Brain Tumor Segmentation from MRI Scans

Yahya Massoud
EECS
University of Ottawa
Ottawa, Canada
ymass049@uottawa.ca

Hamid Reza Aghamiri
EECS
Carleton University
Ottawa, Canada
Hamidrezaaghamiri@cmail.carleton.ca

Mahdi Falahatpisheh
EECS
Carleton University
Ottawa, Canada
mahdi.falahat.pish@gmail.com

***Abstract*—This paper presents an exploration of tumor segmentation from MRI scans, focusing on the efficacy of thresholding and region growing methods, using manual and automatic seed selection. We employed two diverse datasets from the Brain Tumor Segmentation (BraTS) challenge to evaluate our approaches. We meticulously compare the performance of our methods against eight established metrics, encompassing both quantitative and qualitative dimensions. Quantitative analysis involves metrics such as Jaccard similarity and dice coefficient, accuracy, specificity, and runtime, providing a multi-faceted view of the segmentation performance. Qualitative analysis is conducted through visual inspection, ensuring that the segmented tumors align closely with the groundtruth labels. The results demonstrate a superiority of the region growing methods, especially with automatic seed selection, in accurately delineating tumor boundaries. This study contributes significantly to the field of medical image processing, offering insights that could be useful for tumor segmentation.**

***Index Terms*—Biomedical Image Processing, Tumor Segmentation**

I. Introduction

A brain tumor is a mass of tissue that propagates out of control of the normal forces that regulate growth inside the brain. Brain tumors are one of the foremost reasons for the rise in mortality among children and adults. They can be classified as primary or secondary brain tumors depending on the point of origin. Primary originate in brain cells. Secondary originates elsewhere and then spreads to the brain. They can also be classified based on their severity. Benign brain tumors are those that grow slowly and do not metastasize or spread to other body organs and often can be removed and hence are less destructive or curable. They can still cause problems since they can grow big and press on sensitive areas of the brain (the so-called mass effect). Depending on their location, they can be life-threatening. Malignant brain tumors are those with cancerous cells. The rate of growth is fast ranging from months to a few years. Unlike other malignancies, malignant brain tumors rarely spread to other body parts due to the tight junction in the brain and spinal cord.

CT (computed tomography) and MRI (magnetic resonance imaging) are the most common brain tumor imaging technologies. MRI is accurate, non-invasive, and has little risk of radiation. It is capable of differentiating soft tissue with high resolution and is more sensitive in detecting and visualizing subtle changes in tissue density and the physiological alternations associated with the tumor. Medical imaging technology has the potential to improve both the diagnosis and prognosis of brain tumors. Diagnosis is the identification of a disease or condition by its signs, symptoms, and the results of various diagnostic procedures, while prognosis is the prediction or forecast of the likely course or outcome of a disease or ailment.

The modalities of MRI include T1, which are T1-weighted scans that focus on the anatomy, differentiating between grey and white matter. T1c, which are T1-weighted but contrast-enhanced using Gadolinium, highlighting pathological regions like tumors. Gadolinium is a contrast agent used in MRI to enhance the visibility of certain structures. It helps detect abnormalities like tumors or inflammation. T2 which are T2-weighted scans are sensitive to fluid differences, and commonly used for lesion detection. In the context of brain scans, lesions may refer to tumors, cysts, or other abnormal growths. Finally, FLAIR, which is Fluid-Attenuated Inversion Recovery, is another T2-weighted scan that suppresses the fluid signal to see other structures more clearly. Flair performs better in defining the actual volume of the neoplasm. Effective in the detection of Gliomas (such as glioblastoma, astrocytomas, oligodendrogliomas, and ependymomas), which make up 81% of malignant brain tumors in adults.

II. Dataset

The Brain Tumor Segmentation (BraTS) dataset is a widely adopted benchmark for evaluating state-of-the-art methods for segmenting brain tumors (Gliomas) in MRI scans. The dataset contains pre-operative 3D MRI scans with four modalities. It also includes ground-truth human-labeled tumor segmentation masks for three different classes: 1) the entire visible tumor region, 2) the tumor core excluding any swelling, and 3) the enhancing (actively growing) part of the tumor. We will focus on segmenting the entire visible tumor region from MRI scans. Such ground-truth masks are used for model quantitative and qualitative analysis.

MRI data contains four different modalities, namely T1, T1c, T2, and FLAIR. We chose the FLAIR modality in our work as it suppressed fluid for clearer structure visualization. This characteristic of the FLAIR modality is clearly apparent in Figure 1. To construct our datasets, We extract 2D grayscale images of size 240×240 from the 3D MRI scans provided in the BraTS dataset. Then, we produce two datasets to evaluate our segmentation methods: (1) Set 1 containing 12 data points, and (2) Set 2 containing 15 data points. Each data point

is a grayscale brain scan (FLAIR) with its corresponding ground truth segmentation mask which denotes the whole tumor region. Both datasets don't have overlapping data points, MRI scans are unique to each set. Data points in each set were randomly selected from the 1025 samples in the original dataset.

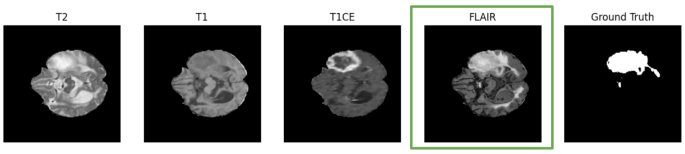

Fig. 1. Example image from the BraTS dataset. Highlited is the FLAIR modality which provide clear structure of the tumor region. On the rightside is a groundtruth mask for a tumor region.

III. Methods

A. Thresholding-based Segmentation

To achieve better results, several image pre-processing steps were applied to the brain images before thresholding.

1) Background Removal: The presence of background pixels (represented by 0 values) can significantly impact the thresholding process. Therefore, the background was removed prior to further processing. This focused the analysis solely on the relevant brain tissue, minimizing the influence of extraneous information.

2) Gaussian Filtering: A Gaussian filter with a kernel size of (5, 5) and sigma of 0 was applied to the brain image. By doing this, the noise in the image is reduced, and any unevenness is smoothed out, thereby simplifying the intensity distribution and facilitating more accurate thresholding. Edge preservation: The chosen kernel size preserved the important edges and boundaries of the brain structures, ensuring they wouldn't be blurred or lost during the filtering process. These preprocessing steps prepared the brain images for optimal performance in the subsequent thresholding stage. In Figure 2, you can see the process of using thresholding.

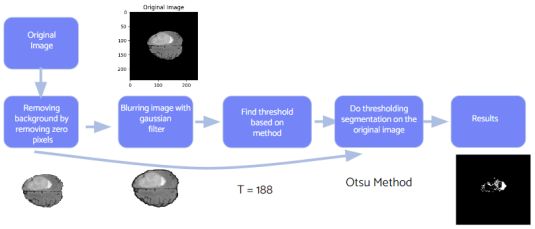

Fig. 2. The process of doing Otsu theresholding method with pre-processing.

3) Thresholding Method: Image thresholding simplifies image analysis by dividing it into distinct regions based on pixel intensity. By setting a threshold value, pixels become classified as belonging to either foreground or background, facilitating the extraction of features and aiding tasks like object detection, segmentation, and image enhancement. At its core, thresholding transforms gray-scale images into binary representations, reducing each pixel to either black or white. This simplification streamlines analysis, enabling efficient processing and object identification within the image. Furthermore, this technique serves as the foundation for numerous computer vision applications, including edge detection and pattern recognition, by facilitating the extraction of key features and maximizing the information retained. Image thresholding simplifies image analysis by dividing it into distinct regions based on pixel intensity. By setting a threshold value, pixels become classified as belonging to either foreground or background, facilitating the extraction of features and aiding tasks like object detection, segmentation, and image enhancement. At its core, thresholding transforms gray-scale images into binary representations, reducing each pixel to either black or white. This simplification streamlines analysis, enabling efficient processing and object identification within the image. Furthermore, this technique serves as the foundation for numerous computer vision applications, including edge detection and pattern recognition, by facilitating the extraction of key features and maximizing the information retained.First we define a threshold value (T): This value determines which pixel intensity values are considered "foreground" (objects) and which ones are considered "background." Then compare each pixel's intensity (I) to the threshold: **[?]**

$$L(x,y) = \begin{cases} 1 & \text{if } I(x,y) \geq T, \\ 0 & \text{if } I(x,y) < T. \end{cases}$$

In this report we used different thereshold value to compare our work with the proposed method:

4) Method 1 - Mean Threshold: In mean thresholding, the threshold value is chosen as the mean intensity of all pixels in the image:

$$T = \frac{1}{MN} \sum_{x=1}^{M} \sum_{y=1}^{N} I(x,y)$$

where:

- M and N are the width and height of the image, respectively.
- $\sum$ is the summation operator over all pixels in the image.

5) Method 2 - Standard Deviation Threshold: This technique builds upon basic thresholding by considering the standard deviation (σ) of pixel intensities. The standard deviation (σ) is calculated using the formula:

$$\sigma = \sqrt{\frac{\sum_{x=1}^{M} \sum_{y=1}^{N} (I(x,y) - \mu)^2}{N}}$$

$I(x,y)$ represents the intensity value of the pixel at location (x,y) in the image, μ is the mean intensity of all pixels in the image. N is the total number of pixels in the image, and $\sum$ denotes the summation over all pixels in the image.

6) Method 3 - User-defined Threshold: In user-defined thresholding, the user manually selects the threshold value based on their visual interpretation of the image. This can be done by observing the image histogram and choosing a value

that separates the foreground and background peaks. We added this method to see if a user chooses a threshold by using three images, what will be the performance for the entire dataset. Will all images follow the same structure? [**?**]

7) Method 4 - Otsu's Method Threshold: Otsu's method is an iterative algorithm that automatically selects the threshold value that minimizes the intra-class variance of the foreground and background classes.Otsu's method for thresholding is defined by the formula:

$$\text{Minimize } \omega_0(T)\sigma_0^2(T) + \omega_1(T)\sigma_1^2(T)$$

where $\omega_0(T)$ and $\omega_1(T)$ are the probabilities of the foreground and background classes, and $\sigma_0^2(T)$ and $\sigma_1^2(T)$ are the variances of the foreground and background classes, respectively. The goal is to find the threshold value T that minimizes the intra-class variance, which is a measure of the spread of pixel intensities within each class.Otsu's method can be implemented using various optimization algorithms, such as dynamic programming or iterative search. In Figure 2 you can see the result of using mentioned theresholding method on the flair image. [**?**]

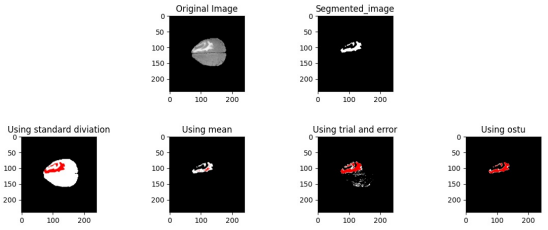

Fig. 3. The result of using theresholding technique on the flair part of the barin image.

B. Region Growing with Manual Seed Selection

In the region-growing segmentation method, the quality of the output image is significantly influenced by two pivotal parameters: the careful selection of an appropriate seed point and the judicious determination of a suitable threshold. The effectiveness of region growing hinges on these key choices. In the following discussion, we will delve into two methods employed to ascertain a well-suited manual seed point and an optimal threshold for the region-growing process. These methods contribute to refining the segmentation approach, ultimately enhancing the overall quality of the output image.

1) Appropriate Manual Seed Selection: Choosing the right seed point is paramount, as it serves as the initial reference from which the segmentation process emanates. An optimal seed point should be representative of the targeted region, ensuring accurate and meaningful segmentation. Conversely, an ill-chosen seed point may lead to inaccuracies or incomplete segmentation.

Our method operates on the principle of region growing, wherein an initial seed point is selected, and the program subsequently conducts the region growing process based on that seed point. Following this, we assess the segmented output

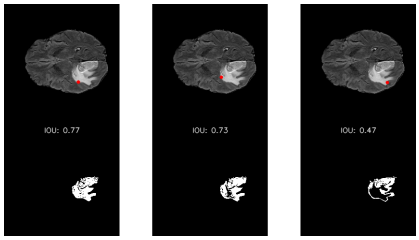

Fig. 4. Tumor segmentation in mode of three different seed points

image obtained from the region growing process against our ground truth, calculating the seed point's Intersection over Union (IoU) score. If the score meets our criteria, the program records the seed point's location in the Evaluation JASON file. Ideally, our method aims to identify the initial seed point yielding an IoU score exceeding 0.8. However, if the first point falls below our expectations, the program allows the user to choose another seed point, continuing the process until a suitable seed point is selected or until the user exit the seed point selection process. Therefore, when none of the selected points surpass this criterion, the algorithm will determine the seed point with the highest IoU.

To perform the region growing process for the seed point, we employed Otsu's method to determine the threshold for each image. The calculated threshold was then utilized in the region growing algorithm.

2) Finding the Best Threshold Value for all Set: In our approach, the threshold serves as another critical parameter, influencing the inclusion of pixels in the growing region. Achieving a careful balance in threshold selection is crucial. Excessive permissiveness may lead to over-segmentation, introducing dissimilar pixels and compromising accuracy. Conversely, undue strictness may result in under-segmentation, causing the omission of vital details in the segmented regions.

Fig. 5. Tumor segmentation for 17, 54, 113, and 127 threshold values.

Our method is tailored to select a singular threshold value that applies uniformly across an entire group of images. This design is implemented to streamline the process and mitigate time consumption, as it eliminates the need to find an individualized threshold value for each image separately within the extensive dataset. Leveraging well-selected seed points from the prior method, we provide the region-growing algorithm with threshold values ranging from 0 to 255. Subsequently, we calculate the average Intersection over Union (IoU) for the segmented images within each image set. Plotting these results allows us to identify the threshold value that optimally enhances segmentation across the image set.

C. Region Growing with Automatic Seed Selection

1) Strategy 1: Preset Threshold: The first strategy that we propose is region growing with a simple thresholding criteria that is based on a preset value. The value that we choose in our experiments is 6. Starting from the seed point, the absolute difference between the seed and its neighboring pixels is going to be calculated, and if it is less than the preset threshold then the region is going to grow, otherwise, the region-growing algorithm is not going to continue exploring this region.

$$|P(S) - P(N)| < T \tag{1}$$

where P(S) is the pixel value at the seed point,
P(N) is the pixel value at a neighboring point, and
T = 6 is the threshold value.

2) Strategy 2: Dynamic Thresholding: The second strategy that we adopt is a dynamic thresholding criteria. This criterion is based on some statistics from the input image, namely the minimum and maximum pixel intensities. First, the minimum pixel value is subtracted from the maximum pixel value, and then the outcome of this subtraction is multiplied by a threshold factor, which is 0.3 in our case. For example, the threshold for a binary image would be computed as $T = (255 - 0) * 0.3 = 76.5$.

$$T = (P_{\text{max}} - P_{\text{min}}) \times \text{factor} \tag{2}$$

where the threshold factor is set to 0.3.

3) Automatic Seed Selection: Seed selection is one of the most challenging parts of any region-growing algorithm, as its choice has a great impact on the final segmentation outcome. In our case, the seed point is going to generate the region of interest and the predicted tumor region. Automatic seed selection can be based on pixel intensity, texture, or other statistical parameters. In our work, we choose the pixel intensity as the deciding factor for seed selection. We select the seed points based on image areas with maximum pixel intensity. We achieve this by iterating through the image in NxN blocks, where N is the dimension of the block size. This is helpful because it helps us process the image in smaller blocks. Then, we extract each block and find the maximum pixel intensity inside it, store it, and store the center coordinates of the block. We do that for every NxN block in the image. After analyzing all the blocks in the whole image, we sorted the center coordinates based on the value of the pixel intensity that was inside the block. Finally, we pick the top three coordinates as our best seeds. This process has the potential to produce very good seed points because the pixel intensity of the tumor region is often brighter compared to other brain regions, which is illustrated in Figure 6.

$$\text{MaxIntensity}(B_{i,j}) = \max_{(x,y) \in B_{i,j}} P(x, y) \tag{3}$$

where P(x, y) is the pixel intensity at coordinates (x, y) in $B_{i,j}$.

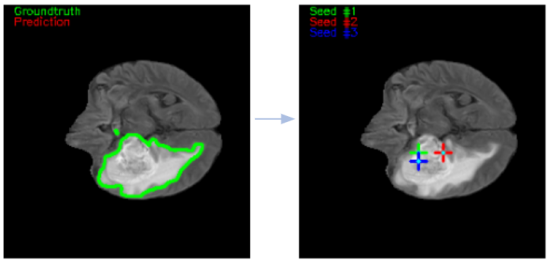

Fig. 6. Left: A FLAIR image from the BraTS that has the tumor region hilghted with green. Right: The top three selected seeds using our automatic seed selection algorithm. The image shows the efficacy of our seed selection process resulting in producing seeds within the target tumor region.

4) Implementation: Now that we have defined the criteria and the seed selection strategy, let's dive into the details of the region-growing algorithm. The region-growing segmentation algorithm is a breadth-first search (BFS) based flood fill algorithm, illustrated in Figure **??**, and that works as follows. Given an image and a seed pixel, the algorithm will start to explore all the neighboring pixels in eight different directions, horizontally, vertically, and diagonally. Then, based on similarity criteria, the algorithm will compare the pixel intensities of the seed points and their neighbors, and grow the region based on the preset criteria. By iteratively performing these steps, the algorithm will be able to segment all pixels that belong to the same class of the seed point based on comparing their pixel intensities.

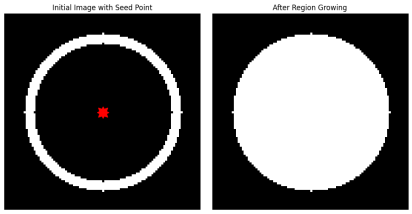

Fig. 7. Left: A simple black circle with white boundaries. Right: The result of applying a region growing algorithm starting from the center pixel as the seed and growing in all directions if the pixel is black. The algorithm stops growing at the white boundaries.

5) Post-processing: We add a post-processing step in our pipeline that has the potential to enhance the performance of segmenting brain tumors, and which works very well for binary segmented images. After we perform the segmentation, we apply a binary hole-filling operation to the segmented output. As the name implies, this operation aims at filling any gaps within the segmented output to increase the region of overlap between the ground truth and the predicted tumor region. This is mainly based on the observation that tumor regions are often represented as a whole part and do not contain any gaps inside them.

$$\text{Hole-Filling}(B) = \text{Fill}(B) \tag{4}$$

where B represents the binary segmented output, the operation **Fill**(B) targets any gaps within B.

Algorithm 1 Region Growing with Automatic Seed Selection, Dynamic Thresholding, and Binary Hole Filling

Dynamic Thresholding
Compute the maximum (P_{max}) and minimum (P_{min}) pixel intensities of the input image.
$T \leftarrow (P_{\text{max}} - P_{\text{min}}) \times 0.3$
Automatic Seed Selection
for each $N \times N$ block $B_{i,j}$ in the image **do**
  Find the maximum pixel intensity $MaxIntensity(B_{i,j})$ in $B_{i,j}$.
  Store $MaxIntensity(B_{i,j})$ and the center coordinates of $B_{i,j}$.
end for
Sort the center coordinates based on $MaxIntensity$ values.
Select the top three coordinates as seed points.
Region Growing
repeat
  for each seed point **do**
    for each neighboring pixel in all directions **do**
      if neighbor's intensity is similar to the seed's intensity based on T **then**
        Include neighbor in the region.
      end if
    end for
  end for
until no more similar neighbors
Post-processing
Apply a binary hole-filling operation to the segmented output.
Output: Segmented image with improved tumor region delineation. =0

D. Evaluation Metrics for Segmentation

1) **Intersection over Union (IoU):** Measures the overlap between the predicted segmentation mask and the ground truth. It's the ratio of the intersection of the two masks to their union. The values of the IoU are between 0-1. Higher scores indicate better segemtantion.

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{TP}{TP + FP + FN} \tag{5}$$

Here, TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

2) **Dice Similarity Score (DSC):** Evaluates pixel-wise agreement between the ground truth and predicted masks. It's similar to IoU but emphasizes the size of the overlap. The values of the DSC are between 0-1. Higher scores indicate better segemtantion.

$$DSC = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{6}$$

3) **Extra Fraction:** Quantifies the extent of over-segmentation, representing the fraction of pixels incorrectly identified as tumor. Lower scores indicate better segemtantion.

$$\text{Extra Fraction} = \frac{FP}{TP + FP} \tag{7}$$

4) **Overlap Fraction:** Measures the fraction of the tumor correctly identified in the prediction. Higher scores indicate better segemtantion.

$$\text{Overlap Fraction} = \frac{TP}{TP + FN} \tag{8}$$

5) **Specificity:** Reflects the ability of the model to correctly identify non-tumor regions. The values of the IoU are between 0-1. Higher scores indicate better segemtantion.

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{9}$$

TN denotes true negatives.

6) **Accuracy:** Represents the overall correctness of the segmentation, i.e., the proportion of correctly identified pixels. The values of the IoU are between 0-1. Higher scores indicate better segemtantion.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

7) **Root Mean Squared Error (RMSE):** Measures the average magnitude of error between the predicted and ground truth masks, pixel-wise. Lower values indicate better segemtantion.

$$RMSE = \sqrt{\frac{\sum (predicted - actual)^2}{N}} \tag{11}$$

N is the total number of pixels.

8) **Runtime:** The time required to process and segment a tumor from an MRI scan. Runtime is typically given in seconds or milliseconds and is often empirically determined rather than expressed as an equation.

$$\text{Runtime (in seconds)} \tag{12}$$

IV. Results

A. Experimental Setup

We perform our experiments on both datasets 1 and 2 with 12 and 15 samples, referred to as Set1 and Set2, respectively. We evaluate our methods' performance on eight evaluation metrics, namely intersection-over-union (IoU), dice similarity score (DSS), extra fraction (EF), overlap fraction (OF), accuracy (Acc), specificity (Sp), and root mean squared error (RMSE). The range of values of IoU, DSS, Acc, OF, and Sp are between 0 and 1, with larger numbers indicating better performance. On the other hand, smaller EF values mean less over-segmentation. All metrics above the 0-1 range are log-scaled between 0 and 1 only when presented on bar charts. In our experiments, we report average scores on all the dataset images, while showing per-image metrics as well for more in-depth quantitative and qualitative analysis.

B. Thresholding-Based Segmentation

By observing the results, we can see that thresholding resulted in perfect segmentation. However, some parts performed better compared to others. We ran the segmentation methods on two different datasets and calculated all metrics for them. In the table below, we can see the results in Tables I and **??**.

	IoU	DS	EF	OF	Acc	Spec	RMSE
SD	0.15	0.26	7.16	0.98	0.84	0.83	0.40
TE	0.54	0.67	1.37	0.69	0.97	0.98	0.14
Mean	0.12	0.20	0.03	0.13	0.97	1.00	0.16
Ostu	0.42	0.56	0.23	0.47	0.98	1.00	0.13

TABLE I
PERFORMANCE METRICS FOR DATA SET 1

	IoU	DS	EF	OF	Acc	Spec	RMSE
SD	0.15	0.26	7.16	0.98	0.84	0.83	0.40
TE	0.54	0.67	1.37	0.69	0.97	0.98	0.14
Mean	0.12	0.20	0.03	0.13	0.97	1.00	0.16
Ostu	0.42	0.56	0.23	0.47	0.98	1.00	0.13

TABLE II
PERFORMANCE METRICS FOR DATA SET 2

As we can see in the tables, the trial and exam methods did a slightly better job than the Otsu method. Standard deviation didn't perform well and mostly failed to predict the correct segmentation.

C. Region Growing with Manual Seed Selection

Observing the plot, we discern that an optimal threshold value yields the most effective segmentation among the images in the set. A lower threshold tends to produce limited tumor segmentation, while a threshold exceeding the optimum introduces more non-tumor pixels into the output. This exploration helps establish a balance, ensuring accurate segmentation results across the entire image group.

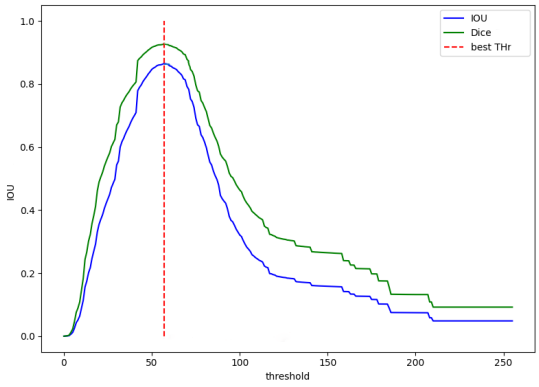

Fig. 8. Graph illustrating the Average IoU for Set1 across Threshold Values from 0 to 255. Notably, at the 57th threshold value, the maximum IoU is recorded at 0.87, accompanied by a Dice similarity of 0.93.

As a result, it is recommended to opt for a threshold value between 54 and 57 for both sets of images. Accordingly, we selected 57 as the threshold for set1 and 54 for set2, conducting region growing for each image based on its respective threshold and the optimal seed point.

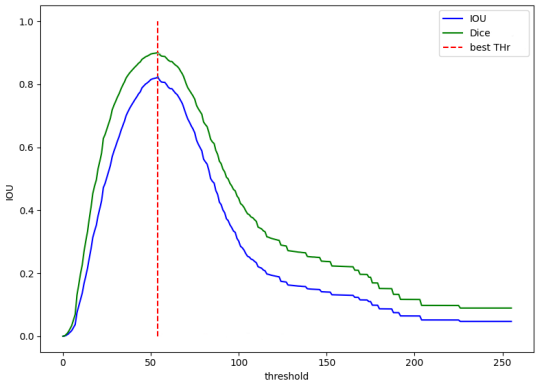

Fig. 9. Graph illustrating the Average IoU for Set2 across Threshold Values from 0 to 255. Notably, at the 54th threshold value, the maximum IoU is recorded at 0.82, accompanied by a Dice similarity of 0.93.

Best Threshold = 57							
ID	IoU	DSS	EF	OF	Acc	Sp	RMSE
1	0.849	0.918	0.059	0.9	0.991	0.996	24.414
2	0.93	0.963	0.032	0.96	0.997	0.998	14.491
3	0.87	0.93	0	0.87	0.991	1	24.159
4	0.776	0.874	0.165	0.905	0.986	0.991	30.014
5	0.777	0.874	0.13	0.879	0.989	0.994	26.156
6	0.89	0.942	0.021	0.909	0.997	0.999	14.607
7	0.87	0.93	0.124	0.978	0.994	0.994	20.382
8	0.85	0.919	0.112	0.945	0.99	0.993	25.786
9	0.831	0.907	0.122	0.932	0.992	0.994	23.327
10	0.884	0.938	0.029	0.91	0.996	0.999	16.008
11	0.906	0.95	0.018	0.922	0.995	0.999	18.434
12	0.946	0.972	0	0.946	0.997	1	13.099
Avg	0.865	0.927	0.068	0.922	0.993	0.997	20.906

Fig. 10. The evaluation table for each image in set1 involving optimum seed point and best threshold value for set1

Best Threshold = 54							
ID	IoU	DSS	EF	OF	Acc	Sp	RMSE
1	0.599	0.749	0.133	0.678	0.982	0.995	34.182
2	0.839	0.913	0.035	0.868	0.99	0.998	24.963
3	0.88	0.936	0.073	0.945	0.994	0.997	19.33
4	0.875	0.933	0	0.875	0.991	1	23.734
5	0.778	0.875	0.14	0.887	0.987	0.992	29.617
6	0.803	0.891	0.074	0.862	0.991	0.997	24.018
7	0.865	0.928	0.008	0.872	0.996	1	16.114
8	0.795	0.886	0.215	0.966	0.989	0.99	26.626
9	0.893	0.944	0	0.893	0.993	1	20.657
10	0.807	0.893	0.163	0.938	0.993	0.995	21.566
11	0.854	0.921	0.032	0.882	0.993	0.999	20.766
12	0.85	0.919	0.051	0.893	0.994	0.998	20.327
13	0.767	0.868	0.262	0.968	0.982	0.983	34.017
14	0.847	0.917	0.019	0.863	0.995	0.999	18.311
15	0.895	0.944	0.03	0.921	0.994	0.998	19.649
Avg	0.823	0.901	0.082	0.887	0.991	0.996	23.591

Fig. 11. The evaluation table for each image in set2 involving optimum seed point and best threshold value for set2

D. Region Growing with Automatic Seed Selection

1) Comparing Thresholding Strategies: First, we assess the performance of both thresholding strategies, namely, preset and dynamic thresholding. We run both methods on both datasets and compute the average metrics values. Figure 12 shows that the dynamic thresholding method consistently outperforms its counterpart on Set1, achieving 0.891 IoU and 0.942 DSS, compared to 0.207 and 0.293 achieved by the preset threshold method. This pattern repeats for all other metrics and when the methods are assessed on Set2, which contains 15 images as shown in Figure 13. Exact numbers for each metric are shown in Table III. Based on this finding, we will focus most of our upcoming experiments on the dynamic thresholded region-growing algorithm.

Set	RG	IoU	DSS	EF	OF	Acc	Sp	RMSE
1	Pre	0.21	0.29	0.56	0.24	0.95	0.98	55.73
1	Dyn	**0.89**	**0.94**	**0.06**	**0.94**	**0.99**	**1.00**	**18.32**
2	Pre	0.13	0.20	2.46	0.26	0.87	0.91	54.81
2	Dyn	**0.80**	**0.87**	**0.33**	**0.92**	**0.98**	**0.99**	**26.95**

TABLE III
COMPARING BOTH THRESHOLDING STRATEGIES FOR REGION GROWING WITH AUTOMATIC SEED SELECTION. 'DYN' INDICATES THE USE OF THE DYNAMIC THRESHOLDING METHOD. 'PRE' INDICATES THE USE OF THE PRESET THRESHOLDING METHOD.

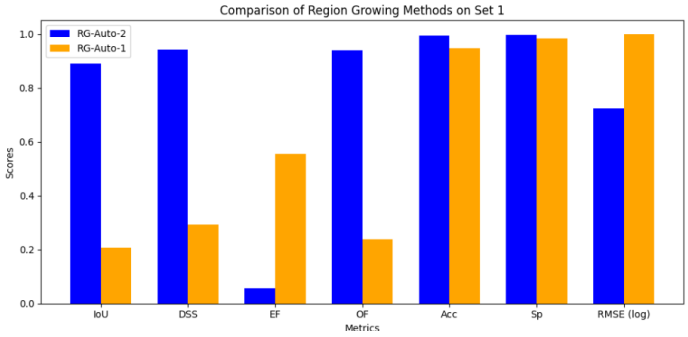

Fig. 12. Blue bars indicate region growing with dynamic thresholding. Yellow bars indicate the preset thresholding strategy. The plot compares the segmentation results of both methods on Set1.

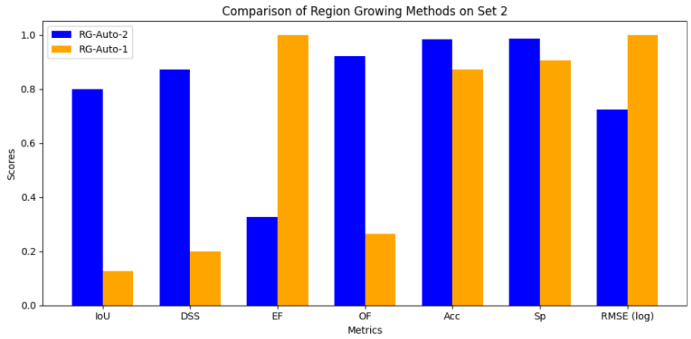

Fig. 13. Blue bars indicate region growing with dynamic thresholding. Yellow bars indicate the preset thresholding strategy. The plot compares the segmentation results of both methods on Set2.

2) Assessing the Impact of Binary Hole Filling: Second, we assess the impact of the post-processing operation of binary hole filling using the dynamic thresholding method. We show the results of performing this experiment on both Set1 and Set2 in Table IV. The table shows how the fill holes operations, denoted as 'fill', have a positive impact on all seven metrics, compared to the absence of this preprocessing step 'nofill'. The performance increases in IoU and DSS are +0.03 and 0.018 for Set1 and +0.025 and +0.012 for Set2. This consistent boost in the segmentation performance is due to the robustness of this post-processing technique in filling unnecessary gaps generated by the segmentation algorithm. By filling these gaps, our algorithm can cover more of the tumor area than originally covered by the region-growing algorithm. Figure 14 visually presents the impact of this technique on two images from Set1 and Set2.

Set	Fill	IoU	DSS	EF	OF	Acc	Sp	RMSE
1	✓	0.89	0.94	0.06	0.94	0.99	1.00	18.32
1	×	0.86	0.92	0.05	0.91	0.99	1.00	55.58
Δ	-	+0.03	+0.02	+0.00	+0.03	+0.00	0.00	+37.26
2	✓	0.80	0.87	0.33	0.92	0.98	0.99	26.95
2	×	0.77	0.86	0.25	0.89	0.98	0.99	54.67
Δ	-	+0.03	+0.01	+0.08	+0.03	-0.00	-0.00	+27.72

TABLE IV
ASSESSING THE IMPACT OF POST-PROCESSING (BINARY HOLE FILLING) USING REGION GROWING WITH DYNAMIC THRESHOLDING.

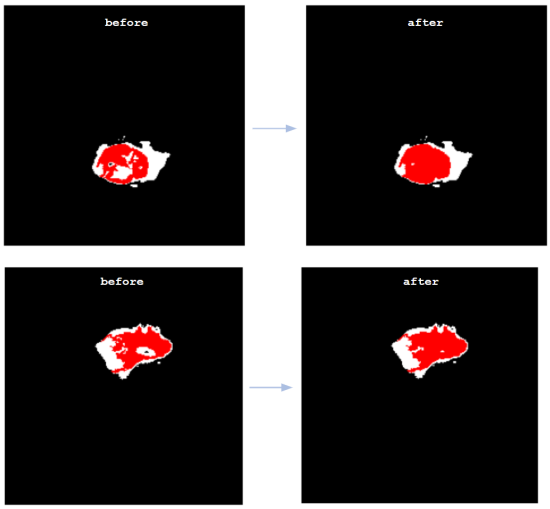

Fig. 14. Red pixels are indicating an overlap between groundtruth and predicted segmetnation. The more red area, the better the segmetnation is. The figure illustrates the positive impact of using the fill holes post-processing techniques on two examples, by effectively filling any unnecessary gaps, which correctly belong to the tumor region most of the times.

3) Experimenting Different Block Sizes for Seed Selection: Seed selection process is a very challenging task as it can significantly affect the segmentation outcome. For this reason, we further assess our automatic seed selection technique by experimenting the same region growing method with dynamic threholsding, but when using different block sizes when searching for seed points. Block sizes used in the experiment

are 4x4, 8x8, and 16x16. In Table V, we show how choosing an 8x8 block size for searching for the best seeds achieves superior results compared to 4x4 and 16x16. That is reflected in the value of all seven evaluation metrics.

Set	Block Size	IoU	DSS	EF	OF	Acc	Sp	RMSE
1	4x4	0.70	0.81	0.08	0.76	0.99	1.00	29.59
1	**8x8**	0.89	0.94	0.06	0.94	0.99	1.00	18.32
1	16x16	0.86	0.93	0.13	0.97	0.99	0.99	20.70
2	4x4	0.67	0.77	0.51	0.80	0.97	0.98	35.63
2	**8x8**	0.80	0.87	0.33	0.92	0.98	0.99	26.95
2	16x16	0.69	0.77	1.17	0.97	0.95	0.95	43.46

TABLE V

EVALUATING THE PERFORMANCE OF REGION GROWING WITH AUTOMATIC SEED SELECTION WITH DIFFERENT BLOCK SIZES USED TO GENERATE THE INITIAL SEED POINTS.

4) Quantitative Analysis: We perform a more in-depth comprehensive analysis of the dynamic thresholding method on both Set1 and Set2, in Tables VI and VII, respectively. We compute per-image metrics and their averages. Our method archives very high-quality segmentation results, reflected in all the seven reported evaluation metrics. For Set1, our method achieves 0.891 IoU and 0.942 DSS. For Set2, our method achieves 0.799 IoU and 0.873 DSS. Both evaluations have very low EF, which means there is only a small portion of over-segmentation. OF is high, which indicates a high overlap between ground truth labels and our produced segments. Finally, RMSE is low, which indicates a low rate of mismatch between ground truth and predicted pixels.

Moreover, we performed a detailed, per-image evaluation. This is important to inspect and analyze failure and success cases in each dataset. In Figures 15 and 16, we show box plots that describe the minimum, maximum, and median scores for the IoU and DSS for both datasets. This is also crucial so that we check the outliers and study the reason behind the segmentation result we got. The worst segmentations for both datasets achieved a DSS of 0.896 and 0.363, respectively, while the best segmentations achieved nearly perfect segmentation, scoring a DSS of 0.988 and 0.981, for sets 1 and 2, respectively. Furthermore, we will qualitatively analyze the worst and best segmentation outcomes and discuss the reasons behind their performance.

ID	IoU	DSS	EF	OF	Acc	Sp	RMSE
1	0.850	0.919	0.070	0.909	0.991	0.996	24.530
2	0.888	0.941	0.009	0.896	0.995	1.000	18.125
3	0.811	0.896	0.229	0.997	0.984	0.983	32.262
4	0.934	0.966	0.059	0.990	0.996	0.997	15.507
5	0.886	0.939	0.006	0.891	0.995	1.000	17.684
6	0.859	0.924	0.105	0.948	0.995	0.997	17.264
7	0.815	0.898	0.036	0.844	0.992	0.998	23.399
8	0.963	0.981	0.022	0.984	0.998	0.999	12.345
9	0.879	0.936	0.024	0.900	0.995	0.999	18.858
10	0.906	0.951	0.078	0.977	0.997	0.997	14.722
11	0.927	0.962	0.028	0.952	0.996	0.998	16.357
12	0.976	0.988	0.004	0.980	0.999	1.000	8.826
Avg	0.891	0.942	0.056	0.939	0.994	0.997	18.323

TABLE VI

PER-IMAGE METRICS AND THEIR AVERAGES FOR SET1 WITH DYNAMIC THRESHOLDING AND FILLING HOLES.

ID	IoU	DSS	EF	OF	Acc	Sp	RMSE
1	0.597	0.748	0.276	0.762	0.980	0.989	36.359
2	0.850	0.919	0.070	0.909	0.991	0.996	24.530
3	0.888	0.941	0.009	0.896	0.995	1.000	18.125
4	0.811	0.896	0.229	0.997	0.984	0.983	32.262
5	0.934	0.966	0.059	0.990	0.996	0.997	15.507
6	0.886	0.939	0.006	0.891	0.995	1.000	17.684
7	0.859	0.924	0.105	0.948	0.995	0.997	17.264
8	0.815	0.898	0.036	0.844	0.992	0.998	23.399
9	0.963	0.981	0.022	0.984	0.998	0.999	12.345
10	0.672	0.804	0.462	0.983	0.985	0.985	31.465
11	0.879	0.936	0.024	0.900	0.995	0.999	18.858
12	0.222	0.363	3.499	0.999	0.859	0.853	95.737
13	0.778	0.875	0.009	0.785	0.986	0.999	29.674
14	0.906	0.951	0.078	0.977	0.997	0.997	14.722
15	0.927	0.962	0.028	0.952	0.996	0.998	16.357
Avg	0.799	0.873	0.327	0.921	0.983	0.986	26.952

TABLE VII

PER-IMAGE METRICS AND THEIR AVERAGES FOR SET2 WITH DYNAMIC THRESHOLDING AND FILLING HOLES.

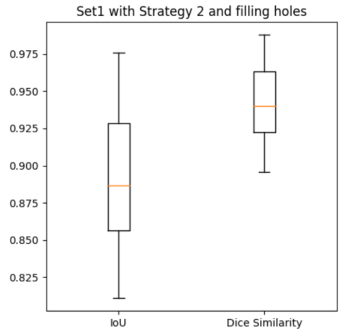

Fig. 15. Box plot that illustrates the per-image statistical evaluation on Set1.

5) Qualitative Analysis: We perform a visual analysis of the quality of our segmentation. In Figure 17, we show the two worst segmentation produced by the dynamic thresholding algorithm. The most common reason for low DSS is over-segmentation, which is apparent in both images. The reason can be that the pixel intensities do not vary a lot between tumor and non-tumor regions, which results in over-segmentation even if the seeds were selected from within the target tumor region. In Figure 18, we show the two best segmentations. The tumor structure is visually clear and the pixel intensities can be

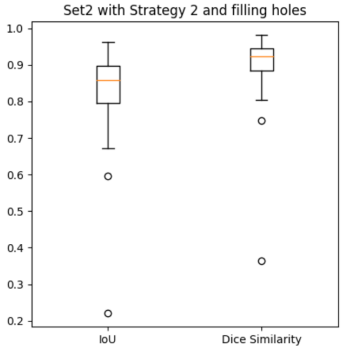

Fig. 16. Box plot that illustrates the per-image statistical evaluation on Set2.

easily differentiated, which makes the segmentation outcome better, achieving a near-perfect DSS of 0.98.

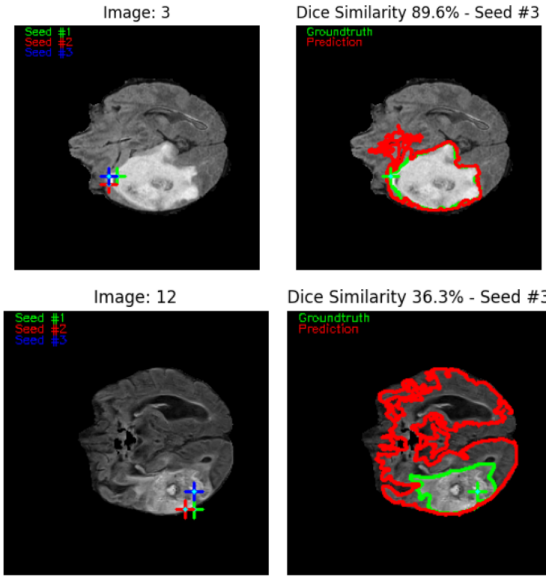

Fig. 17. Worst two segmetnations from Set1 (top row) and Set2 (bottom row) based on the dice similarity score (DSS).

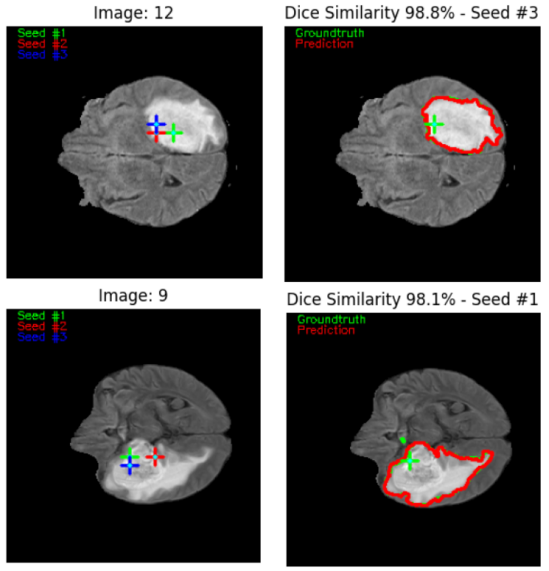

Fig. 18. Best two segmetnations from Set1 (top row) and Set2 (bottom row) based on the dice similarity score (DSS).

E. Computational Complexity

It is crucial to consider the time complexity of the segmentation algorithms and the trade-off between the segmentation quality and its runtime. In Table VIII, we show the runtime in seconds for four of our proposed methods, alongside their best-achieved dice similarity score (DSS). This is considered a good practice to inspect the trade-off and make a decision on which algorithm to use based on the requirements. In most cases, higher quality segmentations are preferred compared to faster execution time, but this depends on the application.

	Runtime (sec)	DSS
RG w/ Dynamic Threshold	0.120	0.891
RG w/ Manual Seed Selection	0.060	0.865
Global Thresholding	0.003	0.535
RG w/ Preset Threshold	0.010	0.207

TABLE VIII
TRADE-OFF BETWEEN RUNTIME IN SECONDS AND DSS SCORE.

F. Effect of Noise on Segmentation

To assess if our proposed segmentation methods are robust to noise, we add a Gaussian noise with a variance of 0.01 to both of our datasets. Then, we apply the best two segmentation methods, namely automatic seed selection with dynamic thresholding and manual seed selection, to the noisy datasets and then evaluate the segmentation performance. However, the impact of noise is more severe when applied to the region growing with manual seed selection. This shows how the dynamic thresholding method is more adaptable and robust to variations in the input when it comes to segmenting the tumor region. Figure 19 shows a bar plot inspecting the IoU for both Set1 and Set2, highlighting how the region growing with manual seed selection is more prone to the added Gaussian noise.

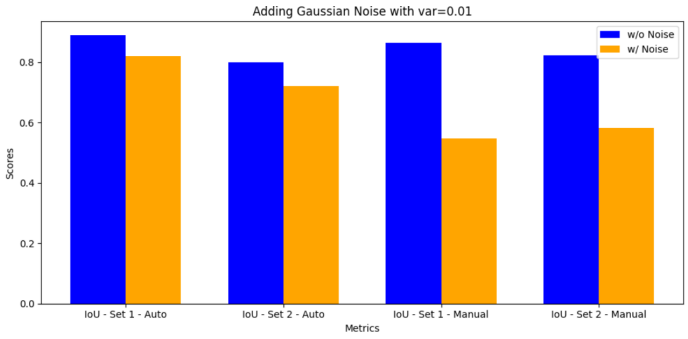

Fig. 19. Blue bars indicate the performance of a specific algorithm without noise. Yellow bars indicate the presence of a Gaussian noise with variance value of 0.01. The plot is comparing the IoU score for automatic seed selection (first 4 bars) vs. manual seed selection (last 4 bars) on both Set1 and Set2.

G. Comparing All Methods

Finally, we compiled this Table IX to compare the results of the best experiments so far on both Set1 and Set2. The variants of region-growing algorithm generally perform better than thresholding-based segmentation. RG with dynamic thresholding performs best on Set1, while manual seed selection performs best on Set2. That being said, both methods are on par when working on the original data. RG with dynamic thresholding proofs being more robust to noisy data as discussed in Subsection IV-F.

Set	Method	IoU	DSS	EF	OF	Acc	Sp	RMSE
1	RG-Auto-2	**0.9**	**0.9**	**0.1**	**0.9**	**1.0**	**1.0**	**18.3**
1	RG-Manual	0.9	0.9	0.1	0.9	0.99	**1.0**	20.9
1	Trial&Error	0.5	0.0	1.4	0.7	0.97	0.98	42.2
1	Otsu	0.4	0.0	0.2	0.5	0.98	1.0	42.2
1	RG-Auto-1	0.2	0.3	0.6	0.2	0.95	0.98	55.7
2	RG-Manual	**0.8**	**0.9**	0.1	0.9	**0.99**	1.0	**23.6**
2	RG-Auto-2	0.8	0.9	0.3	**0.9**	0.98	0.99	27.0
2	Trial&Error	0.5	0.0	0.5	0.7	0.98	0.99	40.0
2	Otsu	0.5	0.0	**0.0**	0.5	0.99	**1.0**	40.0
2	RG-Auto-1	0.1	0.2	2.5	0.3	0.87	0.91	54.8

TABLE IX
ASSESSING THE OVERALL METRICS ON SETS 1 & 2.

V. DISCUSSION

A. MRI Modalities

FLAIR (Fluid-Attenuated Inversion Recovery) MRI significantly enhances brain tumor detection, offering clearer contrast between tumors and edema compared to traditional MRI. This modality effectively suppresses cerebrospinal fluid (CSF) signals, highlighting infiltrative tumor patterns and irregular margins crucial for accurate identification. FLAIR MRI is essential in neuro-oncological imaging, aiding in diagnosis and differentiating post-therapeutic changes, thereby improving clinical decision-making.

B. Post-processing Techniques

Binary hole filling, a post-processing technique in tumor segmentation, effectively addresses cavities in segmented areas, common due to tumor texture heterogeneity. This technique enhances segmentation continuity and accuracy, representing the tumor mass more precisely, crucial for clinical evaluation and treatment planning. Its simplicity and effectiveness render it vital in the segmentation process.

C. Effect of Noise on Segmentation

Our research indicates that manual seed selection and preset thresholding in region growing are more prone to noise, unlike automatic seed selection with dynamic thresholding which offers robustness and adaptability. Automatic seed selection minimizes human error, and dynamic thresholding adjusts to noise variations, ensuring precise and consistent segmentation across different datasets, making it a preferable choice for tumor segmentation.

D. Evaluation Metrics

Intersection over Union (IoU) and Dice Similarity Score (DSS) are effective in evaluating tumor segmentation, accurately assessing overlap and similarity between segmented areas and ground truth. These metrics are less affected by class imbalance, unlike accuracy and specificity, which may not reflect true segmentation performance in imbalanced medical imaging datasets. IoU and DSS provide a more comprehensive evaluation of tumor segmentation.

VI. CONCLUSION

In conclusion, our study demonstrates the effectiveness of region growing methods, especially with automatic seed selection and dynamic thresholding, in MRI-based tumor segmentation. The superiority of these methods is further enhanced by the use of FLAIR modality in MRI, which significantly improves tumor detection. Metrics like IoU and DSS proved more reliable than accuracy and specificity in evaluating segmentation performance, given the challenges of class imbalance and spatial accuracy. Additionally, binary hole filling emerges as a valuable post-processing technique, enhancing the segmentation quality.

REFERENCES

[1] OpenAI. (2023). *ChatGPT* (Mar 14 version) [Large language model]. https://chat.openai.com/chathttps://chat.openai.com/chat

[2] Contributors to Wikimedia projects. "Otsu's Method - Wikipedia." Wikipedia, the Free Encyclopedia, Wikimedia Foundation, Inc., 25 May 2005, https://en.wikipedia.org/wiki/Otsu

[3] G.M.A. R, "Automated Brain Tumor Segmentation Using Multimodal Brain Scans: A Survey Based on Models Submitted to the BRATS 2012-2018 Challenges," \emph{IEEE Reviews in Biomedical Engineering}, no date. Available at: \url{https://pubmed.ncbi.nlm.nih.gov/31613783/} (Accessed: 11 December 2023).

[4] R; G.MA (no date) Automated brain tumor segmentation using Multimodal Brain Scans: A survey based on models submitted to the brats 2012-2018 challenges, IEEE reviews in biomedical engineering.

[5] Biratu, E. S., Schwenker, F., Debelee, T. G., Kebede, S. R., Negera, W. G., & Molla, H. T. (2021). Enhanced region growing for brain tumor MR image segmentation. *Journal of Imaging*, *7*(2), 22. https://doi.org/10.3390/jimaging7020022

[6] Thiruvenkadam, K., Perumal, N. (2015). Brain Tumor Segmentation of MRI Brain Images through FCM clustering and Seeded Region Growing Technique. ResearchGate.

[7] BRATS2020 Dataset (Training + Validation). (2020, July 2). Kaggle.