DETECTION AND DIAGNOSIS OF BRAIN TUMOR

Seema Singh[1], Navaneeth P S[2], Rahul Ramachandra M[3], Rohith K[4], Srikar C H[5]
[1]Associate Professor, Dept. of Electronics and Comm., BMS Institute of Technology and Management, Bangalore, INDIA
[2,3,4,5]Students, Dept. of Electronics and Comm., BMS Institute of Technology and Management, Bangalore, INDIA
Email: seemasingh@bmsit.in[1], navaneethps1995@gmail.com[2], rahulm.ramchandra@gmail.com[3], rohith2205@gmail.com[4], srikarsr@gmail.com[5]

Abstract

Brain Tumor is mass or growth of abnormal cells in or around brain. The chances of survival can be increased if the tumor is detected correctly at its early stage. Conventional methods used manual analysis or segmentation which could lead to wrong classification due to human errors. Due to complexity, automated systems that perform classification with high accuracy are desired. Over the last decade, many algorithms are used which uses techniques of image processing but it is increasingly moving towards machine learning concepts. Here we propose a method that uses support vector machines. The proposed system is designed to identify the tumor region, extract features, perform classification and determine if the tumor is benign or malignant. The proposed method was found to give 98% accuracy of classification.

Index Terms: Brain Tumor, Image segmentation, Feature Extraction, Support vector machine

I. INTRODUCTION

Brain tumor is the growth of abnormal cells in or around brain. Normal cells in the body age and die and are replaced by new cells. But cancer and other tumors disrupt this cycle. Tumor cells grow even though the body doesn't need them and unlike normal old cells they don't die. Thus the tumor continues to grow.

Since brain is a closed area, as the tumor grows, they cause pressure inside the skull which will cause inflammation and swelling which damages the brain. Tumors are of two types: primary and secondary tumors. Primary tumors originate in brain, emerge from various cells that are constituents of brain and central nervous system. They can be benign or malignant. Benign tumors rarely invade other tissues and have clearly defined borders. Most common type in adults is Meningiomas which is benign, astrocytomas. In children, its Medulloblastoma. Secondary tumors are also called as Metastatic tumors. They are malignant tumors which spreads from some other part of body usually lungs. They grow rapidly and invade the healthy tissues. The tumors are graded according to how aggressive they are, with lower-grade tumors often being benign and higher-grade tumors being malignant.

MRI is an imaging technique that produces high quality images of anatomical structures of brain. Image intensity in MRI depends upon four parameters. One is proton density (PD) which is determined by the relative concentration of water molecules. Other three parameters are T1, T2, and T2* relaxation, which reflect different features of the local environment of individual protons. In MRI, boundary of the tumor image is highly asymmetrical and irregular, so image segmentation is a great challenge to analyze the tumor. A key problem in medical imaging is automatically segmenting an image into its constituent heterogeneous processes. Automatic segmentation has the potential to positively impact clinical medicine by freeing physicians from the burden of manual labeling and by providing robust, quantitative measurements to aid in diagnosis and disease modeling.

Manual and Automated tumor segmentation:

The classification of Brain tumor segmentation methods can be made depending on the degree of human interaction as Manual segmentation, Semi-automatic segmentation and Fully automatic segmentation.

Manual segmentation involves delineation of the boundaries of tumor manually and representing region of anatomic structures with various labels. Manual segmentation requires software tools for the ease of drawing regions of interest (ROI), is a tedious and exhausting task. MRI scanners produce multiple 2-D slices and the human expert has to mark tumor regions carefully, otherwise it will generate jaggy images that lead to poor segmentation results.

Semi-automatic segmentation involves human interaction as least as possible. According to Olabarriaga et al [1], the semi-automatic or interactive brain tumor segmentation components consist of computational part, interactive part and the user interface. Since it involves both computer and human's expertise, result depends on both the combination. Efficient segmentation of brain tumor is possible through this strategy but it is also subjected to variations between expert users and within same user.

Fully automatic segmentation does not require intervention of human and segmentation of tumor is determined with the help of computer. It involves the human intelligence and is developed with soft computing techniques, which is a difficult task. Brain tumor segmentation has various properties which reduce the advantage of humans over machines.

These methods are likely to be used for large batch of images in research environment. However; these methods have not gained popularity for clinical practice, due to lack of transparency and interpretability.

II. RELATED WORK

Supervised and Unsupervised Segmentation Image segmentation objective is to segregate the image into mutually exclusive regions, which are similar with respect to pre-defined subsets. This objective can be accomplished using two methods of segmentation methods- Supervised and Unsupervised methods.

Unsupervised segmentation: If for training input vectors, target output is unknown, training method adopted is unsupervised learning. In the previous years, various unsupervised learning methods such as K-means and fuzzy clustering have gained popularity for brain tumor segmentation. The main aim of this type of segmentation is to segment the image into areas that have similar intensity and has well defined anatomic properties. Unsupervised segmentation of brain tumor achieve its anatomic goal by segmenting the image into at least two anatomically regions, one is tumor and other is edema. The advantage of this type is that it can handle very difficult tasks such as brain tumor segmentation; it produces an accurate segmentation of different regions present in heterogeneous tumor. Disadvantages of this segmentation are: number of regions is to be known before; tumors may not be specified clearly. This disadvantage can be avoided using skull stripping. Skull stripping is a pre-processing step to wipe out non-cerebral tissue such as fat, muscle, skin, skull which are not desired region of interest.

Supervised segmentation: In supervised learning, the network is provided with series of sample inputs and output is compared with expected response. It involves both training phase that uses labelled data that maps features to labels and testing phase is used to map labels to unlabeled data. The advantage of this type is that training set can be changed; it can reduce the manual task by providing labelled data. Irrespective of its advantages, it suffers from disadvantages that it requires patient specific training for brain tumor supervised segmentation and also human variability is also a concern.

Segmentation methods: In the segmentation process, accurate delineation of the tumor is responsible for early tumor diagnosis in clinical practice. Manifold approaches for brain tumor segmentation has been proposed. But no standard segmentation technique can produce satisfactory results for all imaging applications. In general, various segmentation techniques are as follows:

Threshold-based methods: It is the convenient and basic technique of image segmentation. It convert gray scale images into binary images.If we consider g(x,y) as the segmented image, we will get two outputs for the corresponding input image f(x,y). According to this technique, g(x,y)=1; if f(x,y) >= T and g(x,y)=0;if f(x,y) nodes. Clarke [2] was the first researcher to propose a supervised classification using an ANN approach for brain tumor segmentation in MR images. Executing ANN for brain tumor segmentation entangle problems of complexity ,the size of network becomes very large, time consuming process and large number of images are required for training the network.

A particular case of ANN is the self-organizing map (SOM). SOM is an unsupervised competitive learning algorithm. SOM automatically organizes itself according to the input data using a similarity factor like Euclidean distance. The brain cortex is organized in such a manner that closer neurons will give answers to the same kind of stimulus; this is one of the reason because of which SOM technique is used in visual pattern recognition.

Fuzzy C means (FCM) FCM is based on clustering which segments one class of data into two or more clusters. It works by casting each data point matching to each cluster center on the basis of distance between the cluster and the data point. The advantages of FCM algorithm are:

a) It gives best result for overlapped data set.

b) It produces comparatively better result than k- algorithm.

c) The application of FCM to MRI data has shown satisfactory results.

FCM is gaining popularity in the research area of brain tumor segmentation. This algorithm produces segmentation images that are clinically neuroanatomic tissue contrast from raw MRI data. A knowledge-based fuzzy clustering approach was proposed and implemented for the segmentation of the MRI images of brain tumor followed by 3-D connected components to build the tumor shape. To improve the accurate detection of stage and size of tumor, a combined method of the k-means and fuzzy c-means algorithms was proposed to deal with the segmentation of brain tumor. However the disadvantages of this method are that it takes more computational time and it is more sensitive to noise.

III. PROPOSED ALGORITHM

The proposed method consists of three steps:

Step 1: Pre-processing (including feature extraction and feature reduction);

Step 2: Training the kernel SVM

Step 3: Submit new MRI brains to the trained kernel SVM, and output the prediction.

Image pre-processing consists of Otsu binarization, K-means clustering and 2D DWT, after which thirteen features are extracted for training and classification of MRI images. Further a GUI is developed to ease the usage.

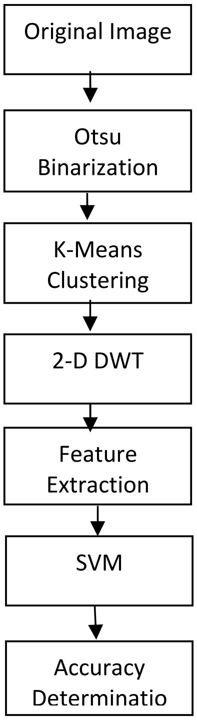


Fig1: Steps involved in brain tumor detection

Image Pre Processing Steps:

A) Otsu Binarization

In Otsu's method we exhaustively search for the threshold that minimizes the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes:

$$\sigma_w^2(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t)$$

Weights $\omega_{0,1}$ are the probabilities of the two classes separated by a threshold t and $\sigma_{0,1}^2$ are variances of these two classes.

B) K Means Algorithm

Clustering is the process of partitioning a group of data points into a small number of clusters. In general, we have n data points xi,i=1...n that have to be partitioned in k clusters. The goal is to assign a cluster to each data point. K-means is a clustering method that aims to find the positions μi,i=1...k of the clusters that minimize the square of the distance from the data points to the cluster. K-means clustering solves

argminc∑i=1k∑x∈cid(x,μi)2=argminc∑i=1k∑x∈ci‖x−μi‖22

where ci is the set of points that belong to cluster i.

C) 2-D Discrete Wavelet Transform

In case of 2D images, the DWT is applied to each dimension separately. There are 4 sub-band (LL, LH, HH, and HL) images at each scale. The sub-band LL is used for next 2D DWT. The LL sub band can be regarded as the approximation component of the image; while the LH, HL, and HH sub bands can be regarded as the detailed components of the image. As the level of decomposition increased, compacter but coarser approximation component was obtained.

D) Feature Reduction using Principal Component Analysis

Excessive features increase computation times and storage memory. Furthermore, they sometimes make classification more complicated, which is called the curse of dimensionality. It is required to reduce the number of features. PCA is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables while retaining most of the variations. It is achieved by transforming the data set to a new set of ordered variables according to their variances or importance. This technique has three effects: it orthogonalizes the components of the input vectors so that uncorrelated with each other, it orders the resulting orthogonal components so that those with the largest variation come first, and eliminates those components contributing the least to the variation in the data set.

Feature Extraction:

Following features are extracted using MATLAB simulation software.

- *Mean*
- *Standard Deviation*
- *Variance*
- *Contrast*
- *Correlation*
- *Energy*
- *Entropy*
- *Homogeneity*
- *RMS*
- *Smothness*
- *Skewness*
- *Kurtosis*
- *Inverse Differnce Method*

SVM Classification

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression problems. Given a set of training examples, each marked for belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a non-probabilistic binary linear classifier.

In addition to performing linear classification, SVMs can efficiently perform a nonlinear classification using what is called the kernel trick, implicitly mapping their inputs into high dimensional feature spaces.

K-Fold Stratified Organization

Since the classifier is trained by a given dataset, so it may achieve high classification accuracy only for this training dataset not yet other independent datasets. To avoid this over fitting, we need to integrate cross validation into our method. Cross validation will not increase the final classification accuracy, but it will make the classifier reliable and can be generalized to other independent datasets.

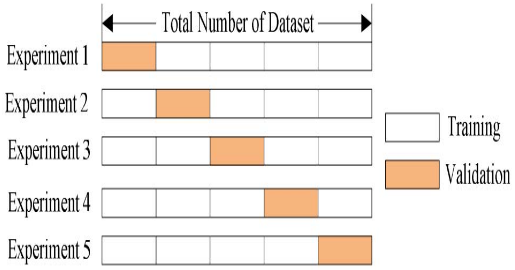


Fig 2: 5- Fold Cross Validation

Cross validation methods consist of three types: Random subsampling, K-fold cross validation, and leave-one-out validation. The K-fold cross validation is applied due to its properties as simple, easy, and using all data for training and validation. The mechanism is to create a K-fold partition of the whole dataset, repeat K times to use K − 1 folds for training and a left fold for validation, and finally average the error rates of K experiments. The schematic diagram of 5-fold cross validation is shown in Fig. The K folds can be purely randomly partitioned, however, some folds may have a quite different distributions from other folds. Therefore, stratified K-fold cross validation was employed, where every fold has nearly the same class distributions. Another challenge is to determine the number of folds. If K is set too large, the bias of the true error rate estimator will be small, but the variance of the estimator will be large and the computation will be time-consuming.

Alternatively, if K is set too small, the computation time will decrease, the variance of the estimator will be small, but the bias of the estimator will be large SVM by each value. Finally we select the optimal K value corresponding to the highest classification accuracy.

IV. RESULTS

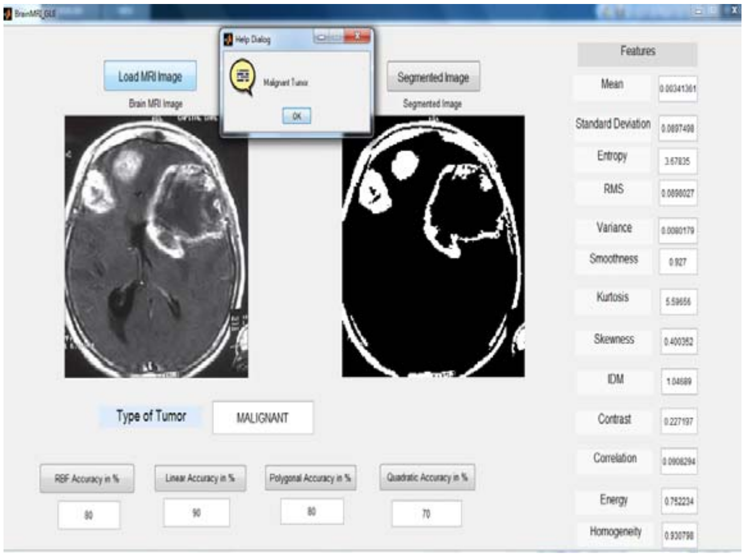
The input MRI image of brain tumor is shown in Fig. 3. Image pre-processing steps are shown in Fig. 4 and Fig. 5. Features are extracted and the numerical values of thirteen features are displayed in GUI as shown in Fig. 6 and 7. One case of Malignant and one case of Benign tumor is shown. Similarly various MRI images were tested and their accuracy statistics are displayed in GUI. The proposed method's classification accuracy is 98 %.

Fig 3: Input brain tumor image

Fig 4: Otsu thresholded image

Fig 5: Segmented Tumor

Graphical user interface was developed for better user experience. The user has to select the brain tumor image. Based on the feature extracted values the image is classified into benign or malignant.

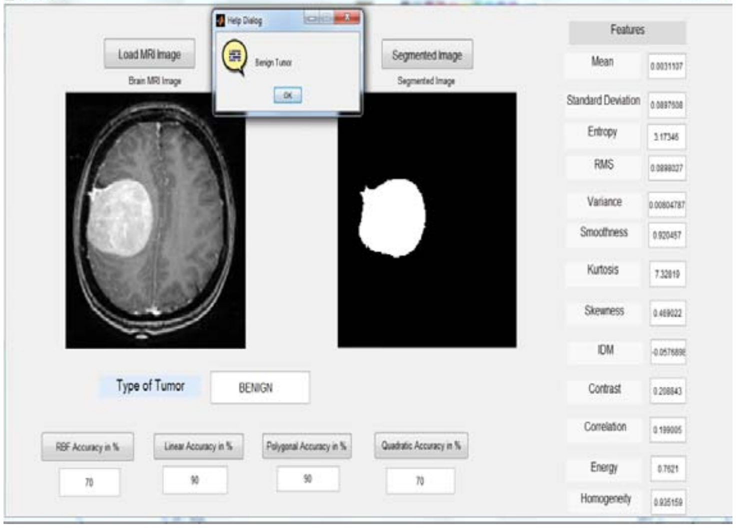


Type of Tumor: MALIGNANT

RBF Accuracy in %: 80
Linear Accuracy in %: 90
Polygonal Accuracy in %: 80
Quadratic Accuracy in %: 70

Mean: 0.00341361
Standard Deviation: 0.0897498
Entropy: 3.67835
RMS: 0.0898027
Variance: 0.0080179
Smoothness: 0.927
Kurtosis: 5.59656
Skewness: 0.400352
IDM: 1.04689
Contrast: 0.227197
Correlation: 0.0906294
Energy: 0.752234
Homogeneity: 0.930798

Fig 6 Malignant Tumor



Type of Tumor: BENIGN

RBF Accuracy in %: 70
Linear Accuracy in %: 90
Polygonal Accuracy in %: 90
Quadratic Accuracy in %: 70

Mean: 0.0031107
Standard Deviation: 0.0897508
Entropy: 3.17346
RMS: 0.0898027
Variance: 0.00804787
Smoothness: 0.920457
Kurtosis: 7.32819
Skewness: 0.469022
IDM: -0.0576898
Contrast: 0.208843
Correlation: 0.199005
Energy: 0.7621
Homogeneity: 0.935159

Fig 7 Benign Tumor

VI CONCLUSION

Current prevalent method used by pathologists for diagnosis of brain tumor is to manually observe the morphological changes in the cells by physical examinations, however manual observation is very subjective and poses several problems. The proposed method of automated diagnosis of brain tumor can reduce such discrepancies and workload of pathologists.

Image pre-processing consists of Otsu binarization, K-means clustering and 2D DWT, after which thirteen features (namely Mean, Standard Deviation, Variance, Contrast, Correlation, Energy, Entropy, Homogeneity, RMS, Smothness, Skewness, Kurtosis, Inverse Differnce Method) are extracted for training and classification of MRI images using SVM. Further a GUI is developed to ease the usage. Objective of detection and diagnosis of brain tumor was successfully achieved by classifying the image into Benign and Malignant.

REFERENCES

[1] S D Olabarriaga, A W M Smeulders, " Interaction in the segmentation of medical images: A survey", ELSEVIER Medical image analysis, 5(2001), pp 127- 142.

[2] L P Clarke, R P Velthuizen, S. Phuphanich, J.D. Schellenberg, J.A. Arrington, M. Silbiger, “MRI: Stability of three supervised segmentation techniques”, Magnetic Resonance Imaging, Vol 11 (1), 1993, pp 95–106

[3] Y Zhang, L Wu “An MR brain image classifier via support vector machine” progress in electromagnetics research, 2012.

[4] Chaplot, S., L. M. Patnaik, and N. R. Jagannathan, “Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network” Biomedical Signal Processing and Control, Vol. 1, No. 1, 86-92, 2006.

[5] El-Dahshan, E.-S. A., T. Hosny, and A.-B. M. Salem, Hybrid intelligent techniques for MRI brain images classification’’ Digital Signal Processing, Vol. 20, No. 2, 433-441, 2010.