Performance Study of Optimizers for Segmentation of Brain Tumors using Atrous Convolution in U-Net

Pranshu Jena
Department of ECE
National Institute of Technology, Rourkela
Odisha, India
pranshujena16@gmail.com

Umesh C. Pati
Department of ECE
National Institute of Technology, Rourkela
Odisha, India
ucpati@nitrkl.ac.in

***Abstract*—The purpose of segmentation of the tumourous region in the brain is to identify the tumorous brain tissues and their location. Manually segmenting the brain tumor is not only time-consuming and error-prone but also increases the rate of mortality. In the time of need for quick and precise segmentation, many convolution neural networks have shown exceptionally good execution. In medical image segmentation, the most favored network that has been used is U-Net. In this work, reinstatement of the 2D convolution layer with a 2D dilated convolution layer or atrous convolution layer with a dilation rate as 2 has been implemented. The performance has been analyzed during the training and testing of the model. Different optimizers like Stochastic Gradient Descent (SGD), SGD along with momentum, Adam, and Nadam (Nesterov Adam) have been used in the U-Net model. MICCAI BraTS 2020 dataset has been used in this work. From the simulated results, it has been observed that the Nadam optimizer along with the atrous convolution layer provides higher accuracy.**

***Keywords*—Brain tumor, Deep learning, Convolution layer, Atrous layer, Optimizer**

I. Introduction

A brain tumor is a bizarre growth of cells. There are two categories of tumors that affect the brain i.e. benign and malignant. Malignant tumors are those which grow aggressively and are mainly cancerous having a high chance to spread to other parts of the body. Malignant tumors are mainly of 3 types Glioma, Meningioma, and Pituitary [1].

Detection of the tumourous regions in the brain is a strenuous task because of the structure and location in the parts of the brain. Glioma is a type of brain tumor that spreads to other parts of the brain and detection of this kind has been a challenging task [1]. In accord with, the data compiled by Global Cancer Statistics in 2020, brain tumour is one of the most lethal cancers and the 21^{st} most prevalent form of cancer worldwide. [2]. For capturing the brain tumor, medical field has been using Magnetic Resonance Imaging (MRI) widely. There 4 types of modalities T1-weighted(T1), T2-weighted(T2), T1 contrast-enhanced (T1ce), and T2-weighted Fluid Attenuated Inversion Recovery(FLAIR), which have been obtained by configurating of MRI scanner. MRI scans can be observed in 3 different planes namely Saggital, Coronal, and Axial. [3]

Precise and error-free segmentation of tumorous regions in the brain from MRI has been a bit obscure task. In many tumor types like Glioma, the appearance of the tumor is hard to identify, as the border is unclear. This problem leads to the spread of the tumor to various parts of the brain making it strenuous to identify the affected tissue from healthy tissue. Nowadays, Encoder-Decoder based models have been playing a key role in the efficient medical image segmentation of brain tumor regions [2]. Deep-learning model U-Net is one among them [4]. In the automated computer-aided methods, there is no perfect technique for the segmentation of tumourous tissue in the brain. Hence, continuously numerous expansions in the field of automatic segmentation have been introduced. The major downside in all the pre-requisite methods is that the segmentation process had been carried out manually or semi-automatic methods resulting in inaccurate segmentation of the ROI, and erroneous localization of the area of the tumor. That would result not only in a decrease in accuracy but also would be a risk in the diagnosis of the brain tumor resulting in a fatality.

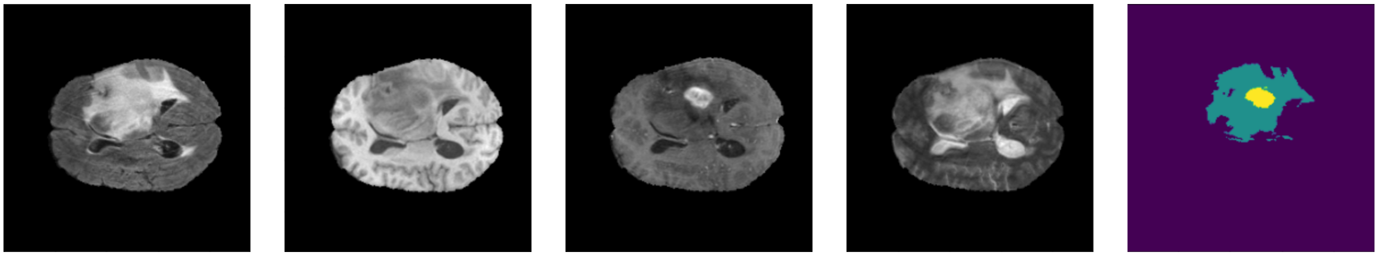
In this paper, the layers of U-net have been increased for more feature maps. The convolution layers has been replaced by atrous convolution layers [5] with a dilation rate of 2. The atrous convolution layer was employed to enlarge the receptive field. The performance of the proposed model has been analyzed by different optimizers.

The rest of the paper is organized as follows. In Section II, the related work is presented. The dataset, network architecture, and evaluation metrics are shown in Section III. Results and Discussion are given in Section IV. In Section V, the Conclusion and future scope are presented.

II. Methodology

A. Dataset

Brats 2020 has been used in this study that had been released in MICCAI BraTS challenge 2020. The dataset has multimodal brain tumor pre-operative MRI scans available in NIfTI format which has 4 sub-categories i.e. T1, T1ce, T2, and FLAIR. It provides additional information which helps in the segmentation of the tumoral region in the brain. The datas have been collected using a wide variety of clinical procedures and scanners coming from 19 different institutions. The segmented tumor has sub-tumoral regions namely necrosis, edema, and enhancing types, where each MRI modality has 155 slices per



(a) (b) (c) (d) (e)

Fig. 1. Visual representation slices of BraTS 2020 dataset (a). FLAIR (b). Native(T1) (c). T1 contrast-enhanced (d). T2 weighted (T2) (e). Segmentation Mask

volume. The preprocessing of dataset has been done by 2D slicing it to 155 slices of size 240×240. The first and last 50 slices are eliminated as they have negligible information. The 2D-sliced MRIs are visually represented in Fig. 1 [6]–[10].

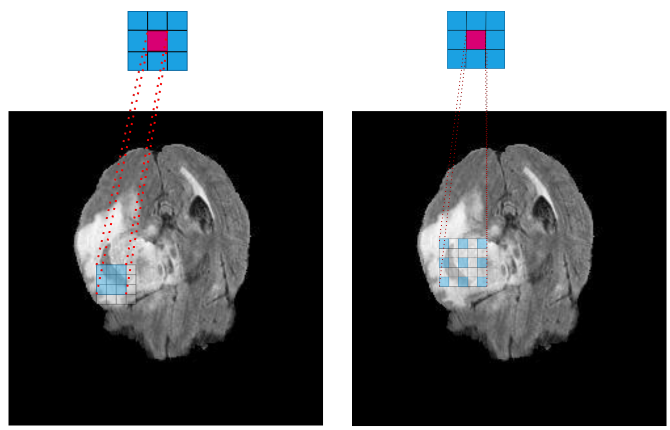
B. Network Architecture

Generally, a CNN network classifies using a label. Images are used as an input and the obtained output is a label to identify if the image is tumorous or non-tumorous. There is a need to spot the area of the tumorous region in the desired output. The obtained output should take localization into account to know the location of the tumor. Ronneberger et. al [4] solved this issue by designing an artificial neural network U-Net which segments and localizes the tumorous region. The network differentiates between the sub-units of the tumor i.e. necrotic, edema, and enhancing.

U-Net is a simple symmetric network that contains 3 parts, the left half is contraction, the right half is expansion and the bottom part is the bottleneck. The convolution layer of 3×3 and max-pooling of 2×2 are present in each layer of the contracting half. In the expanding half the pooling layers are reinstated by upsampling operators, resulting in a boost in the spatial resolution of the feature maps in order to localize the features. Many feature channels in the upsampling part enable the network to provide context data for the next higher-resolution layers. The expansion half is almost symmetrical to the contracting half, giving a visual representation of U- shape. The concatenation had been carried out by skip connections between the contracting and expanding half of those which have the same dimensional properties [4]. The upsampling or deconvolution is carried out along with skip connections so that the original dimension of the image is not lost and to retain the features in downsampling. This holds on to the segmented region and localizes the area of interest. The network only utilizes the valid portion of each convolution and does not employ any Fully Convolution (FC) layers. U-Net is extremely useful for the segmentation of tumorous regions in the brain. More precisely, during the process of segmentation, the feature map comprises the pixels in the input image for which the whole context is available.

The structure of U-Net has been shown in Fig. 3. The feature maps are represented by green boxes. The top of the boxes has been labeled as the number of channels. The light blue arrows denote the concatenated or skip connections that concatenate between the encoder and decoder half. The features map size is denoted at the bottom of the boxes.

The Atrous Convolution layer is a variant of the convolution layer. The main boon of atrous convolution over the convolution layer is that it increases the field of reception without adding any additional parameter. The use of addition of the dilation rate results in the reduction of computation and gives a larger Field of View (FoV) for the feature map. This change of the convolution layer to the atrous convolution layer is useful in feature extraction. It doesn't degrade the resolution of spatial features that aid in better segmentation of required area [5]. Atrous convolution layer introduces holes or zero values by altering the convolutional filter. This relates hybrid layer of convolution and pooling layer. The atrous convolution layer has been represented in Fig. 2.



(a) (b)

Fig. 2. Visual representation of feature extraction in (a).Convolution layer (b). Atrous Convolution layer

C. Proposed Atrous Convoluton U-net

The input images are resized to 128×128 and is fed to the modified network. The convolution layer of U-Net has been replaced by atrous convolution or the dilated convolution layer with a dilation rate as 2. Atrous convolution layer provides a larger FOV. It comparatively extracts larger area for extraction of features. The mask of the tumourous region is the ground truth for the segmentation. The spatial feature is extracted by

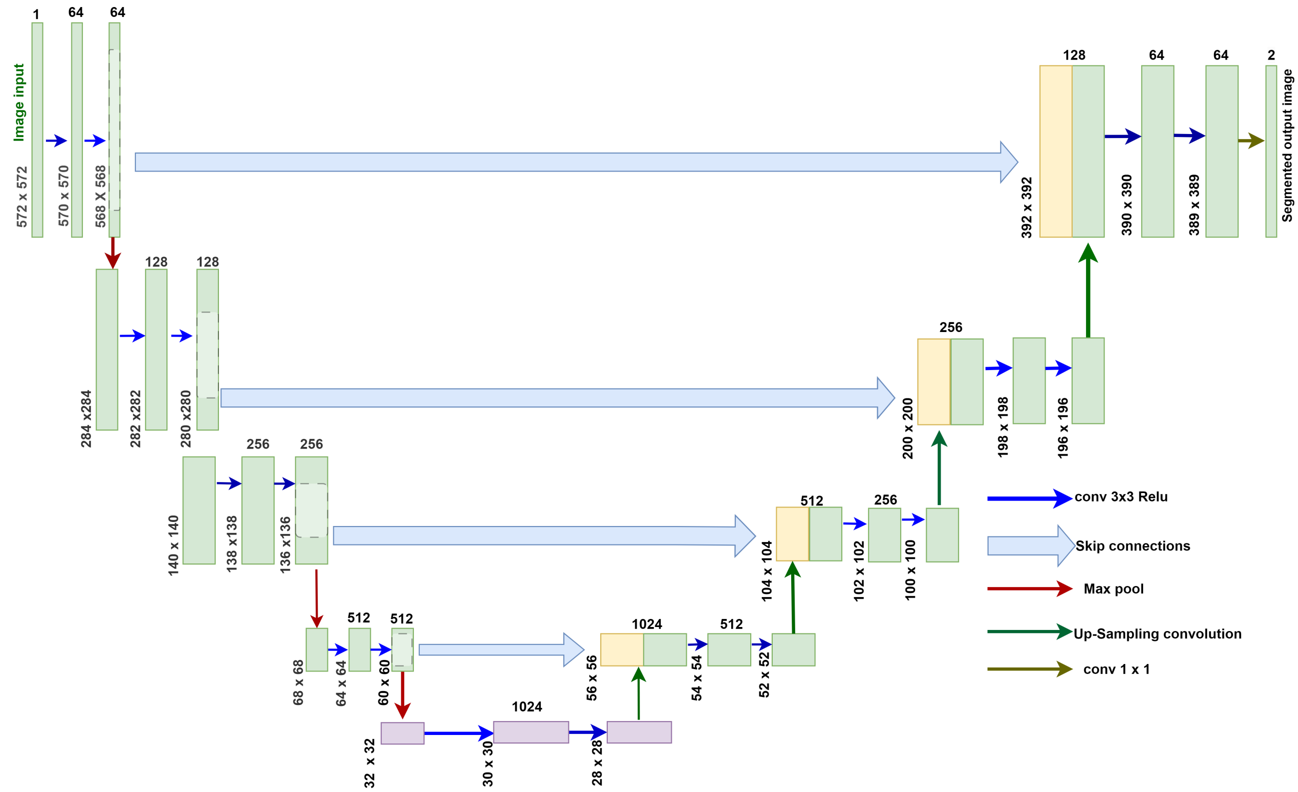


Fig. 3. U-Net Framework

the Atrous U-Net. The advantage of the proposed network compared to the traditional U-Net is that it extracts better features for semantic segmentation. This indeed results in better depth for the extracted features due to large field of reception.

D. Optimizer

- Stochastic Gradient Descent(SGD) - This optimizer has been developed to overcome the gradient descent problem. It uses forward and backward propagation for the computation. SGD updates the weights using the gradient of the loss function with reference to the parameters. The optimizer along with momentum is a slight improvement in SGD which helps the optimizer to converge faster.
- Adam – This optimizer performs better in comparison with other adaptive filters. It considers only first-order gradients with minimal requirement of memory. The optimizer computes the individual rate of adaptive learning for the different parameters from the 1st and 2nd estimates of moments of the gradient. It is the association of 2 other optimizers i.e. Adagrad and RmsProp.
- Nadam- Nadam is a variant of the Adam optimizer. It uses the Adam optimizer along with the Nesterov Accelerated Gradient(NAG).

E. Evaluation Metrics

- Categorical cross-entropy loss – The loss between label and prediction is termed Categorical cross-entropy. It is the gap between the predicted probabilities and the truth mask.

$$L_{CE} = -\sum_{i=1} T_i log(S_i) \tag{1}$$

where, T_i is denoted as ground truth mask and S_i is denoted as predicted area.
- Accuracy- It is a measure that predicts how often the predicted label equals the ground truth label.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

- Dice coefficient – It measures the area of overlapping or the area of intersection between the mask predicted by the network and the ground truth mask. It is generally represented between 0 and 1. 0 represents no overlapping and 1 represents perfect or complete overlapping betwixt the masks of ground truth and prediction.

$$DSC = \frac{2 \times I}{S_p + S_{gt}} \tag{3}$$

where, I is represented as the convergence between the predicted region and ground truth region and S_p is represented as the size of the predicted region and

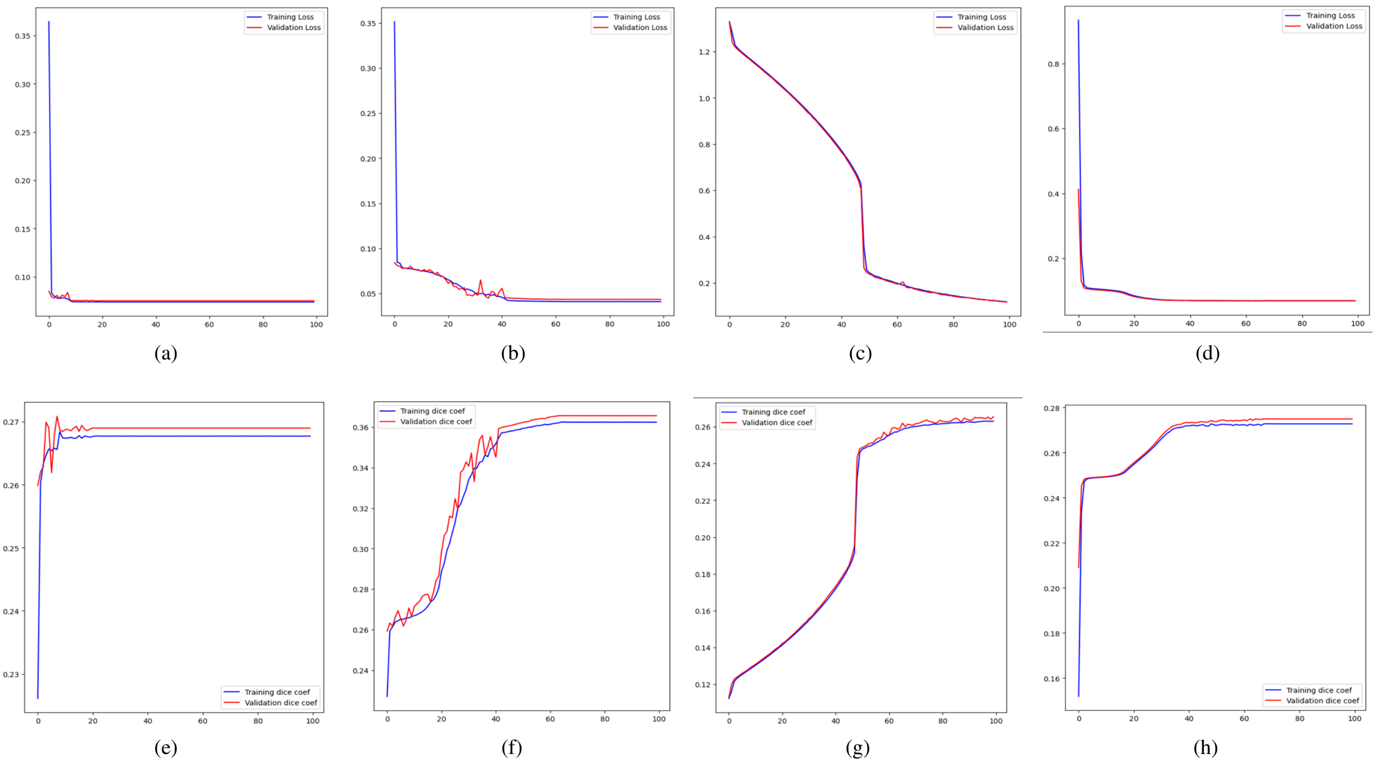


Fig. 4. Loss and Dice coefficient curve for (a) & (e) Adam optimizer, (b) & (f) Nadam optimizer, (c) & (g) Stochastic Gradient Descent(SGD) optimizer, (d) & (h) SGD optimizer with momentum

S_{gt} is represented as the size of ground truth region.

- Precision – It estimates the degree of accuracy in localizing the abnormal pixels in the spatial domain with reference to the given ground truth (provided mask).

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

- Sensitivity – It is the measure to estimate how well an algorithm identifies the Region of Interest(ROI)

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

where, TP is denoted as the number of pixels correctly identified as the object of interest, FN is denoted as the number of pixels that are not detected as part of the object of interest.

III. Result and Discussion

A. Simulation Result

We have used the GPU of NVIDIA GeForce RTX 2080 Ti/PCIe/SSE2. The model has been trained using different optimizers. The results obtained by the optimizers have been compared using various performance parameters. The BraTS2020 dataset has been split into 249 training data, 74 validation data, and 45 testing samples. The training accuracy obtained in Adam, Nadam, SGD, and SGD with momentum have been obtained as 98.42%, 98.67%, 98.31%, and 98.42% respectively, and the validation accuracy obtained by the 4 optimizers have been obtained as 98.26%, 98.68%, 98.33%, and 98.36% respectively,other performance parameters are shown in Fig. 4. The segmented outputs from the proposed model have been shown in Fig. 5.

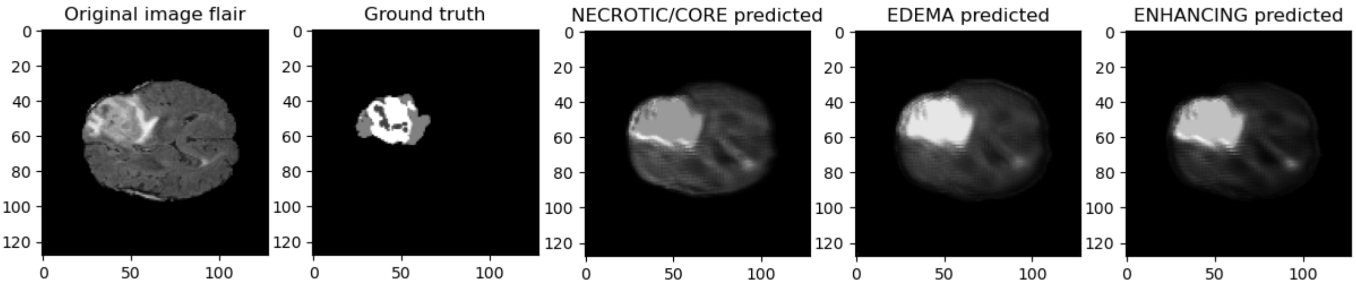
Table I shows the results obtained by testing the U-Net model and reinstating the convolution layer with the atrous convolution layer with a dilation rate as 2. The model has been trained using 100 epochs. It has been tested on the testing dataset of a batch size of 32 and a learning rate of 10^{-4} for all the optimizers.

TABLE I
Comparative Analysis of Performance Parameters on Different Optimizers

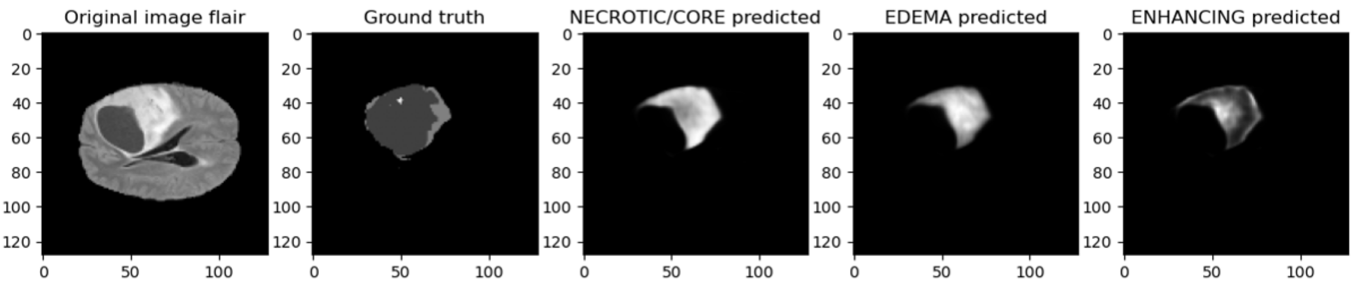
Optimizer	Accuracy	Dice coefficient	Specificity	Sensitivity	Precision
SGD+momentum	0.9821	0.2739	0.9940	0.9820	0.9819
SGD	0.9831	0.2443	0.9945	0.9835	0.9834
ADAM	0.9851	0.3608	0.9965	0.9811	0.9898
NADAM	**0.9978**	**0.3756**	**0.9976**	**0.9849**	**0.9904**

B. Discussion

The convolution layer that has been replaced with the atrous convolution layer with a dilation rate of 2 showed the best accuracy. These chosen optimizers are best performed in the training of the model. The segmentation model has been trained using these optimizers and the results were based on



(a)



(b)

Fig. 5. Visual segmentation results on the trained proposed models on the BRATS 2020.

the various performance parameters. The advantage of replacing the 2D convolution layer with a 2D atrous convolution layer with a dilation rate taken as 2 is that when the tumor dimension, position, circumference, shape, and intensity of the tumorous tissue merges with the normal healthy brain tissues it becomes difficult to extract and differentiate between the tumorous and healthy region. Generally, the small feature map is sometimes not able to collect all the required features. Hence, on adding the dilation rate, the model covers a larger field of feature map without any addition of extra parameters. The atrous convolution layer not only works as a filter but also a sampler. The model has been trained using Adagrad and Adamax but a high loss and low convergence rate was found as compared to the above chosen optimizers.

IV. Conclusion and Future Scope

This work well defines the purpose of the U-Net architecture for segmentation. The model works with different optimizers and it helps to obtain a better accuracy and dice coefficient score when the convolution layer is replaced by the atrous convolution layer. Nadam works better in the modified model as compared to Adam with a learning rate of 10^{-4}, giving comparatively better results than Adam, and efficiently segments the tumorous part. In the future work, efficiency of the segmentation model would be increased. The model can be incorporated with different data augmentation techniques for larger dataset and, hybrid segmentation model along with different dilation rates for better detection of the tumourous region.

References

[1] G. Mohan and M. M. Subashini, "Mri based medical image analysis: Survey on brain tumor grade classification," *Biomedical Signal Processing and Control*, vol. 39, pp. 139–161, 2018.

[2] S. Das, G. K. Nayak, L. Saba, M. Kalra, J. S. Suri, and S. Saxena, "An artificial intelligence framework and its bias for brain tumor segmentation: A narrative review," *Computers in Biology and Medicine*, p. 105273, 2022.

[3] D. Daimary, M. B. Bora, K. Amitab, and D. Kandar, "Brain tumor segmentation from mri images using hybrid convolutional neural networks," *Procedia Computer Science*, vol. 167, pp. 2419–2428, 2020.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[5] Z. Zhou, Z. He, and Y. Jia, "Afpnet: A 3d fully convolutional neural network with atrous-convolution feature pyramid for brain tumor segmentation via mri images," *Neurocomputing*, vol. 402, pp. 235–244, 2020.

[6] H. M. Bjoern, J. Andras, B. Stefan, K.-C. Jayashree, F. Keyvan, K. Justin *et al.*, "The multimodal brain tumor image segmentation benchmark (brats)," *IEEE Trans. Med. Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[7] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani, and C. Davatzikos, "Advancing the cancer genome atlas glioma mri collections with expert and radiomic features," *Scientific data*, vol. 4, no. 1, pp. 1–13, 2017.

[8] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, R. T. Shinohara, C. Berger, S. M. Ha, M. Rozycki *et al.*, "Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge," *arXiv preprint arXiv:1811.02629*, 2018.

[9] B. S. et. al., H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-gbm collection," *The cancer imaging archive*, 07 2017.

[10] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. Kirby, J. Freymann, K. Farahani, and C. Davatzikos, "Segmentation labels and radiomic features for the pre-operative scans of the tcga-lgg collection," *The cancer imaging archive*, vol. 286, 2017.