Brain Tumor Detection: Tracing Brain Tumor Using Machine Learning

Navya Nair[1], Neevan R[2]
[1]Dept of Computer Application
[2]Assistant Professor, Dept of Computer Application
[1, 2] College of Engineering, Trivandrum, Kerala, India

Abstract- *Biomedical Image Processing is a growing and demanding field. CT scans, X-Ray and MRI are its different imaging methods. Even the smallest abnormalities in the human body can be identified using these techniques. With least error possible, medical imaging helps in extracting meaningful and reliable information from the images. MRI is considered as the most safest and reliable medical imaging process available to us.*

A brain tumor is a mass or growth of abnormal cells in our brain. However, primary brain tumors are the second most common childhood cancer and the most common robust childhood neoplasm, accounting for about 20% of all pediatric cancers. For any disease identification and treatment at the proper stage would increase the lifespan of the infected person. Likewise identifying and predicting brain tumor would improve the survival rate of the patient. But diagnosing tumor by a radiologist is a difficult task and time consuming. So this paper aims at making a machine capable of recognizing and tracing its area by using Random Forest Classifier .

Keywords- Brain Tumor, Machine Learning, Random Forest Classifier, Blocking , Trace the brain tumor area.

I. INTRODUCTION

The brain is one of the largest and the most complex organ in human body. One of the major life threatening disease in the world is brain tumor. Brain tumor occurs when abnormal cells form in the brain. The cause of brain tumor is unknown. Hence no prior precautions can be taken to avoid tumor.

The signs and the symptoms of a brain tumor vary greatly in size and the location where it occurs. Headaches, seizures, problems with vision, vomiting and mental alterations are some of the symptoms of brain tumor. Type, location and size of the tumor as well as the patient's age and general health are the different factors on which treatment of brain depends. For children and adults the treatment methods and schedules are different.

Surgery, radiation therapy and chemotherapy are the different methods which are used to treat brain tumorMedical Imaging plays a central role in the diagnosis of brain tumors. Early imaging methods – invasive and sometimes dangerous – such as pneumoencephalography and cerebral angiography have been abandoned in favour of non-invasive, high-resolution techniques, especially magnetic resonance imaging (MRI) and computed tomography (CT) scans, though MRI is typically the reference standard used. In general, Magnetic resonance imaging (MRI) is the first step towards diagonising brain tumor. Once it is detected that tumor is present in the brain, the most common way to determine the type of brain tumor is to look at the results from a sample of tissue after a biopsy or surgery. Neurosurgeons prescribe MRI scans because it helps in detecting smallest of abnormalities. As we have the MRI scans of the brain, using MRI scans we can detect tumor, its size and its location. This is where Computerized Image Processing is needed. This helps in automatically detecting and tracing tumor in brain with various feature extraction method and classifiers. In this paper, Random Forest Classifier is used to detect brain tumor.

II. METHODOLOGY

Brain Tumor detection makes use of brain MRI scans. The system uses C++ and Open CV based implementation to predict whether the input MRI scan is tumours or not. And if tumours the system can trace the area of tumor.

This system consists of five main modules. There is a module which deals with the image acquisition and pre-processing where the binarization, morphological operations are performed. At the end of preprocessing module we obtain an image with external contour representing the skull drawn on it. According to the external contour the image to resized for further processing. The second module is a feature extraction module in which the features such as homogeneity, variance, entropy, contrast etc. are extracted. The extracted features are added to feature vector which is then turned to a matrix for training. The third module is for training the data using Random Forest. Training involves feeding the model with training dataset and training labels. Training and testing are performed to predict whether a MRI image of brain is tumours or not. The fifth module deals with tracing the brain tumor infected area and blocking them.

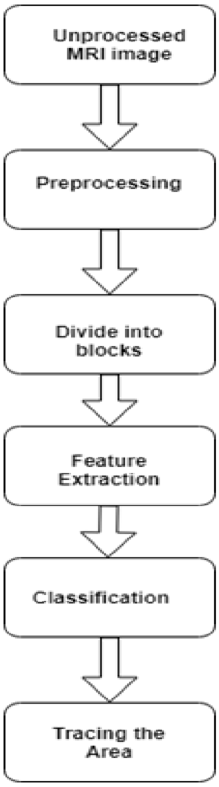

Figure 1: Overall process of the system

III. IMPLEMENTATION DETAILS

A. IMAGE ACQUISITION AND PREPROCESSING

First step is to acquire normal and abnormal CT images are collected from the available database.The given below image shows the obtained data.(fig.(a)-normal image ;fig.(b)tumours image .

Figure 2:
(a) Tumor MRI image of brain
(b)Normal image.

Before processing the image we must ensure that the image is free from unwanted data and to process the data further, need the data to be in the right format so that the results are accurate. This step is known as Preprocessing. Preprocessing involves processes such as gray scale conversion, noise reduction and noise removal, image reconstruction, image enhancement, and may involve steps such as skull removal from an MRI when it comes to medical images. Convertion to gray scale is one of the most common preprocessing practice. A gray scale image is often considered to be only black and white image, but this is not true. A black and white image has only two shades, i.e. black and white, so the intensity may be 1 or 0 at one point. However, a gray scale image consists of shades of grey with no apparent color. This means that every pixel represents the intensity value at that pixel without showing any color. Also unlike a black and white image, a gray scale image has many different shades with white being the lightest shade and black being the darkest.

Once the image is converted to gray scale image morphological operations are applied to dilate the image. Dilation of the image helps in figuring the outer contour or the skull portion of the brain. After dilation of the image we can find the contours of the image by using findContour function and then draw the external contour which represents the skull of the brain. A bounding rectangle is then drawn surrounding the image. The bounding box, is the smallest rectangle that canbe drawn around a set of points such that all the points are inside it, or exactly on one of its sides.Then resize the image around the bounding rectangle which helps in removing the unwanted area around the actual image.

This is the whole preprocessing stage whose outcome is the resized image for further processing. The resized image is then divided into smaller blocks or tiles using small strides. The actual image is these dataset images, pre-processing is applied.

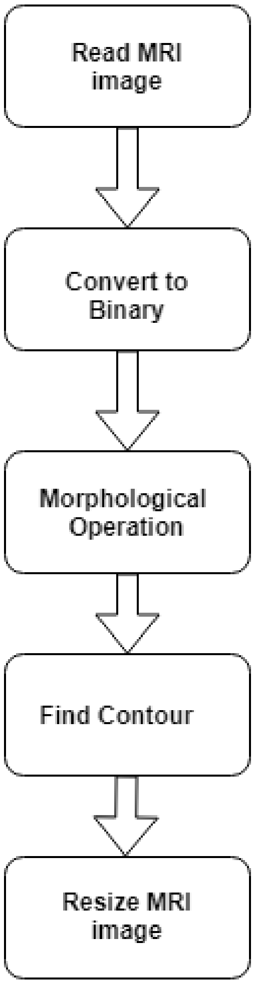

Figure 3: Preprocessing steps .

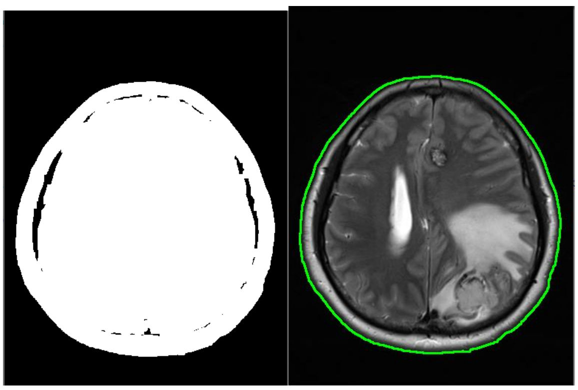

Figure 4 :
a) Dilated image
b) Image after drawing external contour

B. FEATURE EXTRACTION

The features such as contrast, homogeneity, entropy, variance, standard deviation, mean and energy are calculated. For texture classification Gray level co-occurrence matrix (GLCM) has proven to be very useful. Gray level co-occurrence matrix helps in calculating various textual parameters.

Information regarding pixels of pairs are collected by GLCM help understand the details about the overall image content. It is second-order statistics.

First order statistics are values from the original image which include variance. Whereas, second order statics is the relationship between group of two neighbouring pixels. A matrix is built up in degrees (0,45,90,135) at a distance of d=1. GLCM is prepared from grey scale pixels. It is observed how often one grey scale pixel value come horizontally, vertically or diagonally with its next pixel value. Here, we calculate GLCM values with distance=1 in the horizontal direction.

The features extracted are:

1) ENERGY : Energy is needed to do work. Its value range [0 1]. The value of the constant image is 1. The equation of energy is :

$$\sum_{I,J=0} P(i,j)^2$$

2) HOMOGENEITY: In short term, it is known as HOM. It is the calculation of distribution of elements of GLCM and GLCM diagonals. The value is 1 for diagonal GLCM and its range is [0,1]. The equation is :

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

3) CONTRAST: The other name of Contrast is 'Sum of Square Variance.Its defers the calculation of the intensity contrast that link the pixels with the neighbours over the entire image. At constant image the contrast value is 0. Weight increases exponentially (0,1,4,9) as the diagonal persists. Range=[0,size(GLCM,1)-1)^2]

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

The other features that are extracted from the image are:

- Mean: Mean specifies the average of all the dominating frequencies.
- Standard Deviation: Shows how much of the average (mean or expected value) variation or "dispersion" exists. A low standard deviation indicates that data points tend to be close to the average , while a high standard deviation indicates that data points are spread .
- Variance : The variance is a measure of how far a set of numbers is spread out.

For each block sized image these features are extracted and stored in each vector.

C. TRAINING WITH RANDOM FOREST

The image is divided into smaller blocks or tiles and for each block goes through the feature extraction process which results in feature vectors. This vectors are converted to matrix with number of rows meaning the number of training samples and the number of columns are the number of feature we have extracted from each block. This forms the training data with a single column training labels .

Random Forest was used to classify. Random Forest is a supervised algorithm for classification. As the name suggests, with a number of trees, this algorithm creates the forst. The forest gives the results of high precision in the Random Forest Classifier as the number of the tree increases.
A random forest's fundamental idea is to combine many decision trees into one model. Individually, predications made by decision trees (or people) may not be accurate, but together, on average, the predictions are closer to the mark.

Training random forest involves feeding known data along with the previously known values, thus forming a finite trainingset. Classifier then gets the intelligence to classify unknown data from the training set.

D.TESTING AND TRACING THE TUMOR AREA

Testing includes loading the trained model. An MRI image is then provided. This image is then preprocessed and feature extracted and passed to the trained model to predict tumor or not according to its intelligence. As the image is divided into blocks, the blocks tested and the blocks which are affected are traced in rectangles in the original image.

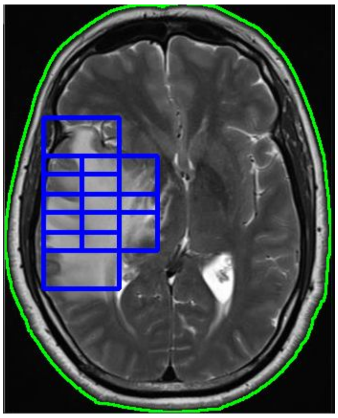

Figure 5: Brain Tumor Detected and Traced

IV. RESULTS

The results for brain tumor detection using Random Forest techniques are as follows:

Sl No	Algoritm used	Input Data	Accuracy
1	Random Forest	MRI	83%

Table 1: Accuracy Table

The inference from the above finding is that the model predicts whether a person have brain tumor or not with around 80 percent accuracy using Random Forest classifier which is considered to have greater accuracy then any other classifiers.

V. CONCLUSION & FUTURE SCOPE

The goal of this paper is to put forth a system that would detect and predict brain tumor .The features like entropy, homogeneity, energy etc seems to have the most predicting power with Random Forest classifier. The experimental results suggest that the proposed method achieves promising results in the detection of brain tumor and its parts. Some unexpected tracing of brain tumor occurs due to blocking of the image and training the model with these models.

The proposed system can be extended by extracting the more critical features from the dataset, the proposed system also can be integrated with health care websites and web apps. With properdatabase, this method can be applied to more diseases .Example: liver diseases, skin cancer, breast cancer identification and classification etc.

VI. ACKNOWLEDGMENT

We are gratified to the Department of Computer Application, College of Engineering, Trivandrum and each & everyone who helped us in carrying out this research work.

REFERENCES

[1] NehaRahi,ShardaVashisth ,Brain Tumor Detection and Classification with Feed Forward Back-Prop Neural Network, The Northcap University Gurgaon.

[2] Er. Kanchan Sharma, Er. Priyanka ,Er. AditiKalsh, Er.Kulbeer Saini, GLCM and its FEATURES , International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE).

[3] Meet Oza,RupalKapdi,Brain Tumor Disease Identification using Random Forest Classifiers,International Journal of Computer Science and Communications.

[4] Sanjeev Thakur, LixitKapoor,A Survey on Brain Tumor Detection Using Image Processing Techniques,Amity School of Engineering and Technology Noida.

[5] Vijay Kumar, Priyanka Gupta, Importance of Statistical Measures in Digital Image Processing,International Journal of Emerging Technology and Advanced Engineering