wl) \quad (1)$$

$$xl + 1 = f(xl) \quad (2)$$

where, yl denotes the output of the l th layer, $F(\cdot)$ is defined as

the residual function, the activation function is denoted $f(\cdot)$, and the identity mapping is $h(\cdot)$, x_{l+1} , and x_l represent the output and the input of the residual unit, respectively.

The Combined loss function:

The optimization of the segmentation process requires a careful selection of the loss function. For that, various performance metrics have been established to assess the quality of the image segmentation results. We now introduce the concept of cross entropy.

The cross entropy (CE): It's computed for each pixel in the objective to determine the entropy of the prediction probability and the ground truth [31]:

$$CE(pt) = -at \log(pt) \tag{3}$$

where, pt denotes the probability of the event t in p .

We have also introduced another loss function which is the Dice coefficient (DL) as another loss function. It measures the level of similarity between predicted and ground truth, no matter how small or large the target value is; this coefficient is given by:

$$DL = 1 - DS \tag{4}$$

Finally, a focal loss indicator (FL) is defined as follows [32]:

$$F_L(pt) = -at(1-pt)^\gamma + \log(pt) \tag{5}$$

where, γ is the focusing parameter and p is the estimated probability ($\gamma \geq 0$ and $p \in [0, 1]$).

It is important to recall that Dice loss and binary cross entropy are generally adopted by scientific community in the evaluation and analyzing of medical image segmentation results. On the other hand, the binary cross entropy is not most adequate in dealing with an imbalanced dataset.

Given the issue of class imbalance, the Focal loss and Dice loss were used to train the proposed Res-Gated-3DUNet network. The ultimate brain image segmentation outcome is evaluated using the following combination of the focal and dice loss, which is defined as the combined loss function (CFD):

$$CFD = FL + (0.1 \times DL) \tag{6}$$

We note the CFD loss is proposed in our architecture to handle multi-class segmentation situations, as is the case in our study.

In conclusion, we have introduced an enhanced network known as the Res-Gated-3DUNet architecture. Derived from the standard UNet model, this architecture replaces the original UNet convolution block with upgraded ResNet^M blocks. Subsequently, we will delve into the performance of our model, emphasizing its key advantages, including its ability to reduce network parameters, mitigate issues related to vanishing gradients, deepen network layers, and ultimately enhance feature propagation.

4. RESULTS AND DISCUSSION

The value of any model is measured by its performance, which greatly depends on its proper parametrization. In our case, the training step is important for the learning process of the proposed network, which will then be tested for its capacity

to properly perform the designated image segmentation.

4.1 Performance metrics

In order to evaluate obtained results, we rely on three metrics: Dice coefficient, specificity, and sensitivity, which we evaluate as follows:

Dice Coefficient: This coefficient measures the similarity between the ground truth and the predicted segmentation. It is composed by the ratio of the intersection of the two sets to their sizes averages.

$$Dice = \frac{2TP}{FN+FP+2TP} \tag{7}$$

The ‘‘Intersection’’ refers to the number of pixels that are correctly classified as tumor regions in both the predicted and ground truth segmentations. The ‘‘Total pixels in predicted’’ and ‘‘Total pixels in ground truth’’ refer to the total number of pixels in the corresponding segmentations.

The Dice coefficient ranges from 0 to 1, where 1 indicates a perfect match between the predicted and ground truth segmentations, and 0 indicates no overlap at all. Higher values indicate better segmentation performance.

Sensitivity: it measures the model’s ability to correctly detect positive instances (tumor regions) among all truly positive instances. So it is the ratio of True Positives to the True Positives and False Negatives, added together.

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

We recall that TP means True Positives, it consists of the number of pixels correctly classified as tumor regions. FN means False Negatives; they are the pixels incorrectly classified as background (missed tumor regions). Sensitivity metric ranges from 0 to 1, where 1 indicates perfect sensitivity (no missed tumor regions) and 0 indicates no true positive detections.

Specificity: Determines the model’s capability to correctly exclude negative instances (healthy tissue) among all truly negative instances. It measures the ratio of True Negatives to the total of True Negatives and False Positives.

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

True Negatives (TN) are the number of pixels correctly classified as background (healthy tissue). False Positives (FP) are the number of pixels incorrectly classified as tumor regions. Similarly, this metric is designed to have a range from 0 to 1, where 1 indicates perfect specificity and 0 indicates no true negative detections.

4.2 Network training and segmentation results

Network training

In all experiments, our network structure was implemented using Keras with a TensorFlow backend, and the training process was executed on NVIDIA GeForce GTX GPUs. The training comprised two hundred epochs, with a patch size of $4 \times 96 \times 96 \times 96$. It's worth noting that two patches were randomly selected for each patient in each epoch. The training process, utilizing Adam optimizers with learning rates of $1e-3$, took approximately 4 hours.

Segmentation results were obtained for the BraTS2020

validation datasets, and Figure 5 illustrates a sample of these results. Brain tumor segmentation was conducted on a randomly selected patient using both the original 3D UNet and our proposed model. The figure displays slices of an MRI image along with their corresponding ground truth and slices of the segmented image predictions from both the 3D UNet and Res-Gated-3DUNet architectures. Notably, when compared with ResNet and 3D UNet, the Res-Gated-3DUNet consistently demonstrated superior performance.

The combined focal dice loss is used to calculate the

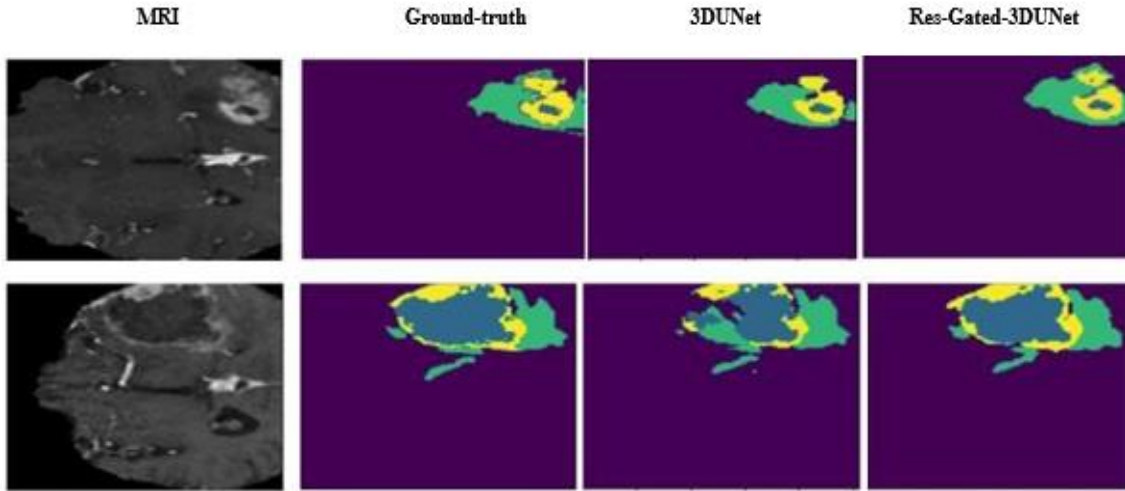


Figure 5. Sample of the segmentation results on the brats2020 validation set

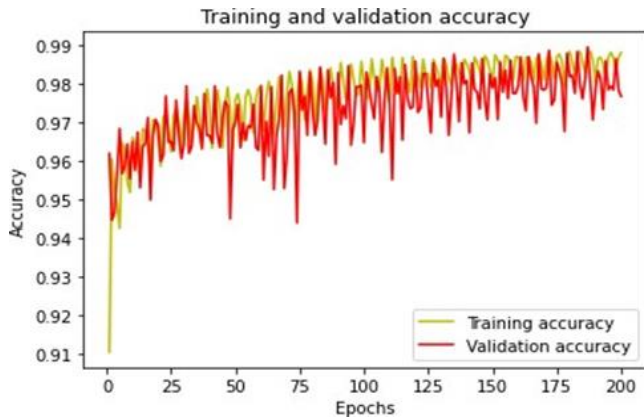


Figure 6. Training accuracy curve on the BraTS2020 dataset

Table 3. Performance of the proposed method on the Brats2020 validation datasets

	TC	ET	WT
Dice	0.854	0.849	0.879
Sensitivity	0.789	0.707	0.825
Specificity	0.999	0.999	0.998

Comparative performance and discussion:

Next, a comparison of the proposed Res-Gated-3DUNet with ResNet and the original 3D UNet method, is presented. We recall that in this study, the used parameters are the same in training process, and training data.

ResNet (Residual Network): is a deep neural network architecture introduced by He et al. [8]. It addresses the vanishing gradient problem in very deep networks by introducing residual connections, allowing layers to learn residual mappings instead of directly learning the desired

training accuracy. We recall in Figure 6 that small epochs lead to higher training accuracy; large epochs lead to a gentler change in the training curve.

The performance metric results of the proposed method on the BraTS2020 validation datasets are shown in Table 3. The obtained values for the dice, sensitivity, and specificity metrics are illustrated for WT, TC, and ET data sets. It's clear that the all TC, ET, and WT sets have dice scores of 0.854, 0.849, and 0.879, respectively.

mappings. The ResNet's skip connections facilitate the flow of gradients during training stage, making it possible to train deeper networks offering thus improved performance. This network architecture has witnessed relatively large success in various computer vision such as object detection and image classification.

3D UNet: is a three-dimensional extension of the U-Net architecture introduced by Çiçek et al. [25]. Similar to U-Net, 3D UNet uses an encoder-decoder structure with skip connections to effectively learn the relevant hierarchical features and capture the essential contextual information in 3D medical images. Due to its ability to handle 3D volumes, and produce reasonably accurate segmentations, this network architecture has been widely used in various medical image segmentation applications, such brain tumor.

Table 4 shows the dice values obtained by our new approach which reflect higher TC, ET, and WT values than those of ResNet models. It surpasses them by 0.119%, 0.125%, and 0.069% respectively. Furthermore, based on the same table, the dice of TC, ET, and WT achieved by our proposed method are also higher, respectively by 0.013%, 0.146%, and 0.044%, than 3D UNet.

Table 4. Comparative analysis of the Brats2020 dataset based on dice metric

	TC	ET	WT
Proposed	0.854	0.849	0.879
3D UNet [25]	0.724	0.703	0.835
ResNet [8]	0.735	0.724	0.810

Using the proposed combined CDF loss, the dice coefficient obtained is of 87% for the WT images, which is significantly performant as shown in Table 5.

Table 5. A comparison of loss function on the basis of average dice metric

Loss Function Value			
Dice Loss	0.84	0.849	0.879
Focal Loss	0.81	0.703	0.835
CDF Loss	0.87	0.724	0.810

Table 6. A comparison of dice scores with other advanced methods on different BraTS datasets

Work Reference	Dice (ET)	Dice (WT)	Dice (TC)
Proposed	0.79	0.87	0.82
[2]	0.65	0.78	0.75
[25]	0.62	0.84	0.73
[27]	0.66	0.85	0.70

One notes clearly that, the developed segmentation method in this work, displays a superior performance compared to that of ResNet and the original 3D UNet. This significant enhancement can be attributed to at least two primary factors. Firstly, our encoder-decoder architecture incorporates ResNet^M blocks, in contrast to the traditional convolution blocks employed by 3D UNet. The ResNet^M blocks contribute to more robust feature propagation compared to conventional blocks. Additionally, the shortcut gated connections within the ResNet^M blocks enhance gradient flow, thereby improving information flow efficiency.

Table 7. Performance comparison of the proposed model with other standard approaches

	Dice			Specificity			Sensitivity		
	TC	ET	WT	TC	ET	WT	TC	ET	WT
Proposed	0.854	0.849	0.879	0.999	0.999	0.998	0.789	0.707	0.825
Unet3d	0.724	0.703	0.835	0.998	0.999	0.998	0.785	0.711	0.867
LMB	0.764	0.716	0.824	0.999	0.999	0.999	0.721	0.695	0.766
CVUARK	0.746	0.640	0.831	0.999	0.999	0.998	0.726	0.660	0.850
Ovgu seg	0.655	0.550	0.773	0.999	0.999	0.998	0.565	0.652	0.778
MiRL	0.741	0.666	0.850	0.999	0.999	0.997	0.745	0.731	0.917

In summary, the proposed method reflected its effectiveness and clear advantages over existing methods. It also proved to capture and learn from the complementary feature information made available through the various image modalities. This allowed the extraction of the most useful features relevant to the target regions. Moreover, for better refinement of the segmentation results, a new effective combined loss function was introduced. Although this method showed promising results, its applications should be can be extended to other research fields and its performances should be further explored.

5. SUMMARY AND CONCLUSION

In this paper, we sought to address the challenging task of MRI image segmentation for brain tumor detection. For that, the world BraTS2020 challenge datasets were used in training and testing a new model for Deep Learning Network applied to image analysis and segmentation.

Specifically, we developed a network architecture, denoted by Res-Gated-3DUNet with improved segmentation accuracy of brain tumor in 3D multimodal images. The proposed network can be regarded as an extension of the 3D U-Net architecture, as it integrated the residual encoder-decoder (ResNet^M) blocks and gated activations. The residual

Table 6 compares the experimental results of our study with advanced methods that used different BraTS datasets but shared the same network parameters. As can be noted, these methods are not better than the proposed Res-Gated-3DUNet in terms of performance metrics.

Our proposed approach is a competitive automatic brain tumor segmentation method. On the basis of the dice coefficient, the Res-Gated-3DUNet model performed better than the four models mentioned above.

In this work, we present an efficient method for combining three MRI sequences into a single image. We also evaluated our Res-Gated-3DUNet model with other teams participating in the BraTS2020 Challenge.

A performance analysis of our method in comparison to some of the most advanced models is outlined in Table 7. The most accurate values in each column are highlighted in bold font.

Additional comparisons with other models and their respective results can be accessed on the online leaderboard through the online evaluation platform. (<https://www.cbica.upenn.edu/BraTS20/lboardValidation.html>.)

Several experiments illustrate the enhanced efficiency of the presented Res-Gated-3DUNet architecture compared to state-of-the-art models such as Unet3dD, LMB, CV_UARK, ovgu_seg, and MiRL. Consequently, incorporating the ResNet^M in the network proves to be an effective means of improving the accuracy the model segmentation.

connections in the encoder and decoder pathways of the 3D U-Net architecture were also incorporated. In fact, the residual connections enable the network to efficiently capture and propagate relevant feature information, thus reducing the vanishing gradient problem and facilitating the training of deeper networks. This has resulted in additional enhancement to the model’s ability to better learn complex hierarchical representations such as those encountered in 3D MRI brain images. Moreover, the proposed model incorporated gated activations within the residual blocks which resulted in the improvement of the information flow. In fact, the gated activations are known to promote feature discrimination, especially in the presence of class imbalance, which is common in medical image segmentation. As a result, feature discrimination was enhanced, which in turn improved the segmentation accuracy, particularly in challenging brain tumour regions.

The implementation of the proposed model proved to be relatively more effective producing balanced and precise segmentation results. A qualitative evaluation of the obtained results confirmed visually better segmentation assessments of the proposed model’s outcome. Moreover, the Res-Gated-3D UNet successfully demarked tumor boundaries with relatively high precision, thus indicating its robust performance in challenging tumorous regions.

Finally, we recall that the integration of a novel loss function, based by a Combined Focal-Dice Loss (CFD), played a useful role in improving the segmentation accuracy. The CFD loss function effectively addressed class imbalance, thus further enhancing the quality of the segmentation results.

One of the main disadvantages of the proposed network architecture is related to its computational complexity, since the deeper learning architecture tends to require more resources during training and inferences stages. Exploring alternative loss functions to further improve the model's performance could bring more benefits. Additionally, we suggest the incorporation of more input modalities to leverage complementary information and enhance the segmentation accuracy. Finally, a generalization of the proposed model to different medical health conditions and datasets is essential to assess its extensibility and applicability beyond the field of brain tumor region segmentation.

REFERENCES

- [1] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, J., Farahani, K., Davatzikos, C. (2017). Segmentation labels and radiomic features for the pre-operative scans of the TCGA-LGG collection. *The Cancer Imaging Archive*, 286. <https://doi.org/10.7937/K9/TCIA.2017.GJQ7R0EF>
- [2] Pereira, S., Pinto, A., Alves, V., Silva, C.A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5): 1240-1251. <https://doi.org/10.1109/TMI.2016.2538465>
- [3] Long, J., Shelhamer, E., Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 3431-3440. <https://doi.org/10.1109/CVPR.2015.7298965>
- [4] Ahmmmed, R., Hossain, M.F. (2016). Tumor detection in brain MRI image using template based k-means and fuzzy c-means clustering algorithm. In 2016 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, pp. 1-6. <https://doi.org/10.1109/ICCCI.2016.7479972>
- [5] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28
- [6] Wang, X.Y., Yang, S., Tang, M.X., Wei, Y.P., Han, X., He, L., Zhang, J. (2020). SK-Unet: An improved U-net model with selective kernel for the segmentation of multi-sequence cardiac MR. In *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges: 10th International Workshop, STACOM 2019, Held in Conjunction with MICCAI, Springer International Publishing, 2019: 246-253*. https://doi.org/10.1007/978-3-030-39074-7_26
- [7] Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q. (2017). Densely connected convolutional networks. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4700-4708. <https://doi.org/10.1109/CVPR.2017.243>
- [8] He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J. (2016). Deep residual learning for image recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, pp. 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [9] Wang, G.T., Li, W.Q., Ourselin, S., Vercauteren, T. (2018). Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: Third International Workshop, BrainLes 2017, Held in Conjunction with MICCAI, 2017: 178-190*. https://doi.org/10.1007/978-3-319-75238-9_16
- [10] Zhou, C.H., Ding, C.X., Lu, Z.T., Wang, X.C., Tao, D.C. (2018). One-pass multi-task convolutional neural networks for efficient brain tumor segmentation. In: Frangi, A., Schnabel, J., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*. MICCAI 2018. Lecture Notes in Computer Science, vol 11072. Springer, Cham. https://doi.org/10.1007/978-3-030-00931-1_73
- [11] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Springer, Cham. https://doi.org/10.1007/978-3-319-46723-8_49
- [12] Menze, B.H., Jakab, A., Bauer, S., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10): 1993-2024. <https://doi.org/10.1109/TMI.2014.2377694>
- [13] Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J.S., Freymann, G.B., Farahani, K., Davatzikos, C. (2017). Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4(1): 1-13. <https://doi.org/10.1038/sdata.2017.117>
- [14] Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R.T., Berger, C., Ha, S.M., Rozycki, M., Prastawa, M., Alberts, E., Lipkova, J., Freymann, J., Kirby, J., Bilello, M., Fathallah-Shaykh, H., Wiest, R., Kirschke, J., Wiestler, B., Colen, R., Kotrotsou, A., Lamontagne, P., Marcus, D., Milchenko, M., Nazeri, A., Weber, M.A., Mahajan, A., Baid, U., Gerstner, E., Kwon, D., Acharya, G., Agarwal, M., Alam, M., Albiol, A., Albiol, A., Albiol, F.J., Alex, V., Allinson, N., Amorim, P.H.A., Amrutkar, A., Anand, G., Andermatt, S., Arbel, T., Arbelaez, P., Avery, A., Azmat, M., Menze, B., 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*. <https://doi.org/10.48550/arXiv.1811.02629>
- [15] RSNA-ASNR-MICCAI Brain Tumor Segmentation (BraTS) Challenge 2021. <http://braintumorsegmentation.org/>
- [16] Stadlbauer, A., Moser, E., Gruber, S., Buslei, R., Nimsky, C., Fahlbusch, R., Ganslandt, O. (2004). Improved

- delineation of brain tumors: An automated method for segmentation based on pathologic changes of ¹H-MRSI metabolites in gliomas. *Neuroimage*, 23(2): 454-461. <https://doi.org/10.1016/j.neuroimage.2004.06.022>
- [17] Węgliński, T., Fabijańska, A. (2011). Brain tumor segmentation from MRI data sets using region growing approach. In *Perspective Technologies and Methods in MEMS Design*, pp. 185-188.
- [18] Muthukrishnan, R., Radha, M. (2011). Edge detection techniques for image segmentation. *International Journal of Computer Science & Information Technology (IJCSIT)*, 3(6): 259-267. <https://doi.org/10.5121/ijcsit.2011.3620>
- [19] Tustison, N.J., Shrinidhi, K.L., Wintermark, M., Durst, C.R., Kandel, B.M., Gee, J.C., Grossman, M.C., Avants, B.B. (2015). Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. *Neuroinformatics*, 13: 209-225. <https://doi.org/10.1007/s12021-014-9245-2>
- [20] Gammoudi, I., Mahjoub, M.A. (2021). Brain tumor segmentation using community detection algorithm. In *2021 International Conference on Cyberworlds (CW)*, Caen, France, pp. 57-63. <https://doi.org/10.1109/CW52790.2021.00016>
- [21] Gammoudi, I., Mahjoub, M.A., Guerdelli, F. (2020). Unsupervised image segmentation based graph clustering methods. *Computación y Sistemas*, 24(3): 969-987. <https://doi.org/10.13053/cys-24-3-3059>
- [22] Suzuki, K. (2017). Survey of deep learning applications to medical image analysis. *Medical Imaging Technology*, 35(4): 212-226.
- [23] Suárez-García, J.G., Hernández-López, J.M., Moreno-Barbosa, E., de Celis-Alonso, B. (2020). A simple model for glioma grading based on texture analysis applied to conventional brain MRI. *PLoS One*, 15(5): e0228972. <https://doi.org/10.1371/journal.pone.0228972>
- [24] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42: 60-88. <https://doi.org/10.1016/j.media.2017.07.005>
- [25] Zhao, X.M., Wu, Y.H., Song, G.D., Li, Z.Y., Zhang, Y.Z., Fan, Y. (2018). A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Medical Image Analysis*, 43: 98-111. <https://doi.org/10.1016/j.media.2017.10.002>
- [26] Bhalerao, M., Thakur, S. (2019). Brain tumor segmentation based on 3D residual U-Net. *International MICCAI Brainlesion Workshop*, Cham: Springer International Publishing, pp. 218-225. https://doi.org/10.1007/978-3-030-46643-5_21
- [27] Zeineldin, R.A., Karar, M.E., Coburger, J., Wirtz, C.R., Burgert, O. (2020). DeepSeg: Deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. *International Journal of Computer Assisted Radiology and Surgery*, 15: 909-920. <https://doi.org/10.1007/s11548-020-02186-z>
- [28] Pereira, S., Pinto, A., Alves, V., Silva, C.A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Transactions on Medical Imaging*, 35(5): 1240-1251. <https://doi.org/10.1109/TMI.2016.2538465>
- [29] Lin, T.Y., Goyal, P., Girshick, R., He, K.M., Dollár, P. (2017). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2): 2980-2988. <https://doi.org/10.1109/TPAMI.2018.2858826>
- [30] Li, C.M., Huang, R., Ding, Z.H., Gatenby, J.C., Metaxas, D.N., Gore, J.C. (2011). A level set method for image segmentation in the presence of intensity inhomogeneities with application to MRI. *IEEE Transactions on Image Processing*, 20(7): 2007-2016. <https://doi.org/10.1109/TIP.2011.2146190>
- [31] Wang, W.C., Li, Z.B., Yue, J., Li, D.L. (2016). Image segmentation incorporating double-mask via graph cuts. *Computers & Electrical Engineering*, 54: 246-254. <https://doi.org/10.1016/j.compeleceng.2016.03.003>
- [32] Lin, T.Y., Goyal, P., Girshick, R., He, K.M., Dollár, P. (2017). Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, pp. 2980-2988. <https://doi.org/10.1109/ICCV.2017.324>