

Research Article

Suchismita Das*, Srijib Bose, Gopal Krishna Nayak, and Sanjay Saxena

Deep learning-based ensemble model for brain tumor segmentation using multi-parametric MR scans

<https://doi.org/10.1515/comp-2022-0242>

received May 28, 2020; accepted April 15, 2022

Abstract: Glioma is a type of fast-growing brain tumor in which the shape, size, and location of the tumor vary from patient to patient. Manual extraction of a region of interest (tumor) with the help of a radiologist is a very difficult and time-consuming task. To overcome this problem, we proposed a fully automated deep learning-based ensemble method of brain tumor segmentation on four different 3D multimodal magnetic resonance imaging (MRI) scans. The segmentation is performed by three most efficient encoder–decoder deep models for segmentation and their results are measured through the well-known segmentation metrics. Then, a statistical analysis of the models was performed and an ensemble model is designed by considering the highest Matthews correlation coefficient using a particular MRI modality. There are two main contributions of the article: first the detailed comparison of the three models, and second proposing an ensemble model by combining the three models based on their segmentation accuracy. The model is evaluated using the brain tumor segmentation (BraTS) 2017 dataset and the *F1* score of the final combined model is found to be 0.92, 0.95, 0.93, and 0.84 for whole tumor, core, enhancing tumor, and edema sub-tumor, respectively. Experimental results show that the model outperforms the state of the art.

Keywords: U-Net, SegNet, MCC, MRI, Glioma

1 Introduction

The task of evaluating and extracting correct information from pathologic seimen is a very difficult due to their subjective and complex nature. From the statistical analysis, 85–90% of all primary central nervous system (CNS) tumors are accounted for by brain tumors. Currently, diagnosis of newly affected brain and CNS cancer cases were found which are responsible for about 3% of all other cancers [1]. In European countries, these cases are 5 times higher [1] than in Asian countries. For early-stage diagnosis and recovery, automatic segmentation of brain tumors is required which is very expensive and time-consuming if done manually. Image processing and computer vision have shown major advances in the automatic extracting region of interest and useful information for diagnosis which helps in further planning of treatment.

Cancerous and non-cancerous brain tumors, when grown, can cause brain damage. A solid neoplasm inside the skull is termed a brain tumor which is caused by uncontrolled and abnormal growth of tissue or cells [2]. Malignant tumors are divided into two types: primary tumors originating within the brain, and secondary tumors, spread to the brain originating from other organs of the body. Glioma is one of the most common types of primary tumor which is developed from glial cells. It is categorized into Low-Grade Glioma (LGG) which grows slowly and becomes highly malignant tumors known as High-Grade Glioma (HGG), which can be life-threatening. The World Health Organization reports [3] that HGG tumors are more critical having a maximum of 2 years of survival rate, while LGG-affected patients can have several years of life expectancy. These tumors are still untreatable [4] in a few cases even having advanced imaging, radiotherapy, and surgical methodology. Among the various imaging techniques, Magnetic resonance imaging (MRI) characteristics are considered one of the preferred ways to diagnose [5–7] a brain tumor as it creates a more detailed and clear picture than CT scans. Different contrast images produced through

* **Corresponding author: Suchismita Das**, Computer Science & Engineering, IIIT Bhubaneswar, Bhubaneswar, Odisha, 751003, India; KIIT University, Odisha, 751024, India, e-mail: suchismita.dasfcs@kiit.ac.in

Srijib Bose: Computer Science & Engineering, KIIT University, Odisha, 751024, India

Gopal Krishna Nayak, Sanjay Saxena: Computer Science & Engineering, IIIT Bhubaneswar, Bhubaneswar, Odisha, 751003, India

multimodal MRI protocols provide complementary information which helps in segmenting the brain tumor and its surrounding tissues that are playing an important role in the diagnosis and treatment of brain tumors. There are mainly three regions in brain tumors such as necrotic and non-enhancing tumor, peritumoral edema, and enhancing tumors. The most common MRI sequences are native or T1-weighted; post-contrast T1-weighted (T1ce), T2-weighted, and Fluid Attenuated Inversion Recovery (FLAIR) volumes. Each one of these above modalities has different contrast which helps in segmenting [5] different regions of the tumors through their characteristics.

Automatic segmentation of the whole brain tumor along with its surrounding regions is the most challenging task and crucial for early diagnosis and treatment progress. There are different methods proposed for automatic segmentation of brain tumors using traditional methods as well as machine learning [8] approaches. As described in ref. [9], machine learning algorithms, such as neural network SVM, clustering, etc., outperform the traditional methods such as threshold-based, region growing, and watershed. But machine learning algorithms are based on selecting the feature vector which influences the performance in a greater way. To overcome this, nowadays deep learning is achieving a great success and drawing researcher's attention in medical image segmentation [10,11], in which the deep features of the images are extracted automatically through a series of convolutional operations and commonly used Convolution neural network (CNN) models are modified in different ways to make it flexible for image segmentation operations.

In this article we ensemble a fully automated deep learning-based brain tumor segmentation technique by combining different encoder and decoder convolutional models [12,13]. In our work, we have considered U-Net and SegNet models [14–17] which are the most widely promising encoding and decoding models for schematic image segmentation. U-Net uses the skip connection during the up-sampling whereas the SegNet uses the pooling indices which reduce the memory usage of the model. A detailed comparative study of these two models along with the encoding–decoding model was performed. The model performances were statistically evaluated on the brain tumor segmentation (BraTS) 2017 dataset for segmenting the different tumor structures [14] grouped into four tumor regions which are used in practical clinical applications:

- Complete tumor: consists of all the intra-tumor classes named as necrosis and non-enhancing, edema, and enhancing tumor in BraTS 2017 dataset.

- The core tumor: consists necrosis and non-enhancing, and enhancing tumor in BraTS 2017 annotation.
- The enhancing tumor: consists of enhancing part of the whole tumor.
- Peritumoral edema tumor: consists of the edema region of the whole tumor.

No single model can extract the different sub-parts of the tumor along with the complete tumor with the same accuracy. Also, the different MRI modalities have different characteristics which help to segment different sub-regions with unequal potential. It means each modality is not performing equally during training to extract the different sub-parts of the whole tumor. To overcome this and take advantage of each modality and model, a method is developed by fusion of features to select the most effective information from different modalities with the highest accuracy. Finally, an ensemble model combines the most effective architecture with the related characterized multimodal MRI scan to produce an aggregated result with maximized accuracy. The main contributions of this article are as follows:

- Automatic brain tumor segmentation method that uses informative image slices taken from 3D multimodal MRI volume to reduce computational time and increase segmentation accuracy.
- Effective extraction of deep features using U-Net, Encoder–Decoder model, and SegNet through each of the multimodal MRI scans.
- A detailed comparative study of U-Net, Encoder–Decoder model, and SegNet for brain tumor segmentation.
- A hybrid model which is a combination of the above existing model based on different modalities so that the result of segmentation can be maximized.

2 Related works

There are still several challenges existing in semi-automatic and fully automatic brain tumor segmentation due to the highly variable characteristics of brain tumors such as size, shape, location, and appearance. Brain tumor segmentation can be broadly categorized as supervised and unsupervised algorithms. Numerous research and development on brain tumor segmentation are going on by using traditional thresholding methods to deep learning techniques [6,18,19]. Here we focused on some recent and closely relevant studies related to the proposed topic.

There are a number of machine learning supervised algorithms and discriminative methods such as SVM,

Decision Tree (DT), K-means Clustering, Peak-Valley algorithms etc., which are used in different medical diagnoses and proposed segmentation as discussed in refs. [9,20]. In all these above methods, the dataset is trained based on selected features which are the most crucial step in machine learning. Currently, a deep learning CNN is attracting attention in the medical domain due to its automatic extraction of features from the given input image. It performs a set of convolutional operations on the input image and finally produces a feature vector that is responsible for characterizing the given image in different fields of research. The deep learning-based U-Net model and SegNet model are the most commonly used methods for researchers due to their performance in medical image segmentation [4,12]. Dong et al. [4] proposed a fully automatic brain tumor segmentation using U-Net and it was evaluated on the HGG and LGG datasets of BraTS 2015. They achieved a good dice score coefficient (DSC) for core and complete but enhancing results was not so good. Similarly, Alqazzaz et al. [14] developed a model to automatically segment brain tumors by focusing on the most effective features and information extraction from different MRI scans using a 3D dataset. Then, the MRI voxels are classified into the tumor and sub-tumor parts using the SegNet model followed by a DT. It takes nearly 3 days on a single NVIDIA GPU Titan XP and gives better accuracy for core and enhanced regions as compared to the state of the art. Yi et al. [12] proposed a 3D CNN deep learning model based on three trained convolutional layers for the extraction of brain tumors. It was evaluated on the BraTS 2015 and BraTS 2013 datasets and found to be an 89% dice score for whole tumor segmentation. A method based on RA-U-Net is developed by Jin et al. [21] in which the tumor is segmented from the extracted volume of interests (VOI) through the proposed model. The method has a similar structure to 3D U-Net which is based on contextual information in the encoding path. Down-sampling, mainly in U-Net architecture, combines the low-level feature maps with high-level ones which helps in accurate semantic segmentation. The method was evaluated on MICCAI 2017 liver tumor segmentation dataset and then was extended to BraTS 2017 brain dataset. It shows satisfactory results for both the datasets. Kaldera et al. [22] proposed a faster region-based CNN (RCNN) deep learning model for Glioma segmentation using MRI. It is fully autonomous in extracting the tumor part and also time and computational cost-effective as it takes 100 images and 23 images of size 128×128 for training and testing, respectively.

An improved FCN [23] incorporated many post-processing methods for automated liver segmentation from

abdominal CT scans. For automatic liver detection, training of the fully convolutional network (FCN) model was done followed by certain post-processing methods which used energy functions like graph-cut-based method, CRF, and level set-based method refinement of segmentation. Taking clues from the U-Net paper, this article has put forward an FCN with 21 layers for the computation of raw probability. The contracting layers or paths consist of three pooling and eight convolutional layers. The expanding route substitutes for the pooling layers with de-convolutional layers accompanied by Rectified Linear Unit (ReLU). The pooling size and the kernel size of the convolution layers are 2×2 and 3×3 , respectively. The high-resolution features are then moved to the respective level of the expanding path for integration with up-sampled features on each level of the contracting path. To enable the network to be accessible for any image size, the images are padded with 1 pixel before being fed to the convolutional layers [24]. By this, the size of the output will be equivalent to that of the original image. Between the ReLU and the convolution layer, they have introduced a Batch Normalization layer. After convolution, normalization of each batch is done with its standard and mean deviation. The major advantage of this framework is: (1) the improvement of FCN for accurate and better segmentation results and (2) it is the comparative study of the performance of various segmentation models based on the post-processing step. The 3DIRCADb database is used for the proposed model.

In a 3D model or convolutional operation, the depth information is captured in detail due to correlation between the slices, but due to less dataset, it may lead to over fitting and also be time-consuming. As the medical data are not available in huge numbers, the 2D slices can be used to avoid over-fitting. U-Net and SegNet models both outperform in image segmentation and the MRI modalities yield a valuable diagnosis in the majority of cases. In this article, we combine different modalities to segment the whole tumor and its subparts using U-Net, SegNet, and Encoder-Decoder models, and compare the models in detail to get better result for automatic segmentation of brain tumors at an early stage.

3 Methodology

The main aim of this study is to develop a brain tumor segmentation algorithm in order to segment the four sub-tumor parts for all the modalities by comparing the existing deep learning models. Here we proposed a two-layered ensemble deep model which segments the whole tumor

and its subparts in three main steps: a data pre-processing step to remove biases and do the normalization of the given MRI dataset, a first layer to train Encoder–Decoder, SegNet, and U-Net models for learning the parameter of segmentation through all four modalities, and an ensemble layer which extracts the maximized feature maps from different model and concatenation of feature maps for better segmentation accuracy.

3.1 Data pre-processing

In this study, we have taken the HGG MRI of the BraTS 2017 dataset for evaluating the results. These MRIs have different artifacts such as motion and intensity inhomogeneity, and are also acquired through different scanners. To validate the algorithm and results, the artifact must be removed through normalization and bias correction.

Initially, to reduce the intensity variation between the surrounding pixels which in turn reduces the noise in an image, a median filter is applied to all 3D MRI modalities of each patient. Then, N4ITK bias field correction [5,25] is applied to all the images of all modalities in order to improve performance by removing unwanted artifacts. Finally, as the intensity values of the images differ from patient to patient, a normalization step is followed which brings the intensities of all the images to a common level. This is done by bringing the mean value of the intensity close to 0 and the standard deviation close to 1. This helps the model to generalize well by removing unwanted biasing. The normalized intensity value I_n of a slice is computed in equation (1) by using the normalization equation, where I is the original intensity value of the slice, μ is the intensity value of I , and σ is the standard deviation of I , given as:

$$I_n = \frac{I - \mu}{\sigma}. \quad (1)$$

As most of the slices of every patient consisted of completely black masks or masks having very little information regarding the tumor, the model is more inclined to learn background or noise instead of the original tumor part. In order to remove the data imbalance, a threshold value [7] of at least 0.007% tumor informative pixels of the entire image is taken. That means for every patient, the image slices that are having at least 400 pixels out of 57,600 pixels regarding the tumor information are considered for training and testing. Additionally, 1% of the border from all 4 sides was trimmed automatically from every image resulting in decreasing the image slice size

from 240×240 to 192×192 . These pre-process not only removed the class imbalance problem but also reduced the training time and memory requirements by a huge margin.

Once the training and testing dataset is created for each of the modalities, it was trained through the U-Net, Encoder–Decoder model, and SegNet models to segment the different tumor regions. In the next section, all the model architecture used in this study is discussed in detail.

3.2 Model architecture

The two-layered ensemble model architecture is shown in Figure 1. This model is based on the three well-known schematic segmentation models which are used widely in medical image analysis. In the first layer, all the models are explored through the multimodal MRI scans, and the maximized feature map is extracted for each individual model. In the second layer, the maximized feature map from each of the model is automatically extracted and it is fused to produce the final segmented output for all the subparts of the tumor.

The U-Net has been one of the most efficient and well-known architectures [16,17] for medical image segmentation. The model considers the full feature map during up-sampling which makes the network larger in terms of memory. To make the model memory-efficient, the next deep model known as the SegNet encoder–decoder model is explored. This model uses pooling indices during up-sampling due to which it takes less memory but performance degrades. To balance this, we explore another deep model encoder–decoder model which is equivalent to SegNet without pooling indices and uses entire feature maps during up-sampling.

In the first layer, the three models are designed and experimented on all the multimodal scans to get the maximized feature map. In our work, the U-Net model consists of a contraction path, a bottleneck region, and an expansion path. The contraction path contains 4 encoding blocks where each one consists of 2 convolutional layers with a Rectified Linear Unit (ReLU) layer followed by a max-pooling layer having strides 2×2 and a kernel size of 2×2 . After each block of the encoding region, the number of feature channels is doubled. The bottleneck layer contains two convolutional layers which act as an intermediary between the encoding and the decoding region. The decoding region contains a transposed convolutional layer followed by a concatenation layer to concatenate with the corresponding

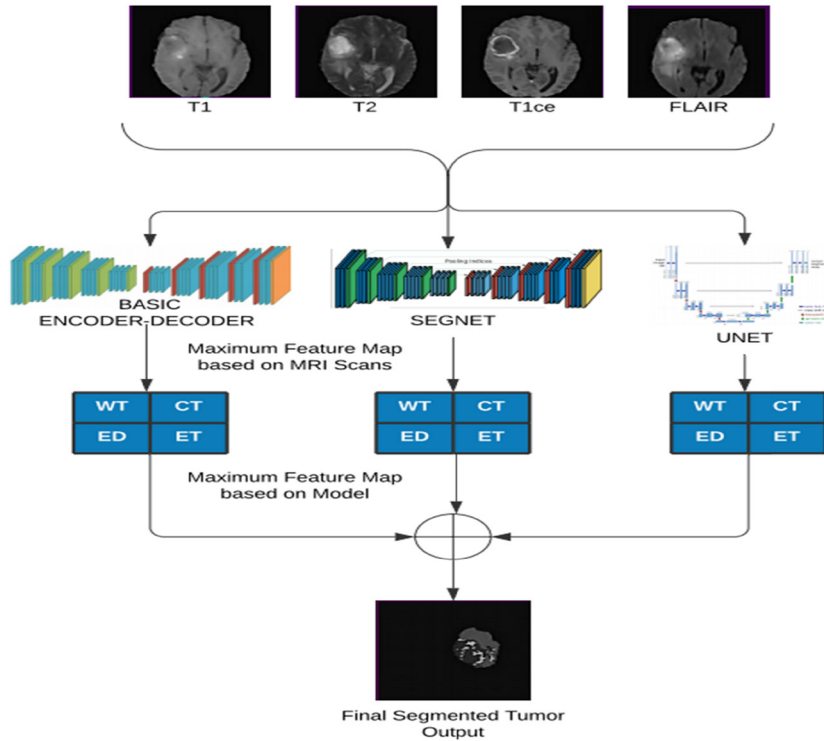


Figure 1: Proposed two-layered ensemble model for WT (Whole tumor), CT (Core tumor), ED (Edema sub part), and ET (enhancing tumor) segmentation.

cropped feature map from the encoding region, and two convolutional layers that reduce the number of feature channels to half. In the encoder–decoder model, a typical encoding-decoding architecture [26] is used. The model can be divided into two steps – the encoding steps followed by the decoding steps. The encoding region consists of five encoding blocks where the number of channels is doubled at each block. The first two encoding blocks are similar to the U-Net model, but the next three encoding blocks contain three convolutional layers followed by a max-pooling layer. The decoding region also consists of five decoding blocks where the number of channels is halved at each block. The first 3 decoding blocks contain an up-sampling layer with pool size 2×2 followed by 3 convolutional layers. The next 2 decoding blocks contain an up-sampling layer followed by 2 convolutional layers. The implemented SegNet model is a typical SegNet model [27] which is similar to the encoding-decoding model. The encoding and decoding blocks are similar to the encoder–decoder model. Instead of the max-pooling layer, a custom layer termed as Max-Pooling-WithArgMax is used which performs typical max-pooling on the feature maps with a stride of 2×2 and kernel size of 2×2 but also stores the index

position of the max-pooling pixels. Similarly, instead of the up-sampling layer, a custom layer termed Max-UpSampling-WithArgMax is used which uses the corresponding index values stored during pooling to up-sample the feature maps. The final output layer of all the models is a convolutional layer of kernel size 1×1 and the number of feature maps are equal to the number of classes and is followed by a sigmoid activation function to get a probability feature map.

Once the best feature map based on MRI scan is computed, in the second layer a feature map fusion is performed based on maximized feature extracted in each of the model. The network connectivity in the ensemble step can be represented as the following equation (2a) and (2b). $x(i)_{[CT,ET]}$ denotes the intermediate segmented output for the core tumor and enhancing tumor and $x(i)_{[WT,ED]}$ (equation (2b)) denotes the intermediate segmented output for the whole tumor and edema subpart.

$$x(i)_{[CT,ET]} = \begin{cases} M(H(x(i-1))), & i = 0-3, \\ H(x(i-1)), & i = 4, \\ H([U(x(i-1), [x(8-i)]), & i = 5-8, \end{cases} \quad (2a)$$

and

$$x(i)[WT,ED] = \begin{cases} M(H1(x(i-1))), & i = 0-1 \\ M(H2(x(i-1))), & i = 2-4, \\ H2(U(x(i-1))), & i = 5-7, \\ H1(U(x(i-1))), & i = 8-9, \end{cases} \quad (2b)$$

where $H(x)$ represents a convolutional layer operation followed by a ReLU activation function applied twice on a feature map x . $M(x)$ and $U(x)$ represent the max-pooling layer and the Conv-transpose layer, respectively. Similarly, $[]$ denotes the concatenation layer along with the skip connection. $H1(x)$ and $H2(x)$ represent the operations (2conv + Batch norm. + ReLU) and (3conv + Batch norm. + ReLU) respectively.

3.3 Activation function and loss function

The activation function is one of the most crucial learning parameters for any of the deep learning models. The activation function controls the output, accuracy, and computational efficiency of training a model. In our methodology, the ReLU and sigmoid, two activation functions are used for all the models. The ReLU function (equation (3)) is defined as:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (3)$$

The sigmoid function (equation (4)) is defined as:

$$f(x) = \frac{1}{1 + \sqrt{e^{-x}}}. \quad (4)$$

The binary cross-entropy loss (BCE) is commonly used in binary classification task (equation (5)). This loss function is computed pixel-wise which compares the actual pixel value y_i (i.e., 0 or 1) with the predicted probability of it being 1 termed as p_i . If y_i is 1, it adds $\log(p_i)$ to the loss and if y_i is 0, it adds $\log(1 - p_i)$ to the loss.

$$\text{BCELoss} = -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)), \quad (5)$$

Another loss function (equation (6)) often used is the dice loss function which measures the overlay between the actual (A) and the predicted mask (B).

$$\text{Dice Loss} = 1 - \frac{2|A \cap B|}{|A| + |B|}. \quad (6)$$

For our segmentation task, we took BCE dice loss as the loss function (equation (7)) which is the addition of Binary cross-entropy and dice loss function. This is done to get a faster convergence and better performance.

$$\begin{aligned} \text{BCE Dice Loss} \\ = & -\frac{1}{N} \sum_{i=1}^N (y_i \log(p_i) + (1 - y_i) \log(1 - p_i)) \\ & + \left(1 - \frac{2|A \cap B|}{|A| + |B|}\right). \end{aligned} \quad (7)$$

3.4 Dataset and training details

Different deep learning model performances have been examined and evaluated using the BraTS 2017 data [9]. Multimodal BraTS has always been focusing on multi-institutional pre-operative MRI scans used for segmentation of Glioma brain tumors which are intrinsically heterogeneous in structural appearance, shape, and histology. There are a total of 210 HGG and 75 LGG a patient subjects available in the BraTS 2017 dataset, whereas this study covers only 210 HGG patient datasets counted by us. Due to the low risk and high survival rate of LGG patients as compared to those with HGG [28], these patients are not included in the evaluation. In BraTS 2017, with the help of different clinical protocols and various scanners, four different multimodal MRI scans known as native (**T1**), post-contrast T1-weighted (**T1ce**), T2-weighted (**T2**), and T2 Fluid Attenuated Inversion Recovery (**FLAIR**) volumes were acquired for each of the patients from 19 multiple institutes. These four types of modalities are taken as input to each of the models separately and the different tumor part is extracted through training. For each of the patients, the ground truth (mask) is given in the dataset verified by experienced neuron-radiologists. In a manually annotated ground truth mask each tumor sub-part: enhancing tumor (ET), the peritumoral edema (ED), and the necrotic and non-enhancing tumor (NCR/NET), is labeled with label 4, label 2, and label 1, respectively, in the referred data set.

The input image slice is trimmed automatically to 192×192 after going through data pre-processing steps. As discussed in the data pre-processing phase, only those image slices are considered for the training which carries at least 0.1% of the information of the tumor in 3D volume. The whole dataset of 210 patients is split into 80% for training and 20% for testing: 170 patients' data were used for training and 40 patients' data were used for testing. The training image dataset is again divided into training and validation data in an 8:2 ratio randomly. The details of the number of image slices taken for training a model for each of the sub-tumor parts are given in Table 1.

Table 1: Details of training and testing image datasets

Tumor subparts/ datasets	Training		No. of image slices for test
	No. of image slices for training	No. of image slices for validation	
Complete	6,946	1,737	2,121
Core	3,816	955	1,393
Enhancing	2,770	693	1,198
Edema	6,124	1,531	1,945

To avoid over-fitting due to less number of images, data augmentation is applied to the whole training dataset. The augmented dataset is created through 90° rotations, horizontal flip, width and height shifting, and zooming of the images, and a shuffled set of image slices are trained through the model in each step of the epoch. To achieve effective and efficient training, the mini-batch size was set to 32 image slices. For the comparison and fine-tuning of the models, the training was carried out for a fixed number of epochs equal to 100. The training was performed using the Google Collaborator platform with GPU back end in a computer having 4GB RAM.

3.5 Evaluation metrics

Several evaluation metrics were used to evaluate the segmented tumor sub-regions with the segmented ground truth. Dice Score compares the overlays between the predicted and the ground truth. Other metrics include accuracy and F1-Score which are the gold standards for classification tasks used to get deeper insights into the predictions.

Dice Score (equation (8)) can be defined as twice the total overlapping area of the predicted mask and the ground truth divided by the sum of the total number of tumor pixels (i.e., pixel value 1 [foreground]) in both the predicted mask and the ground truth).

$$\text{Dice score} = \frac{2 \times \text{Area of overlap}}{\text{Total tumor region}}. \quad (8)$$

Other metrics used require pixel-wise evaluation of the ground truth and the predicted masks. True Positive (TP) can be defined as the total number of positive pixels (belonging to the tumor) in the ground truth which are correctly predicted as a tumor, True Negative (TN) is the total number of negative pixels (belonging to the background) in the ground truth which are correctly predicted negative, False Positive (FP) is the total number of negative pixels which are falsely predicted as positive pixels

and finally False Negative (FN) is defined as the total number of positive pixels which are falsely predicted as negative pixels. Accuracy (equation (9)) can be defined as the ratio of the number of pixels correctly classified to the total number of pixels.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (9)$$

Specificity (true negative rate) measures (equation (10)) the actual number of pixels that do not belong to the tumor class that is correctly identified.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (10)$$

Sensitivity (recall) is the ratio of the total number of correctly predicted foregrounds to the total number of actual foregrounds. It shows (equation (11)) the amount of foreground predicted correctly out of all the foreground.

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (11)$$

F1-Score (equation (12)) is defined as the harmonic mean of precision and recall and is a widely used metric for the classification of imbalanced classes.

$$F1 \text{ score} = \frac{2 \times \text{Precision} \times \text{recall}}{\text{Precision} + \text{recall}}. \quad (12)$$

4 Experimental results and discussion

In this section, we are going to discuss the experimental results, network architecture for all the models, different model parameters, training details, and segmented results. Finally, a detailed comparison of these three deep learning models with different evaluation metrics is performed.

4.1 Effect of network structure

The network structure and the number of trainable and non-trainable parameters of the model are shown in Tables 2 and 3. The Adam algorithm is an adaptive learning rate [29] method that computes individual learning rates for each parameter. In all the models, the Adam optimization [30,31] is used with a default learning rate of 0.001. The model was trained to achieve maximum validation accuracy between the predicted map and the ground truth. Our model is not trained using the mean squared error (MSE) function because it is a non-

Table 2: Detailed structure of U-Net model

Layer type	Output size	Filter size	No. of parameters (param #, total = 24,957,057, trainable params: 24,942,721 and non-trainable params: 14,336)
Input Layer	$192 \times 192 \times 1$	—	0
2 @ (convolution layer + batch normalization)	$192 \times 192 \times 64$	3×3	38,080
max_pooling	$96 \times 96 \times 64$	2×2	0
2 @ (convolution layer + batch normalization)	$96 \times 96 \times 128$	3×3	222,464
max_pooling	$48 \times 48 \times 128$	2×2	0
3 @ (convolution layer + batch normalization)	$48 \times 48 \times 256$	3×3	1,478,400
max_pooling	$24 \times 24 \times 256$	2×2	0
3 @ (convolution layer + batch normalization)	$24 \times 24 \times 512$	3×3	5,905,920
max_pooling	$12 \times 12 \times 512$	2×2	0
3 @ (convolution layer + batch normalization)	$12 \times 12 \times 512$	3×3	7,085,568
max_pooling	$6 \times 6 \times 512$	2×2	0
Up sampling layer	$6 \times 6 \times 512$	2×2	0
3 @ (convolution layer + batch normalization)	$12 \times 12 \times 512$	3×3	7,085,568
Up-sampling layer	$24 \times 24 \times 512$	2×2	0
3 @ (convolution layer + batch normalization)	$24 \times 24 \times 256$	3×3	1,832,064
Up-sampling layer	$48 \times 48 \times 256$	2×2	0
3 @ (convolution layer + batch normalization)	$48 \times 48 \times 128$	3×3	591,744
Up-sampling layer	$96 \times 96 \times 128$	2×2	0
2 @ (convolution layer + batch normalization)	$96 \times 96 \times 64$	3×3	111,232
Up-sampling layer	$192 \times 192 \times 64$	2×2	0
2 @ (convolution layer + batch normalization)	$192 \times 192 \times 64$	3×3	74,368
Output layer	$192 \times 192 \times 1$	3×3	577

convex function for binary classification and is not guaranteed to minimize the cost function during training. The hidden convolutional layer uses ReLU function whereas the output layer uses the sigmoid activation function to threshold the tumor and background.

The network parameter is equivalent to the encoder–decoder model and SegNet model as both uses the same number of encoding and decoding layers. In SegNet, a total of 24,957,057 parameters are used out of which 24,942,721 are trainable parameters and 14,336 are non-trainable parameters. In U-Net, the total number of parameters is 28,070,753, all are trainable parameters. While training through SegNet, its Parameters require 2.9 GB memory for a batch size of 32 images, while U-Net (parameters) takes 3.35 GB memory for the same batch size if each of the parameter takes 4 bytes of memory. There is a difference of 12 MB memory requirement to train

through these different models for a batch size of one input image.

At the same time, we also measured the time taken to train the entire dataset through each of the models. Each modality took 6.8 h on an average for training the model to do segmentation of four subparts: complete, edema, core, and enhancing using the U-Net model. Similarly, the SegNet model takes 16.5 h on an average. and the encoder–decoder model takes 6.7 h on an average for each of the modalities. The U-Net model stores a full encoding feature map which is used in the decoding level with skip connection. Whereas SegNet stores only the pooling indices of feature maps during max-pooling, which in turn requires less memory than U-Net. But through the experiment, it is observed that it takes higher computation time to reconstruct the feature map from pooling indices in the decoding phase.

Table 3: Detailed structure of SegNet and encoder–decoder model

Layer type	Output size	Filter size	No. of training parameters (param #, total = trainable params: 28,070,753)
Input Layer	$192 \times 192 \times 1$	—	0
2 @ convolution layer	$192 \times 192 \times 64$	3×3	37,568
max_pooling	$96 \times 96 \times 64$	2×2	0
2 @ convolution layer	$96 \times 96 \times 128$	3×3	221,440
max_pooling	$48 \times 48 \times 128$	2×2	0
2 @ convolution layer	$48 \times 48 \times 256$	3×3	885,248
max_pooling	$24 \times 24 \times 256$	2×2	0
2 @ convolution layer	$24 \times 24 \times 512$	3×3	3,539,968
max_pooling	$12 \times 12 \times 512$	2×2	0
2 @ convolution layer	$12 \times 12 \times 1024$	3×3	14,157,824
Transposed convolutional layer	$24 \times 24 \times 256$	2×2	1,048,832
Skip connection layer	$24 \times 24 \times 768$	—	0
2 @ convolution layer	$24 \times 24 \times 512$	3×3	5,899,264
Transposed convolutional layer	$48 \times 48 \times 128$	2×2	262,272
Skip connection layer	$48 \times 48 \times 384$	—	0
2 @ convolution layer	$48 \times 48 \times 256$	3×3	1,475,072
Transposed convolutional layer	$96 \times 96 \times 64$	2×2	65,600
Skip connection layer	$96 \times 96 \times 192$	—	0
2 @ convolution layer	$96 \times 96 \times 128$	3×3	368,896
Transposed convolutional layer	$192 \times 192 \times 32$	2×2	16,416
Skip connection layer	$192 \times 192 \times 96$	—	0
2 @ convolution layer	$192 \times 192 \times 64$	3×3	92,228
Output Layer	$192 \times 192 \times 1$	1×1	65

4.2 Performance analysis

The performance of the segmentation is evaluated through the most common and useful metrics [32] for image segmentation: sensitivity, specificity, F1 score, dice score, and

accuracy measurements. In the primary stage of our experiment, we learned that the SegNet model requires less memory as compared to U-Net but the computational time is very large as compared to the other two methods. A confusion matrix (having four values: true positive, false

Table 4: Evaluation metrics for each of the models

	Complete			Core			Enhancing			Edema		
	S	Sp.	Acc	S	Sp.	Acc	S	Sp.	Acc	S	Sp.	Acc
U-net												
FLAIR	0.98	0.94	0.98	0.98	0.84	0.98	0.98	0.59	0.97	0.99	0.75	0.98
T1-weight	0.98	0.83	0.97	0.99	0.76	0.98	0.99	0.56	0.97	0.99	0.65	0.97
T1ce	0.98	0.88	0.98	0.99	0.95	0.99	0.99	0.89	0.99	0.98	0.74	0.97
T2-weight	0.99	0.91	0.98	0.99	0.87	0.98	0.99	0.64	0.98	0.98	0.72	0.97
SegNet												
FLAIR	0.98	0.94	0.98	0.98	0.79	0.98	0.98	0.54	0.97	0.98	0.68	0.97
T1-weight	0.98	0.85	0.97	0.98	0.80	0.98	0.84	0.47	0.70	0.98	0.57	0.96
T1ce	0.98	0.87	0.97	0.99	0.93	0.99	0.99	0.89	0.99	0.98	0.69	0.97
T2-weight	0.99	0.91	0.98	0.99	0.86	0.98	0.99	0.61	0.98	0.98	0.68	0.97
Encoder–decoder model												
FLAIR	0.98	0.92	0.98	0.98	0.85	0.98	0.98	0.51	0.97	0.98	0.70	0.97
T1-weight	0.98	0.81	0.97	0.99	0.79	0.98	0.99	0.51	0.97	0.98	0.62	0.97
T1ce	0.99	0.87	0.98	0.99	0.85	0.98	0.99	0.56	0.97	0.99	0.72	0.99
T2-weight	0.99	0.86	0.99	0.99	0.92	0.99	0.99	0.85	0.97	0.98	0.69	0.99

Table 5: Dice score evaluation for each of the model

	Complete	Core	Enhancing	Edema
U-Net Model				
FLAIR	0.91	0.89	0.73	0.84
T1 weighted	0.86	0.84	0.72	0.76
T1ce	0.86	0.93	0.88	0.78
T2 weighted	0.89	0.89	0.73	0.80
SegNet model				
FLAIR	0.91	0.82	0.66	0.74
T1 weighted	0.86	0.86	0.60	0.72
T1ce	0.83	0.92	0.86	0.73
T2 weighted	0.89	0.85	0.66	0.78
Encoding-decoding model				
FLAIR	0.89	0.87	0.63	0.81
T1 weighted	0.84	0.87	0.66	0.74
T1ce	0.85	0.91	0.84	0.75
T2 weighted	0.85	0.85	0.67	0.83

positive, false-negative, and true negative) is also used to compute the specificity and sensitivity of the segmentation results because it allows visualization of the performance of an algorithm. During training, the confusion matrix for each of the sub-tumor segmentation is computed. Widely used segmentation metrics are sensitivity measures correctly classifying a pixel that belongs to the tumor part, and

specificity measures correctly classifying a pixel that does not belong to the tumor part. These two metrics along with the accuracy of the test image dataset is evaluated by using a confusion matrix. The detailed metrics for each of the models is shown in Table 4.

The above metrics calculate the efficiency of the algorithm regarding correctly classifying a pixel if it belongs to the tumor region. Next we have considered the error of prediction i.e., the analysis is also carried out by measuring the mean squared error of the test dataset. In statistics, the MSE of the model measures the average squared difference between the predicted and actual output. The algorithm having a low MSE value will perform better as compared to another model. MSE is measured for each of the predicted masks with respect to the original ground truth. The overall MSE for a model is computed by taking the average of all individual measurements and found that U-Net is having the lowest MSE value of 0.017026377 whereas SegNet is having the maximum MSE value of 0.021142694. The encoder and decoder model is having an average MSE value of 0.018796785, nearly equal to U-Net model.

Dice score is considered as one of the most common metrics to evaluate the performance of the segmentation algorithm. The model is evaluated here by using

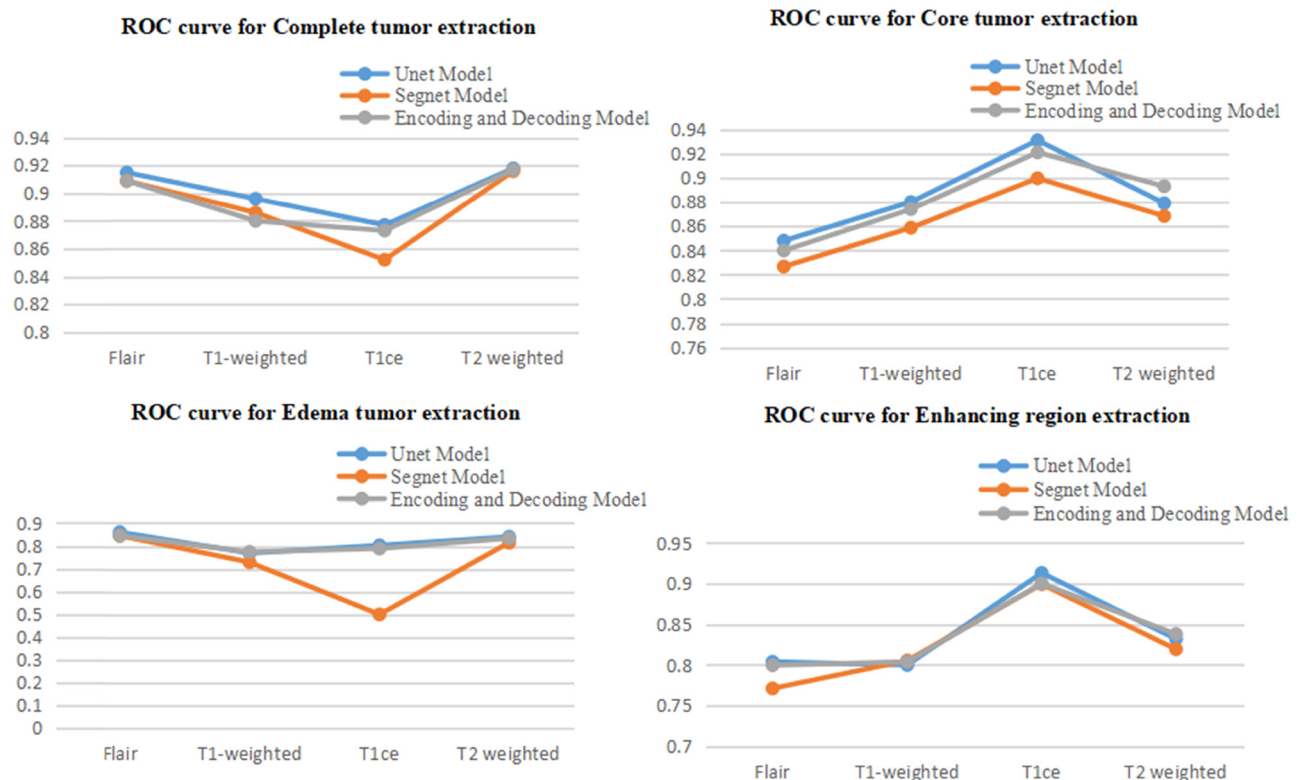


Figure 2: Area under ROC values shown for each of the model for segmenting different sub-tumor parts using FLAIR, T1 weighted, T2 weighted, and T1ce modalities.

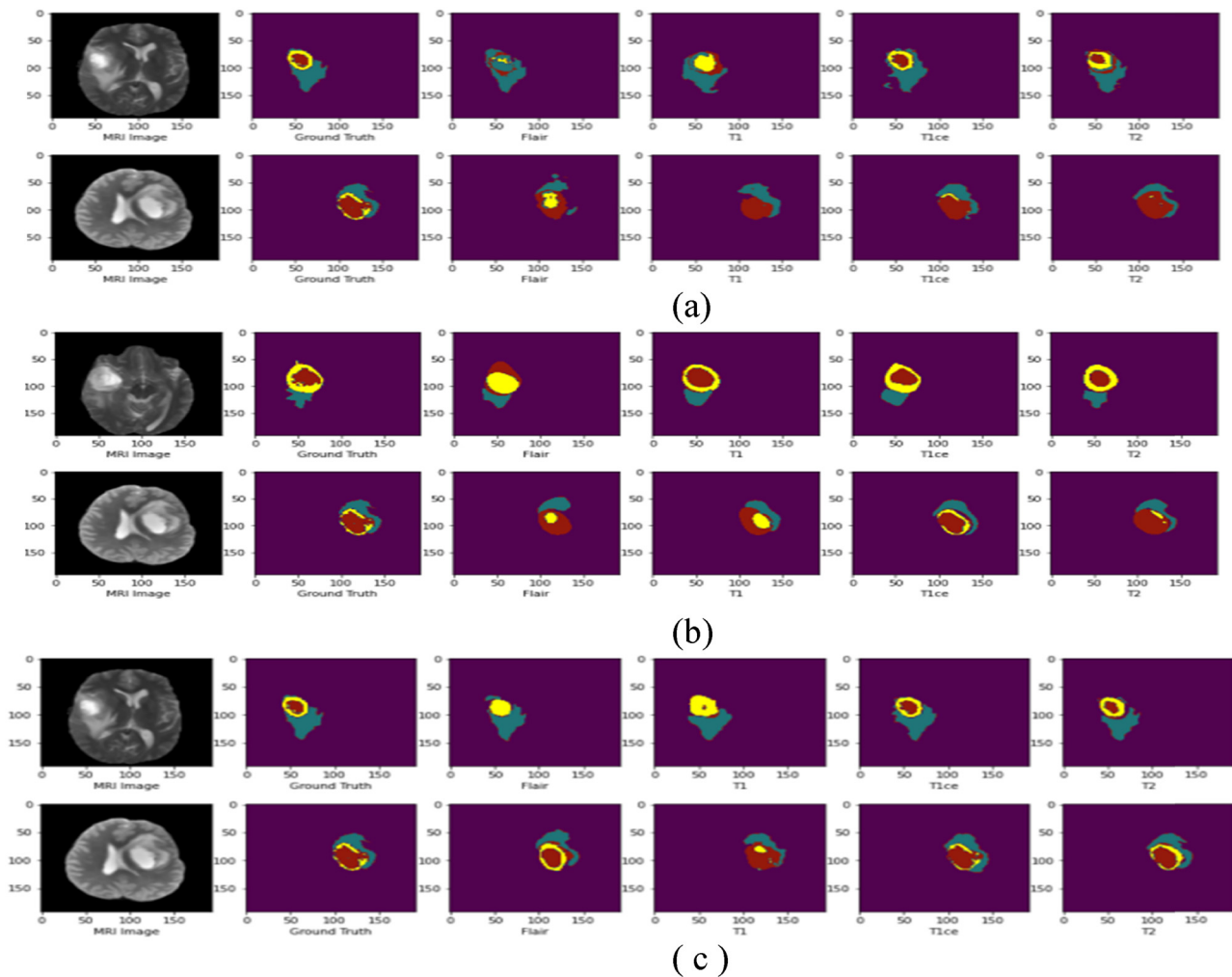


Figure 3: Segmentation results of U-Net, SegNet, encoder–decoder model in (a)–(c). From leftmost, images are: MRI scans, ground truth, tumor-Flair, tumor-T1, tumor-T1ce, and tumor-T2, respectively.

validation dice scores which are considered to be maximized during training instead of MSE. The dice score is measured for individual images and then computed as the overall dice score for a set of validation data as shown in Table 5 below. It is clearly shown that all the models are having maximum DSC for extracting complete sub-tumor using FLAIR modality, core and enhancing sub-parts using T1ce, and edema region using either FLAIR or T2 weighted modality. But U-Net is performing better as compared to the other two models as per the experimented data.

The research also demonstrates the performance in terms of the Receiver Operating Characteristic (ROC) curve taken between false positive rate vs true positive rate and measures the area under the curve (AUC value) (Figure 2) to analyze the performance of classification of the pixel in detail. The AUC can be used to compare the performance of two or more classifiers in absolute terms: the higher the AUC, the better the segmentation/classification.

The details of the AUC value for each of the sub-tumor segmentation through different modalities are shown graphically.

There is a similar trend in AUC values for each of the models during segmenting the sub-tumor region using FLAIR, T1 weighted, T2 weighted, and T1ce modalities. For complete and edema region extraction, FLAIR and T2 weighted modalities perform better as compared to T1 weighted and T1ce. Similarly, for core and enhancing, the performance is better in T1ce as compared to the other three modalities. The result also shows the model efficiency in terms of the ROC curve (AUC) for the sub-tumor part with respect to the different modalities. For overall performance of segmentation, from above all the graphs it is clearly shown that the U-Net model is better as compared to the other two models and SegNet (pooling indices) result is satisfactory in comparison to the state of the art but performance is lower than U-Net model.



Figure 4: F1 score measurement for all three models from FLAIR, T1 weighted, T2 weighted, and T1ce image.

Table 6: Statistical individual model evaluation based on p -value (significance level = 0.05)

	Complete	Core	Edema	Enhancing	p value ($\alpha = 0.05$) between the different MRI modalities
U-Net Model-MCC coefficient					
FLAIR	0.882781597	0.76195301	0.733954421	0.597514232	0.464793942
T1 weighted	0.80834005	0.755741818	0.588498631	0.574054177	
T1ce	0.811308823	0.902201399	0.665461586	0.859285427	
T2 weighted	0.869286394	0.808268511	0.695442371	0.649902579	
p value ($\alpha = 0.05$) between the different sub-tumor class)	0.019135219				
SegNet Model-MCC coefficient					
FLAIR	0.873806396	0.714491313	0.684694879	0.542430024	0.422318467
T1 weighted	0.807497053	0.756842076	0.50458108	0.338464233	
T1ce	0.777448871	0.862175229	0.641156539	0.843895954	
T2 weighted	0.866944349	0.792927035	0.648545161	0.619661342	
p value ($\alpha = 0.05$) between the different sub-tumor class)	0.031750621				
Encoding-decoding model-MCC coefficient					
FLAIR	0.867001123	0.755921778	0.693215883	0.548250198	0.326070166
T1 weighted	0.78168003	0.767959228	0.57776859	0.555870555	
T1ce	0.848352884	0.815648215	0.723083251	0.610734163	
T2 weighted	0.910581218	0.879948605	0.688275931	0.82351615	
p value ($\alpha = 0.05$) between the different sub-tumor class)	0.00730139				

Table 7: Statistical comparison and evaluation between U-Net, SegNet, and encoder–decoder model based on p -value (significance level = 0.05)

Model/sub-tumor parts	Average MCC coefficient				p value ($\alpha = 0.05$) (between the model for segmenting the tumor)
	Complete	Core	Edema	Enhancing	
U-Net model	0.631570379	0.797839851	0.653723194	0.630298253	0.830034404
SegNet model	0.831424167	0.781608913	0.619744415	0.586112888	
Encoder– decoder model	0.851903814	0.804869456	0.670585914	0.634592766	
p -value ($\alpha = 0.05$) (between the model based on different sub-tumor classes)	0.00002432				

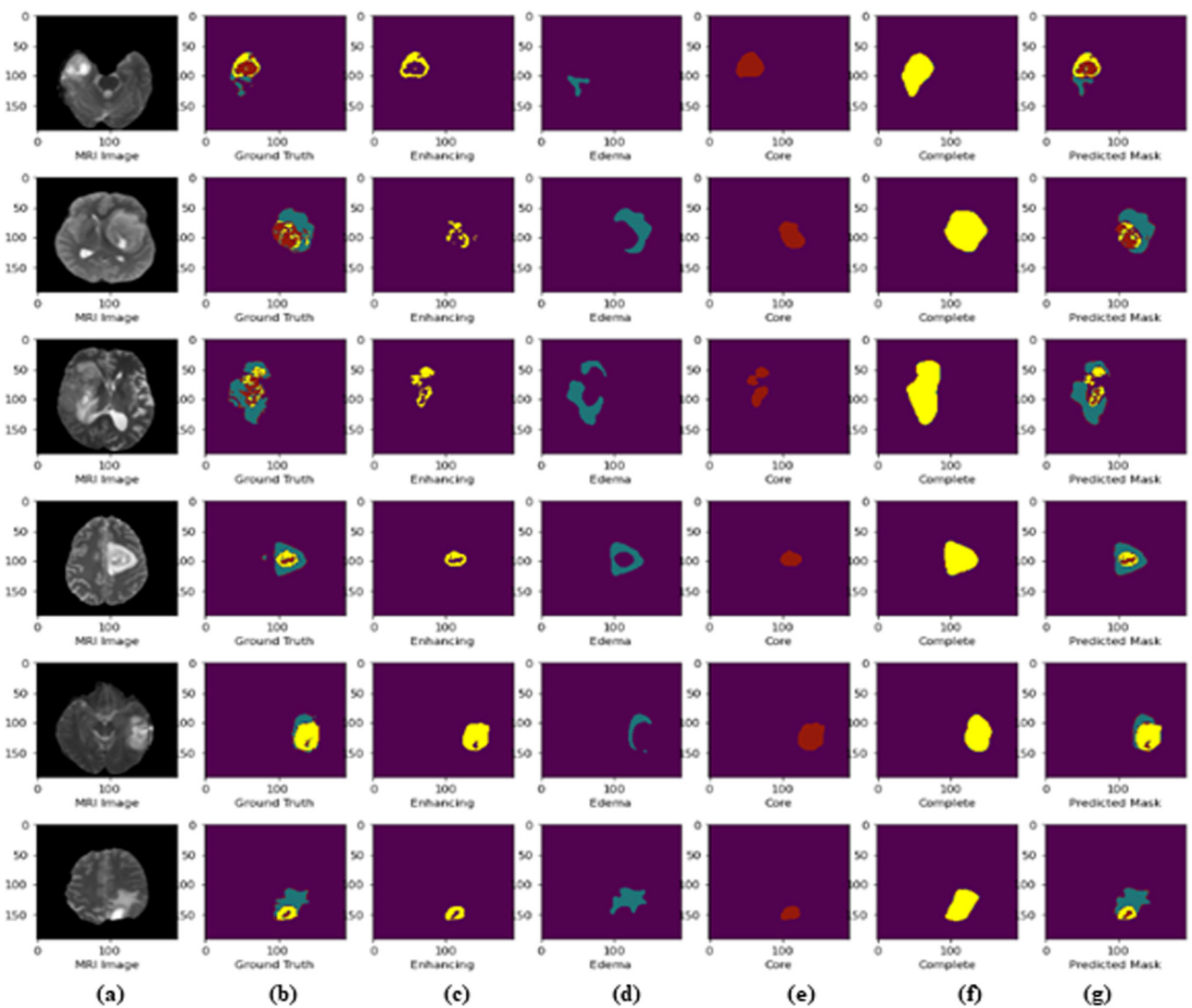


Figure 5: Segmentation results of hybrid models: (a)–(g) MRI slices, ground truth, enhancing part, edema region, core tumor, complete part, and predicted whole mask, respectively.

Table 8: Comparison of our method and other methods

Method	Metric	Dataset	Complete	Core	Enhancing	Edema
Ayşe Demirhan (2014) [33]	DSC	Own dataset	0.61	—	—	0.77
Yi et al. (2016) [12]	DSC	BraTS 2015 training (HGG + LGG)	0.89	0.76	0.80	—
Dong et al. (2017) [4]	DSC	BraTS 2015 training datasets (HGG)	0.88	0.87	0.81	—
Kamnitsas et al. (2017) [24]	DSC	BraTS 2015 training (HGG + LGG)	0.90	0.76	0.73	—
Casamitjana (2017) [23]	F1 score	BraTS 2018 training (HGG + LGG)	0.86	0.68	0.67	—
Xu et al. (2019) [17]	DSC	BraTS 2015 training (HGG + LGG)	0.73	0.62	0.42	—
Alqazzaz et al. (2019) [14]	F1 score	BraTS 2017 training (HGG + LGG)	0.85	0.81	0.79	—
Ensemble model by Mark Lyksborg (2015) [34]	DSC	BraTS 2014 training	0.81	0.697	0.681	—
Ensemble model by Xue Feng (2020)	DSC	BraTS 2018 training (HGG + LGG)	0.9114	0.8304	0.7946	—
Proposed ensemble model	DSC	BraTS 2017 training (HGG)	0.85	0.91	0.88	0.83
	F1 score		0.92	0.95	0.93	0.84

The bold values show the result of the proposed method which is significantly better as compared to existing methods.

The segmentation results are demonstrated in Figure 3 on different test images. Figure 3(a–c) depict some semantic segmentation results in visual form using the U-Net model, SegNet model, and encoder–decoder model, respectively, from an axial view. It shows the different sub-tumor parts extracted from FLAIR, T1 weighted, T2 weighted, and T1ce images along with the original ground truth images.

The U-Net model and encoder–decoder model achieve maximum F1 score of 0.94 and 0.93, respectively, for the complete tumor segmentation using FLAIR modality, whereas in SegNet model, maximum of 0.93 F1 score is achieved using T2 weighted modality. Detailed value of F1 score for each of the models is given in Figure 4 represented graphically.

All the models achieved maximum F1 score for core and enhancing tumor segmentation using T1ce modality (core: 0.95, 0.92, and 0.94 and enhancing: 0.93, 0.92, and 0.91 for U-Net, SegNet, and encoder–decoder models, respectively). For edema region segmentation, U-Net and SegNet achieves 0.87 and 0.84, respectively, using FLAIR modality but encoder–decoder model achieves 0.84 F1 score for both FLAIR and T2 weighted images.

4.3 Statistical analysis

Even if accuracy and F1 score are the most common metrics used in statistics to measure the segmentation performance, sometimes both can be misleading. Both the metrics do not fully consider the size of TP, TN, FP, and FN classes of the confusion matrix in their final computation. To avoid these misleadings and justify our performance comparison, another performance score, the Matthews correlation coefficient (MCC) [35], is considered. It is given in terms of TP, TN, FP, and FN (equation (13)) as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (13)$$

Finally, we conclude our analysis through the MCC metric which is given in Table 6. Along with computed F1 score and accuracy, MCC results also confirm the improvement in the prediction of tumors and show that our segmentation results are very satisfactory with respect to the ground truth.

The statistical analysis is performed through the ANOVA test by using the MCC coefficient value and considering the significance level (α) equal to 0.05. The null hypothesis taken as the mean of each group is equal or there is no significant difference between the performances. From the p values measured for the performance with respect to modalities for U-Net ($=0.464793$), SegNet ($=0.4223184$), and encoder–decoder model ($=0.32607016$), it is clearly shown that there is no such evidence to reject the null hypothesis as all the values are greater than 0.05. Hence, there is no significant difference in segmentation results for the different multimodal MRI scans. Similarly, when the test is performed to analyze the segmentation result for different sub-tumor classes, the p -value is computed as 0.019135 for U-Net, 0.03175 for SegNet, and 0.00730 for the encoder–decoder model. All the values are less than 0.05 which indicates there is a significant difference in performance results when segmenting the different sub-parts of the tumor. This analysis is performed between the groups of individual models. Then, this analysis is extended to measure the performance difference between the models as shown in Table 7.

From the ANOVA test, we got a p -value equal to 0.83003 for the performance of three models for segmenting tumors. As it is clearly shown from the different evaluation metrics that there is no huge gap in the segmentation result while comparing the three CNN models, the test also cooperates with the fact. It shows there is no

significant difference in segmentation results between the three models. When it comes to the individual sub-tumor parts segmentation or whole tumor segmentation, statistics show that there is a difference between the model performances as the p -value is very much less than 0.05.

In the above section, we compare the three different models through different segmentation metrics to analyze the performance. But no single model is performing better in comparison to the other two for segmenting all the sub-tumor parts. Even if there is no such single MRI modality which will give better accuracy as compared to multi-parametric MRI scans. Each model has its own architecture which will reduce either computational time or memory usage. Similarly, each modality has different characteristics and contrast images which help in segmenting different sub-tumor from different modalities with the highest accuracy.

Hence in ensemble models, the maximized feature maps extracted based on MRI scans and models are fused together to segment the whole tumor of a patient which is time and memory-efficient in some contexts. In the second layer of the proposed model, it selects the model with the corresponding modality for predicting the sub-part of the tumor. The visual output of the model is depicted in Figure 5. The result of the proposed ensemble model is compared with other existing results as shown in Table 8.

5 Conclusion

In this article, an ensemble model is designed by combining the existing FCNN model with maximized parameters. It efficiently and automatically segments the whole tumor along with its sub-tumor parts. To increase the efficiency of the model we have used the thresholding in the selection of image slices for training and validation. Our results and analysis show that our method performs better than that in refs. [14,36,37] in segmenting each of the sub-tumors. The edema region is also extracted which is not computed in most of the research articles and got a satisfactory result in comparison to state of the art.

Another contribution of this article is the detailed comparison and analysis of three well-known FCNN models in terms of their performance for brain tumor segmentation. Through the experimental result and statistical analysis, we concluded that out of three, the U-Net model performs better for segmenting the core and enhancing tumor with 0.95 and 0.93 $F1$ scores, respectively. Whereas the encoder-decoder model results better in segmenting complete

and edema tumors with 0.92 and 0.84 $F1$ scores, respectively. The performance of SegNet is satisfactory in comparison to existing methods but lower than that of the models discussed in this article. Due to the pooling indices used in SegNet, the model becomes memory efficient but takes more computational time and less accuracy as it does not store the entire feature map during down-sampling. Along with the models, we compare the different MRI modalities for extracting different sub-tumor regions. Post-contrast T1 weighted (T1ce) MRI scan outperforms the other three in segmenting the core and enhancing the tumor region, while T2 weighted gives the highest accuracy in extracting the whole tumor and edema region. The research can be further extended with performance analysis of the model by changing different learning parameters and also can be evaluated on a different dataset as the characteristics of the patient are changing day by day. The proposed model can be trained parallel to become more time-efficient.

Conflict of interest: Authors state no conflict of interest.

References

- [1] K. K. Farmanfarma, M. Mohammadian, Z. Shahabinia, S. Hassanipour, and H. Salehiniya, "Brain cancer in the world: an epidemiological review," *World Cancer Res. J.*, vol. 6, no. 5, pp. 1–5, 2019.
- [2] D. N. George, H. B. Jehloli, and A. S. Oleiwi, "Brain tumor detection using shape features and machine learning algorithms," *Int. J. Adv. Res. Computer Sci. Softw. Eng.*, vol. 5, no. 10, pp. 454–459, 2015.
- [3] D. N. Louis, A. Perry, G. Reifenberger, A. Von Deimling, D. Figarella-Branger, W. K. Cavenee, et al., "The 2016 World Health Organization classification of tumors of the central nervous system: a summary," *Acta Neuropathol.*, vol. 131, no. 6, pp. 803–820, 2016.
- [4] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in: *Annual Conference on Medical Image Understanding and Analysis*, Springer, 2017, pp. 506–517.
- [5] S. Das, S. Bose, G. K. Nayak, S. C. Satapathy, and S. Saxena, "Brain tumor segmentation and overall survival period prediction in glioblastoma multiforme using radiomic features," *Concurrency Comput.: Pract. Experience*, p. e6501.
- [6] S. Das, "Brain Tumor Segmentation from MRI Images Using Deep Learning Framework," in: *Progress in Computing, Analytics and Networking*, Springer, Singapore, 2020, pp. 105–114.
- [7] S. Das, G. Nayak, S. Saxena, S. C. Satpathy, "Effect of learning parameters on the performance of U-Net Model in segmentation of Brain tumor," *Multimed. Tools Appl.*, pp. 1–19, 2021.
- [8] S. Saxena, P. Mohapatra, and S. Pattnaik, "Brain tumor and its segmentation from brain MRI sequences," in: *Early Detection*

- of *Neurological Disorders Using Machine Learning Systems*, IGI Global, 2019, pp. 39–60.
- [9] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al., “The multimodal brain tumor image segmentation benchmark (BRATS),” *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, Oct 2015.
 - [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst.*, vol. 25, pp. 1097–1105, 2012.
 - [11] R. Girshick, J. Donahue, T. Darrell, and J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation,” in *Proceedings of the IEEE Conference On Computer Vision And Pattern Recognition*, 2014, pp. 580–587.
 - [12] D. Yi, M. Zhou, Z. Chen, and O. Gevaert, “3-D convolutional neural networks for glioblastoma segmentation,” arXiv preprint arXiv:1611.04534, 2016.
 - [13] M. El Adoui, S. A. Mahmoudi, M. A. Larhman, and M. Benjelloun, “MRI breast tumor segmentation using different encoder and decoder CNN architectures,” *Computers*, vol. 8, no. 3, p. 52, 2019.
 - [14] S. Alqazzaz, X. Sun, X. Yang, and L. Nokes, “Automated brain tumor segmentation on multi-modal MR image using SegNet,” *Comput. Vis. Media*, vol. 5, no. 2, pp. 209–219, 2019.
 - [15] G. R. Padalkar and M. B. Khambete, “Analysis of Basic-SegNet architecture with variations in training options,” in: *International Conference on Intelligent Systems Design and Applications*, Springer, Vellore, India, 2018, pp. 727–735.
 - [16] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in: *International Conference on Medical Image Computing And Computer-Assisted Intervention*, Springer, Munich, Germany, 2015, pp. 234–241.
 - [17] F. Xu, H. Ma, J. Sun, R. Wu, X. Liu, and Y. Kong, “LSTM multi-modal UNet for brain tumor segmentation,” in: *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, IEEE, Xiamen, China, 2019, pp. 236–240.
 - [18] S. Saxena, S. Paul, A. Garg, A. Saikia, and A. Datta, “Deep Learning in Computational Neuroscience,” in: *Challenges and Applications for Implementing Machine Learning in Computer Vision*, IGI Global, 2020, pp. 43–63.
 - [19] S. Das, M. K. Swain, G. Nayak, and S. Saxena, “Brain tumor segmentation from 3D MRI slices using cascading convolutional neural network,” in: *Advances in Electronics, Communication and Computing*, Springer, Bhubaneswar, India, 2021, pp. 119–126.
 - [20] F. Rajbdad, M. Aslam, S. Azmat, T. Ali, and S. Khattak, “Automated fiducial points detection using human body segmentation,” *Arab. J. Sci. Eng.*, vol. 43, no. 2, pp. 509–524, 2018.
 - [21] Q. Jin, Z. Meng, C. Sun, H. Cui, and R. Su, “RA-UNet: A hybrid deep attention-aware network to extract liver and tumor in CT scans,” *Front. Bioeng. Biotechnol.*, vol. 8, p. 1471, 2020.
 - [22] H. Kaldera, S. Gunasekara, and M. B. Dissanayake, “MRI based Glioma segmentation using Deep Learning algorithms,” in: *2019 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, Springer, Srilanka, 2019, pp. 51–56.
 - [23] A. Casamitjana, M. Catà, I. Sánchez, M. Combalia, and V. Vilaplana, “Cascaded V-Net using ROI masks for brain tumor segmentation,” in: *International MICCAI Brainlesion Workshop*, Springer, Canada, 2017, pp. 381–391.
 - [24] K. Kamnitsas, C. Ledig, V. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, et al., “Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation,” *Med. Image Anal.*, vol. 36, pp. 61–78, 2017.
 - [25] W. Cong, J. Song, K. Luan, H. Liang, L. Wang, X. Ma, et al., “A modified brain MR image segmentation and bias field estimation model based on local and global information,” *Comput. Math. Methods Med.*, vol. 2016, 1–13, 2016.
 - [26] R. Yasrab, N. Gu, and X. Zhang, “An encoder–decoder based convolution neural network (CNN) for future advanced driver assistance system (ADAS),” *Appl. Sci.*, vol. 7, no. 4, p. 312, 2017.
 - [27] V. Badrinarayanan, A. Kendall, and R. Cipolla, “SegNet: A deep convolutional encoder–decoder architecture for image segmentation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017.
 - [28] S. Banerjee, S. Mitra, F. Masulli, and S. Rovetta, “Deep radiomics for brain tumor detection and classification from multi-sequence MRI,” arXiv preprint arXiv:1903.09240, 2019.
 - [29] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.
 - [30] C. Xue, J. Zhang, J. Xing, Y. Lei, and Y. Sun, “Research on edge detection operator of a convolutional neural network,” in: *2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, IEEE, China, 2019, pp. 49–53.
 - [31] C.-L. Huang, Y. C. Shih, C. M. Lai, V. Y. Y. Chung, W. B. Zhu, W. Yeh, et al., “Optimization of a Convolutional Neural Network Using a Hybrid Algorithm,” in *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE, Hungary, 2019, pp. 1–8.
 - [32] S. Saxena, A. Garg, and P. Mohapatra, “Advanced approaches for medical image segmentation,” in: *Application of Biomedical Engineering in Neuroscience*, Springer, 2019, pp. 153–172.
 - [33] A. Demirhan, M. Törü, and I. Güler, “Segmentation of tumor and edema along with healthy tissues of brain using wavelets and neural networks,” *IEEE Journal of biomedical and health informatics*, vol. 9, no. 4, pp. 1451–1458, 2014.
 - [34] M. Lyksborg, O. Puonti, M. Agn, and R. Larsen, “An ensemble of 2D convolutional neural networks for tumor segmentation,” in: *Scandinavian conference on image analysis*, Springer, Cham, pp. 201–211, 2015.
 - [35] S. Furqan Qadri, D. Ai, G. Hu, M. Ahmad, Y. Huang, Y. Wang, et al., “Automatic deep feature learning via patch-based deep belief network for vertebrae segmentation in CT images,” *Appl. Sci.*, vol. 9, no. 1, p. 69, 2019.
 - [36] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, et al., “Ensembles of multiple models and architectures for robust brain tumour segmentation,” in: *International MICCAI Brainlesion Workshop*, IEEE, Canada, 2017, pp. 450–462.
 - [37] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in: *Proceedings of the IEEE Conference On Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.