A New Approach for Brain Tumor Detection and Area Calculation Using Median Filter, K-Means, SVM and Naïve Bayes Classifier

Sandhya M. Karande[1]
Department of Computer Engineering
Vishwabharati Academy's Collage of Engineering
Ahmednagar, India

Jayapal PC[2]
Department of Computer Engineering
Vishwabharati Academy's Collage of Engineering
Ahmednagar, India

Namrata A. Kale[3]
Department of Computer Engineering
Vishwabharati Academy's Collage of Engineering
Ahmednagar, India

Abstract: Tumor is an uncontrolled growth of tissues in any part of the body. Brain tumor is inherently serious and life-threatening. Huge numbers of deaths have been verified due to the fact of inaccurate detection. Magnetic Resonance images are examined visually for detection of brain tumor producing less accuracy while detecting the stage and size of tumor. To avoid this, we propose a brain tumor detection technique to increase accuracy and reduce the noise and other artifacts in the image. The noise reduction is done by median filtering technique. The accuracy is improved in segmentation by k- means algorithm. In addition support vector machine and Naive bayes classification techniques are compared for more accurate results.

Keywords: Brain tumor, K-means, Magnetic Resonance Imaging (MRI), Pre-processing, Support Vector Machine (SVM), Naïve bayes (NB).

I. INTRODUCTION

The existence of a variety of types of tumors with different characteristics necessitates an individually well-defined specialized treatment as outlined by a neuro-oncologist. Tumor is due to the uncontrolled growth of the tissues in any part of the body. The tumor may be primary or secondary. If it is an origin, then it is known as primary. If the part of the tumor is spread to another place and grown as its own affects CSF (Cerebral Spinal Fluid). The cause of brain tumors is unknown but their occurrence is associated with an abnormal growth of cells inside the skull which may lead to serious impairments and/or life-threatening conditions because of their character and the limited space of the intracranial cavity. Research shows that people affected by brain tumors die due to their inaccurate detection [1]. The MRI images visual evaluation and examination by radiologists is subjective by its nature and is time consuming and prone to errors or omissions, however due to the complexity of information at this point, it cannot be substituted with a fully automated evaluation. Therefore algorithmic image processing can assist radiologists in brain tumor diagnosis in multi-parametric MR images brain tumor detection and segmentation needs to take into account the large variations in appearance and shape of structures.

II. EXISTING METHOD

The existing method is based on the thresholding process. The extracted area of cluster is given to the thresholding process. It applies binary mask over the entire image. It makes the dark pixel become darker and white become brighter. In threshold coding, each transform coefficient is compared with a threshold, If it is less than the threshold value then it is considered as zero else it will be considered as one ,means it consists of only two gray values .That is white as 1 and black as 0.So we cannot

classify the normal and abnormal patient. This is the main drawback of the existing system. Due to that we go for the proposed method for feature extraction.

III. PROPOSED METHOD

The proposed system has mainly five modules: preprocessing, segmentation, Feature extraction, classification and area calculation. In Preprocessing Median filtering is applied on the MR image for noise removal. K-means algorithm is used for segmentation. Gray level Co-Occurrence Matrix is used for feature extraction. And then SVM and naive bayes classifier is applied to detect whether the given image is Normal or Abnormal. Then we compare the accuracy of the proposed system in between SVM and naive bayes. Stage detection step is used to find out the stage of tumor which is marked as abnormal in previous step. See fig.1.
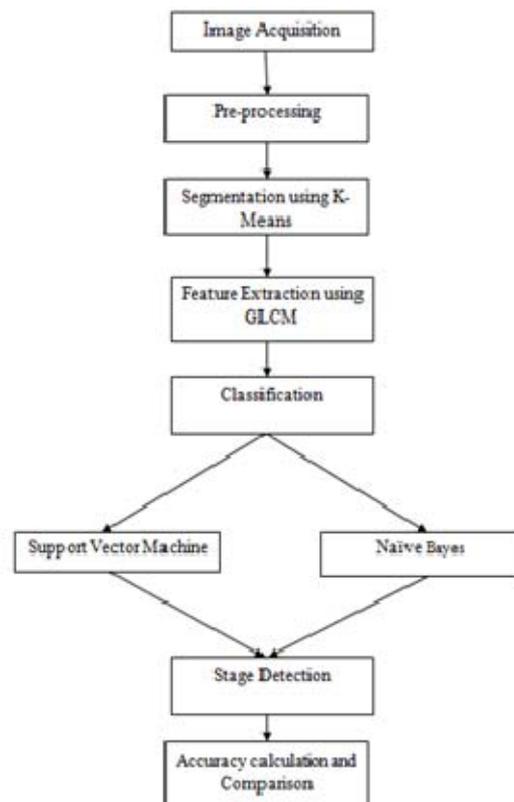

Fig.1 Proposed Method

a) Pre-processing

According to the need of next level that is to enhance the visual appearance of image and to improve the manipulation of dataset, preprocessing of image is required. This performs filtering of noise and other artifacts present in image. The possibilities of arrival of noise in modern MRI scan are very less. It may arrive due to the thermal effect [1]. For preprocessing median filter is used. Median filter acts as noise removal non-linear tool. In this filtering technique each image pixel is replaced by the neighborhood median pixel. Impulse noise is labeled as a pixel which is different from its majority of neighbors, also being structurally unaligned with the pixels similar to it. The median pixel value of those pixels in the neighborhood by which the noise labeling test have passed then replaces these noise pixels. Median filter is robust and widely used in image processing because it preserves edges while removing noise. Another advantage is that it does not create any new intensity image pixel because median is an existing pixel value in the neighborhood window (3 x 3).

In order to perform median filtering, in fig. 2 first window is moved and all the pixels enclosed by the window are shorted. After then median is computed and this value is assigned to center pixel. If the number of elements in K*K window is odd, middle value is assigned as median value, else average of two middle values is assigned as median value [11], [12].
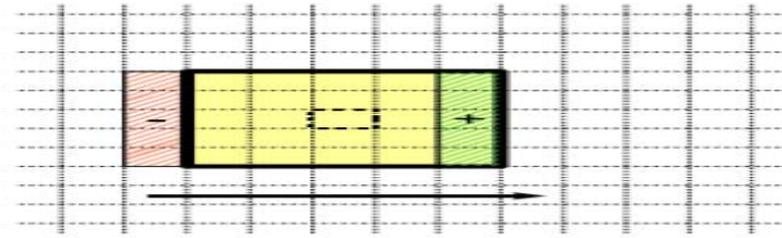
Fig.2 Process of median filter

b)   Segmentation

The MRI image processing using clustering is an efficient method. Clustering technique classifies the objects into different groups, or more specifically, partitioning of a data set into clusters (subsets), so that the data in each cluster (ideally) shares some common trait - often according to some defined distance measurement. Data partitioning is a usual technique for the analysis of statistical data, which is used in many areas, including machine learning, image analysis, pattern recognition, bioinformatics and data mining. The computational task of partitioning the data set into k subsets is often referred to unsupervised learning. There are many approaches of clustering designed for a wide variety of purposes. K-means is a typical clustering algorithm (MacQueen, 1967) [4]. K-means is generally used to determine the natural groupings of pixels present in an image. It is attractive in practice, because it is straightforward and it is generally very fast. It partitions the input dataset into k clusters. Each cluster is represented by an adaptively changing center (also called cluster center), starting from some initial values named seed-points. K-means clustering computes the distances between the inputs (also called input data points) and centers, and assigns inputs to the nearest center. K-means method is an unsupervised clustering method that classifies the input data objects into multiple classes on the basis of their inherent distance from each other [5]. Clustering algorithm assumes that a vector space is formed from the data features and tries to identify natural clustering in them.

1.   Mathematical Representation:

For a given image, compute the cluster means m

$$M = \frac{\sum_{i:C(i)=K} x_i}{N_k} \qquad ,k=1,\ldots,K \qquad\qquad --------------------(1)$$

Calculate the distance between the cluster center to each pixel

$$D(i)= \arg\min \|x_i - M_k\|^2 \qquad ,i=1,\ldots,N \qquad\qquad -------------------(2)$$

Repeat the above two steps until mean value convergence.

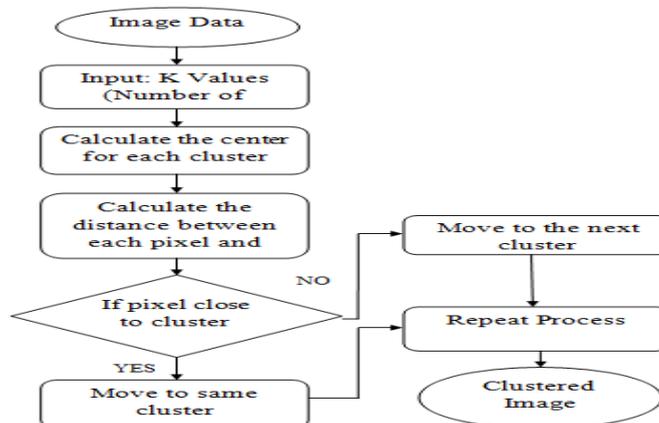### 2.   Flowchart of k-means Algorithm:

Fig.1 Proposed Method

c) *Feature Extraction*

The feature extraction is extracting the cluster which shows the predicted tumor at the k-means output. The extracted cluster is given to the Gray Level Co-occurrence Matrix. Features are said to be properties that describes the whole image. It can also refer as an important piece of information which is relevant for solving the computational task related to specific application. The purpose of feature extraction is to reduce the original dataset by measuring certain features. The extracted features acts as input to classifier by considering the description of relevant properties of image into feature space. GLCM calculates the co-occurrence matrix of an image by computing how often a pixel with a certain intensity 'I' occurs in relation with other pixel 'j' at a certain distance d and orientation. In our paper, statistical features based on image intensity and features from gray level co-occurrence matrix are used to distinguish between normal and abnormal patient.

The Method:

Step 1: Obtain the co-occurrence matrices of each image. The gray levels of the image are reduced to 16. For 0 degree, 45 degree, 90 degree and 135 degree directions, and distances 1, 2, 3, 4 and 5, calculate the corresponding co-occurrence matrix. This produces 20 matrices of 16x16 integer elements per image.

Step 2: Obtain the values for the chosen descriptors. For each co-occurrence matrix, the value for the descriptor (or descriptors) is calculated. For each image, the resulting descriptor values are stored in a matrix where the rows represent the directions 0 degrees, 45 degrees, 90 degrees and 135 degrees and the columns represent distances 1, 2, 3, 4 and 5.

Step 3: Generate image signatures. The image signatures are calculated from the descriptor matrix by averaging the values obtained with the different distances for each direction. Aiming at eliminating dependencies of image rotation, distances are calculated with different "rotations" on the elements of the signatures and the smaller one is adopted.

Step 4: Compare the images through their signatures. The signatures of two images are compared and the distance between them is calculated, using the Euclidean distance function on the elements of the signature [6] [7].

d) *Classification:*

Classification analyses the numerical properties of image features and organize the data into different categories.

1. **Support Vector Machine (SVM):**

SVM is a binary classifier based on supervised learning. It employs two phases of processing- training phase and testing phase. In training phase, characteristic properties of image features are isolated and a unique description of each classification category is created. In testing phase, these features space partitions are used to classify image features. SVM classifies between two classes by constructing a hyperplane in high-dimensional feature space which can be used for classification. Hyperplane can be represented by equation-

$$wx+b=0 \hspace{8cm} \text{-------------------(3)}$$

w is weight vector and normal to hyperplane.

b is bias or threshold.

Linear Separable Binary Classifier is used. Whenever any new image comes for classification the trained data features and new image features are given to SVM classifier. In which the testing features are compared with the new features. If the result of comparing the both is highly matched then we put it to class 1, means the new image is Normal image. Else the features are placed into class 2, means the new image is infected image or abnormal image

2. **Naïve Bayes (NB):**

Naive Bayes classifier is based on the Bayesian theory which is a simple and efficient probability classification method based on supervised classification technique. For each class value it estimates that a given instance belongs to that class [8]. The feature items in one class are assumed to be Independent of other attribute values called class conditional independence [9]. Naive Bayes classifier needs only small amount of training set to estimate the parameters f or classification. The classifier is stated as Naïve Bayes (NB)

P (A|B) = P (B|A) * P (A)/P (B)   -----------------------(4)

Where P (A) is the prior probability or marginal probability of A, P (A|B) is the conditional probability of A, given B called the posterior probability, P (B|A) is the conditional probability of B given A and P(B) is the prior or marginal probability of B which acts as a normalizing constant. This provides the mathematical representation of the way in which the conditional probability of event A given B can be related to the conditional probability of B given A. The probability value of the winning class dominates over that of the others [10].

e)   *Stage Detection:*

Using the binarization method in the stage detection step the area of the tumor is calculated. That means the image having two values either white or black (1 or 0). Here maximum image size is 256x256 jpeg image. We can represent a binary image as a summation of total number of black and white pixels [1].

Image . I = $\sum_{W=0}^{255}\sum_{H=0}^{255}$ [f(0) + f(1)]

Pixels = Height (H) *Width(W) = 256*256

f(1) = black pixel (digit 1)

f(0) = black pixel (digit 0)

No_of_white_pixel,P = $\sum_{W=0}^{255}\sum_{H=0}^{255}$ [f(0)]

Where,

P=Total number of white pixels (height*width)

1 Pixel = 0.264mm

The formula for area calculation is,

Size_of_tumor, S=[($\sqrt{P}$) *0.264] mm^2 --------------------(5)

P = Number of white pixels;

H = height

W = width

IV. CONCLUSION

There are different types of tumors available. They may be mass in the brain or malignant over the brain. Suppose if it is a mass, then K- means algorithm is enough to extract it from the brain cells. If there is any noise present in the MR image it is removed before the K-means process. The noise free image is given as input to the k-means and tumors are extracted from the MRI image. The performance of brain tumor segmentation is evaluated based on K-means clustering. Thus, the pre-processing is done by filtering. Segmentation is done by advanced K-means algorithm. Features are extracted from images using feature extraction algorithm GLCM and classified further using SVM and Naïve Bayes classifier and finally, approximate reasoning method to recognize the tumor shape and stage in MRI image using the binarization.

References

1. Selvakumar J., Lakshmi A., Arivoli, T., "Brain tumor segmentation and its area calculation in brain MR images using K-mean clustering and Fuzzy C mean algorithm", International Conference on Advances in Engineering, Science and Management (ICAESM), 2012, pp. 186 – 190.

2. S.Sathish Kumar, M.Moorthi, M.Madhu3and Dr.R.Amutha "*An Improved method of Segmentation using Fuzzy-Neuro Logic"* Second International

3. S. Murugavalli and V. Rajamani" An improved Implementation of Brain tumor Detection Using Segmentation Based on Neuro Fuzzy Technique", Journal of Computer Science 3 ,2007 october.

4. J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the 5th Berkeley* Symposium on Mathematical Statistics and Probability, Berkeley, University of California Press, 1967, pp. 281-297.

5.] T. Kanungo, D. M. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu, "An efficient k-means clustering algorithm: Analysis and implementation," in *Proceedings of the IEEE Conference on* Computer Vision and Pattern Recognition (CVPR), 2002, pp. 881-892.

6. Mari Partio, Bogdan Cramariuc, Moncef Gabbouj, and Ari Visa " Rock Texture Retrival using Gray level Co-occurrence Matrix" Tampere University of Technology.

7. Joaquim Cezar Felipe, Agma J. M. Traina, Caetano Traina Jr Retrieval by Content of Medical Images Using Texture for Tissue Identification. Institute of Superior Education COC.

8. Li Dan, Liu Lihua, Zhang Zhaoxin (2013) Research of Text Categorization on WEKA. Proc. of Third International Conference on Intelligent System Design and Engineering Applications 1129-1131.

9. J. Han, M. Kamber (2000) Data Mining Concepts and Techniques. Morgan Kauffman Publishers

10. Laszlo Toth, Andras Kocsor, Janos Csirik (2005) On Naive Bayes In Speech Recognition. Int. J.Appl. Math. Comput. Sci 15:287–294

11. Geng Ying, Li Miao, Yuan Yuan and Hu Zelin, "A Study on the Method of Image Pre-Processing for Recognition of Crop Diseases", International Conference on Advanced Computer Control, 2008 IEEE, pp. 202-206.

12. Rafeal C. Gonzalez and Richard E. Woods, "Digital Image Processing", second edition, pearson education.