IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)
ISSN: 2278-2834, ISBN: 2278-8735, PP: 19-23
www.iosrjournals.org

PCA AND PNN ASSISTED AUTOMATED BRAIN TUMOR CLASSIFICATION

Varada S.Kolge [1], Prof.K.V.Kulhalli [2]
[1](Department of Electronics and Telecommunication, Shivaji University, DYPCET, Kolhapur, India)
[2](Department of Information Technology, Shivaji University, DYPCET, Kolhapur, India)

ABSTRACT : *The National Brain Tumor Foundation (NBTF) for research in United States estimated that 29,000 people in the U.S. were diagnosed with primary brain tumor each year and nearly 13000 people died according to the 2007 estimate. This count is continuously increasing. There are about 200 other types of tumors diagnosed in U.K. each year. Conventional methods of brain tumor classification and detection is by human inspection which are having certain disadvantages so Probabilistic Neural Network (PNN) with mathematical technique called Principal Component Analysis(PCA) is used to give more accurate and fast solution. Automated classification of brain tumors is performed in two stages. Feature extraction using PCA and classification using PNN. Probabilistic Neural Network is mathematical analogues of biological neuron system, and they are made up of a parallel interconnected system of nodes called neurons. Depending on the application, the learning scheme is chosen to train the neural network.*

Keywords **-** PRINCIPAL COMPONENT ANALYSIS, PROBABILISTIC NEURAL NETWORK, MEDICAL RESONANCE

I. INTRODUCTION

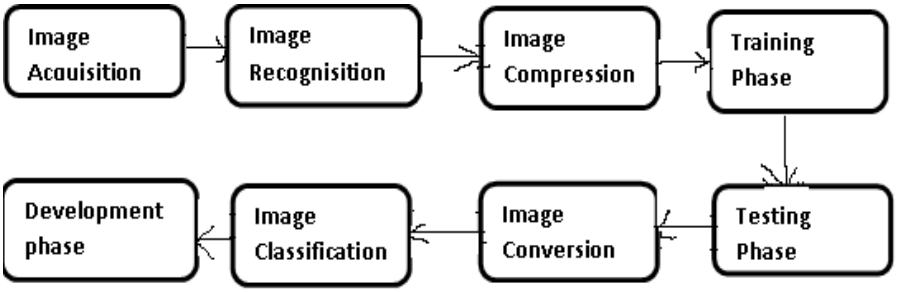
Probabilistic Neural Network (PNN) with image and data processing techniques will be employed to implement an automated brain tumor classification. The conventional method for Medical Resonance (MR) brain images classification and tumors detection is by human inspection. Operator-assisted classification methods are impractical for large amounts of data and are also non-reproducible. Medical Resonance (MR) images contain a noise caused by operator performance which can lead to inaccuracies in classification. The use of artificial intelligent techniques like neural networks, and fuzzy logic has shown great potential in this field. Conventional methods of monitoring and diagnosing the diseases rely on detecting the presence of particular features by a human observer. Due to large number of patients in intensive care units and the need for continuous observation of such conditions, several techniques for automated diagnostic systems have been developed in recent years to attempt to solve this problem. Such techniques work by transforming the mostly qualitative diagnostic criteria into a more objective quantitative feature classification problem. Brain tumor is one of the major causes in increase in mortality among children and adults. A tumor is a mass of tissue that grows out of control of the normal forces that regulates growth. The complex brain tumors can be classified into two general categories depending on the tumor origin, their growth pattern and malignancy. Primary brain tumors are tumors that arise from cells in the brain or the covering of the brain. A secondary or metastatic brain tumor occurs when cancer cells spread to the brain from a primary cancer to the other part of the body. PNN are mathematical analogous to biological neuron system. They are made up of parallel interconnected system of nodes called neurons. Combination of PNN with different types of learning schemes results in a variety of PNN systems. All the PNN systems do not yield a satisfactory result in all the practical applications. Depending on the specific requirement, PNN system is to be designed. This document describes the use of PCA and PNN in automated classification of the brain tumors. PCA is a mathematical technique that is used to reduce the large dimensionality of the data and then PNN can be used for classification of the tumors.

II. PRESENT THEORIES

Artificial neural networks are finding many uses in the medical diagnosis application. According to Qeethara Kadhim Al-Shayea [1].Artificial neural networks provide a powerful tool to help doctors to analyze, model and make sense of complex clinical data across a broad range of medical applications. Most of the applications are providing solution to the classification problems. According to N. Kwak, and C. H. Choi [2] Feature selection plays an important role in classifying systems such as neural networks (NN). The higher performance with lower computational effort is expected with this process. One of the most popular methods for dealing with this problem is the principal component analysis (PCA) method. This method transforms the existing attributes into new ones considered to be crucial. E. D. Ubeyli and I. Guler [3] used feature extraction methods in automated diagnosis of arterial diseases. Since classification is more accurate when the pattern is simplified through representation by important features, feature extraction and selection play an important role in classifying

systems. T.Logeswari and M. Karnan [5] used image segmentation based on the soft computing for improved implementation of the brain tumor detection. The MRI brain image is acquired from patient's database and then Image acquisition, preprocessing, image segmentation is performed for brain tumor detection. Georgiadis. Et all [6] also did the work for improving brain tumor characterization on MRI by probabilistic neural network and non-linear transformation of textural features. According to Chettri, S. R. and Cromp, R.F., the probabilistic neural network architecture can be used for high speed classification of remotely sensed imagery. Probabilistic Neural Network can be applied to remotely sensed data

III. PROPOSED BLOCK DIAGRAM OF THE SYSTEM



Here the automated classification of brain magnetic resonance images by using some prior knowledge like pixel intensity and some anatomical features are proposed [7]. Currently there are no methods widely accepted, therefore automatic and reliable methods for tumor detection are of great need and interest. The application of PNN in the classification of data for MR images problems are not fully utilized yet. These include the clustering and classification techniques especially for MR images problems with huge scale of data and consuming times and energy if done manually. Thus, fully understanding the recognition, classification or clustering techniques is essential to the developments of Neural Network systems particularly in medicine problems

Decision making will be performed in two stages:

1. Feature extraction using the Principal Component Analysis (PCA) and
2. Classification using Probabilistic Neural Network (PNN).

The performance of the PNN classifier will be evaluated in terms of training performance and classification accuracies. Probabilistic Neural Network gives fast and accurate classification and will be a promising tool for classification of the tumors.

Image Acquisition: - Maximum MR Images of brain as possible from the Radiologists, Internet, Medical Atlases, Hospitals or other resources are to be collected.

Image recognition and Image compression: - Mathematical technique of PCA will be used for Image recognition and Image compression.

Training Phase: - Feature vectors for each image from the training set will be extracted in this phase.

Testing Phase: - Feature vector of the test image will be computed in this phase.

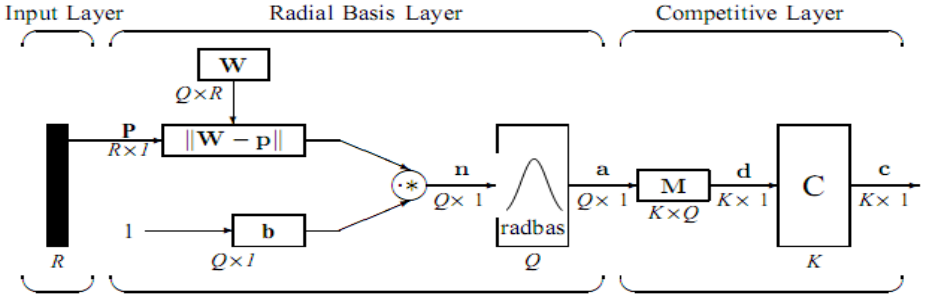
Image Conversion: - MR images will be converted into matrices form using MATLAB or SCILAB as a tool.

Image Classification: - Feed Forward PNN will be used to classify MR images.

Development Phase: - Performance analysis based on the result will be carried out in the development phase.

The block diagram of the above proposed system, is as follows

IV. PROPOSED NETWORK SYSTEM



1) Input Layer: The input vector, denoted as **p**, is presented as the black vertical bar in Fig. 2. Its dimension is *R × 1*.

2) Radial Basis Layer*:* In Radial Basis Layer, the vector distances between input vector p and the weight vector made of each row of weight matrix **W** are calculated. Here, the vector distance is defined as the dot product between two vectors . Assume the dimension of **W** is *Q*×*R*. The dot product between **p** and the *i*-th row of **W** produces the *i*-th element of the distance vector ||**W** □ **p**||, whose dimension is *Q*×*1*, as shown in Fig. 2. The minus symbol, "□", indicates that it is the distance between vectors. Then, the bias vector **b** is combined with ||**W** □ **p**|| by an element-by-element multiplication, represented as "**.***" in Fig. 2. The result is denoted as **n** = ||**W** □ **p**|| **.***p**. The transfer function in PNN has built into a distance criterion with respect to a center. In this paper, it is defined as *radbas(n)* = *e*−*n*2 (1)Each element of **n** is substituted into Eq. 1 and produces corresponding element of **a**, the output vector of Radial Basis Layer. The *i*-th element of **a** can be represented as ai = *radbas*(||**W**i □ **p**|| **.***b**i) (2) where **W**i is the vector made of the *i*-th row of **W** and **b**i is the *i*-th element of bias vector **b**

3) Competitive Layer: There is no bias in Competitive Layer. In Competitive Layer, the vector **a** is firstly multiplied with layer weight matrix **M**, producing an output vector **d**. The competitive function, denoted as **C** in Fig. 2, produces a 1 corresponding to the largest element of **d**, and 0's elsewhere. The output vector of competitive function is denoted as **c**. The index of 1 in **c** is the number of tumor that the system can classify.

V. MATHEMATICAL BACKGROUND REQUIRED FOR PCA

- **Mean**
- **Standard Deviation**
- **Variance**
- **Covariance**
- **Covariance matrix.**
- **Eigen vector (can only be found for square matrix)**
- **Eigen value**
- **Mean**

Notice the symbol $\bar{X}$ (said "X bar") to indicate the mean of the set. All this formula says is "Add up all the numbers and then divide by how many there are".

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- **Standard Deviation**

The Standard Deviation (SD) of a data set is a measure of how spread out the data is.How do we calculate it? The English definition of the SD is: "The average distance from the mean of the data set to a point". The way to calculate it is to compute the squares of the distance from each data point to the mean of the set, add them all up, divide by, (n-1) and take the positive square root. As a formula

$$s = \sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)}}$$

- **Variance**

Variance is another measure of the spread of data in a data set. In fact it is almost identical to the standard deviation. The formula is this:

$$var(X) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})}{(n-1)}$$

- **Covariance**

Standard deviation and variance only operate on 1 dimension, so that you could only calculate the standard deviation for each dimension of the data set *independently* of the other dimensions. However, it is useful to have a similar measure to find out howmuch the dimensions vary from the mean *with respect to each other.* Covariance is such a measure. Covariance is always measured *between* 2 dimensions. If you calculate the covariance between one dimension and *itself*, you get the variance. So, if you had a 3-dimensional data set (x,y,z), then you could measure the covariance between the x and y dimensions, the x and z dimensions, and the y and z dimensions. Measuring the covariance between x and x , or y and y , or z and z would give you the variance

of the < , = and > dimensions respectively.The formula for covariance is very similar to the formula for variance. The formula for variance could also be written like this:

$$cov(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)}$$

- **Covariance matrix.**

covariance is always measured between 2 dimensions. If we have a data set with more than 2 dimensions, there is more than one covariance measurement that can be calculated. For three dimensional (x, y, z) data set calculate cov(x,y), cov(x,z) and cov(y,z) In fact for n dimensional data set, we can calculate total covariance values.

$$\frac{n!}{(n-2)!*2}$$

$$C = \begin{pmatrix} cov(x,x) & cov(x,y) & cov(x,z) \\ cov(y,x) & cov(y,y) & cov(y,z) \\ cov(z,x) & cov(z,y) & cov(z,z) \end{pmatrix}$$

- **Eigen vector (can only be found for square matrix)**

As you know, you can multiply two matrices together, provided they are compatible sizes. Eigenvectors are a special case of this. Consider the two multiplications between a matrix and a vector in In the first example, the resulting vector is not an integer multiple of the original vector, whereas in the second example, the example is exactly 4 times the vector we began with. Why is this? Well, the vector is a vector in 2 dimensional space The other matrix, the square one, can be thought of as a transformation matrix. If you multiply this matrix on the left of a vector, the answer is another vector that is transformed from it's original position. What properties do these eigenvectors have? You should first know that eigenvectors can only be found for *square* matrices. And, not every square matrix has eigenvectors.

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 11 \\ 5 \end{pmatrix}$$

$$\begin{pmatrix} 2 & 3 \\ 2 & 1 \end{pmatrix} \times \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 12 \\ 8 \end{pmatrix} = 4 \times \begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

10. STEPS TO BE FOLLOWED IN PCA

1) Get some data
2) Subtract the mean
3) Calculate the covariance matrix
4) Calculate the eigenvectors and eigenvalues of the covariance matrix
5) Choosing components and forming a feature vector
6) Deriving the new data set

VI. ALGORITHM FOLLOWED IN MATLAB CODING

- MRI images of brain are taken in the database.
- Some of the images are normal, some are benign, some are malignant.
- One MRI image is selected as the input image.
- Resize all the images in one standard format.
- Convert color image into gray image.
- Rename the images. E.g 1.jpg,2.jpg etc.
- Reshape the images from 2D to 1D
- Apply PCA and find feature vectors of all images in the database

VII. CONCLUSION

Automated classification and detection of tumors in different medical images is motivated by the necessity of high accuracy when dealing with a human life. Computer assistance is demanded in all medical applications as it will definitely improve the results of humans. The use of PCA to reduce the dimensionality of the data and the use of PNN for tumor classification will improve the speed and accuracy of the result.

VIII. ACKNOWLEDGEMENT

Authors thank all the Faculty members of Electronics and Telecommunication Dept. and Information Technology Dept, of DYPCET, Shivaji University, Kolhapur, Maharashtra, India for their continuous valuable motivation.

REFERENCES

.

Journal Papers:

[1] Qeethara Kadhim Al-Shayea, "Artificial Neural Networks in Medical Diagnosis" International Journal of Computer Science Issues, Vol 8,Issue 2, March 2011.

[2] N. Kwak, and C. H. Choi, "Input Feature Selection for Classification Problems", IEEE Transactions on Neural Networks, Vol.13, Issue 1, 143–159, 2002.

[3] E. D. Ubeyli and I. Guler, "Feature Extraction from Doppler Ultrasound Signals for Automated Diagnostic Systems", Computers in Biology and Medicine, Vol 35, Issue 9, 735–764, 2005.

[4] D.F. Specht, "Probabilistic Neural Networks for Classification, mapping, or associative memory", Proceedings of IEEE International Conference on Neural Networks, Vol.1, IEEE Press, New York, pp. 525-532, June 1988.

[5] T.Logeswari , and M. Karnan, "An improved implementation of brain tumor detection using segmentation based on soft computing", Journal of Cancer Research and Experimental Oncology Vol 2, Issue 1.

[6] Georgiadis. Et all , " Improving brain tumor characterization on MRI by probabilistic neural networks and non-linear transformation of textural features", Computer Methods and program in biomedicine, vol 89, pp24-32, 2008

[7] Mohd. Fauzi Othman and Mohd. Ariffanan Mohd. Basri, "Probabilistic Neural Network For Brain Tumor Classification", Second International Conference on Intelligent Systems, modeling and simulation.