



dResU-Net: 3D deep residual U-Net based brain tumor segmentation from multimodal MRI



Rehan Raza^{a,b}, Usama Ijaz Bajwa^{a,*}, Yasar Mehmood^a, Muhammad Waqas Anwar^a, M. Hassan Jamal^a

^a Department of Computer Science, COMSATS University Islamabad, Lahore Campus, Pakistan

^b Department of Computer Science, SST, University of Management and Technology, Lahore, Pakistan

ARTICLE INFO

Keywords:
Medical image segmentation
Brain tumor segmentation
Residual U-Net
BraTS 2020
Multimodal MRI
3D deep learning

ABSTRACT

Glioma is the most prevalent and dangerous type of brain tumor which can be life-threatening when its grade is high. The early detection of these tumors can improve and save the life of the patients. The automatic segmentation of brain tumor from magnetic resonance imaging (MRI) plays a vital role in treatment planning and timely diagnosis. Automatic segmentation is a challenging task due to the massive amount of information provided by MRI and the variation in the location and size of the tumor. Therefore, a reliable and authentic method to segment the tumorous region from healthy tissues accurately is an open challenge in the field of deep learning-based medical image analysis. This research paper presents an end-to-end framework for automatic 3D Brain Tumor Segmentation (BTS). The proposed model is a hybrid of the deep residual network and U-Net model (dResU-Net). The residual network is used as an encoder in the proposed architecture with the decoder of the U-Net model to handle the issue of vanishing gradient. The proposed model is designed to take advantage from low-level and high-level features simultaneously for making the prediction. In addition, shortcut connections are employed between residual network to preserve low-level features at each level. Furthermore, skip connections between residual and convolutional blocks in the proposed architecture are used to accelerate the training process. The proposed architecture achieved promising results with the average dice score for the tumor core (TC), whole tumor (WT), and enhancing tumor (ET) on the BraTS 2020 dataset of 0.8357, 0.8660, and 0.8004, respectively. To demonstrate the robustness of the proposed model in real-world clinical settings, validation of the trained model on an external cohort is performed on randomly selected 50 patients of the BraTS 2021 benchmark dataset. The achieved dice scores on the external cohort are 0.8400, 0.8601, and 0.8221 for TC, WT, and ET, respectively. The comparison of results of the proposed approach with the state-of-the-art techniques indicates that dResU-Net can significantly improve the segmentation performance of brain tumor sub-regions.

1. Introduction

Brain tumors are the collection of abnormal cells in the human brain that can damage the nervous system and harm nearby healthy brain tissues. The human brain is one of the most important organs that coordinate among the different parts of the human body. Brain tumors can affect the brain's overall functioning and are considered one of the most dangerous diseases in humans [1]. According to the survey [2], brain tumors cause more than 13,000 death every year in the USA. Brain tumors are classified into primary and secondary brain tumors. The tumor cells that originate inside the brain are called primary brain tumors, while the cells that grow somewhere else in the body and travel through

the blood are called metastatic brain tumors and are termed secondary brain tumors [1,3]. Depending on the severity of tumorous cells, brain tumors are also characterized as malignant and non-malignant. A glioma is a form of primary brain tumor that occurs in the brain and is one of the most prevalent types of brain tumor. According to the World Health Organization (WHO) report, gliomas are graded from level one to four, from the least destructive to the most aggressive type of tumor based on tumor behavior and microscopic images. Grade I and II are slow-growing cases that are classified as low-grade glioma (LGG), whereas grades III and IV are cancerous and aggressive and spread themselves quickly are classified as high-grade glioma (HGG) [4]. HGG is more aggressive, requiring surgery and radiotherapy, and has around two

* Corresponding author.

E-mail addresses: rehan.raza@umt.edu.pk (R. Raza), usamabajwa@cuilahore.edu.pk (U. Ijaz Bajwa).

years of life expectancy, whereas LGG can be benign or malignant and grows more slowly, with a life expectancy of several years [5]. The treatment of gliomas includes chemotherapy, radiation therapy, and surgery. Due to the varying size and diverse location of gliomas, detecting the tumor is a challenging task for radiologists. Since the successful treatment of the tumor is mainly dependent upon the location, type, and grade of the tumor, segmentation of the tumor is necessary for planning the treatment [6].

Medical imaging technology plays a vital role in diagnosing tumors (including their segmentation) and evaluating the response of therapy. Medical imaging mainly includes computed tomography (CT) scan, Ultrasound, examination of X-ray images, and magnetic resonance images (MRI). MRI is the most extensively used modality for brain tumor diagnosis procedures [4,6] due to its sharpness and tissue resolution, and flexibility to set different parameters to obtain distinct anatomical information. MRI is a non-invasive and advanced imaging modality that depends on nuclear magnetic resonance. MRI uses strong magnetic waves ranging from 1.5 T to 3 T and radiofrequency waves to produce a detailed representation of internal body structure and tissues [4,7]. MRI is used to extract detailed information about brain cells without harming the organ being imaged with high ionization and radiation effect [7]. MRI contains different types of image modalities/sequences, from which T1-weighted (T1w), T1-weighted contrast-enhanced (T1ce), Fluid-Attenuated Inversion Recovery (FLAIR), and T2-weighted (T2w) are the most widely used modalities for the diagnosis of the brain tumor. T1w is mainly used to measure healthy tissues. T2w has a bright tumor region, while T1c has a bright tumor border. The FLAIR scan, on the other hand, helps distinguish the edema from the cerebrospinal fluid (CSF) [8]. Accurate and timely segmentation of brain tumor using these imaging modalities can help medical professionals operate safely in tumor treatment, especially in surgery, without damaging healthy parts of the brain. The automated segmentation of a brain tumor from MRI images can help the radiologist to speed up the process and make the results consistent [1]. However, the automated segmentation of the brain tumor and its sub-regions is a challenging task because the tumorous cells may appear in any location inside the brain tissues with varying size, appearance, and shape [6,9].

The most notable computer vision technique is convolutional neural networks (CNN) that automatically learn the high dimensional hierarchical features. While on the other hand, classical machine learning algorithms are dependent on hand-crafted feature engineering. In this study, a novel approach for the brain tumor sub-regions segmentation is being presented that is comprised of a deep residual convolutional neural network and U-Net network called dResU-Net to make the training process faster and more robust. dResU-Net is the modified form of the U-Net network that also contains two parallel paths named encoder and decoder. In this proposed model, residual convolutional blocks are utilized in the encoder part of the U-Net network. The central motivation behind using these blocks in the encoder part is to utilize both low and high-level features by exploiting the skip connections for predicting segmentation masks. Parallel paths are being used that will help in decreasing the training time and provide good model generalization by combining the low and high-level features to improve the whole performance of the proposed model. Furthermore, residual blocks also help to overcome the vanishing gradient problem during the training of the proposed model. The main contribution of this paper is as follows:

- A robust encoder-decoder-based 3D Deep Residual U-Net (dResU-Net) is proposed to address the problem of BTS, which successfully uses the shortcut connections and residual blocks into the U-Net architecture to preserve local contextual response.
- Residual blocks are being utilized in an encoder part of the proposed architecture to extract low and high-level features that will help overcome the vanishing gradient issues during the training process.

- The proposed model is cross-validated on an external dataset (randomly selected 50 patients of the Brats 2021 dataset) to evaluate the robustness of the proposed architecture.

The rest of the article is arranged as follows: Section 2 covers the review of the state-of-the-art techniques on BTS. The proposed method based on 3D deep residual U-Net (dResU-Net) is discussed in Section 3. Section 4 provides details of the experiments and comparisons with the latest techniques on multimodal 3D MRI datasets. Finally, the paper is concluded in Section 5.

2. Literature review

Brain tumor is a life-threatening disease that must be diagnosed and treated in its early stage. Before the deep learning era, classical machine learning techniques were used to learn representations from brain images and were dependent on hand-crafted feature engineering. In recent years, deep learning has emerged in the medical imaging domain, allowing the deep learning model to learn local and global information automatically. This section presents a comprehensive literature review of the existing state-of-the-art methods based on BTS. Manual segmentation is tedious and time-consuming and relies on radiologists' experience. The automatic segmentation from volumetric brain scans has become the need of professionals. Many solutions were proposed for the automatic BTS from MRI images. U-Net is one of the most popular deep learning models mainly designed for biomedical image segmentation. It was first introduced by Ranneberger et al. [10] and gained popularity by achieving good segmentation results, even with few training samples. The architecture of U-Net consists of a U-shape, the left contracting (encoder part that maps raw image pixels to a rich representation) part is responsible for feature extraction, and the right expanding (decoder part that takes the features from the encoder part and produces an output and maps the output back into the original raw image pixels) part is responsible for stellar segmentation. U-Net transmits the contextual information and feature maps from the encoder path to the decoder path through concatenation so that the classifier can decide the predictions. Most of the existing techniques on BTS used 2D or 3D convolutions during the training of the deep CNN model.

2D convolutions do not fully utilize the spatial information contained in medical images, while 3D convolutions need more memory and computational resources. To address this problem, W. Chen et al. [11] introduced a separable 3D U-Net model that used separate 3D convolutions to overcome the memory requirement. In which, each 3D convolution was replaced by two consecutive convolutional layers, one 2D convolution layer that helps in to learn spatial features, and the other one 1D convolution layer was utilized to learn temporal features. The proposed model utilized full 3D brain volume by exploiting 3D convolutions further divided into three branches. The authors introduced separable temporal convolution in the residual inception model, and train their proposed model for each orthogonal view (axial, sagittal, and coronal) separately and by the multi-view fusion technique, combining all the convolutional results for better performance. The proposed model performed well on the BraTS 2018 test dataset in terms of efficiency. For the segmentation, both local and global features are vital for making a decision, but while going deeper, gradients of low-level features (that contain information about boundaries, edges, lines, or dots) become zero. The study of Wang et al. [12] proposed a TransBTS architecture that successfully embedded the transformer in 3D deep CNN model based on encode-decoder architecture. 3D CNN was firstly used to extract local and spatial information maps. For global feature capturing, these extracted features were fed into a transformer. After that, the decoder part combined these local and global features during up-sampling operation to achieve the segmentation results. The experiments were performed on the BraTS 2019 and 2020 datasets, and the proposed model gained comparable results. But their experiments needed comparatively more computational resources and memory. Liu

et al. [13] proposed an encoder-decoder-based modified deep 3D V-Net model based on fewer parameters that utilized low computational resources and memory. To make the network deeper, batch normalization and bottom residual blocks were used in the original V-Net model. The squeeze and excitation (SE) blocks were introduced at each level of the encoder and decoder path. Researchers also embedded 3D deep supervision in a modified model to hasten convergence. Their model performed well on state-of-the-art models while using the BraTS 2017 dataset. Some studies improved the segmentation model's performance by introducing an attention mechanism in the deep learning model. The study of Huang et al. [14] proposed a group cross-channel attention-based residual U-Net (GCAUNET) for BTS to utilize low-level feature maps. These features contain necessary information and can enhance segmentation efficiency. After pre-processing and noise removal, MRI slices were fed into the proposed GCAUNET model. A parallel network path called a detail recovering (DR) path was used to recover fine details and extract local feature maps from the slices of brain tumor. To track the significant features, the attention module was used with group cross-channel. The proposed study utilized benchmark 2017 and 2018 datasets for their experiments and 2D slices were extracted from these two datasets, which only contained the tumorous region. To enhance the effectiveness of the local feature response in MRI images, Zhang et al. [15] introduced attention-based residual U-Net (AResU-Net) for volumetric segmentation of brain tumors. The attention mechanism and residual units were embedded in up-sampling and down-sampling layers to enhance local response during down-sampling and feature recovery during the up-sampling operation. Due to the limited computational resources, experiments were carried out on 2D slices extracted from BraTS 2017 and 2018 datasets.

The study of D Maji et al. [16] introduced an attention-based ResU-Net with a guided decoder for BTS (ARU-GD). At each decoder layer, the model was able to guide the learning process. The attention mechanism was used to focus on relevant features instead of all features that may contain noisy features to pass into the decoder for the segmentation map. The proposed model performed well contrary to other pre-trained models while using the BraTS 2019 dataset. The study [17] introduced a deep residual-based CNN model with two parallel U-Nets for BTS (TPUAR-Net). The suggested model was capable of extracting global and local feature responses at the different levels of upper and lower paths. To accelerate the training process, skip connections were employed between layers and residual blocks. The experiments were carried out using 2D images from the BraTS 2017 dataset, and the model performed well in terms of execution time. In another study by Zhang et al. [18], the authors introduced the modified U-Net architecture with residual attention blocks in the U-Net model. The attention module successfully enhanced the local feature extraction and improved the overall performance of the model. Using skip connections and focusing on salient features information proposed technique accomplished high results on the BraTS 2017, 2018, and 2019 datasets by using 2D images taken from the axial view from a 3D image. Although the proposed 2D model performed well, but it loses contextual information among different slices. To take the contextual benefit from 3D MRI images, M Ghaffari et al. [3] proposed a 3D CNN architecture for segmenting brain tumor sub-regions built on an improved U-Net. At the encoder part researcher utilized the residual blocks that consist of $3 \times 3 \times 3$ convolution and group normalization followed by ReLU activation and identity skip connection. The residual blocks at the encoder end help to learn non-linear residual that provides a deeper architecture to improve the learning process. To make the decoder part lighter, the researcher utilized the self-ensembling approach. This technique helps in reducing the number of feature maps at each level of the decoder part of the network. The proposed model needed more powerful hardware and computational resources. Saqib Qamar et al. [19] presented a hyperdense inception 3D U-Net model where hyperdense connections were used in each residual inception block to learn more complex information from MR images. The proposed model worked as an encoder-decoder model that uses skip

connections to combine feature maps for better segmentation results. The proposed model was validated on the BraTS 2020 dataset. To improve the accuracy, Isensee et al. [20] introduced an improved nnU-Net model. The researchers introduced post-processing and region-based training and data augmentation in the proposed architecture. In region-based training, all three classes of brain tumor were directly optimized instead of an individual class, which significantly improves the performance of the architecture.

The majority of state-of-the-art methodologies for BTS have focused on improving their performance on widely used BraTS benchmark datasets. Despite the success of current segmentation models, most notable U-Net models do not fully exploit 3D volumetric scans of brain tumors because of limited computational resources. Most of the studies used 2D slices/images and a patch-wise strategy for the training of deep learning models [21,22,23]. Moreover, the gradient of low-level features becomes zero while going deeper in a network affects the segmentation results. Most of the existing studies used the U-Net model for the segmentation of MRI images to segment the brain tumor, but the U-Net model spends a lot of time on its execution due to its complex feature extraction way. Furthermore, because of the heterogeneous appearance, shape, and size of the gliomas, segmentation of sub-regions is still challenging [4,9]. Different studies in the literature have also proved the significant improvement in segmentation results by modifying the U-Net model's architecture instead of using the original model. To address these challenges found in the literature, in this research work, a novel end-to-end robust deep learning-based encoder-decoder model is proposed to enhance the performance of the BTS. The proposed approach utilizes the residual network with a 3D U-Net model to overcome the issue of gradient vanishing (this problem arises when the gradient becomes very small during the training of the deep model) by preserving low-level features and transmits them to the corresponding decoder level using adaptive skip connections and helps make the training process smoother and faster. Furthermore, the proposed framework takes benefits from the 3D volume and by using 3D convolution utilizes the complete contextual information from multimodal MRI.

3. Research methodology

In this section, details about the dataset, preprocessing steps, and implementation detail of the proposed method are discussed. Further, this section discusses the proposed deep residual U-Net (dResU-Net) architecture and details of the loss functions used during the training phase. The general flow of the proposed methodology is given in Fig. 1.

3.1. Dataset

The proposed dResU-Net model is trained and evaluated on the publicly available benchmark BraTS 2020 challenge dataset. This dataset is taken from the “*Medical Image Computing and Computer-Assisted Intervention (MICCAI) Multimodal Brain Tumor Segmentation*

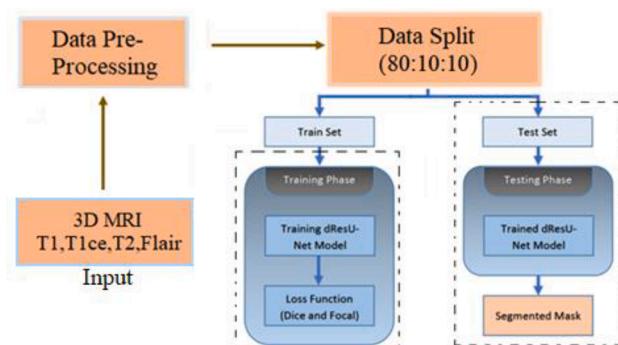


Fig. 1. The Proposed Methodology's General Flow.

Challenge (BraTS) 2020”, collected by medical professionals from the “University of Pennsylvania and UPenn’s Center for Biomedical Image Computing and Analysis (CBICA)” [24]. It contains 3D MRI brain scans from 369 glioma patients, out of which 76 are LGG patients, and the remaining images belong to HGG patients. Each image has a dimension of 240×240 , and there is a total of 155 slices in each 3D scan. Each patient has four different MRI modalities: T2, T1ce, T1, and FLAIR. Fig. 2 represents the sample subjects from the benchmark BraTS 2020 dataset with all four modalities and ground truth. One of the primary visual differences in these modalities is the quantity of the water content in the brain, i.e., cerebrospinal fluid and intensities of tissues. The ground truth segmentation file is manually annotated by one to four neuroradiologists.

Four main classes are associated with this dataset which are:

- Background (Label 0)
- Necrosis and non-enhancing tumor (Label 1)
- Edema (Label 2)
- Enhancing tumor (ET) (Label 4)

Label 4 (ET) has been changed to label 3 in the implementation. The dataset is manually split (class balanced) into 80% of the training data, 10% of the validation data, and 10% of the testing data of both LGG and HGG patients. The distribution of the subjects is given in Table 1.

3.2. Pre-Processing

The dataset (BraTS 2020) has already been passed through different preprocessing phases by the competition organizers before sharing publicly. All the images are already co-registered, skull-stripped, aligned into a common space, and have isotropic resolution [24]. MRI images may contain intensity due to the inhomogeneity of the magnetic field of different scanners. Therefore, data must be pre-processed before it is given to the model for training to improve the segmentation results. Image standardization and normalization were performed on all MR images by subtracting the mean from each voxel and dividing it by their standard deviation, resulting in zero mean and unit variance in each brain image, which is usually called *Z-score normalization* [11]. The equation of image normalization is:

Table 1
Distribution of subjects in the train, validation, and test sets.

BraTS 2020	Train Set		Validation Set		Test Set	
	LGG	HGG	LGG	HGG	LGG	HGG
Patient Distribution	62	233	7	30	7	30
Total Patients	295		37		37	

$$I_{\text{norm}} = \frac{I - \mu_i}{\sigma_i}$$

where I_{norm} and I are normalized input and original input images respectively. Whereas μ_i represents mean value and σ_i represent standard deviation of the input image. Furthermore, all the MR images are resized to $128 \times 128 \times 128$ dimensions from $240 \times 240 \times 155$ due to the limited size of memory. To take advantage of comprehensive information present in four sequences, all the four modalities were stacked together. Training examples with the dimension of $128 \times 128 \times 128 \times 4$ (where 4 represents the four modalities i.e., T1, T1ce, T2, FLAIR) were fed to the model as input at the training time.

3.3. Network architecture

U-Net is one of the most popular architecture for biomedical semantic image segmentation. The architecture of the U-Net gained popularity due to its local and global feature extraction method with varying scales. In addition, it transfers the feature map of each encoder level to its corresponding decoder level using the skip connection so that while deciding the segmentation mask, the classifier can take into account both low-level (that contained information about boundaries and edges) and high-level (that contained information about the object and shape) features. Despite the success of the U-Net model, some limitations make the training process complicated. First of all, while going deeper in a network, the gradient vanishing problem arises during the training because as training proceeds, the gradient norm of the early layers almost approaches to zero. Furthermore, low-level features are equally important when deciding on segmentation masks as they contain information about the boundary, edges, and location of the tumorous region that can help in getting good segmentation results. Although, high-

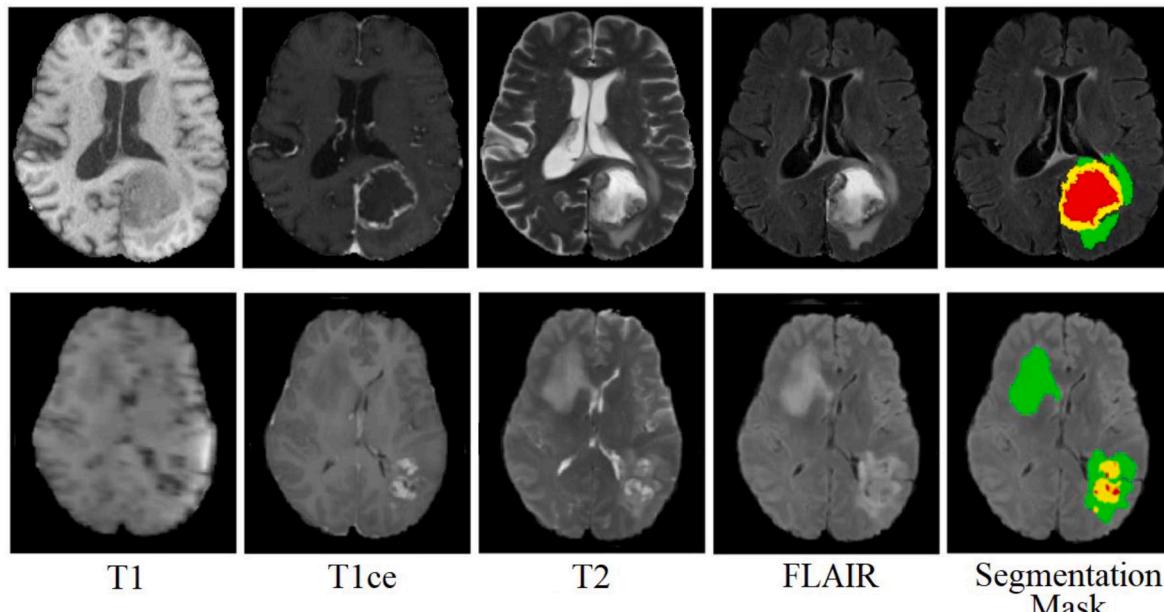


Fig. 2. Sample MRI images and their ground truth with all four modalities (green for Edema, red for Necrosis, and yellow for Enhancing tumor).

level and low-level features are transmitted to the decoder part through skip connection at each level, but while going deeper in a network, during the down-sampling operation, the typical U-Net network has more richer features than low-level features. Therefore, local details and location details get lost in the preceding layers due to the convolution and non-linearity operation.

To overcome the issue of vanishing gradient in the encoder part of the U-Net model during the training of the deeper network, this research introduced an end-to-end encoder-decoder-based deep residual U-Net model (dResU-Net) for the BTS. This model work as a hybridization of residual network and U-Net network by utilizing residual blocks with shortcut connections to preserve low-level features at the encoder path. To preserve low-level features the proposed model used residual blocks at the encoder part of the U-Net model. Residual blocks used unique adaptive skip connections [25] to preserve low-level features and deal with the issue of vanishing gradient problems within the encoder part. The primary motivation behind using the identity mapping (shortcut) connection is to pass the activation of the earlier layer to the later layer directly and combine the input of the residual convolutional block into the output through the concatenation operator. Furthermore, the residual network makes the information transmission smoother and speeds up the model convergence towards the global minima. The proposed architecture of deep Residual U-Net is shown in Fig. 3.

The architecture of the dResU-Net follows a structure of a 3D U-Net [26] like pattern that contains contracting (encoder) and expanding (decoder) part that are interconnected by skip connections. The contracting path contains five levels (where the level denotes the network's depth). The encoder part of the proposed model contains residual blocks (Yellow blocks) at the first four levels for feature extraction. The input to the encoder path of the proposed model is $128 \times 128 \times 128 \times 4$ vowel with the stack of four modalities (T1, T2, T1ce, Flair). Each level comprises a residual block followed by max pool 3D with pool size (2, 2, 2), stride (2, 2, 2), and dropout. Each block of the residual block contains three components and identity mapping. Components 1 and 2 use batch normalization and ReLU activation function after conv3D layers. Component 3 is similar to the preceding two. However, it lacks the ReLU activation function. The input and shortcut are concatenated before

applying the ReLU function. Identity blocks are used in a residual block can be defined by Eq. (i) [27]:

$$b = F(a, \{w_i\}) + a \quad (i)$$

where "a" denotes the input and "b" is for the output vector of the corresponding layer. The function " $F(a, \{w_i\})$ " represents the trained residual mapping that makes the same dimension of input "a" and function "F", whereas " w_i " represents the weights across to corresponding residual unit at the i th layer. The residual block is shown in Fig. 4.

The higher levels have higher feature representation but lower spatial resolution and vice versa. A down-sampling function is employed in the encoder path that reduces the image size, and after that bottleneck, layer is placed. The bottleneck and decoder part contains plain convolutional blocks (dark brown blocks) for predicting segmentation masks. The expanding (decoder) path mainly focuses on recovering the image into its original shape by using the up-sampling or Conv3DTranspose function. The main purpose of the up-sampling path is to find exact segmentation combined with corresponding data from the down-sampling path. The contextual information is transmitted from the contracting path to expanding path using a skip connection. The expanding path resembles the original U-Net model that contains the conv3DTranspose layer that concatenates the up-sampling path with its corresponding down-sampling path. After that, convolutional blocks are being utilized with a 0.1 dropout regularizer. Each convolution block in expanding path contains two conv3D layers followed by batch normalization and ReLU activation function. The final segmentation is performed by a $1 \times 1 \times 1$ convolution layer followed by a softmax activation function between the target classes. The output of the proposed model from the decoder path is $128 \times 128 \times 128 \times 4$ where 4 represents the total no. of the classes i.e., background, WT, TC, and ET.

3.4. Loss function

The efficiency and performance of a deep learning model are determined not only by the architecture but also by the loss function

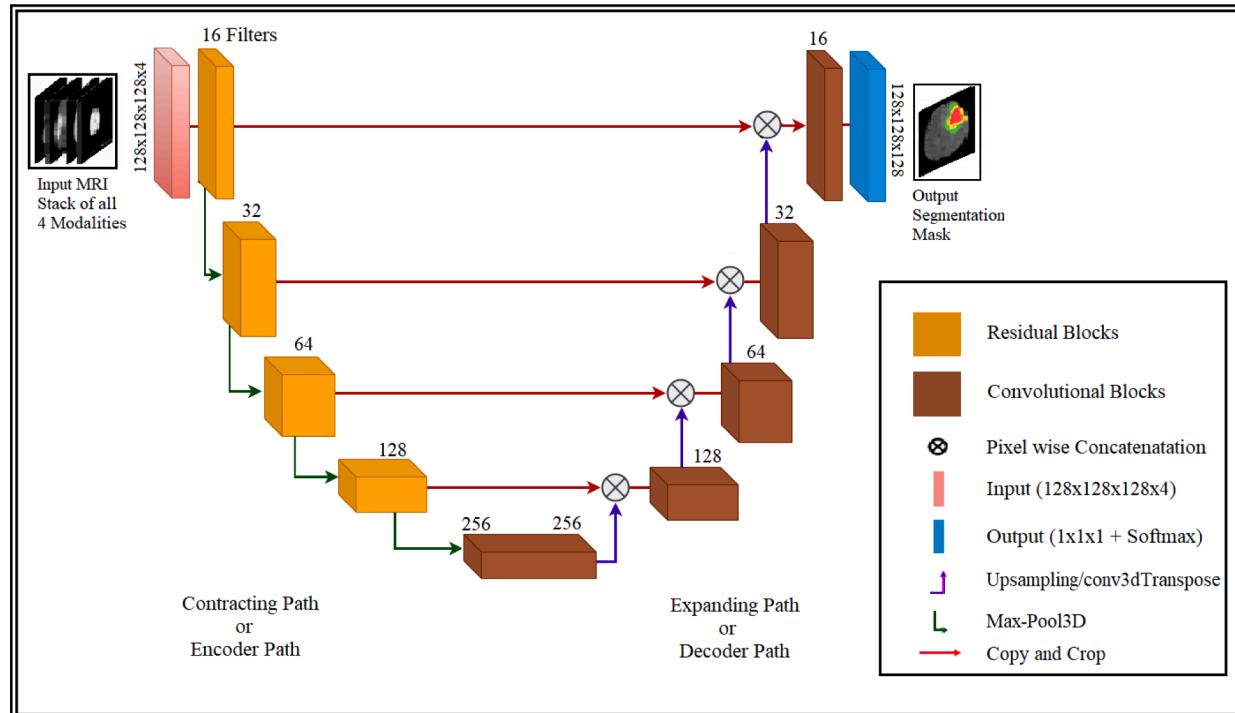


Fig. 3. The proposed architecture of 3D dResU-Net.

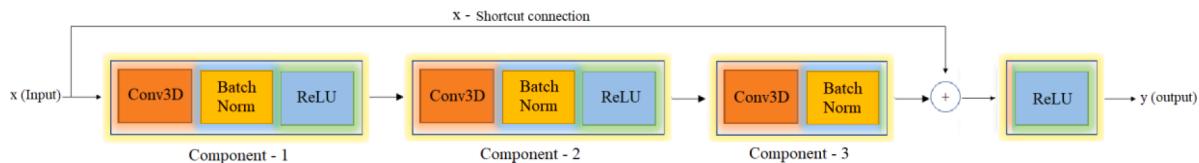


Fig. 4. Residual block with shortcut identity connection [28].

chosen [29]. BTS problem is a high-class imbalance problem where approximately 61% portion of the tumorous region belongs to a whole tumor, 24% from tumor core, and 15% from enhancing tumor class. Therefore, selecting the most suitable loss function is crucial for the effectiveness of the model. To overcome the challenge of the class imbalance problem, this research utilized the combination of focal loss [30] and soft dice loss [31].

Multiclass focal loss for class imbalance problem can be calculated by Eq. (ii).

$$L_{focal}(x, y) = -\frac{1}{N} \sum_{n=1}^N \sum_{c=1}^C (1 - x_{nc})^y y_{nc} \log x_{nc} \quad (\text{ii})$$

where x denotes predicted output and y denote ground truth. C refers to the class and n represents the total no. of the classes.

Multiclass soft dice loss can be calculated by Eq. (iii).

$$L_{dice}(x, y) = 1 - \frac{1}{N} \sum_{c=1}^C \frac{2 \sum_{m,n} x_{cmn} y_{cmn} + \epsilon}{(\sum_{m,n} x_{cmn}^2 + \sum_{m,n} y_{cmn}^2) + \epsilon} \quad (\text{iii})$$

where x = predicted output, y = ground truth, m = voxel, c = class, n refers to total no. of the classes and ϵ is a small number that is added to avoid divisible by zero.

The combination of these losses given in Eq. (iv) is used in the experiment.

$$L_{\text{comb}} = L_{dice} + L_{focal} \quad (\text{iv})$$

The combined loss function is used because of two reasons: The soft dice loss function is used to achieve maximum overlap between the predicted output and segmented mask irrespective of the class. The focal loss is used to down-weight the contribution of easy class and pay more attention to the challenging classes. It also classifies tissue cells regarding their assigned classes.

4. Results and analysis

In this section, the details about the evaluation measures employed to assess the performance of the proposed model, implementation details, results achieved from the proposed method, and comparison with the state-of-the-art methods are discussed.

4.1. Evaluation measure

Dice coefficient score (DSC), specificity, and sensitivity are used to assess the proposed approach. Dice score is the most widely used evaluation metric for the problem of BTS. It measures the overlap region between the predicted segmentation mask and ground truth by penalizing the false negative and false positive. The range of the dice score is between 0 and 1; the higher the dice score value better the segmentation results. Due to clinical applications, the model's performance is evaluated by separating the tumor structure into three sub-regions: WT contains all three cancerous regions, TC includes all tumors except edema, and ET is mostly found in HGG cases and visible in T1ce modality. For a given predicted and ground truth segmentation X and Y , where each tumor voxel is labeled as 1, and non-tumor voxels are 0, the dice score can be calculated as given in equation (v).

$$\text{Dice Score}(X, Y) = 2 \frac{|X \cap Y|}{|X| + |Y|} \quad (\text{v})$$

Sensitivity (True positive rate) represents the ratio between true positive pixels and correctly predicted pixels, whereas specificity (True negative rate) is the percentage of predicted true negative pixels. These measures aid in determining if the tumor regions are over or under-segmented by the model. The sensitivity and specificity can be calculated by the given formulas in Eqs. (vi) and (vii) respectively.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (\text{vi})$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (\text{vii})$$

where TP, FN, TN, and FP refer to a true positive, false negative, true negative, and false positive respectively.

4.2. Implementation detail

The proposed model (dResU-Net) was implemented using Python programming language, Keras library, and TensorFlow as backend. For the experimental purpose, an ADAM optimizer with a learning rate of 0.0001 was used. The activation function ReLU with batch normalization was employed. Batch normalization normally increases the stability of the model and normalizes the network at each layer. The model was trained for 100 epochs on a batch size of 4 due to the limited computational resources. The experiments were conducted on the BraTS 2020 benchmark dataset, from which 80% of data was used for training, 10% data for testing, and 10% data for validation. The model training was carried out on Tesla T4 GPU and 25 GB RAM provided by Google Colab Pro. A number of the experiments were performed on the proposed method to find out the best combination of the hyperparameters. In tuning the hyperparameters, the experiments were started by using smaller filters to collect as much local information as possible and then gradually increased the filter width to reduce the generated feature space width to have more representative information. The dropout layer is used during training, initially, the dropout rate was set to 0.3 and by empirically tuning the dropout value, a 0.1 dropout rate was found best for the final experiment. Moreover, the learning rate was also empirically tuned in different experiments to find out the best learning rate value, starting with the larger learning rate and gradually decreasing the value of the learning rate this approach helps us to reach the global minima much more quickly. The details about the hyperparameter

Table 2
Hyperparameters of the proposed dResU-Net model.

Hyperparameters	Value
Input size	$128 \times 128 \times 128 \times 4$
Learning rate	0.0001
Batch size	4
The hidden layer activation function	ReLU
Optimizer	ADAM
Loss function	Combination of focal and dice loss
No. of epochs	100
Dropout	0.1
Output size	$128 \times 128 \times 128 \times 4$
Output layer activation function	Softmax

setting that was set during the training of the model are given in Table 2.

4.3. Analysis of the results

The proposed dResU-Net was trained and validated on the BraTS 2020 training dataset as the ground truth for the validation data is not available. Therefore, the model is trained on 80% data, validated on 10%, and tested on 10% of the total data. The trained model has been utilized to segment the test images of brain tumors. The proposed method generates a 3D volume of the segmentation mask, which includes tumor regions WT, ET, and TC. Fig. 5 depicts the qualitative results of an MRI image viewed in the axial plane with slices selected at random. The visual depiction confirms that the results are quite close to the WT, CT, and ET ground truth values.

In Table 3, the detail about the dice coefficient values for TC, WT,

Table 3

Quantitative results of train and test set of the proposed architecture on the BraTS 2020 dataset.

Results (BraTS –2020)	Metrics	Result		
		Tumor Core (TC)	Whole Tumor (WT)	Enhancing Tumor (ET)
Train Set	Dice Score	0.9212	0.9212	0.8097
	Sensitivity	0.9868	0.9900	0.9802
	Specificity	0.9864	0.9970	0.9851
Test Set	Dice Score	0.8357	0.8660	0.8004
	Sensitivity	0.9695	0.9786	0.9648
	Specificity	0.9862	0.9936	0.9791

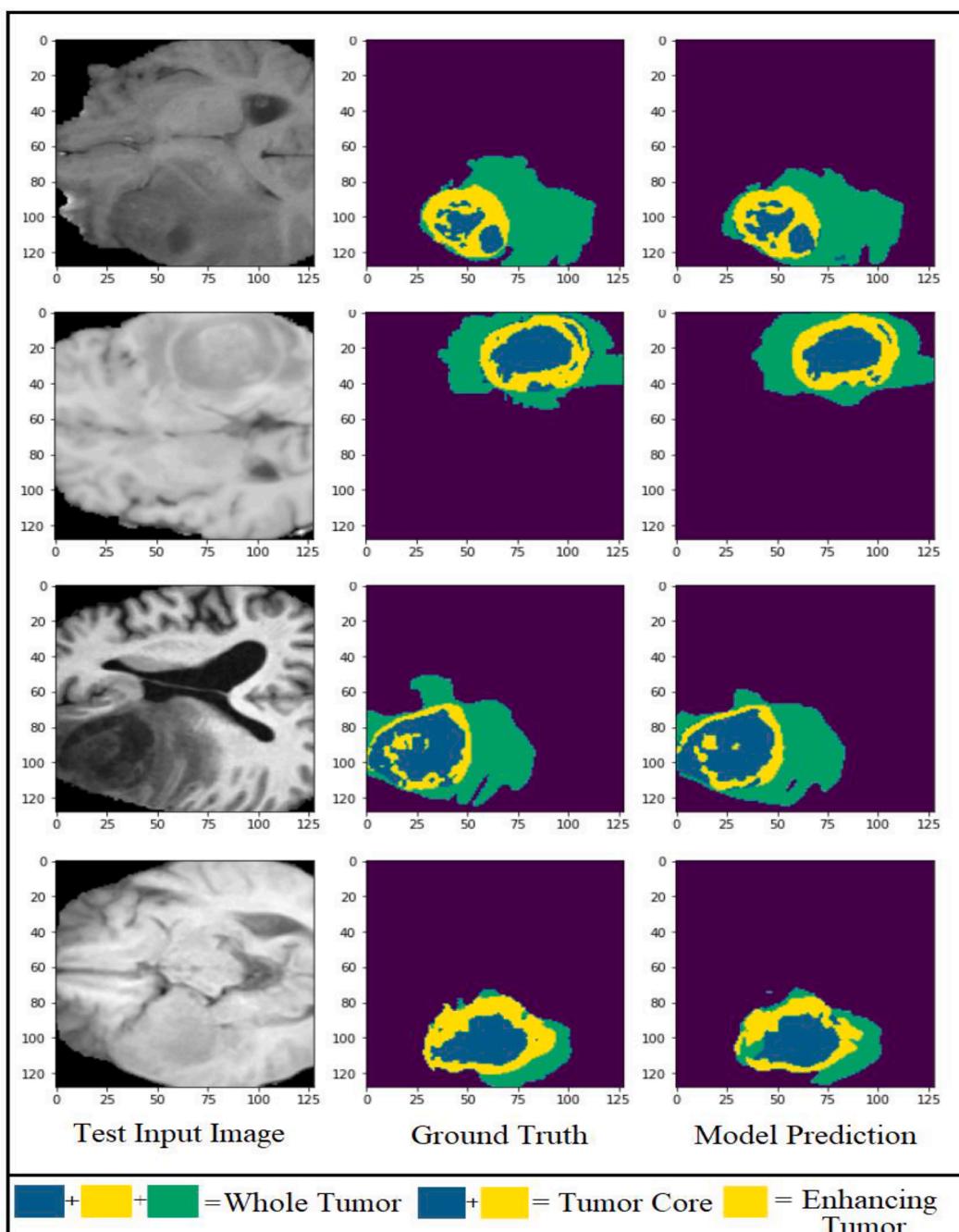


Fig. 5. Qualitative results of the BTS on testing MRI sequences at the axial axis.

and ET of the training and test sets as a result of the proposed approach on the benchmark BraTS 2020 dataset is given.

The dice score per epoch of all the tumorous classes TC, WT, and ET are shown in Fig. 6. The training loss per epoch and validation loss is shown in Fig. 7.

Although the U-Net model has successfully achieved good results, in this study, a modified U-Net model with the residual network is introduced to deal with the challenge of BTS. The proposed model dResU-Net has five levels. The encoder of the model is based on the residual network. While experimenting with the proposed architecture, first of all, two levels of the U-Net model's convolutional layers were replaced with residual convolutional blocks, but there were no significant improvements in the results. After that by employing residual blocks to all levels of the encoder path, significant improvements were seen in the results but the model tends to overfit and the loss of the model starts increasing after a few epochs. As the bottleneck layer of the model has a compressed representation, the aim of the bottleneck layer is that the compressed view should only contain "valuable" data for reconstructing the input (or segmentation map). Therefore, in the final experiments, residual blocks were used only in the first four-level of the proposed model while others remain the same as the U-Net model with the batch normalization and ReLU activation function. The residual convolutional blocks used adaptive skip connections within the encoder part of the U-Net that helps in preserving the low-level feature details (i.e., boundary or edges of the tumorous region) with high-level features. The usage of the skip connection in encoder proves the beneficial to improve the results in this case. Fig. 8 also represents the box plots (box represents the median, whiskers – min/max, first/third quartile, and circles represent the outliers) of segmentation accuracy according to the Dice metric, among the three subregions, WT achieves the best performance compared to TC and ET. ET is much more difficult to segment in some cases with a much higher standard deviation. On the x-axis brain tumor sub-regions i.e., WT, TC, and ET are given, while on the y-axis the value of the dice score is given which is from 0 to 1.

4.4. Comparison with state-of-the-art studies

The proposed model (dResU-Net) is compared to state-of-the-art models for brain tumor semantic segmentation. In comparison to existing segmentation methods, the proposed method is able to segment these individual tumor regions (WT, TC, and ET) nearer to the ground truth, as shown by the findings. The enhancing tumor and its dispersion with necrosis were the most difficult areas to segment. Many existing segmentation models performed poorly when it came to segmenting enhancing tumor and tumor core. The proposed dResU-Net model, on the other hand, successfully segmented these regions. Table 4 also

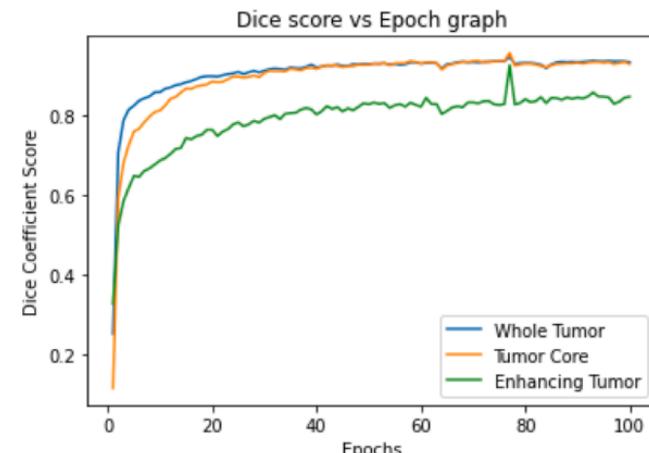


Fig. 6. Dice score of the tumorous classes TC, WT, and ET.

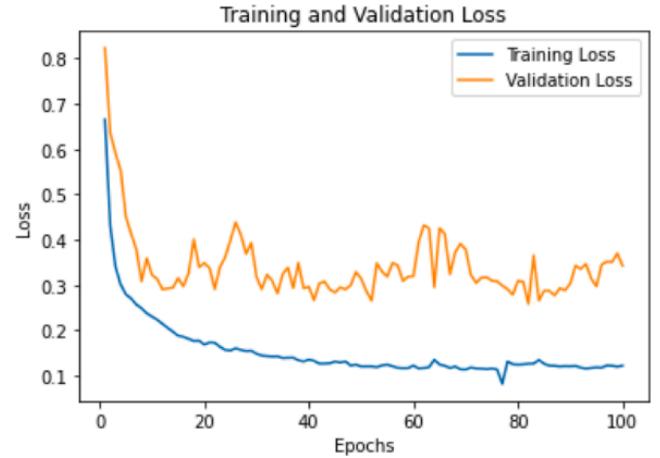


Fig. 7. Training and Validation loss.

depicts the quantitative performance of the proposed 3D dResU-Net, which outperforms the state-of-the-art techniques for tumor core and enhancing tumor classes. In the case of the whole tumor, the proposed method performed well and achieved comparable results with state-of-the-art techniques. Overall, the proposed approach appears to be a better technique for generating segmented images closest to the ground truth.

4.5. Validation on external dataset

According to the study [36], the evaluation of a model on an external dataset can provide an unbiased assessment of its performance. The model that performs well on the development set but performs poorly on the external dataset shows that the model has overfitted the development dataset. Such models are of little use in real-world settings. Therefore, the proposed model was cross-validated on the benchmark BraTS 2021 dataset. A total of 50 patients were chosen at random from the benchmark BraTS 2021 dataset, ensuring that they had not been included in the BraTS 2020 dataset. The mean dice score accomplished from the proposed architecture on cross-validation was 0.8400, 0.8601, and 0.8221 for TC, WT, and ET, respectively. The quantitative results of external cohort validation are given in Table 5.

The qualitative results of MRI sequences (in the external cohort) viewed in the axial plane with slices selected at random from the dataset are shown in Fig. 9. The visual representation confirms that the results are quite close to the WT, TC, and ET ground truth values.

As seen from the cross-dataset validation results, the proposed model generalizes well on unseen data and achieves comparable results. That proves the robustness and effectiveness of the proposed model. The model takes advantage of the 3D image and performs well in segmenting the tumorous area, and the visual results are closer to the ground truth.

4.6. Model complexity

The proposed 3D dResU-Net model has 30.47 M parameters and 374.04G FLOPs which is almost a moderate size model. The size of the model is 334 MB even though in the proposed architecture all the 3D modalities (T1, T1ce, T2, and FLAIR) have been used at a time for the training purpose. The proposed model can be used in any real-world setting. By reducing the layers in the proposed model in a straightforward way and by using each MRI modality separately the complexity of the proposed model can be reduced and make it a lightweight model. Although with 30.47 M parameters the proposed model achieved dice scores of 0.8357, 0.8660, and 0.8004 for TC, WT, and ET, respectively on the BraTS 2020 dataset. Compared with 3D U-Net [26] which has 19.06 M parameters and 1670G FLOPs, our proposed architecture

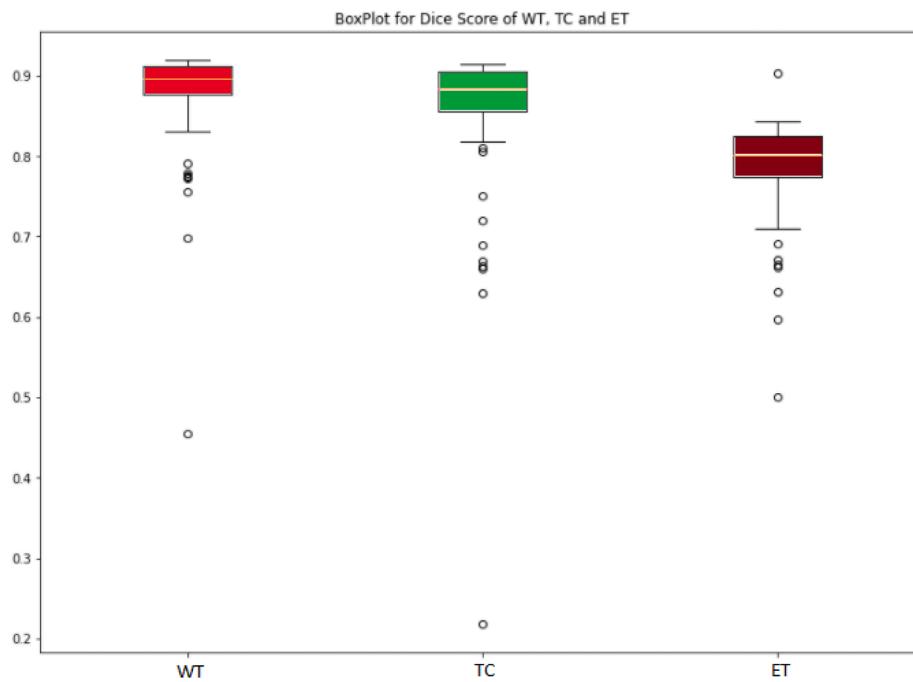


Fig. 8. Boxplot of the Dice metric for proposed dResU-Net model evaluated on the WT, TC, and ET.

Table 4

Comparison of proposed dResU-Net model with state-of-the-art BTS methods on the BraTS 2020 dataset.

Reference/Study	Architecture	Image Dimension	Dataset	Dice Score (DSC)		
				Tumor Core (TC)	Whole Tumor (WT)	Enhancing Tumor (ET)
W. Wang et al. [12]	TransBTS	128 × 128 × 128	BraTS 2020	0.8173	0.9009	0.7873
Ö. Çiçek et al. [26]	3D U-Net	128 × 128 × 128	BraTS 2020	0.7906	0.8411	0.6876
J. Colman et al. [32]	2D Deep Residual U-Net	2D Slices with 240 × 240	BraTS 2020	0.7983	0.8673	0.7514
H. Messaoudi et al. [34]	Modified 3D U-Net by embedding and re-using any 2D classifier network	192 × 160 × 108	BraTS 2020	0.7520	0.8068	0.6959
L. M. Ballestar et al. [33]	3D encoder-decoder based V-Net model	64 × 64 × 64	BraTS 2020	0.7526	0.8463	0.6215
M. Ghaffari et al. [3]	Lightweight decoder and heavy encoder based modified 3D U-Net with residual and dense blocks	128 × 128 × 128	BraTS 2020	0.82	0.90	0.78
F. Wang et al. [35]	3D U-Net with Brain-wise patching strategy	128 × 128 × 128	BraTS 2019	0.798	0.852	0.778
J. Zhang et al. [18]	Attention Guided Residual U-Net	128 × 128 × 128	BraTS 2019	0.777	0.870	0.709
A. Myronenko [37]	Encoder decoder based variational auto-encoder	160 × 192 × 128	Brats 2018	0.8154	0.8839	0.7664
Yang et al. [38]	Residual 3D U-Net with Adversarial Training	96 × 96 × 96	BraTS 2018	0.789	0.869	0.722
Proposed Model	Deep Residual U-Net (DResU-Net)	128 × 128 × 128	BraTS 2020	0.8357	0.8660	0.8004

Table 5

Cross-Validation results on BraTS 2021 benchmark dataset.

Results (BraTS -2021)	Metrics	Result		
		Tumor Core (TC)	Whole Tumor (WT)	Enhancing Tumor (ET)
Cross-Validation (50 Patients)	Dice Score	0.8400	0.8601	0.8221
	Sensitivity	0.9781	0.9817	0.9713
	Specificity	0.9836	0.9945	0.9745

dResU-Net shows great superiority in terms of model complexity. The execution time of our proposed model was 14.75 minutes per epoch and the proposed model was trained for 100 epochs taking approximately 25 hours to complete the training process. The online test time of our proposed model was just 8 seconds for one subject which means that it can be deployed in a real-world clinical setting. By using more powerful computational resources the performance of the model can be further enhanced.

4.7. Ablation study

The proposed technique achieved a higher dice score for each sub-region of brain tumor than the baseline work because of the use of the

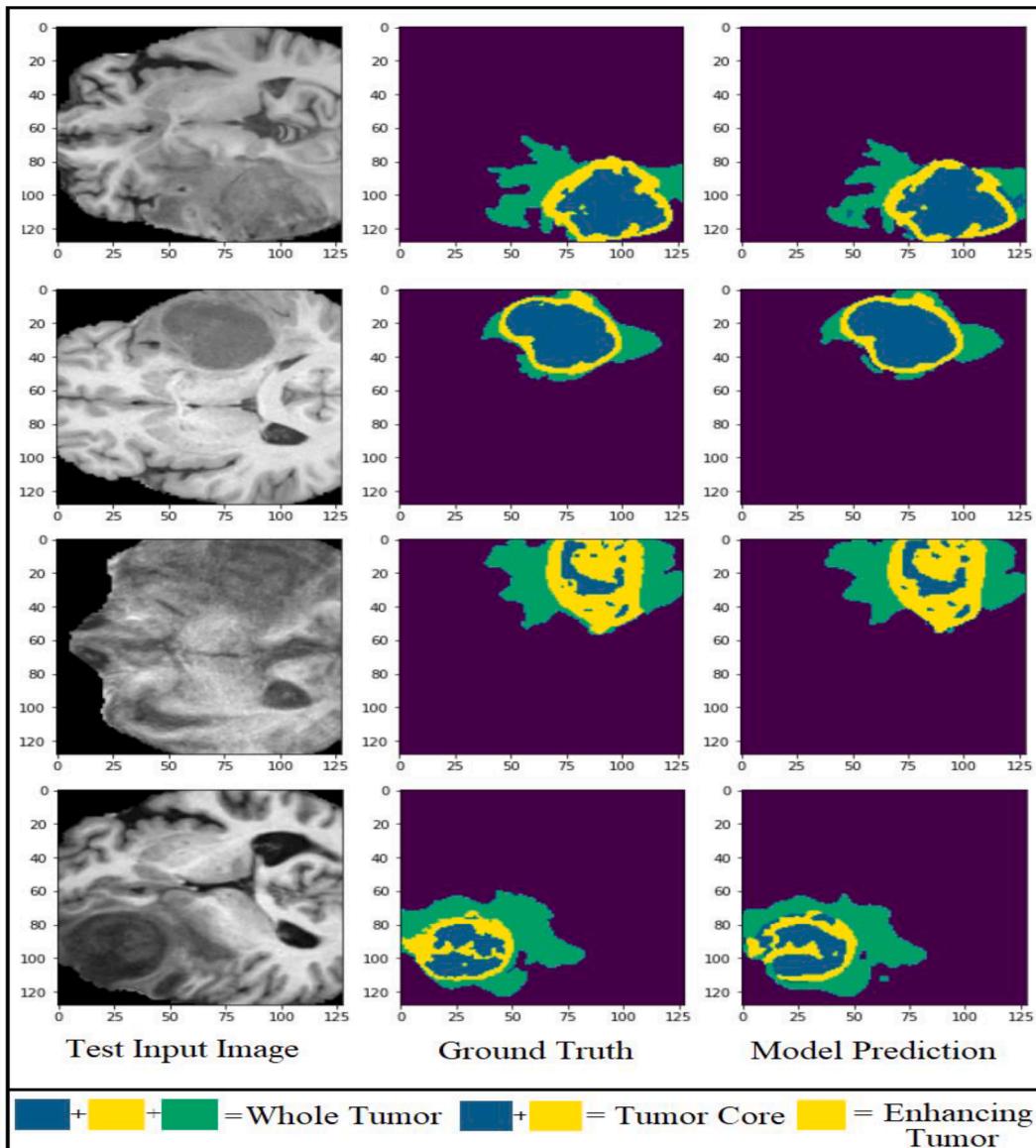


Fig. 9. Qualitative results of the BTS on external cohort MRI sequences at the axial axis.

residual convolutional blocks at the encoder part. Residual convolutional blocks utilized the adaptive skip connections within the encoder part of the proposed model to preserve low-level features with high-level features. The ablation study was conducted to assess the performance of the proposed dResU-Net model that contained adaptive skip connections to preserve low-level features with high-level features to improve the segmentation results with a simple 3D U-Net model that does not contain adaptive skip connections. The influence of these two approaches based on the dice score of all three sub-regions TC, WT, and ET is shown in Table 6. In summary, the proposed dResU-Net model

significantly improves the segmentation accuracy of the ET class which was the most difficult class to segment.

5. Conclusion and future directions

This paper attempts to solve the problem of BTS on 3D MRI images. In this research study, a 3D Deep dResU-Net is proposed to improve the performance of the segmentation of brain tumors from MRI images. The proposed model successfully embedded the residual blocks with identity mapping in the encoder part of the U-Net model to improve the learning process by preserving the local feature response and passing it through the activation from the appropriate level of the encoder side to the decoder side by using skip connections. The proposed model is designed to improve the overall training process and to overcome the vanishing gradient problem. The proposed model was evaluated using the benchmark BraTS 2020 dataset. Experimental results demonstrate that the proposed method outperformed state-of-the-art models and obtained dice scores of 0.8357, 0.8660, and 0.8004 for TC, WT, and ET, respectively, without employing any kind of data augmentation and extensive post-processing. Moreover, to validate the robustness of the proposed architecture, the model was cross-validated on 50 patients randomly

Table 6

Results of the ablation study of the components of the proposed technique on the BraTS 2020 test set.

		TC	WT	ET
Baseline: 3D U-Net	<i>Plain 3D U-Net without adaptive skip connections. (Without preserving low-level features)</i>	0.8005	0.8452	0.7326
Proposed: 3D dResU-Net	<i>Residual blocks at the encoder part to preserve low-level features by using adaptive skip connections.</i>	0.8357	0.8660	0.8004

selected from BraTS 2021 dataset. The cross-validation results show that the model generalizes well on the external dataset. The dice score achieved through cross-validation are 0.8400, 0.8601, and 0.8221 for TC, WT, and ET, respectively. In the future, the performance of the proposed method can be improved by different kinds of augmentation techniques or by using a large benchmark dataset. Furthermore, the size of the dataset can also be increased by using synthetic data generation augmentation techniques. More 3D-based architectures should be explored which keep computational cost in consideration while utilizing the maximum possible contextual information. Moreover, the performance of selected systems can be improved by removing false positive rates using post-processing techniques. In addition, this work can be extended to clinically challenged medical imaging problems and other segmentation applications i.e., Liver tumor segmentation and Kidney tumor segmentation problems, etc.

6. Financial or Non-Financial interests

The authors have no relevant financial or non-financial interests to disclose.

CRediT authorship contribution statement

Rehan Raza: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization, Project administration. **Usama Ijaz Bajwa:** Conceptualization, Methodology, Validation, Formal analysis, Investigation, Writing – review & editing, Supervision. **Yasar Mehmood:** Validation, Investigation, Resources, Supervision. **Muhammad Waqas Anwar:** Writing – review & editing. **M. Hassan Jamal:** Writing – review & editing.

Declaration of Competing Interest

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

References

- [1] Z. Liu et al., “Deep Learning Based Brain Tumor Segmentation: A Survey,” 14(8), pp. 1–21, 2020, [Online]. Available: <http://arxiv.org/abs/2007.09479>.
- [2] A.P. Patel, J.L. Fisher, E. Nichols, F. Abd-Allah, J. Abdela, A. Abdellalim, H. N. Abraha, D. Agius, F. Alahdab, T. Alam, C.A. Allen, N.H. Anber, A. Awasthi, H. Badali, A.B. Belachew, A. Bijani, T. Bjørge, F. Carvalho, F. Catalá-López, J.-Y. Choi, A. Daryani, M.G. Degefa, G.T. Demoz, H.P. Do, M. Dubey, E. Fernandes, I. Filip, K.J. Foreman, A.K. Gebre, Y.C.D. Geramo, N. Hafezi-Nejad, S. Hamidi, J. D. Harvey, H.Y. Hassen, S.A. Hay, S.S.N. Iravani, M. Jakovljevic, R.P. Jha, A. Kasaeian, I.A. Khalil, E.A. Khan, Y.-H. Khang, Y.J. Kim, G. Mengistu, K. A. Mohammad, A.H. Mokdad, G. Nagel, M. Naghavi, G. Naik, H.L.T. Nguyen, L. H. Nguyen, T.H. Nguyen, M.R. Nixon, A.T. Olagunju, D.M. Pereira, G.D. Pinilla-Monsalve, H. Poustchi, M. Qorbani, A. Radfar, R.C. Reiner, G. Rosenthal, H. Safari, S. Safiri, A.M. Samy, S. Sarvi, M.A. Shaikh, M. Sharif, R. Sharma, S. Sheikhhahaei, R. Shirkooohi, J.A. Singh, M. Smith, R. Tabarés-Seisdedos, B.X. Tran, K.B. Tran, I. Ullah, E. Weiderpass, K.G. Weldegewrgs, E.M. Yimer, V. Zadnik, Z. Zaidi, R. G. Ellenbogen, T. Vos, V.L. Feigin, C.J.L. Murray, C. Fitzmaurice, Global, regional, and national burden of brain and other CNS cancer, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016, Lancet Neurol. 18 (4) (2019) 376–393, [https://doi.org/10.1016/S1474-4422\(18\)30468-X](https://doi.org/10.1016/S1474-4422(18)30468-X).
- [3] M. Ghaffari, A. Sowmya, R. Oliver, Automated Brain Tumour Segmentation Using Cascaded 3D Densely-Connected U-Net, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 12658 LNCS (2021) 481–491, https://doi.org/10.1007/978-3-030-72084-1_43.
- [4] E.-E. Mohd. Azhari, M.M. Mohd. Hatta, Z. Zaw Htike, S. Lei Win, Tumor detection in medical imaging: a Survey, Int. J. Adv. Inf. Technol. 4 (1) (2014) 21–30, <https://doi.org/10.5121/ijait.2013.4103>.
- [5] D.N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O.D. Wiestler, P. Kleihues, D.W. Ellison, The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary, Acta Neuropathol. 131 (6) (2016) 803–820, <https://doi.org/10.1007/s00401-016-1545-1>.
- [6] S. A. L. I. R. Al-qazzaz, “Deep Learning-based Brain Tumour Image Segmentation and its Extension to Stroke Lesion Segmentation,” 2020.
- [7] R. V. Tanneedi, P. Pedapati, and S. Johansson, “Brain tumour detection using HOG by SVM,” no. December, 2017, [Online]. Available: www.bth.se.
- [8] S. Bauer, R. Wiest, L.-P. Nolte, M. Reyes, A survey of MRI-based medical image analysis for brain tumor studies, Phys. Med. Biol. 58 (13) (2013) R97–R129, <https://doi.org/10.1088/0031-9155/58/13/R97>.
- [9] U. Baid, S. Talbar, S. Rane, S. Gupta, M.H. Thakur, A. Moiyadi, N. Sable, M. Akolkar, A. Mahajan, A novel approach for fully automatic intra-tumor segmentation with 3D U-net architecture for gliomas, Front. Comput. Neurosci. 14 (2020), <https://doi.org/10.3389/fncom.2020.00010>.
- [10] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, Cham, 2015, pp. 234–241.
- [11] Chen, W., Liu, B., Peng, S., Sun, J., Qiao, X. (2019). S3D-UNet: Separable 3D U-Net for Brain Tumor Segmentation. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds) Brainlesions: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. Lecture Notes in Computer Science(,), vol 11384. Springer, Cham. doi:10.1007/978-3-030-11726-9_32.
- [12] W. Wang, C. Chen, M. Ding, H. Yu, S. Zha, J. Li, TransBTS: multimodal brain tumor segmentation using transformer, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) vol. 12901 LNCS (2021) 109–119, https://doi.org/10.1007/978-3-030-87193-2_11.
- [13] P. Liu, Q. Dou, Q. Wang, P.A. Heng, An encoder-decoder neural network with 3D squeeze-and-excitation and deep supervision for brain tumor segmentation, IEEE Access 8 (2020) 34029–34037, <https://doi.org/10.1109/ACCESS.2020.2973707>.
- [14] Z. Huang, Y. Zhao, Y. Liu, G. Song, GCAUNet: A group cross-channel attention residual UNet for slice based brain tumor segmentation, Biomed. Signal Process. Control 70 (June) (2021) 102958, <https://doi.org/10.1016/j.bspc.2021.102958>.
- [15] J. Zhang, X. Lv, H. Zhang, B. Liu, AResU-Net: Attention residual U-Net for brain tumor segmentation, Symmetry (Basel) 12 (5) (2020) 1–15, <https://doi.org/10.3390/SYM12050721>.
- [16] D. Maji, P. Sigedar, M. Singh, Attention Res-UNet with Guided Decoder for semantic segmentation of brain tumors, Biomed. Signal Process. Control 71 (PA) (2022) 103077, <https://doi.org/10.1016/j.bspc.2021.103077>.
- [17] M.K. Abd-Ellah, A.A.M. Khalaf, A.I. Awad, H.F.A. Hamed, TPUAR-net: two parallel U-net with asymmetric residual-based deep convolutional neural network for brain tumor segmentation, Springer International Publishing, 2019.
- [18] J. Zhang, Z. Jiang, J. Dong, Y. Hou, B. Liu, Attention Gate ResU-Net for Automatic MRI Brain Tumor Segmentation, IEEE Access 8 (2020) 58533–58545, <https://doi.org/10.1109/ACCESS.2020.2983075>.
- [19] S. Qamar, P. Ahmad, L. Shen, HI-Net: Hyperdense Inception 3D UNet for Brain Tumor Segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) vol. 12659 LNCS (2021) 50–57, https://doi.org/10.1007/978-3-030-72087-2_5.
- [20] F. Isensee, P.F. Jäger, P.M. Full, P. Vollmuth, K.H. Maier-Hein, nnU-Net for Brain Tumor Segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) vol. 12659 LNCS (2021) 118–132, https://doi.org/10.1007/978-3-030-72087-2_11.
- [21] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, H. Larochelle, Brain tumor segmentation with Deep Neural Networks, Med. Image Anal. 35 (2017) 18–31.
- [22] M.I. Razzaq, M. Imran, G. Xu, Efficient Brain Tumor Segmentation with Multiscale Two-Pathway-Group Conventional Neural Networks, IEEE J. Biomed. Heal. Informatics 23 (5) (2019) 1911–1919, <https://doi.org/10.1109/JBHI.2018.2874033>.
- [23] M. Yaqub, J. Feng, M. Zia, K. Arshid, K. Jia, Z. Rehman, A. Mehmood, State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images, Brain Sci. 10 (7) (2020) 1–19, <https://doi.org/10.3390/brainsci10070427>.
- [24] B.H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, L. Lanczi, E. Gerstner, M.-A. Weber, T. Arbel, B.B. Avants, N. Ayache, P. Buendia, D.L. Collins, N. Cordier, J.J. Corso, A. Criminisi, T. Das, H. Delingette, C. Demiralp, C.R. Durst, M. Dojat, S. Doyle, J. Festa, F. Forbes, E. Geremia, B. Glocker, P. Golland, X. Guo, A. Hamamci, K. M. Iftekharuddin, R. Jena, N.M. John, E. Konukoglu, D. Lashkari, J.A. Mariz, R. Meier, S. Pereira, D. Precup, S.J. Price, T.R. Raviv, S.M.S. Reza, M. Ryan, D. Sarikaya, L. Schwartz, H.-C. Shin, J. Shotton, C.A. Silva, N. Sousa, N. K. Subbanna, G. Szekely, T.J. Taylor, O.M. Thomas, N.J. Tustison, G. Unal, F. Vasseur, M. Wintermark, D.H. Ye, L. Zhao, B. Zhao, D. Zikic, M. Prastawa, M. Reyes, K. Van Leemput, The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2015) 1993–2024, <https://doi.org/10.1109/TMI.2014.2377694>.
- [25] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. vol. 2016-Decem (2016) 770–778, <https://doi.org/10.1109/CVPR.2016.90>.
- [26] Ö. Çicek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: Learning dense volumetric segmentation from sparse annotation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) vol. 9901 LNCS (2016) 424–432, https://doi.org/10.1007/978-3-319-46723-8_49.
- [27] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: 32nd Int. Conf. Mach. Learn. ICML 2015, 2015, pp. 448–456.
- [28] L.H. Shehab, O.M. Fahmy, S.M. Gasser, M.S. El-Mahallawy, An efficient brain tumor image segmentation based on deep residual networks (ResNets), J. King Saud Univ. - Eng. Sci. 33 (6) (2021) 404–412, <https://doi.org/10.1016/j.jksus.2020.06.001>.
- [29] C.H. Sudre, W. Li, T. Vercauteren, S. Ourselin, M. Jorge Cardoso, Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes

- Bioinformatics) vol. 10553 LNCS (2017) 240–248, https://doi.org/10.1007/978-3-319-67558-9_28.
- [30] F. Milletari, N. Navab, S.A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016, 2016, pp. 565–571, <https://doi.org/10.1109/3DV.2016.79>.
- [31] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, Proc. IEEE Int. Conf. Comput. Vis. vol. 2017-Octob (2017) 2999–3007, <https://doi.org/10.1109/ICCV.2017.324>.
- [32] J. Colman, L. Zhang, W. Duan, and X. Ye, “DR-Unet104 for Multimodal MRI Brain Tumor Segmentation,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12659 LNCS, no. 2021, pp. 410–419, 2021, doi: 10.1007/978-3-030-72087-2_36.
- [33] L. M. Ballestar and V. Vilaplana, “Brain Tumor Segmentation using 3D-CNNs with Uncertainty Estimation,” pp. 1–11, 2020, [Online]. Available: <http://arxiv.org/abs/2009.12188>.
- [34] H. Messaoudi, et al., Efficient Embedding Network for 3D Brain Tumor Segmentation, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell.
- Lect. Notes Bioinformatics) vol. 12658 LNCS (2021) 252–262, https://doi.org/10.1007/978-3-030-72084-1_23.
- [35] F. Wang, R. Jiang, L. Zheng, C. Meng, B. Biswal, 3D U-Net Based Brain Tumor Segmentation and Survival Days Prediction, Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) vol. 11992 LNCS (2020) 131–141, https://doi.org/10.1007/978-3-030-46640-4_13.
- [36] A. Kleppe, O.J. Skrede, S. De Raedt, K. Liestøl, D.J. Kerr, H.E. Danielsen, Designing deep learning studies in cancer diagnostics, Nat. Rev. Cancer 21 (3) (2021) 199–211, <https://doi.org/10.1038/s41568-020-00327-9>.
- [37] A. Myronenko, 3D MRI brain tumor segmentation using autoencoder regularization, in: International MICCAI Brainlesion Workshop, 2018, pp. 311–320.
- [38] H.Y. Yang, J. Yang, Automatic brain tumor segmentation with contour aware residual network and adversarial training, in: International MICCAI Brainlesion Workshop, Springer, Cham, 2018, pp. 267–278.