**APPLICATIONS OF DATA SCIENCE**

**BRAIN TUMOR DETECTION**

**AND**

**CLASSIFICATION**

**Submitted By: LASHIKA ARUNACHALAM**

**Student ID: ARU21537357**

**TABLE OF CONTENT**

[**1. Abstract 3**](#_Toc118972274)

[**2. Introduction 4**](#_Toc118972275)

[**3. Methodology 5**](#_Toc118972276)

[**4. Output 9**](#_Toc118972277)

[**5. Reference 7**](#_Toc118972278)

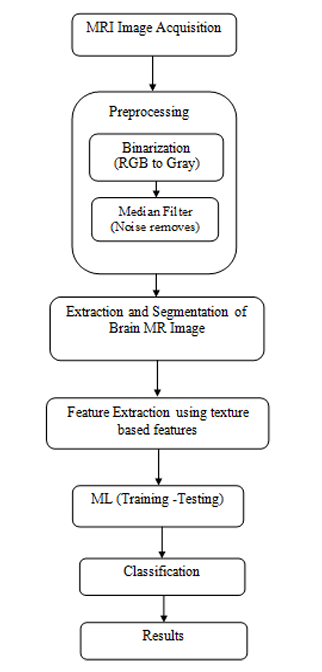
# **Abstract:**

In multiple healthcare evaluation applications, autonomous flaw identification in medical imaging has emerged as a new field. Because of the intricacy and variety of tumors that have not yet been thoroughly characterized, detecting brain tumors is a challenging task. We are currently seeing a significant increase in brain tumor cases. In order to recognize separate tumors from the MRI pictures, developed an automated system. The three-part strategy consists of applying pre-processing techniques to brain magnetic resonance imaging scans, extracting features via Otsu binarization and the gray-level co-occurrence matrix (GLCM), and segmenting the results with K means clustering. Next, identification is performed by the machine learning approach SVM. A dataset of 150 MRI pictures was used for testing the suggested approach. This collection included 30 photographs of a healthy brain, a total of 60 malignant tumors, and 50 MRIs of benign tumors. The findings demonstrate that the suggested approach is capable of correctly recognizing and categorizing carcinomas of the brain, making it a valuable tool for medical practitioners in the detection and management of brain tumors. The suggested approach can also help surgeons diagnose brain tumors more accurately and quickly by speeding up analysis.

# **Introduction:**

There are numerous different cell kinds in human beings. Cells in the body divide and expand in an organized fashion to produce new cells. The aforementioned newly formed cells contribute to the maintenance of the individual’s body functionality and wellness. Certain cell types develop chaotically because they miss their capacity to regulate their own development. Tumors are created when excess cells combine to form a mass of tissue. Tumors may be benign or cancerous. Though benign tumors do not cause carcinoma, malignant tumors do. A report released by the centralized brain tumor registry of the United States (CBTRUS) claims that in 2002, both types of brain tumors had been identified in about 39,550 patients. It suggests that if the prevalence of initial brain tumors is a malignant tumor or benign tumor is 14 per 100,000. Healthcare images gathered by different surgical instruments that employ imaging methods like X-ray, CT scan, and MRI are a crucial component of the wellness diagnostic. The human body's water ions contain atomic quantities of hydrogen, with an MRI procedure, which relies on the measurement of magnetic field vectors that are produced following the proper activation of these atoms by powerful magnets and RF pulses. Since the MRI scan uses minimal radiation, it is far more effective for identification compared to a computed tomography scan. Utilizing MRI, surgeons may assess the brain. The existence of a tumor in the brain may be detected using MRI technology. The individual examination is the traditional technique for finding tumors in MRI images. This process takes a long time. Massive quantities of knowledge are not acceptable. Additionally, noise introduced by human interaction in the MRI can result in erroneous categorization. Since there is a large number of MRI data to evaluate, automated techniques are required for reasons that are more efficient. Automatic use of tumor diagnosis in an MRI scan is required since dealing with human life requires an excessive amount of precision. The use of machine learning algorithms is presented as a powerful automated categorization method for brain MRI. The classification of brain MR images using a supervised machine learning technique.

# **Methodology:**

****

# **Image Acquisition:**

Healthcare diagnostic instruments such as computed tomography (CT) scanning as well as magnetic resonance imaging (MRI) probes are utilized for collecting pictures of the brain in order for the identification of brain tumors. These gadgets provide visuals that can aid in the detection and diagnosis of brain tumors using multiple biological concepts. A powerful magnetic field and radio waves are used in MRI to provide precise pictures of the brain. It can identify various forms of tissue based on their magnetic characteristics, making it possible to find tumors along with additional disorders. X-rays are used in CT scans to provide cross-sectional images of the brain. By monitoring the X-ray absorbance by various brain cells, it can find tumors and other irregularities. The Dataset consists of 150 images from various sources such as Kaggle and various websites.

# **Pre-Processing:**

Pre-processing is required because it improves the image information and highlights some of the aspects that are crucial for subsequent processing. The MIR vision is subjected to the following pre-processing steps: To remove noise from MRI scans of the brain, a Color image is first transformed into a grayscale image and then the median filter is used. For subsequent processing, the noise must be eliminated because great precision is required.

The initial stage in image analysis for brain tumor identification is to turn the acquired image into a grayscale image.  By removing the color features from a color image and keeping only the luminance or brightness data, a grayscale image can be created. This procedure is typically carried out on pictures obtained utilizing CT or MRI scanners in diagnostic imaging.

In hopes of minimizing unwanted noise, median filtering—a frequent data pre-processing technique in image enhancement be utilized in the identification of brain tumors. To get rid of further salt-and-pepper disturbance that might be visible in the picture after translating it to grayscale, median filtering can be used with a square filter among a specific range.

# **Extraction and Segmentation:**

K-means grouping is employed for categorization. A number of the most straightforward methods used to address clustering issues is K-means. A predetermined number of clusters are used in the technique to classify a particular collection of data. k centroids should be defined, a single for every cluster, as the main concept. The separation among these coordinates is as great as feasible. The next step is to take each point from a particular data collection and connect it to the closest centroid. If there aren't any open points, the initial step is finished, and an early groupage is finished. At the barycentre of the clusters produced in the previous phase, k unique centroids are now generated.

The elimination of the skeleton is a crucial process in the diagnosis of brain tumors, and median filtering can be applied as part of this phase to further enhance the picture's clarity. A binary mask is made to segment the skull and remove it from the image after the image has been converted to grayscale and has undergone median filtering to lessen noise.

Thresholding methods is used to produce a binary mask. Using Otsu's method to automatically identify the ideal threshold level. It is transformed into a binary image using Otsu binarization. The procedure generates the best threshold dividing the two groups such that their combined spread (intra-class variance) is as little as possible. It presupposes that the picture to be threshold comprises two classifications of pixels or a bi-modal histogram. In order to separate elements from actual backgrounds when analyzing images, an appropriate threshold of grey level must be chosen. After the binary mask has been constructed, the skull can be eliminated utilizing morphological procedures erosion.

A post-processing procedure called "hole filling" is performed to fix tiny gaps in the segmented image. Using a morphological method known as "fill holes" which occupies any connected components that are totally encompassed by surrounding pixels and do not have any border touching the foreground—accomplishes this. Erosion is used to carry out this surgery and find whether the tumor is present or not.

# **Feature Extraction:**

If an algorithm's intake is too big and inconsistent to analyze, a feature smaller, more accurate set of features—is created. The process of extracting features from a collection of input information is known as feature extraction. The crucial features required for image categorization are collected in this phase. The employed segmented MIR image is utilized to identify texture characteristics that display the texture property of the image. The Gray Level Co-occurrence Matrix, a reliable and effective approach, is used to retrieve those properties.

The GLCM region - based segmentation approach is extremely economical since it minimises the volume of the GLCM, hence lowers the computational cost of the algorithm while maintaining good categorization levels. To identify between a healthy and unhealthy brain, the GLCM traits are utilized. Crucial facts regarding the surface phase structure are contained in texturing. Graystone spatially dependencies-based textural characteristics are generally applicable to picture categorization.

# **Classification:**

If the supervised machine learning algorithm Support Vector Machine (SVM) is frequently applied to classification issues. SVM is utilized to categorise if a picture is normal or contains a tumor in the identification of brain tumors. The SVM method finds the ideal subset of features that effectively divides the two classes of data with the greatest possible range. Support vectors, which are the data points closest to the hyperplane, are very important in pinpointing where the hyperplane is.

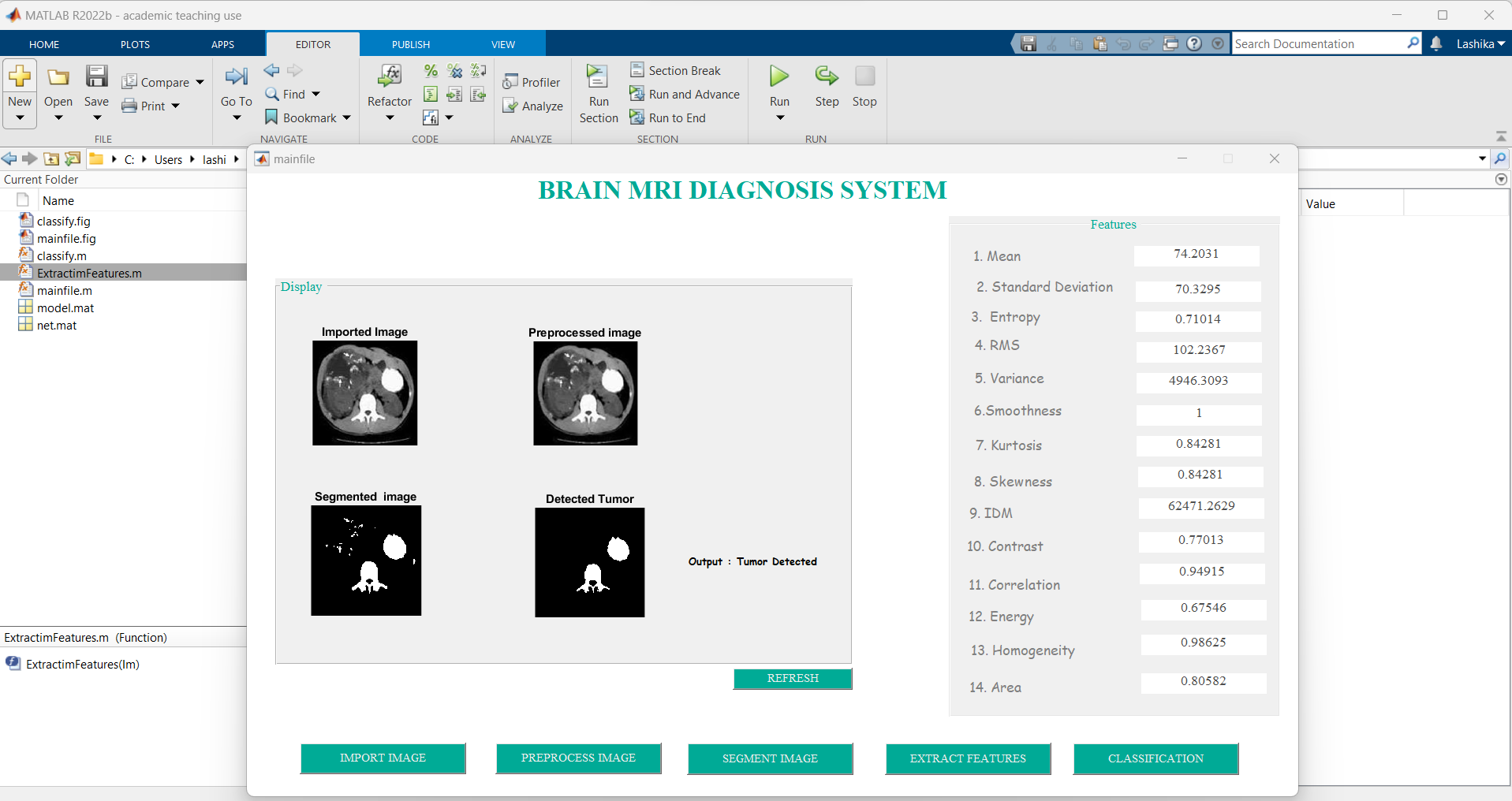
A kernel function is used in SVM to reshape the given data into a higher subspace. The best hyperplane to divide the two classes of data is then discovered using the altered input. The SVM algorithm's effectiveness may be significantly impacted by the kernel function that is selected. Kernels with linear, polynomial, and radial basis functions are frequently utilized.

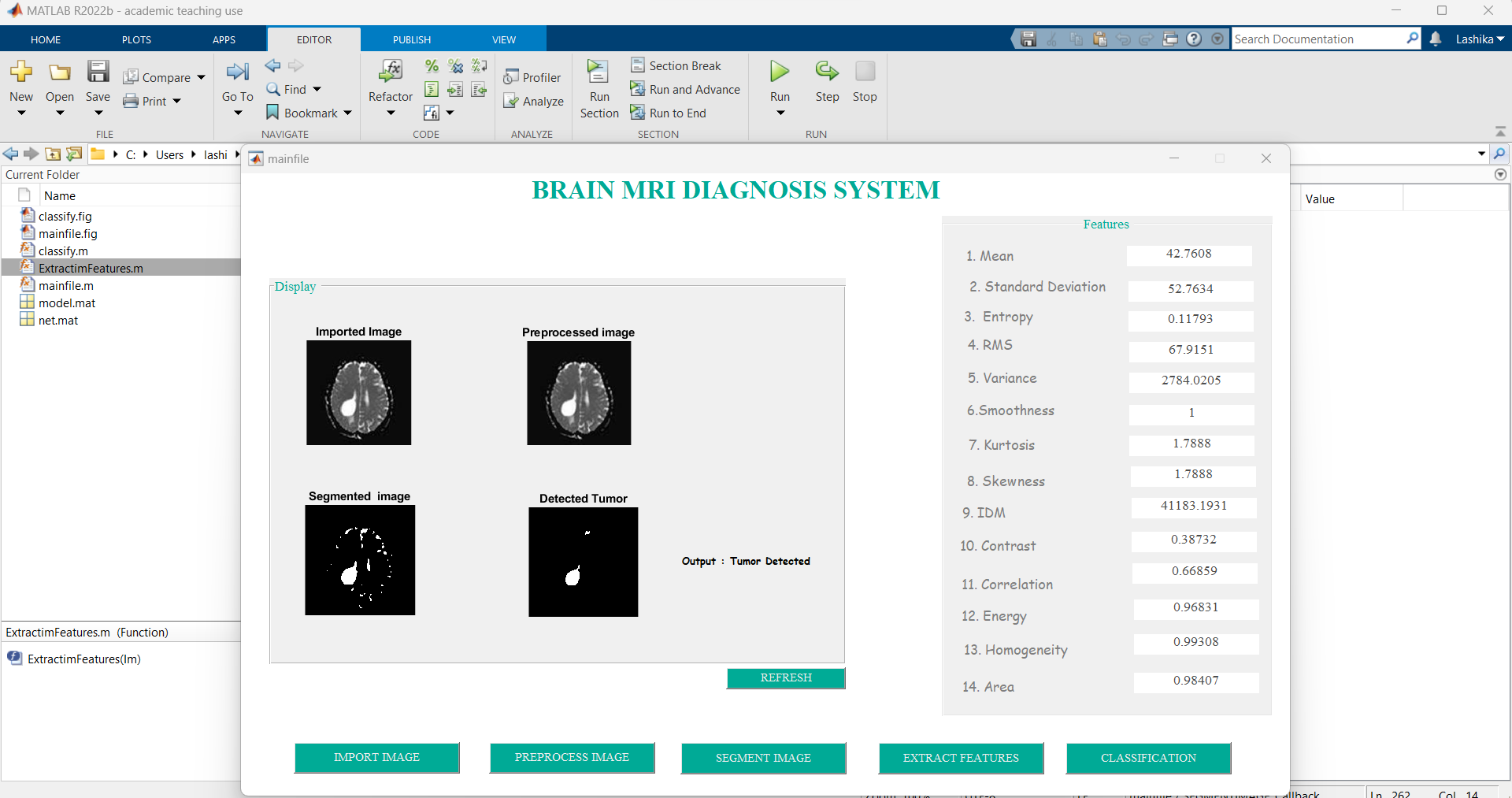
The SVM method finds the ideal hyperplane that divides the two classes of data with the most gap before and after the building phase. After becoming familiar with the hyperplane, one may employ it to categorize fresh datasets by identifying which side of the hyperplane they fall on. The SVM algorithm is trained using a set of tagged photos with the purpose of finding brain tumors. The system is trained to differentiate between benign and malignant images using the tagged pictures. The SVM classifier is to categorize new brain pictures as benign or malignant after it has been developed.

# **Output:**

When mainfile.m is made run, it opens the GUI window to detect whether tumor is present or not. The steps involved are:

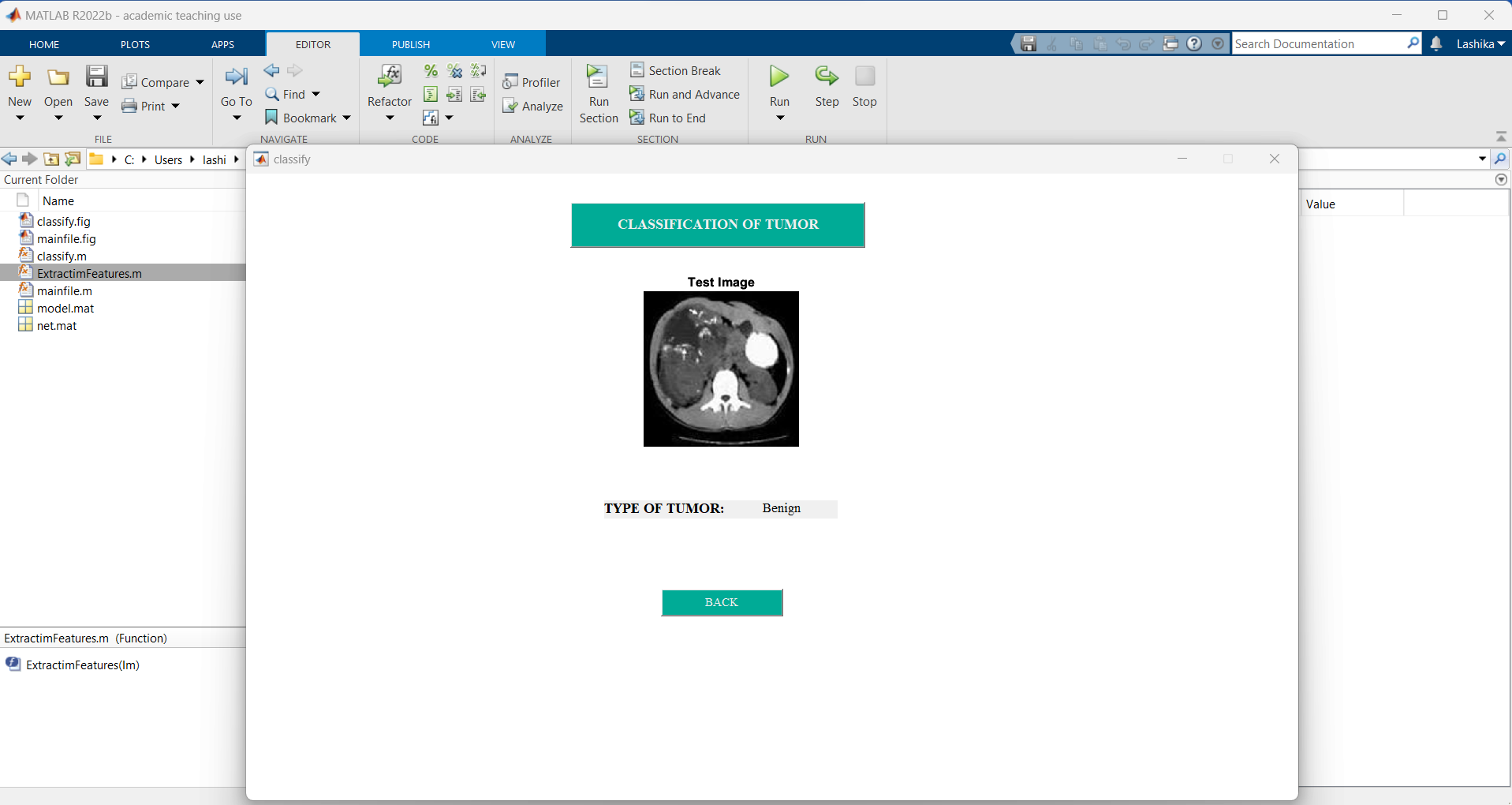
* Select Import Image button to select the image from the folder.
* Select Pre-process button to do the pre-processing operations such as conversion to a grayscale image and applying the median filter.
* Select the Segment image button for performing K means algorithm to find the area, skull removal by applying median filtering, and performing hole filling by using morphological operation Erosion.
* Select Extract Features to calculate and display Mean, Standard Deviation, Entropy, RMS, Variance, Smoothness, Kurtosis, Skewness, Correlation, Energy, Homogeneity, and Area.

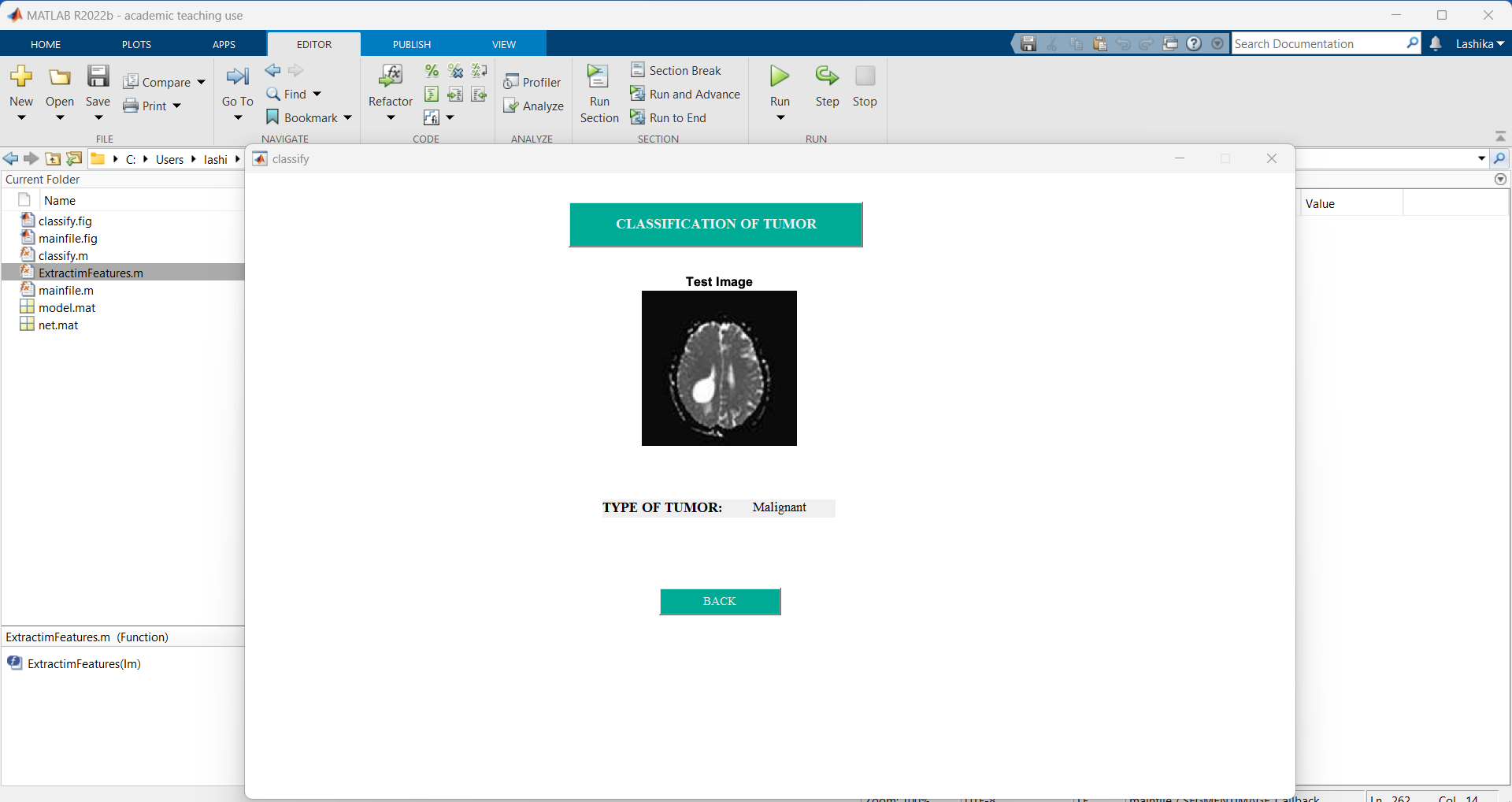




Then Select Classification button to check which type of tumor is it.

* Select Classification button in the window.
* Then Select the Classification of Tumor button in the next window to display the type of tumor.





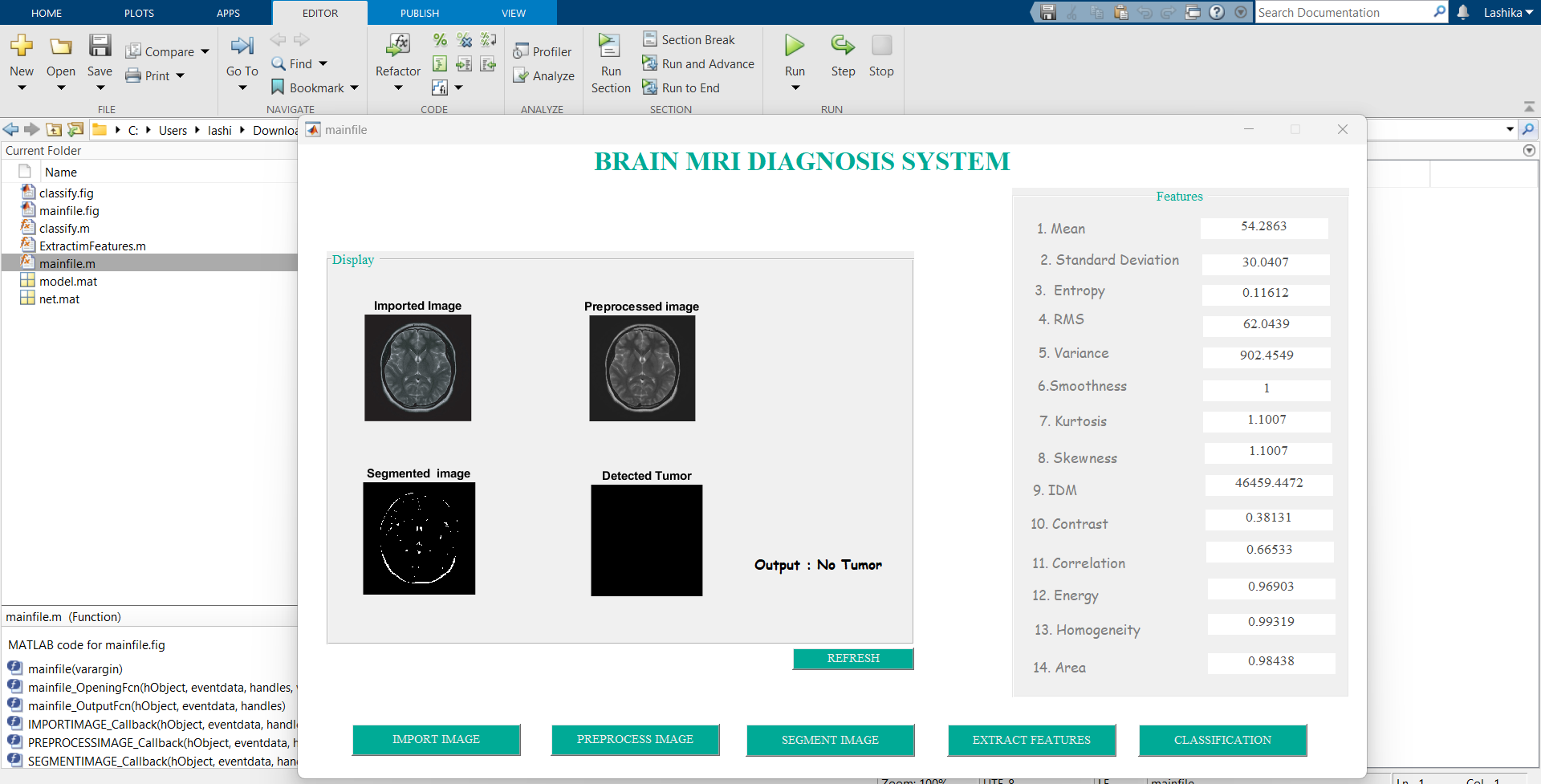


Image: Result showing No Tumor when image is imported.

# **Conclusion:**

# The identification and categorization of brain tumors using a combination of SVM and K-means clustering has yielded superior outcomes. The precision of the categorization has been greatly enhanced by the pre-processing methods applied, including median filtering, skull removal, and morphological operations with hole filling. The tool is easier to use because of the GUI execution, which offers medical professionals a simple user interface. To increase the sensitivity and specificity of the method for the precise identification of brain tumors, additional study and optimization are required. Considered, this method provides a useful tool for aiding in the early detection and diagnosis of brain tumors.

# **Reference:**

**For Pre-Processing:**

1. V. Singhal, A. Kumar, A. Kumar, and S. S. Bedi, "Brain Tumor Detection and Classification using SVM Classifier and Image Processing Techniques," in 2019 2nd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 477-481.
2. Rundo, L., Tangherloni, A., Mauri, G., & Cazzaniga, P. (2019). Pre-processing techniques for medical image analysis. Current Medical Imaging, 15(6), 589-604.
3. Rajan, J., & Ravi, S. (2019). Pre-processing techniques in medical image analysis. In Proceedings of International Conference on Intelligent Computing and Applications (pp. 117-128). Springer.

**For Segmentation:**

1. A. S. Kulkarni and S. P. Shinde, "Brain Tumor Classification using SVM and KNN Classifier," in 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), Pune, India, 2018, pp. 1-6.
2. S. Rana, S. Kumar, and V. K. Singh, "A Review on Brain Tumor Detection and Classification using Machine Learning Techniques," International Journal of Computer Applications, vol. 170, no. 10, pp. 9-14, Jul. 2017.

**For Classification:**

1. S. K. Arora, M. Bhatia, and N. Jain, "Brain Tumor Classification using Hybrid Approach of SVM and Artificial Neural Network," in 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, pp. 1-6.
2. T. Shehata, A. H. Farag, and A. S. Eldeib, "Brain Tumor Detection and Classification Using Multi-Level Features and Support Vector Machines," Journal of Medical Systems, vol. 40, no. 10, pp. 211-217, Sep. 2016.

**For MATLAB:**

1. MathWorks website: <https://www.mathworks.com/>
2. MATLAB Documentation: <https://www.mathworks.com/help/matlab/>
3. MATLAB Tutorials: <https://www.mathworks.com/learn/tutorials/matlab-onramp.html>
4. MATLAB Examples: <https://www.mathworks.com/help/matlab/examples.html>

**YouTube Videos:**

1. MathWorks website: <https://youtu.be/UWdd-_plgK4>

https://www.youtube.com/watch?v=aSHWPiVi8GU

Top of Form