DETECTION AND CLASSIFICATION OF BRAIN TUMOR USING

NAÏVE BAYES AND J48

NORA NAIK[1], MANALI MATONKAR[2], PRATIKSHATANDEL[3],

MAHABALESHWAR ALIAS OMKAR VEREKAR[4] & VEDA SALKAR[5]

[1]*Assistant Professor, Department of Computer Engineering, Agnel Institute of Technology & Design, Goa, India*

[2,3,4,5]*Student, Department of Computer Engineering, Agnel Institute of Technology & Design, Goa, India*

ABSTRACT

Image processing has become an essential component in science and technology with the tremendous influence and impact on modern society. Medical imaging has grown into one of the sub-fields in scientific imaging due to the rapid growth in computerized medical image reconstruction and computer-aided diagnosis. A brain tumor is a mass of cells which grow abnormally within the brain or spinal cord. The proper function of the brain can be dangerously disrupted by tumor. Mostly the method followed in hospitals for diagnosis of tumor is that the physician segments the CT or MRI scan manually to detect a tumor region in the brain which is a manual process. To avoid this problem, the proposed system focuses on detecting and classifying whether brain tumor is malignant or benign based on the features extracted from the tumor region with lesser time and higher accuracy in comparison to the manual analysis.

KEYWORDS: Brain Tumor, Magnetic Resonance Imaging, Cancerous, Benign, Malignant, Image Processing, Data Mining, K-Means, Morphology, Naïve Bayes & J48 Algorithm

Original Article

INTRODUCTION

Major functioning of the body is controlled by human brain, which is the main organ of the human body. A tumor is cumulation of abnormal cells in our brain. Malignant (cancerous) tumors and benign tumors are types of tumors. Malignant tumors can be classified into primary tumors, which are small and slower in growth, and secondary tumors spread in the rest of the body, known as metastasis tumors of brain. Mostly the method followed in hospitals for diagnosis is that the physician manually segments the MRI or the CT scan to detect a brain tumor region. Using image processing techniques the brain cells MRI images will first be analyzed and information will be extracted from those MRI images and further classified by comparing the data obtained with the existing data. The accuracy of this method mostly depends on the experience and skills of the physician. The ever increasing number of cases and shortage of radiologist make this process both expensive and time consuming. The Proposed system consist of two domains i.e. Image processing and Data Mining. The initial step in Image processing is image acquisition followed by image pre-processing, image segmentation, Morphological operation and feature extraction. In order to classify the image into benign or malignant we make use of two classification algorithms: Naïve Bayes and J48.

PROPOSED SYSTEM

The proposed system consists of two main domains i.e. image processing and data mining. The input to the proposed system is MR image, which are stored in a specific folder on the system (computer) and its output is whether the tumor region detected is benign or malignant. The proposed system begins with images acquired from the system followed by image pre-processing, image segmentation, thresholding, morphological operation and feature extraction. In order to classify the image into benign or malignant we make use of two classification algorithms: Naïve Bayes and J48. The flowchart of the system proposed is shown in figure 1 below
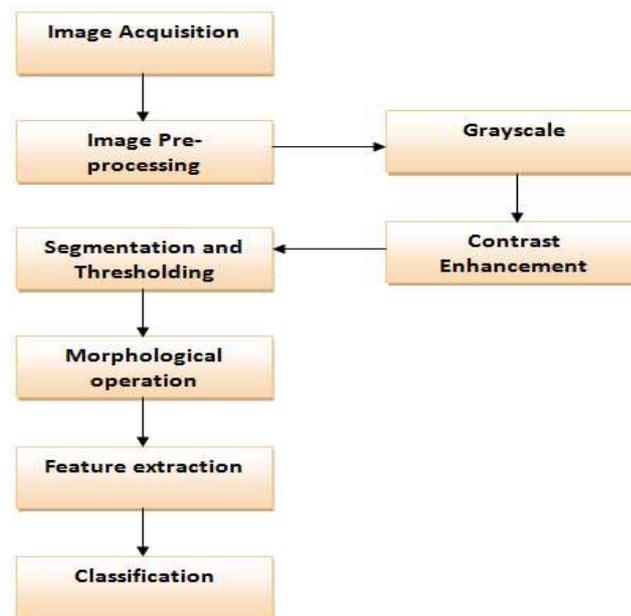
Figure 1: Proposed System Flowchart

Image Acquisition

This is the starting step in image processing. This deals with acquiring images by acquisition method. For the proposed system we make use of brain MRI images acquired from hospitals and radiology scan centers.

Image Preprocessing

In image pre-processing we perform different operations on image to improve the image quality. In order to enhance the quality of an image, it is needed to perform the operations such as reduction or removal of noise and contrast enhancement. The contrast enhancement technique brings out the information that exists within the low dynamic range of that gray level image.

Image Segmentation

Segmentation includes clustering which is a technique for finding similar groups in data called clusters. Homogenous classes or clusters are formed by partitioning the data points in to different clusters. In the proposed system, K-means clustering Algorithm was used.

The algorithm is as follows:

- Let the input image contains of $x_1,..,x_m$ data points, let k denotes no. of clusters.

- Choose $c_1,..,c_k$ cluster centers.

- Distance between each of pixel and cluster center is found.

- J denotes distance function and it is given by the formula:$J=|x_i-c_j|$ for i=1,..,N and for j=1,..,k where $|x_i-c_j|$.

- Using the given relation distribute the data points x among the k clusters $x\varepsilon C_j$ if $|x-c_j|<|x-c_i|$ for i=1,2,..,k,i≠j, where $c_j$denotes the set of data points who secluster center is c_j.

- Updated cluster center is given as,$C_j=\sum_{x\varepsilon cj}x$, for i=1,..,k, where m_i is the number of objects in the data set c_i where c_i is the ith cluster and $c_j$is the center of the cluster c_i.

- Repeat Step 4 to Step 7 until convergence is met.

Thresholding

Thresholding is an efficient technique for image binarization. In our proposed system we have used Otsu's thresholding algorithm.

The algorithm is as follows:

- Compute histogram and probabilities of each intensity level.

- Set up initial weight w_i which is the probabilities of the two sets by an initialthreshold t and$u_i$which is the mean.

- Step through all possible thresholds t=1,......, maximum intensity

- Update w_i and u_i.

- Compute Variance.

- Threshold Desired corresponds to the maximum variance computed above.

  For each computed value

  if(pixel value)>Threshold value, then pixel. Value=255

  else

   pixel value=0

Morphological Operations

Morphological operations are used for analysis. It uses local operations for shape modification of an object in an image. It can be used to remove unwanted effects in segmentation post-processing. The interaction of the image with a structuring element S are described by erosion, dilation, opening and closing are the morphological operations. The structuring element is a small binary image having a small matrix of pixels, each with a value of zero or one. The matrix dimensions determine the *size* of the structuring element. The *shape* of structuring element is specified by the pattern of ones and zeros. The structuring element used for the project is disk-shaped structuring element due to its isotropy, disks and spheres are very attractive structuring elements.
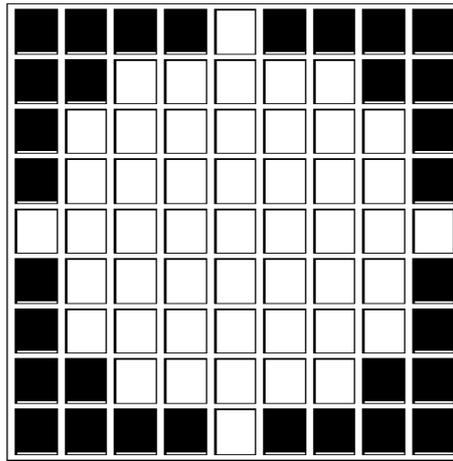
Figure 2: Disk-Shaped Structuring Element

Dilation

In dilation the size of object is increased by filling in holes, connecting areas and broken areas that are separated by spaces smaller than the size of the structuring element. Dilation is denoted with the symbol$\oplus$. The dilation formula is defined as:

$$A \oplus B = \bigcup_{b \in B} A_b$$

Erosion

In Erosion the size of object is decreasedand removes smaller objects by subtracting objects smaller than the structuring element. Erosion is denoted by the symbol Θ and it is given by the formula.

$$A \Theta B = \bigcap_{b \in B} A_{-b}$$

Closing

In closing we first perform dilation followed by erosion and it is given by formula:

$$A \cdot B = \left((A \oplus B)\Theta B\right)$$

Opening

In opening we first perform erosion followed by dilation.

$$A \circ B = \left((A \Theta B) \oplus B\right)$$

Feature Extraction

Feature extraction refers to the transfer of the input data i.e. MR images, into a different set of features. In the proposed system, the feature extraction unit houses the algorithms which are used to detect and isolate various desired features of the tumor region. This unit is most important because the accuracy of this process determines the performance of the classification algorithm used. [5]The features extracted are as follows below which will be used for the classification process of tumors identified:

Contrast

The contrast is determined by difference in the color and the brightness of objects.

$$\sum_{i,j=0}^{N-1} P i, j(i-j)^2$$

Homogeneity

The homogeneity is defined as the state of being homogeneous.

$$\sum_{i,j=0}^{N-1} \frac{Pij}{1+(i+j)^2}$$

Entropy

This feature is a statistical measure of random values used to characterize the texture of the input image.

$$\sum_{i,j=0}^{N-1} -\ln(Pij)\, Pij$$

Energy

This feature describes how grey levels are distributed.

$$\sum_{i,j=0}^{N-1} (Pij)^2$$

Correlation

It is a measure of linear relationship between the objects. Correlation is computed into what is known as the correlation coefficient, which ranges between -1 to +1.

$$\sum_{i,j=0}^{N-1} P_{i,j} \frac{(i-\mu)(j-\mu)}{\sigma^2}$$

Tumor Classification

The features extracted in the Feature Extraction unit are used by the Naïve Bayes and J48 classifiers to classify the tumor as benign or malignant tumor.

Naïve Bayes classifier works on the principle of conditional probability as given by the Bayes theorem. Bayes theorem gives the conditional probability of an event A given another event B has occurred. The problem with the Naïve Bayes is that if the number of features *n* is large then it is infeasible basing such a model on probability tables. We therefore reformulate the model to make it more tractable. Naïve Bayes classifier only works well with discrete values. Since we have dataset of continuous attribute values we cannot make use of Naïve Bayes algorithm. Gaussian Naïve Bayes is higher version of Naïve Bayes which can be applied on continuous dataset.

Gaussian Classification

In Gaussian Naïve Bayes algorithm we calculate mean, variance and standard deviation to calculate Gaussian probability density function (PDF).

$$\text{Mean } (\mu) = \frac{1}{n} * \sum_{i=1}^{n} xi$$

$$\text{Variance } (\sigma^2) = \frac{1}{n-1} * \sum_{i=1}^{n} (xi - \mu)^2$$

Standard deviation $(\sigma) = \sqrt{Variance}$

Algorithm

- Divide the continuous data set into the number of class variable.

- For each attribute estimate mean, standard deviation and variance for all the class variables.

- Calculate Gaussian probability density function for each attribute in the new data for all the class variable.

    Gaussian probability density function $= \frac{1}{\sqrt{2*\pi*\sigma}} * e^{\frac{-(x-\mu)^2}{2*\sigma^2}}$

- Multiply each Gaussian probability density function of each attribute in the data for all class variables.

- Whichever class has the highest Gaussian probability density function will be assigned to the new data.

J48 Algorithm

J48 is an open source Java implementation of the C4.5 algorithm. The C4.5 is implemented in WEKA, an open source software which is issued under the GNU General Public License. WEKA contains different machine learning algorithms required for data mining tasks like data preparation, classification, regression etc.

Imports used: weka.classifiers.trees.J48, weka. core. Instances.

Functions used: build Classifier(), classify Instance().

public void build Classifier(Instances instances)throws java. lang. Exception and generates the classifier specified by:

   build Classifier in the interface Classifier

Parameters:

   Instances are the data to train the classifier with.

public double classify Instance(Instance instance) throws java. lang. Exception that classifies an instance specified by:

   Classify Instance in interface Classifier

Overrides:

   Classify Instance in class Abstract Classifier

Parameters:

   Instance - the instance to classify

Returns:

   The classification for the instance

Experimental Results of Proposed System

The MRI image is given as an input to the proposed system. If the MRI image is RGB image then the image undergoes preprocessing where the image is converted to a grayscale image using averaging algorithm and contrast enhancement is being performed using contrast stretching. The image is then segmented using K-means Algorithm in order to extract the layer containing the tumor region. If the image is already preprocessed then it directly given to the segmentation stage. Thresholding is performed using Otsu's Algorithm to convert the gray scale image into binary image. Features are being extracted which are required to classify the type of tumor. The two classification algorithm used i.e. Naïve bayes and J48 are being compared in order to conclude which of these two algorithms provides an accurate result.

The proposed technique for preprocessing the tumor is applied to the images from the public dataset. The results of the preprocessing algorithm mentioned in section 2 is applied to figure 2, are shown in figure 3, 4 and 5.
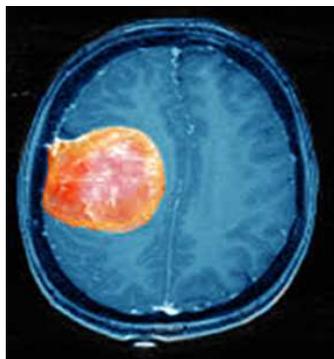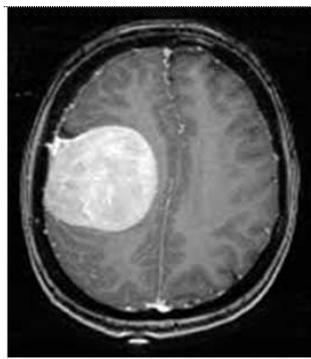

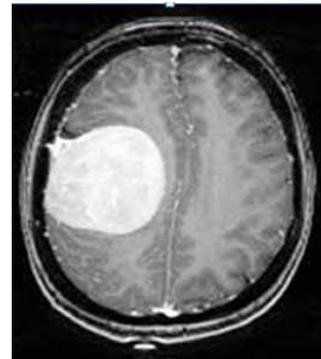**Figure 3: Original Image**   **Figure 4: Image after Gray**   **Figure 5: Image after**
                                             Scale Conversion   **Contrast Stretching**
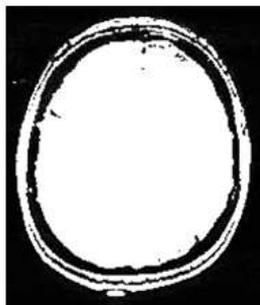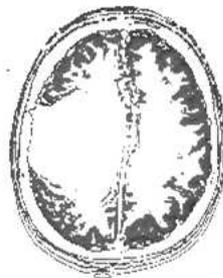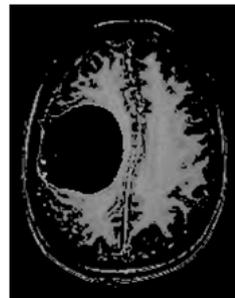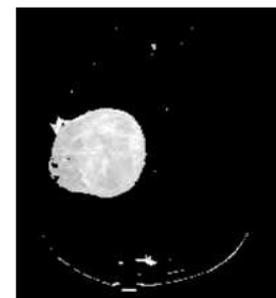

Figure 6: Image Segmentation using K-Means Algorithm

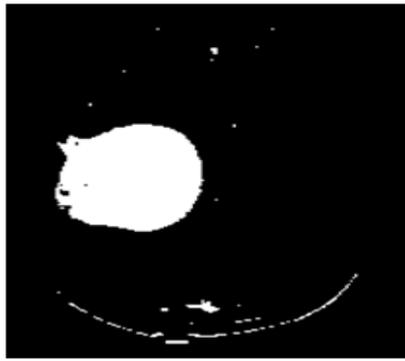The above images show the result of preprocessing and segmentation module.

Figure 7: Threshold Operation Performed   **Figure 8: Morphological Operation Performed**

Figure 9: Output after Feature Extraction   **Figure 10: Output after Classification**

These features are extracted as shown in figure 9 and fed to the Naïve Bayes and J48 classifiers in order to classify the tumor as shown in figure 10.

CONCLUSIONS

The proposed system focuses on detecting and classifying whether the brain tumor is benign or malignant based on abnormality of cell growth in the brain. In RGB to gray scale conversion the averaging algorithm is easier, efficient and more widely used algorithm. Contrast stretching is favorable for MRI images in enhancing the image. Binary image is preferred since operation becomes easier and efficient. Morphological operation is further performed in order to focus on the tumor region which helps in expansion and reduction of image pixels consisting four methods i.e. dilation, erosion, opening and closing. To classify the tumor, we need to extract features of tumor which is essential to distinguish the result to compare using two classification algorithms i.e. Naïve Bayes and J48.

Since time taken for detection in J48 is less which adds to its benefits. With methods like segmentation and thresholding, outputs obtain have greater accuracy and efficiency. We have used 110 MRI images for testing the system and out of 110 images J48 Algorithm gave correct result for 101 images whereas Naïve Bayes Algorithm gave correct result for only 72 images. The accuracy of algorithms based on our results is 91.818% (J48 Algorithm) and 65.45%(Naïve Bayes Algorithm).

We achieve more prominent results in J48 algorithm in comparison to Naïve Bayes algorithm which concludes that J48 algorithm is more accurate than Naïve Bayes algorithm in detecting the type of tumor in MR images.

REFERENCES

1. *Nameirakpam Dhanachandra∗, Khumanthem Manglem and Yambem Jina Chanu, Image Segmentation using K-means Clustering Algorithm andSubtractive Clustering Algorithm, Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Procedia Computer Science 54 (2015) 764 – 771.*

2. *Neeraj Battish Department of CSECGC-COE, Dapinder Kaur Department of CSECGC-COELandran, Mohali, Analysis of Pre-processing Methods using MRI Images, International Journal of Modern Computer Science (IJMCS) ISSN: 2320-7868 (Online)Volume 5, Issue 2, April 2017.*

3. *V. Velusamy, Dr. M. Karnan, Dr. R. Sivakumar, Dr. N. Nandhagopal, Enhancement Techniques and Methods for MRIA Review, V. Velusamy et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 397-403.*

4. *Poonam1, Jyotika Pruthi Computer Science and Engineering, Itm University, Gurgaon, India, Review of Image Processing Techniques for Automatic Detection of Tumor in Human Brain, Poonam et al, International Journal of Computer Science and Mobile Computing Vol.2 Issue. 11, November- 2013, pg. 117-122.*

5. *Poonam1SivaSankari.S, Sindhu.M, Sangeetha.R, ShenbagaRajan.AU.G. Student, Department of Computer Engineering, PSRR Engineering College, Sivakasi, Tamilnadu, India, Associate Professor, Department of Computer Engineering, PSRR Engineering College, Sivakasi, Tamilnadu, India, Feature Extraction of Brain Tumor Using MRI, International Journal of Innovative Research in Science, Engineering and Technology, 2018: Volume 7, Issue 6.*

6. *Jay Gholap Dept. of Computer Engineering College of Engineering, Pune, Maharashtra, India, Performance Tuning of j48 Algorithm For Prediction of Soil Fertility, Asian Journal of Computer Science And Information Technology 2: 8 (2012) 251– 252,2013: ISSN 2249-5126*

7. *Dr. Neeraj Bhargava, Girja Sharma Dept. of Computer Science, School of Engg. & System Sciences, MDS University, Ajmer, India, Dr. Ritu Bhargava Dept. of MCA, Govt. Women Engineering College, Ajmer, India, Manish Mathuria Dept. of C.E. & I. T., Govt. Engineering College, Ajmer, India, Decision Tree Analysis on J48 Algorithm for Data Mining, International Journal of Advanced Research in Computer Science and Software Engineering Volume 3, Issue 6, June 2013,ISSN: 2277 128X.*

8. *Deepshri Hebbalkar, Smita Naik, Anuj Divkar, Avinkumar Pednekar, Valerie Menezes & Shreedatta Sawant, Agnel Institute of Technology and Design, Goa, India, Breast Cancer Image Segmentation Using EKStrap and FCS Algorithm, International Journal of Computer Science Engineering and Information Technology Research Volume 7, Issue 3, May – June 2017, ISSN 2249-6831.*

9. *Nora Naik, Alan Pereira, Shruti Dhekne, Shubham Kadam & Ankita Kalangutkar, Agnel Institute of Technology & Design, Goa, India, Detection of Brain Tumor Using Enhanced K-strange Points Clustering and Morphological Filtering, International Journal of Computer Science Engineering and Information Technology Research Volume 7, Issue 3, May – June 2017, ISSN 2249-6831.*

10. *Nora Naik, Peter Braganza, Aaron Cordeiro, Ronan D'Souza & Rayster Fernandes, Agnel Institute of Technology & Design, Goa, India, Leukemia Prediction Using Random Forest Algorithm, International Journal of Computer Science Engineering and Information Technology Research Volume 8, Issue 3, Aug 2018, ISSN 2249-6831.*