Classifying Brain Anomalies Using PCA And SVM

Rosy Kumari[1], Rishi Kumar Soni[2]

[1]CSE Dept. GGITS, Jabalpur
rsjhilmil@gmail.com
[2]Asst. Prof CSE Dept. GGITS, Jabalpur
rishi.soni.m@gmail.com

Abstract: *In this research paper an automated intelligent classification system is proposed which caters the need for classification of image slices after identifying abnormal MRI volume, for anomalies identification. Features are extracted by the use of Principal component Analysis(PCA).SVM classifier is to group items that have similar feature values into two categories as normal or abnormal.RBF kernel function is to classify non-linear datas. Experimental results shows that the proposed system have high classification accuracy of 98% and outperformed all other classifiers tested. Software used is MATLAB R2012.*

Keywords: *MRI, PCA, SVM, Kernel.*

I. INTRODUCTION

Medical images are usually obtained by X- rays and recent years by Magnetic Resonance Imaging . MRI is used as a valuable tool in the clinical and surgical environment because of its characteristics like superior soft tissue differentiation high spatial resolution and contrast. It does not use harmful ionizing radiation to patients. Magnetic Resonance Images are examined by radiologists based on visual interpretation of the films to identify the presence of abnormal tissue. The sensitivity of the human eye in interpreting large numbers of images decreases with increasing number of cases, particularly when only a small number of slices are affected. Hence there is a need for automated systems for analysis and classification of such medical images.

The MRI may contain both normal slices and defective slices. The defective or abnormal slices are identified and separated from the normal slices and then these defective slices are further investigated for the detection of tumor tissues.The motivation behind this paper is to develop a machine classification process for evaluating the classification performance of different classifiers to this problem in terms of statistical performance measure. The categorization of slices into normal and abnormal is done using statistical features of images such as mean, variance,and co-occurrence based textural features of image such as energy, entropy, difference moment, inverse difference moment and correlation.For comparative analysis, SVM Classifier with linear and nonlinear type of kernels, are also implemented using the same data sets.

II. PROPOSED METHOD

This section explains the proposed system design and methodology which concerns its fundamental design and the features of each parts of proposed method.The proposed method consists of phases such as Feature extraction from MRI data set using PCA. Feature extraction using principal component analysis (PCA) selects most valuable features. This section also explains the classification stage in which support vector machine is used for classification. Dataset used in this work consists of 256×256 pixels, T2-Weighted MRI brain images. These images are taken from Ishaan MRI & CT Scan Centre of original patients.In this study dataset having 60 images in which 45 images are abnormal having tumor and 15 images are normal. This database contains images of axial plane. The following figure shows Normal and Abnormal brain images.
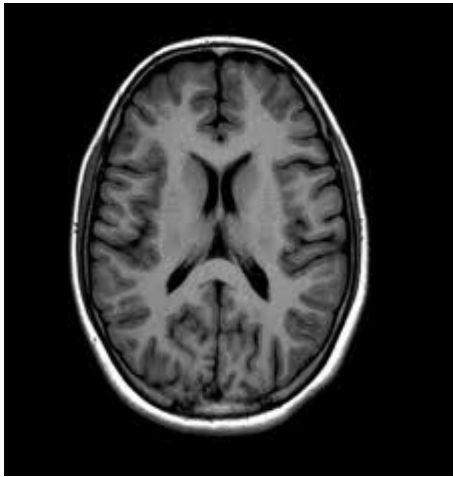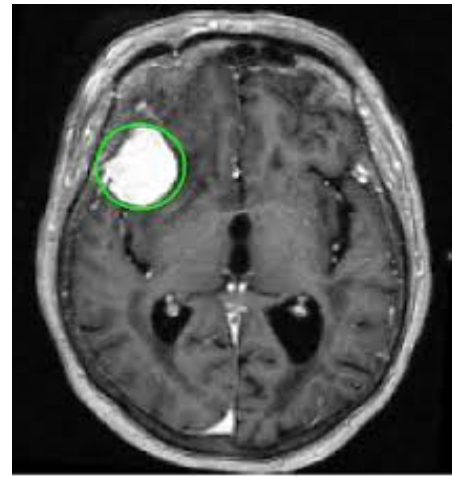
Fig2.1:NormalBrainMRI

Fig 2.2 Abnormal Brain MRI Image

III. FEATURE EXTRACTION

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes
poorly to new samples. In this paper Principal Component Analysis is formulated to obtain statistical texture features. It is a way of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analysing data.

The other main advantage of PCA is that once you have found these patterns in the data, and you compress the data, ie. by reducing the number of dimensions, without much loss of information. This technique used in image compression . The algorithm is described below

IV. CLASSIFICATION

Classification is one of the most important tasks for different application such as text categorization, tone recognition, image classification, micro-array gene expression, proteins structure predictions, data

STEP 1: Organize a data set as an m*n matrix, where m is the number of measurement types and n is the number of trials.

STEP 2: Subtract off the mean for each measurement type.

STEP 3: Calculate the covariance matrix.

STEP 4: Calculate the eigenvectors and eigenvalues of the covariance matrix.

STEP 5: Choose the components and form a feature vector.

STEP 6: Derive the new dataset.

PCA can be used to approximate the original data with lower dimensional feature vectors. The basic approach is to compute the eigen vectors of the covariance matrix of the original data, and approximate it by a linear combination of the leading eigen vectors. By using PCA procedure the test image cab be identified by first, projecting the image onto the eigen face to obtain the corresponding set of weights, and then comparing with the set of weights of the faces in the training set.

Classification etc. Most of the existing supervised classification methods are based on

traditional statistics, which can provide ideal results when sample size is tending to infinity. It employs two phases of processing- training phase and testing phase.

In training phase, characteristic properties of image features are isolated and a unique description of each classification category is created. In testing phase, these features space partitions are used to classify image features.

4.1 SUPPORT VECTOR MACHINE

SVMs are set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classification. A special property of SVM is SVM simultaneously minimize the empirical classification error and maximize the geometric margin. Two parallel hyperplanes are constructed on each side of the hyperplane that separate the data. These separating hyperplane is the hyperplane that maximize the distance between the two parallel hyperplanes. An assumption is made

$f(x) = w\,T\,x + b = 0$,

where w is the unit vector and b is a constant.

In Figure below, SVM classification with a hyperplane that minimizes the separating margin between the two classes are indicated by data points marked by "X" s and "O"s. Support vectors are elements of the training set that lie on the boundary hyperplanes of the two classes.
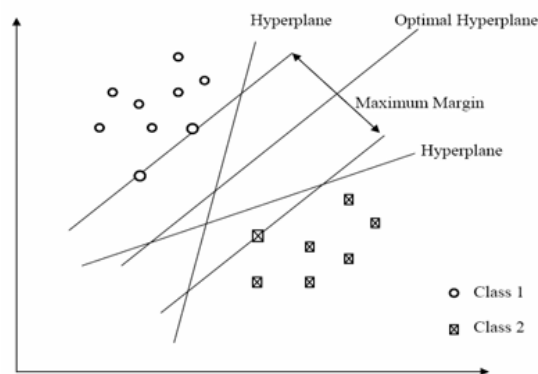
that the larger the margin or distance between these parallel hyperplanes the better the generalization error of the classifier will be.Let us begin with the simplest case, in which the training patterns are linearly separable. That is, there exists a linear function of the form $f(x) = w\,T\,x + b$ (1)

such that for each training example xi, the function yields ( ) $\geq 0\ if\ x$ for $= +1, i\ y$

and f(xi)<0 for yi=-1.

In other words, training examples from the two different classes are separated by the hyperplane

Fig 4.1: Linear SVM

4.2 KERNEL SELECTION OF SVM

A special property of SVM is SVM simultaneously minimize the empirical classification error and maximize the geometric margin. So SVM called Maximum Margin Classifiers.SVM is based on the Structural risk Minimization (SRM). SVM map input vector to a higher dimensional space where a maximal separating hyper plane is constructed. There are many kernel functions in SVM, so how to select a good kernel function is also a research issue. However for general purposes there are some popular kernel functions

(a). Linear kernel
(b).Polynomial kernel
(c).RBF kernel
(d).Sigmoid kernel

In these popular kernel functions, RBF is the main kernel function because of following reasons :

1. The RBF kernel nonlinearly maps samples into a higher dimensional space unlike to linear kernel.
2. The RBF kernel has less hyper parameters than the polynomial kernel.

3. The RBF kernel has less numerical difficulties.

V. RESULT

The result is discussed below which comprises of 60 images out of which 45 images are cancerous and the rest 15 images are non cancerous. Image read from database is further classified as cancerous or non cancerous which is discussed below.
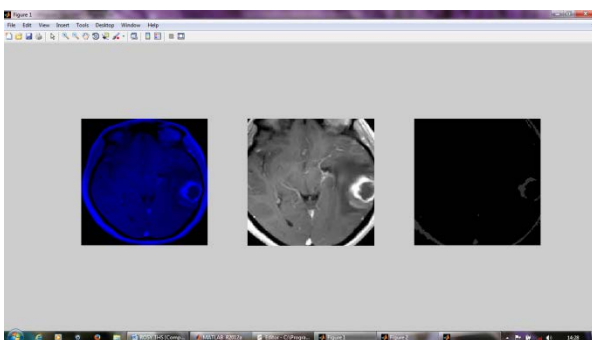

Fig 5.1: Query images
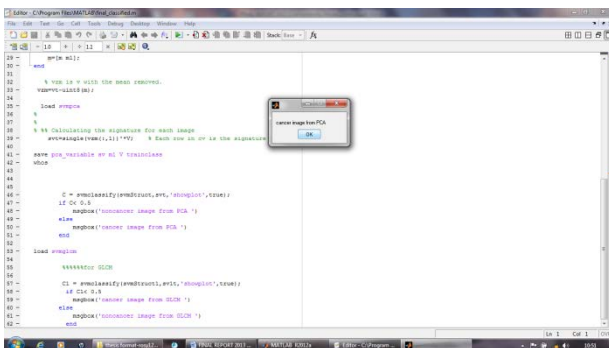

Fig 5.2: RGB to gray image

Fig 5.3: Cancerous MRI

VI. CONCLUSION

The approach proposed for feature extraction using principal component analysis yields better performance than gray level cooccurence matrix as PCA gives 100% accuracy compared to GLCM which is only 82% accurate. Further SVM as a classifier for classification of brain images provides a good classification efficiency. The results indicate that SVM approach for classification yielded better performance than other supervised classifiers such as ANN and K-NN.

The sensitivity, specificity and accuracy is also improved. The proposed approach is computationally effective and yields good result. This automated analysis system could be further used for classification of images with different pathological condition,types and disease status.

Feature extraction	Sensitivity	Specificity	Accuracy
PCA	92	96.43	100
GLCM	90	92.56	82.35

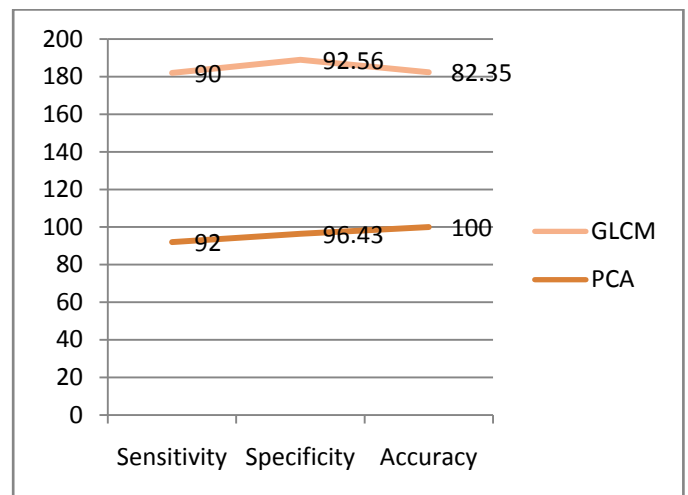Table 5.4 : Comparative study of feature extraction techniques

Fig 5.5: Graph showing performance of PCA and GLCM

Researchers are also using different imaging techniques to measure different types of brain activity

following strokes and traumatic head injuries. New imaging techniques bring new means for peering into the human body, helping to reduce the need for more invasive diagnostic and treatment procedures. Depending on the imaging technique and what diagnosis is being considered, image

processing and analysis can be used to determine the diameter, volume and vasculature of a tumor or organ;

flow parameters of blood or other fluids and microscopic changes that have yet to raise any otherwise discernible flags.

The proposed approach is computationally effective and yields good result. This automated analysis system could be further used for classification of images with different pathological condition,types and disease status. The future work is to improve the classification accuracy by extracting more features and increasing the training data set.

REFERENCES

01.T. Joachims, "Transductive inference for text classification using support vector machines," published at the Int. Conf. Machine learning, Solvenia, 1990.

02.Y.S. Hwang and S.Y. Bang,"An efficient method to construct a radial basis function neural network classifier ", Neural Networks, vol.10, no. 8, pp. 1495-1503,1997.

03.P.C. Chen and T.Pavlidri, Segmentation by texture using a co-occurence matrix and a split and merge algorithm" Computer Graphics And Image Processing, 10,pp.172-182, 1979

04.C.J.C.Burges, " A tutorial onsupport vector machines for pattern recognition", Data mining and knowledge discovery.2(2)L121-167,1998.

05.J.A. Freeman and D.M. Skapura,"Neural Networks Algorithm, Applications and Programming Techniques", Addison-Wesley Publishing Company,2002.

06.J.C. Russ, "The Image Processing Handbook", Raleigh, NC: CRC Press,1995.

07.N. Cristianini and J. Shawe-Taylor," An introduction to Support vector machine,U.K, Cambridge Univ. Press,2000.

08.S.C. Bushong," Magnetic Resonance Images physical and biological principles".Mosby Year Book, New York,NY,1996.

09 .B.Scholkopf and A.J. Smola," Learning with Kernels",MIT Press,2002.

10.A.K.Jain, "Fundamentals of digital Image Processing",Prentice Hall of India,1997.

11.A.E.Lashkari, A neural network basedmethod for brain abnormality detection in MR images using Gabor Wavelets, International journal of computer Applications, vol.4 no.7,2010.

12. Ramesh Babu Durai C.,Balaji V.,Improved content based Image retrieval using SMO and SVM classification technique , European journal of scientific Research, vol. 69, no.4,pp.560-564,2012.

13.Fei Peng Kehong Yuan Shu Feng Wufan Chen, Proceeding of international conference on Bioinformatics, pp.2495-2498,2008.

14.Ali Reza Fallah, Mohammad Pooyan, Hassan Khotanlou, A new approach for classification of human brain CT images based on morphological operations, J. Biomedica Science and Engineering,Vol.3, pp.78-82,2010.