GCSER-UNet: An Innovative Fusion of Multi-scale Residual Learning and Novel Global Context-aware Squeeze and Excite Mechanism for Enhanced Brain Tumor Segmentation from Magnetic Resonance Images

Sourjya Mukherjee *Student Member, IEEE*, Ananya Bhattacharjee, *Student Member, IEEE*, R Murugan, *Member, IEEE*

***Abstract*—Brain cancer's severity necessitates precise brain tumor segmentation via 3D MRI, crucial for diagnosis, monitoring, and effective treatment planning. Manual identification, burdened by high costs, labor, and error risks, highlights the need for automated methods. In this study, we introduce the Global Context-aware Squeeze and Excite Residual UNet (GCSER-UNet), integrating the innovative Global Context-aware Squeeze and Excite (GCSE) mechanism. GCSE facilitates a fusion of spatial and channel-wise attention, enhancing the model's capacity to capture intricate spatial dependencies and contextual information. GCSER-UNet efficiently extracts tumor segments from multimodal MRI slices, delivering exceptional performance. Evaluations on benchmark databases exhibit its superiority, achieving a notable 94% dice score on the TCGA LGG dataset, surpassing the state-of-the-art dice score of 91.8%. In the BraTS 2020 dataset, the proposed GCSER-UNet ensemble approach yielded dice scores of 95%, 92%, and 90% for the tumor regions—Whole Tumor (W), Tumor Core (T), and Enhancing Tumor (E), respectively. By comparison, the current state-of-the-art dice scores were 94%, 93%, and 88%. These compelling outcomes highlight the efficacy of GCSER-UNet in precise brain tumor segmentation, promising enhanced efficiency and precision, thus advancing brain cancer management and treatment planning.**

***Index Terms*—Semantic segmentation, U-Net, medical image segmentation, brain tumor segmentation, deep convolutional neural networks.**

I. Introduction

BRAIN tumor is an uncontrolled proliferation of somatic cells caused by a series of accumulating random mutations in critical genes that regulate cell growth and differentiation in the human brain [1]. Despite their rarity, brain tumors are highly lethal malignancies. The main body of the tumor can originate in the brain or it can manifest itself as a secondary growth with its primary origins in other organs [2]. Glioma, one of the most rapidly advancing types of brain tumor arising from glial cells in the brain, has been the focus of most recent brain tumor segmentation research. The typical survival duration for patients with glioblastoma, a particularly aggressive type of glioma, is fewer than 14 months [3]. Therefore, early detection of such malignancies can help implement a proper therapeutic regimen and more effective planning of surgeries.[3].

Due to its superior soft tissue contrast and ubiquitous accessibility, magnetic resonance imaging (MRI) is widely considered as the industry standard for detecting brain tumors [4]. However, precise segmentation is difficult due to the features of brain tumors [4]. To improve the distinction between the various tumor subregions and between tumor and non-tumor tissue, data from a variety of complementary 3D MRI modalities, including T1, T1-ce (T1 with contrast agent), T2, and FLAIR (Fluid Attenuation Inversion Recovery), are combined. The T1-ce modality highlights the tumor boundary, while the T2 modality emphasizes the tumor area. FLAIR scans aid in distinguishing edema from cerebrospinal fluid (CSF) [5]. Integration of data from different MRI modalities can enhance the region-wise segmentation of brain tumors.

Manually segmenting MR scans is a laborious and error-prone process due to the significant variability in brain tumor location, size, and appearance [6]. Consequently, computer-assisted tumor segmentation has emerged as a highly sought-after solution. Automated segmentation methods enable the segmentation of brain tumors into different classes, such as a necrotic tumor, enhancing tumor, tumor core, and edema, without requiring human intervention.

Convolutional Neural Networks (CNNs) have found extensive applications in the field of biomedical imaging, particularly in segmentation tasks [7]. U-Net, a prominent CNN-based method, has been widely adopted for medical image segmentation [8]. Residual Networks (ResNets) help mitigate data loss during propagation by incorporating skip connections parallel to convolutional layers [9]. The squeeze-and-excitation (SE) network also introduces a content-aware mechanism that adaptively weighs each channel, enhancing the representation of important features [10]. Atrous convolution, which captures multiscale data and enables precise feature resolution in deep convolutional neural networks, has also demonstrated significant utility [11]. Motivated by the achievements of these networks, we propose the GCSER-UNet segmentation

This research was funded by the Ministry of Human Resource Development of the Government of India and NVIDIA Academic Grant as part of a Philosophy degree at the National Institute of Technology Silchar.

This work did not involve human subjects or animals in its research.

Sourjya Mukherjee, Ananya Bhattacharjee, and R Murugan are with Bio-Medical Imaging Laboratory, Department of Electronics and Communication Engineering, National Institute of Technology Silchar, Silchar, Assam, India, 788010.

architecture.

This paper introduces the GCSER-UNet segmentation architecture, an innovative ensemble model that amalgamates key elements from the UNet architecture, Residual Networks, Atrous Spatial Pyramidal Pooling (ASPP), and introduces the novel Global Context-aware Squeeze and Excite (GCSE) mechanism. The proposed study encompasses substantial enhancements to the foundational UNet architecture, coupled with the integration of the advanced GCSE mechanism, effectively boosting the model's segmentation capabilities. In addition, to address the complexity and intricacies of brain tumor segmentation, the ensemble approach involves training three parallel models, each performing binary segmentation for the W, T, and E tumor classes, respectively. This approach is justified by the requirement for enhanced model capacity, considering the vast number of samples and the overlapping regions prevalent in the tumor classes. Furthermore, the utilization of the ensemble approach ensures more precise delineation of each tumor class, facilitating accurate segmentation outcomes

This paper presents a novel segmentation architecture called GCSER-UNet, that amalgamates key elements from the UNet architecture, Residual Networks, and Atrous Spatial Pyramidal Pooling (ASPP), and introduces the novel Global Context-aware Squeeze and Excite (GCSE) mechanism. The proposed study offers three main contributions. Firstly, we evaluate the performance of the proposed model by leveraging the synergy of all four aforementioned aspects. Secondly, while the proposed architecture is rooted in the fundamental principles of the U-Net paradigm, notable structural refinements have been meticulously devised for both the contracting and expanding pathways. These advancements are complemented by the seamless integration of the proposed GCSE mechanism. Consequently, the resulting architecture exhibits a marked advancement in the quality of segmentation outputs. Thirdly, we implement an ensemble of three parallel models, all trained with the same architecture, for the segmentation of each tumor class: W, T, and E respectively. The decision to use binary segmentation for the three tumor classes, rather than a multiclass approach, aims to bolster the model's capacity for each class, given the extensive sample size. Additionally, the presence of overlapping regions within the tumor classes renders multiclass segmentation challenging. Binary segmentation allows focused delineation of unique class characteristics, ensuring precise segmentation while mitigating the complexities associated with inter-class boundaries, thus optimizing overall model performance.. The contributions of this work are as follows:

1) A novel 2D CNN model based on the U-Net paradigm 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) The extensive utilization of the Residual Block (Res-block) in the encoder and decoder and long skip connections enhances the feature mapping across the encoder and decoder. Also, it minimizes the performance degradation caused due to the vanishing/exploding gradients problem.
3) The novel Global Context-aware Squeeze and Excite (GCSE) blocks, integrated post each Res-block module in the encoder and decoder networks, dynamically recalibrate channel-wise features while seamlessly fusing channel-wise and spatial attention. This comprehensive approach overcomes the limitations of relying solely on channel-wise attention, significantly enhancing the model's capacity to capture intricate inter-channel dependencies and vital global context information.
4) The receptive field of the suggested model for multi-scale information capture was enhanced by the ASPP applied at the bottleneck area of the encoder.

The remainder of the paper is organized as follows. Section II is devoted to a survey of the literature. Section III delves into the materials and methods used in this study, including the model's construction and mathematical expressions. Section IV goes over the experiments and their results. The conclusion finishes Section V.

II. Literature Review

This section has discussed several previous works that have provided valuable insights into developing the proposed model.

Ronneberger et al.[8] created the U-Net model for biological image segmentation based on the encoder-decoder paradigm. Sundaresan et al.[12] implemented a tri-planar architecture that consists of three two-dimensional UNet models, one used for each. MR plane(coronal, sagittal, axial). The multichannel input for this architecture came from the modalities Fluid Attenuation Inversion Recovery, T1, T1ce, and T2 slices. On the BraTS 2020 dataset, the ensemble was trained and assessed, yielding dice coefficient (DC) values of 83.8%, 89.9%, and 85.3% for the tumor subregions E, W, and T, respectively. Incorporating a 3D ASPP module into a three-dimensional UNet, Y. Xu et al. [13] obtained dice DC values of 76.9%, 87.1%, and 77.9% for the tumor subcategories E, W, and T, respectively for the BraTS 2017 dataset. Varghese et al. [14] implemented a 23-layer deep fully connected 2D CNN based on the encoder-decoder framework. On the local testing set, of the BraTS 2017 dataset, this method produced DC values of 84.3%, 84.1%, and 77.3% for the tumor subcategories W, T, and, E, respectively. Ding et al.[15] developed a new two-dimensional architecture called the SMCSR-Net, which is based on the stacked UNet paradigm, and trained it on the BraTS 2015 datasets. On the local testing set, this technique produced DC values of 83%, 67%, and 59% for the tumor subcategories W, T, and E, respectively. Zhang et al.[16] greatly enhanced the performance of the standard two-dimensional UNet [8] by introducing residual building blocks to the original framework and adding gated attention units to the decoder. This approach yielded DC values of 87%, 77%, and 72% on tumor subcategories W, T, and E, respectively, on the local testing set of the BraTS 2017 dataset. Ilyas et al.[17] proposed the DANet, which employs a novel weight alignment technique using attention modules with multiple

dilation rates between the encoder and decoder skip links to promote enhanced feature mapping. On the BraTS 2018 dataset, this method produced DC values of 88%, 76%, and 65% for tumor subclasses W, T, and E, respectively. Ashraf et al.[18] proposed the ZNet for the semantic segmentation of low-grade glioblastomas. The proposed framework was trained and evaluated on the TCGA LGG dataset yielding a DC value of 91.5% on the testing data. A residual UNet created by Santosh et al.[19] was trained and evaluated on the TCGA LGG dataset, and the testing dataset produced a DC value of 90%. Using a pre-trained VGG-16 backbone from the imagenet dataset and training data from the TCGA LGG dataset, Sourodip et al.[20] constructed a U-Net model and obtained a DC value of 91.6% on test data.

In conclusion, several original methods for segmenting brain tumors have been explored, and some general conclusions may be made. Most of the works covered so far have been either trained and evaluated on datasets mostly made up of high-grade glioma (HG) volumes or exclusively on low-grade glioma (LG) slices. Due to their broad proliferation throughout healthy brain tissue, HGs are easier to segment than LGs, albeit class-based segmentation of tumor regions may prove to be quite difficult. An effective brain tumor segmentation model should function equally effectively on low-grade-glioma (LG) and high-grade-glioma (HG) volumes. This comes up as a major research gap in this field of study. In order to get over this constraint and show the effectiveness of the proposed model, it was trained and validated independently using the BraTS 2020 dataset and the TCGA LGG dataset. The findings show great promise for the segmentation of both LGs and HGs.

III. Materials and Methodology

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

A. Materials

In this study, two datasets have been used that are described in detail.

1) BraTS 2020 training dataset

The BraTS 2020 database [21], [22], [23] contains MRI volumes from 293 patients with HGs and 76 patients with LGs. T1, T1-enhanced (T1-ce), T2, and FLAIR MR volumes, along with physician segmentation outputs, were provided for each patient. The tumors have been manually categorized into three classes: edematous tissue, tissue having necrosis, and enhancing tumor region. Figure 1 illustrates exemplary 2D multimodal MR slices from the aforementioned dataset for patient volume no. 001, as well as the accompanying multiclass segmentation output produced by a physician..

2) TCGA LGG dataset

The TCGA-LGG dataset acquired from The Cancer Imaging Archive (TCIA) [10] 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, 1373 images had tumor regions and 2556 showed normal brain tissue. Each image has three channels-

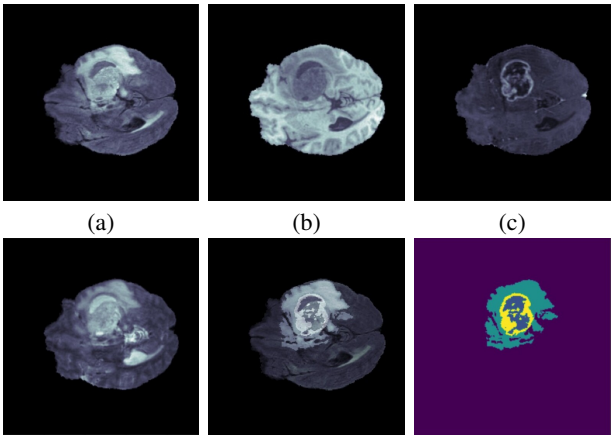


Fig. 1: Representative multimodal 2D slices from the BraTS 2020 training dataset and corresponding segmentation ground truth. (a) FLAIR modality, (b) T1 modality, (c) T1-ce modality, (d) T2 modality, (e) FLAIR modality superimposed with multiclass segmentation mask, (f) Multiclass segmentation mask.

T1, FLAIR, T2, and T1-ce. However, several slices were missing the T1 and T1-ce channels. In such cases, the FLAIR modality replaced 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.

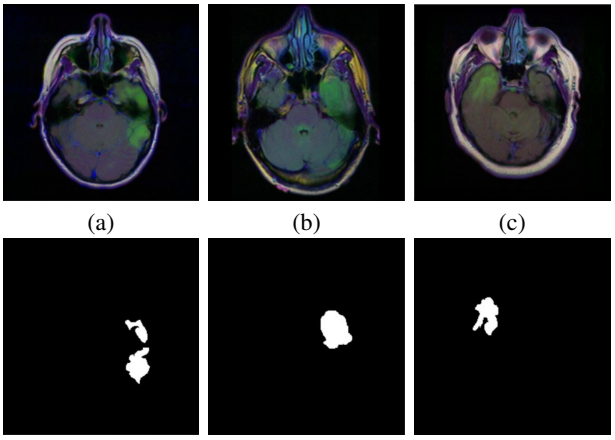


Fig. 2: Representative FLAIR 2D slices from the TCGA LGG dataset with the corresponding segmented ground truths. (d), (e) and (f) represent the corresponding ground truths for the FLAIR slices (a), (b), and (c) respectively.

B. Methodology

An overview of the methodology workflow for the implementation of the proposed method is provided in this section. Fig. 3 is a schematic representation of the workflow plan. The 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) Thresholding generated outputs and evaluation.

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

1) Preprocessing

From the BraTS 2020 training dataset, 2D slices were derived from each multimodal MR and their corresponding mask volumes. Slices containing at least one positive pixel in the corresponding mask slice were retained for subsequent processing, while the rest were discarded. 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: W, T, and E, 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. Figure 4(a) shows a schematic illustration of this technique.

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 (W). The FLAIR, T1-ce, and T2 MR slices are combined into a single 2D multichannel 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 256×256×3 prior to training. No additional preprocessing was performed to maintain the high-dimensional characteristics for optimal model performance. Fig. 4(b) shows a schematic illustration of this technique.

2) The proposed main architecture

The proposed network derives its name "GCSER-UNet" based on the repeated use of **s**queeze and **e**xcite blocks and employment of the **r**esidual learning approach at both the encoder and decoder. Fig. 5c shows a schematic illustration of the encoder and decoder model architecture. Fig 5a shows the Res block architecture which has been thoroughly discussed in the subsections that follow.

a) Res block

The residual block, along with its skip connection, has been termed the Res-block. The Res-block has been implemented as an integral architectural aspect of both the contractive and expansive paths, as shown in Fig. 5c. The residual network in each Res-block incorporates two sequential Conv2D layers, each preceded by Batch-Normalization and ReLU activation layers. Batch-Normalization aids in regularization by minimizing shifts in network activation distribution resulting from parameter fluctuations during training. The Dropout layer randomly weights node outputs from hidden layers to 0 with a predetermined rate 'r' (here 0.1) at each training step, which builds resilience to overfitting. Node outputs not set to 0 are scaled up by 1/(1 - r) such that the sum over all inputs remains constant. From Fig. 5a, the output of the Res block is the sum of the output of the convolutional path $Z_2(X)$ and the skip connection X. The final output R can be shown as eqation1. *(Here $BN()$, $\delta()$, and $D()$ represent the Batch Normalization, Relu Activation and Dropout operations respectively)*

$$Z_1(X) = W_1 \times \delta(BN(X)) \tag{1}$$
$$Z_2(X) = W_2 \times \delta(BN(Z_d)),\ where\ Z_d = D(Z_1) \tag{2}$$
$$R(X) = Z_2(X) + X \tag{3}$$
$$Z_2(x) = R(X) - X \tag{4}$$

The rearrangement of the terms in Equation 3 yields the expression for $Z_2(X)$, denoting the difference between the desired output and the input, commonly referred to as the *residue*, as indicated in Equation 4. Kaiming et al. [9] suggest that the network finds it easier to learn this "residue," supported by the observations from Equations 3 and 2. For $R(x)$ to represent the identity function X, the residue $Z_2(X)$ must be nullified, necessitating the weight matrix W_2 to approach zero. In contrast, in the absence of the skip connection $(R(x) = Z_2(x))$, the adjustments to the weights and bias values are imperative to conform to the identity function. Learning an identity function from scratch for a non-residual network is notably challenging, exacerbated by the non-linearity within the layers, leading to the degradation problem. Figure 5a presents a schematic representation of the sequential layers within a Res-block, with further specifics available in Table I. Here, f_i denotes the number of filters in the i^{th} encoder block, k represents the kernel size, d signifies the dropout percentage and s corresponds to the stride. To maintain consistent input and output feature map dimensions, zero padding was applied for each convolution.

TABLE I: Sequence of layers in the Res-block, where layer 8 represents the skip connection.

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

b) GCSE Block

The Global Context-aware Squeeze and Excitation (GCSE) block represents an advanced extension of the conventional Squeeze-and-Excitation (SE) block, designed to capture comprehensive global context information within neural networks. Initially, in the Squeeze stage, the mean $(M_{(k)})$ and standard deviation $(S_{(k)})$ for the k^{th} channel are computed from the input tensor (X) through equations 5 and 6.

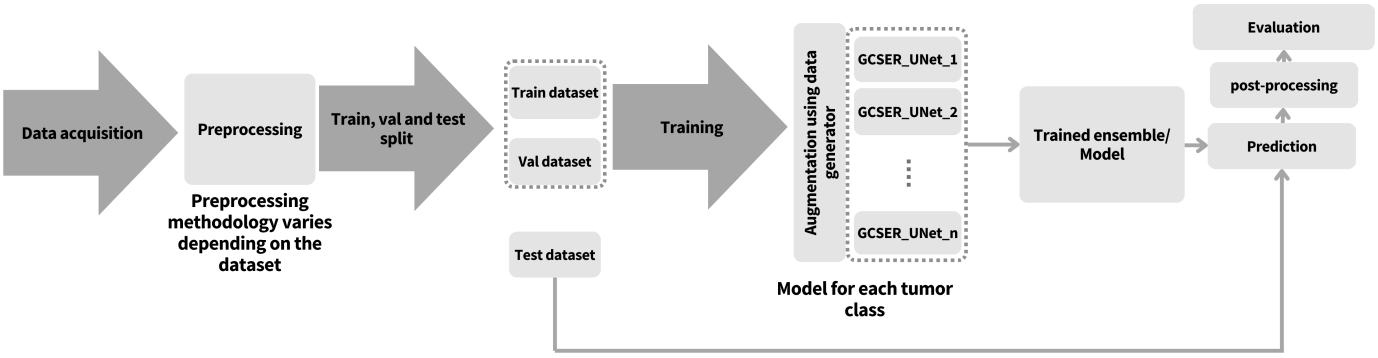
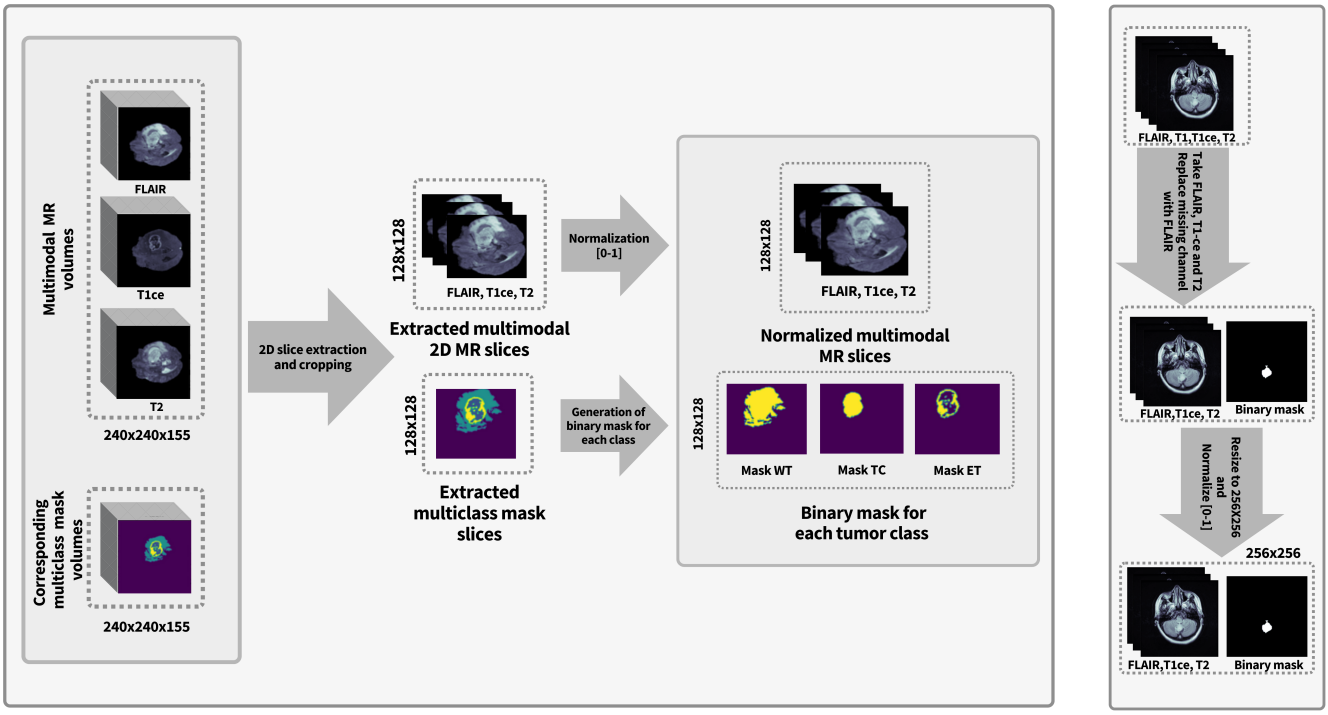


Fig. 3: Schematic representation of the methodology workflow.



(a) (b)

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

$$M_{(k)} = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X(i,j,k) \tag{5}$$

$$S_{(k)} = \sqrt{\frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} (X(i,j,k) - M_{(k)}^2)} \tag{6}$$

These operations extract vital channel-wise statistics, enabling the model to understand essential channel dependencies. Subsequently, in the Excitation stage, the channel-wise attention (A_c) is obtained using the equation 7.

$$A_c = \sigma(W_2 \delta(W_1 \begin{bmatrix} M \\ S \end{bmatrix})) \tag{7}$$

where σ and δ represent the sigmoid and ReLU activation functions, respectively. The parameter ratio (r) controls the dimensionality reduction in the intermediate layers, impacting the expressiveness and computational complexity of the GCSE block. Simultaneously, the spatial attention (A_s) is derived through equation 8.

$$A_s = \sigma(W_3 \delta(\frac{1}{n} \sum_{k=1}^{n} X(i,j,k))) \tag{8}$$

The combination stage integrates the outputs of the channel-wise and spatial attention, generating a comprehensive attention map (A) given by equation 9. Finally, the feature rescaling stage dynamically reweights the input features (X) according

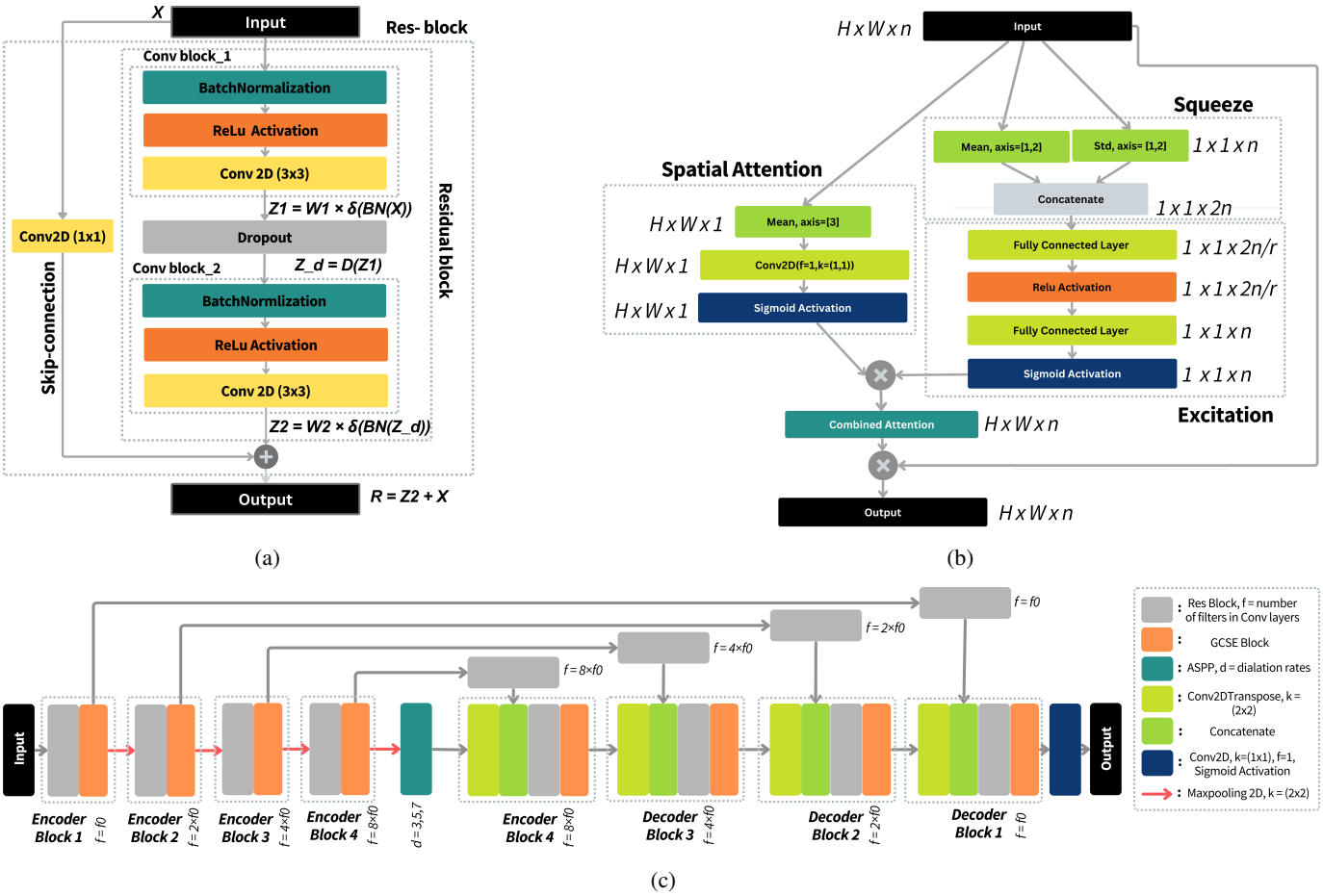


Fig. 5: (a) Schematic of Res-Block, (b) Schematic of the proposed GCSE mechanism, (c) Structural overview of the proposed GCSER-UNet

to the attention map (A), as given by equation 10.

$$A = A_c \odot A_s \tag{9}$$

$$Y = X \odot A \tag{10}$$

Through these intricately orchestrated mathematical operations, the GCSE block significantly enhances the network's ability to discern complex patterns and salient features within the data, thereby fostering an enriched understanding of global context and intricate relationships.

c) Encoder

The encoder of the GCSER-UNet is predominantly a cascade of structurally repeating units known as encoder blocks, with each successive unit in the cascade having double the number of feature channels compared to its preceding counterpart. Each encoder block is composed of a Res-block followed by a GCSE block, with the integration of the GCSE blocks significantly enhancing model performance through dynamic feature recalibration. This recalibration facilitates nuanced channel-wise and spatial attention fusion, vital for capturing complex inter-channel dependencies and extracting essential global context information, thereby ensuring precise and accurate brain tumor segmentation.

The output from each encoder block follows two paths: the first path, known as the (long) skip connection, directs the output feature map through a Res-block and then to the corresponding decoder block. The second path subjects the output feature map to a MaxPooling2D operation with a stride of 2, effectively reducing its spatial dimensions by half before feeding it as input to the subsequent encoder block. This sequential downsampling and increase in the number of feature channels enhance the model's receptive field, facilitating the extraction of more contextual information.

The pooled output from the final encoder block undergoes processing via an Atrous Spatial Pyramidal Pooling (ASPP) module, incorporating atrous convolutions at multiple dilation rates and spatial pyramidal pooling. This integration allows the model to capture contextual information at different scales, a crucial aspect for brain tumor segmentation tasks, given the complex nature of brain MR images, which possess abundant local intricacies and virtually limitless macro target expansion. The proposed model utilizes four parallel 3×3 convolutions with dilation rates of 3, 5, and 7, respectively, within the ASPP module, with the image-level feature generated using global average pooling. The features from all branches are upsampled to the input size using bilinear interpolation, concatenated, and passed through another 1×1 Conv2D layer, with the resulting output provided to the decoder.

d) Decoder

The decoder, like the encoder, is made up of structurally repeating units. There is a corresponding decoder block for each encoder block. Each decoder block produces a twofold increase in the size of the input feature map by performing a 2×2 Conv2DTranspose operation on it. Following that, the upsampled output is concatenated with the skip feature maps acquired from the corresponding encoder block. Following concatenation, the resulting feature maps are sent via a Res-block (see section III-B2a) and then a GCSE unit. A 1×1 Conv2D operation with $sigmoid$ activation on the output of the last decoder block creates the final output.

IV. Experimental results

A. Implementation detail

The code was written in TensorFlow version 2.6.4, and the models were trained on a cloud-based server using an NVIDIA Tesla P100 GPU.

B. Experimental settings

A 75:15:10 train-validation-test split ratio was employed for each dataset.*All the results presented in this paper are derived from this test set*. 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 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 was used for training the models. The models took about 50 epochs to converge for both datasets.

C. Evaluation metrics

The model's effectiveness was assessed using the following metrics:

(1) **Loss function** As shown in eqn 11, the loss function L_{FD} utilized in this work is a mix of Dice loss and Focal loss.

$$L_{FD} = L_{Dice} + L_{Focal} \tag{11}$$

where L_{Dice} and L_{Focal} are Dice loss and Focal loss, respectively.
Dice loss may be defined as stated in eqn 12 and is dependent on the Dice coefficient.

$$L_{Dice} = 1 - \text{Dice coeff} \tag{12}$$

Focal loss is a binary cross-entropy loss version that tackles the issue of class imbalance with the conventional cross-entropy loss by down-weighting the contribution of easy-to-categorize samples, allowing learning of tougher cases. It is defined in eqn 13.

$$L_{Focal(P_T)} = \alpha(1 - E_t)^{\gamma} \cdot L_{BCE(P,y)} \tag{13}$$

Here, L_{BCE} is the Binary Cross Entropy loss. The likelihood of correctly guessing the class of ground truth, E_t, is defined as shown in eqn 14.

$$E_t = \begin{cases} E, & \text{if } y = 1 \\ 1 - E, & \text{if } y = 0 \end{cases} \tag{14}$$

The degree to which easy-to-classify cases are down-weighted so that learning the challenging instances can receive more attention is controlled by the parameters α and γ. For γ = 0, the Focal loss is simplified to the Binary Cross Entropy loss.

(2) **Dice coefficient**:
The Dice coefficient is calculated by taking twice the intersection of the two sets and dividing it by the sum of their sizes. In the context of image segmentation, the sets represent the pixels or voxels within the predicted and ground truth regions. Mathematically it can be represented as shown in eqn 15.

$$Dice = \frac{2 \times |T \cap \overline{T}|}{|T \cup \overline{T}|} \qquad where,\ Dice \in [0,1] \tag{15}$$

Here, Class T: tumor, Class $\overline{T}$: non-tumor. It can also be represented as eqn 16.

$$Dice = \frac{2 \times Pos}{2 \times Pos + \overline{Pos} + \overline{Neg}}\ where,\ Dice \in [0,1] \tag{16}$$

Here Pos = instances from true positive class, $\overline{Pos}$ = instances from false positive class, Neg = instances from true negative class, $\overline{Neg}$ = instances from false negative class.

(3) **IoU**:
IoU quantifies the extent of agreement or overlap between the predicted location of an object (determined by a machine learning model) and the ground truth location (manually labeled). It provides a measure of how well the predicted region aligns with the actual region of the object. It can be represented as shown in 17.

$$IoU = \frac{|T \cap \overline{T}|}{|T \cup \overline{T}|} = \frac{Pos}{Pos + \overline{Pos} + \overline{Neg}}\ where,\ IoU \in [0,1] \tag{17}$$

where, Class T: tumor and Class $\overline{T}$: non-tumor.
(4) **Sensitivity:**
The metric sensitivity measures how well a model can predict true positives for each class that is provided. It is represented as shown in eqn 18.

$$Sensitivity = \frac{Pos}{Pos + \overline{Neg}} \tag{18}$$

(5) **Specificity:**
The metric of specificity measures the ability of a model to predict true negatives for each class that is provided. It is represented by the eqn 19.

$$Specificity = \frac{Neg}{Neg + \overline{Pos}} \tag{19}$$

D. Ablation Results

1) Performance on the TCGA LGG dataset

Among the 1373 subjects exhibiting positive mask values, a selection of 1098 subjects was utilized for training purposes. Additionally, 138 subjects were allocated for validation, while the remaining 137 subjects were reserved for testing. The

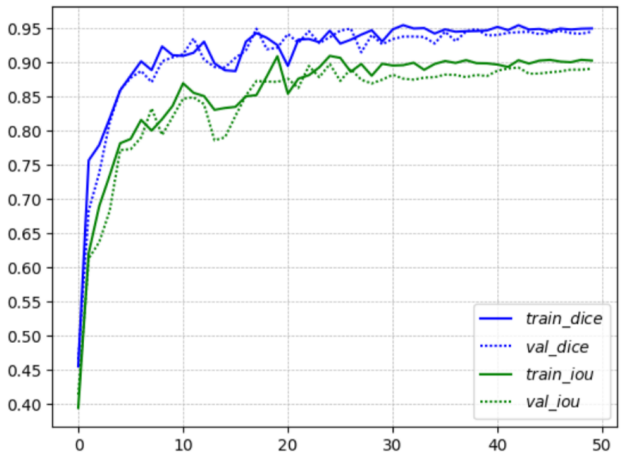


Fig. 6: Training-validation curves for Dice Coefficient and IoU for TCGA LGG dataset

model took around 50 epochs to converge. The training-validation curves can be seen in Fig 6.

To assess the impact of specific components within the proposed architecture on segmentation performance, the segmentation results on the test set were analyzed using the following configurations: (i) U-Net (ii) Res-UNet, (iii) SE-Res-UNet, (iv) SE-Res-UNet + ASPP, and (v) GCSER-UNet (*Here 'SE' denotes the original squeeze and excite mechanism as introduced by [10]*). The performance metric values for the ablation study are available in Table II.

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

Model	Dice	IoU	Sensitivity	Specificity
UNet	0.77	0.693	0.781	0.9991
Res-UNet	0.83	0.722	0.820	0.9995
SE-Res-UNet	0.88	0.803	0.881	0.9996
SE-Res-UNet + ASPP	0.901	0.83	0.907	0.9997
Proposed GCSER-UNet	**0.94**	**0.89**	**0.977**	**0.9997**

The performance gains achieved by virtue of the architectural enhancements to the UNet leading to the GCSER-UNet can be clearly seen from the comparison of the segmentation masks produced by the various models presented in Fig 7.

2) Performance on the BraTS 2020 dataset

Slices from the complete set of 369 labeled subjects in the BraTS 2020 dataset were extracted in the form of 2D images and masks. Only images with associated positive mask values were selected for subsequent processing and training. For each class, only slices with positive mask values were taken into account. The obtained set of slices underwent further processing and a train-validation-test split of 75:15:10 before the training phase. Each model for the ensemble took approximately 50 epochs to reach convergence. The training-validation curves can be seen in fig 8. To assess the efficacy of each individual building block of the model, an ablation study akin to that carried out for the TCGA LGG dataset was conducted. Table III reports the ablation study's findings based on the testing data, and fig 9 displays some of the qualitative segmentation results from the ablation study.

V. Discussions

This section presents an evaluation of the effectiveness of the GCSER-UNet model in comparison to state-of-the-art techniques. Additionally, the significant outcomes and implications resulting from the proposed model are outlined.

A. An evaluation of the GCSER-UNet in comparison to state-of-the-art techniques.

The performance comparison of the suggested model has been laid out with respect to both TCGA LGG dataset and the BraTS dataset.

1) TCGA LGG dataset

A comprehensive comparative analysis was conducted to assess the efficacy of the suggested method in comparison to several state-of-the-art techniques, as reported in Table IV. The results indicate that the GCSER-UNet model achieved the highest dice score of 92.3% and mean IOU score of 85.5%, surpassing the performance of existing techniques.

2) BraTS 2020 dataset

The trained ensemble was evaluated against some of the state-of-the-art techniques and the comparative findings have been enlisted in Table V. The proposed GCSER-UNet ensemble method obtained the best mean dice score of 92%, 87% and 84% for tumor subcategories-W, T, and, E respectively.

The followings are the outlines of the most important findings obtained from the proposed model:

1) A single GCSER-UNet model was trained for the binary segmentation of brain tumors belonging to TCGA LGG dataset whereas an ensemble of three GCSER-UNets has been used for the multiclass segmentation of brain tumors on the BraTS 2020 dataset.
2) The suggested architecture achieved 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.
3) The gains in performance are a consequence of the improvements made to the original UNet[8] framework. A number of studies that improved the performance of the UNet [8] by introducing architectural improvements were examined. Taking cues from these works[12], [13], [14], [20] a multitude of changes were introduced to the UNet[8] to develop a model that caters effectively to the task of brain tumor segmentation.
4) The diverse features of tumor subcategories (W, T, and E) manifest distinctly in MR slices acquired from different modalities. To optimize segmentation results, the GCSE blocks introduce channel-wise weighting, ensuring that channels with heightened tumor-related information exert a more substantial influence on the output feature map. Simultaneously, the spatial attention mechanism facilitates effective discrimination between tumor and non-tumor tissue, enhancing the model's segmentation accuracy.

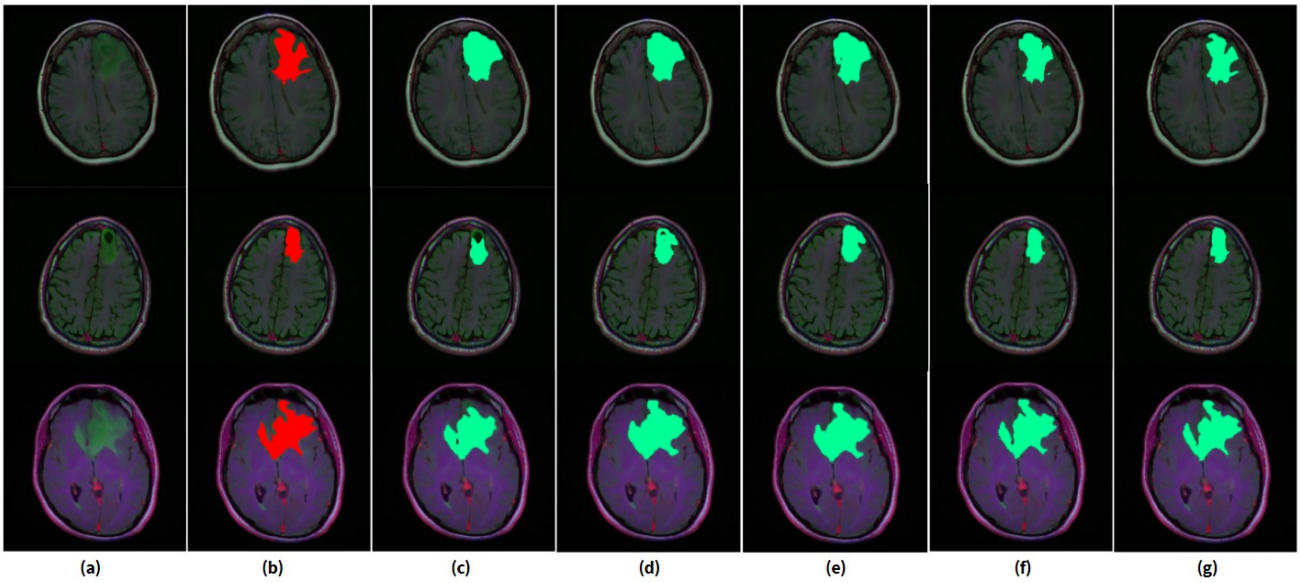


Fig. 7: A comparison of the segmentation results produced by the GCSER-UNet and different variants of the UNet on random slices from the TCGA LGG testing data. (a) Original FLAIR image, (b) FLAIR image with ground-truth, (c) Predicted mask: UNet, (d) Predicted mask: Res-UNet, (e) Predicted mask: SE-Res-UNet, (f) Predicted mask: SE-Res-UNet+ASPP, (g) Predicted mask: GCSER-UNet

TABLE III: Performance evaluation of the proposed ensemble against several UNet architecture variations using test data from the BraTS 2020 dataset.

Architecture	Dice			IoU			Sensitivity			Specificity		
	W	T	E	W	T	E	W	T	E	W	T	E
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	0.88	0.86	0.88	0.82	0.79	0.93	0.92	0.89	0.9993	0.9994	0.9992
Proposed GCSER-UNet	**0.95**	**0.92**	**0.90**	**0.91**	**0.90**	**0.87**	**0.98**	**0.98**	**0.95**	**0.9997**	**0.9997**	**0.9997**

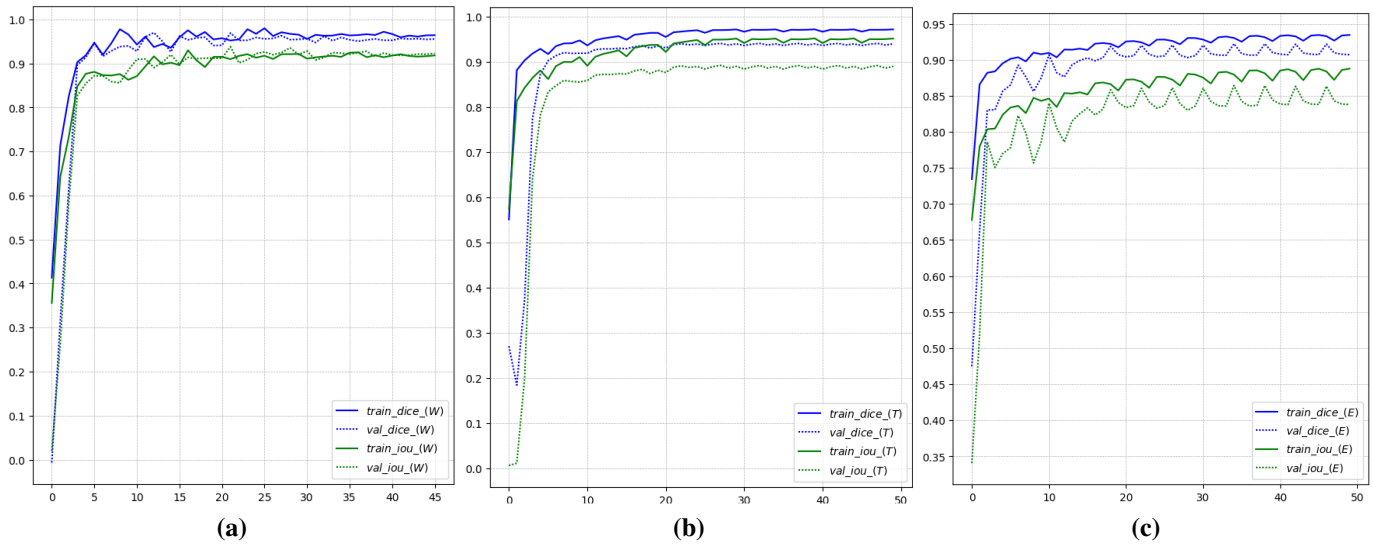


Fig. 8: Traning validation curves for the metric Dice Coefficient and IoU for (a) GCSER-UNet_W, (b) GCSER-UNet_T, and (c) GCSER-UNet_E for the BraTS 2020 dataset

5) The Residual building blocks were used in both the contractive as well as the expansive path to tackle the degradation problem.
6) 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 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.

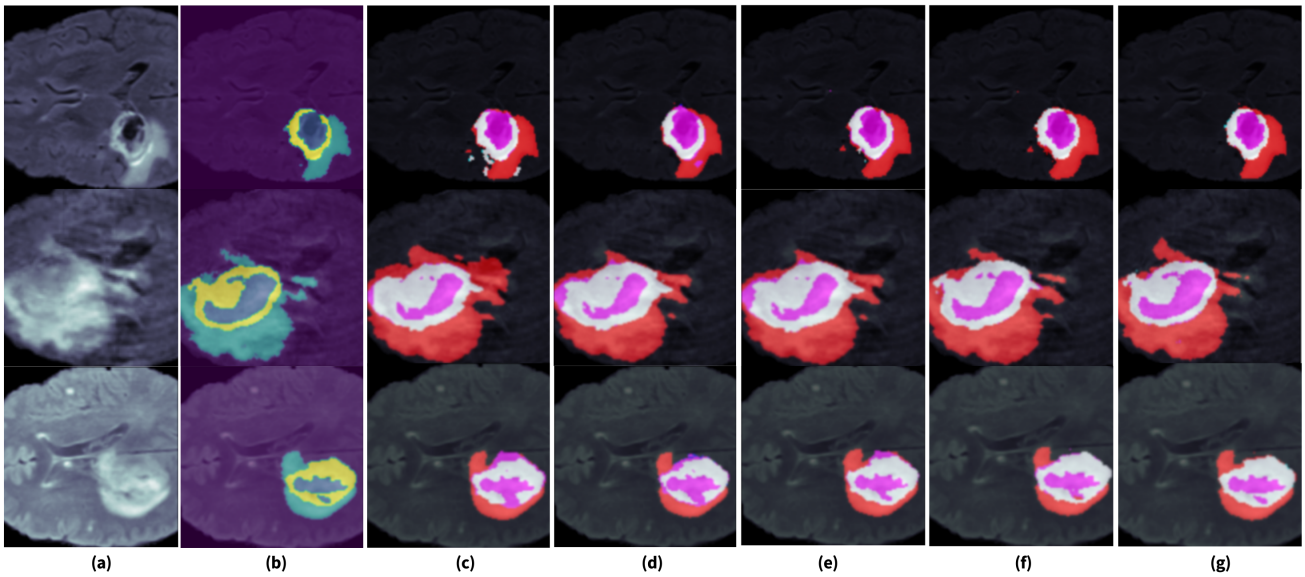


Fig. 9: A comparison of the segmentation results produced by the GCSER-UNet and different variants of the UNet on random slices from the BraTS 2020 testing data. (a) Original FLAIR image, (b) FLAIR image with ground-truth, (c) Predicted mask: UNet, (d) Predicted mask: Res-UNet, (e) Predicted mask: SE-Res-UNet, (f) Predicted mask: SE-Res-UNet+ASPP, (g) Predicted mask: GCSER-UNet

TABLE IV: Results using the TCGA LGG test set for the suggested method and comparison with other state-of-the-art techniques.

Method	Dice	IoU
Santosh et al.[19]	0.905	0.829
Buda et al.[24]	0.915	0.840
Venkata et al.[25]	0.870	0.780
Naser et al.[26]	0.905	0.829
Sourodip et al.[20]	0.918	0.826
Proposed GCSER-UNet	**0.94**	**0.89**

TABLE V: Comparison of the results of the proposed ensemble (GCSER-UNet_W, GCSER-UNet_T, and GCSER-UNet_E) with some of the most advanced techniques on the local test set of the BraTS 2020 training dataset

Method	Dataset	Dice		
		W	T	E
Sundaresan et al.[12]	BraTS 2020	0.89	0.85	0.83
Yanwu Xu et al.[13]	BraTS 2017	0.87	0.78	0.77
Varghese et al.[14]	BraTS 2017	0.84	0.84	0.77
Ding et al.[15]	BraTS 2015	0.83	0.67	0.59
Zhang et al.[16]	BraTS 2017	0.87	0.77	0.72
Ilyas et al.[17]	BraTS 2018	0.88	0.76	0.65
Ballestar et al.[27]	BraTS 2020	0.85	0.85	0.77
Findon et al.[28]	BraTS 2020	0.91	0.84	0.77
Aboelenein et al.[29]	BraTS 2020	0.87	0.84	0.80
(nn_UNet) Hou et al.[30]	BraTS 2020	0.94	**0.93**	0.88
Proposed GCSER-UNet ensemble	**BraTS 2020**	**0.95**	**0.92**	**0.90**

VI. Conclusion

This paper introduces the application of the unique 2D CNN architecture, GCSER-UNet, for precise brain tumor segmentation. The proposed method exhibits notable enhancements over existing techniques, as evidenced by the achieved dice scores (DC) of 95%, 92%, and 90% for tumor subregions W, T, and E, respectively, on the BraTS 2020 dataset. Notably, these results surpass the current state-of-the-art benchmark DC values of 93.93%, 92.82%, and 88.37% (nn_UNet [30]). Furthermore, when tested on the TCGA LGG dataset, the model demonstrated a remarkable DC value of 94.3%, outperforming the current state-of-the-art benchmark score of 91.8%. These impressive results can be attributed to the novel GCSE mechanism, effectively facilitating precise feature extraction and segmentation.

It is essential to note that all comparisons presented in this paper are based on the outcomes reported on their respective local test sets by the other authors. While the proposed model demonstrates its efficacy in accurately segmenting high-grade (HG) and low-grade (LG) volumes, future research endeavors could explore the integration of contextual information from multiple planes, an aspect that remains unexplored in this study. Despite the potential rise in processing costs, this avenue of research holds promise for further refining the segmentation effectiveness of the proposed approach, offering a viable strategy to address the limitations observed in this study.

Acknowledgement

All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this paper.

Code availibility

All code and implementation details can be found in https://github.com/Sourjya261/Brain_tumor_segmentation which will be made publicly available post-publication.

References

[1] Anna M Di Giacomo, Maximilian J Mair, Michele Ceccarelli, Andrea Anichini, Ramy Ibrahim, Michael Weller, Michael Lahn, Alexander MM Eggermont, Bernard Fox, and Michele Maio. Immunotherapy for brain metastases and primary brain tumors. *European Journal of Cancer*, 179:113–120, 2023.

[2] Xiaofei She, Shijun Shen, Guang Chen, Yaqun Gao, Junxian Ma, Yaohui Gao, Yingdi Liu, Guoli Gao, Yan Zhao, Chunyan Wang, et al. Immune surveillance of brain metastatic cancer cells is mediated by ifitm1. *The EMBO Journal*, page e111112, 2023.

[3] Hanne Blakstad, Jorunn Brekke, Mohummad Aminur Rahman, Victoria Smith Arnesen, Hrvoje Miletic, Petter Brandal, Stein Atle Lie, Martha Chekenya, and Dorota Goplen. Survival in a consecutive series of 467 glioblastoma patients: Association with prognostic factors and treatment at recurrence at two independent institutions. *PloS one*, 18(2):e0281166, 2023.

[4] Ikram Hasan, Shubham Roy, Bing Guo, Shiwei Du, Tao Wei, and Chunqi Chang. Recent progress in nanomedicines for imaging and therapy of brain tumors. *Biomaterials Science*, 2023.

[5] Zachary S Mayo, Ahmed Halima, James R Broughman, Timothy D Smile, Martin C Tom, Erin S Murphy, John H Suh, Simon S Lo, Gene H Barnett, Guiyun Wu, et al. Radiation necrosis or tumor progression? a review of the radiographic modalities used in the diagnosis of cerebral radiation necrosis. *Journal of Neuro-Oncology*, pages 1–9, 2023.

[6] Zhihua Liu, Lei Tong, Long Chen, Zheheng Jiang, Feixiang Zhou, Qianni Zhang, Xiangrong Zhang, Yaochu Jin, and Huiyu Zhou. Deep learning based brain tumor segmentation: a survey. *Complex & Intelligent Systems*, 9(1):1001–1026, 2023.

[7] Srigiri Krishnapriya and Yepuganti Karuna. A survey of deep learning for mri brain tumor segmentation methods: Trends, challenges, and future directions. *Health and Technology*, pages 1–21, 2023.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] Naveed Ilyas, Yoonguu Song, Aamir Raja, and Boreom Lee. Hybriddanet: An encoder-decoder based hybrid weights alignment with multi-dilated attention network for automatic brain tumor segmentation. *IEEE Access*, 10:122658–122669, 2022.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] 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.

[23] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, 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.

[24] 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.

[25] Pattabiraman Ventakasubbu and Parvathi Ramasubramanian. Deep learning-based brain tumour segmentation. *IETE Journal of Research*, pages 1–9, 2021.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] Qingfan Hou, Zhuofei Wang, Jiao Wang, Jian Jiang, and Yanjun Peng. Diffraction block in extended nn-unet for brain tumor segmentation. In *International MICCAI Brainlesion Workshop*, pages 174–185. Springer, 2022.