BRAIN TUMOR SEGMENTATION USING MODIFIED U-NET MODEL

Saumya B

DESE, Indian Institute of Science

ABSTRACT

Brain tumor segmentation is crucial for diagnosis and treatment planning, yet challenges such as class imbalance and boundary ambiguity limit existing methods. This study proposes a modified U-Net incorporating focal loss to address class imbalance and Monte Carlo Dropout for uncertainty estimation, providing interpretable predictions. The model was evaluated on MRI data, with experiments focusing on focal loss parameter tuning and studying the impact of three data augmentation techniques: Horizontal Flip, Rotation, and Scaling. The proposed model achieves comparable results to state-of-the-art models, making it suitable for clinical applications.

Index Terms— brain tumor segmentation, data augmentation, u-net, focal loss

1. INTRODUCTION

Brain tumors are among the most challenging medical conditions to diagnose and treat, often requiring precise identification of tumor boundaries for effective treatment planning. Magnetic Resonance Imaging (MRI) is a widely used imaging modality for detecting brain tumors, providing detailed anatomical information essential for accurate diagnosis. However, manual delineation of tumor regions by radiologists is time-consuming, prone to inter-observer variability, and difficult to scale in clinical settings. These challenges highlight the need for automated methods that can accurately segment brain tumors and provide interpretable predictions.

Deep learning-based approaches, particularly convolutional neural networks (CNNs), have shown significant promise in medical image segmentation tasks due to their ability to learn hierarchical spatial features. Among these, the U-Net architecture is widely regarded as the gold standard for biomedical image segmentation, owing to its encoder-decoder structure with skip connections that effectively capture fine-grained spatial details[4]. While U-Net models have demonstrated high segmentation accuracy, challenges persist in handling imbalanced datasets, segmenting tumor boundaries, and quantifying prediction uncertainty—critical factors in medical applications where decision confidence is paramount.

This study proposes a U-Net-based model designed to address these challenges by incorporating focal loss to handle class imbalance and Monte Carlo Dropout for uncertainty estimation. Focal loss helps the model focus on difficult-to-classify regions, such as tumor boundaries, by dynamically adjusting the loss contribution based on prediction difficulty. Monte Carlo Dropout enables the quantification of uncertainty in the model's predictions, providing insights into regions where the model is less confident, such as ambiguous or noisy tumor edges. To further enhance the model's performance, three data augmentation techniques—Horizontal Flip, Rotation, and Scaling—were evaluated to improve generalization and reduce uncertainty. The effects of these augmentations were analyzed with respect to segmentation accuracy, boundary uncertainty, and overall model confidence.

This paper provides a comprehensive analysis of the proposed model's performance and highlights the role of uncertainty estimation and data augmentation in improving segmentation accuracy and confidence. The findings have significant implications for developing robust and interpretable deep learning models for brain tumor segmentation in clinical applications.

2. DATASET

For this study, a brain tumor dataset containing 3064 T1-weighted contrast-enhanced MRI images was used. The data was collected from Nanfang Hospital and Tianjing Medical University, China, from 2005 to 2010, by Jun Cheng, who originally used the dataset in his study [1] and [2], and had uploaded the entire dataset with its metadata to Figshare. The dataset consists of T1-weighted contrast-enhanced MRI scans from 233 patients, and has three kinds of brain tumors - 708 cases of Meningiomas, 1426 gliomas and 930 cases of pituitary tumors. The tumor border was manually delineated by three experienced radiologists. A copy of this dataset was taken from Kaggle [3], where the scans and corresponding binary masks were uploaded as 256x256 pixel images. A summary of the details of the dataset, along with sample images, are given in Table 1.

3. EXPERIMENTAL SETUP

The objectives of this study were threefold:

Tumor type	Dimensions	No. of images
Meningioma	256x256	708
Glioma	256x256	1426
Pituitary	256x256	930

Table 1. Dataset details

- To demonstrate the effectiveness of the proposed variation of the U-Net model in segmenting tumor regions from MRI scans.
- To analyze the impact of varying focal loss parameters on model performance.
- To evaluate the effect of three different data augmentation techniques on model performance.

To achieve these objectives, the study was conducted in two phases, using the same U-Net model:

1. *Focal Loss Parameter Tuning:*
   In the initial set of experiments, the goal was to identify the optimal parameters for focal loss that yielded the best segmentation performance. These experiments were conducted using the original dataset without applying any data augmentation techniques.

2. *Data Augmentation Analysis:*
   After determining the best-performing focal loss parameters, they were fixed for the subsequent experiments. Three different data augmentation techniques were then applied individually to the dataset, and the model's performance was evaluated for each augmentation technique.

   This systematic approach ensured a comprehensive analysis of both focal loss parameter tuning and the impact of data augmentation on the model's segmentation performance.

4. HARDWARE

All training was performed using Google Colab's free TPU, the TPUv2-8. A tensorflow library was used to enable the model to utilise the TPU during training.

5. PRE-PROCESSING

The preprocessing steps began by loading both the images and masks in grayscale format, hence reducing the complexity by using fewer channels compared to color images. The images and masks were then resized to a standardized size (256x256) to ensure uniformity across the dataset. Following this, the pixel values were normalized by dividing each pixel by 255, which not only facilitates faster convergence [6] during training but also makes the computations more efficient and less memory-intensive. Since the model is designed to work with binary masks, an important step was performed to verify that all mask pixel values were either 0 or 1. Any pixel with a value between 0 and 1 (non-binary) was converted to 0 if it was greater than 0 but less than 1. Finally, the dataset, which contained 3064 samples, was randomly split into training, validation, and test sets, with 1838 samples assigned to the training set, and 613 samples each assigned to the validation and test sets.

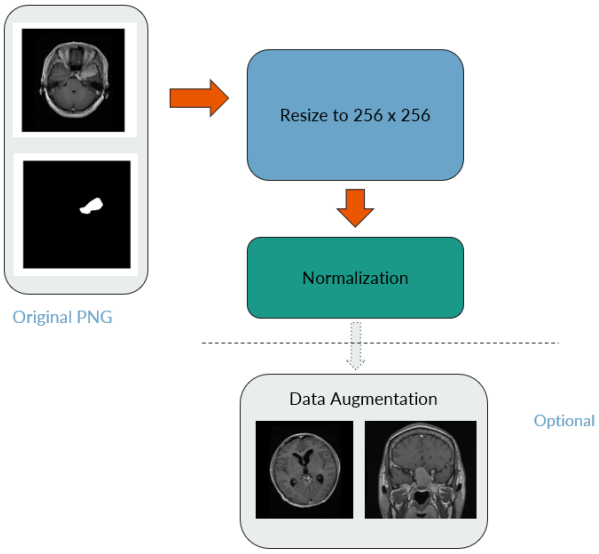

Fig. 1. Pre-processing workflow

6. MODEL ARCHITECTURE

6.1. Modified U-Net architecture

The proposed model is based on the U-Net architecture, originally designed for biomedical image segmentation tasks. The U-Net architecture was selected due to its effectiveness in biomedical image segmentation, particularly in tasks with limited labeled data and the need for precise spatial localization, which is critical for tumor segmentation [5]. It employs an encoder-decoder structure with skip connections to preserve spatial information. The encoder extracts hierarchical features through downsampling, while the decoder reconstructs the segmentation mask by upsampling the feature maps.

The encoder consists of four downsampling blocks, each with two convolutional layers (with padding = "same" and kernel size of 3x3) to balance the need for fine-grained segmentation accuracy and model generalization [10]. This is followed by a max pooling layer with a pool size of 2x2. The convolutional layers use ReLU activation and He normal initialization. ReLU was chosen because it introduces

non-linearity, avoids vanishing gradient issues, and promotes faster convergence, while He initialization complements it by maintaining variance throughout the network, ensuring stable and efficient training. After the pooling layers, a dropout layer with a rate of 0.3 is applied to regularize the network and prevent overfitting. The number of filters doubles in each block, starting from 64 in the first block and increasing to 1024 in the bottleneck layer.

The decoder mirrors the encoder with four upsampling blocks. Each block begins with a transposed convolution layer to upsample the feature maps, followed by skip connections that concatenate feature maps from the corresponding encoder block, preserving spatial details lost during downsampling. These are followed by two convolutional layers with kernel size 3x3 and with ReLU activation and He initialization. A dropout layer with the same rate of 0.3 is applied after the concatenation step to maintain regularization and support uncertainty estimation. At the bottleneck, the deepest part of the U-Net, two convolutional layers with 1024 filters capture high-level feature representations essential for accurate segmentation. The output layer is a 1×1 convolution with a sigmoid activation function, producing a binary segmentation map to identify tumor regions.

To enable uncertainty estimation, the dropout layers are configured as Monte Carlo Dropout during inference. This allows for multiple stochastic forward passes, providing spatial uncertainty maps and evaluating the model's predictive confidence. This architecture combines the power of U-Net's hierarchical feature extraction with the added ability to quantify uncertainty, ensuring robust and interpretable predictions.

Layer	Details	
Input	Image input with shape `(height, width, channels)`	
Encoder (Contracting Path)	1. **Conv2D Layers**: 64, 128, 256, 512 filters, each followed by ReLU activation. Kernel size = 3 x 3	x 4
	2. **MaxPooling**: 2x2 operation for downsampling.	
	3. **Dropout**: Regularization after pooling layers. Dropout = 0.3	
Bottleneck	1. **Conv2D Layer**: 1024 filters with ReLU activation.	
	2. **Dropout**: Applied for additional regularization.	
Decoder (Expansive Path)	1. **Transposed Convolutions**: 2x2 upsampling.	x 4
	2. **Skip Connections**: Merge with corresponding encoder features.	
	3. **Conv2D Layers**: 512, 256, 128 filters, each followed by ReLU activation.	
Output	1. **Conv2D Layer**: Final 1x1 convolution with sigmoid activation for binary segmentation.	

Fig. 2. Model summary

6.2. Choosing focal loss as loss function

Focal loss was selected as the loss function for this model due to its effectiveness in addressing class imbalance, which is a common issue in segmentation tasks. In this case, the segmentation masks contain significantly fewer foreground pixels (tumor regions) compared to the background pixels. Standard loss functions, such as cross-entropy loss, may struggle in such scenarios because the loss tends to be dominated by the majority class (background). Focal Loss mitigates this issue by dynamically scaling the contribution of each pixel's loss based on its classification difficulty. Hard-to-classify examples (e.g., noisy textures, small tumor regions, or partial objects) are given higher weights, while easy-to-classify examples (e.g., background pixels) are down-weighted, allowing the model to focus on learning challenging patterns.

Focal Loss is an extension of Cross-Entropy Loss, formulated as:

$$F_L(p_t) = -\alpha\,(1-p_t)^{\gamma}\log(p_t)$$

where:

p_t is the model's predicted probability for the true class

α is a weighting factor for balancing the importance of different classes

$(1-p_t)$ is the modulating factor, which gives more weight to misclassified or hard examples

γ is the focusing parameter that adjusts how much the loss focuses on hard examples.

The parameters α and γ play crucial roles in the behavior of Focal Loss. α is the weighting factor which controls the balance between the foreground and background classes. For instance, setting a higher α for the tumor (foreground) class ensures that the model assigns greater importance to learning from tumor pixels, which are typically underrepresented. γ is the focusing parameter and determines the degree of focus on hard examples. When γ=0, Focal Loss is equivalent to standard Cross-Entropy Loss. Increasing γ(e.g., 2.0) amplifies the contribution of harder examples while reducing the weight of well-classified ones. Higher values of γ are particularly useful in datasets with severe class imbalance.

6.3. Hyperparameters

The Adam optimizer (Adaptive Momentum Estimator) was chosen for training the model due to its effectiveness in biomedical image segmentation tasks [8] and its ability to balance computational efficiency, minimal memory requirements, and ease of implementation [9]. Unlike traditional stochastic gradient descent, Adam adapts the learning rate dynamically for each parameter based on past gradients, making it well-suited for complex models like U-Net. The learning rate is a critical hyperparameter that directly affects training stability and speed. A learning rate that is too low can result in slow convergence, while one that is too high may cause the model to miss the optimal solution. For this model, an initial learning rate of 1e-4 was selected, which provided a good balance between convergence speed and stability.

The rest of the hyperparameters chosen are given in detail in the Table 2.

Category	Hyperparameter	Value
Model architecture	Input Shape	(256,256,1)
	No. of filters	64-> 128 -> 256 -> 512 -> 1024
	Kernel Size	(3,3)
	Kernel Size	(3,3)
	Activation function	ReLU
	Kernel Initializer	He Normal
	Dropout Rate	0.3
	Final Activation	Sigmoid
Loss and Metrics	Loss function	Focal Loss
	Optimizer	Adam
	Training metrics	Accuracy
Training Parameters	Batch Size	8
	Learning Rate	1e-4
	No. of epochs	200
Uncertainty Analysis	MC Dropout Samples	50

Table 2. List of hyperparameters chosen

6.4. Metrics

The evaluation focused on two major quality factors, segmentation quality and prediction reliability. Segmentation quality was assessed using standard metrics such as Dice Coefficient, Intersection Over Union (IoU), precision and recall. These metrics provide a quantitative measure of the overlap and accuracy of the predicted segmentation masks compared to the ground truth.

To evaluate the reliability of the predictions, a detailed uncertainty and boundary analysis was performed. Additionally, these metrics were also helpful in understanding the effects of data augmentation on the model performance. This was done with the help of overall uncertainty metrics and boundary-specific metrics. Overall uncertainty metrics included measures such as mean uncertainty (average of uncertainty calculated over all the voxels of the image), max uncertainty (maximum uncertainty observed across the entire image) and visualizing the uncertainty distribution. The boundary-specific metrics considered are mean boundary uncertainty (average of uncertainties calculated at the tumor boundaries), std. boundary uncertainty (standard deviation of boundary uncertainty values), max boundary uncertainty (maximum uncertainty observed at the boundary) and total boundary length.

6.5. Procedure followed for uncertainty and boundary analysis

For uncertainty analysis, Monte Carlo Dropout was employed to estimate the model's confidence in its predictions. During inference, dropout layers were replaced with Monte Carlo Dropout layers and were kept active, allowing for multiple stochastic forward passes of the model. By aggregating the predictions from these passes, the mean prediction and uncertainty map were calculated, with higher uncertainty values indicating regions where the model was less confident, such as tumor boundaries or noisy areas. The uncertainty distribution across the image was also visualized.

Additionally, boundary analysis focused on the tumor's edges, which are often challenging to segment accurately. Boundary-specific uncertainty metrics were computed, including mean boundary uncertainty, standard deviation of boundary uncertainty, maximum boundary uncertainty, and total boundary length. These metrics provided valuable insights into the model's performance at the tumor boundaries, highlighting areas of high uncertainty and helping to assess the model's ability to precisely segment these critical regions.

7. TRAINING

The model was trained for 200 epochs using Adam optimizer with a learning rate of 1e-04. Validation was done using the standard training-validation-test split of 60%-20%-20%. During training, accuracy was used as the primary evaluation metric, which provides a general measure of how well the model's predictions align with the ground truth. However, for a more comprehensive evaluation of segmentation performance, additional metrics were computed on the test dataset, including the Dice coefficient, Intersection over Union (IoU), precision, and recall. These metrics are particularly relevant for assessing the quality of segmentation masks, especially in imbalanced datasets where pixel-level accuracy may not fully reflect model performance.

8. EXPERIMENTS CONDUCTED

Phase 1: Focal Loss Parameter Tuning

In this phase of the experiments, the α and γ parameters for focal loss were varied to evaluate their impact on segmentation performance. These experiments were conducted using the original dataset, without applying any data augmentation techniques.

Two sets of parameter combinations were tested:

(i) α=0.25, γ=2.0

(ii) α=2.0, γ=0.75

For the first set of parameters (α=0.25, γ=2.0), the lower α places more weight on the background class, while the higher γ emphasizes hard-to-classify pixels, such as tumor boundaries or ambiguous regions. In contrast, the second set of parameters (α=2.0, γ=0.75) increases the importance of the minority class (tumor) with a higher α, while the smaller γ of 0.75 reduces the focus on difficult examples and instead prioritizes the overall distribution, giving weightage to easier-to-classify regions too.

Phase 2: Data Augmentation Evaluation

In this phase, three different data augmentation techniques were evaluated on the proposed model, while keeping the focal loss parameters constant at α=0.25 and γ=2.0. The techniques evaluated are as follows:

- Horizontal Flip
- Rotation
- Scaling

The details of these augmentation techniques are provided in Table 3. It is important to note that, since this is a segmentation task, the augmentation transformations were applied to both the image and its corresponding mask to ensure consistency between the input and target.

Technique	% of training dataset applied on	Parameters
Horizontal Flip	50%	none
Rotation	50 %	Angle: $\pm 15^\circ$
Random Scaling	50 %	Range: 0.8 - 1.2

Table 3. Data Augmentation experiment details

9. RESULTS AND DISCUSSION

This section presents the results of the experiments conducted to evaluate the segmentation performance of the proposed model, the impact of focal loss parameter tuning, the effect of data augmentation techniques, and insights from uncertainty and boundary analysis. The proposed model was evaluated using multiple segmentation metrics, including accuracy, loss, Dice coefficient, Intersection over Union (IoU), precision, and recall. Experiments were conducted using both the original dataset and augmented datasets to assess the impact of data augmentation techniques on performance. The model showed competitive results across all metrics, with Horizontal Flip emerging as the most effective data augmentation technique, achieving the highest Dice coefficient and IoU scores. Rotation also contributed positively, whereas Scaling had minimal impact on model performance.

9.1. Impact of focal loss parameters

The results obtained in the first set of experiments of tuning the focal loss are presented in Table 4. In the first case, the configuration prioritized hard-to-classify examples, particularly tumor boundaries, resulting in better performance. In the second case, the configuration emphasized the minor class (tumor regions) more but focused less on boundary details, leading to weaker performance in challenging regions.

The experiments demonstrated that the choice of focal loss parameters play a significant role in balancing foreground and background segmentation accuracy.

9.2. Impact of Data Augmentation

Three data augmentation techniques - Horizontal flip, Rotation and Scaling - were experimented with to evaluate their effect on model generalization and robustness. The results of the experiments are summarized in the Table 5.

Horizontal flip consistently improved performance across all metrics, making it the most effective augmentation technique. Rotation improved the Dice coefficient and IoU, indicating its effectiveness. Scaling, however, showed negligible or no improvement in segmentation performance, suggesting that size variations are less significant for this dataset. Visualization of ground truth and predictions further supported these findings (refer figure 3), with horizontal flip and rotation demonstrating clearer and more accurate segmentation results compared to the original dataset or when using Scaling.

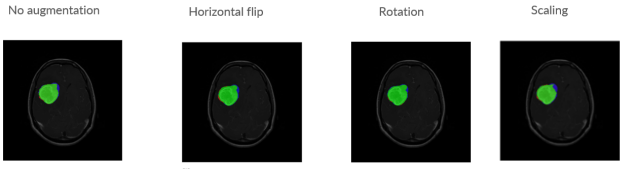

Fig. 3. Ground Truth v/s Model Prediction

9.3. Inference drawn from Training graphs

Figure 4 shows the training graphs of the different augmentation techniques and the baseline model. As seen from the graphs, horizontal flip and rotation helps in achieving a lower final loss when compared to the baseline model. Horizontal flip has a more positive effect and can be seen to help in achieving faster convergence and lower final loss. Rotation performed better than baseline but not as good as horizontal flip. On the contrary, the training graph for scaling technique shows a higher final loss when compared to baseline and shows a lot of fluctuations, hence suggesting that it is not very suitable for this task.

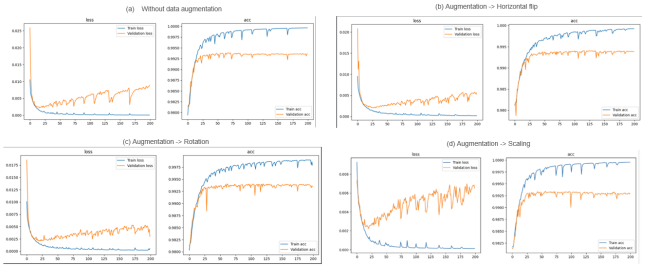

Fig. 4. Training graphs

9.4. Uncertainty and Boundary analysis

To analyze model confidence, Monte Carlo Dropout was implemented, enabling multiple stochastic forward passes during inference. Mean uncertainty and maximum uncertainty were computed, with higher uncertainty values observed at tumor boundaries and ambiguous regions (refer to Table 6). Uncertainty distribution visualizations revealed that augmentation techniques reduced overall uncertainty, particularly at the boundaries. Boundary-specific metrics—including mean boundary uncertainty and maximum boundary uncertainty—highlighted that applying augmentations like Horizontal Flip and Rotation decreased uncertainty at tumor

Loss function	Parameters	Accuracy	Loss	Precision	Recall	IoU	Dice Co-efficient
Focal Loss	α=0.25 γ=2.0	0.9941	0.0082	0.9014	0.7681	0.7082	0.7867
Focal Loss	α=2.0 γ=0.75	0.9939	0.0154	0.8778	0.7789	0.7004	0.7839

Table 4. Results for loss parameter experiments

Augmentation type	Accuracy	Loss	Precision	Recall	IoU	Dice Co-efficient
None	0.9941	0.0082	0.9014	0.7681	0.7082	0.7867
Horizontal Flip	0.9942	0.0053	0.9001	0.7779	0.7152	0.8041
Rotation	0.9940	0.0029	0.8774	0.7892	0.7090	0.7955
Random Scaling	0.9934	0.0064	0.9097	0.7106	0.6643	0.7486

Table 5. Results for Data augmentation experiments

boundaries. This indicates that these techniques not only improve segmentation accuracy but also enhance the model's confidence in its predictions. Hence, it can be concluded that performing augmentations helps improve model confidence and reliability.

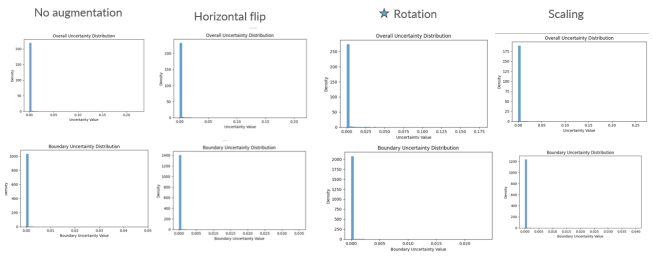

Fig. 5. Uncertainty distribution for different augmentation

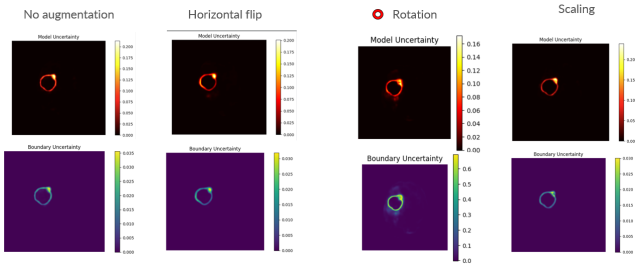

Fig. 6. Uncertainty visualizations for different augmentation

Augmentation	Mean Boundary Uncertainty	Std Boundary Uncertainty	Max Boundary Uncertainty	Mean Total Uncertainty
None	0.0116	0.0374	0.2672	0.0032
Horizontal Flip	0.0107	0.0367	0.2876	0.0035
Rotation	0.0040	0.0190	0.1995	0.0019
Random Scaling	0.0060	0.0278	0.2877	0.0025

Table 6. Results from Uncertainty analysis

9.5. Comparison with state-of-the-art

The proposed model achieved comparable performance with state-of-the-art methods for brain tumor segmentation. A brief comparison with state-of-the-art models is shown in Table 7.

Model	Precision	Recall	IoU	Dice co-efficient
[Ours]	0.9001	0.7779	0.7152	0.8041
Arafat et al. [2023]	).82	0.74	0.68	0.94
Gupta wt al. [2021]	0.89	0.91	-	0.90

Table 7. Comparison with SOTA

10. CONCLUSION AND FUTURE WORK

The aim of this study was to demonstrate the effectiveness of the proposed model for brain tumor segmentation and to explore the impacts of loss parameters and data augmentation on model performance and confidence. The proposed model demonstrated strong segmentation performance, particularly with data augmentation. Changes in focal loss parameters significantly impacted the model behavior, with better results obtained when parameters were tuned to give more weightage on minority class (tumor region) and hard-to-classify examples. Among the augmentation techniques, Horizontal flip was the most effective, followed by Rotation, while Scaling showed minimal improvement. Uncertainty analysis revealed that augmentation reduces boundary uncertainty, with Horizontal flip giving the most positive effect. Hence it can be concluded that augmentation not only helps to improve segmentation performance, it also increases model's confidence in its predictions. Overall, the results suggest that the proposed model is robust, interpretable and capable of comparable performance to SOTA methods.

In the future, this work can be extended to include evaluation of complex data augmentation techniques like elastic deformations, modality-specific transformations, etc. Artificially synthesized data using generative models like GANs can also be experimented with to evaluate its effect on model performance when used as an augmentation technique. Fur-

thermore, classification of tumor type can also be integrated into the project pipeline such that when given an image, the model segments it and classifies the tumor into specific categories.

11. REFERENCES

[1] Cheng, Jun, et al. "Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition." *PloS one* 10.10 (2015)

[2] Cheng, Jun, et al. "Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation." *PloS one* 11.6 (2016)

[3] The Kaggle dataset link

[4] Ronneberger, O., Fischer, P., Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W., Frangi, A. (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI* 2015

[5] Du, Getao, Xu Cao, Jimin Liang, Xueli Chen, and Yonghua Zhan. "Medical Image Segmentation based on U-Net: A Review." *Journal of Imaging Science Technology* 64, no. 2 (2020).

[6] Chen, Xinyi, Xiang Liu, Yuke Wu, Zhenglei Wang, and Shuo Hong Wang. "Research related to the diagnosis of prostate cancer based on machine learning medical images: A review." *International journal of medical informatics* 181 (2024): 105279.

[7] Ross, T-YLPG, and G. K. H. P. Dollár. "Focal loss for dense object detection." *IEEE conference on computer vision and pattern recognition*, pp. 2980-2988. 2017.

[8] Yaqub, Muhammad, Jinchao Feng, M. Sultan Zia, Kaleem Arshid, Kebin Jia, Zaka Ur Rehman, and Atif Mehmood. "State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images." *Brain Sciences* 10, no. 7 (2020): 427.

[9] Kingma, Diederik P. "Adam: A method for stochastic optimization." *arXiv preprint* arXiv:1412.6980 (2014).

[10] Pereira, Sérgio, Adriano Pinto, Victor Alves, and Carlos A. Silva. "Brain tumor segmentation using convolutional neural networks in MRI images." *IEEE transactions on medical imaging* 35, no. 5 (2016): 1240-1251.