

## Brain Tumour Detection Using Neural Network Classifier and K-Means Clustering Algorithm for Classification and Segmentation

*M. Latha and R. Surya*

Department of Computer Science and Engineering, Vels University, Chennai, India

---

**Abstract:** The aim of this work is to present an automated method that assists diagnosis of normal and abnormal MR images. The diagnosis method consists of four stages, pre-processing of MR images, skull Stripping, feature extraction, feature reduction and classification. After histogram equalization of image, the features are extracted based on Dual-Tree Complex wavelet transformation (DTCWT). Then the features are reduced using principal component analysis (PCA). In the last stage two classification methods, k-means clustering and Probabilistic neural network (PNN) are employed. Our work is the modification and extension of the previous studies on the diagnosis of brain diseases, while we obtain better classification rate with the less number of features and we also use larger and rather different database.

**Key words:** Magnetic resonance imaging • Pattern recognition • Classification • Dual-Tree Complex wavelet feature extraction • Probabilistic neural networks

---

### INTRODUCTION

In the recent century, increasingly use of sophisticated laboratory tests has made diagnosis a sensitive and accurate issue. The use of computer technology in medical decision support is now widespread and pervasive across a wide range of medical area, such as cancer research, gastroenterology, heart diseases, etc [1]. In studying human brain, magnetic resonance imaging (MRI) plays an important role in progressive researches. The rich information that MR Images provide about the soft tissue anatomy has dramatically improved the quality of brain pathology diagnosis and treatment. However, the amount of data is far too much for manual interpretation and hence there is a great need for automated image analysis tools [2]. Pattern recognition techniques are being increasingly employed in magnetic resonance imaging (MRI) data analysis [3]. The automatic classification of brain magnetic resonance images is vital process for separating healthy subjects and subjects with various brain diseases like, cerebrovascular, Alzheimer, brain tumour, inflammatory, etc.

In this study, our goal is to achieve higher classification rate in order to diagnosis normal images from those with brain abnormality. First, the MR images are pre-processed to set the mean intensity of pixels to

same level. In this way the dark images become light like other images. Second, features are extracted using wavelet transformation. Wavelets seem to be a suitable tool for this task, because they allow analysis of images at various levels of resolution [4, 5]. Third, principal component analysis (PCA) is used for reducing the number of features and also increasing discrimination between classes. Principal component analysis is appealing since it effectively reduces the dimensionality of the data and therefore reduces the computational cost of analyzing new data [6]. Finally, three pattern recognition methods *k*-NN, parzen and ANN are used for classification. The results indicate fully classification of data. Our work is the extension and modification of the method introduced in. But, our case is different, the database contains more and different images, we use preprocessing step and additional classifier; the number of features obtained by PCA for maximum classification rate is less and we obtain better classification rate. However, DWT, PCA and classifiers are commonly used steps in pattern recognition problems. The organization of this paper is as follows. Section 2 shortly describes the image dataset, steps of brain diagnosis method including pre-processing, feature extraction and reduction. Section 3 contains classification stage. Results and comparison with previous works are presented in section 4. Section 5 summarizes the approach and presents future works.

**MATERIALS AND METHODS**

**Imaging Data:** The diagnosis methods have been implemented on a real human brain MRI dataset. The protocol includes high resolution axial, T1, T2-weighted  $256 \times 256$  pixel images. The dataset contains 125 normal subjects and 41 abnormal MR images including cerebrovascular, Alzheimer, brain tumor, inflammatory, infectious and degenerative diseases. These dataset were collected from Harvard Medical School website [6-11] and Laboratory of Neuro Imaging (LONI) website. Fig. 1 shows some examples from normal and abnormal subjects.

**Pre-Processing:** Some images of dataset were dark rather than others. This is because of data acquisition scanner problems. The Scans are corrected for intensity non uniformity using histogram equalization. Fig. 2 depicts an example of dark MRI after and before pre-processing.

**Skull Stripping:** Removal of the skull and other non-brain tissue like dura and eyes from anatomical images, which could otherwise complicate e.g. coregistration and normalization steps. Skull stripping can be part of the tissue segmentation (e.g. in SPM) but is mostly done by specialized algorithms that delineate the brain boundary. See [1] for a comparison of some brain extraction algorithms (BSE, BET, SPM and McStrip), which suggests that all algorithms perform well in general but results highly depend on the particular dataset.

Skull stripping, also known as whole brain segmentation, is an important step to remove the non-cerebral tissue such as skin, skull, fat, muscle and connective tissues. An algorithm is used to remove the non-cerebral tissue such as skin, skull, fat, muscle and connective tissue.

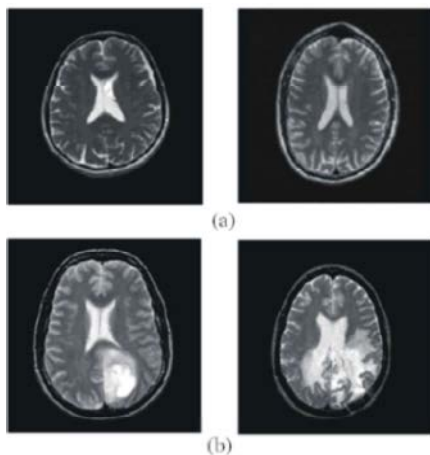


Fig. 1: Examples from (a) normal (b) abnormal subjects

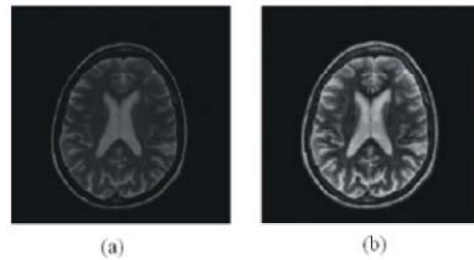


Fig. 2: (a) Original image (b) Image after pre-processing

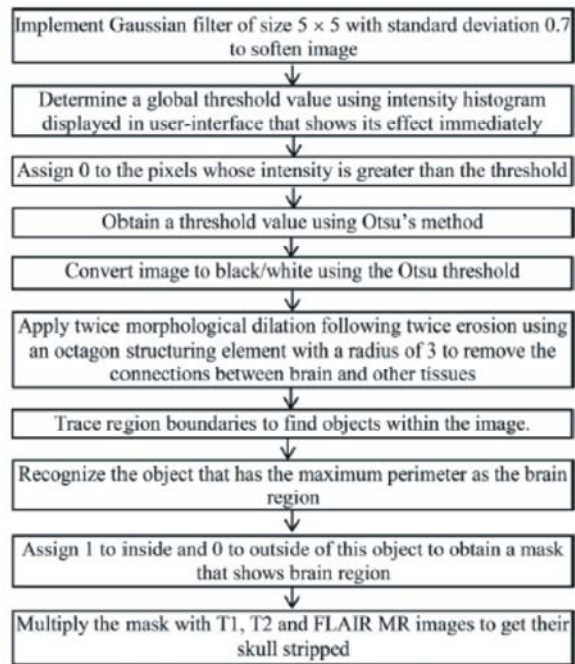


Fig. 3: Stages of the skull stripping algorithm

The algorithm that has been used here for skull stripping totally has ten steps which include Gaussian Filter and Otsus Method and Morphological dialation.

**Feature Extraction:** We use wavelet coefficients for generating the initial features. Wavelet transform is traditionally used for feature extraction. The provision of localized frequency information about a function of a signal, is the main advantage of wavelets and is particularly beneficial for classification. Earlier, wavelets have been used as a feature extraction method for discrimination [7].

In two-dimensional wavelet transform a scaling function  $\phi(X, Y)$  and three wavelets  $\varphi_1(x, y)$  measures variations along columns,  $\varphi_2(X, Y)$  responds to variation along rows,  $\varphi_3(x, y)$  corresponds to variations along diagonls, are required. The Dual-Tree Complex wavelet transform image  $f(x, y)$  of size  $M*N$  is then,

$$w(j, m, n) = 1/\sqrt{MN} \sum_{x=0}^{M-1} * \sum_{Y=0}^{N-1} (X, Y) \varphi(j, m, n)^{(x, y)}$$

$$w(k, m, n) = 1/\sqrt{MN} \sum_{x=0}^{M-1} * \sum_{Y=0}^{N-1} (X, Y) \varphi(k, m, n)^{(x, y)}$$

where I identifies the directional wavelets (i={H, V, D}) and k is the arbitrary starting scale. The w(j, m, n) coefficients define an approximation of f(x, y) and scale j. The w(k, m, n) coefficients add horizontal, vertical and diagonal details for scale j>=k [8]. Fig. 3 shows the process in block diagram form. There are several different kinds of wavelets which have gained popularity throughout the development of wavelet analysis. One important discrete wavelet is the Haar wavelet. Basically, it is one period of a square wave. Because of its simplicity, it is often the wavelet to be chosen [9].

We use three scale Haar (H2) basis functions for DTCWT feature extraction. Thus, for the image of size 256\*256, we use the approximation coefficient of the third level as the features. Thus, the number of features used in this stage would be 1024.

$$H2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}$$

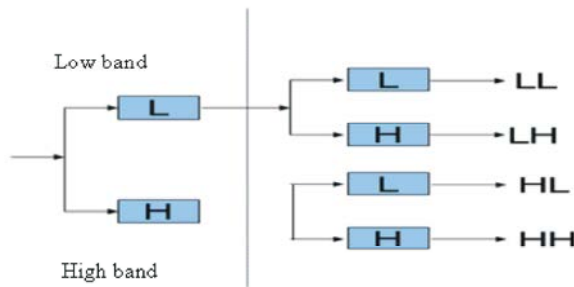


Fig. 4: Dual-Tree Complex Wavelet Transform



Fig. 5: Dual-Tree Complex wavelet transform of one example MR image

**Feature Reduction:** Measurement cost and classification accuracy are two predominant reasons for minimizing the dimensionality of the pattern representation (i.e., the number of features). Classifiers which are built on the selected representation can use less memory and be faster by utilizing the limited feature set. Linear transforms, due to their simplicity, have been widely used for feature extraction. These transforms create a smaller set of features from linear combination of the initial features.

One of the best known linear feature extractor is the principal component analysis (PCA) or Karhunen-Loeve expansion [10]. The basic approach in principal components is conceptually quite simple. First, the d-dimensional mean vector  $\mu$  and d\*d covariance matrix  $\Sigma$  are computed for the full data set. Next, the eigenvectors and eigenvalues are computed and sorted according to decreasing eigenvalue. Call these eigenvectors  $e_1$  with  $\gamma_1$ ,  $e_2$  with  $\gamma_2$  and so on. Next, the largest k such eigenvectors are chosen. In practice, this is done by looking at a spectrum of eigenvectors. Form a k\*k matrix A whose columns consist of the K eigenvectors. Preprocess data according to:

$$X = A(x-\mu)$$

Since PCA uses the most expressive features (eigenvectors with the largest eigenvalues), it effectively approximates the data by a linear subspace using the mean squared error criterion [10]. In our work the number of features after PCA that achieves the maximum accuracy for k-means is 6 and for PNN is 3.

**Classification**

**K-Means Clustering:** K-means clustering algorithm is one of the simplest pattern recognition classification techniques. The algorithm for the clustering rule is summarized as follows. Given an unknown feature vector and a distance measure, then:

- Out of the training vectors, identify the k nearest Neighbors with same features, irrespective of class label.
- Combine those features having same values- Clusters

**Probabilistic-Neural Networks:** Probabilistic (PNN) and General Regression Neural Networks (GRNN) have similar architectures, but there is a fundamental difference: Probabilistic networks perform classification where the target variable is categorical, whereas general regression neural networks perform regression where the target

variable is continuous. If you select a PNN/GRNN network, DTREG will automatically select the correct type of network based on the type of target variable.

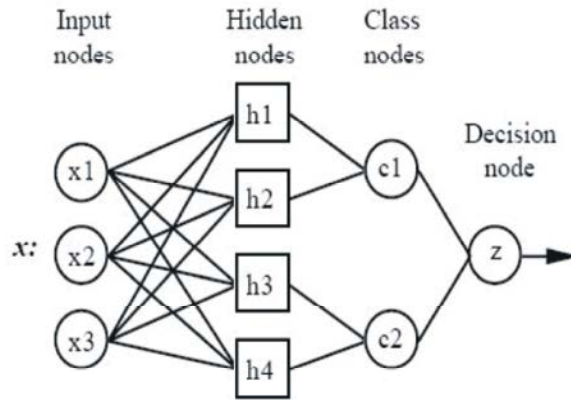


Fig. 6: Architecture of a PNN

**All PNN Networks Have Four Layers:**

- **Input layer-** There is one neuron in the input layer for each predictor variable. In the case of categorical variables, N-1 neurons are used where N is the number of categories. The input neurons (or processing before the input layer) standardizes the range of the values by subtracting the median and dividing by the interquartile range. The input neurons then feed the values to each of the neurons in the hidden layer.
- **Hidden layer-** This layer has one neuron for each case in the training data set. The neuron stores the values of the predictor variables for the case along with the target value. When presented with the x vector of input values from the input layer, a hidden neuron computes the Euclidean distance of the test case from the neuron's center point and then applies

the RBF kernel function using the sigma value(s). The resulting value is passed to the neurons in the pattern layer.

- **Pattern layer / Summation layer-** The next layer in the network is different for PNN networks and for GRNN networks. For PNN networks there is one pattern neuron for each category of the target variable. The actual target category of each training case is stored with each hidden neuron; the weighted value coming out of a hidden neuron is fed only to the pattern neuron that corresponds to the hidden neuron's category. The pattern neurons add the values for the class they represent (hence, it is a weighted vote for that category).

For GRNN networks, there are only two neurons in the pattern layer. One neuron is the denominator summation unit the other is the numerator summation unit. The denominator summation unit adds up the weight values coming from each of the hidden neurons. The numerator summation unit adds up the weight values multiplied by the actual target value for each hidden neuron.

- **Decision layer-** The decision layer is different for PNN and GRNN networks. For PNN networks, the decision layer compares the weighted votes for each target category accumulated in the pattern layer and uses the largest vote to predict the target category.

For GRNN networks, the decision layer divides the value accumulated in the numerator summation unit by the value in the denominator summation unit and uses the result as the predicted target value. The following diagram is actual diagram or propose network used in our project.

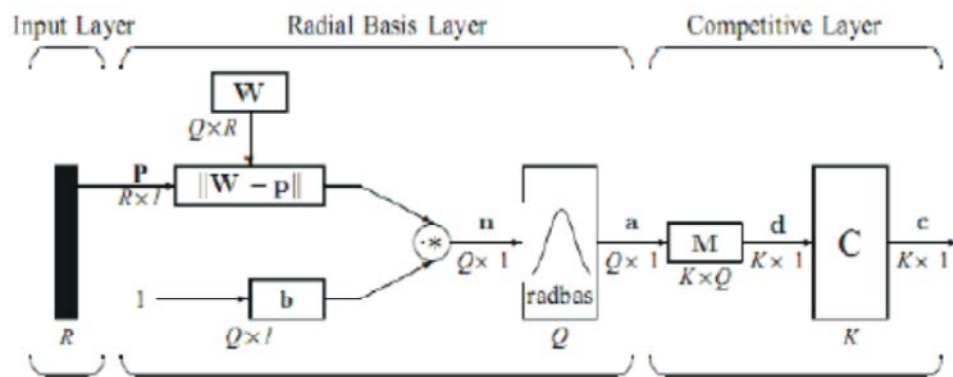


Fig. 7: Proposed PNN Structure

**RESULTS AND DISCUSSION**

We apply a supervised method for the diagnosis of normal and abnormal MRI brain images. As mentioned the method employs four stages: preprocessing, skull stripping, feature extraction, feature reduction and classification. After histogram equalization of images, the first three levels coefficients of decomposition of MRI with Haar are computed to extract the features. Then, the 3<sup>rd</sup> approximation component and all detailed components are used as the wavelet coefficients. These coefficients are used for feature extraction. The dimension of feature vector is 1024 and we use principal component analysis for reducing the number of features which leads to increase the accuracy rates of classification. Two classifiers based on Pattern recognition methods, k-means clustering and PNN are used for classification.

Table 1: Classification Results

Classifier	Number of Features	Total Number of Images	Number of images in training		Number of Images in testing		Correct Classification
			19	21	106	20	
k-means+DTCWT	5	166	19	21	106	20	98.6%
PNN+DTCWT	5	166	19	21	106	20	98.4%

Table 2: Comparison with other methods

Method	Techniques used for Classification	Correct Classification Ratio
Our Proposed Work	DTCWT+K-means+PNN	98%
El-Dahshan et al. [3]	DWT+PCA+ANN	95%
Ays.e Demirhan [1]	SWT+SOM	96.8%

The experimental results of the classifiers are compared in Table 1, which shows the number of features extracted by PCA for each classifier, number of normal and abnormal images used for training and testing, number of images misclassified with each classifier and finally the percentage of correct classification ratio for the two different class images. The analysis of the experimental results shows that correction classification ratio 98.4% achieved with the PNN Classifier and 98.6% with K-means Clustering [11].

**CONCLUSION**

In this study, we used the Probabilistic neural network was used for classification based on unsupervised learning using multi wavelet based co occurrence features and target vectors and k-means Clustering. The neural network training and classification was done efficiently by using Matlab neural network toolbox within less time. Finally this system is very useful to diagnose the magnetic resonance brain images for tumour diagnosis. Correct Classification of more than 98.5% in case of PNN and 93% in case of K-means Clustering demonstrate the efficiency of our method.

This method has been applied only to T1 and T2-Weighed images and can be applied to other MRI with more than one slice of brain MRI in order to achieve better accuracy. And also many features can be added other than the existing ones we considered for our proposed work.

**REFERENCES**

- Demirhan Ayse, Mustafa Toru and Inan Guler, 2015. Segmentation of Tumor and Edema Along With Healthy Tissues of Brain Using Wavelets and Neural Networks, IEEE Journal of Biomedical and Health Informatics, 19(4).
- Nazem-Zadeh M.R., E. Davoodi-Bojd and H. Soltanian-Zadeh, 2011. Atlasbased fiber bundle segmentation using principal diffusion directions and spherical harmonic coefficients, NeuroImage, 54: 146-164.
- El-Dahshan, E.A., T. Hosny and A.M. Salem, 2010. Hybrid intelligent techniques for MRI brain images classification, Digital Signal Processing, 20: 433-441.
- Angelini, E.D., O. Clatz, E. Mandonnet, E. Konukoglu, L. Capelle and H. Duffau, 2007. Glioma dynamics and computational models: A review of segmentation, registration and in silico growth algorithms and their clinical applications, Curr. Med. Imag. Rev., 3(4): 262-276.
- Langleben, D.D. and G.M. Segall, 2000. PET in differentiation of recurrent brain tumor from radiation injury, J. Nucl. Med., 41: 1861-1867.
- Demirhan, A. and I. Guler, 2010. Image segmentation using self-organizing maps and gray level co-occurrence matrices, J. Fac. Eng. Arch. Gazi Univ., 25(2): 285-291.

7. Kaus, M., S.K. Warfield, F.A. Jolesz and R. Kikinis, 1999. Adaptive template moderated brain tumor segmentation in MRI, in Proc. Bildverarbeitung für die Medizin, pp: 102-106.
8. Straka, M., A.L. Cruz, A. Kochl, M. Sramek, M.E. Groller and D. Fleischmann, 2003. 3D watershed transform combined with a probabilistic atlas for medical image segmentation, in Proc. 2003 MIT, pp: 1-8.
9. Chang, H.H., D.J. Valentino, G.R. Duckwiler and A.W. Toga, 2007. Segmentation of brain MR images using a charged fluid model, IEEE Trans.Biomed. Eng., 54(10): 1798-1813.
10. Corso, J.J., E. Sharon, S. Dube, S. El-Saden, U. Sinha and A. Yuille, 2008. Efficient multilevel brain tumor segmentation with integrated Bayesian model classification, IEEETrans.Med. Imag, 27(5): 629-640.
11. Corso, J.J., E. Sharon, S. Dube, S. El-Saden, U. Sinha and A. Yuille, 2008. Efficient multilevel brain tumor segmentation with integrated Bayesian model classification, IEEETrans.Med. Imag., 27(5): 629-640.