

SER-Net: A Novel 2D Squeeze-and-Excite Residual UNet Architecture for Semantic Segmentation of Brain Tumors from MR Images

Sourjya Mukherjee, Ananya Bhattacharjee, and R. Murugan

Department of Electronics and Communications Engineering, National Institute of Technology Silchar, Silchar, Assam, 788010, India

Abstract—Brain cancer is a deadly disease. Brain tumor segmentation using 3D magnetic resonance imaging (MRI) is essential for the disease's diagnosis, surveillance, and therapy planning. Manual delineation techniques are expensive, time-consuming, labor-intensive, and subject to human error. Thus, there is a great need for the development of technologies capable of automatic segmentation of brain tumors. In this paper, a novel encoder-decoder based Squeeze and Excite Residual UNet (SER-Net) has been proposed, which can automatically segment the tumor subregions from multi-modal MRI slices. The performance of the proposed model was evaluated on the TCGA LGG dataset and the BraTS 2020 dataset. The proposed method achieved a dice score of 92.4% on the TCGA LGG dataset outperforming the current state-of-the-art dice score of 91.8%. On the BraTS 2020 dataset, an ensemble approach using the proposed architecture achieved a dice score of 92.4%, 87.7%, 85.2% for tumor subregions - whole tumor(WT), Tumor core(TC), and Enhancing Tumor(ET) respectively outperforming the current state-of-the-art dice scores of 91.8%, 87.3%, 83.3%. Hence, the proposed SER-Net model is apt for brain tumor segmentation.

Index Terms—Brain Tumor Segmentation, U-Net, Residual Networks, Squeeze and Excite Block.

I. INTRODUCTION

CANCER is an unchecked somatic cell proliferation brought on by an accumulating series of random mutations in critical genes that regulate cell growth and differentiation [1], [2]. A clinical tumor will develop when the mutation effectively avoids immunosurveillance [3]. Brain tumors are among the most deadly malignancies, despite their rarity. Depending on where they first developed, they can be classified as primary or secondary brain tumors. Primary tumors originate from brain cells whereas secondary ones originate from malignancies in other parts of the body that have metastasized to the brain. The majority of current brain tumor segmentation research revolves around a special type of tumor called glioma which originates from the glial cells in the brain. The average survival time for glioblastoma patients is less than 14 months [4]. Timely diagnosis of brain tumors is thus vital to ensuring appropriate treatment planning, surgery, and follow-up visits [4]. In addition to having some characteristics in common with tumors found elsewhere in

the body and presenting similar diagnostic and therapeutic obstacles, brain tumors also provide unique challenges because of the characteristics of the organ they are located in.

The blood-brain barrier (BBB), which separates the majority of the brain from the blood, controls which chemicals can travel through it far more tightly than most other organs do [5]. As a result, many tracers that easily reach tumors in other parts of the body would only reach brain tumors once the tumor has disrupted the BBB. Thus, the key diagnostic marker for brain tumors is the rupture of the blood-brain barrier (BBB), which is easily seen on contrast-enhanced MRI and computed tomography (CT) [5]. MRI is regarded as the industry-standard method for brain tumor detection because of its excellent soft tissue contrast and widespread accessibility. An extremely strong magnetic field and radio frequency signals are used in MRI, a non-invasive in vivo imaging technology, to stimulate target tissues and provide interior images of those tissues. However, the characteristics of brain tumors make correct segmentation difficult [6]. Usually, several complimentary 3D MRI modalities are acquired such as T1, T1 with contrast agent (T1-ce), T2, and Fluid Attenuation Inversion Recovery (FLAIR) to emphasize different tissue properties and areas of tumor spread. T1-ce modality has a bright tumor border, and T2 modality has a bright tumor area. On the other hand, the FLAIR scan aids in isolating the edema from the cerebrospinal fluid (CSF) [7]. Integrating data from different MRI modalities can thus help improve the region-wise segmentation of brain tumors.

The manual segmentation of medical images is a taxing and error-prone process that is known to have a low inter-professional agreement because of the quantity of brain tumor images that are currently being generated in clinics and the high variability in the location, size, and appearance of brain tumors [8]. Thus, computer-assisted tumor segmentation has become an increasingly sought-after solution to this problem. Computer-assisted methods [9], [10] can be broadly classified into two categories: semiautomatic brain tumor segmentation and automatic brain tumor segmentation. The semiautomatic method entails the expert providing some initial information, such as the tumor's location and precise delineation, and offloading the computational effort to a machine. The amount of time spent by human specialists is reduced by this method. But as MR images are frequently produced in clinics, completely

automated segmentation methods are necessary. The automatic segmentation methods require no human intervention and can segment the brain tumor into different classes such as necrotic tumor, enhancing tumor, tumor core, and edema. In recent years, convolutional neural networks (CNNs) have emerged as an effective tool for automatic medical image segmentation tasks, such as organ and lesion segmentation, and have attained previously unheard-of levels of accuracy in the segmentation of brain tumors. However, the problem of accurate and repeatable segmentation results and the characterization of anomalies is still a work in progress. In an attempt to overcome this, we describe in this paper a novel 2D convolutional neural network (CNN) based on the U-Net [11] framework. The proposed model was trained and evaluated on the BraTS 2020 [12], [13], [14] dataset and the TCGA LGG [15] dataset achieving exemplary segmentation results. Although the proposed architecture is based around the U-Net paradigm a multitude of architectural enhancements were made to the building blocks of the contractive as well as the expansive path leading to segmentation outputs that are on par with many of the state-of-the-art methods. An ensemble of three parallelly trained models (with the same architecture) has been implemented for the segmentation of each class: Whole Tumor(WT), Tumor Core(TC), and Enhancing Tumor(ET) respectively. This differs from many of the previous implementations of the U-Net model for the multiclass segmentation of multimodal brain MR images, where a single model is trained to produce the multiclass segmentation output. The (ensemble) model performs with great accuracy on each tumor class due to the ensemble approach, producing accurate segmentation outputs. The novelties and contributions associated with this work are:

- 1) A novel encoder-decoder based 2D CNN model for the segmentation of brain tumors is implemented, which achieved better results than most 3D state-of-the-art variants with a much lower computational cost.
- 2) A novel residual network (Res-block) design comprising of Batch-Normalization, ReLU Activation, Dropout and Convolutional layers has been proposed. The extensive utilization of this module (Res-block) in the encoder, decoder as well as across the long skip connections helps minimize performance degradation due to vanishing/exploding gradients and enhances the feature mapping across the encoder and decoder.
- 3) Squeeze and excite blocks have been implemented after each Res-block module in the encoder and decoder networks to improve the representational power of the proposed model by enabling it to perform dynamic channel-wise feature re-calibration.
- 4) Atrous spatial pyramidal pooling (ASPP) has been implemented into the encoder's bottleneck region to expand the proposed model's receptive field for capturing multi-scale information.

II. LITERATURE REVIEW

In this section, a number of previous works have been discussed that have provided valuable insights into the development of the proposed model. In the results section, the

proposed ensemble would be evaluated against some of these works.

Ronneberger et al.[11] developed the U-Net model based on the encoder-decoder paradigm for the segmentation of biomedical images. By analyzing the previous best models used in the Brats competition, the extensive use of UNet is undeniable. Following this trend, the proposed CNN was built around the U-Net framework. Sundaresan et al.[16] used a tri-planar architecture consisting of three 2D U-Nets, one for each plane (coronal, sagittal, axial), taking FLAIR, T1, T1-CE, and T2 slices as input channels. The ensemble was trained and evaluated on the BraTS 2020 dataset. The dataset was split into training-validation-testing datasets with a split ratio of 255-37-73 subjects. On the testing dataset, the tri-planar ensemble achieved a dice score of 0.833, 0.899, and 0.853 for the tumor subregions ET, WT, and TC respectively. Yanwu Xu et al.[17] incorporated a 3D ASPP module into a modified 3D U-Net leading to a considerable enlargement in the receptive field of the model enabling it to capture large-scale information without the introduction of extra parameters. Two variants of the aforementioned model- one with ASPP and one without -were trained and evaluated on the BraTS 2017 training dataset yielding results that quantitatively demonstrated the effectiveness of the ASPP module. The model proposed by Yanwu Xu et al.[17] achieved dice scores of 0.769, 0.871, and 0.779 on tumor classes ET, WT, and TC respectively on the testing set. Varghese et al[18] implemented a 23-layer deep FCNN based on the U-Net[12] framework for multiclass segmentation of brain tumors on 2-D slices extracted from patient volumes of the BraTS 2017 training dataset. This approach yielded a dice score of 0.84, 0.84, and 0.77 on the local testing set for the classes Whole tumor(WT), Tumor core(TC), and Enhancing tumor(ET) respectively. Ding et al.[19] introduced a novel 2D framework called Stack Multi-Connection Simple Reducing_Net(SMCSRNet) which is based on the stacked UNet paradigm. The Network is obtained by stacking basic UNet-like building blocks called Simple Reducing_Net(SRNet) and adding a series of bridge connections within the cascade. This method was trained and evaluated on the BraTS 2015 dataset and obtained dice scores of 0.83, 0.67, and 0.59 on tumor classes WT, TC, and ET respectively on the local testing set. This approach proved to be highly efficient due to the low computation cost but failed to obtain state-of-the-art results. Zhang et al.[20] improved the performance of the standard 2D UNet [11] by introducing residual building blocks to the original framework and adding gated attention units to the decoder. This approach yielded dice scores of 0.87, 0.77, and 0.72 on tumor classes WT, TC, and ET respectively on the local testing set of BraTS 2017 dataset, which is significantly higher than the scores obtained by the standard UNet[11]. Ilyas et al.[21] proposed a novel framework called the hybrid DANet which employs hybrid weight alignment with multi-dilated attention modules between encoder-decoder skip connections in order to facilitate improved feature mapping. This approach yielded dice scores of 0.88, 0.76, and 0.65 on tumor classes WT, TC, and ET respectively on the BraTS 2018 dataset. Ashraf et al.[22] proposed a novel 2D framework called the ZNet for

the semantic segmentation of low-grade glioblastomas. The proposed framework was trained and tested on the TCGA LGG dataset yielding a dice score of 0.915 for the testing set. Similar to the previous approach Santosh et al.[23] developed a residual UNet which was trained and tested on the TCGA LGG dataset yielding a dice score of 0.90 on the testing dataset. Sourodip et al.[24] Implemented a U-Net model with a VGG-16 backbone pre-trained on the imangenet dataset and trained it on the TCGA LGG dataset obtaining a dice coefficient of 0.916 on testing data. This approach stands as the current state-of-the-art for the TCGA LGG dataset. However, when it comes to the segmentation of high-grade gliomas (HGG) the effectiveness of this method is not known.

So far a number of unique approaches to brain tumor segmentation have been discussed and a number of inferences can be drawn in general. To summarize, it can be said that the majority of the works covered so far have either been trained and evaluated on datasets mostly made up of HGG volumes (an overwhelmingly large percentage of positive mask pixels in the BraTS datasets come from HGG volumes), or exclusively on LGG slices. Due to their broad proliferation throughout healthy brain tissue, high-grade gliomas (HGG) are easier to segment than low-grade gliomas (LGG), albeit class-based segmentation of tumor regions may prove to be quite difficult. An effective brain tumor segmentation model should function equally effectively on both LGG and HGG volumes. In order to assess the effectiveness of a model it is imperative that it be evaluated independently on a LGG and a HGG dataset. The methods proposed by [16], [17], [18], [19], [20], [21] have been trained and tested on the BraTS datasets exclusively with an overwhelmingly large majority of mask pixels belonging to HGG.[22], [23], [24] on the other hand have only been trained and tested on LGG slices. This comes up as a major research gap in this field of study. To overcome this limitation and to prove the effectiveness of the proposed model it was trained and tested independently on the BraTS 2020 dataset and the TCGA LGG dataset yielding promising results for the segmentation of low-grade gliomas as well as high-grade gliomas.

Another interesting observation is that of all the 2D networks discussed so far the work proposed by Sunderson et al.[16] is the only one that tries to integrate 3D context by taking into account the features from the three MR planes: coronal, axial, and sagittal. Despite this innovative approach, the ablation study shows that the tri-planar ensemble's performance is hardly seen to improve in comparison to the model trained exclusively on the axial slices. The fourth model added to the ensemble that segments the tumor core from axial slices, however, significantly boosts the performance of the network. Taking inspiration from this the proposed model has been designed based on the ensemble approach where individual models have been used for the segmentation of each tumor class from the axial slices. The model architecture also draws inspiration from the modifications made to the original UNet[11] framework as seen in [17], [18], [19], [20], [21] based on which a number of architectural enhancements have been made to achieve exemplary results.

III. MATERIALS AND METHODOLOGY

This section discusses in detail the datasets used, workflow, preprocessing techniques, and the model architecture employed in sequential order.

A. Datasets

BraTS 2020 training dataset: The BraTS 2020 training dataset[12], [13], [14] is comprised of 293 high-grade and 76 low-grade glioma patient MRI volumes. For each patient T1, T1 contrast-enhanced (T1-ce), T2, and Fluid Attenuated Inverse Recovery (FLAIR) MR volumes, along with expert tumor segmentation are provided. Each brain tumor is manually delineated into 3 classes: edema, necrotic/non-enhancing core, and enhancing tumor core. Fig. 1 depicts representative 2D multimodal MR slices for patient volume no. 001 from the BraTS 2020 training dataset along with the corresponding multiclass segmented ground truth.

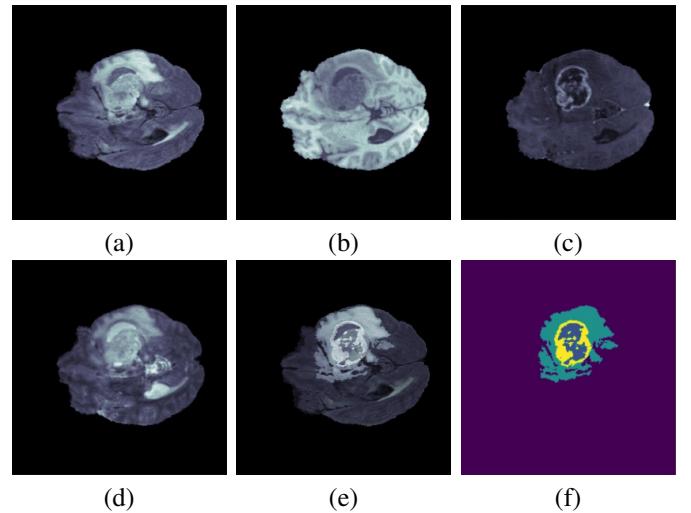


Fig. 1: Representative multimodal 2D slices from the BraTS 2020 training dataset and corresponding segmentation ground truth. (a) FLAIR modality, (b) T1 modality, (c) T1-ce modality, (d) T2 modality (e) FLAIR modality superimposed with multiclass segmentation mask (f) Multiclass segmentation mask.

TCGA LGG dataset: The TCGA-LGG dataset acquired from The Cancer Imaging Archive (TCIA) [25] was used for training and validating the model on LGG slices exclusively. The dataset comprises a total of 3,929 images with corresponding binary segmentation masks. Of these 2556 images had tumor regions and 1373 showed normal brain tissue. Each image consists of three channels- T1, FLAIR, T2, and T1-ce. However, several slices were missing the T1 and T1-ce channels. In such cases, the FLAIR modality was used to replace the missing channels. Some of the representative MR slices from the TCGA LGG dataset along with their corresponding ground truths is shown in Fig. 2.

B. Methodology

An overview of the methodology workflow for the implementation of the proposed method is provided in this section. Fig. 3

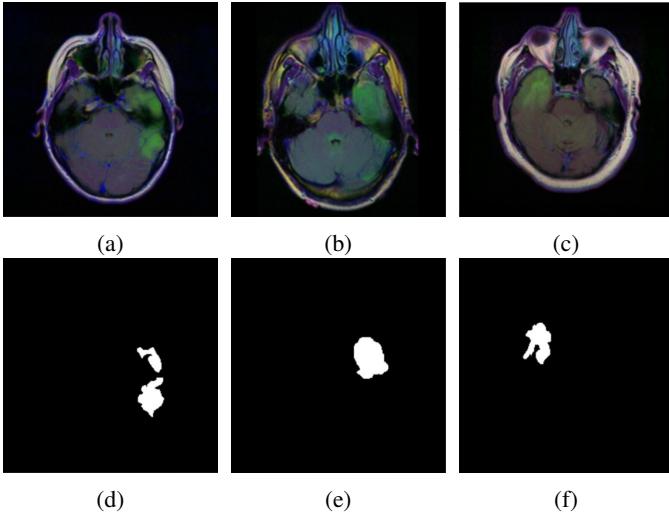


Fig. 2: Representative FLAIR 2D slices from the TCGA LGG dataset with the corresponding segmented ground truths. (d), (e) and (f) represent the corresponding ground truths for the FLAIR slices (a), (b), and (c) respectively.

is a schematic representation of the workflow plan. Following are the steps:

- 1) Pre-processing. Due to the two datasets' varied natures, two distinct pre-processing approaches have been adopted.
- 2) Splitting processed data into train test and validation datasets.
- 3) Training models for each tumor class while introducing augmentation via a data-generator.
- 4) Generation of test set segmentation outputs using trained model/ensemble.
- 5) Application of post-processing scheme.
- 6) Final output generation and evaluation.

The following sections go into great depth about the implementation of these steps.

C. Preprocessing:

For the BraTS 2020 training dataset 100 2D slices were extracted from each multimodal MR and corresponding mask volume. The extracted 2D MR and mask slices were then cropped to dimensions 128×128 following which the multiclass masks were modified to represent the tumor classes whole tumor(WT), tumor core(TC), and enhancing tumor(ET) respectively prior to training. The MRI slices of modalities FLAIR, T1-ce, and T2 were then merged to form 3-channel images. Each image and corresponding mask pair for the respective classes were then normalized to [0-1] and then fed to the corresponding models for training. A schematic representation of this scheme can be seen in Fig. 4(a). The TCGA LGG dataset consists of 2D MR slices of modalities T1, T1-ce, T2 and FLAIR along with their respective binary masks representing class whole tumor(WT). The FLAIR, T1-ce, and T2 MR slices are combined into a single 2D multi-channel image. In case of missing MR modalities, they were

replaced with the FLAIR modality. Finally, the multichannel MR images and their corresponding binary masks were normalized to [0-1] and resized to dimensions $256 \times 256 \times 3$ prior to training. No further preprocessing was carried out to preserve the high-dimensional features for optimized model performance. A schematic representation of this scheme has been represented in Fig. 4(b).

D. The proposed architecture:

The architectural details of SER-Net have been discussed in this section. The network derives its name "SER Net" based on the repeated use of **squeeze** and **excite** blocks and employment of the **residual learning** approach at both the encoder and decoder. Fig. 5a shows a schematic illustration of the model architecture.

1) Res block: This section introduces the idea of residual learning[26] and discusses its advantages following which the architectural details of the Res block used in the SER-Net have been analyzed in depth.

During training, a neural network sees what error it is making and tries to minimize it by adjusting its internal weights. It is guided in this by the first derivative (or gradient) of the error. Unfortunately, with many layers, data from the lower levels get lost and the gradient tends to be spread too thin across all layers and the network converges slowly or not at all. This is termed the degradation problem. Residual networks[26] help prevent the loss of data during propagation by adding skip connections in parallel with the convolutional layers. Skip connections convey the signal as is, then recombine it with the data that has been transformed by one or more convolutional layers. The combining operation is a simple element-wise addition. The element-wise addition will work only if the dimensions of the data at the skip connection output and the output from the residual block (the sequence of layers straddled by a skip connection) is called a residual block. The residual block along with its skip connection has been termed the Res-block) remains the same. To remedy this a 1×1 convolutional layer is introduced across the skip connection to increase the number of feature channels to match the shape of the residual block's output.

As can be seen in Fig. 5c. the output of the block is the sum of the output of the convolutional path $C_2(x)$ and the skip connection x . The final output $f(x)$ can be shown as eqn 1

$$f(x) = C_2(x) + x \quad (1)$$

$$f(x) = x + W_2g(W_1g(x) + b_1) + b_2 \quad (2)$$

$$C_2(x) = f(x) - x \quad (3)$$

the terms of eqn 1 can be rearranged to obtain the expression for $C_2(x)$, which is clearly difference between the desired output and the input(also called the residue), as can be seen in eqn 3. Kaiming et al[26] argue that this "residue" is easier for the network to learn. This is evident from eqn 1 and eqn 2. For $f(x)$ to become identity function x , the residue $C_2(x)$ must become zero i.e the weight matrix W_2 and bias b_2 must become 0. On the other hand in the absence of the skip connection $f(x) = C_2(x)$, the weights and bias values have to be modified so that it will correspond to the identity function.

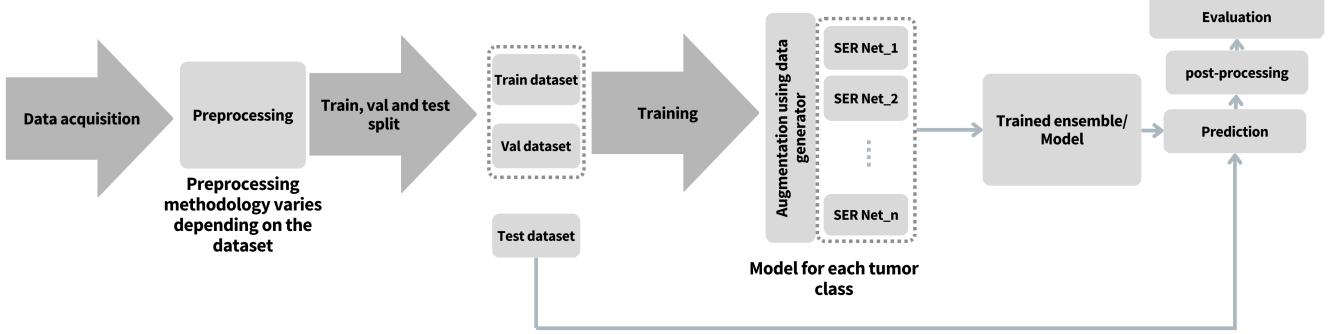


Fig. 3: Schematic representation of the methodology workflow.

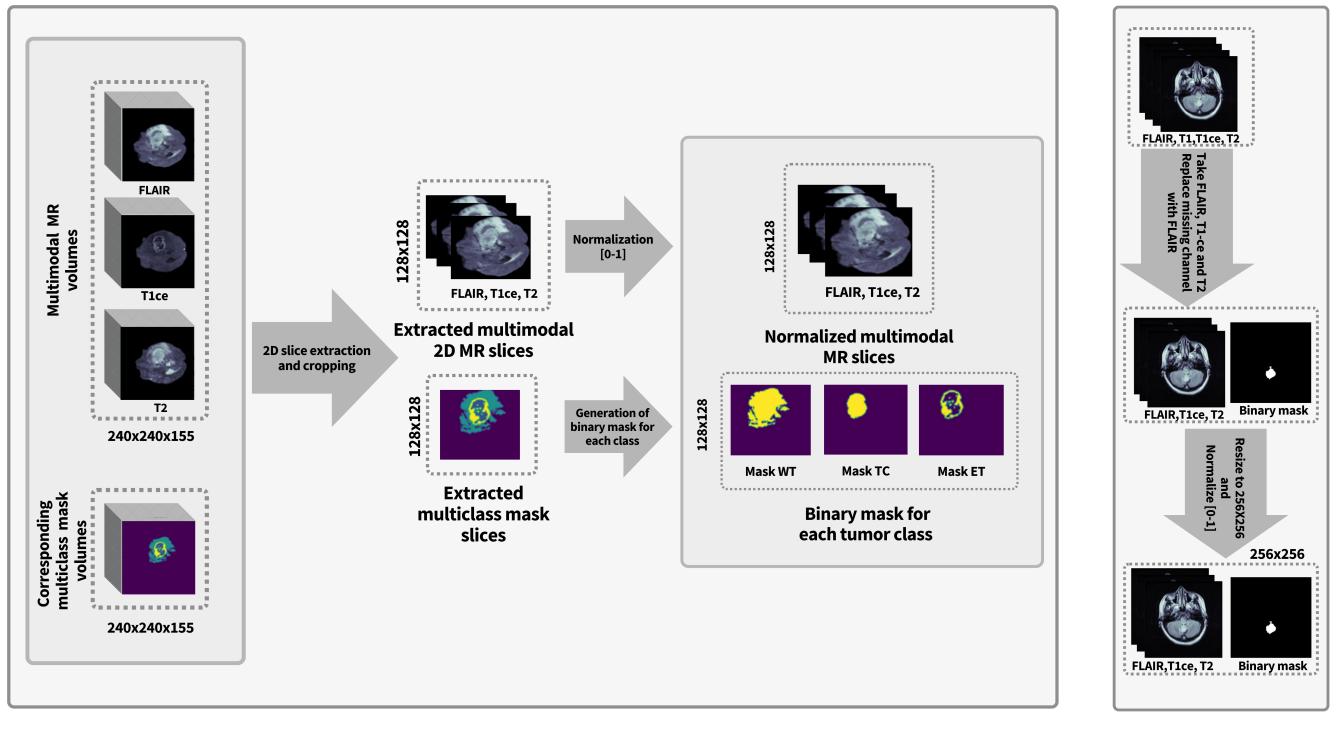


Fig. 4: Pre-processing methodology. (a) Preprocessing methodology for the BraTS 2020 dataset. (b) Preprocessing methodology for the TCGA LGG dataset

It is very difficult to learn identity function from scratch for a non residual network, exacerbated by the non-linearity in the layers which results in the degradation problem.

Taking inspiration from the advantages of the residual learning approach the Res-block (see Fig.5c) has been implemented as an integral architectural aspect of both the encoder and decoder. The residual network in each Res-block consists of two convolutional layers each preceded by a batch normalization and a relu activation layer in sequence. The batch normalization layers help reduce the internal co-variate shift and also regularize the model. The Dropout layer randomly sets input units to 0 with a predetermined ‘rate’ (here 0.1) at each step during training time, which helps prevent overfitting. Inputs not set to 0 are scaled up by $1/(1 - \text{rate})$ such that the

sum over all inputs is unchanged. Fig.5c shows a schematic for the sequence layers in a Res-block further details to which can be found in Table I.

2) *Squeeze and excite block*: CNNs use their convolutional filters to extract hierachal information from images. Lower layers find trivial pieces of context like edges or high frequencies, while upper layers can detect faces, text, or other complex geometrical shapes. They extract whatever is necessary to solve a task efficiently. All of this works by fusing the spatial and channel information of an image. The different filters will first find spatial features in each input channel before adding the information across all available output channels. The network weights each of its channels equally when creating the output feature maps. The squeeze-and-excitation (SE) network

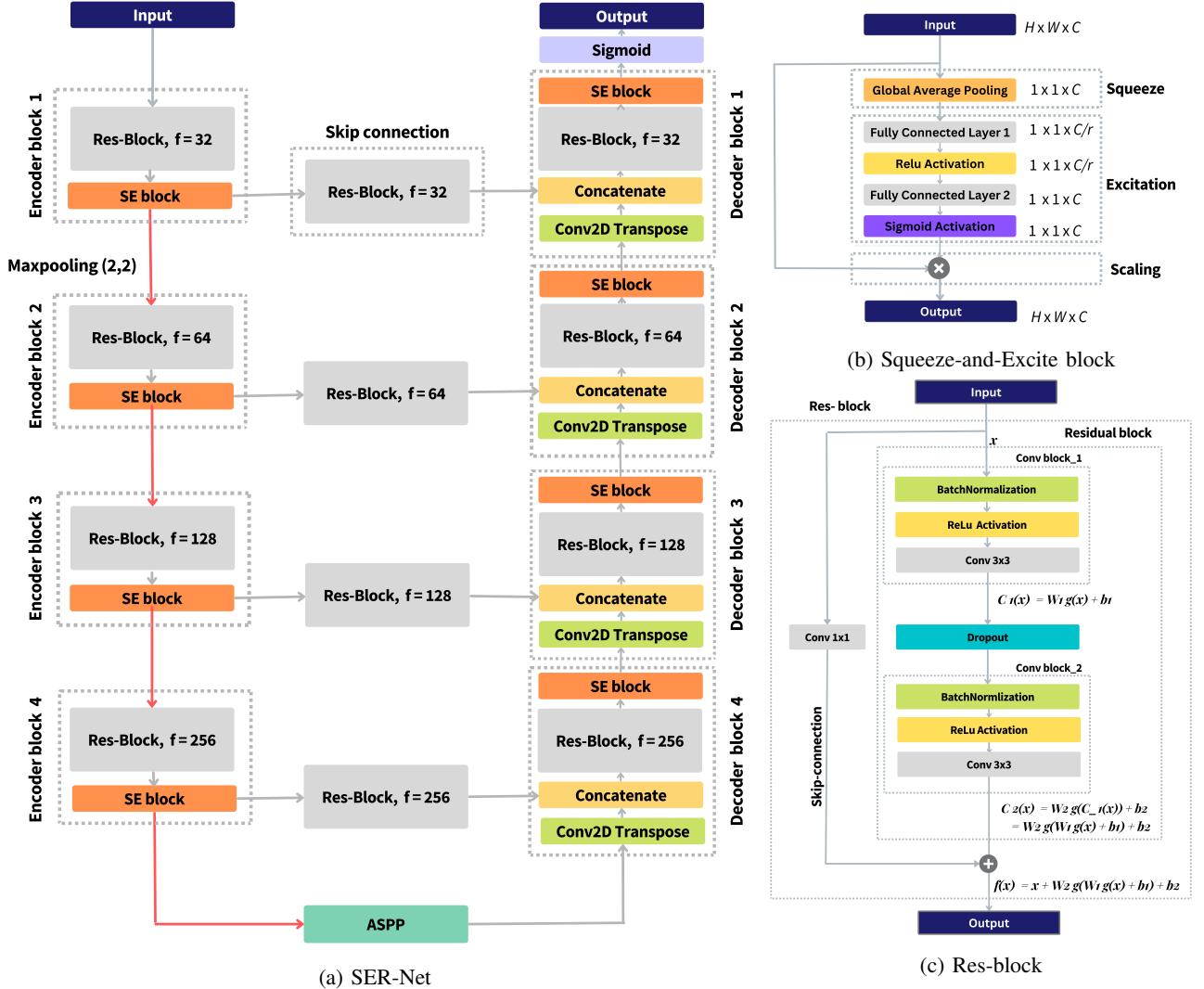


Fig. 5: Illustrative representation of (a) proposed model architecture, (b) SE-block, (c) Res-block.

TABLE I: Sequence of layers in the Res-block, layer 8 represents the skip connection. Here f_i stands for the number of filters in the i^{th} encoder block, k is the kernel size, d is the dropout percentage, and s refers to the stride. For each convolution zero padding was used to keep the dimensions of the input and output feature maps the same.

Layer no.	Layer type	connected to layer
1	BatchNormalization	input
2	ReLU Activation	1
3	Conv2D($f_{-i} = i \times 32, k = 3, s = 1$)	2
4	Dropout($d=0.1$)	3
5	BatchNormalization	4
6	ReLU Activation	5
7	Conv2D($f_{-i} = i \times 32, k = 3, s = 1$)	6
8	Conv2D($f_{-i} = i \times 32, k = 1, s = 1$)	input,9
9	Addition	7, 8

developed by Hu et al.[25] adds a content-aware mechanism to weigh each channel adaptively. The SE block first estimates the importance of each channel (feature map) and then weights each channel according to its importance. This ensures that the

channel with more relevant spatial information gets a higher contribution in the output feature map compared to the other channels. The SE block has two branches. One branch just conveys input x while the other branch computes a modulation weight for each input channel. The computation of modulation weight is carried out in two phases- *Squeeze* and *Excitation* (see Fig. 5b) following which the input channels are scaled with the corresponding modulation weights. Let the tensor x of dimensions $H \times W \times C$ be the input to SE block. In the squeeze phase, a scalar descriptor is computed for each channel of input x . This channel descriptor is computed using global average pooling. If the channel shape is $H \times W$ and the i^{th} channel is denoted as c_i then channel descriptor is:

$$\frac{1}{H \times W} \sum_{m=1}^H \sum_{n=1}^W c_i[m, n] \quad (4)$$

The output of the squeeze operation has the dimensions- $1 \times 1 \times C$. At the excitation phase channel descriptors computed at the squeeze phase are fed to fully connected layer 1. Here dimension of channel C is reduced by a factor r . Hence, the

output of the fully connected layer has shape- $1 \times 1 \times \frac{C}{r}$. In the proposed model the hyperparameter r has a default value of 8). There is another fully connected layer after the Rectified Linear unit where the dimensions of C is restored to the original value. Hence output of fully connected layer 2 has shape- $1 \times 1 \times C$. At the end of the excitation phase, *sigmoid* activation is used to map modulation weights to $[0, 1]$ interval. After excitation phase channels are scaled with corresponding modulation weights. The final output shape is $H \times W \times C$.

3) *Atrous spatial pyramidal pooling*: Multiple parallel atrous convolutions with various rates make up the ASPP. It combines atrous convolution and spatial pyramidal pooling, and for more precise classification, it may gather contextual information at different scales. The brain MR images not only have plenty of local specifics but also have an almost limitless macro target expansion. As a result, in order to extract multiscale features for brain tumor extraction, ASPP is required. Deep convolutional neural networks' features can be resolved more precisely via atrous convolution, which can also be used to change the receptive field to capture multiscale data[27]. For each pixel i on the output y and filter W , atrous convolution is applied to the input x as shown in eqn 5

$$y[i] = \sum_k x[i + r.k]W[k] \quad (5)$$

where the atrous rate r determines the stride of sampling the input image. Atrous convolution is the process of combining the input x with the filters created by interspersing $r - 1$ zeros between the values of two successive filters. The receptive field of the filter can be changed by varying the rate r . The proposed model employs an ASPP module that consists of four parallel atrous convolutions with various atrous rates. The ASPP module specifically consists of - (a) a single 1×1 convolution, three parallel 3×3 convolutions with dilation rates 3,5, and 7 respectively, and (b) an image-level feature created by global average pooling. The resulting features from all of the branches are bilinearly upsampled to the input size and then concatenated and passed through another 1×1 convolution. ASPP is applied to the feature map produced by the encoder part, and the resulting feature map is fed into the decoder part. The schematic representation of ASPP can be seen in Fig. 6a.

4) *Encoder*: The encoder of the SER-Net is primarily a cascade of structurally repeating units called encoder blocks where each successive unit in the cascade has double the number of feature channels compared to the preceding one. Each encoder block comprises of a Res-block followed by a squeeze-and-excite block, the architectural details of which can be found in section III-D1, III-D2, and Table I respectively. The output from each encoder block is channeled across two paths - The first path takes the output feature map across a Res-block and feeds the output thereof to a decoder block operating with the same number of feature channels. This path is called the (long) skip connection. The second path subjects the output feature map to a 2×2 maxpooling operation with a stride of 2 thereby reducing its spatial dimensions by half following which it is fed as input to the next encoder block.

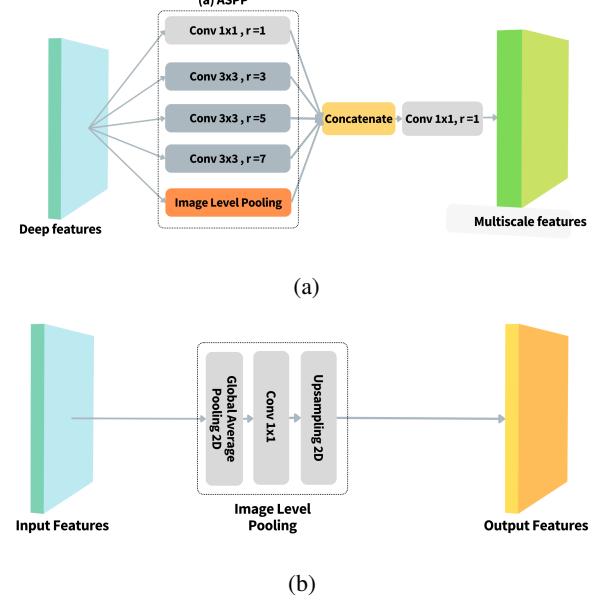


Fig. 6: (a) Illustrative representation of the ASPP block used in the bottle-neck region of SER Net. (b) Illustrative representation of the sequence of operations followed in image level pooling.

This successive downsampling and increase in the number of feature channels helps improve the receptive field of the model thereby enabling the model to extract more contextual information.

5) *Decoder*: The decoder is composed of structurally repeating blocks like the encoder. For each encoder block there is a corresponding decoder block. Each decoder block performs a 2×2 bi-linear up-sampling on the input feature map doubling its dimensions. The upsampled output is then concatenated with the skip features obtained from the corresponding encoder block. After concatenation the feature maps are then passed through a Res-block (see section III-D1) followed by a squeeze and excitation unit. The output of the last decoder block undergoes a 1×1 convolution with *sigmoid* activation producing the final output.

E. Evaluation metrics

The following metrics were used to evaluate the performance of the model:

(1) **Loss function** The datasets used for the training and evaluation of the proposed model suffer heavily from class imbalance. (Overwhelming majority of pixels in the images represent the non-tumor class). Extraction of equal training samples from each class is a typical strategy to adjust for data imbalance. However, the drawback of this strategy is that it may favor rare classes and does not fully utilize the information richness of the images. To overcome this challenge Focal-Dice Loss(L_{FD}) was used as the loss metric during the training process. It is defined as shown in eqn 6.

$$L_{FD} = L_{Dice} + L_{Focal} \quad (6)$$

where L_{Focal} is Focal loss. Focal loss is a variant of the binary cross entropy loss that addresses the issue of class imbalance with the standard cross entropy loss by down-weighting the contribution of easy examples enabling learning of harder examples. It is defined as shown in eqn 7.

$$L_{Focal}(p_t) = \alpha(1 - p_t)^\gamma \cdot L_{BCE}(p, y) \quad (7)$$

Here, L_{BCE} is the Binary Cross Entropy loss. The probability of predicting the ground truth class, p_t , is defined as shown in eqn 8.

$$p_t = \begin{cases} p, & \text{if } y = 1 \\ 1 - p, & \text{if } y = 0 \end{cases} \quad (8)$$

The class weights and degree of down-weighting of easy-to-classify pixels are controlled, respectively, by the parameters α and γ . The Focal loss is reduced to the Binary Cross Entropy loss when $\gamma = 0$. L_{Dice} is the Dice loss. It is based on the Dice coefficient, (see eqn 11) and can be defined as shown in eqn 9.

$$L_{Dice} = 1 - \text{Dice coeff} \quad (9)$$

(2) Dice coefficient:

Dice coefficient is a parameter used to gauge a model's segmentation performance. It can be used to assess how well a predicted segmentation matches the ground truth in terms of pixel-level agreement. It is determined by dividing the total area of overlap between the two images (predicted output and ground truth) divided by number of pixels in both images. Mathematically it can be represented as shown in eqn 10.

$$DC = \frac{2 \times |X \cap Y|}{|X \cup Y|} \quad \text{where, } DC \in [0, 1] \quad (10)$$

Here X and Y are binary classes representing tumor and non-tumor classes respectively. It can also be represented as eqn 11.

$$DC = \frac{2 \times TP}{2TP + FP + FN} \quad \text{where, } DC \in [0, 1] \quad (11)$$

Here TP = true positives, FP = False positives, FN = False negatives

(3) IoU:

IoU is defined as the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the two. The mean IoU of an image is determined for binary (two classes) or multi-class segmentation by averaging the IoU of each class. It is represented as shown in eqn 12.

$$IoU = \frac{|X \cap Y|}{|X \cup Y|} = \frac{TP}{TP + FP + FN} \quad \text{where, } IoU \in [0, 1] \quad (12)$$

Here X and Y are binary classes representing tumor and non-tumor classes respectively.

(4) Sensitivity:

The metric sensitivity assesses a model's capacity to forecast true positives for each available class. It is represented as shown in eqn 13.

$$Sensitivity = \frac{TP}{TP + FN} \quad (13)$$

(5) Specificity:

The metric specificity assesses a model's capacity to forecast true negatives for each available class. It is represented as shown in eqn 14.

$$Specificity = \frac{TN}{TN + FP} \quad (14)$$

F. Implementation details:

The code was implemented using TensorFlow version 2.6.4 and the models were trained using an NVIDIA Tesla P100 GPU on a cloud-based server. For each dataset, a train-validation split of ratio 75:10:15 was followed. Image-data-generator was used to introduce augmentations such as random-flip (horizontal and vertical), random-rotate, random-zoom, etc. during the training process. The models were trained using the Adam optimizer with an initial learning rate of 0.001 and *reduce on plateau* (for val loss) with *decay factor* = 0.2 and *patience* = 5. The model trained on the TCGA LGG dataset took approximately 70 epochs to reach convergence while each model for the ensemble trained on the BraTS 2020 dataset took around 35 epochs to converge.

IV. EXPERIMENTAL RESULTS

Performance on the TCGA LGG dataset:

The TCGA LGG dataset has 1373 subjects with positive mask values. Of these, a total of 1098 subjects were used for training, 138 for validation, and 137 for testing. The model took around 70 epochs to converge. The training-validation curves can be seen in fig 7. In order to determine the effect of individual components of the proposed architecture on the segmentation performance, the segmentation results on the test set have been evaluated with the following components : (i) UNet only, (ii) Res-UNet, (iii) SE-Res-UNet, (iv) SE-Res-UNet + ASPP, and (v) SE-Res-UNet + ASPP + Residual skip connections (SER-Net). The values of the performance metrics for the ablation study are shown in II. The performance gains achieved by virtue of the architectural enhancements to the UNet leading to the SER-Net can be clearly seen from the comparison of the segmentation masks produced by the various models in fig 8.

TABLE II: Comparison of the proposed model's performance with various other UNet architectural variants on testing data of TCGA LGG dataset.

Model	Mean Dice score	Mean IoU score	Sensitivity	Specificity
UNet	0.77	0.693	0.781	0.9991
Res-UNet	0.83	0.722	0.820	0.9995
SE-Res-UNet	0.87	0.795	0.881	0.9996
SE-Res-UNet + ASPP	0.901	0.812	0.907	0.9997
SER-Net	0.924	0.855	0.933	0.9997

Further, a detailed comparative analysis of the performance of the proposed method with some of the state-of-the-art methods was carried out. The details can be seen in Table III.

Performance on the BraTS 2020 dataset:

Out of the total 369 labeled subjects from the BraTS 2020

TABLE III: Results of the proposed method on the TCGA LGG test set and comparison with other state-of-the-art methods.

Method	Mean Dice score	Mean IoU score	Sensitivity	Specificity
Ashraf et al.[22]	0.915	-	-	-
Santosh et al.[23]	0.905	0.829	-	-
Buda et al.[28]	0.915	0.840	-	-
Venkata et al.[29]	0.870	0.780	-	-
Naser et al.[30]	0.905	0.829	-	-
Sourodiip et al.[24]	0.918	0.826	-	-
SER-Net	0.923	0.855	0.933	0.9997

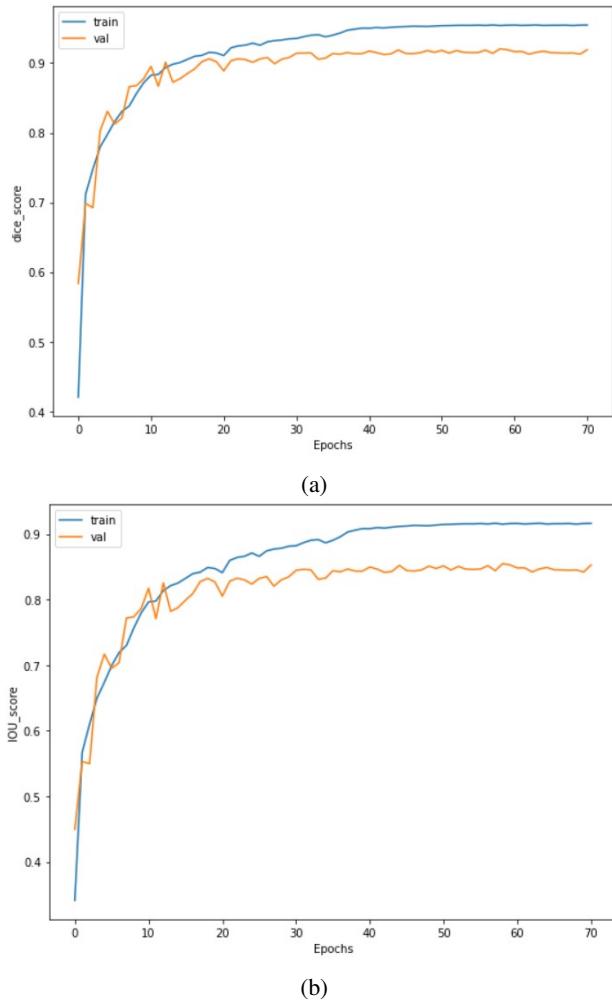


Fig. 7: (a) Training-validation curve for the dice score metric on the TCGA LGG dataset. (b) The training-validation curve for the IoU score metric on the TCGA LGG dataset.

dataset 255 were used for training, 37 for validation, and 77 for testing. Each model for the ensemble took approximately 35 epochs to reach convergence. The training-validation curves can be seen in fig 9. To evaluate the efficacy of each individual component of the model, an ablation study akin to that carried out for the TCGA LGG dataset was conducted. Table IV reports the ablation study's findings based on the testing data, and fig 10 displays some of the qualitative segmentation results from the ablation study.

The trained ensemble was further evaluated against some of

the state-of-the-art methods and the comparative findings have been enlisted in Table V.

V. DISCUSSIONS AND CONCLUSION

In this study, SER-Net, a novel encoder-decoder based CNN architecture, for brain tumor segmentation has been proposed. For the TCGA LGG dataset a single SER-Net model was trained for the binary segmentation of brain tumors whereas an ensemble of three SER-Nets has been used for the multiclass segmentation of brain tumors on the BraTS 2020 dataset. The proposed method yielded dice scores of 92.4%, 87.7%, 85.2% for tumor subregions - whole tumor(WT), Tumor core(TC), and Enhancing Tumor(ET) respectively on the BraTS 2020 dataset[12], [13], [14] outperforming the current state-of-the-art dice scores of 89.9%, 85.3%, 83.3%. On the TCGA LGG dataset, the model achieved a dice score of 92.4% again outperforming the current state-of-the-art dice score which is 91.8%.

The beauty of this work is that it achieves dice scores higher than many of the state-of-the-art 3D approaches without relying on any 3D context from the data indicating that effective volumetric segmentation can be achieved purely based on the planar context. The gains in performance are a consequence of the improvements made to the original UNet[11] framework. A number of studies that improved the performance of the UNet [11] by introducing architectural improvements were examined. Taking cues from these works[16], [17], [18], [24] a multitude of changes were introduced to the UNet[11] to develop a model that caters effectively to the task of brain tumor segmentation. The features of the various tumor classes (WT, TC, and ET) are more pronounced in MR slices of different modalities. Channel-wise weighting is thus introduced using SE blocks[25] to improve segmentation results. Similarly, residual building blocks were used in both the contractive as well as the expansive path to tackle the degradation problem. Taking into account the fact that brain MR images not only have plenty of local specifics but also have an almost limitless macro target expansion, ASPP (see sec III-D3) was integrated to extract the multiscale features.

Despite obtaining exemplary segmentation results, the 2D segmentation approach stands as one of the major limitations of the proposed work. In this work, the possibility of integrating context from multiple planes remains unexplored and might be taken up as future works. This might considerably improve the segmentation performance of the proposed method albeit with the added computational cost.

VI. ACKNOWLEDGEMENTS

The authors of this paper declare that the proposed work did not utilize any pre-trained models nor additional datasets other than those listed in section ?? and that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The work was done under the Biomedical Imaging Laboratory (BIOMIL), National Institute of Technology Silchar, Silchar, Assam, 788010, India.

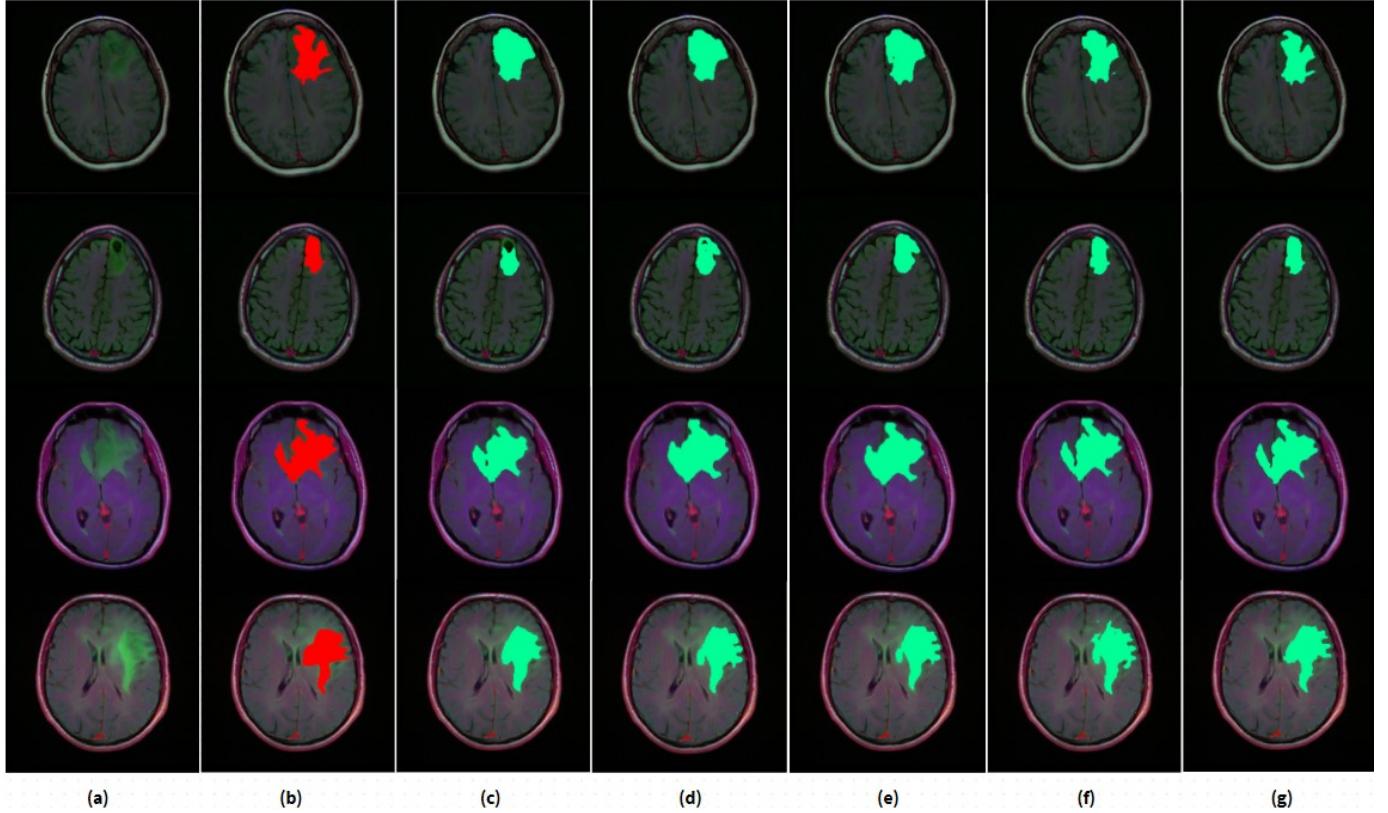


Fig. 8: A comparison of the segmentation results produced by the SER-Net and different variants of the UNet on random slices from the TCGA LGG testing data. (a) Original FLAIR image, (b) FLAIR image with ground-truth, (c) Predicted mask: UNet, (d) Predicted mask: Res-UNet, (e) Predicted mask: SE-Res-UNet, (f) Predicted mask: SE-Res-UNet+ASPP, (g) Predicted mask: SER-Net

TABLE IV: Comparison of the proposed ensemble’s performance with various other UNet architectural variants on testing data from the BraTS 2020 dataset.

Architecture	Mean Dice			Mean IoU			Sensitivity			Specificity		
	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET
U-Net	0.85	0.80	0.76	0.79	0.73	0.71	0.88	0.83	0.84	0.9991	0.9993	0.9992
Res-UNet	0.88	0.85	0.80	0.82	0.75	0.74	0.90	0.87	0.86	0.9992	0.9995	0.9993
SE-Res-UNet	0.90	0.86	0.82	0.86	0.79	0.76	0.93	0.90	0.87	0.9994	0.9995	0.9991
SE-Res-UNet + ASPP	0.91	0.86	0.83	0.87	0.80	0.78	0.93	0.91	0.89	0.9993	0.9994	0.9992
SER-Net	0.92	0.87	0.85	0.88	0.81	0.80	0.95	0.90	0.88	0.9997	0.9995	0.9994

TABLE V: Comparison of the results of the proposed ensemble (SER-Net_WT, SER-Net_TC, and SER-Net_ET) on the local test set of BraTS 2020 training dataset with some of the state-of-the-art methods

Method	Dataset	Mean Dice			Mean IoU			Sensitivity			Specificity		
		WT	TC	ET									
Sundaresan et al.[16]	BraTS 2020	0.89	0.85	0.83	-	-	-	0.88	0.83	0.84	0.99	0.99	0.99
Yanwu Xu et al.[17]	BraTS 2017	0.87	0.78	0.77	-	-	-	-	-	-	-	-	-
Varghese et al.[18]	BraTS 2017	0.84	0.84	0.77	-	-	-	-	-	-	-	-	-
Ding et al.[19]	BraTS 2015	0.83	0.67	0.59	-	-	-	-	-	-	-	-	-
Zhang et al.[20]	BraTS 2017	0.87	0.77	0.72	-	-	-	-	-	-	-	-	-
Ilyas et al.[21]	BraTS 2018	0.88	0.76	0.65	-	-	-	0.89	0.74	0.66	0.99	0.99	0.99
Ballestar et al.[31]	BraTS 2020	0.85	0.85	0.77	-	-	-	-	-	-	-	-	-
Findon et al.[32]	BraTS 2020	0.91	0.84	0.77	-	-	-	-	-	-	-	-	-
Aboelenein et al.[33]	BraTS 2020	0.87	0.84	0.80	-	-	-	-	-	-	-	-	-
SER-Net ensemble	BraTS 2020	0.92	0.87	0.85	0.88	0.81	0.80	0.95	0.90	0.88	0.99	0.99	0.99

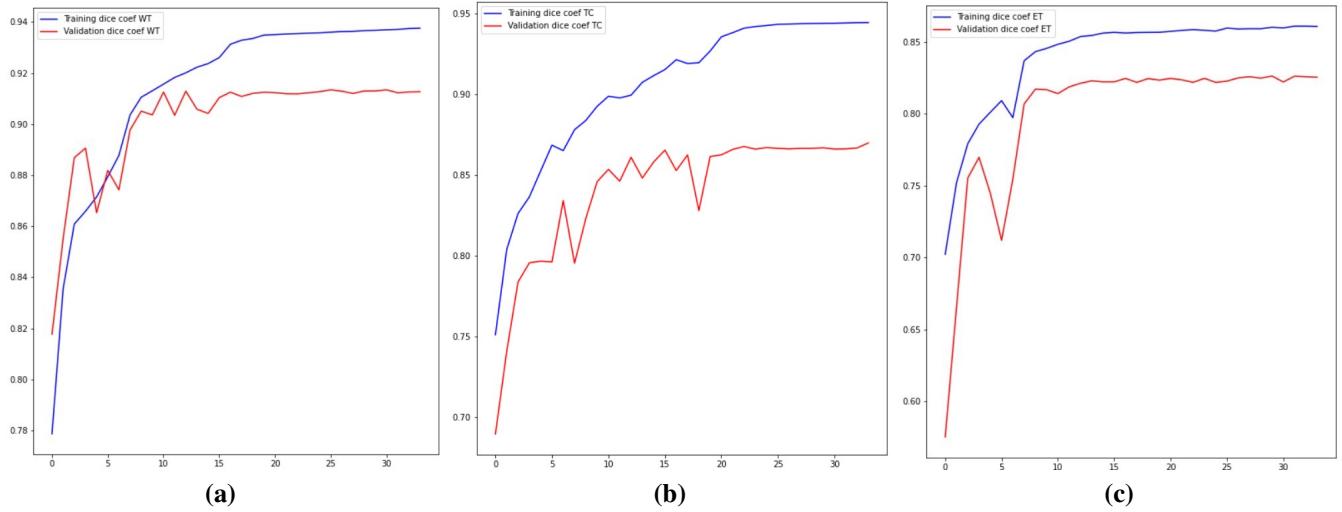


Fig. 9: Traning validation curves for the metric dice coefficient for (a) SER-Net_WT, (b) SER-Net_TC, and (c) SER-Net_ET on the BraTS 2020 dataset

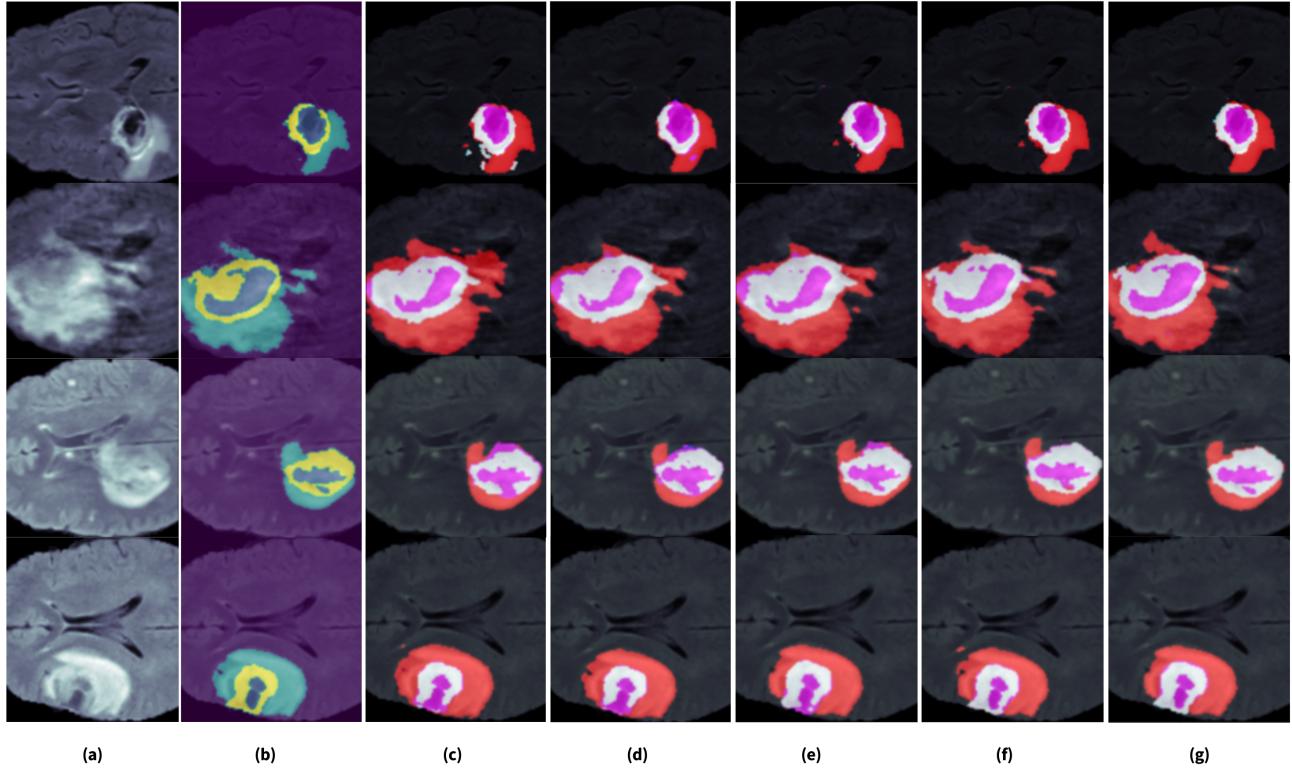


Fig. 10: A comparison of the segmentation results produced by the SER-Net and different variants of the UNet on random slices from the BraTS 2020 testing data. (a) Original FLAIR image, (b) FLAIR image with ground-truth, (c) Predicted mask: UNet, (d) Predicted mask: Res-UNet, (e) Predicted mask: SE-Res-UNet, (f) Predicted mask: SE-Res-UNet+ASPP, (g) Predicted mask: SER-Net

VII. CODE AVAILABILITY

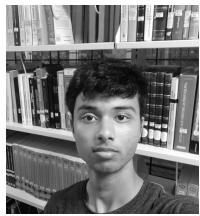
All code and implementation details can be found in https://github.com/Sourjya261/Brain_tumor_segmentation which will be made publicly available post-publication. For further details please contact sourjya20_ug@ece.nits.ac.in or muru-

gan.rmn@ece.nits.ac.in.

REFERENCES

- [1] Mark A Vickers. Jak2 617v ζ f-positive polycythemia rubra vera maintained by approximately 18 stochastic stem-cell divisions per year,

- explaining age of onset by a single rate-limiting mutation. *Blood, The Journal of the American Society of Hematology*, 110(5):1675–1680, 2007.
- [2] OI Olopade and G Pichert. Cancer genetics in oncology practice. *Annals of oncology*, 12(7):895–908, 2001.
- [3] Gavin P Dunn, Allen T Bruce, Hiroaki Ikeda, Lloyd J Old, and Robert D Schreiber. Cancer immunoediting: from immunosurveillance to tumor escape. *Nature immunology*, 3(11):991–998, 2002.
- [4] Karl Herholz, Karl-Josef Langen, Christiaan Schiepers, and James M Mountz. Brain tumors. In *Seminars in nuclear medicine*, volume 42, pages 356–370. Elsevier, 2012.
- [5] Erwin G Van Meir, Costas G Hadjipanayis, Andrew D Norden, Hui-Kuo Shu, Patrick Y Wen, and Jeffrey J Olson. Exciting new advances in neuro-oncology: the avenue to a cure for malignant glioma. *CA: a cancer journal for clinicians*, 60(3):166–193, 2010.
- [6] Mina Ghaffari, Arcot Sowmya, and Ruth Oliver. Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the brats 2012–2018 challenges. *IEEE reviews in biomedical engineering*, 13:156–168, 2019.
- [7] Stefan Bauer, Roland Wiest, Lutz-P Nolte, and Mauricio Reyes. A survey of mri-based medical image analysis for brain tumor studies. *Physics in Medicine & Biology*, 58(13):R97, 2013.
- [8] Brent Foster, Ulas Bagci, Awais Mansoor, Ziyue Xu, and Daniel J Mollura. A review on segmentation of positron emission tomography images. *Computers in biology and medicine*, 50:76–96, 2014.
- [9] Jin Liu, Min Li, Jianxin Wang, Fangxiang Wu, Tianming Liu, and Yi Pan. A survey of mri-based brain tumor segmentation methods. *Tsinghua science and technology*, 19(6):578–595, 2014.
- [10] Dzung L Pham, Chenyang Xu, and Jerry L Prince. A survey of current methods in medical image segmentation. *Annual review of biomedical engineering*, 2(3):315–337, 2000.
- [11] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [12] H Menze Bjoern, Jakab Andras, Bauer Stefan, Kalpathy-Cramer Jayashree, Farahani Keyvan, Kirby Justin, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Trans. Med. Imaging*, 34(10):1993–2024, 2015.
- [13] Christopher T Lloyd, Alessandro Sorichetta, and Andrew J Tatem. High resolution global gridded data for use in population studies. *Scientific data*, 4(1):1–17, 2017.
- [14] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozyczyk, et al. 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*, 2018.
- [15] Maciej A Mazurowski, Kal Clark, Nicholas M Czarnek, Parisa Shamsesfandabadi, Katherine B Peters, and Ashirbani Saha. Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with the cancer genome atlas data. *Journal of neuro-oncology*, 133(1):27–35, 2017.
- [16] Vaanathi Sundaresan, Ludovica Griffanti, and Mark Jenkinson. Brain tumour segmentation using a triplanar ensemble of u-nets on mr images. In *International MICCAI Brainlesion Workshop*, pages 340–353. Springer, 2021.
- [17] Yanwu Xu, Mingming Gong, Huan Fu, Dacheng Tao, Kun Zhang, and Kayhan Batmanghelich. Multi-scale masked 3-d u-net for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 222–233. Springer, 2019.
- [18] Varghese Alex, Mohammed Safwan, and Ganapathy Krishnamurthi. Automatic segmentation and overall survival prediction in gliomas using fully convolutional neural network and texture analysis. In *International MICCAI Brainlesion Workshop*, pages 216–225. Springer, 2017.
- [19] Yi Ding, Fujuan Chen, Yang Zhao, Zhixing Wu, Chao Zhang, and Dongyuan Wu. A stacked multi-connection simple reducing net for brain tumor segmentation. *IEEE Access*, 7:104011–104024, 2019.
- [20] Jianxin Zhang, Zongkang Jiang, Jing Dong, Yaqing Hou, and Bin Liu. Attention gate resu-net for automatic mri brain tumor segmentation. *IEEE Access*, 8:58533–58545, 2020.
- [21] Naveed Ilyas, Yoonguu Song, Aamir Raja, and Boreom Lee. Hybridanet: An encoder-decoder based hybrid weights alignment with multi-dilated attention network for automatic brain tumor segmentation. *IEEE Access*, 10:122658–122669, 2022.
- [22] Mohammad Ashraf Ottom, Hanif Abdul Rahman, and Ivo D Dinov. Znet: deep learning approach for 2d mri brain tumor segmentation. *IEEE Journal of Translational Engineering in Health and Medicine*, 10:1–8, 2022.
- [23] P Santosh Kumar, VP Sakthivel, Manda Raju, and PD Satya. Brain tumor segmentation of the flair mri images using novel resunet. *Biomedical Signal Processing and Control*, 82:104586, 2023.
- [24] Sourodip Ghosh, Aunkit Chaki, and KC Santosh. Improved u-net architecture with vgg-16 for brain tumor segmentation. *Physical and Engineering Sciences in Medicine*, 44(3):703–712, 2021.
- [25] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.
- [26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [27] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.
- [28] Mateusz Buda, Ashirbani Saha, and Maciej A Mazurowski. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Computers in biology and medicine*, 109:218–225, 2019.
- [29] Pattabiraman Ventakasubbu and Parvathi Ramasubramanian. Deep learning-based brain tumour segmentation. *IETE Journal of Research*, pages 1–9, 2021.
- [30] Marco Domenico Cirillo, David Abramian, and Anders Eklund. Vox2vox: 3d-gan for brain tumour segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I* 6, pages 274–284. Springer, 2021.
- [31] Laura Mora Ballestar and Veronica Vilaplana. Mri brain tumor segmentation and uncertainty estimation using 3d-unet architectures. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part I* 6, pages 376–390. Springer, 2021.
- [32] Lucas Fidon, Sébastien Ourselin, and Tom Vercauteren. Generalized wasserstein dice score, distributionally robust deep learning, and ranger for brain tumor segmentation: Brats 2020 challenge. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 6th International Workshop, BrainLes 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Revised Selected Papers, Part II* 6, pages 200–214. Springer, 2021.
- [33] Nagwa M AboElenein, Piao Songhao, and Ahmed Afifi. Irdnu-net: Inception residual dense nested u-net for brain tumor segmentation. *Multimedia Tools and Applications*, 81(17):24041–24057, 2022.
- [34] B Nitesh, A Madhuri, B Sai Manogna, K Naga Jogendra Babu, N Ishwarya, and G Mohan Trivedra. Brain tumor segmentation using u-net based on inception. In *2022 Second International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, pages 908–914. IEEE, 2022.
- [35] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.
- [36] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.



Sourjya Mukherjee is currently pursuing a B-Tech degree with the Department of Electronics and Communications Engineering, National Institute of Technology, Silchar, Assam, 788010, India. His main research interests are Machine Learning and Deep Learning in Medical Applications.



Ananya Bhattacharjee is born in Assam, India in 1993. She is currently pursuing a Ph.D. degree in the Department of Electronics & Communication Engineering at the National Institute of Technology Silchar, Assam, India. She received her M.Tech degree in Communication Engineering from the North Eastern Regional Institute of Science and Technology, Arunachal Pradesh, India in 2018. She has published in reputed journals such as IEEE Transactions on Radiation and Plasma Medical Sciences, and Physical and Engineering Sciences in Medicine. Her research interest includes Medical Image Processing, Lung Cancer Detection, Machine Learning, Deep Learning, and Artificial Intelligence.



R. Murugan R. Murugan received his B.E. degree in Electronics and Communication Engineering and M.E. degree in Embedded System Technologies from Anna University, Chennai, Tamilnadu, in 2005, and 2010 respectively. He received his Ph.D. degree from Information and Communication Engineering, Centre for Research, Anna University, Chennai, Tamilnadu, India. He is working as an Assistant Professor in the Department of Electronics and Communication Engineering, National Institute of Technology Silchar since 15th June 2018. He published more than 26 journal publications including IEEE Transactions on Parallel and Distributed Systems, and IEEE Transactions on Radiation and Plasma Medical Sciences, 22 conference proceedings, 2 books, 9 book chapters, and 4 patents in his credit. His area of interest includes Bio-medical signal and image processing, medical imaging, Retinal image analysis, computer vision, pattern recognition, machine learning, and deep learning.