



Research Journal of Pharmaceutical, Biological and Chemical Sciences

Novel Technique for Efficient Selection of Support Vectors to Enhance the Classification Accuracy of SVM.

E Rajinikanth*, and N Prabakaran.

School of Electrical and Electronics Sathyabama University, Chennai, Tamilnadu, India

ABSTRACT

Magnetic Resonance Imaging (MRI) based Brain Imaging is the most effective and reliable method for early detection of Brain tumor. In this paper we discuss a novel algorithm in brain tumor diagnosis of MRI Images. Our algorithm includes Segmentation, Features Extraction and Classification of Benign and Malignant Brain Tumors. This algorithm shows an efficient way in differentiating malignant tumors from benign tumors. The segmentation technique we use here is Fuzzy C Means Algorithm which helps us to segment the tumor region. We use GLCM Feature extraction technique to extract features from both the malignant and the benign tumors as well. Our train database has features of 50 Malignant and 50 Benign tumors. We use Non-Linear SVM to classify the tumors and the accuracy we found was 93%.

Keywords: Brain Tumors, Segmentation, Feature Extraction, Non-Linear SVM

**Corresponding author*

INTRODUCTION

Brain Cancer is one of the most common diseases, (i.e) uncontrolled growth of abnormal cells in the brain, which lead to death, especially in developed and developing countries. The brain tumors can be classified into two types malignant or cancerous and benign or non-cancerous. Malignant tumors can be divided into primary tumors, which is started within the brain and spread from somewhere else called as brain metastasis. Following are the symptoms of brain tumors: Headaches, seizures, problem with vision, vomiting and mental changes. The metastatic or secondary brain tumors are more common than primary brain tumors. In Australia the average economic cost of a case of brain cancer is \$1.9 million, the greatest of any type of cancer. MRI is currently the most effective imaging for cancer screening. A major class of problems in medical sciences involves the diagnosis of a disease, based upon various tests performed upon the patient [4]. When several tests are involved, the ultimate diagnosis may be difficult to obtain, even for a medical expert. The physician is interested in ascertaining whether the patient under examination exhibits the symptoms of a benign case or a malignant one [6]. In MRI images the brain tumors appear either darker than brain tissue or same intensity as brain tissue. Some tumors can be discovered because the patient has symptoms, other scan be showed on an imaging scan [9].

The most common primary brain tumors are

1. Gliomas(50.4%)
2. Meningiomas(20.8%)
3. Pituitary adenomas(15%)
4. Nerve Sheath tumors(8%)

In this paper, image processing tool is used to identify the region of lesions and SVM classifier is used to classify the brain images.

MATERIALS AND METHODS

The classification of brain images using SVM consists of the following steps: The brain images are collected from the MRI and the images are segmented for identifying the region where the tumor is present using Fuzzy C-means clustering techniques. The segmented images are classified using Support Vector Machine algorithm to find whether the tumor is cancerous or non-cancerous and analyze the performance of the classifier.

Image Acquisition:

Images are acquired from the GE Medical System's 3T Signal HDX Machine with high-resolution, T1-weighted and 3D turbo-flash sequence features.

Segmentation:

The process of dividing a digital image into multiple segments that is, set of pixels and similar pixels in a region according to some homogeneity criteria such as color, intensity or texture in order to locate and identify objects and boundaries in an image [2].In the present study, the pixels that have similar attributes are grouped together to get distinctly identifiable segmented regions. Though there are many segmentation methods like clustering, gradient-computation and region-growing are feasible. Some people have suggested that the clustering [3] is one of the simplest and widely used strategies adopted in segmentation of grey-level images and it is used in this study.

Segmentation by clustering:

Clustering is an unsupervised learning task, where it suggests that one needs to identify a finite set of categories known as clusters to classify the pixels. This is the technique to access the relationships between patterns of the data by organizing the patterns into clusters or groups such that the patterns within a cluster are more similar to each other than patterns belonging to different clusters. Segmentation by means of

clustering is adopted in this study since the numbers of groups needs to be identified from the brain images are known a priori.

Fuzzy C Means Clustering:

Fuzzy clustering depends on whether a pattern data belongs exclusively to a single cluster or to several clusters with different degrees of correspondence. The fuzzy C-means clustering (FCM) algorithm attempts to partition a total collection of elements of an image into a collection of fuzzy clusters with respect to a set of declared criteria. This algorithm involves assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point. More the data nearer to cluster-center more is its membership of belongingness towards the particular cluster center. The FCM algorithm can be summarized by following steps and Fig 1-Fig.4 shows their process as follows

Step 1: Fix the number of cluster and initialize the centers by random points from data set (based on observation).

$$c_j = \frac{\sum_{i=1}^N (u_{ij}^m)^* x_i}{\sum_{i=1}^N u_{ij}^m} \quad u_{ij} = \frac{1}{\sum_{k=1}^c \left[\frac{d_{ij}}{d_{jk}} \right]^{\left(\frac{2}{m-1} \right)}}$$

Step 2: Update the membership degrees by using

Where u_{ij} represents the membership of i^{th} cluster to j^{th} cluster d_{ij} represents Euclidean distance between i^{th} data and j^{th} cluster.

Step 3: Update centers using

Step 4: Repeat steps 2 and 3 until convergence. The convergence of this algorithm will be reached when the change in membership values is less than a given center value (threshold).

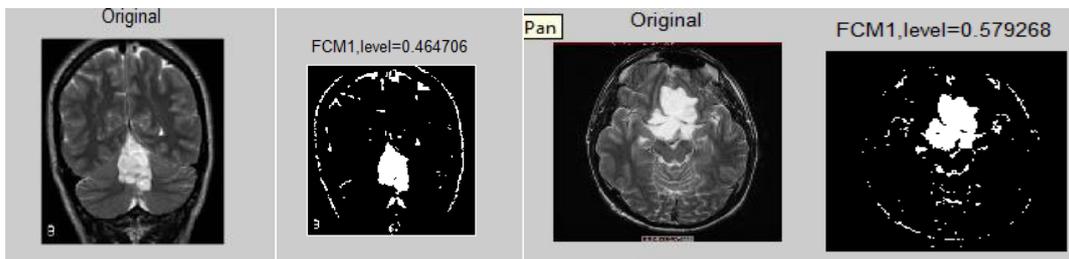


Fig.1

Fig.2

Fig. 3

Fig. 4

Fig.1- Original Image (Benign),

Fig.2- Segmented image using FCM

Fig. 3- Original Image (Malignant),

Fig. 4- Segmented Image using FCM

Feature Extraction:

A variety of feature sets represents the essential characteristics to be extracted to decide whether the lesion is benign or malignant [8].The brain image contains many features like auto correlation, correlation, energy, entropy, contrast, sum of variance, sum average. To classify the cancerous and non cancerous cells, we have to examine all the features mentioned above. In this approach we examined a set of 21 features. The Grey Level Co-occurrence Matrix method was used for feature extraction [5].

Classification of Image:

The technique which is used for classification of brain image is the Support Vector Machine (SVM). It is one of the machine-learning methods, based on the statistical learning theory, which aims to minimize the errors in the data set [7]. SVM is a promising technique for detection in a medical imaging application.

Theoretically, the advantage of SVM is that by choosing a specific hyperplane among many data in the feature space, the over fitting problem of the training data is reduced.

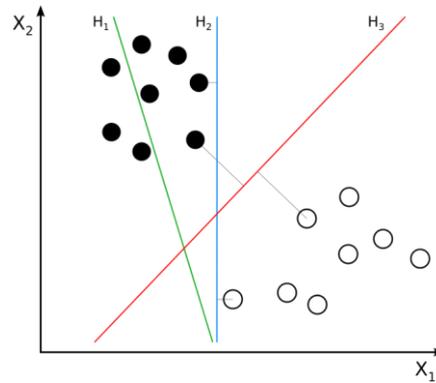


Fig 5: An optimal separating hyper plane

The distance between the nearest of the positive and negative data points and Fig.5 the separating hyperplane is called the margin of the SVM classifier [4]. The set of 8 images can be trained and 4 images can be tested. The train sets and the test sets are created. The SVM train instruction is used for training the classifier. This instruction accepts at least two input arguments. The instruction SVM classify is used to classify the test set using the knowledge gained from the training. Totally 12 images are segmented from that 8 images are trained and 4 images are tested.

Performance Analysis:

The most important thing is to analyze the performance of the SVM classifier. It is usually measured by sensitivity, specificity and Accuracy.

Sensitivity:

Sensitivity is the measure of how reliable a system is at making positive identifications, or, in other words, correctly identifying that which is inspected as being specifically that which is sought.[1]

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

Specificity:

Specificity is a measure of how well a system can make a negative identification, or indicate when something inspected is not what is being sought, but something else. A classification system with high specificity will rarely make the mistake of identifying what is being inspected as what is being sought. It is defined as the ratio of the number of true negatives, to the sum of true negatives and false positives.

$$\text{Specificity} = \frac{TN}{FP + TN}$$

Accuracy:

Accuracy which measures the global performance of the algorithm about the correct decisions.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

RESULTS AND DISCUSSION

In this Paper, the benign and malignant tumors are classified under the MATLAB environment. Fig. 6 shows the confusion matrix which is used to elevate the performance of the classifiers.

The brain images are segmented and the segmented images are underwent feature extraction. The images are classified using SVM classifier and performance of the classifier is analyzed.

Totally 12 brain images are taken for classification. The overall images are divided into 8 images for training and 4 images for testing. The normal and abnormal classification results of the test set can be presented as follows:

Number of test images: 8

Number of train images: 4

Sensitivity = $TP / TP + FN \times 100 = 97.826 \%$ Specificity = $TN / FP + TN \times 100 = 89.091 \%$

Accuracy = $TP + TN / TP + TN + FP + FN \times 100 = 93\%$

| | | Identified | |
|--------|-----|------------|----|
| | | Yes | No |
| Actual | Yes | TP | FN |
| | No | FP | TN |

Fig 6: Confusion Matrix

The future work can be done with more number of samples and adding more parameters for improving the performance which would help in defining the fine details of the brain tumor.

REFERENCES

- [1] Lee N.; "Computer-Aided Diagnostic Systems For Digital Mammograms", Louisiana State University, December, 2006.
- [2] RajeshwarDass, Priyanka and Swapna Devi, Image Segmentation techniques , International Journal of electronics & Communication technology, Vol.3,pp. 66-70, 2012.
- [3] Mansur R., Chhaya S. and Vandana S. 'Implementation of hard C-means clustering algorithm for medical image segmentation', Jr. of information, knowledge and research in electronics and communication engineering, Vol. 2,pp. 436-440, 2013.
- [4] NaseerM.Basheer and Mustafa H. Mohammed, Classification of breast masses in Digital Mammograms using Support Vector Machines, International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 3, pp. 57-63, 2013.
- [5] Aswini Kumar Mohanty, SwapnasiktaBeberta and Saroj KumarLenka, Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram, International Journal of Engineering Research and Applications, Vol. 1, pp. 687-693,
- [6] Reyes C. and Sipper M.;"A fuzzy-genetic approach to breast cancer diagnosis", Artificial Intelligence in Medicine, Vol. 17, pp. 131-155, 1999. [7] Martins L. et al; "Detection of Masses in Digital Mammograms using K-means and Support Vector Machine",
- [7] Cruz C.; "Automatic Analysis of Mammography Images: Enhancement and Segmentation Techniques", Engineering Faculty-Porto University, Porto, July 2011.
- [8] Mariam Ahmed Sheha and Mai S. Mabrouk, Pigmented skin lesion Diagnosis Using Gwometric and Chromatic Features., International Biomedical Engineering Conference, pp.115-200, 2014.
- [9] PV Nandhakumar and E.Rajinikanth, MRI Brain Abnormality using MLRPNN, Research Journal of Pharmaceutical, Biological and Chemical Sciences, pp.1584-1594, 2015