Classification of Brain Tumor Based On Data Mining Techniques

Aniket Sharma, Mrinal Jadhav, Mansi Kharya, Priyanka Gondhale

[1] *Student, Information Technology, Jayawantrao Sawant College of Engineering, Maharashtra,India*
[2] *Student, Information Technology, Jayawantrao Sawant College of Engineering, Maharashtra,India*
[3] *Student, Information Technology, Jayawantrao Sawant College of Engineering, Maharashtra,India*
[4] *Student, Information Technology, Jayawantrao Sawant College of Engineering, Maharashtra,India*

ABSTRACT

Image processing techniques are most widely used in medical imaging to identify the affected area through X-ray, CT scan, MRI scan. Image processing is used to detect and identify the inner most portion of the human body. Here we have focused on the brain tumor detection techniques. Major brain tumors are not diagnosed until after symptoms appear. The main objective of the work is to explore various techniques to detect brain tumor using MRI in an efficient way. It has been found that the most of existing methods has ignored the poor quality images like images with noise or poor brightness. Also, tumor detections using MRI images are a challenging task, due to the complex structure of the brain. The methodology includes image pre-processing, image segmentation, the detection and clustering of the tumor zone based on morphological operations. Using K-Means Clustering Algorithm and classification of tumor on basis of severity using Naïve Bayes Classifier.The tumor is extracted from MRI image and its exact shape and position are determined. The amount of area calculated from the cluster is used to display the type of the tumor.

Keyword : - *Pre-processing, Segmentation, Clustering, Classification, Brain Tumor, K-Means Clustering,Naïve Bayes Classifier.*

1. INTRODUCTION

Nowadays, brain tumor is one the main reason for increasing mortality among adults and kids. It's been concluded in the research of the majority of the western world that number of individuals who are suffering and dying from brain tumors continues to be increased to 300 a year during past few decades [1]. More than 612,000 Americans will likely be identified as having a brain tumor and also over millions of people in the United States are living with brain tumors that have yet to be detected [based on Nation's Brain Tumor Foundation (NBTF) and American Brain Tumor Association (ABTA)]. As this number is candidate to increase, tools and methods to detect, extract the tumors and also to analyze their behavior are increasingly widespread and must consider the type of tumor, the kind of images to be utilized and depending there from the several approaches to use or develop. The proposed algorithm is to detect the troubleness and filter the affected area of brain. In this paper we can detect the tumor with visualization, preprocessing, segmentation, clustering, classification operations. From the MRI images the information such as tumors location provided by radiologist, an easy way to diagnose the tumor and plan the surgical approach for its removal. Normally the Brain can be viewed by the MRI scan or CT scan. In this paper the MRI scanned image is taken for the entire process. The MRI scan is safer than CT scan for diagnosis. It will not affect the human body. Because it doesn't use any radiation. It is based on the magnetic field and radio waves. There are various types of algorithm were developed for brain tumor detection. But they may have some drawback in detection and clustering. CT scan or MRI that is directed into intracranial cavity produces a complete image of a brain. This image is visually examined by the physician for detection and diagnosis of brain tumor. However, this method of detection resists the accurate determination of stage and size of a tumor. In this paper, a process for brain tumor classification, depending on the analysis of Magnetic Resonance (MR) images and Magnetic Resonance Spectroscopy (MRS) data collected for patients with benign and malignant tumors will be proposed. The aim is to achieve a high accuracy in discriminating the two types of tumors through a combination of several techniques for image de-noising, image segmentation, feature clustering and classification.MRI images of brain are widely used nowadays to predict tumor and analyzing the risk level of the tumor as normal or benign or malignant. The images

that are generated using the MRI images are not very clear for diagnosis due to the presence of noise. The noises are added to the image due to various reasons such as, the conductivity of an object that is subjected to scan and imaging (inductive losses); thermal noises are added to MR images from the conductivity of system's hardware. Especially, brain images of MRI suffer from low Signal-to-noise Ratio (SNR) and Contrast-to-noise (CNR). These issues become an obstacle for effective image analysis, feature clustering, and quantitative measurements. Therefore, it becomes mandatory to remove the noise in order to improve the quality of the MR image and other imaging modalities. This process acts as the preprocessing step for predicting the risk level of tumor in a scanned image. Denoising is a crucial issue since there are maximum possibilities for removal of required information along with the noise. Noise removal can be carried out using various filtering techniques. A good and efficient noise removal technique should remove maximum noise as well as it should retain the important feature as much as possible. Plenty of filtering methods are available to denoise the image through making a determination of exact version of pixels. Therefore, noises are removed depending on the requirement using an appropriate filter. The main objective of the proposed technique of this paper is to predict the risk level of brain tumor from an MRI image.

1.1 TYPES OF BRAIN TUMOR

There are several types of brain tumor it can be classified as benign and malignant (non-cancerous).

1) Types of Tumor:

There are three common types of tumor: *1)* Benign; *2)* Pre-Malignant; *3)* Malignant (cancer can only be malignant).

a) *Benign Tumor:* A benign tumor is a tumor is the one that does not expand in an abrupt way; it doesn't affect its neighboring healthy tissues and also does not expand to non-adjacent tissues. Moles are the common example of benign tumors.

b) *Pre-Malignant Tumor:*Premalignant Tumor is a precancerous stage, considered as a disease, if not properly treated it may lead to cancer.

c) *Malignant Tumor:* Malignancy is the type of tumor, that grows worse with the passage of time and ultimately results in the death of a person.

2. LITERATURE REVIEW

The image segmentation is a division or separation of the image into region of similar features. In this paper we will discuss technique of preprocessing, segmentation, clustering and classification. This approaches will give improvements in segmentation performance that can be achieved by combining methods from distinct regions of the tumor image. This paper deals with a new image, segmentation and classification technique combining region growing and detection of edges the combination of this two methods helps to avoid characteristic segmentation errors and noise removal of an image which occurs when using region growing are edge detection seperatively. Several other authors defined the brain tumor detection using filter in and segmentation with different algorithm and methods.

MRI is basically used to detect the differences in the body tissues which have a considerably better technique as compared to CT scan. Thus, this technique becomes a special technique especially for the brain tumor detection and cancer imaging. This paper provides a review of image-based tumor detection. This paper deals with a new image, segmentation and classification technique combining region growing and detection of edges the combination of this two methods helps to avoid characteristic segmentation errors and noise removal of an image which occurs when using region growing are edge detection seperatively. Several other authors defined the brain tumor detection using filter in and segmentation with different algorithm and methods. There are a large number of methodologies that present almost optimal results, but only for well-segmented images [6]. Some of these techniques achieved error rate above 10% in the segmentation phase only. A brief survey was carried out among all the aforementioned areas.

3. PROPOSED APPROACH

The proposed approach focuses on the image preprocessing technique such as segmentation, clustering and classification. The segmentation is used to segment and divide the region. The clustering is used to cluster the information from the segmented image. Classification is used to compare the extracted information with the data set.

The algorithm is based on filtering of brain images which is used for removing the noise over an image, segmented with filtering images, clustering the information from segmented image and compare that information with the data set. Fig. 1 portrays the general workflow of the prediction process.

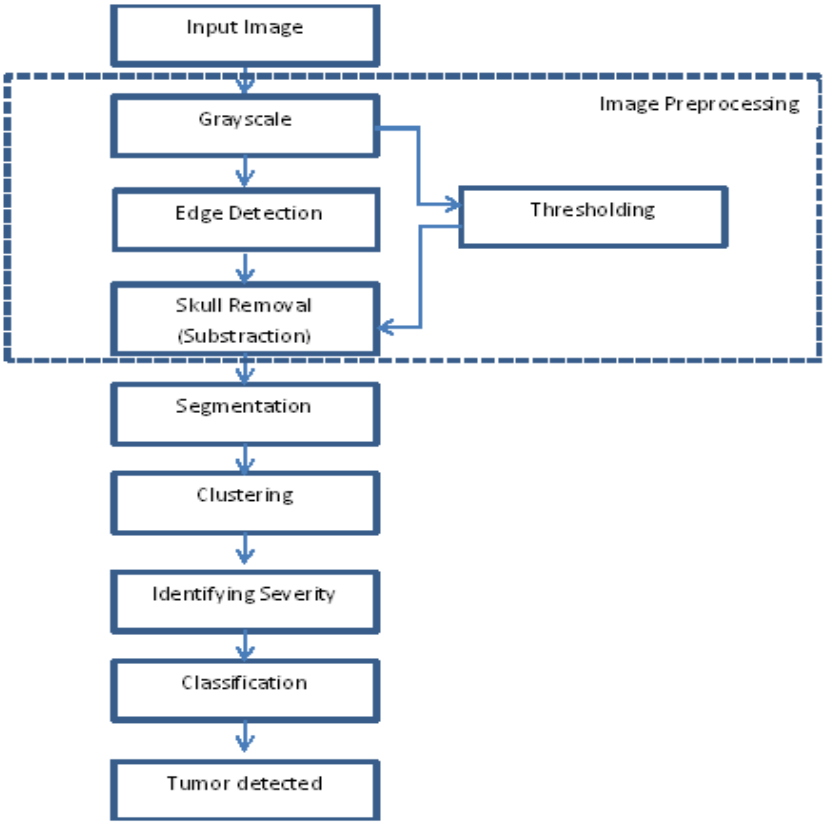

Chart -1:Workflow of the prediction process.

4. IMAGE PRE-PROCESSING

The preprocessing step converts the input image for the need of next level. It performs noise filtering and artifacts in the image and sharpening the edge in the image. RGB to gray conversion and reshaping of the image takes place here. Grayscale, Thresholding, Median filter is used for noise removal. The pre-processing is to improve the image quality to make it ready to further processing by removing or reducing the unrelated and surplus parts in the background of the images which are complicated to interpret. It will prepare the image for the next two-process segmentation and feature clustering. The noise and high frequency components are removed by filters. It does not increase image information content. This methods use the considerable redundancy in images.

5. SEGMENTATION

Segmentation is carried out by block segmentation.Image Segmentation partitions an image into set of regions. The region represents meaningful areas in an image or be the set of border pixels grouped into structures such as line segments, edges etc. The segmentation has two objectives. First to decompose an image into regions for further analysis and to perform a change of representation of an image for faster analysis. Different types of segmentation techniques are used for segmentation. Segmentation algorithm is based on the properties of gray level values of pixels. The different types of segmentation techniques are: Edge based segmentation, Threshold Based

Segmentation Region Based Segmentation, Clustering, Matching. Image segmentation has played an important role in computer vision especially for biomedical application. The result of image segmentation is a set of segments that collectively cover the entire image or a set of contours extracted from the image. Its accuracy but very elusive is very crucial in areas as medical, remote sensing and image retrieval where it may contribute to save, sustain and protect human life. This paper presents the analysis and implementation using Java features and one best result can be selected for any algorithm using the Subjective evaluation.

6. CLUSTERING

The feature clustering is used to extract the cluster that shows the predicted tumor at the segmentation output. The result of image segmentation is a set of segments that collectively cover the entire image or a set of contours extracted from the image. Its accuracy but very elusive is very crucial in areas as medical, remote sensing and image retrieval where it may contribute to save, sustain . The k-means method is a widely used clustering technique that seeks to minimize the average squared distance between points in the same cluster. Although it offers accuracy guarantees ,its simplicity and speed are very appealing in practice. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them.
The points are clustered around centroids µi

$$V = \sum_{i=1}^{k} \sum_{x_j \in S_i} (x_j - \mu_i)^2$$

∀i = 1 . . . k which are obtained by minimizing the objective.

7. CLASSIFICATION

An accurate classification of brain tumor grading is very important in diagnosis because, it defines prognosis and treatment decision for the patient. Classification phase refers to image categorization. This contains a dataset with predefined or previously detected patterns that are going to be compared with newly detected patterns. This will help to classify new pattern into a desired category.

Here we have used Naïve Bayes classifier. Bayesian classifier is a statistical classifier. Bayesian classifier is based on Bayes theorem. Bayesian classification is used for classifying objects into associated classes based on the attributes of those objects. Attributes of object are considered as independent of each other in Naïve Bayes classification.

8. MODELING TECHNIQUES

8.1 GRAYSCALE

In photography, computing, and colorimetry, a grayscale image is one in which the value of each pixel is a single sample representing only an amount of light, that is, it carries only intensity information. Images of this sort, also known as black-and-white or monochrome, are composed exclusively of shades of gray, varying from black at the weakest intensity to white at the strongest.Grayscale images are distinct from one-bit bi-tonal black-and-white images, which in the context of computer imaging are images with only two colors, black and white . Grayscale images have many shades of gray in between.Grayscale images can be the result of measuring the intensity of light at each pixel according to a particular weighted combination of frequencies (or wavelengths), and in such cases they

8.2 THRESHOLD SELECTION

The key parameter in image segmentation using thresholding technique is the choice of selecting threshold value T. In case of manual thresholding method, the threshold value T can be selected by the user with the help of image histogram. This method is generally accomplished by a tool that allows the user to select the threshold value T based on choice. In case of automatic threshold selection method, the value of T can be chosen based on histogram, clustering, variance, means etc. It is one of the simplest and fastest segmentation methods based on the assumption that images are formed from regions with different gray levels.

8.3 K-MEANS ALGORITHM

The purpose of the k-means algorithm is to cluster the data. K-means algorithm is one of the simplest partitions clustering methods. K-Means is the unsupervised learning algorithm for clusters. Grouping of pixels is done according to the same characteristics. In the k-means algorithm initially, we have to define the number of clusters k. Then k-cluster center is chosen randomly. The distance between the each pixel to each cluster centers is calculated. The distance may be of simple Euclidean function. A single pixel is compared to all cluster centers using the distance formula. The pixel is moved to the particular cluster which has the shortest distance among all. Then the centroid is re-estimated. Again each pixel is compared to all centroids. The process continuous until the center converges.

Algorithm:

1. Give the no of cluster value as ‘k’.
2. Randomly choose the ‘k’ cluster centers
3. Calculate mean or center of the cluster
4. Calculate the distance between each pixel to each cluster center
5. If the distance is close enough to the center then move to that cluster.
6. Otherwise, move to next cluster.
7. Re-estimate the center.

8.4 NAIVE BAYES THEOREM

Mathematically Bayes Theorem can be defined as

1. Let X be the data object
2. Let H be some hypothesis, such as X belongs to class C
3. P(H) is known as the prior probability
4. P(X|H) is probability that the hypothesis H holds given the observed data object X
5. P(H|X) is called posterior probability

Baye’s theorem provides a way of calculating posterior probability.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

P(X|H) is posterior probability of X conditioned on H. P(X) is prior probability of X. Posterior probabilities are class density estimates. Accuracy of classifier depends on class density estimates. As much accurate class density estimate as much higher accuracy is achieved.

9. EXPERIMENTAL OUTCOMES

First of all input image or original image is shown here, Fig. 1 shows original images which has brain tumor. Grey Scale and Threshold segmentation is applied on these images which contains brain tumor. The results are shown in the respective figure.

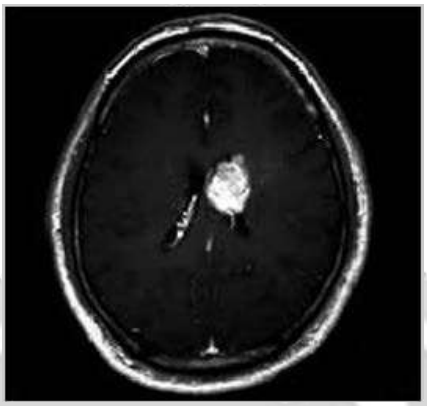

Chart-2: Original Image

Chart-3: Histogram Equalized of Original Image

In the following figure white spot is shown, which is the result of threshold segmentation applied on the images. This is basically the area with the intensity values higher than the defined threshold. High intensity areas mostly comprises of tumors. So through threshold segmentation we can specify the location of tumor.

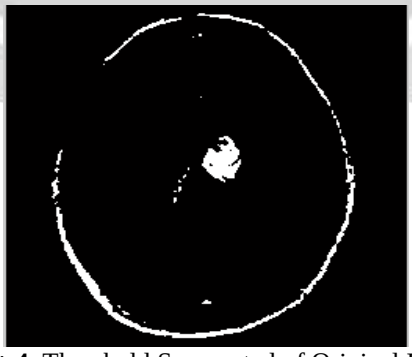

Chart-4: Threshold Segmented of Original Image

Then there is a technique called K-Mean segmentation which is applied on the resulted image obtained after threshold segmentation. Only the portion which contains tumor is highlighted here. The portions with the high intensity values are detected through threshold segmentation. These portions are marked through K-Mean segmentation methods.

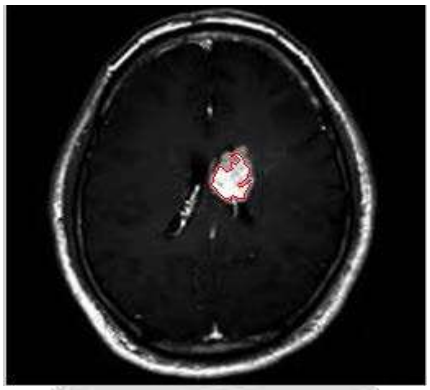

Chart-5:Detected tumor in original image

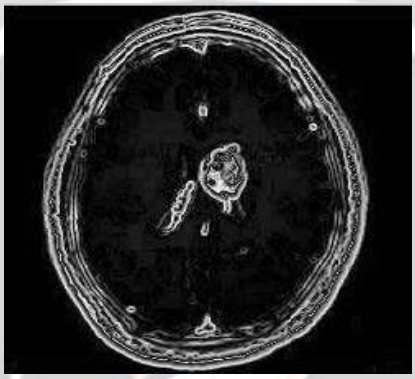

Chart-6: Image after filtering

Following Figure contains the resulted image obtained after applying functions like ROI.

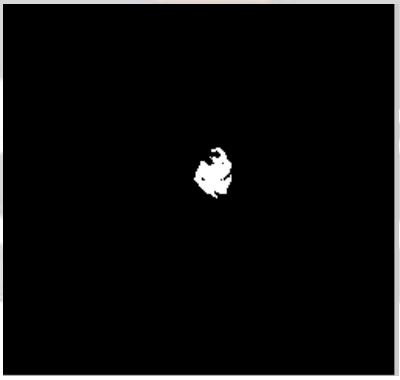

Chart- 7 :Tumor Detected

10. PERFRORMANCE ANALYSIS

As number of input images varies then according to the histogram and threshold values obtained, the accuracy changes. If we put threshold value nearby 80 to 100% then accuracy is achieved up to 95 to 98%.as threshold value increased we observed in the figure accuracy is decreased accordingly.

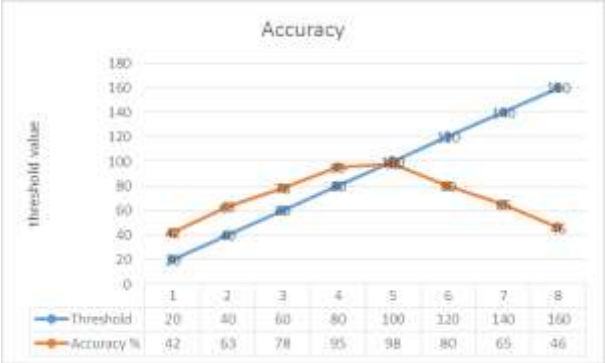

Chart-8: Performance Analysis

This expected output for detection of tumor achieves maximum 94% to 97 % accuracy for given set of MRI brain Images.

11. CONCLUSION AND FUTURE WORK

This research was conducted to detect brain tumor using medical imaging techniques. The main technique used was K-mean clustering & segmentation, which is done using a method based on threshold segmentation and morphological operators. The proposed segmentation method was experimented with MRI scanned images of human brains: thus locating tumor in the images. This technique gives efficient results as compared to previous researches. Experiments are applied on various images and results were extraordinary. Our proposed research is easy to execute and thus can be managed easily. In this work, preprocessing an image gives the result of an input image of tumor area. Segmentation, clustering and classification process removes the noise over an image and regain the smoothed clear an image of the tumor. Using the K-means clustering algorithm we detect the affected region over an tumor image. In near future we can implement several algorithm and techniques to detect the tumor more effectively.

12.REFERENCES

[1]Asra Aslam, Ekram Khan, M.M.SufyanBeg ,Improved edge detection algorithm for brain tumor segmentation,Aligarh Muslim University(AMU),Aligarh,202002,2015.

[2] P. Dhage, M. Phegade, S. Shah, Watershed segmentation brain tumor detection, in*: IEEE 2015 International Conference on Pervasive Computing (ICPC), 2015, pp. 1–5.*

[3] *K.Padmavathi, C.Megala, Detection of brain tumor with Filtering Techniques ,2015.*

[4] S. Kannan, A new segmentation system for brain MR images based on fuzzy techniques, *Applied Soft Computing 8 (4) (2014) 1599–1606.*

[5] S. Chandra, R. Bhat, H. Singh, A PSO based method for detection of brain,2013.

[6] tumors from MRI. In *NaBIC World Congress on Nature & Biologically Inspired Computing, (2013), 666-671.*

[7] D. Wei, C. Li, Y. Sun, Medical image segmentation and its application in cardiac MRI, *Biomedical Image Understanding, Methods and Applications (2015) 47–89.*

[8] J. Canny, A computational approach to edge detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence, (6) (2011) 679–*698.