International Journal of Computational and Experimental Science and Engineering (IJCESEN)

Vol. 11-No.2 (2025) pp. 3146-3159
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Decoding GBM Tumor Dynamics: AI-Driven Segmentation with RANO Criterion Validation for the Prediction of Radiotherapy Outcomes

Serhat Ünalan[1]*, Didem Karaçetin[2,3], Ümit Tüzün[4], Osman Günay[5]

[1]Yildiz Technical University, Electrics and Electronics Faculty, Biomedical Engineering Department, 34230, Istanbul - Turkiye
* **Corresponding Author Email:** serhtunalan07@gmail.com - **ORCID:** 0009-0000-4328-2114

[2]Department of Radiation Oncology, Neolife Medical Center, 34330, Istanbul -Turkiye
[3]Department of Radiation Oncology, Biruni University Hospital, 34290, Istanbul – Turkiye
Email: didemkaracetin@gmail.com - **ORCID:** 0000-0001-5359-5958

[4]Department of Radiology, Neolife Medical Center, 34330, Istanbul -Turkiye
Email: umit.tuzun@neolife.com.tr - **ORCID:** 0009-0002-9881-3255

[5]Yildiz Technical University, Electrics and Electronics Faculty, Biomedical Engineering Department, 34230, Istanbul-Turkiye
Email: ogunay@yildiz.edu.tr - **ORCID:** 0000-0003-0760-554X

Article Info:

DOI: 10.22399/ijcesen.1630
Received: 12 January 2025
Accepted: 27 April 2025

Keywords

GBM,
Radiotherapy,
Artificial Intelligence,
Segmentation,
RANO.

Abstract:

Glioblastoma Multiforme (GBM) is a very aggressive brain tumor which has poor prognosis despite wide range of treatment modalities with specific enhancements in radiotherapy. Correct evaluation of tumor response to treatment is crucial for guiding treatment decision-making for patients. Despite the wide application of deep learning models for tumor segmentation and evaluation, their fundamental complexity has cast doubt on whether a simpler, traditional approach can yield insights of comparable reliability. A retrospective analysis was performed using MRI data from 18 GBM patients who had radiotherapy. An experienced radiologist evaluated all pre- and post-treatment MRI's and provided RANO scores to determine the tumor response. Multiparametric MRI sequences were segmented using Otsu's thresholding and GMM methods across sagittal, coronal, and axial planes. Dice Similarity Coefficients (DSC) and Intensity Distribution Scores (IDS) were used to evaluate tumor changes, with low DSC and high IDS values indicating successful treatment. The segmentation and statistical results were then compared with RANO scores to confirm the findings. The results demonstrated different tumor dynamics among patients, highlighting the variability in treatment outcomes. DSC and IDS offered additional insights into tumor alterations, where low DSC and high IDS values were determined as signs of successful radiotherapy. Both techniques effectively predicted outcomes with notable alterations, showcasing their capability for evaluating radiotherapy effectiveness in GBM treatment. This method provides a more straightforward, budget-friendly option compared to deep learning, yielding valuable understanding of tumor dynamics. Future research should prioritize confirming these results in more extensive groups by integrating advanced AI techniques.

1. Introduction

Main Glioblastoma multiforme (GBM), classified as grade IV by the World Health Organization (WHO), is the most frequent and aggressive primary malignant brain tumor in adults [1]. GBM comprises about 15% of all CNS tumors and is distinguished by rapid growth, extreme invasiveness and ultimate recurrence. The heterogeneous histopathology including necrosis and microvascular proliferation of GBM drives it is poor prognosis [2]. Clinically, patients typically display head symptoms, seizures, focal neurological deficits depending on the location of the tumor and it is stage. In spite of active therapeutic interference, the median overall survival is

restricted for GBM patients, in the range of 12 to 15 months after the diagnosis [3].

Radiotherapy is central in the treatment of GBM and constitutes the standard of care, along with maximal safe surgical resection and concurrent chemotherapy with temozolomide [4]. Radiotherapy is a localized treatment modality, which is mainly used to eliminate residual tumor cells that persist after surgery taking advantage of its mechanism to induce DNA damage and inhibit proliferation. The rise of sophisticated radiation delivery methods such as intensity-modulated radiotherapy (IMRT) and stereotactic radiosurgery (SRS) has allowed delivering radiation more accurately to the malignant tissues and sparing normal brain structures around the tumors [5]. However, due to the invasive nature of GBM, microscopic residual tumors can form beyond the visible tumor margins, making them resistant to radiotherapy. Moreover, the impact of radiation on adjacent healthy brain tissue, such as edema and radiation necrosis, complicates treatment decisions and follow-up imaging evaluations [6].

Another major clinical challenge is monitoring the response to radiotherapy in GBM. Traditional imaging modalities, including magnetic resonance imaging (MRI), have limitation because of their lack of reliability in differentiating true tumor progression versus treatment-related changes, for example, pseudo-progression [7]. The vagueness of treatment response adds uncertainty to clinical decision-making and perpetuates delays in therapy modification, indicating that novel methods are needed to adequately quantify treatment response.

Artificial intelligence (AI) has a potential solution for these challenges, particularly in the application of AI in neuroradiology. All of these factors support the incorporation of AI technology into imaging-based methods, especially deep learning algorithms, for enhanced tumor segmentation, feature extraction, and longitudinal analyses over time of the imaging data. These tools can also offer new quantitative perspectives of tumor dynamics, including tumor volume variability and changes in peritumoral niches, with superior accuracy and reproducibility [8,9]. These powerful capabilities gained wider practical relevance in the realm of radiotherapy, where a synergistic integration of clinical readouts could potentially lead to the early identification of treatment response or resistance that drives adaptive treatment strategies and ultimately improves treatment outcomes.

In this retrospective study, we aim to incorporate a mathematical and statistical framework to investigate and independent correlation with radiotherapy response in GBM patients from pre- and post-radiotherapy imaging data. Therefore, this study aims to use quantitative measurements of volumetric and morphological changes in tumor and peritumoral regions to extract mathematical models and statistical metrics that reflect the impact of treatment. This has led to the development of these methods, which can be used to create reliable frameworks, that are interpretable and can assess the radiotherapy treatment outcomes, owing to a more objective and precise evaluation of the therapeutic effectiveness, contributing to the management of GBM.

2. Material and Methods

The changes in tumor features were studied using pre and post radiotherapy magnetic resonance (MR) images of patients who were diagnosed with Glioblastoma Multiforme (GBM). This study was performed as a retrospective analysis and approved by the ethical committee of Biruni University Hospital (2024-BİAEK/04-48). We extracted data from the PACS (Picture archiving and communication system) for around 750 patients using this ethical clearance. From this data, we found 18 patients who met the inclusion criteria. In addition, for strengthening our data set for analysis, data were collected at Special Neolife Medicine Center. The criteria for the inclusion of patients can be outlined as follows: (i) Patients diagnosed with GBM and treated with radiotherapy, (ii) patients with complete MR images obtained pre-radiotherapy and post-radiotherapy, (iii) patients with MR images in DICOM (Digital Imaging and Communications in Medicine) format and sufficient quality for analytical purposes, (iv) adult patients aged 18 or older. This study was retrospective and allowed us to evaluate existing MRI data to determine the evolution of tumor features with radiotherapy. The selected dataset has also provided a rich basis to assess segmentation methods and clinical relevance.

Image preprocessing was performed on the aforementioned DICOM files in order to maintain uniformity and increase the data quality for segmentation as well as conducting further analysis. Preprocessing included rescaling of MR images to a uniform intensity range, denoising to remove noise, and registration (spatial alignment) for cross-scan consistency. These steps were necessary to reduce the impact of variances in image acquisition and to allow for standardization of the dataset for subsequent processing.

The tumor regions were delineated using two different segmentation methods: Otsu's thresholding method and Gaussian Mixture Model (GMM)-based segmentation. Otsu's segmentation was carried out for binary segmentation of tumor

regions in the pre- and post-treatment MR images. Namely, this approach determines an optimal threshold value such that the inter-class variance between the foreground (tumor) and background (non-tumor) regions of the image is maximized [10]. Assuming there are two main pixel intensity distributions in the image, each for a respective class, the algorithm proceeds to define separation boundaries. In our implementation, the threshold which maximizes inter-class variance is chosen as it provides the best discrimination between tumor and non-tumor areas by incrementally trying out threshold values. The threshold T is determined mathematically by maximizing the between-class variance (σ_B^2), expressed as:

$$\sigma_B^2(T) = \omega_1(T)\omega_2(T)[\mu_1(T) - \mu_2(T)]^2 \qquad (1)$$

where $\omega_1(T)$ and $\omega_2(T)$ are the probabilities of the two classes separated by T, and $\mu_1(T)$ and $\mu_2(T)$ are the mean intensities of the respective classes. Thus, the segmentation was based on statistically most significant threshold.

Prior to segmentation, some preprocessing methods were used to enhance image quality and make the segmentation algorithm robust. To enhance contrast, histogram equalization was applied, allowing pixel intensities to be spread more uniformly, which aided in distinguishing tumor from non-tumor regions. Furthermore, contrast stretching was applied using the 2nd and 98th percentiles of the pixel intensity values. This linear rescaling highlighted the key elements represented in the images while minimizing the effects of outlier intensities, allowing us to effectively utilize Otsu's method. After pre-processing, Otsu's method was implemented on each MR images. The threshold value at which the optical density of the images reached its minimum was calculated, and binary images were produced using that threshold and all the pixels with intensity greater than were treated as foreground pixels and the rest were treated as background pixels. The generated binary segmentation maps were encoded into matrices and tumor regions were denoted with a label of 1 and non-tumor regions with a label of 0.

On the other hand, GMM segmentation method offers a probabilistic framework to categorize image pixels into various components, utilizing the statistical properties of pixel intensities. This approach represents the pixel intensity values as a combination of Gaussian distributions, where each Gaussian represents a distinct tissue type or area within the brain tumor dataset [11]. The process of segmentation starts with converting the image pixels into a one-dimensional vector and utilizing the "GaussianMixture" class from the `sklearn` library. The quantity of components, n, indicates the number of Gaussian distributions to be fitted, providing adaptability in modeling varied tissue areas. This parameter was configured to 4, enabling the algorithm to categorize the image into four separate clusters: necrotic tissue, enhancing tumor, non-enhancing tumor, and healthy brain or background.

In mathematical form, the GMM assumes the probability density function $p(x)$ of the pixel intensities, x, can be modeled as a weighted sum of n Gaussian components:

$$p(x) = \sum_{i=1}^{n} \pi_i N\left(x|\mu_i, \sigma_i^2\right) \qquad (2)$$

where π_i are the mixture weights (summing to 1), μ_i are the means, and σ_i^2 are the variances of the i-th Gaussian component. The Expectation-Maximization (EM) algorithm estimates iteratively these parameters to maximize the likelihood of observed data. After fitting the GMM, the segmentation is done by assigning each pixel to the component with the highest posterior probability, given by Bayes' theorem:

$$label\ (x) = \arg(\max)_i \frac{\pi_i N\left(x|\mu_i,\sigma_i^2\right)}{\sum_{j=1}^{n} \pi_j N\left(x|\mu_j,\sigma_j^2\right)} \qquad (3)$$

In order to ensure computational efficiency, the input images are down-sampled to a fixed resolution (e.g., 128×128) before GMM fitting. The output segmentation label (after segmentation) is then up-sampled to the original image size using nearest-neighbor interpolation to preserve discrete labels. By reducing the amount of data while retaining enough detail for segmentation, this preprocessing step minimizes the amount of computation required.

The code implementation incorporates GMM segmentation into a pipeline designed for processing MR images before and after radiotherapy. The image data of each patient is processed concurrently to enhance throughput, utilizing the `joblib` library for multi-core processing. The generated segmentations are saved as `.npy` files for subsequent analyses. In contrast to Otsu's method, which is fundamentally deterministic and presumes a single-modal intensity distribution in each segmented area, GMM segmentation considers the likelihood of multimodal distributions and offers probabilistic assessments of each pixel's class affiliation. This ability is especially beneficial when managing

intricate, diverse tissues in areas affected by tumors.

Quantitative parameters were computed to analyze tumor changes between pre- and post-treatment imaging and segmentation performance. The spatial overlap between binary segmentation maps of images taken before and after treatment was measured using the Dice Similarity Coefficient (DSC) [12]. DSC has the following mathematical definition:

$$DSC = 2\,\frac{|A \cap B|}{|A|+|B|} \qquad (4)$$

where A and B are the binary segmentation results for pre- and post-treatment images, respectively. The numerator |A∩B| represents the number of overlapping pixels in the segmented regions, while the denominator |A|+|B| corresponds to the sum of the pixels in both regions. DSC varies between 0 and 1, where values nearing 1 suggest a higher level of overlap and consequently better consistency in tumor segmentation. To guarantee precise calculations, pre- and post-treatment images were trimmed to their smallest intersecting area prior to calculating DSC.

We also calculated the intensity distribution between pre- and post-treatment scans using an Intensity Distribution Score (IDS). For each image, pixel intensity histograms were calculated and compared with distinct images using the Bhattacharyya distance. The IDS calculated as this distance provides a measure of the overlap of the intensity distributions, hence a value close to 0 means high similarity and a value close to 1 corresponds with a high degree of dissimilarity [13]. Statistically, the Bhattacharyya distance, D_B, between two normalized histograms H_1 and H_2 is defined as:

$$D_B = -\ln\left(\sum_{i=1}^{n}\sqrt{H_1(i)H_2(i)}\right) \qquad (5)$$

where $H_1(i)$ and $H_2(i)$ are the probabilities of intensity i in the respective histograms. For IDS, $1 - D_B$ was used to simplify interpretation, with values closer to 1 reflecting more substantial treatment-induced changes in intensity distributions.

The findings were illustrated for a group of five randomly chosen patients, emphasizing segmentation results for scans conducted before and after treatment. Binary segmentation maps were presented alongside their original and processed versions, facilitating qualitative evaluation of tumor boundary definition. Quantitative measurements, such as DSC and IDS, were computed for each patient, and the findings were gathered for additional statistical examination. The metrics offered significant understanding of how treatment affects tumor morphology, along with the dependability of the segmentation methodology. This layered strategy integrated statistical precision, preprocessing improvements, and strong segmentation methods to guarantee precise and significant analysis of tumor alterations in glioblastoma patients receiving radiotherapy.

The mean and standard deviation of DSC values were calculated for each patient in order to compare the segmentation techniques' performance and examine the distribution of DSC and IDS scores among patients. This allowed for an evaluation of the segmentation methods' consistency and dependability. Based on the ranges of their mean DSC scores, the patients were further divided into groups, and bar charts, pie charts, and histograms were used to show the frequency distribution of these groups. In order to find patterns in intensity fluctuations, the distribution was evaluated after the scores were similarly divided into predetermined ranges for IDS analysis.

Patients were grouped according to their score ranges, and scatter plots were created to find out the association between mean DSC scores and their accompanying standard deviations. The segmentation performance for several patients was revealed by this graphic, which also indicated patterns and outliers. Furthermore, each patient's temporal variations in DSC scores across image pairs were shown, allowing for an evaluation of the segmentation techniques' consistency over time. These evaluations provide a thorough assessment of the segmentation methods and their suitability for characterizing GBM tumors.

To validate the findings, an experienced radiologist evaluated all of the pre- and post- MR images and determined the RANO (Response Assessment in Neuro-Oncology) scores for each patient. RANO criteria offer a uniform approach for assessing how a tumor responds to therapy, relying on MRI results [14]. These scores classify treatment results into four categories:

- RANO 1 (Complete Response): Signifies a total eradication of the tumor without requiring corticosteroids, demonstrating an outstanding therapeutic outcome.
- RANO 2 (Partial Response): Indicates a notable decrease (≥50%) in tumor dimensions, implying a favorable reaction to radiotherapy.
- RANO 3 (Stable Disease): Shows no significant alteration in tumor size, suggesting that the treatment did not alleviate or exacerbate the tumor load.

- RANO 4 (Progression): Indicates a ≥25% rise in tumor size or the emergence of new lesions, implying a treatment outcome that is ineffective.

3. Results and Discussions

The dataset included a total of 8,889 DICOM images, which included 3,990 pre-radiotherapy images and 4,899 post-radiotherapy images. The images were obtained from 18 patients who satisfied the inclusion criteria. Examples of raw DICOM images from four randomly selected patients before and after radiotherapy are illustrated in Figure 1. We used multiparametric MRI sequences, including T2-weighted (T2W), T1-weighted (T1W), and contrast-enhanced T1-weighted (T1ce) images in sagittal, coronal, and axial planes. For further segmentation and analysis, these raw images were utilized. The graphic illustrates how different anatomical features and imaging quality vary throughout individuals, necessitating meticulous preprocessing and segmentation to guarantee a consistent assessment of tumor alterations. Furthermore, Figure 2 demonstrates the results of Otsu's segmentation applied to MRI slices of multiple patients of pre- and post-treatment. It illustrates the binary segmentation masks generated by Otsu's thresholding, where white regions represent the segmented areas and black regions correspond to the background. This effectively highlights distinct regions of interest in the MRI slices. This visualization emphasizes the alignment between the

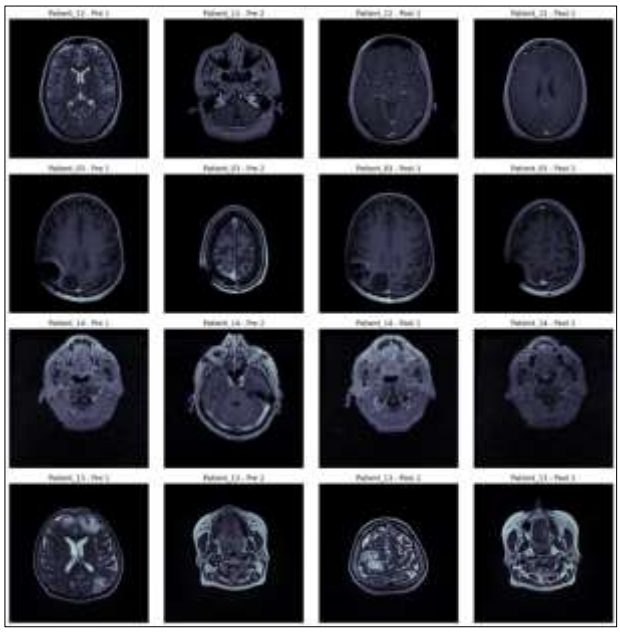

Figure 1. *Raw multiparametric MR images of four randomly selected patients, showing pre- and post-radiotherapy scans.*

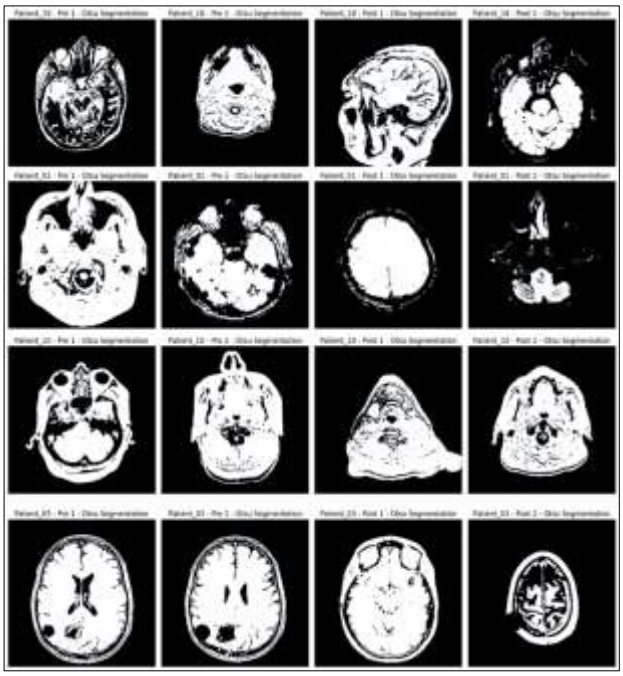

Figure 2. *Otsu's method segmentation results for four randomly selected patients' MR images. The figure displays pre- and post-radiotherapy segmented images for each patient, processed using multiparametric MRI sequences.*

segmented regions and the anatomical structures within the MRI, offering a comprehensive perspective on the segmentation results. Randomly selected samples from four patients across the dataset are presented, demonstrating consistent segmentation performance while reflecting anatomical and imaging variations. Multiparametric MR images from randomly selected patients, including T2W, T1W, and T1ce sequences in both pre- and post-radiotherapy scans, were segmented with the Gaussian Mixture Model (GMM). A representative example from four patients presenting their segmentation results is shown in Figure 3, where the colors denote clustered intensity segments among tumor areas and surrounding tissues. Outputs might not represent the overall segmentation performance due to the randomness in the selection of patients and images being sampled. However, these findings show the use of GMM to distinguish different areas of interest in the MRI data, reflecting the intensity changes between the scans before and after the radiotherapy. Second, all the segmented pre- and post-radiotherapy scans were used for further quantitative evaluation, including Dice Similarity Coefficient (DSC) and Intensity Distribution Scores (IDS) calculation. These metrics revealed consistency and intensity-based differences in the segmented tumor volumes between the various MRI modalities and time points.

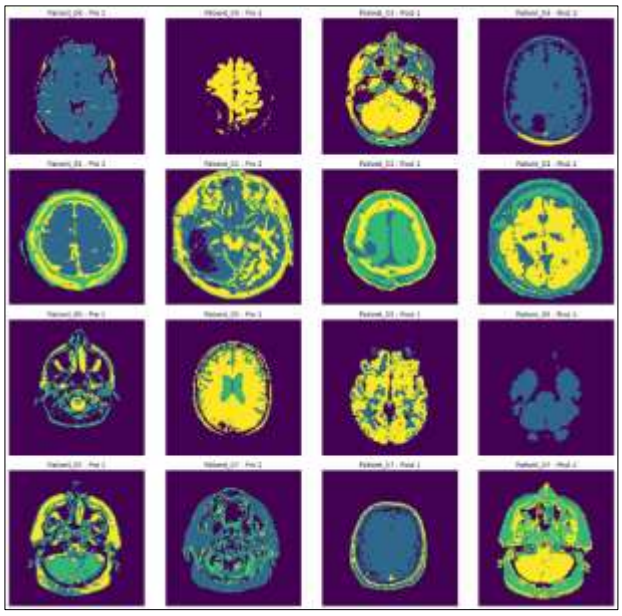

Figure 3. *GMM-based segmentation results on pre- and post-radiotherapy MR images for four randomly selected patients, showing intensity cluster classifications with a colormap of "viridis".*

3.1 Evaluation Metrics for Otsu's Segmented Images

The DSC of Otsu segmented pre and post radiotherapy images ranged between 0 and 1 as shown in figure 4. The DSC mean values for individual patients were also widely spread reflecting patient-specific segmentation performance differences. For example, pre- and post-segmented regions of Patient_09 and Patient_14 overlapped better and therefore had higher DSC while lower DSC (Patient_16 and Patient_18) indicate lower agreement or greater variability. Error bars illustrate the range of DSC scores for each patient, which demonstrates the variability in segmentation quality across the

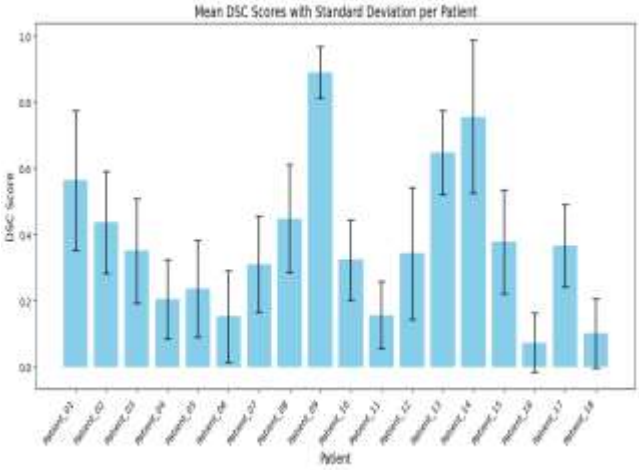

Figure 4. *Mean dice similarity coefficient scores (blue vertical bars) with standard deviations (black lines)*

dataset. This variation may correlate for example with tumor morphology, imaging quality, or segmentation difficulty. Figure 5 illustrates the Dice Similarity Coefficient (DSC) scores for various patients, comparing pre- and post-radiotherapy image pairs, and emphasizes the three patients with the highest and lowest scores. Patients exhibiting the highest DSC scores (marked in red, green, and yellow) show a consistently strong resemblance between pre- and post-treatment images with elevated scores. This suggests slight alterations in the images, implying that the treatment probably did not work. Conversely, patients who have the lowest DSC scores (marked in dark brown, grey, and black) demonstrate scores close to 0, indicating substantial variances between pre- and post-treatment images. This indicates that the treatment might have worked, given that significant changes took place. The other patients, shown by dashed lines, display moderate or fluctuating DSC scores, suggesting varied results. These results imply that DSC scores may serve as an important measure for assessing treatment efficacy, where lower scores reflect more significant alterations in tumor segmentation and possibly better treatment results.

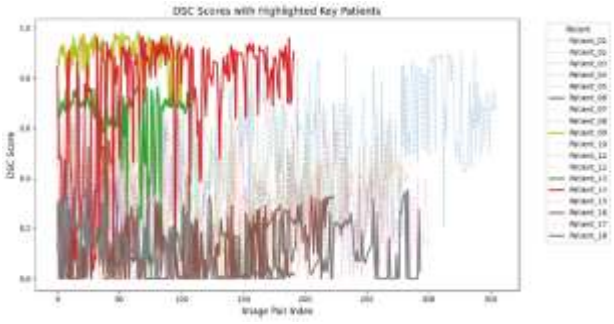

Figure 5. *Image pair index vs. DSC scores of pre- and post- images. Three patients with highest and lowest DSC scores are highlighted.*

The scatter plot in Figure 6 depicts the correlation between the mean and standard deviation of DSC scores for every patient, sorted by their mean score ranges. Patients exhibiting higher average DSC scores usually show reduced variability (standard deviation), indicated by the grouping of points. The color-coding clearly emphasizes score ranges, with distinct separations between groups evident. The detailed values of this plot are provided in Table 1. The bar chart illustrated in Figure 7 illustrates the average Intensity Distribution Scores (IDS) along with their standard deviations for every patient, indicating the differences in intensity distribution between segmented images taken before and after treatment. Patients_04 and Patient_17 exhibit the highest average IDS values of 0.46 and 0.33, respectively, with standard deviations of 0.15 and

Table 1. *Mean and standard deviation values of patients (Otsu's segmented) according to mean DSC scores*

Range	Patient	Mean	Std
0.25 - 0	Patient_04	0,205716	0,120245
0.25 - 0	Patient_05	0,236853	0,147213
0.25 - 0	Patient_06	0,153666	0,13822
0.25 - 0	Patient_11	0,157233	0,101912
0.25 - 0	Patient_16	0,073495	0,089431
0.25 - 0	Patient_18	0,102145	0,105265
0.5 - 0.25	Patient_02	0,438745	0,154307
0.5 - 0.25	Patient_03	0,351615	0,157635
0.5 - 0.25	Patient_07	0,310756	0,14558
0.5 - 0.25	Patient_08	0,44861	0,163759
0.5 - 0.25	Patient_10	0,324367	0,120712
0.5 - 0.25	Patient_12	0,344076	0,198819
0.5 - 0.25	Patient_15	0,378815	0,156853
0.5 - 0.25	Patient_17	0,366739	0,123899
0.75 - 0.5	Patient_01	0,564524	0,211738
0.75 - 0.5	Patient_13	0,648	0,12701
1 - 0.75	Patient_09	0,890691	0,078436
1 - 0.75	Patient_14	0,756669	0,232002

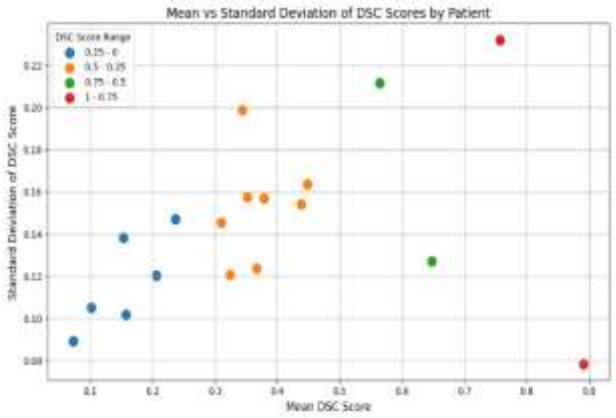

Figure 6. *Scatter plot of 18 patients illustrating the mean DSC scores (x-axis) versus standard deviations (y-axis).*

0.10, reflecting significant differences in intensity distribution. Conversely, Patient_09 and Patient_13 exhibit the lowest mean IDS values at 0.01 and 0.02, with minimal variations of 0.001 and 0.01, indicating slight intensity fluctuation. Patients like Patient_05 (average IDS: 0.23, standard deviation: 0.10) and Patient_15 (average IDS: 0.25, standard deviation: 0.08) exhibit moderate ratings, indicating intermediate intensity variations. The variation in deviations, exemplified by Patient_08 with a standard deviation of 0.19 versus Patient_12 at 0.06, underscores disparities in the consistency of intra-patient intensity distribution.

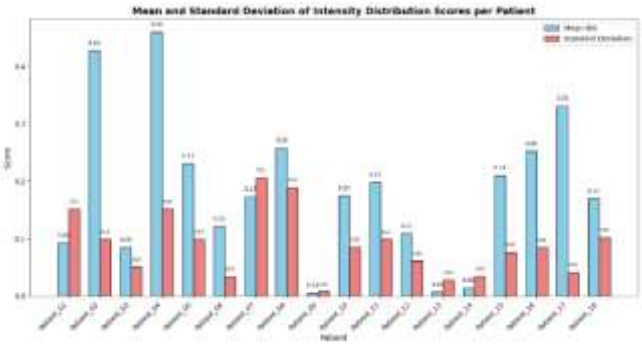

Figure 7. *Mean and standard deviation of IDS of each patient with blue bars representing mean values while red bars representing standard deviations.*

IDS of the 18 patients were examined, categorizing them into four unique score intervals: 0.01–0.12, 0.12–0.23, 0.23–0.35, and 0.35–0.46, which included 6, 7, 3, and 2 patients, respectively. The pie chart given in Figure 8 displays the percentage breakdown of patients across these score ranges, demonstrating that the largest group of patients (38.9%) is found in the 0.12–0.23 range, with 33.3% in the 0.01–0.12 range. Patients falling within the 0.23–0.35 and 0.35–0.46 ranges represent 16.7% and 11.1% of the cohort, respectively, suggesting that elevated IDS values are less frequent.

The scatter plot given in Figure 9 offers additional understanding, illustrating the connection between the mean IDS and standard deviation for each patient, grouped by score ranges. Patients displaying a higher average IDS, generally between 0.35 and 0.46, show comparatively less variability, indicated by reduced standard deviations. On the other hand, individuals in the lowest scoring range (0.01–0.12) show varying standard deviations, indicating variability in IDS patterns. This trend indicates that patients exhibiting higher IDS values show more uniform intensity distributions, whereas

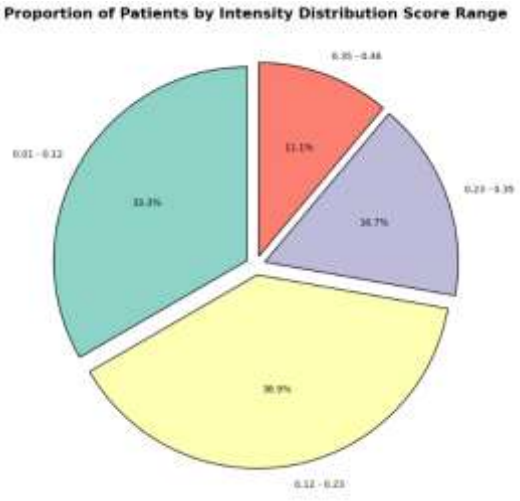

Figure 8. *Percentages of patients by IDS with specified subgroups*

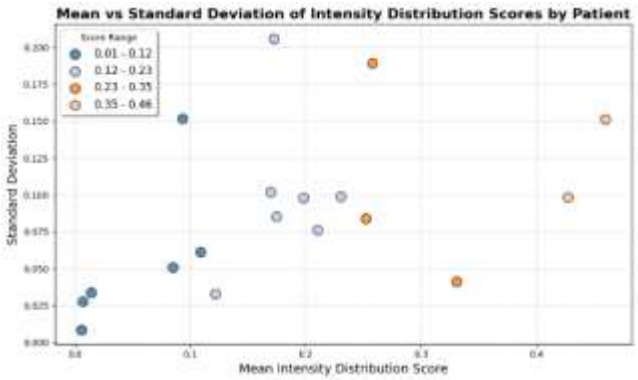

Figure 9. *Scatter plot of mean IDS of patients versus standard deviation.*

those with lower scores face increased variability. These visualizations offer an in-depth perspective on the IDS distribution among patients, highlighting intensity and variability patterns that may guide additional analyses of underlying conditions or segmentation effectiveness.

3.2 Evaluation Metrics for GMM Segmented Images

Using Gaussian Mixture Model (GMM) segmentation, the Dice Similarity Coefficient (DSC) scores for each patient are illustrated over a number of image pair indices in the Figure 10. With various colors representing the 18 patients, each point reflects the DSC score for a particular image pair of a particular patient. High structural similarity between pre- and post-therapy images is indicated by a majority of DSC scores being focused in the upper range (0.7–1.0). For certain patients or image pairings, this suggests that the segmented regions have only slight morphological alterations. Nonetheless, a sizable portion of ratings are below 0.5, with some approaching 0.0. major variations between pre- and post-therapy images are demonstrated by these lower scores, which may be a reflection of major therapy-induced alterations, tumor growth, or other anatomical differences. Patients' DSC scores demonstrated a variety of trends. Some, like Patient_02 and Patient_18, exhibit minimal variation between pre- and post-therapy states, continuously maintaining high DSC ratings throughout all image pairs. Conversely, patients with DSC values that range from great similarity to significant divergence, such as Patient_04 and Patient_11, exhibit significant variability. In many situations, this variety may indicate dynamic tumor activities or diverse therapeutic responses. Furthermore, certain patients—like Patient_07 and Patient_13—consistently exhibit lower DSC ratings across picture pairs, which may be a sign of severe therapeutic side effects or difficult imaging circumstances. There are also discernible temporal or sequential dependencies. Patient_16, for example, has a decreasing trend in DSC values over the image pair indices, which can indicate increasing alterations as treatment goes on. For certain pairings of images, the appearance of outliers with DSC scores close to 0.0 indicates significant variances between the pre- and post-therapy images. To find potential underlying causes, such as notable tumor shrinkage, new tumor growth, or imaging irregularities, these cases need detailed evaluation.

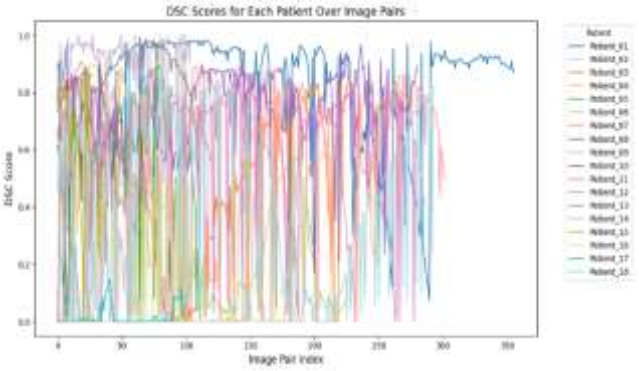

Figure 10. *DSC scores of GMM segmented pre- and post- therapy image pair indexes*

DSC scores for the pre- and post- radiotherapy images from each patient are shown in Figure 11. While Patients 01, 08, 09 and 10 show consistently high DSC scores, Patients 03, 04, 07 and 11 show limited variability. Patient 16 and 17 also have predominantly low scores, indicating significant changes. These variances embody distinct structural changes among various patients.

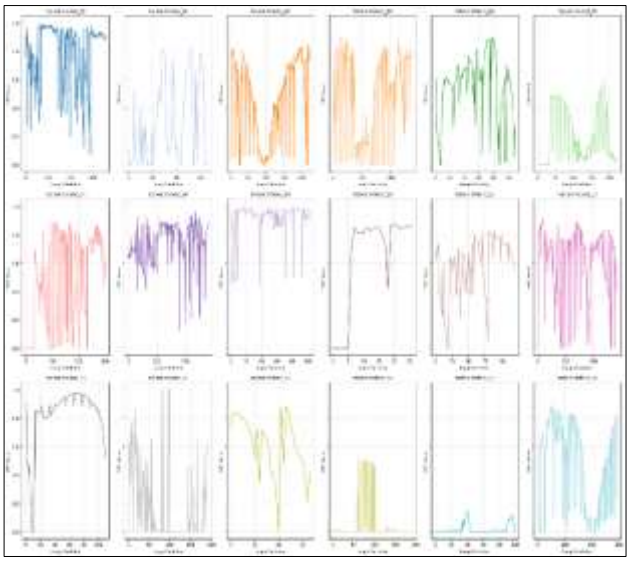

Figure 11. *DSC scores of GMM segmented image pair indexes for each patient separately.*

Figure 12 illustrates the percentage of patients categorized by average DSC score ranges (GMM). The majority of patients (38.9%) are within the 0.75–0.5 range, suggesting moderate similarity. Scores in the lower ranges (0.25–0 and 0.5–0.25) represent 22.2% each, whereas the highest range (1–0.75) comprises 16.7% of patients, indicating variability in segmentation results. The specific mean and standard deviation values for each patient is provided in Table 2.

Table 2. *Mean and standard deviation values of patients (GMM segmented) according to mean DSC scores*

Range	Patient	Mean	Std
0.25 - 0	Patient_06	0.220979	0.196961
0.25 - 0	Patient_14	0.174759	0.266865
0.25 - 0	Patient_16	0.064041	0.160468
0.25 - 0	Patient_17	0.017998	0.036674
0.5 - 25	Patient_02	0.296327	0.297387
0.5 - 0.25	Patient_03	0.453709	0.283054
0.5 - 0.25	Patient_07	0.494161	0.318920
0.5 - 0.25	Patient_18	0.480168	0.303567
0.75 - 0.5	Patient_04	0.510406	0.298773
0.75 - 0.5	Patient_05	0.547195	0.233237
0.75 - 0.5	Patient_08	0.731772	0.147899
0.75 - 0.5	Patient_10	0.630400	0.354381
0.75 - 0.5	Patient_11	0.575500	0.213715
0.75 - 0.5	Patient_12	0.626682	0.228277
0.75 - 0.5	Patient_15	0.609753	0.215147
1 - 0.75	Patient_01	0.782270	0.217551
1 - 0.75	Patient_09	0.901926	0.142500
1 - 0.75	Patient_13	0.817502	0.219824

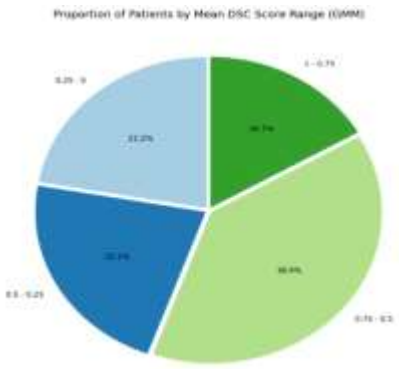

Figure 12. *Pie chart distribution of mean DSC scores for GMM segmented images*

The bar chart shown in Figure 13 illustrates the average and standard deviation of IDS for each patient, organized using GMM. Individuals like Patient_02, Patient_08, and Patient_15 show the highest average scores, suggesting stronger intensity distributions, whereas Patient_13 and Patient_16 display significantly lower average scores. The standard deviation, illustrated by the red bars, emphasizes the differences in scores among patients. In patients like Patient_08 and Patient_14, greater variability is noted, while patients like Patient_11 exhibit low variability. Overall, a clear positive correlation exists, where higher average values are frequently linked to larger standard deviations, indicating a wider dispersion in the scores for those patients. This examination offers insights into the distribution and uniformity of intensity scores within the patient group.

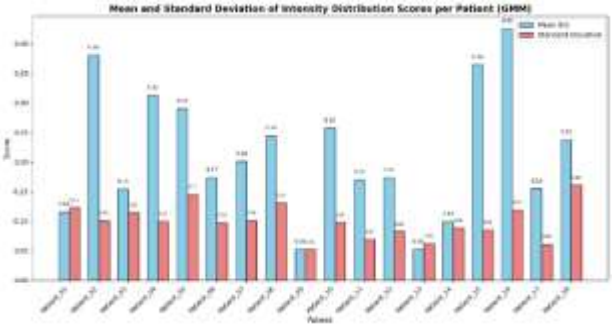

Figure 13. *Mean and standard deviation scores of IDS for GMM segmented images*

3.3 Comparison of RANO Scores with the findings

The assumption that low DSC values and high IDS values are associated with positive treatment outcomes (RANO 1 or RANO 2) and high DSC values and low IDS values are associated with negative treatment response (RANO 3 or RANO 4) was to be validated in the present study. This is done by comparing the segmentation-based predictions from both Otsu's thresholding and GMM methods with RANO score by an experienced radiologist which are provided with patient IDs in Table 3. (CR: Complete Response; PR: Partial Response; S: Stable; P: Progression)

Table 3. *RANO scores of each patient assessed by an experienced radiologist*

Patients	RANO
02, 05	1 - CR
06, 07, 08, 10, 15	2 - PR
01, 13, 14, 17, 18	3 - S
03, 04, 09, 11, 12, 16	4 - P

For the low DSC and high IDS which were expected to be RANO 1 or 2, Otsu's method identified seven patients, four of whom (57%) had matching outcomes. However, three patients (43%) did not align with this assumption, as they showed stable or progressive disease (RANO 3 or 4). On the other hand, GMM performed in predicting

responders with 50% of patients correctly classified as RANO 1 or 2. Both techniques correctly classified all patients in this group as stable or progressive (RANO 3 or 4), showing the validity of this pattern. All of these results are provided in Table 4.

Table 4. *The agreements of our findings with the RANO scores*

Segmentation Method	Low DSC & High IDS (Expected: RANO 1 or 2)	Actual RANO Score	Agreement	High DSC & Low IDS (Expected: RANO 3 or 4)	Actual RANO Score	Agreement
Otsu	Patient 6	RANO 2	✓	Patient 1	RANO 3	✓
	Patient 16	RANO 4	✗	Patient 9	RANO 4	✓
	Patient 4	RANO 4	✗	Patient 13	RANO 3	✓
	Patient 5	RANO 1	✓	Patient 14	RANO 3	✓
	Patient 11	RANO 2	✓	Patient 3	RANO 4	✓
	Patient 18	RANO 3	✗	Patient 12	RANO 4	✓
	Patient 2	RANO 1	✓			
GMM	Patient 6	RANO 2	✓	Patient 1	RANO 3	✓
	Patient 14	RANO 3	✗	Patient 13	RANO 3	✓
	Patient 17	RANO 3	✗	Patient 14	RANO 3	✓
	Patient 2	RANO 1	✓	Patient 3	RANO 4	✓
				Patient 12	RANO 4	✓

3.4 Discussions

Radiotherapy is one of the cornerstones treatment modalities for Glioblastoma Multiforme (GBM) and is used with the goal of reducing tumor size and prolonging survival. However, its efficacy is hampered by the highly heterogeneous infiltrative nature of GBM, which generally leads to incomplete tumor control and subsequent recurrence. Radiotherapy response is usually assessed based on their size and morphology changes on imaging (e.g. MRI). Though such assessments offer important insights, they are inherently subjective and may differ widely between clinicians, especially when determining the treatment effect versus tumor progression. Thus, there is an increasing demand for more accurate, measurable, reproducible ways to evaluate tumor dynamics, the manner in which tumor cells respond to radiotherapy and imaging tools with objective performance in this setting.

Artificial intelligence (AI)-driven segmentation techniques have emerged as a transformative solution to these challenges, capitalizing on their high accuracy and efficiency in processing complex imaging data [15,16]. In contrast to manual delineation, artificial intelligence–based methods, such as deep learning algorithms and automated clustering techniques, allow for both consistent and reproducible segmentation of GBM tumor regions. Manually difficult-to-recognize nuances radiological changes as pseudo-progression or necrosis, have been identified using convolutional neural networks (CNN) and hybrid approaches. Havaei et al. (2017) created a deep learning model with convolutional neural networks (CNNs) for automating brain tumor segmentation in multi-modal MRI, showcasing excellent accuracy and resilience across various datasets, emphasizing the promise for reliable and repeatable outcomes [17]. CNNs and hybrid methods have been utilized to detect manually challenging nuances in radiological changes, including pseudo-progression and necrosis. This is especially significant in GBM, where it is vital to differentiate between actual tumor advancement and changes associated with treatment for effective clinical management. In another study, Myronenko (2018) introduced a 3D U-Net model with variational autoencoder (VAE) regularization, focusing on the difficulty of segmenting necrotic areas and enhancing tumor core regions, demonstrating enhanced capability in distinguishing these intricate tissue types [18]. These studies showcase the capability of AI to enhance both segmentation precision and consistency, while also assisting in the difficult process of characterizing intricate tumor microenvironments. In addition, unsupervised approaches such as Otsu's

thresholding and Gaussian Mixture Models (GMMs) are promising for segmenting tumors without relying heavily on large annotated datasets, making them useful in clinical settings [19,20]. Bridging these autonomously derived layers of treatment response with the outcomes of radiotherapy will further enlighten the field and lead to more personalized intervention for patients with GBM.

The main objective of this research was to assess whether complex deep learning approaches are necessary to predict radiotherapy output with respect to Glioblastoma Multiforme (GBM), or if standard segmentation methods combined with mathematical and statistical metrics are sufficient. Tumor heterogeneity and infiltrative nature of GBM tumors further complicate segmentation process of tumor regions in pre- and post-radiotherapy MRI scans and hence, it is crucial for tumor segmentation accurate to assess the treatment efficacy. Despite the fact that deep learning approaches can unlock the state-of-the-art capabilities, they are resource-hungry and demand large amounts of labeled data, so the question remains whether simpler methodologies can provide similar interpretations.

In order to fully address this, both Otsu's thresholding and Gaussian Mixture model (GMM) segmentation, common segmentation techniques, were utilized in this study. This procedure was selected because of its simplicity and its non-reliance on extensive amounts of training data. The DSC and intensity distribution measures were computed for the segmented MRI data from both methods to quantify their agreement or discordance with the pre-treatment and post-treatment tumor regions. The hypothesis was that a high similarity between pre- and post-treatment images suggest a poor treatment efficacy whereas large differences indicate a successful reduction of the tumor. As such, this strategy gives us insight into how efficiently more traditional approaches can be applied to this task, perhaps making it a more tractable, representative alternative to deep learning applications in clinical settings.

The assessment of similarity or dissimilarity between pre- and post-radiotherapy images based on DSC and IDS relies on their quantitative measurements. DSC assesses the spatial correlation between two segmentations, with values from 0 to 1. A DSC value near 1 signifies high similarity, indicating that the tumor area has not altered much after treatment, which might suggest reduced treatment efficacy. In contrast, reduced DSC values indicate increased dissimilarity, implying notable alterations in tumor volume or structure, which may suggest successful radiotherapy. In the same way, IDS assesses the intensity-related variations in the segmented areas. A lower IDS value indicates greater similarity in intensity distributions, while higher IDS values reflect significant differences, which relate to alterations in tumor characteristics following treatment. By evaluating these metrics in segmented images taken before and after radiotherapy, one can quantitatively measure the extent of tumor response to the treatment. Successful treatment outcomes are linked to a combination of low DSC values, which indicate major structural changes, and high IDS values, which indicate notable intensity differences, according to an analysis of a particular patient's data. This method aids in locating instances when the tumor has been considerably damaged by radiation.

According to the Otsu method, patients who have low DSC scores are 4, 5, 6, 11, 16, and 18. Patients with medium DSC scores include 2, 3, 7, 8, 10, 12, 15, and 17, whereas high DSC scores were noted in patients 1, 13, 9, and 14. In the GMM method, the patients with low DSC scores are numbered 6, 14, 16, and 17. Patients 2, 3, 7, and 18 exhibit medium DSC scores, while those with high DSC scores are 4, 5, 8, 10, 11, 12, 15, 1, 9, and 13.

Concerning Otsu IDS scores, the patients with extremely low IDS scores (Q1) are 1, 3, 9, 12, 13, and 14. Individuals with low IDS scores (Q2) comprise 5, 6, 7, 10, 11, 15, and 18, whereas those with medium IDS scores (Q3) are represented by patients 8, 16, and 17. Patients 2 and 4 show high IDS scores (Q4). In the GMM method, patients 1, 9, 13, and 14 exhibit extremely low IDS scores (Q1). Patients 3, 6, 7, 11, 12, 17, and 18 show low IDS scores (Q2). Patients 4, 5, 8, and 10 are linked with medium IDS scores (Q3), while high IDS scores (Q4) are observed in patients 2, 15, and 16.

Individuals with low DSC scores and high IDS scores might demonstrate the capability for effective therapy. In the Otsu method, the patients who exhibited low DSC scores are 4, 5, 6, 11, 16, and 18, whereas those with high IDS values are 2 and 4. In the same way, according to the GMM approach, patients with low DSC scores are 6, 14, 16, and 17, while those with high IDS scores are 2, 15, and 16. These patients might be linked to the successful treatment results due to the combination of these evaluation metrics.

Patients exhibiting high DSC values and low IDS values for both Otsu and GMM segmentation techniques might suggest that the treatment was ineffective. In the Otsu method, patients exhibiting high DSC values are 1, 13, 9, and 14, whereas those having low IDS values are 1, 3, 9, 12, 13, and 14. In the same way, for the GMM approach, high DSC

values are noted in patients 4, 5, 8, 10, 11, 12, 15, 1, 9, and 13, whereas low IDS values are seen in patients 1, 9, 13, and 14. The rationale for this is that high DSC values typically indicate a strong resemblance between images taken before and after treatment, implying that the tumor or impacted region may not have noticeably altered or reacted to the therapy. Moreover, low IDS values usually suggest minimal variation or alteration in the images, supporting the notion that the treatment did not significantly affect the tumor or lesion. In such scenarios, the significant dissimilarity noted with high DSC values combined with low IDS values suggests that the treatment may not have been as effective as anticipated, showing little change or reduction in the tumor's features after radiotherapy.

According to the results of comparison this study's findings with the actual RANO scores demonstrated that the two segmentation approaches had comparable performance in predicting cases of RANO 1 and RANO 2, and detected patients with poor radiotherapy response. Collectively, these observations suggest that high DSC and low IDS values correlate with good RANO scores and support the ability of DSC and IDS to detect non-responders early so that treatment strategies can be adapted accordingly. For RANO 1 and RANO 2 patients, respectively, it is found that Otsu's method had higher accuracy in predicting responders, correctly identifying 57% (4 out of 7) of responders, versus 43% (3 out of 7) than GMM. To enhance treatment responder detection, it may be beneficial to incorporate deep learning-based segmentation, radiomic features, or advanced intensity-based parameters. These results emphasize that the assessment of the radiotherapy response in GBM will benefit from new methods and larger datasets.

The small sample size is the major limitation of this study. While the dataset includes pre- and post-radiotherapy MRI scans from 18 patients, this small cohort may not reflect the heterogeneity and variation found in the field of GBM. This tumor displays heterogeneity in behavior, response to therapy, and patient outcomes. A larger dataset would have granted us much better statistical power, better exploration of the significance and highlighting of the differences between segmentation methods, as well as a finer exploring the relationship between DSC and IDS scores and treatment response. Broadening the patient extension across a wider range of GBM presentations and therapy response may also enhance the generalizability of the findings.

Another major challenge is the dependence on retrospective data. Although these types of studies are useful, they can often lack data completeness and standardization. Differences in MRI acquisition protocols, imaging quality, and methods of performing radiotherapy between institutions may lead to variations in the analysis. Segmentation accuracy may be susceptible to differences in the resolution and positioning of MRI sequences, especially for more complex methods. Resolving these would either necessitate the standardization of imaging protocols or supplementary integration of sophisticated image preprocessing methods to standardize the data.

Additionally, the research focused exclusively on Otsu's and GMM segmentation methods, both of which are valuable but do not encompass all the segmentation algorithms currently available. More sophisticated techniques, like segmentation utilizing deep learning, could also enhance sensitivity to alterations in low-grade tumors. It's important to mention that these methods typically require more data for both training and validation, leading to the need for a larger dataset. Furthermore, the lack of ground truth annotations, such as human-segmented masks provided by clinical experts, complicates the direct evaluation of the performance of these methods. Utilizing this ground truth data in upcoming research will allow for more precise validation of the employed segmentation techniques.

The shortcomings of current research emphasize the need for future studies to include larger and more varied datasets, consistency in imaging techniques, and investigate various segmentation methods to enhance both the predictive power and dependability of treatment outcome assessments in GBM.

4. Conclusions

In this study, we examined GBM tumor dynamics using pre- and post- radiotherapy MRI data with Otsu and GMM segmentation methods. Patients with low DSC and high IDS values predicted as significant tumor response to treatment, highlighting these metrics as indicators of therapeutic efficacy. Through validation by comparing the results of this study with actual RANO scores, it was determined that the Otsu method demonstrates superior accuracy in estimating complete and partial responses compared to the GMM method. However, both methods exhibit comparable performance in identifying stable disease and progression responses. Although promising, the study's small sample size highlights the necessity for additional research with larger data to confirm these findings. Developing such further researches could improve

predictive modeling and aid in creating more tailored treatment strategies for GBM.

Author Statements:

- **Ethical approval:** This study was performed as a retrospective analysis and approved by the ethical committee of Biruni University Hospital (2024-BİAEK/04-48).
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** S. Unalan: methodology, coding, writing; D. Karacetin: data acquisition, data interpretation, review; O. Gunay: data acquisition, reviewing; U. Tuzun: RANO evaluation, data interpretation.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

References

[1] Jovčevska, I., Kočevar, N., & Komel, R. (2013). Glioma and glioblastoma-how much do we (not) know?. *Molecular and clinical oncology*, 1(6), 935-941.

[2] Urbańska, K., Sokołowska, J., Szmidt, M., & Sysa, P. (2014). Glioblastoma multiforme – an overview. *Contemporary Oncology/Współczesna Onkologia,* 18(5), 307-312. https://doi.org/10.5114/wo.2014.40559

[3] Roth, J. G., & Elvidge, A. R. (1960). Glioblastoma multiforme: a clinical survey. *Journal of neurosurgery*, 17(4), 736-750.

[4] Stupp, R., Mason, W. P., Van Den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J., ... & Mirimanoff, R. O. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *New England journal of medicine*, 352(10), 987-996.

[5] Ylanan, A. M. D., Pascual, J. S. G., Cruz-Lim, E. M. D., Ignacio, K. H. D., Cañal, J. P. A., & Khu, K. J. O. (2021). Intraoperative radiotherapy for glioblastoma: A systematic review of techniques and outcomes. *Journal of Clinical Neuroscience*, 93, 36-41.

[6] Wu, C. X., Lin, G. S., Lin, Z. X., Zhang, J. D., Liu, S. Y., & Zhou, C. F. (2015). Peritumoral edema shown by MRI predicts poor clinical outcome in glioblastoma. *World journal of surgical oncology*, 13, 1-9.

[7] Young, R. J., Gupta, A., Shah, A. D., Graber, J. J., Zhang, Z., Shi, W., ... & Omuro, A. M. P. (2011). Potential utility of conventional MRI signs in diagnosing pseudoprogression in glioblastoma. *Neurology*, 76(22), 1918-1924.

[8] Sengul, A., Ünalan, S., Can, S., Gunay, O., Karaçetin, D., & Sayyed, M. I. (2024). Volumetric rigid MR-CT registration for glioblastoma in radiation oncology: A Novel approach. *Journal of Radiation Research and Applied Sciences*, 17(1), 100798.

[9] Zlochower, A., Chow, D. S., Chang, P., Khatri, D., Boockvar, J. A., & Filippi, C. G. (2020). Deep learning AI applications in the imaging of glioma. *Topics in Magnetic Resonance Imaging*, 29(2), 115-00.

[10] Zhang, J., & Hu, J. (2008, December). Image segmentation based on 2D Otsu method with histogram analysis. *In 2008 international conference on computer science and software engineering (Vol. 6, pp. 105-108). IEEE.*

[11] Reynolds, D. A. (2009). Gaussian mixture models. *Encyclopedia of biometrics*, 741(659-663).

[12] Thada, V., & Jaglan, V. (2013). Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm. *International Journal of Innovations in Engineering and Technology*, 2(4), 202-205.

[13] Jacobsen, N., Deistung, A., Timmann, D., Goericke, S. L., Reichenbach, J. R., & Güllmar, D. (2019). Analysis of intensity normalization for optimal segmentation performance of a fully convolutional neural network. *Zeitschrift für Medizinische Physik*, 29(2), 128-138.

[14] Gaillard F, Glick Y, Campos A, et al. RANO criteria for glioma. *Reference article, Radiopaedia.org (Accessed on 23 Feb 2025)* https://doi.org/10.53347/rID-29155

[15] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: *18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241)*. Springer International Publishing.

[16] Milletari, F., Navab, N., & Ahmadi, S. A. (2016, October). V-net: Fully convolutional neural networks for volumetric medical image segmentation. *In 2016 fourth international conference on 3D vision (3DV)* (pp. 565-571). Ieee.

[17] Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., Pal, C., Jodoin, P. M., & Larochelle, H. (2017). Brain tumor segmentation with Deep Neural Networks. *Medical image analysis*, 35, 18–31. https://doi.org/10.1016/j.media.2016.05.004

[18] Myronenko, A. (2019). 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds) Brainlesion:

Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2018. *Lecture Notes in Computer Science(), vol 11384. Springer, Cham*. https://doi.org/10.1007/978-3-030-11726-9_28

[19] Ahmed, M. N., Yamany, S. M., Mohamed, N., Farag, A. A., & Moriarty, T. (2002). A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE transactions on medical imaging*, 21(3), 193–199. https://doi.org/10.1109/42.996338

[20] Li, C.H., & Tam, P.K. (1998). An iterative algorithm for minimum cross entropy thresholding. *Pattern Recognit. Lett.*, 19, 771-776.