Automated Multimodal Brain Tumor Segmentation using Deep Convolutional Neural Networks

Course Project: Foundations of Machine Learning (CS 725)

Sunaina Saxena
Department of EE
IIT Bombay
Mumbai, India
213070001@iitb.ac.in

Nihar Mahesh Gupte
Department of EE
IIT Bombay
Mumbai, India
213070002@iitb.ac.in

Harsh Diwakar
Department of EE
IIT Bombay
Mumbai, India
213070018@iitb.ac.in

Mohit Kumar Meena
Department of EE
IIT Bombay
Mumbai, India
213070021@iitb.ac.in

***Abstract*—Glioma brain tumors are the most common primary brain malignancies, with different degrees of aggressiveness and has variable prognosis. Currently, MRI segmentation of gliomas is largely based on examination of biological tissues in order to observe the appearance of infected cells in microscopic details. Manual tumor segmentation is a tedious, time-intensive task that requires a human expert to delineate components,Therefore manual segmentation often fraught with inaccurate readings and results. So in this project, A fully automated deep learning methods namely, 2D UNet and 3D UNet were implemented to segment out brain tumors into their sub-components. Experimentation of the model were carried out on BraTS 2018 challenge dataset and the results on 2D UNet are found to be more impressive whereas 3D UNet needs more training to perform better which couldn't be done with available dataset. The models are promising to generate good results when applied to independent clincial dataset of both LGG (low-grade glioma) and HGG (High-grade Glioma) patients.**

I. Introduction

Glioma brain tumors are caused due to mass or growth of abnormal cells in the brain, they begin in glial cells that surround nerve cells and help them function. If the growth rate is higher then can turn to be life threatening, Moreover from an study[1], about 33% of overall brain tumors are gliomas. Glial cells are the support cells in the brain that help to keep neurons in place and functioning well. Gliomas form when these glial cells mutate and grow out of control. They are classified as astrocytoma, or, oligodendroglioma, or ependymoma. At times, the tumor can be a combination of all of these too. Medical image segmentation for detection of brain tumor from the magnetic resonance images or from other medical imaging modalities is a very important process for deciding right therapy at the right time because the earlier the detection, the faster the treatment can be started. Studies based on its detection, classification and cure are at an onset at this instant. In this fast phased world, a computerized system for brain tumor detection and classification is a priori to save time and proceed into the next series of medications according to the achieved result. MRI images are preferred in our computerized system since it can accurately comprehend different tissues. Based on aggressiveness, they can be categorized into two basic grades: low-grade gliomas (LGG) that tend to exhibit benign tendencies and indicate a better prognosis for the patient, and high grade gliomas (HGG) that are malignant and more aggressive. With the development of medical imaging, brain tumors can be imaged by various Magnetic Resonance (MR) sequences, such as T1-weighted, contrast enhanced T1-weighted (T1c), T2-weighted and Fluid Attenuation Inversion Recovery (FLAIR) images. Different sequences can provide complementary information to analyze different subregions of gliomas. For example, T2 and FLAIR highlight the tumor with peritumoral edema, designated whole tumor". T1 and T1c highlight the tumor without peritumoral edema, designated tumor core" as per [2]. An enhancing region of the tumor core with hyper-intensity can also be observed in T1c, designated enhancing tumor core" as per [2]. this segmentation task is challenging because we cannot apply image augmentation to the multidimensional sequence since this could lead to inaccurate segmentation mask also the size, shape, and localization of brain tumors have considerable variations among patients. This limits the usability and usefulness of prior information about shape and location that are widely used for robust segmentation of many other anatomical structures the boundaries between adjacent structures are often ambiguous due to the smooth intensity gradients, partial volume effects and bias field artifacts.

The sections in this report are organised as follows, literature review is done in section II then the foremost and most important task of data preprocessing is discussed in section III, methodologies employed are briefly justified in section IV followed by results and conclusions in section V, lastly the discussions are documented in section VI.

II. Related works

Polly, P., Shil K, et.al. in their paper [3] proposed computerized system to differentiate between normal brain and abnormal brain with tumor in MRI images and also further classified the abnormal brain tumors into HGG or LGG tumors, they used K-means for the segmentation technique for clustering, they had used Discrete Wavelet Transform (DWT)

and Principal Component analysis (PCA) for feature extraction and feature reduction respectively.

Guotai W, Wenqi Li, et.al. in their paper [4] proposed cascaded fully convolutional neural networks to segment multimodal Magnetic Resonance (MR) images with brain tumor into background and three hierarchical regions: whole tumor, tumor core and enhancing tumor core. They designed a cascaded network to decompose the multi-class segmentation problem into a sequence of three binary segmentation problems according to the subregion hierarchy. They segmented the enhancing core based on the bounding box of the tumor core segmentation result. Their networks consist of multiple layers of anisotropic and dilated convolution filters, and combined with multiview fusion to reduce false positives.

III. Data Preparation

Dataset used in our project is BRATS 2018 [2]. It contains brain MRI of high grade glioma (HGG) patients as well as low grade glioma (LGG) patients, a total amounting to 210 HGG patients and 75 LGG patients. Each MRI is multimodal, consisting of 4 channels, T1, T2, Flare and T1 CE (Contrast Enhanced) as shown in figure 2 and it's corresponding segmentation mask is shown in figure 3. They are stored in nii.gz format,the dataset was handled with the help of SimpleITK module of python. A single MRI has 4 channels, each channel consists of 155 slices and each individual slice is of the dimensionality 240x240. For the 2D UNet, we created three batches, each with 70 MRIs, each MRI is first loaded in the form of a numpy array, and concatenated and rearranged to form a 5 dimensional data (N, 155, 240, 240, 4), N represents the number of patients in each batch. One batch is used for training, and other two for testing. Furthermore, each slice contains redundant information (background) and the volume has few slices which do not have any segmentation. So we have cropped each slice of the volume, as well as reduced few slices of the volume. Finally, for each batch, we get the image in the format (N, 95, 192, 192, 4). The ground truth consists of segmented volume with the labels : 0 for background, 1 for edema, 2 for enhancing tumor, and 3 for non enhancing tumor. At the time of training, one hot encoding was used to split a single segmented volume into 4 channels. Thus, the 2D UNet was trained in a fashion to take in one 4 channel 2 dimensional slice and produce 4 channel 2 dimensional output, each representing segmentation of one of the four channels. As for the 3D UNet, the approach towards dealing with the data completely changes, since now we tend to exploit the correlation between each slice of a mode of MRI as well. For this, after loading the MRI in the form of 5 dimensional numpy array, for a given 4 channel (155, 240, 240, 4) volume, we extract out a (16,160,160,4) volume such that the extracted volume has atmost 95% background and not more, i.e. atleast 5% of the segmented ground truth. This is carried out by an iterative algorithm explained as follows :

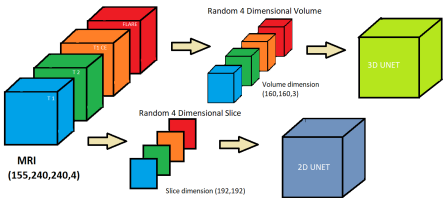

Figure 1: Data preparation

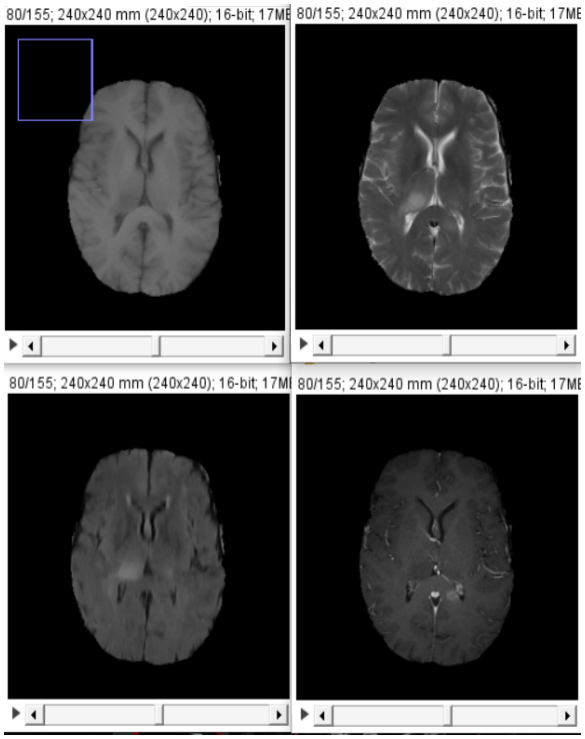

Figure 2: Images with one multimodal sequence. plot 1-4 represents T1, T2, FLAIR and T1 (ce) respectively (Tool used: ImageJ).

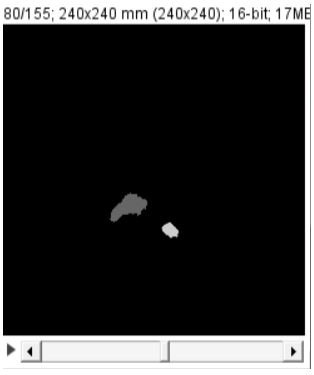

Figure 3: Segemtation mask corresponding to the modalities shown in Figure 2(Tool used: ImageJ).

1) Random (x,y,z) corresponding to 160x160x16x4 volume selected from one hot encoded volume.
2) Check for number of pixels labelled 0 (Background)

3) Check if these pixels account at most 95% of the total pixels.
4) If <95%, the volume 160x160x16x4 is desired one. Go to step 6.
5) If >95% and current iteration < max iterations, vary the centre coordinates (x,y,z) of the sub volume to get a new one hot encoded volume. Go to step 2. If current iterations > max iterations, go to step 6.
6) Use the same (x,y,z) to get the subvolumes from all the 4 channels. Repeat for each MRI data sample.

IV. Methodology

We have worked on two models which are popular for segmentation task because of their complex architecture design, specifically 2D UNet and 3D UNet, which are discussed in this section.

A. 2D UNet model

U-net was originally invented and first used for biomedical image segmentation because of speciality of performing image localization by predicting image pixel by pixel. Its architecture can be broadly thought of as an encoder network followed by a decoder network. Unlike classification where the end result of the the deep network is the only important thing, semantic segmentation and regression not only requires discrimination at pixel level but also a mechanism to project the discriminative features learnt at different stages of the encoder onto the pixel space.

- Batch normalization is also implemented in order to diminsh the reliance of gradients on the scale of the parameters also to make model less delicate to hyper-parameter tuning.
- The encoder is the first half in the architecture diagram, It is usually pre-trained but we had trained this model on our dataset and then applied convolution blocks followed by a maxpool downsampling to encode the input image into feature representations at multiple different levels.
- The decoder is the second half of the architecture. The goal is to semantically project the discriminative features (lower resolution) learnt by the encoder onto the pixel space (higher resolution) to get a dense regression. The decoder consists of upsampling and concatenation followed by regular convolution operations.
- Instead of Tanh, logistic, arctan or Sigmoid as activation function it uses ReLU function which reduce likelihood of vanishing gradient problem.
- It trains faster than other deeper architectures.

UNet architecture as shown in figure 4 comprises of two 3 × 3 convolutions, followed by Rectified Linear Unit (ReLu) and 2×2 maximum pooling operations with the stride of 2 for down sampling path. In Up sampling path, 2 × 2 transposed convolution operation taken place for reducing the feature channels. Convolution path Skip connections also introduced in the UNet architecture [5]. This connection is used to skip the features from the contracting path to the expanding path in order to recover the spatial feature lost during down sampling operations as shown in figure 4. So,the regression is very fast and accurate when compared with other regression methods.Specifically, we would like to upsample it to meet the same size with the corresponding concatenation blocks from the left.You may see the gray and green arrows, where we concatenate two feature maps together. The main contribution of U-Net in this sense is that while upsampling in the network we are also concatenating the higher resolution feature maps from the encoder network with the upsampled features in order to better learn representations with following convolutions [5]. Upsampling is a sparse operation we need a good prior from earlier stages to better represent the localization. The parameters used for the U-Net model Figure 5 is represented in Table I.

Table I: Parameter description of the model

Description of parameter for U-Net	
Functions used	**Description**
Activation function (Input)	ReLU
Activation function (Output)	Softmax
Optimizer	Adam
Loss function	Dice coefficient

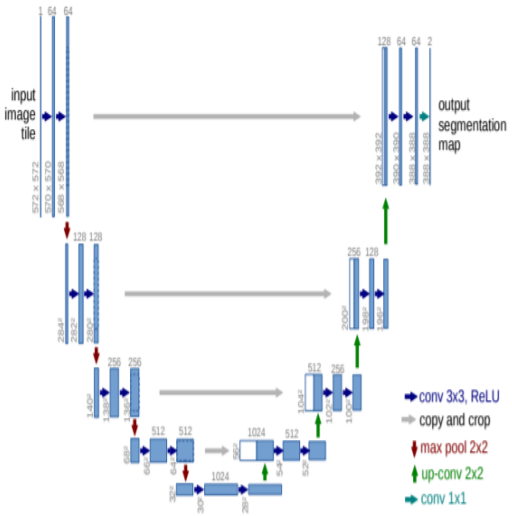

Figure 4: Architecture of UNet model (ideal)[5]

The 3D UNet [6] model essentially makes use of the fact that there exists correlation among the slices of a volume, and thus rather than operating with 2D slices, it essentially operates with 3D volumes by the means of 3D Convolution. 3D U-Net architecture is quite similar to 2D U-Net architecture except for the fact that it takes 3D volumes as input and processes them with corresponding 3D operations, in particular, 3D convolutions, 3D max pooling, and 3D up-convolutional layers. The architecture of basic 3D Unet is shown in figure 7. Implemented 3D U-Net

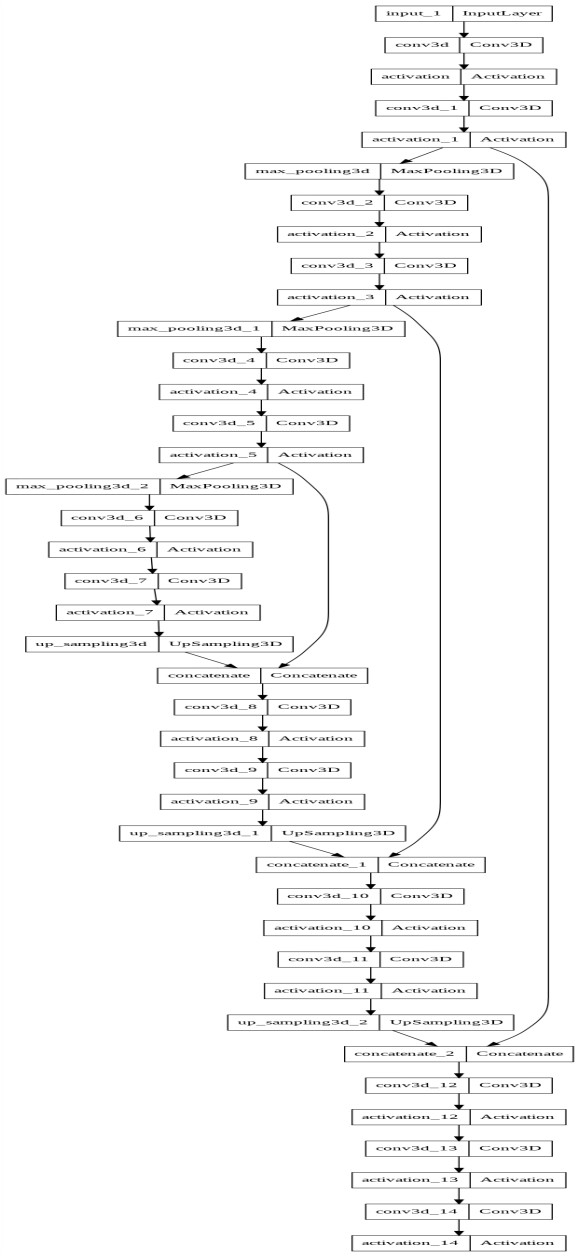

Figure 5: Architecture of UNet model (implemented)

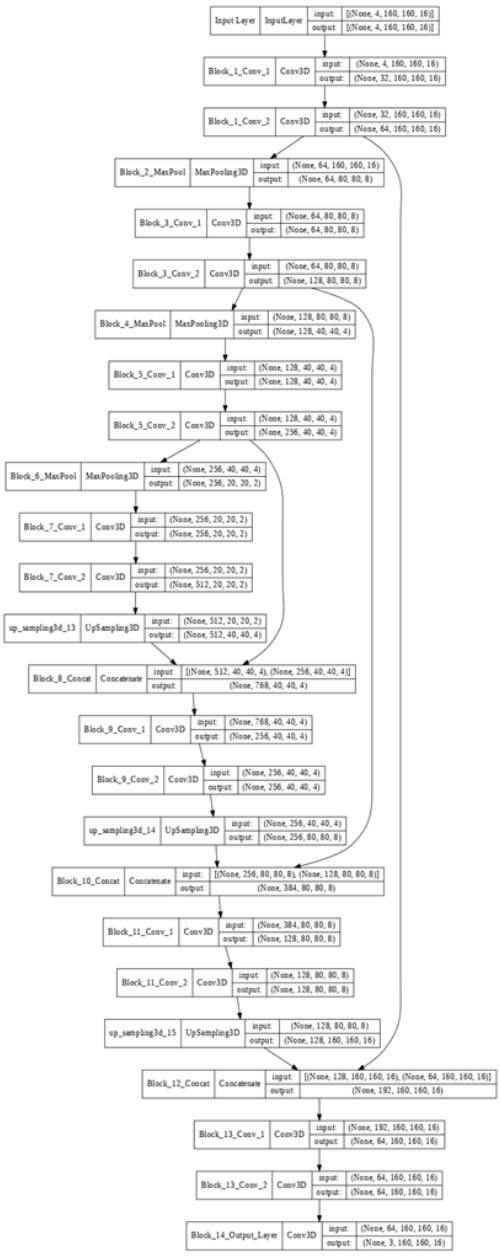

Figure 6: Architecture of 3D U-Net (implemented)

architecture shown in figure 6.

The encoder path of the architecture consists of several blocks.The first block consists of 2 convolution layers with number of filters being 32 and 64 respectively, kernel size being (3,3,3). The second block is essentially max pooling block with a stride of 2 and pool size of (2,2,2) to reduce the size from (None,64,160,160,16) to (None,64,80,80,8). The pattern of block 1 and block 2 is repeated, i.e. block 3 and 5 are similar to block 1 but with filter numbers 128,256 and 256,512 respectively, and block 4 and 6 is same as block 2. The output of block 6 has the shape (None,256,20,20,2). The next block 7 consists of 2 convolution layers with kernel

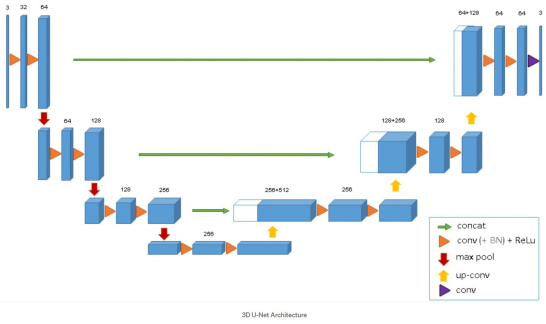

Figure 7: Architecture of 3D U-Net, image source : [6]

size (3,3,3) and filter numbers being 256 and 512, to obtain an output with shape (None,512,20,20,2). Now, this output is upsampled to get an output of shape (None,512,40,40,4), which is concatenated with the output of block 5 to get output with shape (None, 768,40,40,4). This connection is called skip connection. Block 9 has 2 convolution layers again of size (3,3,3) with 256 filters each. Output is upsampled, and concatenated by block 10 with block 3, and the process is again repeated, block 11 and 13 have 2 convolution layers each with filter numbers 128,128 and 64,64 respectively. Block 10 and 12 concatenate the output of block 9 with block 3 and block 11 with block 1 respectively, with the necessary upsampling in between. Block 13 has two convolution layers with 64,64 and 64,64 filters each, followed by block 14, the final block which has 3 filters only to obtain the 3 segmented labels. The final output obtained is thus of the shape (None,3,160,160,16), 3 being the number of segmented channels in the output.

V. Experiment and Results

The model comparison between 2D UNet and 3D UNet is shown below.

Comparison between 2D UNet and 3D UNet		
Model	Train/Test	Dice Coefficient
2D UNet	Train	0.99
2D UNet	Test	0.96
3D UNet	Train	0.59
3D UNet	Test	0.40

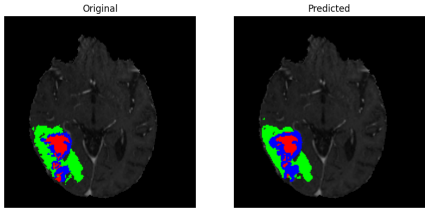

Figure 8: 2D UNet result 1

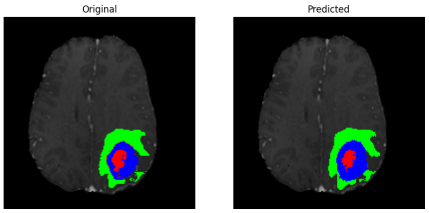

Figure 9: 2D UNet result 2

For testing, the MRI was selected from the HGG test batch. The segmentation has three labels. Enhancing tumor (red), non enhancing tumor (blue) and edema (green). The results were also obtained for LGG test sample for segmentation. 3D UNet was not able to segment the enhanced tumor region clearly. 2D UNet did not produce good results with LGG data.

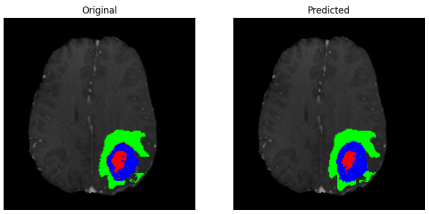

Figure 10: 2D UNet result 3

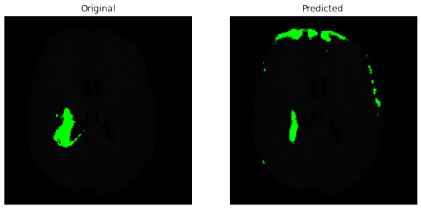

Figure 11: 2D UNet result for a slice of LGG MRI

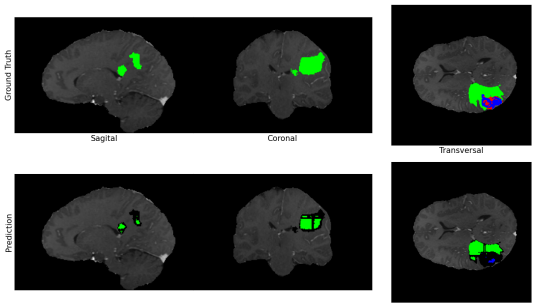

Figure 12: 3D UNet result 1

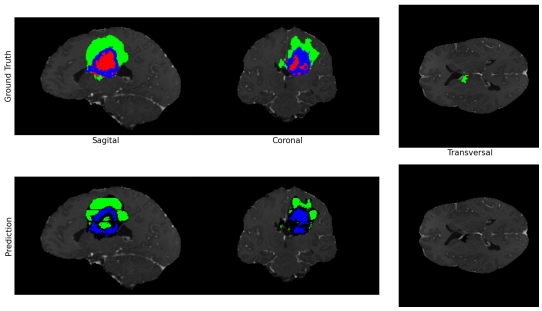

Figure 13: 3D UNet result 2

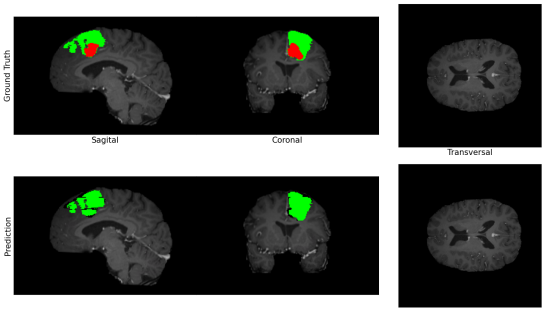

Figure 14: 3D UNet result for LGG

VI. Discussion and Conclusion

The results clearly indicate that 2D UNet has outperformed 3D UNet. A major reason for this, can be less amount of data for 3D UNet, where each volume is treated as single training sample, while in 2D UNet, each slice is treated as a single training sample, so as a result, 96 training samples were obtained from one MRI itself. One more observation that supports this point is that, the enhancing tumor (red segment) which is noticeable in 2D UNet, is not at all segmented in 3D UNet because the proportion of ground truth of enhancing tumor was very less compared to the non enhancing tumor (blue segment) and edema (green segment), thus lack of data was the undoing for 3D UNet. On the contradictory, the 2D UNet has a good segmentation output. The edema is segmented neatly for 3D UNet even when applied for LGG MRI, even though the model wasn't trained on LGG data.

Acknowledgement

We would like to thank Prof. Preethi Jyothi, Dept. of Computer Science and Engineering, IIT Bombay for her keen efforts during teaching that helped throughout the project and also helped to avoid and correct several mistakes while implementing the project.

Contributions

This is a team project and each member has putted nearly equal efforts. In detailed manner, Nihar Mahesh Gupte had worked on data preparation and architecture design of 3D UNet, Sunaina Saxena had worked on architecture design of 2D UNet and reviewed papers regarding the work done on this topic in past, Mohit Kumar Meena had worked on data preparation and implementation and evaluation of 2D UNET model, lastly, Harsh Diwakar had implemented 3D model and created data pipeline that helped in the execution of both the models.

References

[1] John hopkins, "Brain tumors," https://www.hopkinsmedicine.org/health/conditions-and-diseases/gliomas/, accessed 26-11-2021.

[2] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE transactions on medical imaging*, vol. 34, no. 10, pp. 1993–2024, 2014.

[3] F. Polly, S. Shil, M. A. Hossain, A. Ayman, and Y. M. Jang, "Detection and classification of hgg and lgg brain tumor using machine learning," in *2018 International Conference on Information Networking (ICOIN)*. IEEE, 2018, pp. 813–817.

[4] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *International MICCAI brainlesion workshop*. Springer, 2017, pp. 178–190.

[5] 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, 2015, pp. 234–241.

[6] Towards Data S, "3d unet," https://towardsdatascience.com/review-3d-u-net-volumetric-segmentation-medical-image-segmentation-\8b592560fac1, accessed 26-11-2021.