International Journal of Advanced Technology in Engineering and Science
Vol. No.4, Issue No. 03, March 2016
www.ijates.com
ijates
ISSN 2348 - 7550

MRI BRAIN IMAGE CLASSIFICATION USING POLYNOMIAL KERNEL PRINCIPAL COMPONENT ANALYSIS WITH NEURAL NETWORK

Mahipal Singh Choudhry[1], Shikha Misra[2], Rajiv Kapoor[3]

[1,3]ECE Department, Delhi Technological University, (India)

[2]ECE Department, Sachdeva Institute of Technology, (India)

ABSTRACT

Magnetic Resonance (MR) Imaging has come up as widely accepted and revolutionary innovation in field of medical science and brain imaging especially. A new method is proposed here for MRI brain image classification using Polynomial Kernel Principle Component Analysis (KPCA) with Neural Network. In this paper, we are having various stages namely pre-processing, feature extraction, feature reduction and classification of MRI brain images. Here for improving the MRI image quality, imadjust function is used as pre-processing stage of MRI image. In second stage, features are reduced by Polynomial Kernel Principle Component Analysis. In last stage, MRI images are classified as normal or abnormal image by Artificial Neural Network. Different feature reduction methods like PCA, LDA, SVD, Gaussian KPCA and Polynomial KPCA with P (power of kernel) = 2, 3, 4 and 5 are used. The results show that classification rate of 99.8 % is achieved for p = 4 of KPCA.

Keywords: *ANN, DWT, FCM, k-NN, KPCA, LDA, MRI, SOM, SVD, SVM.*

I. INTRODUCTION

MRI is a powerful medical imaging technique to provide detail and reliable information about human brain. MRI is preferred over other imaging methods for human brain imaging because it does not involve any ionizing radiation and it is used in non-invasively form [6].In proposed method, firstly preprocessing of the MR images are employed using imadjust MATLAB function and then feature extraction is done by discrete Wavelet Transform (DWT). In Wavelet Transform, different frequencies are examined with different resolutions [9] and DWT coefficients are used as feature vectors of image. The wavelet coefficients are extracted from MR images by DWT in form of localized frequency information about the functions of image that is used for classification [4].

There are different types of Wavelet Transform functions with their strength and limitations [1]. Haar wavelet functions are used in proposed method due to suitability for brain MR images. To increase the processing speed, feature reduction method is used by transforming high dimensional input feature space into a lower-dimensional feature space. There are various techniques like Principal Component Analysis [10], Kernel PCA [18], Linear Discriminant Analysis [11] and Singular value Decomposition [14] are used for feature reduction**.** There are different types of Kernel PCA. In this method, Kernel PCA with variable power is specially used as feature reduction method because it does not involve any nonlinear optimization [12] [6].

MR image classification techniques have been broadly classified in two classes of supervised techniques and unsupervised techniques. Artificial Neural Networks (ANN) [2] [13], Support Vector Machine (SVM) [6], k-Nearest Neighbor (k-NN) [13] and feed forward back propagation [17] are important supervised techniques and Self-Organization Map (SOM) [6] and Fuzzy C-Means (FCM) [8] are typical unsupervised techniques. Out of these classifiers, ANN, k-NN, SVM and FCM are commonly used to obtain higher classification rate for detection of abnormality of brain. Neural networks have many advantages like requirement of less number of training sets, having large numbers of training algorithms and capability of detecting nonlinear relation between independent and independent parameters [19] etc. Due to these advantages, Neural Network is used in proposed method for MRI brain image classification.

This paper is structured in following parts. The part 2 presents the methodology of paper. The part 3 presents KPCA as feature reduction method. For classification of MRI brain image, artificial neural network is used in part 4.The experimental results are shown in part 5 and in last part conclusion is described.

II. METHODOLOGY

The suggested methodology has different stages like MRI data acquisition, pre-processing of MRI images, feature extraction from images, dimension reduction and MRI images classification as normal or abnormal. The complete methodology is shown in Fig.1.

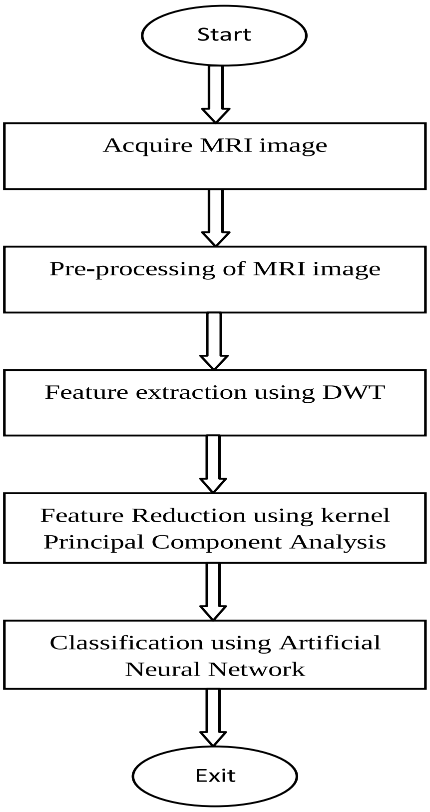

Fig. 1 Methodology

2.1 Data Acquisition

In this paper, we implemented diagnosis method on real MRI data of human brain with high resolution axial images of size 256×256 pixel. Total 120 MR brain images from input dataset are used for training and testing purpose. Out of 120 MR images, we are taken 50 as training images and 70 as testing images. For training five sets of ten images are used. One set of training data has normal MR images. The four sets have abnormal or brain tumor infected MR images in which tumor location are different.

2.2 Pre-Processing of MRI Images

Some MRI images are darker than others due to data acquisition scanner problems. For improving the quality of image, image enhancement techniques are employed. Intensity adjustment is one of the important methods for image enhancement technique, in which the intensity value of original image are mapped into a new range of intensity value. In this paper, gamma correction [7], which is implemented in MATLAB using imadjust function, is used to improve MRI images quality.

2.3 Feature Extraction using DWT

For feature extraction from MR brain images Short Time Fourier Transform (STFT) and Wavelet Transform can be used. In STFT, both time and frequency are represented in limited precision. Precision of STFT is determined by window size. Once specific window size is chosen for time window, it will be same for all frequency. Wavelets are short time localized waves with zero average value and possibility of time shifting, flexible etc. DWT is widely used for feature extraction because in case of Wavelet Transform, different frequencies are examined with different resolutions [4]. In proposed method, DWT are used for features extraction from MR images. The features are represented in form of DWT coefficients [18]. MR brain image is applied through series of half band LPF and HPF. The output of HPF is detail coefficients and output of LPF is approximation coefficients [1].The image can be represented with different resolution levels by wavelet transform. Two level decomposition using DWT is explained in Fig.2.

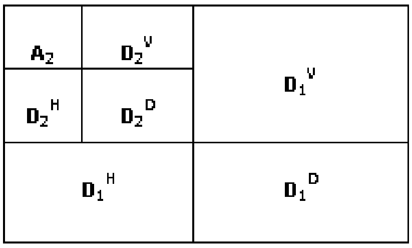

Fig.2 Two Levels Decomposition using DWT

Where, D_2^V, D_2^H, D_2^D are detail of image of input image I_j when resolution is j.

III. KERNEL PRINCIPAL COMPONENT ANALYSIS

To increase the processing speed, feature reduction method is used. The features reduction techniques are applied on features which are extracted by DWT in the form of wavelet coefficients. There are different methods for feature reductions like PCA, KPCA, LDA and SVD. Complex structure of data, which is used for classification, may not be preserved in LDA and SVD. If discriminatory information is contained in the variance of data instead of mean, then LDA will not work properly. In this case PCA is preferred. If the distributions are non-Gaussian, PCA may not find the best direction for discriminating between two classes. It is difficult to visualize data in dimensions greater than three in PCA. In this case KPCA is preferred over PCA. KPCA applies similar theory as the PCA and it projects data points on low dimensional subspace that captures the highest possible amount of variance in the data. In PCA, data points are separated linearly in original space (referred to as Rd) but in KPCA, data points are separated in a high dimensional space which is known as feature space (referred to as F), by using a mapping function "Ψ" as shown in Fig. 3.

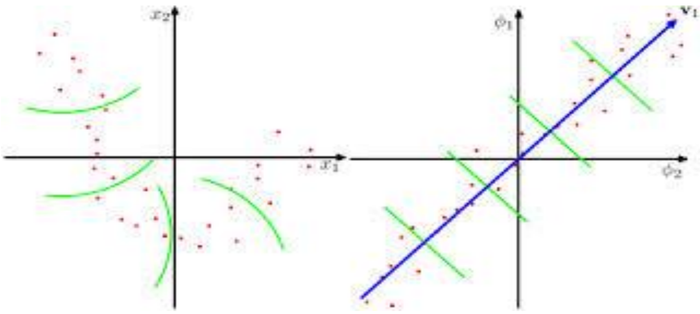

Fig. 3 (a) Nonlinearly distributed data before applying KPCA (b) Data into linearly distributed feature space after applying KPCA

Due to limitation of above methods, KPCA is used in this research paper. Advantages of KPCA are:

1. It does not require nonlinear optimization, knowledge of the network architecture or the number of dimensions.
2. A new point can be quickly projected onto a pre-computed basis.
3. It gets eigenvector with higher variance (principal component) than PCA.

In PCA, covariance matrix is constructed for feature reduction. It is difficult to construct covariance matrix in feature space due to nonlinear projection. Due to this reason kernel trick is used to develop the kernel matrix Ke (x, x') = Ψ(x).Ψ(x') without explicitly doing the mapping [16] [18].The data in the feature space is projected onto a low-dimensional subspace, spanned by the eigenvectors that capture most of the variance. One important fact is that without knowing the mapping "φ" or the feature space "F", KPCA is applied to data input. Instead computations are performed on the inner product of pairs of points which are stored in a kernel matrix. The procedure of working with the data in feature space without knowing the mapping "Ψ" is known as "the kernel trick" and is a central part of the kernel PCA method. There are normally two types of kernels used in KPCA: the polynomial kernel and Gaussian kernel.

3.1 Polynomial Kernel

$$Ke\,(x_j, x_k) = (x_j^T x_k)^p \qquad (1)$$

Or

$$Ke\,(x_j, x_k) = (x_j^T x_k + c)^p \qquad (2)$$

Where, p denotes polynomial's order and c (constant) >0

3.2 Gaussian Kernel

$$Ke\,(x_j, x_k) = \exp\left(-\left|x_j - x_k\right|^2 / 2\sigma^2\right) \qquad (3)$$

Or

$$Ke\,(x_j, x_k) = \exp\,(-\sigma |x_j - x_k|^2) \qquad - (4)$$

To kernalize the procedure, $x_i \rightarrow \Psi\,(x_i)$ and

$$Ke_{jk} = \Psi\,(x_j)\,\Psi\,(x_j) \qquad - (5)$$

If data are non-centered set, then inner product computation is applied for centering data.

$$(\widetilde{Ke})_{jk} = (\Psi(x_j) - \bar{\Psi})(\Psi(x_k) - \bar{\Psi}) \qquad - (6)$$

Where, $\widetilde{Ke}$ denotes the centered kernel matrix

$\bar{\Psi}$ represents the center point in the feature space which is written by

$$\bar{\Psi} = \frac{1}{n}\sum_{i=1}^{n} \Psi\,(x_i) \qquad - (7)$$

The eigenvectors of $\tilde{k}$ can be obtained by solving the given equation

$$\widetilde{Ke}\,V = \lambda V \qquad - (8)$$

Where normalized eigenvectors are stored in the column of ‘V’ and “λ” is a matrix in which the non-diagonal values are zero and diagonal consist of the corresponding Eigen values. The eigenvectors in V are arranged in descending order according to the size of their corresponding Eigen value. Also the Eigen values in λ are sorted in descending order [12].

Using a similar calculation, this can be expressed easily in terms of Ke_{ij}.

$$K_{e1} = K_e - 1_N K_e - K_e 1_N + 1_N K_e 1_N \qquad - (9)$$

Where, 1_N is NXN matrix in which all elements equal to 1/N. In this paper, polynomial kernels, with varying order (p = 2, 3, 4) are used to achieve good classification efficiency.

IV. ARTIFICIAL NEURAL NETWORK (ANN)

ANN is highly interconnected numbers of neurons which are arranged processing into different layers [2]. ANN consists of many layers and each layer contains one or more neuron. A single neuron has many inputs; these inputs are multiply with weights and weighted inputs are summed up by summer and then activation function is applied. The biological neuron can be modeled as an artificial neuron [19]. This is explained in Fig. 4 with following steps.

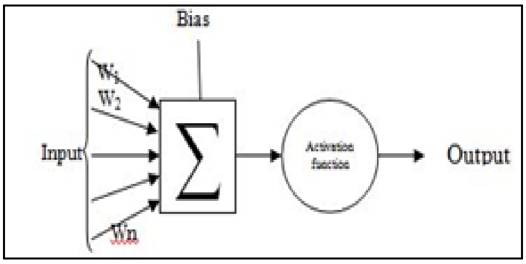

Fig. 4 Single Neuron Model

1. The synapses works as weight element in neuron. When a signal is applied at the input of synapse, the input signal is multiplied by the synaptic weight w_k. The weight w_k is positive or negative depending upon type of synapse.
2. A summer output is the weighted sum of input signal.

3. The output of summer is applied to an activation function which is used to limit the amplitude of the neuron's output. The commonly used normalized amplitude range are [0, 1] or [-1, 1]. A bias is used for altering the input of the activation function.

Two fundamental kinds of networks are occurred in neural networks i.e. using feedback and without using feedback. If the output is calculated for each input value, then this is called the network without feedback. In networks with feedback, the output values are applied back to input for tracking the input value. If the information flows in forward direction, then it is called feed forward network [17]. It is shown in Fig.5.

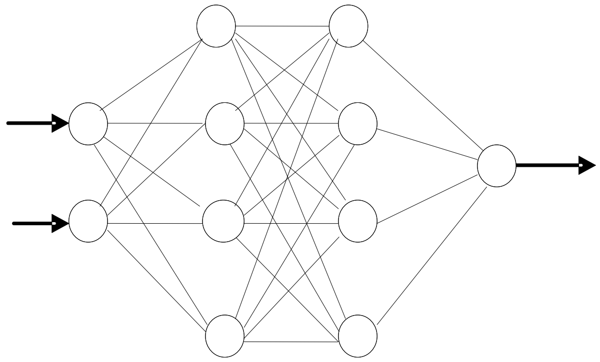

Fig. 5 Feed Forward Neural Networks

There are different types of feed forward network like multilayer perceptions (MLP) and radial basis function (RBF) etc. In MLPs, classes are separated via hyper-planes and in RBFs via hyper-spheres. MLPs have input layer, one or more hidden layers and an output layer. In radial basis function network, a single hidden layer is used which has radial basis activation function for hidden neuron. MLPs have one or more hidden layers but RBFs have only one hidden layer. RBFs require more hidden neurons which lead to curse of dimensionality. The perception learning rule is used for single layer neuron but for more than one layer neuron networks, this training algorithm is not suitable because the output of hidden layer is not available for calculating the output and updating the weight so for multilayer perceptions, perception learning algorithm is not used. For this, error back propagation learning is employed, in which training is done in a supervised manner.

In this paper, multilayer perceptions network with error back propagation learning is used. Back propagation learning rule is based on the error-correction learning rule [13]. In this learning rule, firstly input signal (vector) are applied into input layer and this signal is flowed in forward direction layer by layer through hidden layer to output layer. Then output vector is generated for applied input vector which is known as actual response of the neural network. This is called forward pass. In forward pass, weights of neuron are fixed. Now actual output is compared with desire output or target output. If the difference between desire output and actual output is not equal to zero, then error signal will be generated. This error signal is propagated in backward direction so this is called as "error back propagation". Now according to perceptions error correction rule, weights are updated for minimizing the error signal. In backward pass, weights of neurons are modified so that actual response tends to desire output [5] [13]. Forward NN have three layers like input, hidden and output layer. In this paper, seven different intensity based features after reduced by KPCA are used so seven input elements are used in first layer.

V. RESULTS

For testing seven sets of ten images are used. Three set of training data contains normal MR images & rest four sets contain abnormal or brain tumor infected MR images. The result in terms of classification accuracy for ANN with different feature reduction method is given in Table 1 for MR brain images.

Table 1: Comparison of polynomial KPCA with other feature reduction techniques

Classifier(ANN) with different feature reduction methods	Classification Accuracy (%)
PCA	86.7 %
LDA	93.3%
SVD	80.2%
KPCA (Gaussian)	86.7 %
KPCA(polynomial p=2)	93.7%
KPCA(polynomial p=3)	96.7%
KPCA(polynomial p=4)	99.8 %
KPCA(polynomial p=5)	86.7 %

From above table, it is clear that KPCA with ANN for power p = 4 gives maximum classification accuracy (99.8 %).

VI. CONCLUSION

This work proposed the efficient method for detection of normal or abnormal brain images. Various classification rates are obtained using different power of applied kernel. Confusion matrix obtained by classification gives, 93.7% accuracy with p=2, 96.7% accuracy with p=3, 99.8% accuracy with p=4, 86.7% accuracy with p=5. From above result, it can be concluded that KPCA with kernel's power = 4 gives best classification result. Here, in this paper, discrete wavelet transform (DWT) is used as feature extraction technique from brain MRI images. There are various others transforms and techniques available for extracting image features so in future work, different feature extraction methods can be used to achieve higher classification accuracy for large MRI data.

VII. ACKNOWLEDGEMENTS

Special thank to Institute of Nuclear Medicine & Allied Science (INMAS), Defence Research & Development Organization (DRDO), Govt of India, for providing real data set of brain MR images.

REFERENCES

[1] Stephane Mallat, "A WAVELET TOUR OF SIGNAL PROCESSING" Second Edition, Ecole Polytechnique, Paris Courant Institute, New York University, 1999, Elsevier (USA)

[2] S.Haykin, Neural Networks, A Comprehensive Foundation, Prentice Hall, 1999.

[3] L.M. Fletcher-Heath, L.O. Hall, D.B. Goldgof, and F.R. Murtagh, "Automatic segmentation of non-enhancing brain tumors in magnetic resonance images", Artif. Intell. Med 21, 2001, 43-63.

[4] K. Karibasappa, S. Patnaik, "Face recognition by ANN using wavelet transform coeff," IE (India) Journal of Computer Eng., vol . 85, pp. 17–23, 2004.

[5] M. O'Farrell, E. Lewis, C. Flanagan, and N. Jackman,"Comparison of k-NN and neural network methods in the classification of spectral data from an optical fibre-based, sensor system used for quality control in the food industry", Sens. Actuators B: Chemical 111–112C (2005) 354–362.

[6] S.Chaplot, L.M.Patnaik and N.R.Jagannathan "Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network", Biomed Signal Process. Control, 1 (2006) 86–92.

[7] Gerard Blanchet, and Maurice Charbit, "Digital Signal and Image Processing using MATLAB", TK5102.9.B545 2006 ,621.382'2--dc22

[8] M.Maitra, A.Chatterjee "Hybrid multiresolution Slantlet transform and fuzzy c-means clustering approach for normal-pathological brain MR image segregation", Med. Engg. Phys. (2007), doi: 10.1016/j medengphy .2007 .06.009.

[9] S. Kara,and F. Dirgenali, "A system to diagnose atherosclerosis via wavelet transforms, principal component analysis and artificial neural networks", Expert Syst.Appl, 32 (2007).

[10] Noramalina Abdullah, Lee Wee Chuen,and Umi Kalthum Ngah Khairul Azman, "Improvement of MRI Brain Classification using Principal Component Analysis", 2011 IEEE International Conference on Control System, Computing and Engg.

[11] Quanquan Gu, Zhenhui Li, and Jiawei Han,"Linear Discriminant Dimensionality Reduction", Department of Computer Science, University of Illinois at Urbana- Champaign Urbana, IL 61801, US.

[12] Daniel Olsson, Master's Thesis Report Applications and Implementation of Kernel Principal Component Analysis to Specific Data Sets, January 28, 2011

[13] N.Hema Rajini, and R.Bhavani ,"Classification of MRI Brain Images using k- Nearest Neighbor and Artificial Neural Network", IEEE-International Conference on Recent Trends in Information Technology, ICRTIT 2011, MIT, Anna University,

[14] Davi Marco Lyra-Leite, Joao Paulo Carvalho Lustosa da Costa, and Joao LuizAzevedo de Carvalho," Improved MRI reconstruction and denoising using SVD-based low-rank aproximation", Digital Signal Processing Group Department of Electrical Engineering University of Brasilia Brasilia-DF, Brazil.

[15] Harvard Medical School, Web, data available at http://med.harvard.edu/AANLIB/.

[15] Quan Wang," Kernel Principal Component Analysis and its Applications in Face Recognition and Active Shape Models"arXiv: 1207.3538v1 [cs.CV] 15 Jul 2012

[17] S.N.Deepa,and B.Aruna Devi, Artificial Neural Networks design for Classification of Brain Tumour" 2012 International Conference on Computer Communication and Informatics (ICCCI -2012), Jan. 10 – 12, 2012, Coimbatore, India

[18] Devvi Sarwinda,and Aniati M Arymurthy,"Feature Selection Using Kernel PCA for Alzhemier's Disease Detection with 3D MR Images of Brain", ICACSIS 2013 ISBN: 978-979-1421-19-5

[19] Walaa Hussein Ibrahim, Ahmed Abdel Rhman Ahmed Osman and Yusra Ibrahim Mohamed, "MRI Brain Image Classification Using Neural Networks", 2013 International Conference On Computing, Electrical And Electronics Engineering(Icceee),978-1-4673-6232,2013.

[20] Jagdish H Pujar Pallavi S. Gurjal, Shambhavi D.S, Kiran S. Kunnur."Medical Image Segmentation based on Vigorous Smoothing and Edge Detection Ideology", International Journal of Electrical and Computer Engineering 5:2 2010