A Novel Approach for Feature Extraction and Selection on MRI Images for Brain Tumor Classification

V.P.Gladis Pushpa Rathi[1] and Dr.S.Palani[2]

[1]Department of Computer Science and Engineering, Sudharsan Engineering
College Sathiyamangalam, Pudukkottai , India
gladispushparathi@gmail.com
[2]Department of Electronics and Communication Engineering , Sudharsan
Engineering College Sathiyamangalam, Pudukkottai , India
palani_keeranur@yahoo.co.in

ABSTRACT

Feature extraction is a method of capturing visual content of an image. The feature extraction is the process to represent raw image in its reduced form to facilitate decision making such as pattern classification. The objective of this paper is to present a novel method of feature selection and extraction. This approach combines the Intensity, Texture, shape based features and classifies the tumor as white matter, Gray matter, CSF, abnormal and normal area. The experiment is performed on 140 tumor contained brain MR images from the Internet Brain Segmentation Repository. PCA and Linear Discriminant Analysis (LDA) were applied on the training sets. The Support Vector Machine (SVM) classifier served as a comparison of nonlinear techniques Vs linear ones. PCA and LDA methods are used to reduce the number of features used. The feature selection using the proposed technique is more beneficial as it analyses the data according to grouping class variable and gives reduced feature set with high classification accuracy.

KEYWORDS

Linear Discriminant Analysis, BrainTumor, Shape, Intensity, Texture, PCA, SVM, MRI

1. INTRODUCTION

Early and accurate diagnosis of brain tumor is the key for implementing successful therapy and treatment planning. However the Diagnosis is a very challenging task due to the large variance and complexity of tumor characterization in images, such as size, shape, location and intensities and can only be performed by professional neuroradiologists. In the recent past several research works have been done for the diagnosis and treatment of brain tumor. MRI is the viable option now for the study of tumor in soft tissues. The method clearly finds tumor types, size and location. MRI is a magnetic field which builds up a picture and has no known side effects related to radiation exposure. It has much higher details in soft tissues. Researcher had proposed various features for classifying tumor in MRI. The statistical, Intensity, Symmetry, Texture features etc,

which utilize gray value of tumors are used here for classifying the tumor. However the gray values of MRI tend to change due to over –enhancement or in the presence of noise.

In image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (much data, but not much information) then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

 Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data, one of the major problems stems from the number of variables is involved. 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. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy.

Feature selection is the technique of selecting a subset of relevant features for building robust learning models By removing most irrelevant and redundant features from the data, feature selection helps improve the performance of learning models by:

- Alleviating the effect of the curse of dimensionality.
- Enhancing generalization capability.
- Speeding up learning process.
- Improving model interpretability.
-

Feature selection also helps people to acquire better understanding about their data by telling them which are the important features and how they are related with each other.

This Paper is organized as follows. In section 2 we describe normalization, and feature extraction and selection. In section 3 comparative analysis of PCA and LDA is described. In section 4 tumor classification and experimental results are discussed. The conclusions are given in section 5.

2. PROPOSED METHOD

The architecture of our system is illustrated in Figure 1.The major components of our system are Brain tumor Database, Normalisation, Feature selection, Feature extraction and Classification.
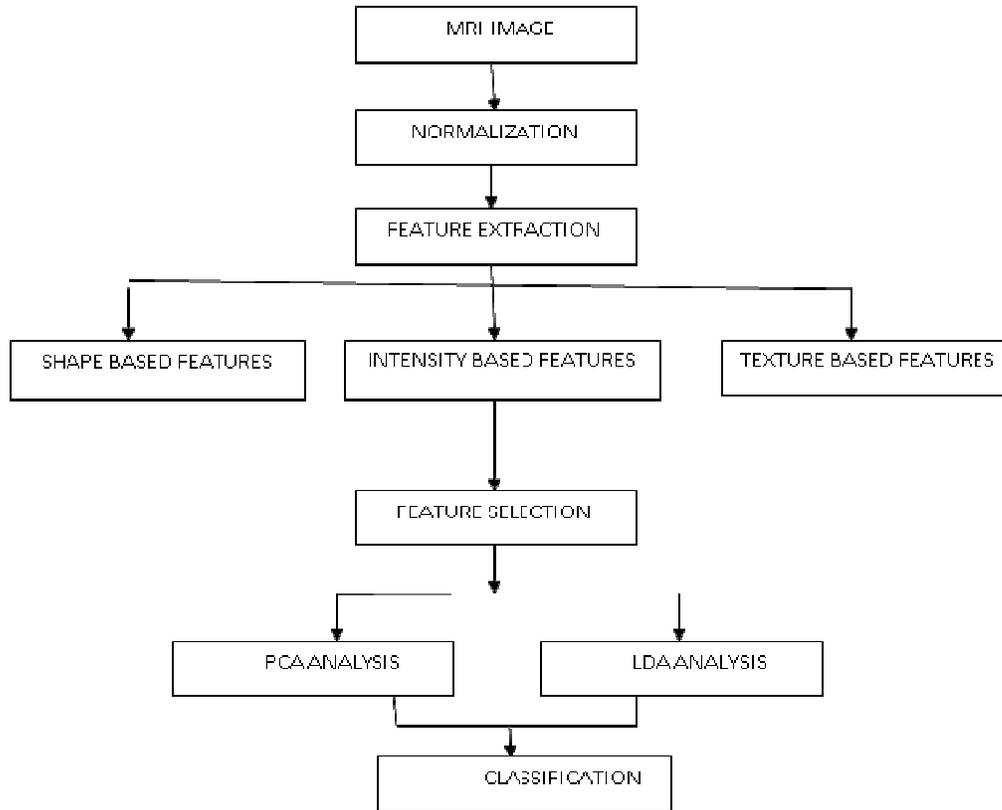
Figure 1. Architecture of proposed method

2.1. Data Description

Experiments are conducted on MR images collected from 20 different patients with gliomas. Each patient has 3 sequences of MR images T1, T2 and FLAIR. Each volume contains 24 slices in axial plain with 5 mm slice thickness. MR imaging was performed on 3.0T siemens devices. The imaging conditions of different protocols are; T1 weighted, T2 weighted, and Flair weighted. The MRI image data description of the proposed method is shown in table 1.Each set of features are individually normalized to the range of 0 to 255.

Table 1. Data Description

Attribute	Description	Value
Age	Age in Years	17 to 83
Sex	Sex	Men -46, Women -52
Matrix size	Size of the matrix	192*256*192
Voxel size	Size of the voxel	0.98*0.98*1mm
Sequences	MRI image sequences	Axial 3D T1 weighted , Sagittal 3D T2 weighted , Fluid Attenuated Inversion Recovery (FLAIR)

2.2. Normalization

Initially, these MRI images are normalized to gray level values from 0 to 1 and the features are extracted from the normalized images. Since normalization reduces the dynamic range of the intensity values, feature extraction is made much simpler.

2.3. Feature Extraction

Features, the characteristics of the objects of interest, if selected carefully are representative of the maximum relevant information that the image has to offer for a complete characterization of a lesion. Feature extraction methodologies analyse objects and images to extract the most prominent features that are representative of the various classes of objects. Features are used as inputs to classifiers that assign them to the class that they represent. The purpose of feature extraction is to reduce the original data by measuring certain properties, or features, that distinguish one input pattern from another pattern. The extracted feature should provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the image into feature vectors. In this proposed method we extract the following features.

> Shape Features   - circularity, irregularity, Area, Perimeter, Shape Index
> Intensity features – Mean , Variance, Standard Variance, Median Intensity, Skewness, and Kurtosis
> Texture features   –Contrast, Correlation, Entropy, Energy, Homogeneity, cluster shade, sum of square variance.

Accordingly, 3 kinds of features are extracted, which describe the structure information of intensity, shape, and texture. These features certainly have some redundancy, but the purpose of this step is to find the potential by useful features. In the next step the feature selection will be performed to reduce the redundancy.

2.4. Feature Selection

Feature selection (also known as subset selection) is a process commonly used in machine learning, wherein a subset of the features available from the data is selected for application of a learning algorithm. The best subset contains the least number of dimensions that contributes to high accuracy; we discard the remaining, unimportant dimensions.

2.4.1. Forward Selection

Start with no variables and add them one by one, at each step adding the one that decreases the error the most, until any further addition does not significantly decrease the error. We use a simple ranking based feature selection criterion, a two –tailed t-test, which measures the significance of a difference of means between two distributions, and therefore evaluates the discriminative power of each individual feature in separating two classes. The features are assumed to come from normal distributions with unknown, but equal variances. Since the correlation among features has been completely ignored in this feature ranking method, redundant features can be inevitably selected, which ultimately affects the classification results. Therefore, we use this feature ranking method to select the more discriminative feature, e.g.by applying a cut-off ratio (p value<0.1), and then apply a feature subset selection method on the reduced feature space, as detailed below

2.4.2. Backward Selection

Start with all the variables and remove them one by one, at each step removing the one that decreases the error the most (or increases it only slightly), until any further removal increases the error significantly. To reduce over fitting, the error referred to above is the error on a validation set that is distinct from the training set. The support vector machine recursive feature elimination algorithm is applied to find a subset of features that optimizes the performance of the classifier. This algorithm determines the ranking of the features based on a backward sequential selection method that remove one feature at a time. At each time, the removed feature makes the variation of SVM based leave-one-out error bound smallest, compared to removing other features.

2.5. Classification

There are many possible techniques for classification of data. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are the two commonly used techniques for data classification and dimensionality reduction. Linear Discriminant Analysis easily handles the case where the within-class frequencies are unequal and their performance has been examined on randomly generated test data. This method maximizes the ratio of between-class variance to the within-class variance in any particular data set thereby guaranteeing maximal separability. The use of Linear Discriminant Analysis for data classification is applied to classification problem in speech recognition We decided to implement an algorithm for LDA in hopes of providing better classification compared to Principal Components Analysis. The prime difference between LDA and PCA is that PCA does more of feature classification and LDA does data classification. In PCA, the shape and location of the original data sets change when transformed to a different space whereas LDA doesn't change the location but only tries to provide more class separability and draw a decision region between the given classes. The classification process is divided into the training phase and the testing phase. In the training phase known data are given. In the testing phase, unknown data are given and the classification is performed using the classifier after training. The accuracy of the classification depends on the efficiency of the training.

2.5.1. Linear Discriminant Analysis

LDA methods are used in statistics, pattern recognition, and machine learning to find a linear combination of features. LDA attempts to express one dependent variable as a linear combination of other features or measurements. LDA is also closely related to PCA and factor analysis in that they both look for linear combination of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA on the other hand does not take into account of any difference in class, and factor analysis builds the feature. Combination is based on differences rather than similarities. LDA searches for those vectors in the underlying space that best discriminable among classes. More formally given a number of independent features relative to which the data is described, LDA creates a linear combination of those which yields the largest mean differences between the desired classes. We define two measures: 1) one is called within- class scatter matrix as given by

$$Sw = \sum_{j=1}^{c} \sum_{i=1}^{Nj} \left(x_i^{\ j} - \mu_j \right)\left(x_i^{\ j} - \mu_j \right)^T \qquad (1)$$

where $x_i^{\ j}$ is the ith sample of class j, μ_j is the mean of class j, c is the number of classes, and μ_j is the number of samples in class j and 2)between class scatter matrix

$$Sb = \sum_{j=1}^{c} \left(\mu_j - \mu\right)\left(\mu_j - \mu\right)^T \tag{2}$$

where μ represents the mean of all classes.

2.5.2. Support Vector Machine

Support vector machines are a state of the art pattern recognition technique grown up from statistical learning theory. The basic idea of applying SVMs for solving classification problems can be stated briefly as follows: a) Transform the input space to higher dimension feature space through a non-linear mapping function and b) Construct the separating hyperplane with maximum distance from the closest points of the training set.

In the case of linear separable data, the SVM tries to find among all hyper planes that minimize the training error, the one that separates the training data with maximum distance from their closest points

$$w \bullet x + b = 0 \tag{3}$$

with w and b weight and bias parameters respectively.

In order to define the maximal margin hyperplane (MMH) the following constrains must be fulfilled:

$$\text{Minimize } \frac{1}{2} \| w \|^2 \ with y_i \left(w \bullet x_i + b\right) \geq 1 \tag{4}$$

This is a classic nonlinear optimization problem with inequality constraints. It can be solved by the karush-kuhn-Tucker (KKT) theorem by introducing Lagrange multipliers

$$\text{maximize} \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j a_i a_j x_i^{T} x_j \tag{5}$$

$$\text{subject to } \sum_{i=1}^{l} a_i y_i = 0 \ and \ a_i \geq 0 \tag{6}$$

The solution of w is:

$$w = \sum_{i=1}^{l} a_i y_i x_i \tag{7}$$

The only nonzero solutions define those training data (usually a small percentage of the initial data set) that are necessary to form the MMH and are called support vectors. The optimal hyper plane theory is generalized for non-linear overlapping data by the transformation of the input vectors into a higher dimensional feature space through a mapping function

$$x_i \in R^n \rightarrow z(x) = \left[a_1 \Phi_1(x), a_2 \Phi_2(x), \dots a_n \Phi_n(x)\right]^T \in R^f$$

The KKT conditions transform to

$$\text{Maximize } \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j a_i a_j K(x_i x_j) \tag{8}$$

$$\text{subject to } \sum_{i=1}^{l} a_i y_i = 0 \, and a_i \geq 0 \tag{9}$$

The optimization problem is solved using the MATLAB optimization toolbox

2.6. Experiment Results

In all the selected 60 features, there are 22 Intensity based features, 5 Shape based features, 33 texture based features . It is found that there are 3 kinds of features extracted in our work and are all useful for the classification. Besides, the distribution of T1, T2, and FLAIR are 10, 20,30 respectively. It means FLAIR provides the most information for tumor segmentation, T2 provides less and T1provides the least. This result is in accordance with the conclusion in Medical Imaging that FLAIR and T2 are more sensitive in pathological discrimination than T1. The distribution of selected features is shown in table 3.

Table 3: Distribution of Selected Features

Features	T1	T2	FLAIR	TOTAL
Intensity	6	5	11	22
Shape	1	1	3	5
Texture	8	5	20	33
Total	10	20	30	60

Efficiency or accuracy of the classifiers for each texture analysis method is analysed based on the error rate. This error rate can be described by the terms true and false positive and true and false negative as follows:

True Positive (TP) : The test result is positive in the presence of the clinical abnormality.

True Negative (TN): The test result is negative in the absence of the clinical abnormality.

False Positive (FP): The test result is positive in the absence of the clinical abnormality.

False Negative (FN): The test result is negative in the presence of the clinical abnormality

$$FP= \text{false positive pixels number /tumor size} \tag{10}$$

$$FN=\text{false negative pixel number / tumor size} \tag{11}$$

$$\text{Correct rate=FP+FN} \tag{12}$$

The average correct rate by the presented method is 97.82% with FP of 1.0% and FN of 2.50%. All the features produce classification accuracy of 98.87% using LDA. The extracted four PCA components are classified using LDA and SVM classification and the accuracy achieved is 96%. . The overall accuracy percentage details are shown in fig 1.
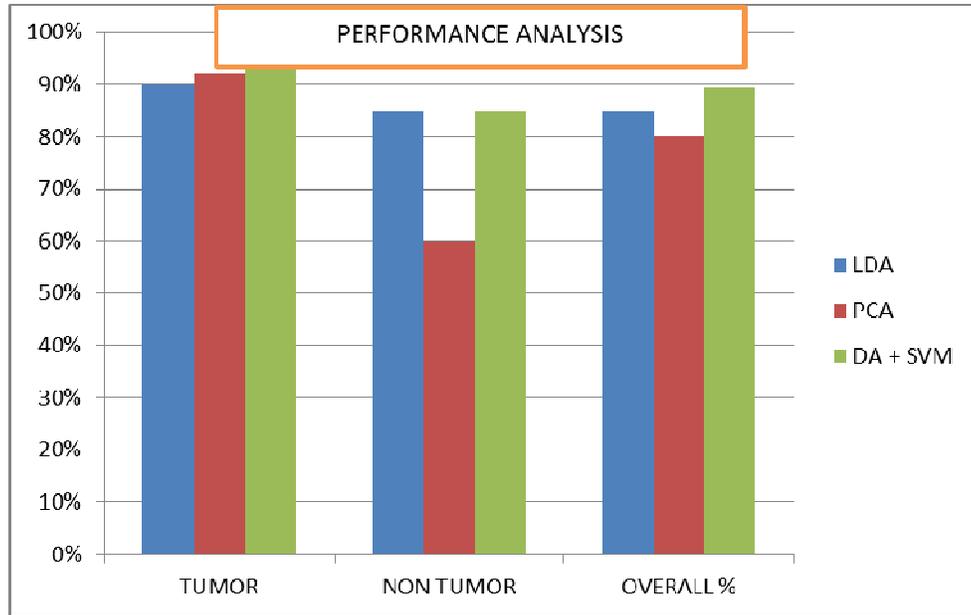

Fig: 1 overall accuracy performance of LDA, PCA and the proposed method

The comparative analysis of the proposed method and the existing algorithms are shown in table4.

Table 4.Comparative analysis

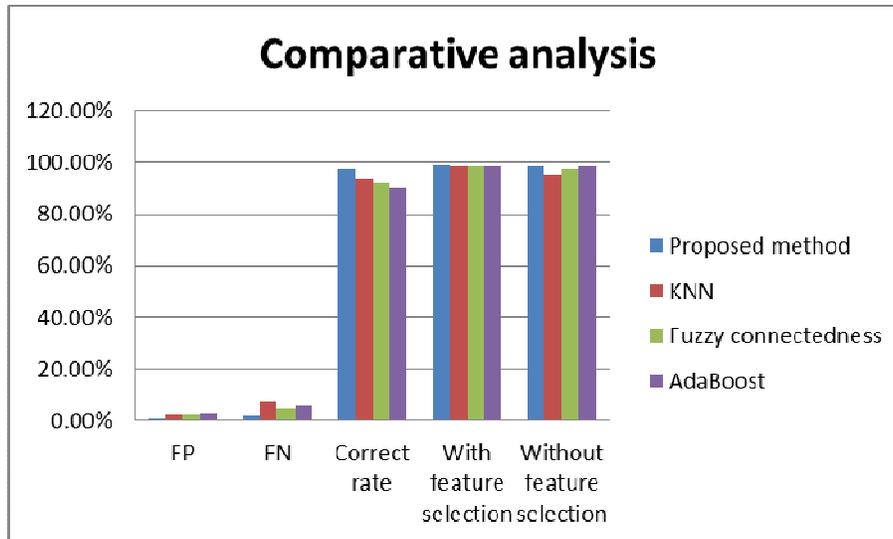| Classification accuracy	FP	FN	Correct rate	With feature selection	Without feature selection
Proposed method	1.00%	2.50%	97.82%	98.87%	98.77%
KNN	2.75%	7.51%	93.50%	98.48%	95.47%
Fuzzy connectedness	2.95%	5.02%	92.04%	98.35%	97.47%
AdaBoost	3.15%	6.07%	90.05%	98.74%	98.55%

Fig: 2 Comparative analysis of existing algorithms and the proposed method

3. CONCLUSIONS

The method mainly includes the steps, Image collection, Normalization, Intensity, shape and Texture feature extraction, feature selection and classification. The proposed method performs better than the existing algorithms. The shape, Intensity and Texture features are extracted and used for classification. Vital features are selected using LDA. The results are compared with dimension reduction techniques like PCA. The number of features selected or features extracted by PCA and The classification accuracy by SVM is 98.87%. It is expected that the information of new imaging technique fMRI and the Image MOMENTS when added into the scheme will give more accurate results which will be taken for future work.

ACKNOWLEDGEMENTS

REFERENCES

[1] K.M. Iftekharuddin, On techniques in fractal analysis and their applications in brian MRI, in: T.L. Cornelius (Ed.), Medical imaging systems: technology and applications, Analysis and Computational Methods, vol. 1, World Scientific Publications, 2005, ISBN 981-256-993-6.

[2] L.P. Clarke, R.P. Velthuizen, M.A. Camacho, J.J. Heine, M. Vaidyanathan, L.O. Hall, R.W. Thatcher, M.L. Silbiger, MRI segmentation: methods and applications, Magn. Reson. Imaging 13 (3) (1995) 343-368.

[3]   J.C. Bezdek, L.O. Hall, L.P. Clarke, Review of MR image segmentation techniques using pattern recognition, Med. Phys. 20 (4)(1993)1033-1048.

[4]   H.S. Zadech, H.S. Windham, Optimal linear transformation for MRI feature extraction, IEEE Trans. Med. Imaging 15 (1996) 749-767.

[5]   H.S. Zadech, J.P. Windham, A comparative analysis of several transformations for enhancement and segmentation of magnetic resonance image scene sequences, IEEE Trans. Med. Imaging 11 (N3) (1992) 302-318.

[6]   D. Wang, D.M. Doddrell, A segmentation-based partial-volume-compensated method for an accurate measurement of lateral ventricular volumes on T1-weighted magnetic resonance images, Magn. Reson. Imaging 19 (2001) 267-272.

[7]   X. Zeng, L.H. Staib, R.T. Schultz, J.S. Duncan, Segmentation and measurement of the cortex from 3-D MR images using coupled-surfaces propagation, IEEE Trans. Med. Imaging 18 (10) (1999) 927-937.

[8]   M. Just, M. Thelen, Tissue characterization with T1, T2 and proton density values: results in 160 patients with brain tumors,Radiology 169 (1988) 779-785.

[9]   M.W. Vannier, R.L. Butterfield, D. Jordan, W.A. Murphy, R.G. Levitt, M. Gado, Multispectral analysis of magnetic resonanceimages, Radiology 154 (1) (1985) 221-224.

[10] M.W. Vannier, T.K. Pilgram, C.M. Speidel, L.R. Neumann, D.L. Rickman, L.D. Schertz, Validation of magnetic resonance imaging(MRI) multispectral tissue classification, Comput. Med. Imaging Graph. 15 (4) (1991) 217-223.

Dr.S Palani graduated from P.S.G. College of Technology, Coimbatore, did his post graduate studies at IIT Kharagpur   and the doctoral degree from Regional Engineering College Trichy. He is a faculty member of the department of Electrical and Electronics Engineeri ng, Sudharsan Engineering College, Pudukkottai, India. He has more than 40 years of teaching experience. His field of Interest includes Control Systems, Electrical Engineering and Digital signal Processing.

V.P.GladisPushpaRathi graduated from Cape Institute of Technology, Tirunelveli, did her post graduate studies at M.S.University Tirunelveli and is now doing her Ph.D in AnnaUnivers ity,Trichy. She is a faculty member of the department of Computer Science and Engineering, Sudharsan Engineering College. She has 5 years of teaching experience. Her field of Interest includes Digital Image Processing, Soft computing, and Datamining.