Enhancing Brain Tumor Identification: A U-Net Approach for Segmentation

Maidul Islam[1], Farjana Alam[2], Farhan Bin Bahar[3], Mukarrama Tun Fabia[4], Md Sabbir Hossain[5], Farah Binta Haque[6], and Annajiat Alim Rasel[7]

[1] Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
maidul.islam@g.bracu.ac.bd
[2] Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
farjana.alam@g.bracu.ac.bd
[3] Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
farhan.bin.bahar@g.bracu.ac.bd
[4] Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
mukarrama.tun.fabia@g.bracu.ac.bd
[5] Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
md.sabbir.hossain1@g.bracu.ac.bd
[6] Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
farah.binta.haque@g.bracu.ac.bd
[7] Brac University, 66 Mohakhali, Dhaka 1212, Bangladesh
annajiat@gmail.com

Abstract. Brain tumors are either benign or malicious. Benign brain tumors have consistent showing in construction and do not have participating cancer cells. On the other hand, malicious types of tumors are unpredictable in their structure and contain large numbers of active ones. Benign tumors can be further differentiated in two primary variations-gliomas and meningiomas. Malignant ones are which grows rapidly while benign tumor cells copy living cells and proliferate slowly. Image segmentation and classification helps in image comprehension, feature extraction and prediction. Tissue stratification, localization of tumors, tumor magnitude etc are the applications done through it. The accurate and reproducible measurement and morphology of tumors are critical for brain tumor diagnosis and proper treatment. Our proposed method involves using the architecture of 3 Dimensional U-net to properly section brain tumors from BraTS 2023 data. The 3D U-net model is commonly known for its good segmentation because it uses the additional spatial information in the 3D data. Our proposed architecture of 3D U-net model has been applied to the BraTS 2023 MRI image dataset which achieved an accuracy and Intersection over Union score of 99.38% and 85.71% respectively.

Keywords: Brain tumor · Image segmentation · Image classification · Image classification · 3D U-net · BraTS 2023.

1 Introduction

Intracranial tumors, another name for brain tumors, are abnormal clusters of tissue in which cells grow and proliferate at an unstable rate which remain unchecked by defense mechanisms in the brain. Though there are 150 different kinds of tumors, primary and metastatic ones are the most common. Finding a brain tumor manually is very time-consuming and challenging. Because of this, an automated process is necessary to detect and separate the brain tumor. The method utilized is known as brain tumor segmentation and classification, and it is meant to address the aforementioned issue and find brain tumors early. Brain tumor segmentation is a popular way of segmenting the brain tumor using deep learning. It is a part of biomedical image processing in which the process retrieves useful data from scans, segments brain scans correctly, classifies the tumor, and accurately predicts the result. Through achieving these objectives, the task of identifying the tumor along with its type is done. 3D U-Net architecture is like U-Net architecture which mainly extends its concept to 3D data. Mainly, for segmentation in two dimensions, U-Net architecture is a popularly used method. The letter U resembles architecture which includes downsampling and upsampling. 3D U-Net uses this architecture to work with data that are three-dimensional. This 3D U-Net is beneficial to work with medical images like CT scans and MRIs. The main task of segmentation of these medical images is to identify specific regions within the data which can be an organ or also it can be a tumor. And so, 3D U-Net is advantageous because it can capture spatial dependencies in all three dimensions. For tasks that require accurate localization and segmentation in three dimensions, 3D U-Net is well suited.

2 Literature review

In this paper [1], for accurate brain tumor segmentation, a combined and hybrid approach based on machine learning classifiers and deep CNN is presented in order to reduce time,improve accuracy, and reduce human error. Three public datasets were used for the experiments. First one was from Kaggle which contained about 3174 Brain MRI images consisting of 926 glioma, 901 pituitary tumors and 937 meningioma. The second dataset chosen from Figshare contained 3064 T1-weighted contrast-enhanced brain MRI images from 233 persons which had 1426 glioma, 708 meningioma, and 930 pituitary tumor scans. The third one was a combination of splits of the two datasets. Initially, CNN is employed to extract the feature map from the illuminated regions of a brain MRI image. Then, a faster regional CNN is created for tumor area localisation, followed by RPN. Finally, deep CNN and machine learning classifiers are used in series for refining the classification and segmentation process. The procedure is iterated to shape the layout of the forthcoming deep convolutional neural network (CNN) as well as the machine learning classifiers. After the trials, the results showed the three designated tumor boundaries - glioma, meningioma and pituitary. The accuracy for the kaggle dataset was 98.3% and for the Figshare dataset it was

98%. The Dice similarity coefficient was over 97% in both the cases. In conclusion, the outcomes received from the experiments were more than satisfactory as it used image registration technique. Linear and non-linear image registration refined classifier precision by 4–5% in contrast to no image registration. In spite of being mostly positive, the process has some negatives. Computation time is long as the algorithms and running time are huge. More work will be done in order to make the model more efficient and accurate. Therefore, the model is highly recommended for detection of tumors in the brain.

In this paper [2], the author proposes a novel deep learning approach which is called Znet. It is used here for segmenting 2D brain tumors in the magnetic resonance images. In the paper, the authors demonstrate the potentials of their proposed model and demonstrate how it will assist in the medical treatment technologies in the case of brain tumor treatments. The dataset they used is based on "The Cancer Genome Atlas Low-Grade Glioma (TCGA-LGG)" dataset. The dataset basically consists of 110 patients with low-grade glioma. To expand the training dataset's size and enhance the model's performance, the authors employed data augmentation techniques. Znet is basically a convolutional neural network architecture consisting of an encoder-decoder having skip connections. This skip connection which is between the decoder and the encoder networks is for helping to preserve the spatial information. It also improves the accuracy of segmentation. According to the authors of this paper, their proposed model "Znet" outperforms all other methods in case of segmentation accuracy. For the mean dice similarity coefficient, the result shows very high values. During their model training they got a dice score of 0.96 and for their independent testing dataset, a dice score of 0.92 they got. Their other evaluational measures are also very high such as they got the F1 score of 0.81, Pixel accuracy of 0.996 and the Matthews correlation coefficient (MCC) of 0.81.

Arkapravo et al. [3] proposes an algorithm that uses CNN and deep learning methods to segment the brain tumor. The main objective of the authors is to improve patient care and fasten the medical process. The authors got the dataset for the experiment from the BraTS dataset. This dataset included 369 MRI scans. Of them, 293 are from glioblastoma and 76 are from lower-grade glioma. The authors also acknowledged that the assessment can only be done by using online tools of the BraTS challenge. Dice scores were used to measure the performance of the study by taking the segmented tumor area as P1 and the actual tumor area as T1. It was then calculated using online tools. The preprocessing of the data was done by taking all the input images in 128X128X3 dimensions. The methodology authors used to train and process the data is a 9-layer CNN method. Using ground truth segmentation, the assessment of the dataset was done. Many data augmentation techniques were used like rotation, flipping, and scaling to improve the dataset. For the experiment, 2892 images were taken, and after using RMSprop as an optimizer and softmax activation function, an accuracy of 99.74% was achieved after dropping some images to prevent overfitting. The main strength of the paper was the proposed algorithm which gained a much higher accuracy than other results obtained by other researchers. The limitations of the paper are that no other algorithm was compared with the proposed algorithm and also other details of the algorithm like time and space complexity were not analyzed properly which can limit the actual capability of the algorithm proposed.

In this paper [4], the authors have used the BraTS dataset, a popular dataset regarding Brain Tumor Segmentation. The research paper consists of three modules. First one is the module of semantic segmentation. To overcome the limitations of the transformer, they have used a shifted patch tokenization. It helps to remove noise disturbance and enables the use of small size datasets. Secondly, they have used an edge detection module. Convolutional Neural Networks or CNNs from the FLAIR and T1ce help to get a clear picture along the tumor edges from MRI scans. A regularization layer, 2x2 max pooling layer, 3x3 convolutional layer have been used . Lastly, in order to combine the deep semantic and edge detention, they have used a Multi Feature Inference Block. The accuracy of the trained model is seen through comparison, starting with SwinTrans and eventually adding SPD, ED, MFIB respectively. The complete model gives the most accurate result. The paper lacked a few things such as they could not provide BraTS 2023 data set. The convolutional neural network faces difficulties in distinguishing global dependencies. Last but not the least, the authors have achieved an accurate and reliable method by creating a fusion between deep semantics and edge detection.

Ahmet et al. [5] proposed of using U-net to separate portions of brain tumor from MRI images. They have used tumor localization and enhancement methods for this task. The datasets were collected from Brain Tumor Segmentation (BraTS) Challenge from years of 2012, 2019, and 2020. The BraTS dataset consist of MRI images of T1, T1c, T2, FLAIR modalities. The dataset from 2012 contains 3,725 and 1,908 high grade and low grade, 2019 contains 40,145 and 11,780 high grade and low grade, 2020 contains a total of 57,195 2D FLAIR images. The authors preprocessed the dataset by filtering with a 5x5 mean filter the tumorous images which helped them to get a clear and noiseless image. Additionally, to make the tumors more visible, they used tumor localization with no parameters and improve techniques . After that, the U-net architecture is applied for the segmentation task. In the U-net architecture, the encoder network derive data from the source image by using 3x3 convolutional layers, ReLU, and a 2x2 maxpool operation having 2 stride. On the other hand, the decoder networks up samples and reconstructs the segmentation map by using 2x2 up-convolution, 3x3 convolution and ReLU. The final result is passed on a 1x1 convolution layer and pixel-wise classification is achieved. Finally, this localization and enhancement technique along with U-net architecture helped them to get an accurate dice scores of 0.94, 0.85, 0.87 and 0.88 respectively for the 2012 HGG, 2012 LGG, 2019 and 2020 BraTS datasets.

MRI image is used in identifying and judging gliomas which are also known as the deadliest tumour because of its high mortality rate, and hence, proper identification and segmentation of the tumours is number one priority from the scans of MRI. To get the required result faster and more accurately, Deep CNN is used for achieving complex function mapping. A caveat to Deep CNN is that it does not make use of the fact that most scans are left-right symmetric based and also of the information and knowledge regarding prior medical diagnosis. Hence, DCSNN or Deep Convolutional Symmetric Neural Network is proposed to consider the considerations of the two main difficulties mentioned above. In this paper [6], the dataset from BRATS 2015 was used and it contained two sets- Training Set and Testing set . The four variations T1, T1c, T2, and T2 weighted with FLAIR are included for all the samples from the BRATS 2015 dataset. Training set consists of test patient datasets which is 274 in number along with the ground truths, and on the other hand, 110 samples of testing set containing both HGG and LGG grades are collected. The method is influenced by DCNN and the name of the network is called 'Baseline Network' which contains pathways of bottom up and top down. Bottom up pathways help in collecting hierarchical traits of the different types of morphological brain scans obtained and the top down is DCSNN which receives outcome of the pathway of bottom up as given input and produces finely tuned features which are grained via coarse grained features. These attributes are connected by lateral connection. Feature maps obtained by the second level divides the receiving levels and measures similarity by comparing the feature vectors. Majority of tumor regions are left-right asymmetries, which is taken as prior knowledge . The Deep Convulutional SNN creates a Left-Right Similarity Mask or LRSM in short, along with the proposed 'Baseline Network'. The traits obtained via the bottom-up pathway of our DCSNN are synthetic characteristics of the modal MRI images, which are taken as input for the DCSNN. Hence , the LRSM obtains an accurate portrayal of asymmetrical situations of four types of modality in brain images. Dice Similarity Factor metric is used in judging the results and the results obtained were an average DSC of 0.852. Adding to it, the process took only 11 seconds to divide whereas other processes take 25 seconds to 3 minutes. Furthermore, two more networks for loss gain function. As a result, the method proposed works more efficiently and accurately than other DCNN methods.

This paper [7] is based on an adversarial learning weighted training system for segmentation. By adding extra data on the patients' data, more adversarial examples are generated and as a result, the network learns about variable information linked with the segmentation outcomes. The paper is heavily based on the datasets provided by BRaTS for the multimodal brain tumor segmentation competition in 2021. The dataset consists of 1251 MRI images which consists of T1, T1c, T2,T2 FLAIR. From these weighted sequences, 4 separate tumor areas can be found which can be further divided into Enhancing Tumor (ET) , Non Enhancing Tumor (NET), Necrotic Tumor(NCR) and Peritumoral Edema (ED). With the addition of ED, the whole tumor weight can be gained. The procedure is based on three networks - segmentation network, critic network and virtual adversarial training network. Pathways are included in the segmentation network which consists of down sampling and up sampling and hence, makes it having the visuals of a U-loop network architecture. The critic network is designated a convolution adversarial network. Markovian patchGan architecture is implemented in the critic network which creates confidence scores for produced masks and these masks are produced and bears a similarity to ground truth masks. Lastly, the VAT training system produces predictions such that wrong predictions can be avoided by segmentation networks based on new data provided by the patients. VAT improves the accuracy of the model by adding warnings generated against the virtual adversarial direction. In conclusion, a smart and efficient way to improve 3D U-Net training is implemented through adversarial learning. The segment network is made robust and accurate via adversarial warnings and limitations and generating adversarial examples by adapting the min-max approach and by tweaking the GAN architecture is achieved. Dice scores achieved for ET was 97.25 and for total tumor weight was 99.1492%. As a result, it can be seen that the VAT is crucial for increased performance in the brain tumor segmentation tasks.

The main objective of this [8] paper is to use the best features of different architecture of CNNs and merge them into one hybrid CNN which is expected to give higher accuracy. Different architectures are SegNets, U-Nets, Resnet18. Seg-UNet, U-SegNet, and Res-SegNet are the hybridization of those architectures and these are described in this paper. Seg-Unet is a mixture of the advantageous features of SegNet5 and U-Net. SegNet contains an Encoder and a Decoder. It also has an extra convolutional layer in each block. Hybridization of it is done by adding a skip connection in the convolutional block. This skip connection is taken from U-Net. Res-SegNet is another hybridization and it is the combination of ResNet18 and SegNet5. Here a skip connection is added from Res-SegNet to Seg-Net5 which works to recapture the information by element-wise addition. U-SegNet is another hybrid architecture derived from U-Net and Seg-Net3. It is also made hybrid by adding a skip connection from U-Net. It segments the MRI into 4 classes. The dataset used in this was the BraTS dataset and it was used for training and testing. The dataset mainly contains the MRI images and ground truths which are the outputs. These are verified by specialized doctors. It contains modalities of 4 types. These are T1, T1ce , T2 and Flair. The result was obtained by considering five parameters. It was seen that hybrid architectures expectedly performed better than other popular architectures. U-Segnet performed better than U-net and Seg-Net3. Similarly, all other hybrid architectures performed better than the separate architecture from which they were merged. From all the models, it was seen that Seg-Unet performed the best as it gives better-segmented output. Though hybrid architecture takes more time for training, overall it gives better results and thus it is more accurate than other architectures of deep learning.

The paper [9] focuses on achieving a flexible brain tumor segmentation system. It proposes a pre-processing method, Then the Cascade Convolutional Method is proposed and then a Distance Wise Attention mechanism is also proposed in this paper. Brats 2018 Dataset is used as the dataset. Four types of modalities were used. They are T1, Flair, T1c and T2. These four datasets were used because they have a characteristic to detect some parts of the brain tumor. The dataset contains 75 cases with LGG and 210 cases with HGG. Also, this dataset contains patient's clinical data with different variables. The general pre-processing method is not followed here. Rather than using the whole image, a small part of the image is used for preprocessing. Because of this, a very deep convolutional model is not used, and true negative results decrease a lot. CNN architecture is widely used to segment the brain tumor. However, the fuzzy segmentation outcome is one of its major lackings. So, in this paper, the Cascade CNN model has been proposed. This model mainly combines local and global modalities of different MRIs. Also, the solution to the overfitting problem is given by proposing a distance-wise attention mechanism. Distance-wise dependency on each slice of four modalities is explored by this distance-wise attention mechanism. The parameters considered for the result were EC, TC, and WT. DSC mainly calculates the difference between ground truth and predicted truth. The outcomes were acquired through the utilization of Dice Similarity and Sensitivity metrics. It is seen that adding the attention mechanism and preprocessing method increased the accuracy by a lot. By adding those two, the Dice Score went from 0.2531 to 0.8756 for End, 0.2796 to 0.8550 for Whole, and 0.2143 to 0.8715 for Core. By using the methods proposed almost all the criteria improved. But the sensitivity value at the core didn't change at all.

This paper [10] is made with a joint contribution of two optimizations. Fuzzy algorithm helps to create a structure of each layer of the brain. Brainstorm optimization mainly focuses at the center and provides a better outcome. The primary objective of this paper is to provide maximum accuracy as well as identification of the source of the disease. MRI dataset is used to retrieve the image from its source. The quality of the image is upgraded and classified using labels. The information from the image is extracted by texture analysis. The BRATS 2018 dataset is used to conduct the analysis. Modalities such as T1, T2w, Flair are used. The dataset contains MRI scans of a number of patients with sequence settings to identify the tumor. Labeling is done using several colors depending on the condition of the tumor. The BSO algorithm creates a number of clusters and identifies the best solution from the clusters. This identification of the best resembles human decisions which ultimately named this algorithm Brain-storm optimization. The following optimization continues creating new solutions from two clusters using Gaussian random. FCM is a pattern recognition method which divides the clusters and works on the center of every cluster. More precise information is achieved using feature extraction. The brain is composed of several complex parts including GM, WM. GLCM is used to extract important features such as texture, color from these parts. Higher entropy values indicate clearer and more informative images. The accuracy of the results are determined by F1 score, sensitivity, and precision. The paper shows that the FBSO technique provides the highest number of accuracy and F1 score than other techniques like GSO,WSO. The combination of fuzzy and brain storm optimization methods helps address challenges and enhances the accuracy and efficiency of medical image analysis, thereby leading to improved patient care and better outcomes in diagnosing and treating brain tumors.

This paper [11] presents an innovative and architecture based way of fully automated segmentation using DNN. The dataset of BRATS 2013 is used. The dataset cannot provide a three dimensional image. A two dimensional way is taken with the modalities T1, T1C,T2 and FLAIR. The preprocessing technique is similar to the Brats challenge champion. The process removes the maximum minimum strength and uses N41TK and modalities like T1. The main objective of this paper is to use deep neural networks to attain maximum accuracy. A number of architectural approaches are taken for identification and segmentation of brain tumors. CNN lacks to include the local dependencies. To overcome the problem, DNN is used as a cascade architecture. The CNN is a multiple layered hierarchical feature which is connected and forms a stack. A two phased training procedure is introduced to enhance the ability of the model. This local and global path leads to an output prioritizing the subject and details respectively. The experiment is categorized into different levels of tumors. The region of tumors are identified and divided and experimented. The experiment includes precision ,sensitivity, specificity. The segmentation outputs multiple cascade architectures and their dice calculation. The result shows that the two phase CNN provides the most accuracy during the first stage of training. During the second stage, Input Cascade CNN is the fastest among other cascades and compared with other methods, it provides efficiency and accuracy among previous years methodologies. The updated version of Brats 2015 provides an even more accurate result. The two stage cascade CNN training enhances the output multiple times regarding segmentation of brain tumors in the medical sector.

3 Dataset description

The dataset from BraTS 2023 is used. BraTS dataset is widely used in the field of medical image analysis for the segmentation of brain tumors in MRI scans. MRI images usually don't have a singular origin point similar to a geometric coordinate system. MRI images are mainly composed of a lot of individual voxels which are also known as 3D pixels and each of them represents a specific point in the 3D space within the image. The objective of this dataset is to provide enough information to provide efficient and accurate results. MRI sequences capture specific information of the tissue of the body. Brain consists of complex parts like Gray Matter, White Matter, CSF. Modalities like T1, T1 Contrast, T2, FLAIR are used to give distinct information. In T1 and T2, CSF appears dark and other soft tissues have variations of gray. T1 contrast enhances the visibility of tumors. In FLAIR, CSF appears as dark which helps to visualize abnormalities easily. 1251 training data has been used to train the model, containint a shape of 240X240X155 which basically means there are 155 slices of 240X240 images.

4 Preprocessing

In the BraTS 2023 dataset, we have 1251 images in each class which are T1, T1c, T2, FLAIR and Masks or Ground truth. Since the T1 image does not provide meaningful information that could contribute to segmenting the brain tumors, we excluded the T1 images for the training part and combined the T1c, T2 and FLAIR images into one volume so that we can perform the training more effectively. Firstly, All the images were in the shape of 240X240X155. Before combining the images into one volume, the 3D image data was converted into a 2D array where the last dimension was preserved and each row of that 2D array worked like a 3D pixel or voxel. So, The images which are created by voxels have three channels which are T1 contrasted, T2 and FLAIR. These three channels are altogether creating volume of the image. While a pixel can basically represent a single or 2D grid, a voxel can represent the same element in a 3D grid. Voxels play an important role in the analysis and also in the representation of 3D MRI image data as a voxel is considered as a fundamental element in the 3D grid which is used to represent the volumetric data. After the conversion, the shape of each image became 57600X155. This operation was performed so that the normalizations can be applied to each pixel independently. Secondly, normalizations were applied to the images so that the pixel values of the images range from 0 to 1 to minimize the computational complexities. Finally, the images were converted to their original shape after normalization which is 240X240X155.

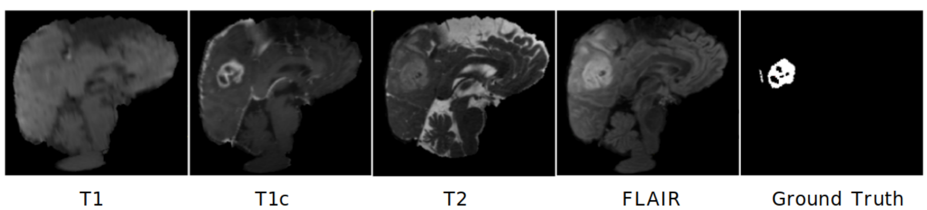


Fig. 1: Different volumes of MRI brain images from BraTS 2023 data

This normalization procedure was implemented in all the image classes except the ground truths or masks. Additionally, after performing the normalization, all the images except T1 and Mask were combined into one volume to make the segmentation task more robust. After combining the images into one volume, the input shape became 240X240X155X3 where the last 3 indicates the T1c, T2 and FLAIR images. Moreover, in the MRI images, there were a number of empty pixels surrounded by the brain images, which was unnecessary for the segmentation task. That is why, the combined images were cropped and shaped into 128X128X128X3 to reduce the computation power and focus on the parts that were necessary for the segmentation operation. Additionally, some of the image masks were completely empty and carried no contribution to the segmentation tasks. The annotated masks contain 4 labels such as 1, 2, 3 and 0 which means NCR, ED, ET and everything else respectively. To make the computation more robust, the masks containing 1,2 and 3 labels were considered as useful labels and if the total volume labels are at least 1% of the whole mask, the images were selected for the training process. Other images were ignored because they carried no value in the segmentation process. Finally, after performing these preprocessing, 1151 useful images among the 1251 images were selected for the model training process.

5 Methodology

The main objective of this research is to perform segmentation tasks on the BraTS 2023 dataset. The dataset is a combination of multimodal images in 3D nifty formats. Since it is a multimodal dataset containing T1, T1c, T2 and FLAIR images, our approach is to slice the 3D images into multiple 2D images and then combine some modalities into one image for performing the segmentation task. For the segmentation of BraTS 2023 dataset, T1 provides minimal information in terms of brain tumor segmentation. This is why, combining only the T1c, T2 and FLAIR images will help to have less computation power and achieve high accuracy. Many types of convolutional neural network models can be implemented for performing the segmentation tasks but among the state-of-the-art methods, the U-net architecture is one. The traditional U-net architecture can be modified by changing its layers, number of filters, adding skip connections, introducing residual connections, implementing different ReLU activation functions and many more. Moreover, many hybrid models combining different architectures can also be implemented like Res-SegNet which is a combination of SegNet and ResNet18 for the segmentation task.

After the preprocessing part, 1151 useful images among the 1251 images were selected for the segmentation task and we splitted the training and validation images into 75:25 ratio which is basically selecting 863 images for the training data, and 288 images as validation data. The ratio selection has been done based on our dataset. To perform well with a good accuracy we need a good amount of data for our model to learn from. So, we tried to keep the training data set as large as possible ensuring no biases due to the smaller test data set. A larger training data set provides more data for our model to learn which will result in better performance with more accuracy.

While training neural networks, a smaller learning rate makes the training process stable but slow. In contrast, an larger learning rate accelerates the training procedure but introduces instability, as the model exhibits a behaviour to overshoot the most favorable solutions, causing the training to diverge. The learning rate was decided based on the method which the researchers and the practitioners follow. The model was trained for a few epochs by gradually increasing or decreasing the learning rate and the loss, accuracy as well as the IoU score was monitored. Finally the learning rate 0.0001 is selected which minimizes the loss and maximizes the IoU and accuracy.

In our proposed 3D-unet architecture, a kernel initializer named “He uniform” was used for the training task. A unet model is basically the combination of input layer, contraction path, expansive path and an output layer. Initially, an input layer is defined for the image shape of 128X128X128X3X4 where the 128

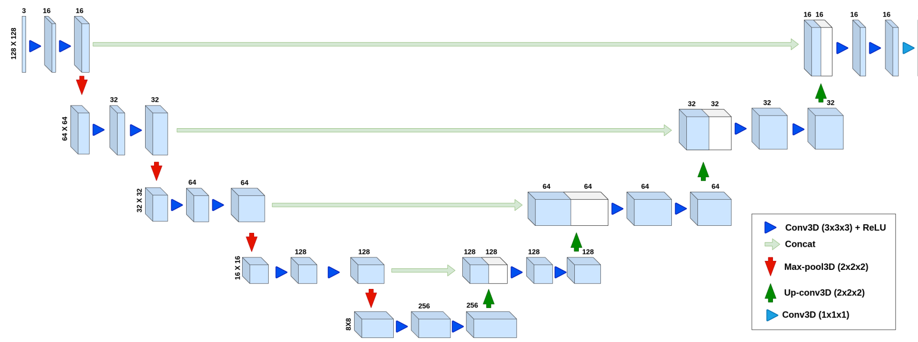
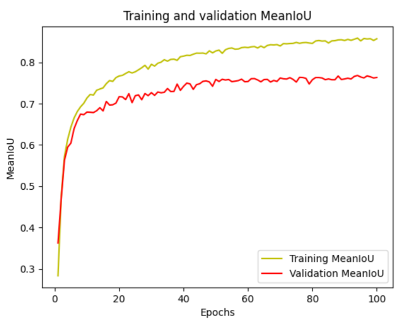


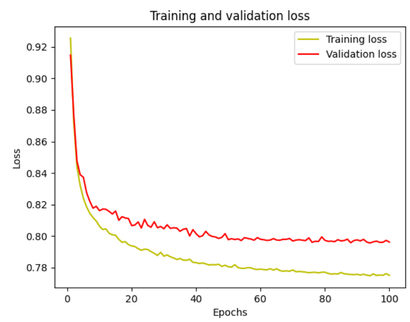
Fig. 2: Our proposed 3D-unet architecture

means image height, width and depth of that image. 3 and 4 correspond to the image channels and the image classes (T1c, T2, FLAIR and Masks), which are basically the dimension of the image volumes. In the contraction path of the 3D-Unet architecture, the first layer consists of one 3D convolution layer. In the first convolution layer, the architecture was started using 16 filters which basically means the number of patterns the layer will try to learn. In this layer a kernel size of 3X3X3 was used to perform the convolution operation in that layer, which will extract the necessary information like edges, textures, shapes and many more from the input data. The activation function employed was ReLU, which stands for Rectified Linear Unit. In terms of padding, the input and output dimensions were set to the same by making the padding necessary when needed. Following each convolutional layer, a dropout layer was incorporated to safeguard the model against overfitting. Again, another 3D convolution layer same as before is added to the model. After that, the convolutional layer's output was directed to a MaxPooling 3D layer with a 2X2X2 pooling window, intended to reduce the size of the feature maps in relation to their spatial dimensions. After the first convolution and max pooling layer, the second layer followed the same pattern but in this case, the convolution operation consisted of 32 filters, MaxPooling3D remains the same. This similar operation was performed 3 more times, each time increasing the 3D convolution layer by the multiplication of 2 upto 256 filters. This is how the contraction path of the 3D-unet architecture was developed. In the expansive path, the low resolution feature maps we got from the contraction path are upsampled and merged with high-resolution feature maps. Firstly, a 3D Transpose Convolution layer is being applied to the model having a filter of 128 and with 2X2X2 kernel and a stride of 2 is applied. The output is then concatenated with the output of the contraction path. After that, one 3D convolution layer is applied with a filter size of 128 and kernel size of 3X3X3 followed by a dropout layer for handling overfitting. Again, another 3D convolution of the same size is applied to the model. This scenario repeats 3 more times having a 3D Transpose Convolution layer, two 3D convolution layers along with a dropout layer and finally the expansive path is completed. In the expansive path, the

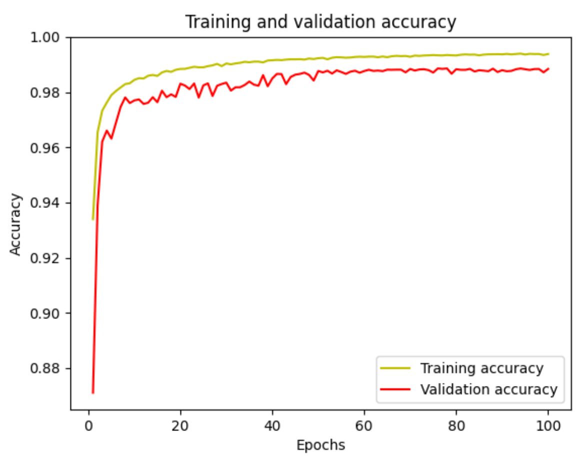
upsampling layers are completed.
Additionally, after the expansive path of a 3D u-net architecture, a final 3D convolution layer is applied to the mode with 1X1X1 kernel along with the number of classes, which is 4. The final activation function used is softmax because of the multiclass semantic segmentation task. Finally, the preprocessed MRI image data is feed to the model along with 100 epochs and a batch size of 1 because of limited hardwares, which took a good amount of training time but ended up with minimizing loss and maximizing the MeanIoU as well as the accuracy.



(a) MeanIoU plot.



(b) Loss plot.



(c) Accuracy plot.

Fig. 3: Training Progress Visualization of the proposed 3D-unet architecture

The fig. 3 represents training, validation and meanIoU curves during the training process. During the 100 epochs, the training and validation loss gradually decreased. On the other hand, duing these 100 epochs, the training and validation accuracy as well as the meanIoU increased, resulting to a stable training and accuracy of the 3D-unet model.

6 Results

The results we achieved after training our model was quite satisfactory. The 3D-unet architecture was trained with 100 epochs and a batch size of only 1 because of hardware limitations. The training process took around 2 days in a machine consisting of 16 GB RAM and 4 GB GPU. After training process, the IoU score we achieved was 85.71%. On the other hand, after completing the epochs the loss reduced to 77.51% and the accuracy gradually improved with the epochs which ended in 99.38%

The fig. 4 represents the testing image, testing label and the prediction on the test image. It is clearly visible from the image that the testing label and the prediction on test image are nearly similar, representing a stable result in the prediction of MRI brain tumor images with the 3D-unet architecture.

References

1. Haq, E.U., Jianjun, H., Huarong, X., Li, K., Weng, L.: A hybrid approach based on deep cnn and machine learning classifiers for the tumor segmentation and classification in brain mri. Computational and Mathematical Methods in Medicine 2022 (2022)
2. Ottom, M.A., Rahman, H.A., Dinov, I.D.: Znet: deep learning approach for 2d mri brain tumor segmentation. IEEE Journal of Translational Engineering in Health and Medicine 10, 1–8 (2022)
3. Chattopadhyay, A., Maitra, M.: Mri-based brain tumour image detection using cnn based deep learning method. Neuroscience informatics 2(4), 100060 (2022)
4. Zhu, Z., He, X., Qi, G., Li, Y., Cong, B., Liu, Y.: Brain tumor segmentation based on the fusion of deep semantics and edge information in multimodal mri. Information Fusion 91, 376–387 (2023)
5. Ilhan, A., Sekeroglu, B., Abiyev, R.: Brain tumor segmentation in mri images using nonparametric localization and enhancement methods with u-net. International journal of computer assisted radiology and surgery 17(3), 589–600 (2022)
6. Chen, H., Qin, Z., Ding, Y., Tian, L., Qin, Z.: Brain tumor segmentation with deep convolutional symmetric neural network. Neurocomputing 392, 305–313 (2020)
7. Peiris, H., Chen, Z., Egan, G., Harandi, M.: Reciprocal adversarial learning for brain tumor segmentation: a solution to brats challenge 2021 segmentation task. In: International MICCAI Brainlesion Workshop. pp. 171–181. Springer (2021)
8. Daimary, D., Bora, M.B., Amitab, K., Kandar, D.: Brain tumor segmentation from mri images using hybrid convolutional neural networks. Procedia Computer Science 167, 2419–2428 (2020)
9. Ranjbarzadeh, R., Bagherian Kasgari, A., Jafarzadeh Ghoushchi, S., Anari, S., Naseri, M., Bendechache, M.: Brain tumor segmentation based on deep learning and an attention mechanism using mri multi-modalities brain images. Scientific Reports 11(1), 10930 (2021)

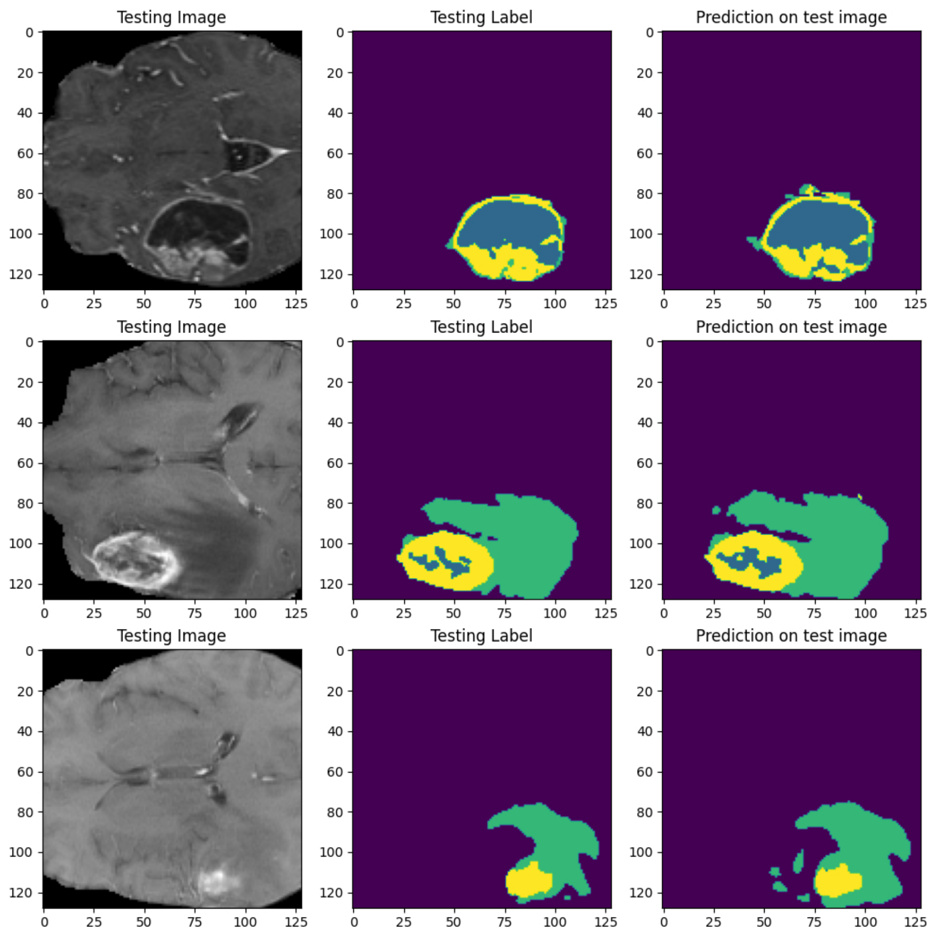


Fig. 4: Comparison of a 3D MRI image at it's corresponding testing and prediction labels

10. Narmatha, C., Eljack, S.M., Tuka, A.A.R.M., Manimurugan, S., Mustafa, M.: A hybrid fuzzy brain-storm optimization algorithm for the classification of brain tumor mri images. Journal of ambient intelligence and humanized computing pp. 1–9 (2020)
11. Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P.M., Larochelle, H.: Brain tumor segmentation with deep neural networks. Medical image analysis 35, 18–31 (2017)