ISSN 2349-4506
Impact Factor: 2.785

Global Journal of Engineering Science and Research Management

ADVANCED K-MEANS ALGORITHM FOR BRAIN TUMOR DETECTION USING NAIVE BAYES CLASSIFIER

Veena Bai K*, Dr. Niharika Kumar
* MTech CSE, Department of Computer Science and Engineering, B.N.M. Institute of Technology, Bengaluru, India
Associate Professor, Department of Computer Science and Engineering, B.N.M. Institute of Technology, Bengaluru, India

DOI: 10.5281/zenodo.814827

KEYWORDS: Brain Tumor, MRI, K-means, segmentation, Naive Bayes.

ABSTRACT
In health care centers and hospitals, millions of medical images have been generated daily. Analysis has been done manually with an increasing number of images. Brain tumor segmentation in Magnetic resonance imaging (MRI) has been recent area of research in the field of medical diagnosis. Accurate segmentation of brain tumors is an important task and it is challenging problem. K-means clustering algorithm is the most popular and widely-used partitional clustering algorithm in practice. However, traditional k-means algorithm suffers from sensitive initial selection of cluster centers, and it is not easy to specify the number of clusters in advance. Here an Advanced k-means algorithm is proposed for segmentation that can automatically split and merge clusters which incorporate the new ideas in dealing with huge scale of medical image data. Then features are extracted from the segmented image and its efficiency is increased by using Naive Bayes classifier and is classified into normal or abnormal images.

INTRODUCTION
Brain tumor is the abnormal growth of tissues in the brain. It occurs to any person at any age. In today's environment there are various techniques to identify the tumor. Tumor is of three types, they are primary, malignant and extended. Primary tumor starts in brain cells and spread to other parts of the body. Malignant tumor starts at specific part of the brain and doesn't spread. It is sometimes dangerous and sometimes even causes death. Extended tumor starts at any part of the body and spreads to the brain.

There is different techniques like computed tomography (CT), Magnetic Resonance Imaging (MRI) etc, to diagnosis the brain tumor. Magnetic resonance imaging (MRI) is the imaging technique used to visualize the internal parts of the body. MR images are the most important tool for early detection of brain tumor. But MRI interpretation is based on the opinion of the radiologist. The data obtained from MRI will help to diagnosis and treat the disease well.

Accurate visualization of the MR images is very important. Hence image processing technique like segmentation is the most important tool used to make format of image to make it easier to analyse and understand. This manual segmentation of image faces many problems; hence various clustering methods are used for this purpose.

RELATED WORK
The brain tumor detection techniques are rapidly increasing in research and advanced ideas have been proposed for its detection.

Charutha.S and M.J. Jayashree proposed a Texture based region growing and Cellular automata based edge detection for efficient detection of brain tumor. The segmentation methods will be performed separately and integrated. In texture based region growing, the region growing is carried by taking the intensity and texture. In cellular automata edge detection the exact and clear edges are detected. Finally both the methods are integrated in order to segment the tumor exactly. But the main problem is time complexity of performing both the methods of segmentation.

R.Preetha and G.R.Suresh proposed fuzzy c-means algorithm which is a soft clustering method of segmentation. Here data elements can belong to more than one cluster and for each data elements memberships are assigned. Thus FCM gives clustered image of the brain from which features are extracted. The accuracy of segmentation is identified by SVM classifier.

Natarajan.P et al. discussed about threshold segmentation to enhance the image quality. Thresholding uses gray scale images to form binary images and is applied after image enhancement. Morphological operations are applied on the image obtained after segmentation and final technique subtraction is applied to get exact tumor part. The main problem with this method is threshold value needs to be specified for segmentation.

Janki Naik and Sagar Patel had discussed about decision tree classification. Pre-processing is done, texture features are extracted from the CT scan images and then association rules are applied for mining. Later efficiency of the system is increased by using decision tree classification. The main issue in decision tree classification is time complexity as it iterative process.

Ashwini A. Mandwe and Anisa Anjum presented a method called K-means clustering algorithm for segmentation. Pre-processing of MR images is done then K-means algorithm is applied for segmentation. There may be a chance of wrongly clustered regions hence morphological filtering will be applied. K-means algorithm gives the efficient results compared to other clustering algorithms but main drawback it is very sensitive to initial selection of cluster centers, and it is not easy to specify the number of clusters in advance.

PROPOSED SYSTEM

There are two phases in the proposed system. Training phase and Testing phase. The overall architecture of the proposed system is as shown in the Fig.1. The various MR images are gathered and stored in the image database. After the image database is ready, apply different image processing techniques for both training and testing phase. The different modules in the proposed system are Pre-processing, Segmentation, Feature extraction and Classification.

The pre-processing, segmentation and feature extraction phases are same for both training and testing phase. Later trained and tested feature vales are given to the Naive Bayes classifier.

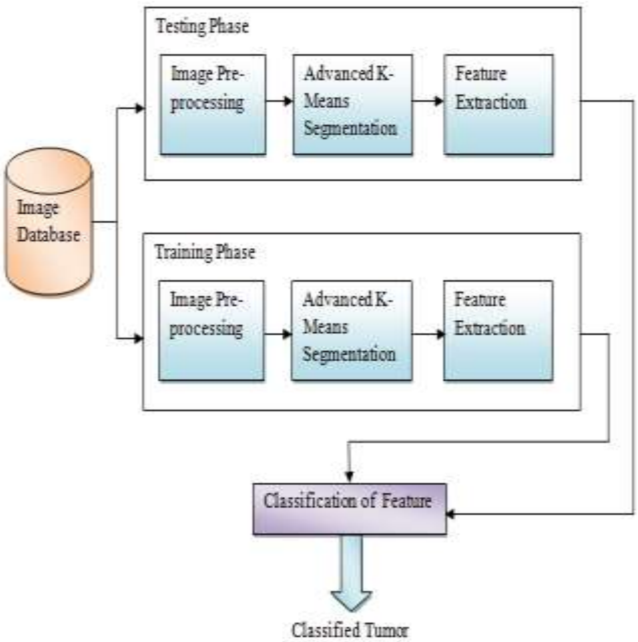


Fig.1 Architecture of Proposed System

Pre-processing

Pre-processing is the method of removing unwanted data from the MR images. The main objective is to enhance the MR image data quality for further processes. The MR images contain the noise generated during MRI scanning. It should be removed in order to improve quality of images for efficient detection of tumor.
The various steps involved in pre-processing are:

Median filtering

It is the first step of any pre-processing stage, which is often used to remove salt-and-pepper noise. Median filtering is often used in digital image processing which preserves the edges while eliminating noise.

The main idea behind working of median filter is, the median filter window size is fixed and it passes through the image pixels bit by bit replacing each pixel entry with median of neighbouring entries until all pixels are covered. The median is first calculated by first sorting all pixel values in ascending order and obtaining the middle value and replacing it. This filtering enhances image quality for further process.

Sobel Edge Detection

Sobel edge detection algorithm is particularly used for creating an image emphasizing edges. The main idea of working is, first filtered image is read then gradient vector along x-direction and y-direction of sobel-operator is read. The gradient vector along x-direction is multiplied with input image. Finally the horizontal and vertical edge images are added to get perfect and accurate MRI image.

Contrast Enhancement

The filtered image and edge detected image is added to get the enhanced image. This enhanced image quality and accuracy is more than actual input image.

Binarization

Binarization is a method of converting a pixel image into binary image. Here threshold is set automatically or specified manually, depending on the threshold value the pixels below is converted to 1 and above threshold value is converted to 0.

Morphological Opening

Another important step of pre-processing is morphological opening (skull removing). Two morphological operations called erosion and dilation are carried out. The opening is erosion followed by dilation by using the same structuring element. The erosion finds the places, marks those positions and element fits inside the image. The dilation fills back in places where element fits inside the object. The opening removes small objects, objects with protrusions, and connections between objects.

Advanced K-means algorithm

Cluster analysis is an important task of data mining in MRI brain tumor detection. K-means algorithm is the most popular and widely-used partitional clustering algorithm in practice. In the proposed system a robust k-means algorithm that can automatically split and merge clusters which incorporates the new ideas in dealing with huge scale of medical image data is proposed. This novel algorithm not only addresses the sensitivity in selecting initial cluster centers, but also is resilient to the initial number of clusters.

Algorithm

Step 1: Read the filtered image.
Step 2: Convert matrix value to an array.
→arr1 [i][j]=arr[i*n + j]
Step 3: Calculate average value of pixels i.e. Seed point S
→S (average) = Sum of all pixels/No. of pixels
Step 4: Find the Euclidean distance between seed point and array i.e. dist D
→D (dist) = $sqrt(\Sigma_{i=1} (x_i - y_i)^2)$
Step 5: Calculate distance threshold, sum of the distances i.e. distth DT

DT=D1+D2+D3+…… Dn

Step 6: Select pixels which are lesser than distance threshold i.e. D< DT = q(qualified).

Step 7: Calculate the mean of q pixels i.e. new seed NS

→NS (mean) = sum (q)/n

Step 8: If S==NS

→Empty the pixels assigned to qualified cluster.

→array (q) = []

→Save the new seed as center value.

center (c) = NS

else update S= NS

Repeat from step 4 to step 8

Step 9: Sort the center values in ascending order.

→C=sort(c)

Step 10: Remove centers whose difference is very small.

Step 11: find the distance between input image and center value.

→Distance = ((vector-centers). ^2)

Step 12: Segment the tumor image using the minimum distance from the label image.

→lb = min(distance)

→Label=reshape (lb, size (gray))

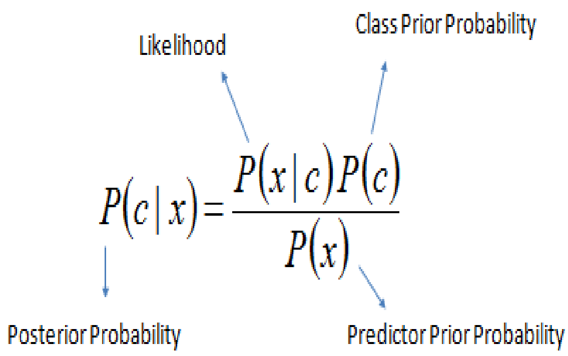
Step 13: Output segmented results.

GLCM Feature Extraction

Feature extraction is very critical role in brain tumor classification system, more sensitive information is obtained by applying different feature on the brain image. Gray-level co-occurrence matrix (GLCM) is the statistical method of examining the textures that considers the spatial relationship of the pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. Using GLCM matrix other 27 features are calculated like AutoCorrelation, Contrast, Mcorrelation, PCorrelation, Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Entropy, MHomogeneity, PHomogeneity, Maximum probability, Variance, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure of correlation1, Informaiton measure of correlation2, Inverse difference, Inverse difference normalized, Inverse difference moment normalized, Skewness, Standard Deviation and Kurtosis.

Classification

Image classification is perhaps the most important part of digital image analysis. The Naive Bayesian classifier is based on Bayes' theorem with independence assumptions between predictors. A Naive Bayesian model is easy to build, with no complicated iterative parameter estimation which makes it particularly useful for very large datasets. Bayes theorem provides a way of calculating the posterior probability, $P(c|x)$, from $P(c)$, $P(x)$, and $P(x|c)$. Naive Bayes classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the values of other predictors. This assumption is called class conditional independence.



Using this probability function the images are classified. The probability value of images below this value is considered as normal image and above this value is considered as abnormal image containing tumor.

RESULTS AND ANALYSIS

The performance of the proposed system is determined by confusion matrix. This matrix describes all possible outcomes of a prediction results in table structure. The possible outcomes of a two class prediction be represented as True positive (TP), True negative (TN), False Positive (FP) and False Negative (FN). The classifications of normal and abnormal images are accurately classified as True Positive and True Negative respectively. A False Positive is when the outcome is incorrectly classified as positive when it is a negative. False Positive is the False alarm in the classification process. A false negative is when the outcome is incorrectly predicted as negative when it should have been in fact positive.

In our system consider,
TP= Number of Abnormal images correctly classified.
TN= Number of Normal images correctly classified
FP= Number of Normal images classified as Abnormal.
FN= Number of Abnormal images classified as Normal.

Sensitivity: The probability of the test finding the abnormal case among all abnormal cases.

$$\frac{TP}{TP+FN}$$

Specificity: The probability of the test finding the normal case among all normal cases.

$$\frac{TN}{TN+FP}$$

Accuracy: The fraction of test results those are correct.

$$\frac{TP+TN}{TP+FN+TN+FP}$$

Using these equations the accuracy of the proposed system is calculated.
Results of the proposed system:
Sensitivity: 86.20%
Specificity: 92.39%
Accuracy: 90%

The comparative analysis of the existing and proposed system is as shown in below table,

	Proposed System	Base Paper
Methodology	K-means+Naive Bayes	Association rules+Naive Bayes
Sample size	125	124
Image types	2	2
Sensitivity	86.20%	89%
Specificity	92.39%	82%
Accuracy	90%	88.23%

CONCLUSION

"Advanced K-means algorithm for Medical image for Brain tumor Classification" is used to get accurate and efficient result. Using K-means algorithm the exact tumor is found, the GLCM features are extracted and with Naive Bayes classification technique tumor have been found as well as classified in Normal or Abnormal class. The proposed system has accuracy of 90%, sensitivity of 86.20% and specificity of 92.39% were found in classification of brain tumor using Naive Bayes classifier. This will produce result into normal or abnormal in efficient way.

REFERENCES

1. Charutha S and M. J. Jayashree, "An Efficient brain tumor detection by integrating modified texture based region growing and cellular automata edge detection" , *International Conference Control, Instrumentation, Communication and Computational Technologies (ICCICCT),* pp. 1193 – 1199, 2013.
2. R.Preetha and G.R.Suresh "Performance Analysis of Fuzzy C Means Algorithm in Automated Detection of Brain Tumor" World Congress on Computing and Communication Technologies, Pages 30-33, 2014.
3. Natarajan et. al., "Tumor Detection using threshold operation in MRI Brain Images", IEEE International Conference on Computational Intelligence and Computing Research, pp: 368 - 372, 2012.
4. Janki Naik and Sagar Patel "Tumor Detection and Classification using Decision Tree in Brain MRI." International Journal of Computer Science and Network Security (IJCSNS), VOL.14 No.6, June 2014.
5. Janki Naik and Sagar Patel "Tumor Detection and Classification using Decision Tree in Brain MRI." International Journal of Computer Science and Network Security (IJCSNS), VOL.14 No.6, June 2014.