A Robust Brain MRI Classification with GLCM Features

Sahar Jafarpour
Electrical Engineering Department
Urmia University, Iran

Zahra Sedghi
Electrical Engineering Department
Urmia University, Iran

Mehdi Chehel Amirani
Electrical Engineering Department
Urmia University, Iran

ABSTRACT
Automated and accurate classification of brain MRI is such important that leads us to present a new robust classification technique for analyzing magnetic response images. The proposed method consists of three stages, namely, feature extraction, dimensionality reduction, and classification. We use gray level co-occurrence matrix (GLCM) to extract features from brain MRI and for selecting the best features, PCA+LDA is implemented. The classifiers goal is to classify subjects as normal and abnormal brain MRI. A classification with a success of 100% for two normal and abnormal classes is obtained by the both classifiers based on artificial neural network (ANN) and k-nearest neighbor (k-NN). The proposed method leads to a robust and effective technique, which reduces the computational complexity, and the operational time compared with other recent works.

General Terms
Pattern Recognition, Classification.

Keywords
Brain MRI, Feature extraction, GLCM, ANN, KNN.

1. INTRODUCTION
Nowadays widespread and universal use of computer technology in medical decision support covers a wide range of medical area, such as cancer research, hart diseases, gastroenterology, brain diseases etc. In the recent century, the Computer-Aided Diagnosis (CAD) [1] is progressively becoming an essential area for intelligent systems. Magnetic resonance imaging (MRI) is a valuable diagnostic study, which is a non-aggressive, nonradioactive and pain-free method of assessing the human body, especially the brain.
The importance and necessity in accurate brain pathology diagnosis and treatment requires more accuracy in automatically classifying MRI images in distinguishing disease without human interference.
In CDA there is a challenging process for automatically classifying MRI in normal and abnormal classes. For this goal researchers have proposed a lot of approaches which fall into two categories. One category contains supervised classification techniques such as artificial neural networks (ANN) [2, 3] and support vector machine (SVM) [4]. The other category has unsupervised classification techniques such as self-organization map (SOM) [4] and fuzzy c-means [5]. Since the goal of this study is to design a more efficient and accurate classifier and on the other hand, supervised classifiers in the term of classification accuracy has a better performance than unsupervised classifiers, we use supervised machine learning algorithms (ANN and K-NN) in our proposed method.
Most of works for classifying MRI are based on pattern recognition methods that their main issue is to extract effective features, often by utilizing Digital Wavelet Transformation (DWT) [2, 4, 6] or Co-occurrences Matrix [7].
Gray Level Co-occurrences Matrix (GLCM) introduced in [7,8,9] is used to extract features. GLCM has less computational complexity in comparison to other methods like wavelet transform. The main objective of this paper is to extract effective features using GLCM [7, 10, 11, 16]. These features are completely explained in [8, 9] and [12].
Projection of original feature space, through a transformation, into a smaller subspace is what feature reduction methods do. Linear Discriminant Analysis (LDA) [13, 14] and Principal Component Analysis (PCA) [13, 14] are two major methods which extract new features in different areas. The features extracted by feature extraction methods might have correlations. In the proposed method we use a process of PCA+LDA [13] that leads to the best uncorrelated effective features. In addition to decreasing dimensionality in this method, complexity and time cost are decreased in a satisfied range. In the next step to perform the classification on the input data an artificial neural network (ANN) and a K-nearest neighbor (K-NN) [15] classifiers are used. The proposed method deals with an efficient feature extraction tool and a robust classifier which results in a more robust and accurate automated MRI normal/abnormal brain images classification.
The structure of this paper is organized as following: Section 2 has a short description of our method, which consists of database, feature extraction and feature reduction. Classification methods are presented in section 3. Discussion and comparison with previous works are presented in section 4. Section 5 concludes this paper.

2. METHODOLOGY
The proposed method as illustrated in Fig. 1 is based on the following techniques: Gray Level Co-occurrence Matrix (GLCM), the principle components analysis (PCA), Liner discriminant analysis (LDA), artificial neural network (ANN) and K-nearest neighbor (K-NN). It consists of three stages: feature extraction stage, feature reduction stage and classification stage. KNN and ANN classifiers with three classes as normal, tumoral and MS are used in classification stage.

2.1 Data Base
Two different databases are used in this paper. The first database covers 120 real human brain MRIs with 41 normal and 79 abnormal images, which 43 of them are MS and 36 are tumoral. This data base is collected from the Harvard Medical School website [17]. A sample of each set is illustrated in figure 2. The second used data base is a collection from

Laboratory of Neuro Imaging (LONI) website [18] and Harvard Medical School website [17] which contains 121 normal images and 41 abnormal images which are tomural, MS, Alzheimer, inflammatory, infectious and degenerative diseases. Some samples are illustrated in Figure 3. Both data bases contents are the T2-weighted MR brain images in axial plane and 256×256 in-plane resolution. In the following for comfort we call the first data base Harvard data base and the second one as LONI data base.

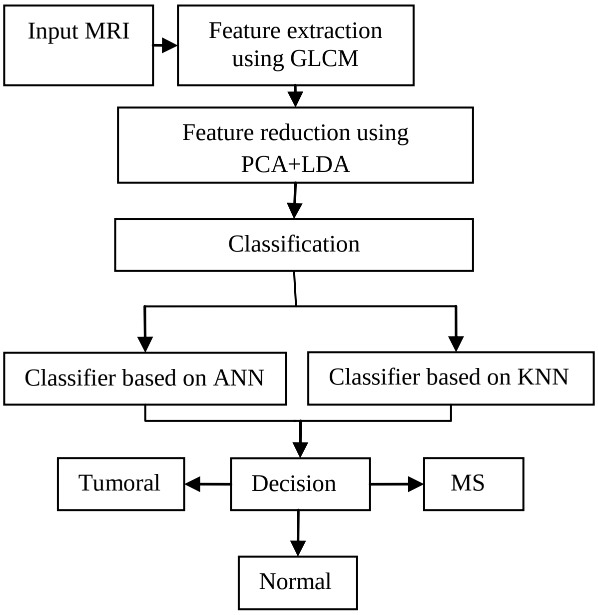

Fig 1: Methodology of the proposed method

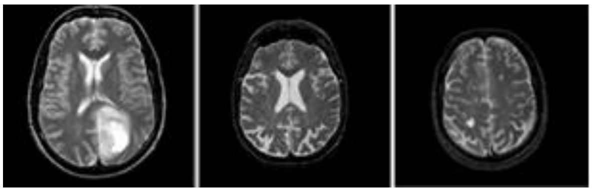

Fig 2: Harvard database samples (a) tumoral (b) normal (c) MS

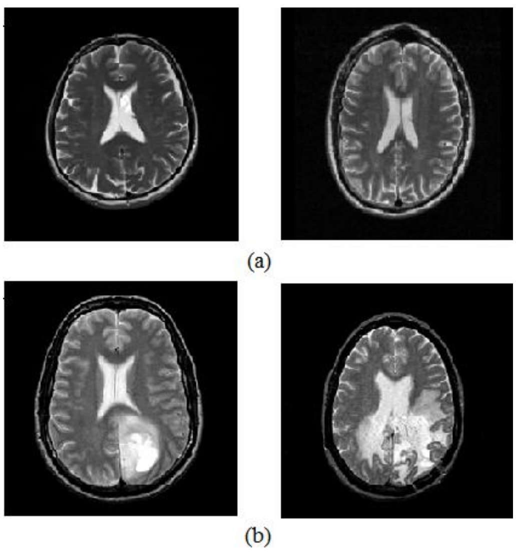

Fig 3: LONI database (a) normal (b) abnormal

2.2 Feature extraction

Features of an image are the properties that completely describe the image. The problem in most previous works is the lack of effective feature selection strategies. We base our feature extraction on the gray level co-occurrence-matrix (GLCM) which is a textural feature extraction based on Haralick et al. Method [8, 9]. This statistic method calculates the co-occurrence-matrix of each image in the database by computing how often a pixel with a certain intensity i occurs in relation with another pixel j at a certain distance d and orientation θ. In this paper the matrix is calculated for only one direction (θ=0) and one distance (d=1). The grey level co-occurrence matrix reveals certain features affecting to the spatial distribution of the grey levels in an image object.
The thirteen Haralick texture features computed from each co-occurrence matrix, produces the set of feature vectors. In addition to these features explained in [7,8,9] which are commonly used in researches, two other features, explained in [12], named cluster shad and cluster prominence, have an effective influence in classification accuracy and are used in our method. Therefore, the feature vector of each image contains fifteen effective GLCM features.
According to the formula below, GLCM is normalized.

$$P(i,j) = \frac{V_{i,j}}{\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} V_{i,j}}$$

The number of columns (j) and rows (i) is equal to the number of gray levels (G) used in image and each matrix element V(i, j), the value of cell (i, j), is normalized as P(i, j).

The features extracted from GLCM are as following:

- Mean:

  $$\mu_i = \sum_{i,j=0}^{G-1} iP(i,j)$$

  $$\mu_i = \sum_{i,j=0}^{G-1} jP(i,j)$$

- Variance

  $$\sigma_i = \sum_{i,j=0}^{G-1}(i-\mu_i)\,P(i,j)$$

  $$\sigma_j = \sum_{i,j=0}^{G-1}\left(j-\mu_j\right)P(i,j)$$

- Entropy:

  $$\text{Entropy} = -\sum_{i,j=0}^{G-1} P(i,j)\log\left(P(i,j)\right)$$

  Higher entropy values are extracted from homogeneous scenes, and lower ones are from inhomogeneous scenes.

- Dissimilarity:

  $$\text{Dissimilarity} = \sum_{i,j=0}^{G-1}|i-j|\,P(i,j)$$

  Dissimilarity is similar to GLCM contrast and it is high if the local region has high contrast.

- Contrast:

  $$\text{contrast} = \sum_{i,j=0}^{G-1}(i-j)^2 P(i,j)$$

  Contrast or local intensity variation measures the distance from the mean diagonal of gray-level co-occurrence matrix and the more the distance the higher the weight that is assigned to P(i, j),so contrast exponentially increases when i-j increases.

- Inverse Difference Moment (Homogeneity):

  $$\text{IDM} = \sum_{i,j=0}^{G-1}\frac{P(i,j)}{1+(i-j)^2}$$

IDM measures the closeness of distribution of GLCM elements to main diagonal. The more concentration along main diagonal in GLCM leads to more homogeneous area and therefore higher values for IDM.

- Correlation:

$$\text{correlation} = \sum_{i,j=0}^{G-1} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j}$$

It measures the gray level linear dependency between neighboring pixels.

- Energy:

$$\text{ENRGY} = \sum_{i,j=0}^{G-1} (P(i,j))^2$$

- Cluster shade

$$\text{SHADE} = \sum_{i,j=0}^{G-1} \left(i + j - \mu_i - \mu_j\right)^3 P(i,j)$$

- Cluster prominence

$$\text{PROM} = \sum_{i,j=0}^{G-1} \left(i + j - \mu_i - \mu_j\right)^4 P(i,j)$$

- Sum entropy

$$\text{SENT} = -\sum_{i=0}^{2G-2} P_{x+y}(i) \log(P_{x+y}(i))$$

- Sum average

$$\text{AVRE} = \sum_{i=0}^{2G-2} iP_{x+y}(i)$$

- Difference entropy

$$\text{DENT} = -\sum_{i=0}^{G-1} P_{x+y}(i) \log(P_{x+y}(i))$$

- Sum variance

$$\text{SVAR} = \sum_{i=0}^{2G-2} (i - SENT)^2 P_{x+y}(i)$$

- Information measures of correlation

In the previse works using GLCM, a few common and popular features were extracted for training the classifier which resulted in less accuracy rate in compare with other feature extractor methods like wavelet transformation. However, in our proposed method extracting powerful features made a robust classification with a growth in accuracy rate.

2.3 Feature Reduction

Lower computational time and less memory are two predominant reasons for minimizing the dimensionality of pattern representation. Principal Component Analysis (PCA), an unsupervised linear method, finds a set of the most representative projection vectors such that the projected samples preserve the most information about original samples. Linear Discriminant Analysis (LDA), a supervised linear discriminator, uses the class information and finds a set of vectors that maximizes the between-class scatter while minimizing the within-class scatter [14].

In the proposed method by using PCA+LDA, we obtain a combining process. The first processing step is PCA transformation without dimension reduction, in other words, as in [14] all the eigenvalues are kept in a matrix. Then numbers of eigenvalues, which have highest and effective values, are computed. Figure 4 illustrates that only two features from the fifteen features have the essential information required. In this graph the average cumulative sum of the eigenvalues, obtained from PCA, is depicted against the number of eigenvalues. It shows that the sum of two largest eigenvalues has the value of 99.99 percentages of the whole eigenvalues. This means that the values after third eigenvalue are enough small to not affect the results. Therefore, we have an action of LDA in second step where feature matrix dimensionality reduction discounts features from 15 to 2. Limiting the feature vectors by such a combining process leads to an increase in accuracy rates and a decrease in complexity and computational time.

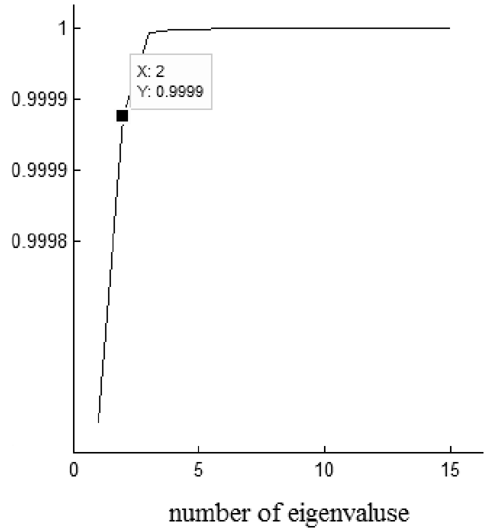

Fig 4: Cumulative sum of eigenvalues per sum of all the eigenvalues

3. CLASSIFICATION

3.1 Artificial Neural Network (ANN)

Over the last few years an increasingly interest has emerged in Neural networks. In this classifier whereas no information needed about the probability distribution and the a priori probabilities of different classes, neural networks are widely used in pattern classification. In this method we use a single-hidden-layer back propagation neural network which is adopted with sigmoid neurons in the hidden layer and linear neuron in the output layer. It consists of two neurons in the input layer, three as hidden layer and one for output layer as depicted in Figure 5.

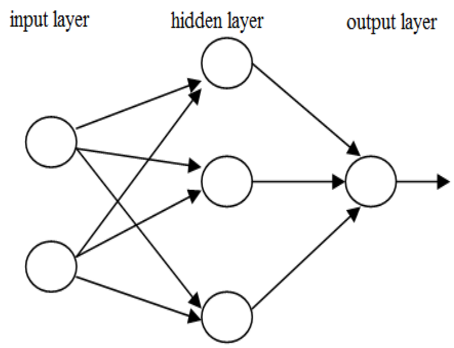

Fig 5: A two-layer ANN used for brain classification

3.2 K-nearest Neighbor (K-NN)

The k- nearest neighbor classifier is a simple supervised classifier that has yield good performance for optimal values of k. This classifier computes the distance from the unlabeled data to every training data point and selects the best k neighbors with the shortest distance. No requirement for training process makes this classifiers implementation simple.
In this work, the Euclidean distance is used for distance metric and after trials and errors, k=1 is the best value obtained.

4. Comparison and Experimental Results

This experiment has applied a supervised method for diagnosis normal, tumoral and MS brain images. As mentioned, the algorithm employs three stages: feature extraction, feature reduction and classification. After the features computed with GLCM, PCA+LDA minimizes the dimension of feature vector as possible which leads to an increase in the classification accuracy and a decrease in computational complexity. Two classifiers KNN and ANN are used for classifying the images in a set of two or three classes.
Harvard database is used in two different categories first as a set of three classes: normal, tumoral and MS brain images and second as a set of two classes: normal and abnormal brain images. For accurate results the proposed method is implemented twice. Once the method is trained by 65% of the whole data and the other time 50% of the data are used for training. Table 1 and Table 2 illustrate the results of the classification. The results demonstrated in tables for both implementations, express that classification accuracy in the first category is obtained 100% for the normal and tumoral sets and for MS is 92.86%. In other words normal and abnormal sets have completely separated from each other and the only one miss classified image is in the MS class. Computational time is an important factor in different algorithms. This time contains two parts: First, feature extraction time and second, feature reduction plus classification time. In the proposed method, computation time is resulted from an average of 1000 repetitions, which is computed for each image in the both classifiers. Average computation time for feature extraction is obtained 0.0246 seconds. For feature reduction with KNN, the time is obtained 22µs and with ANN, it is obtained 28ms. Therefore, the average computational time is about 0.025s, which is better in compare to similar works done with same dataset. For verifying the robustness of our proposed method, we apply the second database (LONI database). The results obtained from LONI database with two sets of 65% and 50% of the whole database as training sets, are the same as the results obtained from Harvard database with the accuracy of 100% in the normal and abnormal category.

Table 1. Confusing matrix of normal and abnormal (for both of datasets)

Abnormal	normal	input image / Classification
0%	100%	normal
100%	0%	Abnormal

Table 2. Confusing matrix for normal, MS and tumoral categories (With both of KNN and ANN classifier)

MS class	Tumoral class	Normal class	Input image / Classification
0%	0%	100%	Normal class
7.14%	100%	0%	Tumoral class
92.86%	0%	0%	MS class

For decreasing the computational complexity, instead of using DWT, GLCM is used. The previous works with GLCM could not reach to the maximum rate in classification but in the present method by extracting the best features from gray level co-occurrence matrix that were not used in previous works, we obtained the maximum possible correct rate.
Scatter plots depicted in Figure 6 and Figure 7 illustrate the robustness of the proposed method in classifying normal and abnormal classes. The two best features selected from the feature reduction step, completely separates the normal and abnormal classes from each other in the both databases. In Harvard database, as depicted in Figure 7, two images in MS set are near by the tumoral set, which seems to have two miss classified images. However, with implementing ANN with suitable weights and KNN with the best k, only one image is misclassified and we receive to an accuracy of 92.86%.

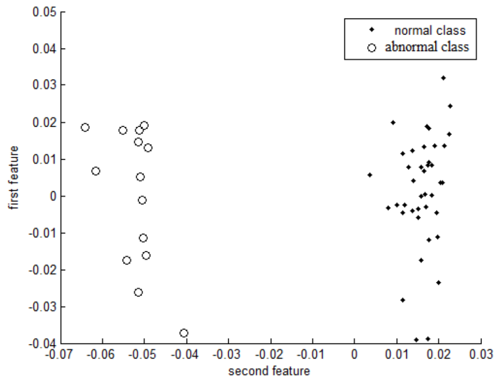

Fig 6: Scatter plot for LONI database

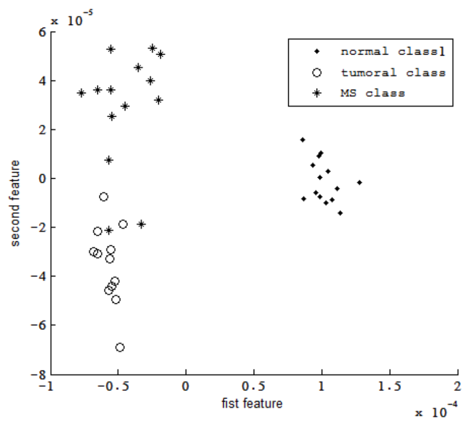

Fig 7: Scatter plot for Harvard database

In many researches for MRI, digital wavelet transformation (DWT) is used for feature extraction, which leads in good results. The comparison of different methods is illustrated in Table 3.

Table 3. Comparison with other methods reported in the literature for same dataset

Techniques Used for Classification	P (%)
Proposed Method GLCM+ PCA+ LDA+ k-NN	100
Proposed method GLCM+PCA+LDA+ANN with computational time 0.025 s	100
DWT+PCA+ANN [2]	97
DWT+ PCA+ *k*-NN [2]	98
DWT+SOM [4]	94
DWT+SVM with linear kernel [4]	96
DWT+SVM with radial basis function based kernel [4]	98
DWT+ PCA +BPNN with Computational time 0.0451 s [6]	100
GLCM+PCA+SVM [7]	95

5. CONCLUTION

In this study, we have developed a medical decision support system with three class sets as normal, tumoral and MS MRIs.

This automatic detection system which is designed by gray level co-occurrence matrix (GLCM), principal component analysis (PCA), Linear discriminant analysis (LDA) and supervised learning methods (ANN and k-NN) obtains very satisfactory and promising results to assist the diagnosis of brain disease without hesitation.

The methodology developed in this study is based on using the effective image features and employing a hybrid feature reduction technique toward distinguishing normal, MS and tumoral brain MRIs. Our work produces 100% accuracy rate for normal, 100% for tumoral and 92.86% for MS images with both KNN and ANN classifiers. In addition, with employing LONI database we obtain an accuracy of 100% in classifying normal and abnormal MRIs, which confirms previous records. Other improved factor in this research is computation time. The addition of maximum accuracy, computation time for feature extraction and classification of each image is only 0.025s. GLCM according to its low computational complexity and low computational time leads in better results than DWT methods.

The proposed method applied to only on T2-weighted images and just for diagnosis of two kind of brain abnormality. So future works should focus on selecting other effective features from the gray level co-occurrence matrix and employing the proposed method on the other types of MRIs such as T1-weighted, proton-density weighted, and diffusion-weighted MRIs and multiple-class classifications for brain MRIs should be extended for more kind of brain disease like Alzheimer, sarcoma etc.

6. REFERENCES

[1] Gorunescu, F., 2007. Data Mining Techniques in Computer-Aided Diagnosis: Non-Invasive Cancer Detection. World Academy of Science, Engineering and Technology 34.

[2] Dahshan, A. E., Salem, and Younis, T. H., 2009. A hybrid technique for automatic MRI brain images classification, BABES–BOLYAI, Informatica, Volume LIV.

[3] Abdolmaleki, P., Mihara, F., Masuda, K., and Buadu, L. D., 1977. Neural networks analysis of astrocytic gliomas from MRI appearances, Cancer Letters 118, pp. 69-78.

[4] Chaplot, S., Patnaik, L.M., and Jagannathan, N.R., 2006. Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network, Biomedical Signal Processing and Control 1, pp. 86-92.

[5] Maitra, M., Chatterjee, A., 2007. Hybrid multi resolution Slantlet transform and fuzzy c- means clustering approach for normal-pathological brain MR image segregation, Med Eng Phys, doi:10.1016/j. medengphy.

[6] Zhang, Y., Dong, Z., Wu, L. and Wang, SH., 2011. A hybrid method for MRI brain image classification, Vol. 38, pp. 10049-10053.

[7] Nazari, M.R. and Fatemizadeh, E, 2010. A CBIR system for human brain magnetic resonance image indexing. International Journal of Computer Applications, vol. 7, pp. 33–37.

[8] Haralick, R. M., 1979. Statistical and structural approach to texture, proceeding of the IEEE, vol. 67, pp. 786-804.

[9] Haralick, R. M., Shanmugam, K. and Dinstein, I., 1973. Textural Features for Image Classification, IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-3, pp. 610-621.

[10] Albregtsen, F., 2008. Statistical Texture Measures Computed from Gray Level Co-occurrence Matrices

[11] Albregtsen.F. , 2008. Statistical Texture Measures Computed from Gray Level Co-occurrence Matrices

[12] Soh, L.K., 1999. Texture Analysis of SAR Sea Ice Imagery Using Gray Level Co-Occurrence Matrices.

[13] Duda, R. O., Hart, P. E., and Stork, D. G., 2001. Pattern Classification, New York: John Wiley and Sons.

[14] You, M., Chun, CH., Bu, J., Lu, J. and Tao, J. Emotion Recognition From Speech Signals Combining PCA And LDA

[15] Cocosco, C., Zijdenbose, A. P.,and Evans A. C., 2003. A fully automatic and robust brain tissue classification method, Medical Image Analysis 7, pp.513-527.

[16] Joshi, J. and Phadke, A.C., 2010. Feature Extraction and Texture Classification in MRI.

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

[18] Laboratory of Neuro Imaging (LONI), web, http://www.loni.ucla.edu/.