DAH-U-Net: A modified U-Net for Semantic Segmentation of MRI images for brain tumour detection

Mohankrishna Potnuru* and Dr. B.Suribabu Naick+

* Department of EECE, GIT, GITAM (Deemed to be University), Andhrapradesh, India
E-mails: mpotnuru@gitam.in[1], sbhukya@gitam.edu[2]
*Corresponding author Phone no.: +91-9014338355

Abstract

Using sophisticated image processing techniques on brain MR images for medical image segmentation significantly improves the ability to detect tumors. It takes a lot of time and requires a doctor's training and experience to manually segment a brain tumor. To address this issue, we proposed a modification in U-Net architecture called DAH-U-net that combines residual blocks, a rebuilt Atrous Spatial Pyramid Pooling (ASPP), and depth-wise convolutions. Also, a hybrid loss function which is explicitly aware of the boundaries is another thing we suggested. Experiments were conducted on two publicly available datasets and showed improved results in some metrics. For BraTS 2017 dataset the model achieved a segmentation accuracy of 93.2% and dice score of 0.945 for the tumor code(TC). For the BraTS 2018 dataset ,it achieved an accuracy of 94.3% and dice score 0.942 for the tumour core (TC), as compare to existing semantic segmentation models.

Key Words: Semantic segmentation; U-Net architecture, DAH-U-Net, high-grade glioma, Brain tumour Detection.

1   Introduction

The most common malignant tumor in guys appears to be gliomas. They often fall into one of two grades and begin in the glial central nervous system of the brain: Low-Grade Gliomas (LGG) develop more slowly and have a better chance of patient survival than High-Grade Gliomas (HGG), which grow swiftly and are consequently more aggressive [1]. The pathology The most prevalent malignant brain tumor is a glioma, which contains three subregions: the tumor core (TC), the enhancing core (EC), and the entire tumor (WT). It is essential to employ magnetic resonance imaging (MRI) when diagnosing brain tumors and when determining the best course of treatment. pictures created using fluid attenuation inversion recovery (FLAIR), contrast-enhanced T1-weighted pictures, and T1 and T2 weighted imaging techniques can all be used to detect brain tumors. T1 and T1ce images offer improved distinction for the tumor central region (excluding fibrotic edoema) compared to T2 and

FLAIR images, which tend to accentuate the entire tumor region (particularly fibrotic edoema) [2].

In order to analyze different subareas of brain tumors, these diverse genomes that offer complementary information might be combined. For repeatable and accurate tumour evaluation, which can support clinical detection, care plans, and monitoring [3], it is essential to effectively divide brain tumors and regional groupings using multi-modal MRI [4]. On the other hand, brain tumors and multi-modal MRI are still difficult for automatic systems to distinguish accurately. This is actually the situation because the scans frequently show ambiguous borders between healthy tissue and brain tumors. Additionally, whereas the liver [4] and the heart [5] were previously divided into segments using geometric factors, the size, volume, and location of brain tumors differ greatly between individuals [6]. As a result, it is difficult to classify brain tumors accurately given their past location and form.

Deep learning, machine learning, and convolutional neural networks (CNNs) have recently emerged as effective techniques for detecting brain tumors. Contrary to supervised machine learning techniques like decision trees [7] and SVM classifiers [8], CNNs may identify and categorize the most important features even without the need for manually constructed and evaluated features. It is quite difficult to create a suitable network architecture and training method for CNN-based categorization. Even yet, employing a two-dimensional CNN for 3D volumes necessitates fast GPUs and lots of RAM [9]. This type of models ignore three dimensional information, which reduces the effectiveness of segmentation. Although three-dimensional CNNs may be essential components of good three-dimensional properties, their storage requirements may limit the size, thickness, or quantity of input patches or attributes that they can manage [10]. In contrast to 2D CNNs, recently created 2.5D CNNs [11] may benefit from inter-slice properties and require less storage than their 3D counterparts. The entire tumor, the tumor core, and the expanding tumor core all have a hierarchical structure [12]. By using the segmentation of the complete tumor (tumor core) to influence the segment of the tumor center (improving tumor center), it might be able to eliminate false positive outcomes.

Below is a list of our key contributions:

- A modification to U-Net architecture with Atrous and depthwise convolution.

- We present a brand-new loss function that is based on metric recall, focal and cross entropy loss. Recall loss compares the instantaneous training recall performance of each class with its standard cross entropy and focal loss.

- The suggested hybrid loss develops a better semantic segmentation model that offers enhanced and balanced accuracy and Intersection over Union (IOU) performance.

- The suggested loss enhances feature learning for image segmentation along with the proposed modification in U-Net architecture.

The rest of the paper is organized as followed: In section 3, detailed architecture of proposed semantic segmentation model followed with results and discussion in section 4. Finally, conclusions at end of the section 4.

2   Related work

In recent years, a variety of deep learning frameworks for image processing applications, including as VGG16Net [13], Alexnet [14], and DenseNet [15], have been introduced. Because of its remarkable ability to detect feature maps dynamically, Deep Convolutional neural Networks are often used in domains like medical radiography and genetics [16]. The use of deep learning for computer-assisted tumor identification has also received considerable media attention. The development of neural network-based designs for brain tumor diagnostics in recent

decades has been significantly influenced by the Medical Image Computing and Computer-Assisted Invention (MICCAI) and Brain Tumor Segmentation (BraTS) challenges. For image segmentation, there are two distinct deep learning methods: CNN and Fully Convolutional Network (FCN). CNN-based techniques use the small patch classification approach for segmentation and detection. Using a segmented approach based on the CNN model, Havaei et al. divided the areas containing brain tumors from 2D MRI scans and then extracted important contextual information from the resulting solution of various sizes [18].
.

Recent initiatives have focused on the restricted responsive field/context modeling difficulty in FCN. Dilated convolution is used by DeepLab [18] and Dilation [19] to enlarge the region of interest. On the other hand, context modeling is the main emphasis of PSPNet [20] and DeepLabV2 [21]. The former supplies the ASPP modules, which use pyramid dilated convolutions with varying dilation rates, while the former offers the PPM component for getting contextual data for various zones. The most popular RGB-D datasets, as well as traditional machine learning-based techniques and deep learning-based network approaches, were summarized by the authors in [22] for RGB-D segmentation. An in-depth analysis of neural network topologies, datasets, and evaluation techniques for semantic segmentation of real-world photographs using deep learning models is given in [23].

Technique	Advantages	Limitations
Alexnet [14]	The network is simple and easy to train, and it can realize image classification.	Unable to process image semantics segmentation and multi-classification
FCN [32]	Becoming full convolutional layer (without connected layer).	Low accuracy of feature maps with high GPU occupation.
SegNet [10]	The first symmetric network.	Slow speed.
U-Net [18]	The structure is simple like the letter U with less parameter. Suitable for object detection in small number of medical images.	Difficult to obtain uniform standard of sub-sampling and up-sampling.
DeepLab [21]	Using atrous convolutional layer.	The atrous convolution layer occupied high GPU.
PSPNet [20]	Using the Pyramid pooling module to identify the prior information; Fantastic understanding and high identification of complex scenes	Base backbone of ResNet101 made processing speed slow

Table 1: Advantages and limitations of the techniques

[24] provided a high-level overview of radiography-specific deep learning-based medical image processing techniques. A description of state-of-the-art deep learning techniques for medical picture segmentation, along with information on network construction and model testing, was provided in [25]. In order to improve segmentation applications, [26] reviewed the literature on label noise in deep learning-based medical image processing and evaluated existing approaches on 3 diagnostic imaging databases. An overview of the methods suggested for integrating medical images from various sources was provided for the analysis of medical images [27]. For the purpose of using machine learning for clinical image segmentation, authors in [28] examined fully supervised, poorly supervised, and transfer learning methodologies as well as traditional approaches for addressing data scarcity and class imbalances.

Using five categories—explanatory, poorly supervised, classification strategies, and proactive reinforcement

learning—[29] studied methods for dealing with low sample counts in the medical field. In order to address the problems of sparse tags as well as poor tags (e.g., noisy tags, picture-level labels, sparse footnotes, etc.), [30] provided a study of pertinent literature in brain tumor segmentation. Similar to this, a variety of evaluations from the literature on object recognition [31] can be used to obtain the component in question's preliminary coordinates. The advantages and cons of the strategies utilized are listed in the Table 1 .

3   Methods and Materials

In this paper, we present Depthwise convolution, Atrous and Hybrid loss (DAH-U-Net), an upgraded version of the U-Net-based backbone network that combines residual blocks, a rebuilt atrous spatial pyramid pooling (ASPP), and depth-wise convolutions. Also, a hybrid loss function which is explicitly aware of the boundary is another thing we suggest.
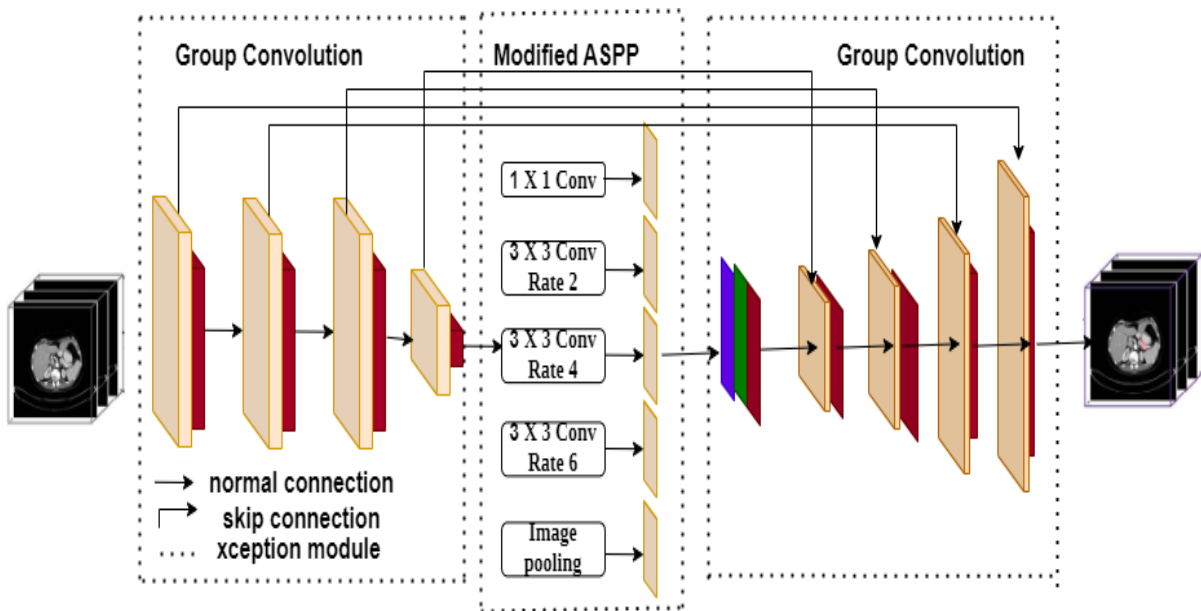
3.1   Model Architecture

Figure 1: Proposed model architecture

Our model utilizes a U-shaped decoder and encoder layout, as seen in Fig 1, which enhances the fundamental U-Net layout in a number of ways.

- First, we use group convolution to substitute the ordinary convolutions in both decoder and encoder section of layout, with a notable exception of the first most layer, such that the intra and interchannel correlation information is individually extracted throughout each level's of encoding process [33]. In order to separately capture the specific changes among neighboring slices of images, which is beneficial for more precise segmentation, we use the overlay of neighboring slices holding the foreground as the input of our model, which is based on this structure. In principle, the channels should be handled individually, it is preferable to avoid mapping them together.

- Second, we have included a residual architecture [34] between neighboring convolution blocks, which helps to reduce the loss incurred during down-sampling process.
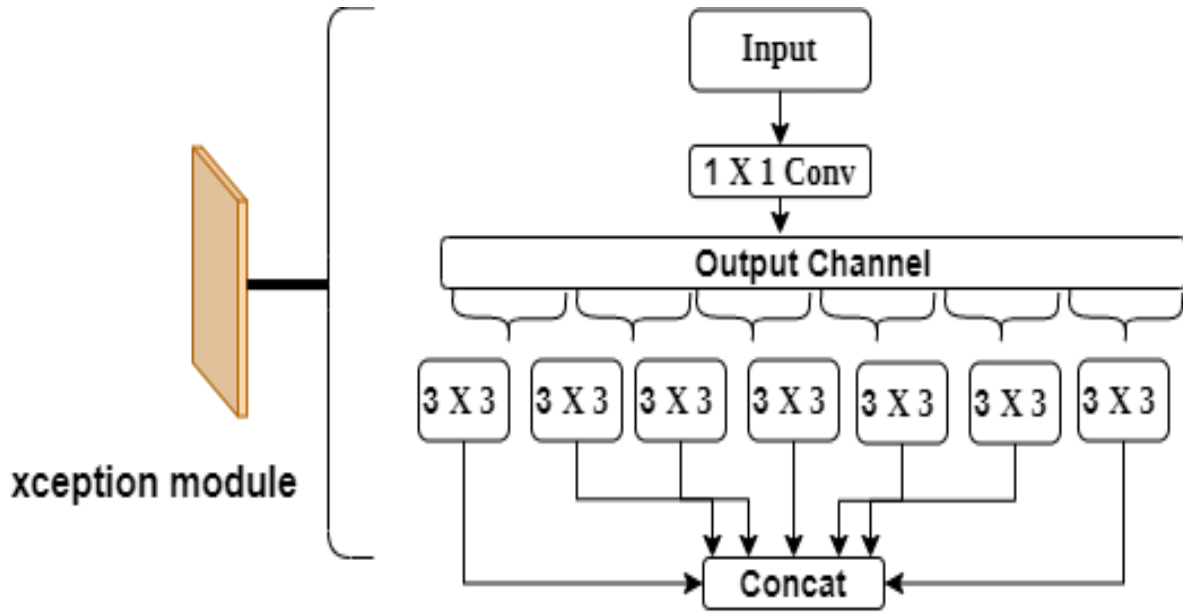
Figure 2: Xception module

- Third, to minimize risk of data loss while down-sampling, we established a bottleneck layer utilizing ASPP [35], which is crucial for retrieving multi-level relevant information. Convolution procedures are carried out simultaneously on the feature maps created during the encoding stage with various dilation rates in order to gather the environment of the image at several scales and produce more precise foreground location data [36].

Existing approaches might not be capable of recognizing the presence of the tumor when given a hard input since the occupancy of tumor in CT images is very small and is override with the neighbor organs like liver, stomach etc., and also size of tumor is flexible and variable with respect to patients. For such a small and complex structured organs, extraction of multi-level and single level contextual information is crucial. In the decoding stage of model, we applied group deconvolution [37] at every stage of each layers to get back original input image size and 1×1 convolution was applied to limit the number of feature maps to two.

3.2   Depth-Wise Separable Convolution

At, encoder section of model, we substitute the ordinary convolution with a particular group convolution called depth-wise separable convolution. Ordinary convolution is a simultaneous mapping of spatial and correlation information of an individual channels [38]. Although these two types of input are connected, depth-wise convolution in Inception allows the correlations between the two to be broken [39]. According to the Inception hypothesis, the two correlations can really be mapped individually to produce superior results since, they are independent of one another [40].

As CT Brain images have many slices, an individual mapping of information across channels makes more sense. In our approach, we employ the Xception which is extreme case of Inception method, in which, the number of groups in the group convolution is equal to the number of input channels. With this process, totally decoupling of spatial correlation between channels and the inter-channel correlation occur. The input feature map undergoes channel-by-channel linear transformation using 1×1 convolution, and the resultant feature map is fed through a series of 3×3 convolutions. Because there are exactly as many groups in our grouped convolution as there are input channels, each filter in this convolution process has a convolution kernel of 33, meaning that each channel of the input feature map is only convolved by one kernel with a size of 331. The

output feature map is created by stacking the outputs of these filters. Because there are exactly as many groups in grouped convolution as there are input channels, each filter in this convolution process has a convolution kernel of 3 × 3, meaning that each channel of the input feature map is only convolved by one kernel with a size of 3 3 × 1. The output feature map is created by stacking the outputs of these filters.

In calculation of model parameters, lets assume P is the feature maps at input layer of model and Q feature maps at output layer with kernel size of convolution layer is 3, then number of parameters of model is 3×3×P×Q. In case of depth-wise separable convolution, the number of parameters are addition of parameters of depth-wise and point wise (Pg = $P_{depthwise}$ + $P_{pointwise}$ = (3 x 3 x Q) + (1 x 1 x Q x P)). From these observations, our model parameters are less as compared to ordinary U-Net and it enhance the feature maps which helps for good classification of tumor and non-tumor regions of brain. In addition to this, our model has two level down-sampling with replacement of ordinary convolution with depth-wise separable structure Xception as in Fig. 2. To achieve information decoupling, convolution kernels with the same amount of input channels are employed in each down-sampling phase, followed by an ordinary convolution to increase the amount of feature maps produced. If ordinary convolution is being used entirely, the a whole down-sampling approach involves 1, 261, 532 3×3 convolution kernels, whereas our enhanced structure requires only 679, 329 3×3 convolution kernels.

3.3 ASPP Module

Brain tumor images typically have slightly out of focus boundaries and are easily confused with neighboring soft tissue structures, especially because it holds a comparatively small area in a CT image with such a complex background and less than 1.5% in a 2D image. This makes determining whether or not the tumor occurs in the image much more difficult. Just above all, the existing models are incapable of extracting sufficient information about the position of the tumor, which is heavily dependent on the image's overall context. To enhance the capability of extracting features in our architecture, we use an ASPP plugin with atrous convolution. The ASPP plugin, uses numerous concurrent atrous convolution layers with various sampling rates and takes its design cues from the spatial pyramid. At various scales, the feature map simultaneously captures the context. We presume that the deep and superficial features of medical images are significant in the case where the medical image doesn't already consist noise, any unwanted information or complex background, and that the fusion of various levels of features can result in improved judgment in this situation. The ASPP module that we employ mostly consists of the components shown in Fig. 2:

- This architecture has one ordinary and three atrous convolution layers with size 1×1 and 3×3 respectively. Filters will deteriorate into a straightforward 1×1 convolution with only the filter center functioning when the dilated rate is almost equal to the feature map size. The dilated ratio of the original module is therefore scaled to (2, 4, 6).

- Features of the image are obtained by applying average pooling globally and these extracted features are passed through a convolution layer of size 1×1 which performs bi-linear interpolation to get back the original image of same size/resolution as like input image.

- In order to create a new extracted features with 256 layers, the four different types of extracted features from the previous two steps are combined in the channel dimension and thereafter routed through 1×1 convolution for fusion.

- The ASPP unit, which can collect multi-level location information more effectively and has higher classification and learning skills to recognize and localize the tumor, partially corrects the standard U-Net's deficiency in describing information. Additionally, if the dilate rate approaches or even exceeds the size of the input feature map, it will collapse into 1×1 convolution, and a dilate rate that is too high will not allow for pixel-level output. For these reasons, we adopt a reduced dilate rate of (2, 4, 6).

3.4   Loss function

Loss functions determine the amount by which the forecasts differ from the actual. Lower loss values signify higher accuracy of model, whilst higher loss values indicate predictions of model are fairly correct. Always model should looks for minimization of loss function as feasible, ideally it must be near to zero. Models takes the loss function to learn the trainable parameters, like biases and weights and these are updated in each iteration based on gradient descent process. Cross entropy is one loss function mostly used for binary classification but some times fails for multi-class classification problems. It defined from the concept of entropy in mathematics which is defined as

$$Entropy = P_i log(P_i) \tag{1}$$

$$CrossEntropy = -\sum_{i}^{n} Y_i log(P_i) \tag{2}$$

Where Y is the ground true or actual output and P is the model predicated output. Due to the fact that probabilities might range from 0 to 1, the logarithm of this produce negative value so in order make the entropy positive we included negative in front of equation. In case of class imbalanced problems, as loss function followed by gradient descent mostly focus on majority class, which will cause the weights to update in a way that increases the model's confidence in forecasting the majority class while decreasing its concentration on the minority classes. Solution to this issue is focal loss [41]

3.4.1   Focal loss:

By concentrating on the instances where the model fails instead of the ones where it can reliably forecast, focal loss makes sure that forecasts on challenging examples get better over time instead of getting too confident with simple ones. This is accomplished by a process termed "Down Weighting" in focal loss. By lessening the impact of simple instances on the loss function, down weighting emphasizes the importance of difficult examples. By incorporating a modulating factor $(\mu)\mu = (1 - P_i)_\gamma$ into the Cross-Entropy loss, this method can be put into practice.

$$Focalloss = -\sum_{i}^{n} \alpha(1 - P_i)^\gamma log(P_i) \tag{3}$$

Where the cross validation tunable focusing parameter is called γ (Gamma) and α is weighing factor. The behavior of Focal Loss for various values of γ and α is depicted in Fig. 3. From figure, the following observations are made

- Since the p_i is low for the sample that was incorrectly classified, the μ is near to or exactly 1 which leads loss function remains unaltered and became a cross-entropy loss.

- The μ will tend to zero as the model's confidence level rises, which is indicated by π is equal to 1, which will reduce the loss amount for correctly categorized cases. In order to lessen the impact of the simple instances on the loss function, γ re-scale the μ such that the easy instances are down-weighted more than the hard ones. To our dataset focal loss performance is better with γ value equal to 2.

- Focal Loss is the same as Cross Entropy when γ is equal to 0.

3.4.2   Recall loss:

To solve the imbalance dataset issue, we used a novel recall-based performance-balanced loss called recall loss. Based on the value of recall of model during training, the model weights are updated towards minimization of loss function of that class. As opposed to the hard example mining method in the focal loss, it is an illustration of hard class mining. The recall loss, in contrast to focal loss and other losses, dynamically modifies its weights
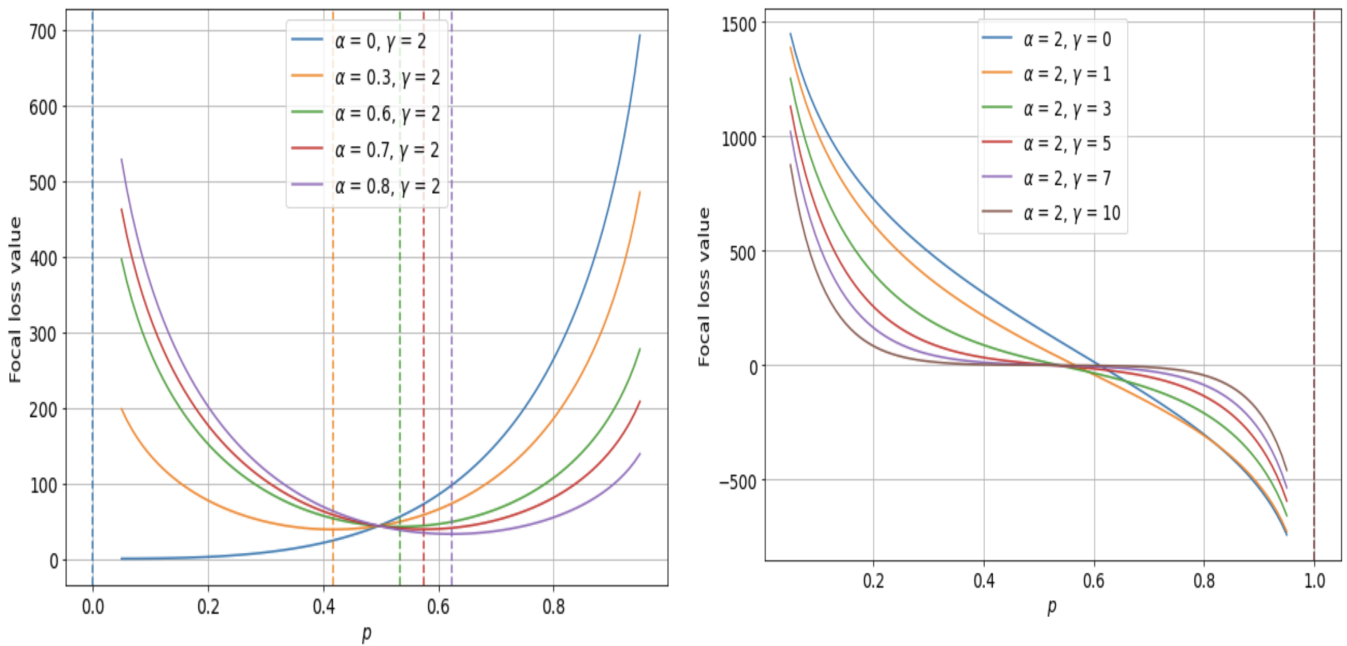
Figure 3: focal loss variation with respect to a) α and b) γ

with training based on per-class recall value. At the cost of Intersection over Union (IOU), which takes false positives into account in semantic segmentation, the CB loss increases accuracy. The precision and recall of each class may be successfully balanced by our recall loss, which enhances accuracy while maintaining a competitive IOU.

$$RecallLoss = -\sum_{i=1}^{C} \frac{FN_c}{FN_c + TP_c} N_c log(P^c) \tag{4}$$

Where N_C, FN_C and TP_C are number of samples, false positive and true positive rate of class C respectively. The same formula is applicable for other classes if applicable. It is possible to use the recall-loss as the standard cross entropy which is weighted by the class-wise FN_C. The second realization is that minority classes, which have greater FN_C, are probably harder to categorize than large classes, which have lower FN_C. Gradients of minority classes will therefore be raised and gradients of majority classes will be repressed, analogous to inverse frequency loss [42].

3.4.3   Hybrid loss function:

In our work we took the advantages and disadvantages of above said three loss functions for better training of model and better performance of the model. In our case, loss function is the sum of above three loss functions as in Fig. 4.

4   Results

4.1   Dataset information

There is a detailed discussion of the openly accessible standard datasets used in this study. The DAH-U-Net architecture was evaluated using two benchmark datasets. The datasets under consideration are BraTS 2017 and BraTS 2018 [44][45][46] (https://www.med.upenn.edu/cbica/brats2020/data.html). Images from 285
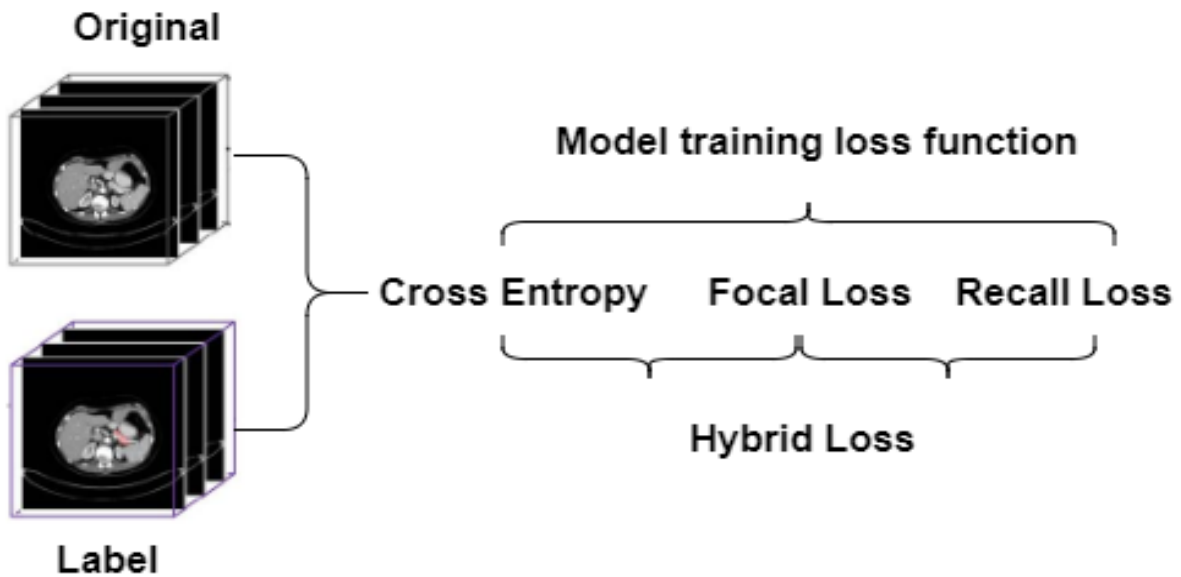
Figure 4: Proposed training loss function

glioma patients, including 210 HGG instances and the remaining LGG cases, are included in the BraTS 2017 collection. The validation dataset for BraTS 2017 also contains pictures of 46 patients with unknown grades. Since the labels of the validation dataset and the ground truths of the data sets were assigned by experts, only the BraTS website can produce the conclusions. The dataset is in the form of NIfTI files (.nii.gz) which contains multimodal scans (images taken at different direction and position). It has four 'channels' of information – 4 different volumes of the same region include Native (T1), Post-contrast T1-weighted (T1CE), T2-weighted (T2), and T2 Fluid Attenuated Inversion Recovery (FLAIR) volumes. All the imaging datasets have been segmented manually and were approved by experienced neuro-radiologists. Annotations (labels): Label 0: Unlabeled volume, Label 1: Necrotic and non-enhancing tumour core (NCR/NET), Label 2: Peritumoural edema (ED), Label 3: Missing (No pixels in all the volumes contain label 3), Label 4: GD-enhancing tumour (ET).

4.2   Pre-processing of dataset

In this section, downloaded images are processed for unzip, scaling all volumes by using MinMaxScaler (https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html) and installation of NiBabel library to handle nii files (https://pypi.org/project/nibabel/). Also, in this step, a single multi-channel volume is obtained by combining the three non-native volumes (T2, T1CE and Flair). In labels of the dataset, label number 3 is missing, so reassign pixels of value 4 to value 3. All the volumes are resized to 96x96x96 by cropping the volumes and remove useless blank regions around the actual volume of interest. In order to maximize training on labeled volumes, drop all volumes where the amount of annotated data is less that certain percentage. Finally split image and label volumes into train and validation datasets.

4.3   Experimental Setup:

The model is implemented in Matlab 2019 with CUDA available service and version of cude is 5.0. Stochastic gradient descent with the Adam optimizer (learning rate = 1e-4) was used to minimize the loss function (cross entropy). On an NVIDIA GTX Titan X GPU with cuDNN v5.0, each training takes about 6 hours. The DAH-"U-Net" architecture also enables networks to be trained on GPUs with as little as 3GB of RAM. Even though
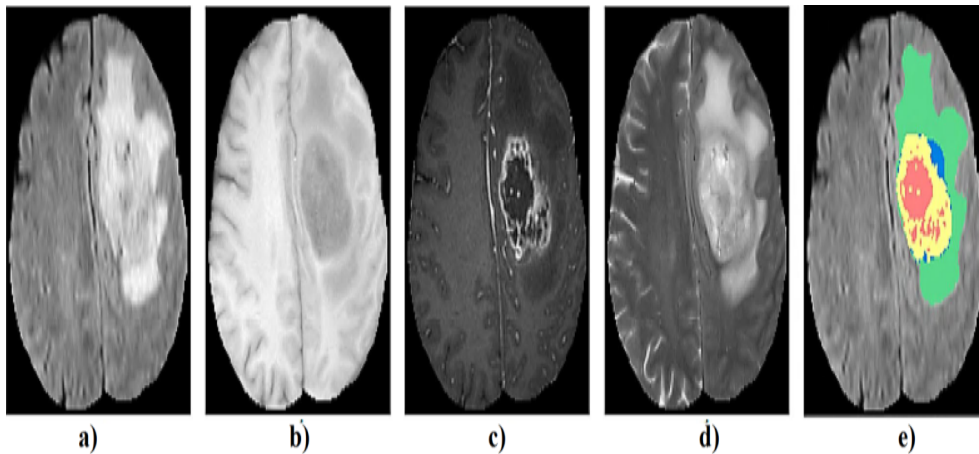
Figure 5: Sample dataset: a) FLAIR, b) T1, c) T1C, d) T2, and e) ground truth

the sizes of the volumes in the digested databases prevent dense training on entire volumes for this size of structure, dense interpretation on entire volumes still seems to be possible because it just takes a forward pass and so consumes less storage. Volume segmentation takes about 30 seconds throughout this manner, but it needs 12GB ofGPU memory.

5   Discussion

In this section, we conducted quantitative and qualitative research. Figure 5 shows some sample images of dataset and visual variation with FLAIR, T1, T1C and T2. The performance evaluation in terms of statistics is covered in quantitative analysis. The visual quality of the results is examined in qualitative analysis. We chose the Dice score as the figure of merit for measuring DAH-"U-Net" performance. The previous state-of-the-art procedures used dice score; therefore it will help us get a better result. A quantitative comparison of existing state-of-the-art and DAH-"U-Net" architectures is performed. The Dice score indicates the degree of resemblance between sets R and S, which may be stated numerically as

$$Dicescore = \frac{2X|R \cap S|}{|R| + |S|} \qquad (5)$$

Architecture	tumor core (TC)	whole tumor (WT)	enhancing core (EC)
Seg-Net [10]	0.840	0.720	0.620
U-Net [21]	0.831	0.801	0.750
Densely CNN [33]	0.720	0.830	0.810
FCNN [35]	0.865	0.864	0.816
BU-Net [36]	0.901	0.867	0.835
Shuffle UNet [43]	0.943	0.902	0.910
DAH U-Net	0.945	0.908	0.915

Table 2: Comparitive results in terms of dice score with other nets for HGG data

The suggested model was first tested on the BraTS 2017 dataset, which contains 210 cases. Eighty percent of these samples have been used for training, while the remaining twenty percent were used for testing. The models are optimized with the proposed loss function. The BraTS challenge defines the training and testing scenarios. Table 2 compares the results obtained by DAH-"U-Net" to those obtained by other approaches.

Architecture	tumor core (TC)	whole tumor (WT)	enhancing core (EC)
Seg-Net [10]	0.833	0.703	0.496
U-Net [21]	0.870	0.762	0.700
PSP-Net [37]	0.809	0.701	0.554
Novel-Net [35]	0.876	0.763	0.642
BU-Net [36]	0.892	0.783	0.736
Shuffle UNet [43]	0.932	0.834	0.832
DAH U-Net	0.936	0.835	0.835

Table 3: Comparitive results in terms of dice score with other nets for BraTS 2017 dataset(57 MRI scans)

Cost functions, optimizers, and other co-factors were used in all of the architectures in a similar way. When compared to its baseline U-Net, DAH-"U-Net" achieved gains of 5.5 percent, 7.8 percent, and 5.7 percent for whole, core, and enhancing tumour segmentation, respectively. The suggested model exceeded four current state-of-the-art strategies in terms of Dice score, outperforming them all in terms of detection accuracy for data.

Results for the entire dataset of BraTS 2017 were collected for further assessment of DAH-"U-Net". 228 MRI scans have been used for training and the rest 57 MRI scans have been used for testing in this research. Table 3 shows the findings obtained with BU-Net and their comparison to the best known approaches. BU-Net receives the best findings from the literature in cases of enhancing tumour and core tumour. In comparison, our suggested model showed a 0.32 percent and 0.49 percent performance gain for enhancing tumour and core tumour, respectively as in Table 4.

Architecture	tumor core (TC)	whole tumor (WT)	enhancing core (EC)
U-Net [21]	0.860	0.790	0.767
3DU-Net [38]	0.885	0.718	0.760
Ensemble Net [39]	0.881	0.777	0.773
TTA [35]	0.873	0.783	0.754
S3DU-Net [40]	0.894	0.831	0.749
MCC [41]	0.882	0.748	0.718
BU-Net [36]	0.901	0.837	0.788
Shuffle-U-Net [43]	0.938	0.872	0.802
DAH-U-Net	0.935	0.875	0.806

Table 4: Comparitive results in terms of dice score with other nets for BraTS 2018 dataset(66 MRI scans)

Model	PPV	NPV	SE	SP	AC	F1
U-Net [21]	0.583	0.538	0.615	0.734	0.703	0.761
3DU-Net [38]	0.632	0.555	0.655	0.749	0.756	0.738
Ensemble Net [39]	0.739	0.571	0.631	0.746	0.752	0.792
TTA [35]	0.745	0.626	0.775	0.792	0.762	0.809
S3DU-Net [40]	0.799	0.621	0.761	0.769	0.836	0.833
MCC [41]	0.801	0.643	0.797	0.742	0.857	0.825
BU-Net [36]	0.814	0.652	0.877	0.854	0.893	0.889
Shuffle-U-Net [43]	0.842	0.759	0.872	0.928	0.928	0.881
DAH-U-Net	0.852	0.801	0.884	0.935	0.932	0.882

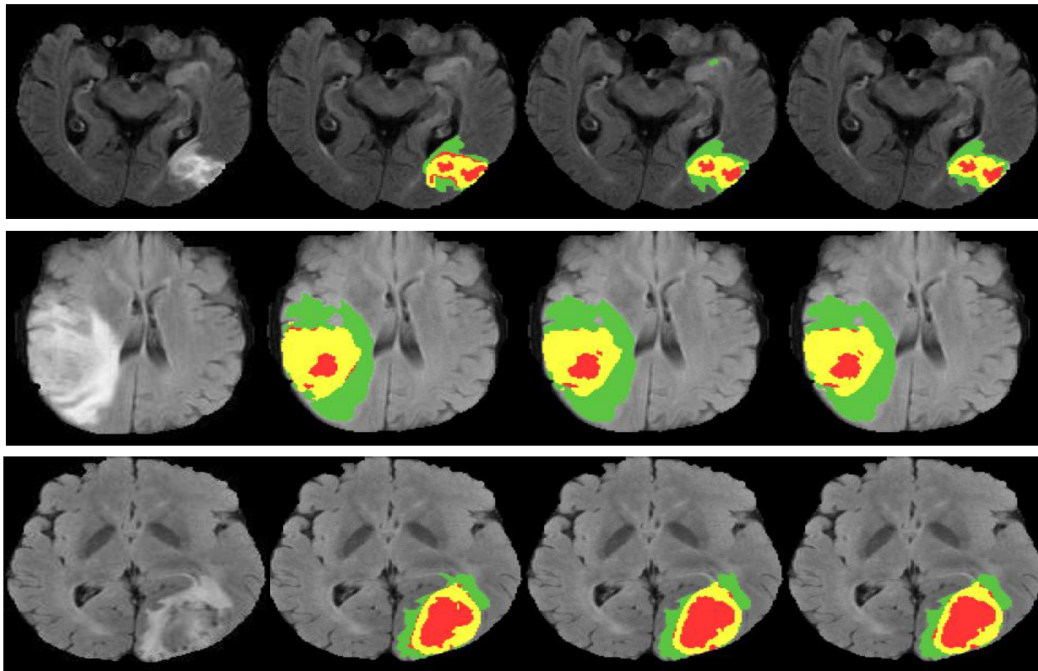Table 5: Various metrics obtained with various nets for BraTS 2017 dataset

Figure 6: a) Original image a) Ground truth b) Segmented results with Shuffle-U-Net [36] c) Segmented results with DAH-U-Net. Three different colors are used to depict three different types of tumours. Necrosis and non-enhancing tumours are shown by red; edoema is represented by green; and an enhancing tumour is represented by yellow.

Table 5 shows the various metrics used for evolution of the DAH-U-Net. The percentage of true positives classified correctly by a diagnostic test is known as sensitivity. It demonstrates how effective the test is at detecting disease. The proportion of true negatives correctly determined by a diagnostic test is known as specificity.

Model	PPV	NPV	SE	SP	AC	F1
U-Net [21]	0.874	0.664	0.769	0.706	0.795	0.811
3DU-Net [38]	0.837	0.654	0.798	0.793	0.798	0.836
Ensemble Net [39]	0.848	0.679	0.824	0.806	0.791	0.859
TTA [35]	0.856	0.705	0.843	0.827	0.775	0.816
S3DU-Net [40]	0.846	0.713	0.862	0.819	0.774	0.866
MCC [41]	0.806	0.762	0.854	0.811	0.839	0.884
BU-Net [36]	0.840	0.749	0.874	0.858	0.918	0.896
Shuffle-U-Net [43]	0.867	0.863	0.905	0.867	0.932	0.944
DAH-U-Net	0.869	0.869	0.912	0.875	0.939	0.950

Table 6: Various metrics obtained with various nets for BraTS 2018 dataset

It indicates how well the test detects normal (negative) conditions. The proportion of true outcomes in a population, whether true positive or true negative, is known as accuracy. It assesses the accuracy of a diagnostic test on a given condition. In terms of accuracy, table 5 displays the effectiveness of the suggested U-Net approach and the existing models. According to the findings, the DAH-U-Net model has an average accuracy of 93.2 percent, whereas the present models approach only has an accuracy of 90.5 percent. In similar, the DAH-U-Net outperforms in terms of other metrics as in table 5 and 6.

Table 5: Metrics and formulas

Metrics	Description	Equation
Accuracy (AC)	The percentage of input images' weight that is correctly segmented is measured by accuracy.	$\text{Accuracy} = \dfrac{T_P + T_N}{T_P + F_P + T_N + F_N} \times 100$
Specificity (SP)	The segmentation process of input images' proportion is how exactly performed for negative results are measured by the specificity.	$\text{Specificity} = \dfrac{T_N}{T_N + F_P} \times 100$
Sensitivity (SE)	The segmentation process of input images' proportion is how exactly performed for positive results are measured by the sensitivity.	$\text{Sensitivity} = \dfrac{T_P}{F_N + T_P}$
positive prediction value (PPV)	The number of pixels obtained in the segmented images as the positive which are actually positive and it is called as the positive predictive value that is estimated by below equation	$\text{PPV} = \dfrac{\text{TP}}{\text{FP} + \text{TP}}$
negative prediction value (NPV)	The number of pixels obtained in the segmented images as the negative which are actually negative and it is called as the positive predictive value	$\text{NPV} = \dfrac{\text{TN}}{\text{FN} + \text{TN}}$
F1-score	F1-score represents the single measure that indicates the network model accuracy for specific input images.	$F1 - score = \dfrac{T_P}{T_P + \frac{1}{2}(F_P + F_N)}$

T1 scans are typically utilized to detect healthy tissues, while T2 images are used to define the edema region, which displays a strong signal on the image. The strong signal of the acquired contrast agent (gadolinium ions) in the active cell zone of the tumour tissue in T1-Gd images easily distinguishes the tumour border. Because necrotic cells do not interact with the contrast agent, they can be seen in a low-intensity part of the tumour core, allowing them to be easily distinguished from the active cell region in almost the same series. From figure 7 and figure 8, the segmented regions obtained with DAH-U-Net are much similar to the label or mask or ground truth.

5.1   Importance of loss function

Loss functions are necessary aspects of a machine learning approach, and they're utilized to train the learner's attributes. Because the ideal required parameters are determined by minimizing the mean amount of the loss provided a labeled training dataset, selecting the right loss function is critical. In this paper, we suggest extending a deep CNN with a combination of the two loss function to boost simultaneously robustness and measuring accuracy. From figure 9, it is observed that hybrid loss function which is combination of cross entropy, recall and dice coefficient loss got good accuracy as compared to individual loss function in both datasets (BraTS 2017 and 2018). So in this article we used hybrid loss function for optimization of model by selecting appropriate
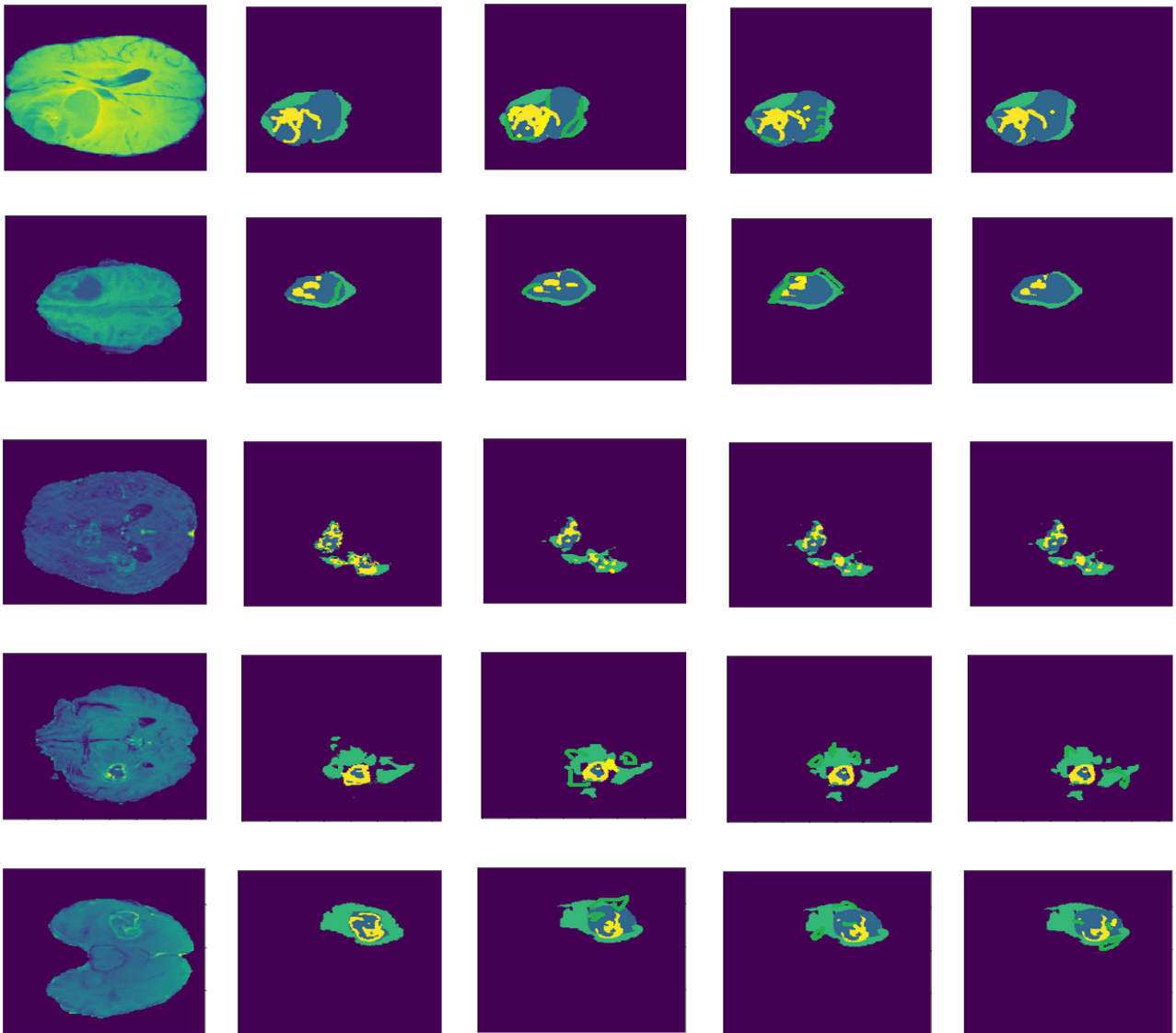
Figure 7: Segmented results for BraTS 2017 a) FLAIR b) GE c) LABEL d) BU-"U-Net" e) DAH-"U-Net

weights of fully connected neural network and optimization of kernel or filter elements.

6   Conclusion

On two publicly available datasets, with data augmentation on the training set, the proposed Shuffle-U-Net, DAH-U-Net and other models were assessed. The outcomes of the trial were contrasted with those of cutting-edge techniques. The findings showed that the suggested DAH-U-Net exhibits greater Dice, accuracy, and recall than other models with less computing parameters, and offers a potential strategy for semantic segmentation of tiny tissues. Visual examination of the overlapped maps reveals that the suggested DAH-U-Net can successfully fit the human segmentation, further supporting the efficacy of our approach.

On the BraTS dataset, the suggested architecture has a segmentation accuracy of 93.2 percent and a dice score of 0.945 in case of tumour core (TC) for BraTS 2017. When trained on the BraTS dataset and validated and evaluated on BraTS 2018, the model achieves accuracy and estimated dice scores of 94.3 percent and 0.942
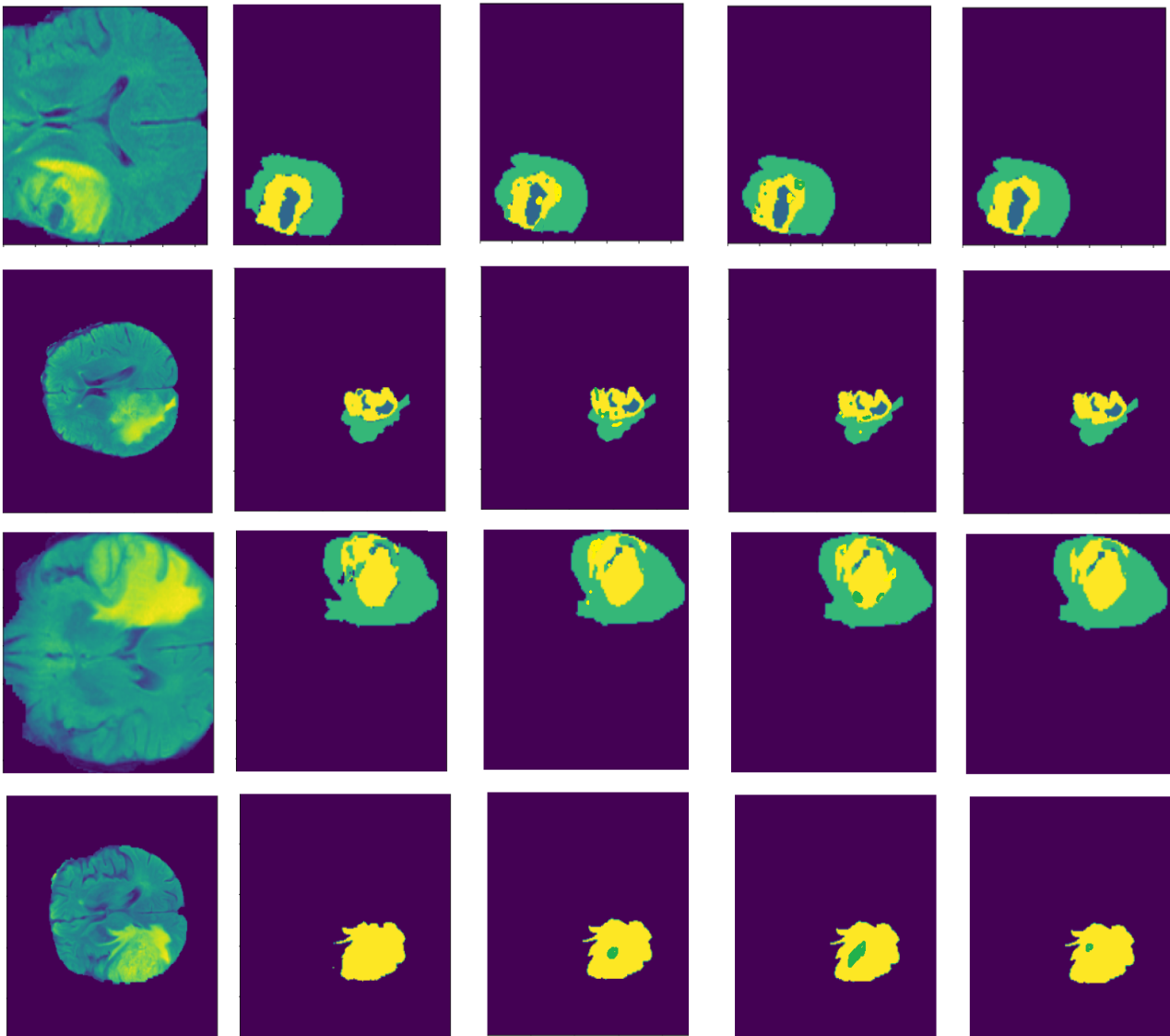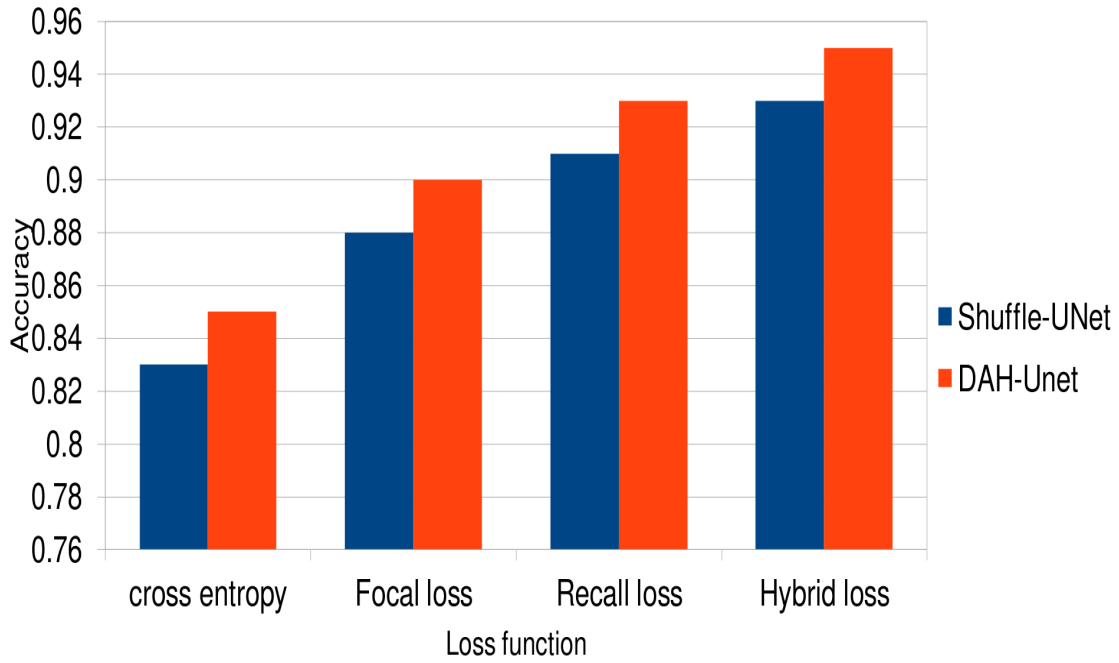
Figure 8: Segmented results for BraTS 2018 a) FLAIR b) T1 c) LABEL d) BU-U-Net e) DAH-U-Net

in case of tumour core (TC) respectively. The DAH-"U-Net" model outperforms the existing methods in terms of quality and quantitative analysis. Future scope of the article can be extended by including advanced loss functions and proposing a new architectures.

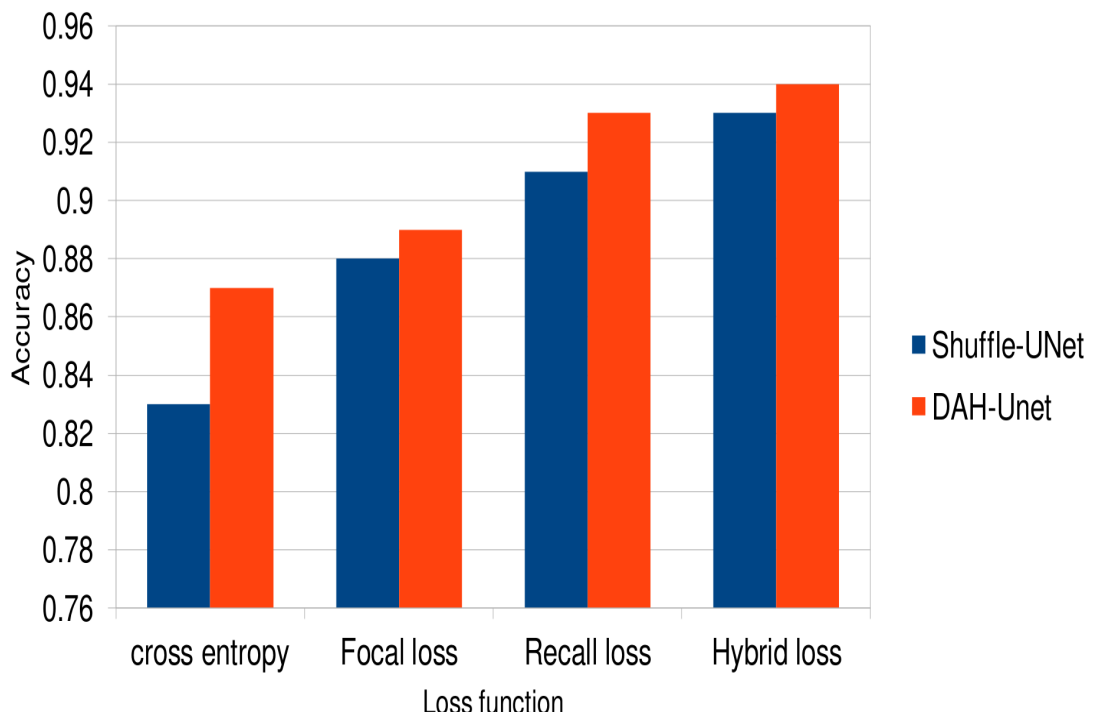## Effect of loss on accuraacy

Effect of loss on Accuracy

Figure 9: Effect of loss function on accuracy a) for BraTS 2017 b) 2018 dataset

References

[1] Louizos, Christos and Max Welling. "Structured and Efficient Variational Deep Learning with Matrix Gaussian Posteriors." International Conference on Machine Learning (2016). `https://doi.org/10.48550/arXiv.1603.04733`

[2] Menze BH, Van Leemput K, Lashkari D, Weber MA, Ayache N, Golland P. A generative model for brain tumor segmentation in multi-modal images. Med Image Comput Comput Assist Interv. 2010;13(Pt 2):151-9. doi: 10.1007/978-3-642-15745-5_19. PMID: 20879310; PMCID: PMC3050038. `https://doi.org/10.1007/978-3-642-15745-5_19`

[3] Bakas, Spyridon et al. "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge." ArXiv abs/1811.02629 (2018): n. pag. `https://arxiv.org/abs/1811.02629`

[4] Wang, G., Zhang, S., Xie, H., Metaxas, D. N., & Gu, L. (2015). A homotopy-based sparse representation for fast and accurate shape prior modeling in liver surgical planning. Medical image analysis, 19(1), 176 186. `https://doi.org/10.1016/j.media.2014.10.003`

[5] Grosgeorge, Damienetal. "Graphcutsegmentation with astatistical shape model in cardiac MRI." Comput. Vis. Image Underst. 117 (2013): 1027-1035. `https://doi.org/10.1016/j.cviu.2013.01.014`

[6] Karri, C., & Jena, U. (2016). Fast vector quantization using a Bat algorithm for image compression. Engi neering Science and Technology, an International Journal, 19(2), 769-781. `https://doi.org/10.1016/j.jestch.2015.11.003`

[7] Kotowski, K., Adamski, S., Malara, W., Machura, B., Zarudzki, L., Nalepa, J. (2021). Segmenting Brain Tumors from MRI Using Cascaded 3D U-Nets. In: Crimi, A., Bakas, S. (eds) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2020. Lecture Notes in Computer Science(), vol 12659. Springer, Cham. `https://doi.org/10.1007/978-3-030-72087-2_23`

[8] Joseph, Nisha et al. "Patient Specific Brain Tumor Segmentation using Context Sensitive Feature Extraction in MR Images." International Journal of Computing and Digital Systems 9 (2020): 1091-1097. `https://dx.doi.org/10.12785/ijcds/090607`

[9] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017 Jan;35:18-31. doi: 10.1016/j.media.2016.05.004. Epub 2016 May 19. PMID: 27310171. `https://doi.org/10.1016/j.media.2016.05.004`

[10] C. Karri and M. S. R. Naidu, "Deep Learning Algorithms for Secure Robot Face Recognition in Cloud Environments," 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom), Exeter, United Kingdom, 2020, pp. 1021-1028. `https://doi.org/10.1109/ISPA-BDCloud-SocialCom-SustainCom51426.2020.00154`

[11] Geng, Y., Ren, Y., Hou, R., Han, S., Rubin, G. D., & Lo, J. Y. (2019, March). 2.5 D CNN model for detecting lung disease using weak supervision. In Medical Imaging 2019: Computer-Aided Diagnosis (Vol. 10950, p. 109503O). International Society for Optics and Photonics. `https://doi.org/10.1117/12.2513631`

[12] Roth, H.R. et al. (2014). A New 2.5D Representation for Lymph Node Detection Using Random Sets of Deep Convolutional Neural Network Observations. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014. MICCAI 2014. Lecture Notes in Computer Science, vol 8673. Springer, Cham. `https://doi.org/10.1007/978-3-319-10404-1_65`

[13] Karri, C.; Cheikhrouhou, O.; Harbaoui, A.; Zaguia, A.; Hamam, H. Privacy Preserving Face Recognition in Cloud Robotics: A Comparative Study. Appl. Sci. 2021, 11, 6522. `https://doi.org/10.3390/app11146522`

[14] Karri, C. Secure robot face recognition in cloud environments. Multimed Tools Appl 80, 18611–18626 (2021). `https://doi.org/10.1007/s11042-020-10253-5`

[15] G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, "Densely Connected Convolutional Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269, doi: 10.1109/CVPR.2017.243. `https://doi.org/10.1109/CVPR.2017.243`

[16] Z. Abbas, H. Tayara and K. t. Chong, "SpineNet-6mA: A Novel Deep Learning Tool for Predicting DNA N6-Methyladenine Sites in Genomes, " in IEEE Access, vol. 8, pp. 201450-201457, 2020, doi:10.1109/ACCESS.2020.3036090. `https://doi.org/10.1109/ACCESS.2020.3036090`

[17] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017 Jan;35:18-31. doi: 10.1016/j.media.2016.05.004. Epub 2016 May 19. PMID: 27310171. `https://doi.org/10.1016/j.media.2016.05.004`

[18] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018 Apr;40(4):834-848. doi: 10.1109/TPAMI.2017.2699184. Epub 2017 Apr 27. PMID: 28463186. `https://doi.org/10.1109/tpami.2017.2699184`

[19] Gashi, D., Pereira, M., & Vterkovska, V. (2017). Multi-scale context aggregation by dilated convolutions machine learning-project. `https://doi.org/10.48550/arXiv.1511.07122`

[20] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network, " 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 6230-6239, doi: 10.1109/CVPR.2017.660. `https://doi.org/10.1109/CVPR.2017.660`

[21] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848. `https://doi.org/10.1109/tpami.2017.2699184`

[22] Hu, Y., Chen, Z., & Lin, W. (2018, July). RGB-D semantic segmentation: a review. In 2018 IEEE Inter national Conference on Multimedia & Expo Workshops (ICMEW) (pp. 1-6). IEEE. `https://doi.org/10.1109/ICMEW.2018.8551554`

[23] Lateef, F., & Ruichek, Y. (2019). Survey on semantic segmentation using deep learning techniques. Neurocomputing, 338, 321-348. `https://doi.org/10.1016/j.neucom.2019.02.003`

[24] Goceri, E., & Goceri, N. (2017). Deep learning in medical image analysis: recent advances and future trends. `https://doi.org/10.1146/annurev-bioeng-071516-044442`

[25] Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep learning techniques for medical image segmentation: achievements and challenges. Journal of digital imaging, 32(4), 582-596. `https://doi.org/10.1007/s10278-019-00227-x`

[26] Karimi, D., Dou, H., Warfield, S. K., & Gholipour, A. (2020). Deep learning with noisy labels: Exploring techniques and remedies in medical image analysis. Medical Image Analysis, 65, 101759. `https://doi.org/10.1016/j.media.2020.101759`

[27] Zhou, T., Ruan, S., & Canu, S. (2019). A review: Deep learning for medical image segmentation using multi-modality fusion. Array, 3, 100004 `https://doi.org/10.1016/j.array.2019.100004`

[28] Baheti, B., Innani, S., Gajre, S., & Talbar, S. (2020). Eff-unet: A novel architecture for semantic segmentation in unstructured environment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 358-359). `https://doi.org/10.1109/CVPRW50498.2020.00187`

[29] Zhang, P., Zhong, Y., Deng, Y., Tang, X., & Li, X. (2019). A survey on deep learning of small sample in biomedical image analysis. arXiv preprint arXiv:1908.00473. `https://doi.org/10.48550/arXiv.1908.00473`

[30] Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. Medical Image Analysis, 63, 101693. `https://doi.org/10.1016/j.media.2020.101693`

[31] Liu, C., Chen, L. C., Schroff, F., Adam, H., Hua, W., Yuille, A. L., & Fei-Fei, L. (2019). Autodeeplab: Hierarchical neural architecture search for semantic image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 82-92). `https://doi.org/10.48550/arXiv.1901.02985`

[32] Chen, L., Wu, Y., DSouza, A. M., Abidin, A. Z., Wism¨uller, A., & Xu, C. (2018, March). MRI tumour segmentation with densely connected 3D CNN. In Medical Imaging 2018: Image Processing (Vol. 10574, p.105741F). International Society for Optics and Photonics. `https://doi.org/10.48550/arXiv.1802.02427`

[33] Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In Deep learning in medical image analysis and multimodal learning for clinical decision support (pp. 3-11). Springer, Cham. `https://doi.org/10.1007/978-3-030-00889-5_1`

[34] P. Hu et al., "Automatic Pancreas Segmentation in CT Images With Distance- Based Saliency-Aware DenseASPP Network, " in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 5, pp. 1601-1611, May 2021, doi: 10.1109/JBHI.2020.302346. `https://doi.org/10.1109/JBHI.2020.3023462`

[35] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell. 2018 Apr;40(4):834-848. doi: 10.1109/TPAMI.2017.2699184. Epub 2017 Apr 27. PMID: 28463186. `https://doi.org/10.1109/tpami.2017.2699184`

[36] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., ... & Wu, J. (2020, May). Unet 3+: A full-scale connected unet for medical image segmentation. In ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp.1055-1059). IEEE. `https://doi.org/10.1109/ICASSP40776.2020.9053405`

[37] Roth, H., Oda, M., Shimizu, N., Oda, H., Hayashi, Y., Kitasaka, T., ... & Mori, K. (2018, March). Towards dense volumetric pancreas segmentation in CT using 3D fully convolutional networks. In Medical Imaging 2018: Image Processing (Vol. 10574, pp.59-64). SPIE. `https://doi.org/10.1117/12.2293499`

[38] Alhichri, H., Alswayed, A. S., Bazi, Y., Ammour, N., & Alajlan, N. A.(2021). Classification of remote sensing images using EfficientNet-B3 CNN model with attention. IEEE access, 9, 14078-14094. `https://doi.org/10.1109/ACCESS.2021.3051085`

[39] Roth, H. R., Lu, L., Farag, A., Shin, H. C., Liu, J., Turkbey, E. B.,& Summers, R. M. (2015, October). Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation. In Interna tional conference on medical image computing and computer-assisted intervention (pp. 556-564). Springer, Cham. `https://doi.org/10.48550/arXiv.1506.06448`

[40] Cai, J., Lu, L., Xie, Y., Xing, F., Yang, L. (2017). Pancreas Segmentation in MRI Using Graph-Based Decision Fusion on Convolutional Neural Networks. In: Descoteaux, M., Maier-Hein, L., Franz, A., Jannin, P., Collins, D., Duchesne, S. (eds) Medical Image Computing and Computer Assisted Intervention - MICCAI 2017. MICCAI 2017. Lecture Notes in Computer Science(), vol 10435. Springer, Cham. `https://doi.org/10.1007/978-3-319-66179-7_77`

[41] T. -Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, "Focal Loss for Dense Object Detection, " 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2999-3007, doi: 10.1109/ICCV.2017.324. `https://doi.org/10.1109/ICCV.2017.324`

[42] Tian, Junjiao et al. "Striking the Right Balance: Recall Loss for Semantic Segmentation." 2022 International Conference on Robotics and Automation (ICRA) (2021): 5063-5069. `https://doi.org/10.1109/icra46639.2022.9811702`

[43] Potnuru, M., Naick, B.S. Semantic Segmentation of MRI Images for Brain Tumour Detection with ShuffleNet-Based UNet. SN COMPUT. SCI. 4, 445 (2023). `https://doi.org/10.1007/s42979-023-01878-y`

[44] B. H. Menze et al., "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS), " in IEEE Transactions on Medical Imaging, vol. 34, no. 10, pp. 1993-2024, Oct. 2015, doi: 10.1109/TMI.2014.2377694. `https://doi.org/10.1109/TMI.2014.2377694`

[45] Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci Data. 2017 Sep 5;4:170117. doi: 10.1038/sdata.2017.117. PMID: 28872634; PMCID: PMC5685212. `https://doi.org/10.1038/sdata.2017.117`

[46] S. Bakas, M. Reyes, A. Jakab, S. Bauer, M. Rempfler, A. Crimi, et al., "Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Pre diction in the BRATS Challenge", arXiv preprint arXiv:1811.02629 (2018 `https://doi.org/10.48550/arXiv.1811.02629`