Deep Learning-based Brain Tumor Segmentation Using MRI

Fazil Kagdi
fazil.kagdi@ucf.edu

Abstract

In this study, we followed a deep learning-based segmentation model for brain tumor detection using a 3D U-Net architecture.[Wan+21] The robustness and generalization of the model were tested using a five-fold cross-validation strategy on different data splits. The model was trained in two phases, each lasting 15 epochs, with the second phase resuming from pre-saved checkpoints to continue learning and enhancing performance. Dice loss function was employed in optimizing segmentation training accuracy, and the performance of the model was calculated based on Dice similarity coefficients and Hausdorff Distance (HD95) metrics of three tumor subregions, which are Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT). For each fold, the model that had the highest Dice scores and lowest HD95 values was selected as optimal. The history of losses was also recorded and saved for analysis later. The outcome was ultimately reported as a mean and standard deviation of Dice scores and HD95 values of the five folds. This repetitive training approach allowed the model to refine its learning, improving segmentation accuracy in brain tumor detection.

1. Introduction

Brain tumor segmentation is critical in medical image assessment, offering diagnostic aid, treatment planning, and disease monitoring. Human radiologist segmentation following standard protocols is labor-intensive and disrupts observer variability. Advances in deep learning, particularly convolutional neural networks (CNNs), are a useful automation tool for the process of segmentation with benefits of increased efficiency and accuracy.

In this work, we employed a 3D U-Net model for segmenting brain tumors using five-fold cross-validation for model generalization. We utilized multi-modal MRI scans with labels pointing towards three major subregions of a tumor: Enhancing Tumor (ET), Tumor Core (TC), and Whole Tumor (WT). Model training went through two stages of 15 epochs each, where the second stage was an extension of pre-saved checkpoints for additional finetuning of acquired representations.

We utilized the Dice loss function for the optimization of the segmentation process and measured model performance in terms of Dice similarity coefficients and Hausdorff Distance (HD95) measures. The best model for each fold was determined according to highest Dice scores and lowest HD95 measures. History of loss was monitored and saved for subsequent inspection. This repeated training approach enabled the model to gradually enhance segmentation accuracy and resistance.

The outcomes demonstrate the effectiveness of the suggested approach for segmenting brain tumors and quantify its performance on different folds. This work aids in pushing forward automatic brain tumor segmentation towards more reliable and effective clinical application.

2. Implementation Details

The proposed brain tumor segmentation system was implemented in Python, using a suite of advanced libraries tailored for medical image analysis and deep learning. The heart of the project is based on the MONAI framework to create a 3D U-Net architecture model, with the PyTorch library being used as the main deep learning package for building the models, training, and validation. In addition, MedPy is used for the computation of evaluation metrics like the Dice similarity coefficient and the Hausdorff Distance (HD95), and scikit-learn offers data handling functions like K-Fold cross-validation. This robust set of tools ensures an efficient and reproducible pipeline for constructing high-quality segmentation models.

Data preprocessing is a critical component of the segmentation pipeline. The data set includes volumetric MRI scans in NIfTI format with each scan and an accompanying segmentation mask. The preprocessing pipeline begins by loading these files using the nibabel library. At every scan, image data is obtained and normalized through a MinMaxScaler that normalizes the voxel intensity values between 0 and 1. This is done per-image, such that the differences in intensities from various scans will not have a negative impact on training the model.

Tumor Region	Dice Score (Mean ± Std)	Hausdorff Distance (Mean ± Std)
Enhancing Tumor (ET)	0.4034 ± 0.0255	13.1251 ± 2.1798
Tumor Core (TC)	0.6216 ± 0.0198	7.4116 ± 0.6882
Whole Tumor (WT)	0.7174 ± 0.0189	6.9547 ± 0.7114

Table 1. Final segmentation performance for different tumor subregions using Dice Score and Hausdorff Distance.

For further preprocessing to remove input data, both masks and MRI images are cropped spatially. Cropping process is intentionally defined in such a way as to obtain the focal area of the brain by cropping unnecessary background regions. Specifically, cropping images and masks into sizes 128 × 128 × 128 limits the region of interest while reducing the computational load. After cropping, the information of the images gets reshaped and converted to PyTorch tensors. The tensor of the image is reorganised from the native form into a structure of (4, 128, 128, 128) where the first dimension represents the four modalities of the MRI. The segmentation masks, in contrast, are retained in the form of single-channel tensors in a structure of (128, 128, 128), where voxel value is equal to a class label.

The segmentation model is based on a 3D U-Net architecture specified in MONAI. The model employs an encoder-decoder architecture with skip connections that efficiently merge high-resolution features of the contracting path with the upsampled outputs of the expansive path. The network is configured with an input of 4 channels and outputs 4 channels, corresponding to the different segmentation classes: background, edema, non-enhancing tumor, and enhancing tumor. The model architecture is defined by a series of convolutional layers with feature maps being 16, 32, 64, 128, and 256 at increasing stages, with downsampling achieved through a stride of 2 in each step. The use of residual units throughout the network also boosts the flow of gradients and makes stable learning of features easier. [Bns]

For model training, we used a five-fold cross-validation to ensure our model learns well and generalizes on various data splits. In each fold, the network was trained for more than 15 epochs in two stages, with the second stage resuming training from pre-saved checkpoints in order to enhance the learned parameters further. The training process was facilitated through the Dice loss function, chosen for its application in dealing with class imbalance that is common to medical segmentation problems. Training loss was tracked on a per-epoch basis and stored for analysis and visualization in the future so that one can closely observe how the model converged over time.[Car+22]

3. Experimentation & Results

The experiments were conducted with a 5-fold cross-validation strategy, wherein for every fold there were two training phases of 15 epochs each. In the second phase, the model was trained further from the best checkpoint obtained during the first phase to fine-tune its parameters. The training loss curves of the training phases of all five folds are represented in Figures 1 and 2. For all folds, loss was decreasing persistently with improving network through epochs, indicating effective learning of intrinsic tumor characteristics. We could have trained the model for more epochs but due to limited time availibility we could only do 30 epochs.

For assessing segmentation robustness, two crucial metrics were applied: the Dice Similarity Coefficient (Dice) and Hausdorff Distance (HD95). The Dice coefficient is utilized to measure overlap between ground truth and predicted tumor regions, and it ranges from 0 to 1. The higher the Dice score, the better the segmentation because the prediction is closely approximating the ground truth. The Hausdorff Distance (HD95) is interested in boundary precision by assessing how much the ground truth and predicted boundaries deviate from each other. The smaller the value of HD95, the closer the predicted contours of the tumor are to the real tumor boundaries.

Table 1 shows the final averaged Dice and HD95 for all three tumor subregions—Whole Tumor (WT), Tumor Core (TC), and Enhancing Tumor (ET)—averaged over the five folds. The network worked best on the Whole Tumor region as indicated by a mean Dice of 0.7174 ± 0.0189 and a mean HD95 of 6.9547 ± 0.7114. This result shows that the network performed best in identifying the larger, more extensive tumor regions. The Tumor Core region followed with a mean Dice of 0.6216 ± 0.0198 and a mean HD95 of 7.4116 ± 0.6882, demonstrating a moderately high segmentation accuracy. On the contrary, the most difficult to segment was the Enhancing Tumor region, evidenced by a lesser Dice score of 0.4034 ± 0.0255 and greater HD95 of 13.1251 ± 2.1798. These are consistent with long-standing difficulties with segmenting more diminutive and less symmetrical enhancing regions of brain tumors. Table 2 shows avg Dice scores and Hausdorff Distance score for each individual fold across all 30 epochs.

Overall, the two-phase training method revealed that

Fold	Dice (ET)	Dice (TC)	Dice (WT)	HD95 (ET)	HD95 (TC)	HD95 (WT)
1/5 (Epoch 30)	0.4293	0.6537	0.7443	13.1991	7.6212	7.1047
2/5 (Epoch 30)	0.3938	0.6102	0.7160	13.9106	7.6898	7.0813
3/5 (Epoch 30)	0.4208	0.6172	0.7037	11.8094	8.0606	7.5557
4/5 (Epoch 30)	0.3659	0.6114	0.7244	18.7866	7.6895	6.8990
5/5 (Epoch 30)	0.3455	0.5946	0.6828	15.7585	8.3843	8.0824

Table 2. Dice Scores and HD95 values for 5 Folds (Epoch 30)

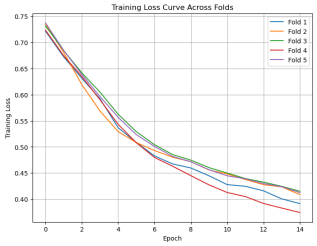

Figure 1. Dice loss curves for the first 15 epochs

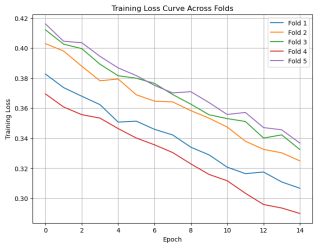

Figure 2. Dice loss curves for the 15-30 epochs

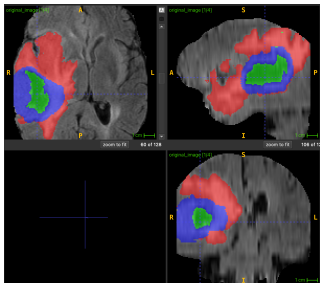

Figure 3. Ground Truth Visualization using ITK snap

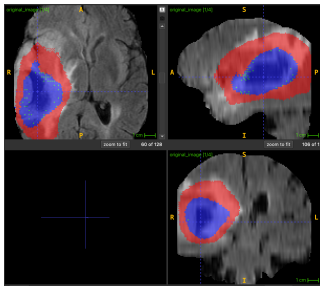

Figure 4. Predicted mask Visualization using ITK snap

continuous training from the optimal checkpoints led to incremental progress, as reflected by the constantly falling values of loss and steady improvement in Dice scores. Although the accuracy of segmentation for the greater tumor regions was improved, future work can work towards optimizing the model architecture or using high-level loss functions in order to boost performance on the smaller, refining tumor segments.

4. Visualization

Its visible from Fig. 3 and Fig. 4 the difference between the ground truth and predicted mask. We used ITK-snap software to visualize the masks on the actual image. To better visualize the brain tumor segmentation data, we began by randomly sampling the dataset, capturing both the original input image and its ground truth mask. Since the dataset consists of multi-modal MRI images with four channels, we checked to ensure the image was properly transformed before visualizing. We then used the model to create a predicted segmentation mask. In order to maintain uniformity in ITK-SNAP, we made sure the spatial resolution of ground truth and predicted masks were preserved meticulously when reformating the original image shape into (height, width, depth, modalities). For the purpose of visualization, we extracted a random slice on the depth axis and displayed it along with its corresponding ground truth and predicted masks. In addition, the processed data was stored under NIfTI format (.nii.gz), enabling usage within ITK-SNAP to continue further.

The outputs of the visualization indicated that the model was effective in segmenting the whole tumor (WT) region, wherein there was considerable correlation between predicted mask and the ground truth. However, it was still not easy to discern the enhancing tumor (ET) since the model struggled to adequately define its margins. This indicates that while the model captures the larger tumor region effectively, high-resolution segmentation of the enhancing tumor is not as accurate, possibly due to class imbalance or minor intensity variation. Overall, results are with the average Dice Score and Hausdorff Distance (HD) obtained, corroborating the consistency of the model's performance.

5. Conclusion

In conclusion, the brain tumor segmentation model was good, particularly when segmenting the whole tumor (WT) region, where the predicted masks closely resembled the ground truth. The detection of the enhancing tumor (ET) was somewhat more challenging, with the model not doing as well in defining its boundaries, possibly due to minor variations in the data or class imbalance. Even though there were some challenges, overall results were according to the average Dice score and Hausdorff Distance (HD), that is, the model is stable and provides helpful segmentation results.

While the model produced promising results, it could be improved further with additional training. Because of the long training times involved, however, this was not feasible within the time limit. Future research could, nonetheless, be dedicated to overcoming the model's shortcomings in ET detection, for instance, by means of data augmentation, class rebalancing, or more advanced architectures.

References

[Wan+21] Wenxuan Wang et al. "TransBTS: Multimodal Brain Tumor Segmentation Using Transformer". In: *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. Springer. 2021, pp. 109–119.

[Car+22] M. Jorge Cardoso et al. *MONAI: An open-source framework for deep learning in healthcare*. 2022. arXiv: 2211.02701 [cs.LG]. URL: https://arxiv.org/abs/2211.02701.

[Bns] Bnsreenu. *Python_for_microscopists/231_234_BraTa2020_Unet_segmentation at master · bnsreenu/python_for_microscopists*. URL: https://github.com/bnsreenu/python_for_microscopists/tree/master/231_234_BraTa2020_Unet_segmentation.