ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Pattern Recognition Using PCA algorithms.

G Uma Maheswari*, EP Ephzibah, and V Mareeswari.

School of Information Technology and Engineering, VIT University, Vellore-14, India

ABSTRACT

To distinct among various brain pathologies the information output from spectroscopy is one of the very important pointers. The attitude is to examine the effectiveness of a pattern recognition system in the task of discerning between meningiomas and metastatic brain cancers. The information used for this analysis was a set of Magnetic Resonance Spectroscopy images copied from internet with verified tumors. In this analysis we have applied Discrete Wavelet Transform for the attribute extraction and Principal Component Analysis algorithm for attribute reduction and Support Vector Machine for classification. The system was accomplished using an external cross verification process.

Keywords: Brain cancers, Magnetic Resonance Image, Magnetic Resonance Spectroscopy, Pattern classification, volumetric textural features

**Corresponding author*

INTRODUCTION

Brain cancer is frequent in adults; it grows mainly in brain or cerebellum and in very rare case in the meninges. It is necessary to detect the tumor in time and also need to examine and conclude that what actually it is, so that appropriate treatment can be provided on time .For all these we need not only require noninvasive technique of diagnosis but also sophisticated management of the information system. We can use only the available medically available data for analysis and knowledge withdrawal. The data may be in various forms images, database or signals. This information is very important boon for identifying the pattern and making intelligent decisions. The cerebral metastases are the most common brain tumors affecting adults. These grow generally in the brain or cerebellum and less frequently in the meninges [1]. However, solitary dural metastases are the second most frequent meningeal lesions, and from an imaging point of view these are difficult to discerning from meningiomas [2]. Both of these canceral forms can uptake magnetic resonance imaging (MRI) contrast agent in a like manner [3]. Many literatures are present which say about the problem of differential diagnosis between these two types of tumors. The majority of this analysis has come to the result that there is a radiographic matching between these two types of tumors while according to some of them radiological images of metastases even display a “dural tail,” a sign usually associated with meningiomas. In particular, According to Tagle et al. [1] four cases of isolated meningeal metastases where in all of them a meningioma has been considered as the main preoperative diagnosis. Lath et al. [2] have recognized a case of extra-axial oligoastrocytomas, 92% for oligodendrogliomas, 96% for meningiomas and 98% for gliomas by employing a combination of four textural features (the averaged values of pixel intensities within the tumor boundaries from four sequences — T1, T2, Proton Density [PD] and gadolinium [GD] enhanced T1) and 10 spectroscopic features (one from each quantified metabolite from the MR spectra) and the LS-SVM classification algorithm. In another study , Devos t al.[13] by combining the similar textural and spectroscopic features, 99% of mean classification accuracy achieved for discriminating low- from high-grade gliomas, low- from high-grade tumors, meningiomas from gliomas and Grade II from Grade III gliomas. Though both of these studies, which have trusted on the same set of data [20], have used a common methodology; they have (a) utilized one textural feature (average value), thus it is depriving their system, from higher-order, information-rich, textural features, and (b) have used attributes of Magnetic Resonance Spectroscopy data. The organization of the paper is as follows

In section 2 System organization is given, section 3 Preprocessing methods is discussed, section 4 attribute extraction using DWT is done, section 5 feature reduction using PCA algorithm is performed, section 6 classification using Support Vector Machine is done finally conclusion is drawn in Section 7.

SYSTEM ORGANIZATION

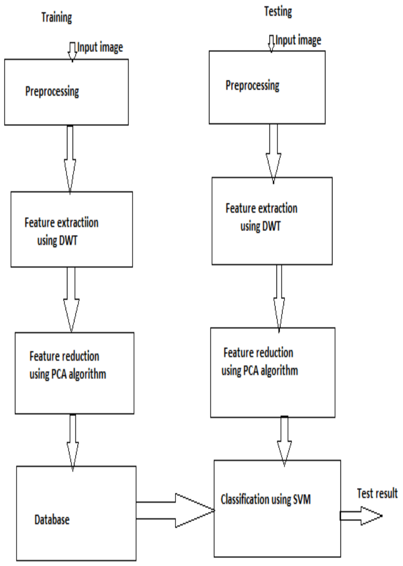


Fig 1: System organization

As given in the organization diagram, the Magnetic Resonance Spectroscopy images were first preprocessed in training phase and then feature extraction is performed in the DWT of the image. The features extracted from DWT are of high dimensionality, due to this reason dimensionality reduction PCA algorithm is used. Finally all the data extracted after Principal Component Analysis algorithm process is stored in the database. In testing phase the MRS image which is to be tested for classification is also processed by Preprocessing, DWT and PCA. After this the classification is done for the test result.

Steps for preprocessing

Resize

If the input image is larger in size, processing is easier. To perform the comparison process, logical we need to resize the image. The size of the image used (256x256), the images which are not of this size were resized first it is an optional step.

Enhancement

In an input image there is usually a chance for the presence of various types of noises and this can affect our conclusion significantly. So we have to select a technique according to the type of noise present in the image. In this paper we have used median filter for noise removal.

Segmentation

We have to select the area in the image so that only that part of the image will be forwarded for DWT processing. In other way we can say that in an image there is certain portions which of our interest and other parts of the image are of less interest for us so it is also a very major and necessary point for this analysis procedure .Instead we can go for segmentation also. Here the implementation is done using the inbuilt function of Matlab “imcrop()” which passes the specific portion of the image for attribute extraction and then the other tasks are performed.

Discrete Wavelet Transform

The discrete wavelet transform plays a vital role in image processing. Here during the process of training, Discrete Wavelet Transformation is used for attribute extraction of Magnetic Resonance Spectroscopy data and the feature reduction is done using PCA Algorithm. For each of the MRS images the resultant data was stored in the database. In this way a robust database was created with a number of samples and the data related to each of its attributes. Extracting relevant data will be a very difficult task in the presence of noise and these noises increase the complexity of the process geometrically [14]. To avoid the above problems use of wavelet is being proposed. As an example filtering processes based on the continuous Wavelet Transform method were proposed to cut off the lactate signal from overlapping lipid resonances. An explorative analysis of the use of wavelet techniques in this specific area can be found in (García-Gómez et al., 2009). (MRS signal-based brain tumour diagnostic analysis), with a similar setting to the one used in. Consider a signal x (t), the signal after transformation is $w(\tau)$, this signal is a function of the translation parameter τ and scale s .Here the signal energy is normalized at every scale by dividing the wavelet coefficients by $\frac{1}{\sqrt{|s|}}$ .This ensures that the wavelets will have the same energy at every scale. It is possible to reconstruct the original signal by using the inverse CWT. It is defined by

$$x(t) = \frac{1}{C_\psi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(\tau, s) \frac{1}{s^2} \psi\left(\frac{t-\tau}{s}\right) d\tau ds. \qquad (2)$$

$$C_\psi = \int_0^\infty \frac{|\psi(f)|^2}{f} df < \infty.$$

The term Wavelet means a small piece of wave like Sine and Cosine functions keep repeating their behavior after certain intervals up to infinity. The wavelet exists only for a finite domain; its value at all other points is zero. The discrete wavelet transform is a method in which we convolve the signal against a particular instance of the wavelet at various time and positions. The wavelet transformation is a join of time and frequency domain since we can model both change in frequency and change in time. During the process of wavelet transform the scale and the step spacing between the wavelets is increased by a factor of two in each step.

For example if we use wavelet of scale 1 and place those by one sample apart. In the second pass, scale them to two samples, and place those by two sample points.

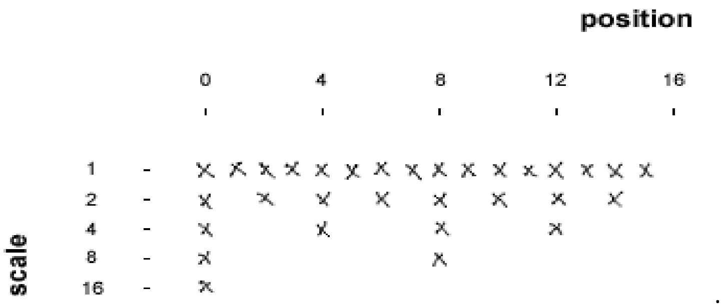


Fig2: symbolic representation

The same number of calculations for each convolution, but less actually needed to be done in each pass, in the first pass [9], where the wavelets are at a scale to take up a single sample point is unnecessary. From this we can understand that the transformation takes up the same amount of space as the original data. We take a low pass (scaling function) and a high pass (wavelet function) version of the curve and split the high pass and low pass information. The wavelet transform is treated as a filter bank, and we iterate along the low pass sub band. The low pass data is treated as a signal in its own right and is subdivided into its own low and high sub bands.

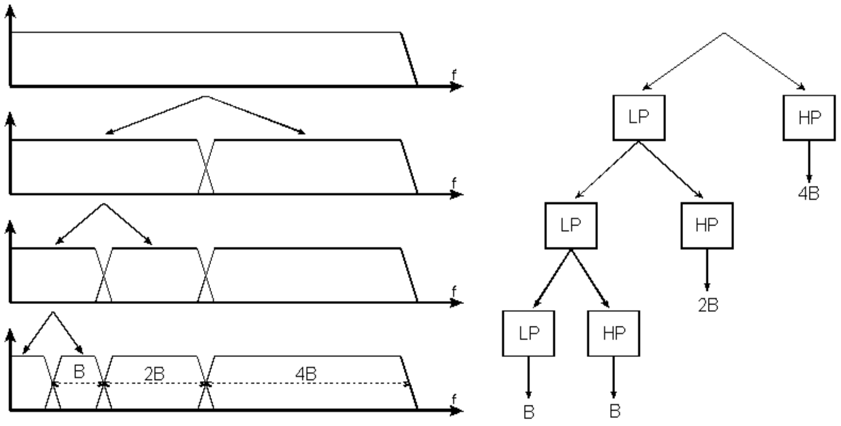
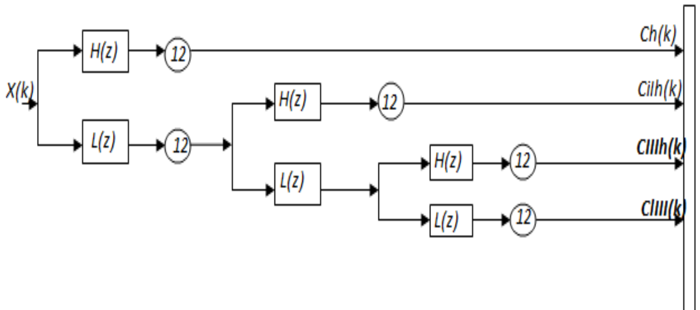
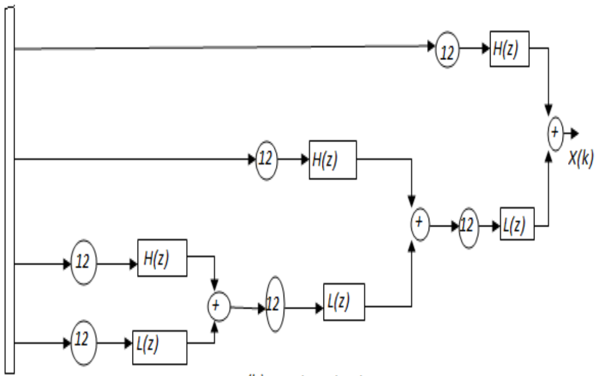


fig 3 : Discrete Wavelet Transform

At last, we can keep going down until the low pass and high pass bands are a single valued, but how many steps to go it depends on the application and on images so we need to decide it manually. Many real time applications need to be redefined in a discrete domain [8]. The process is being performed in a discrete frequency domain. The implementation of DWT is done by the octave filter bank, in a cascade arrangement of low-pass L(z) and high-pass H(z) filters, followed by sub sampling, as illustrated in Fig. 2.



(a) Analysis bank



(b) Synthesis bank

Fig4: Analysis and synthesis

The original signal can be reconstructed by using the synthesis filter bank where the signals were sampled and were passed through the filters L0(z) and H0(z). Here we can find some rounding errors, but preferably it leads to the restoration of the original image if none of the coefficients was altered.

Attribute reduction using Principle Component Analysis

After processing and performing the filtering operation in MRS data we get some data. In image processing it is termed as diagonal component, Average component, Vertical component and Horizontal component. By using the obtained decomposition coefficients as input we perform dimensionality reduction process, using Principle Component Analysis. This will help the subsequent classification process, as by this the unnecessary information will be removed from the input. The output of the DWT is the input of this process it will remove the unneeded information from the input. Suppose that the input is Z is in the form of n-dimensional vector We set the input z to be a n-dimension random vector Whose means are zero and x = { x1 , x2 , ..., x,}is a transform matrix whose dimension is n X m (m < n) [10]. y = W T z is the random vector after performing the linear transformation. Thus y is termed as first m principal components of the random vector z, if

$$x_1 = \arg\max_{v_i}\{E(v_i{}^T z)^2\} \quad (3)$$

And the n-dimension vector V, satisfies

$$v_i{}^T v_f = 0$$
$$\forall_j \neq i$$

And $$v_i^T v_f = 1$$

i=1,2, ...,m x, is the i th principal direction of the random vector z and here operator E denotes expectation. Traditionally, transform matrix x got through the Eigen value decomposition operation to the covariance matrix of input random vector z[13]. Recently, it has been found that principal directions defined by (1) and (2) has several equal definitions[12], minimum criteria for signal reconstruction is one used frequently. Set a n-dimension random vector U = xy to be reconstruction of z, then e = z - U is the reconstruction error. The cost function is defined as function

$$j_1(x) = E\left\|e^2\right\| = e\left\|z - u^2\right\| \quad (4)$$

For practical samples set, the estimation of the cost function will be

$$j_1(x) = \frac{1}{N}\sum_{i=1}^{N}\left\|x_1 - XX^T Z_i\right\|^2 \qquad (5)$$

Support Vector Machine classification

It is a statistically robust method based on risk minimization [4]. In this the classifier is trained by finding an optimal separating hyper plane it maximizes the margin between two classes of data in the induced feature space. Without losing the generality, suppose that there are m instances of training data[11]. Each of the instances consists of an $(x_{i,} y_i)$ here each of $x_{i,} \in IR^N$ it is a vector in the it has the attribute of ith instance $y_i \in \{+1, -1\}$ it is the class label for the instance The objective of the using SVM is to find optimal separating hyper plane W. X+b=0 between two classes of data. It issued to classify the data which is available is two classes .for classifying the testing instance X, decision function is

$$f(x) = W.X + b \qquad (6)$$

Corresponding classifier is sgn(f(x)). By using the SVM we find the optimal separating hyper plane by solving the following quadratic programming optimization problem

$$\arg\min_{w,b\xi 2} \frac{1}{2}\left\|w\right\|^2 + c\sum_{i=1}^{m}\xi_i \qquad (7)$$

Subject to $$y_i(W.X_i + b) \geq 1 - \xi_{i,}\xi_{i,}$$ for i = 1,. . .m;

In objective function minimizing $\frac{1}{2}\left\|w^2\right\|$ is for maximizing the margin two classes of data (w.x+b=1and w.b+x=-1) these two constraints are put on the positive and negative labels respectively the variables ξ_i, i=1….m are called slacks each ξ_i denotes the extent of x_i which is falling outside its corresponding region[5] . C is termed as cost parameter, it is a positive constant specified by the user during process. It denotes the penalty of slacks.

Objective function of the optimization problem is a tradeoff between maximizing the margin and minimizing the slacks. Larger the value of C means higher the penalty of slacks; it will result in less slacks but a smaller margin. The value of C is usually determined by cross validation. The optimization is usually done in dual form by applying the Lagrange multipliers and KKT-conditions [6], [7]. Solving this dual problem is equivalent to solving the primal problem [12].By replacing the dot products with kernel functions, the kernelized dual form of the SVM's optimization problem is

$$\arg\min_{\alpha} \frac{1}{2}\sum_{i.j=1}^{m}\alpha_i\alpha_j y_i y_j K(x_i x_j) - \sum_{i=1}^{m}\alpha_i$$

$$\text{Subject to} \quad \sum_{i=1}^{m}\alpha_i y_i = 0, 0 \le \alpha_i \le c \quad \text{for i=1….m;} \qquad (8)$$

Since $w = \sum_{i=1}^{m}\alpha_i y_i x_i$ in case of duality, the kernelized decision function in the dual form is

$$f(x) = \sum_{i=1}^{m}\alpha_i y_i K(x_i, x) + b$$

The bias term b used here can be calculated by using KKT-complementarities conditions [6], [7]. Optimization problem is solved using kernel trick, SVM imply maps data into a high-dimensional space and it finds an optimal separating hyper plane there. Testing is also done in kernel induced high-dimensional space by applying the kernelized decision function. The kernel induced mapping and high dimensional space are called feature mapping and feature space, respectively. The actual dot product is called linear kernel i.e. $K(X,Y) = X.Y$ In linear kernel, optimal separating hyper plane is found in the original space without feature mapping. The feature mapping of nonlinear kernel functions could be very complex and we may not even know the actual mapping.

CONCLUSION

It is very difficult for medical experts to detect brain cancer in hospital environment, as it is a critical disease. So it is essential to identify it on time and accurately. In this context most of the decisions are taken by a combination of doctors unknowingly the correctness of the decision depends on the experience and the background knowledge, and from the information they have. The brain tumors in human brain are very different in types and it requires an accurate analysis. Of tumor, it is very essential to select those tailored therapies that can maximize the probability of survival. The availability of information about the cancer in the form of signal and image makes it easy to use in the computer-based system. When we analyze the signals and images through a computer based system it is important to remove the noise and reduce the dimensionality so that the analysis can be done with only the useful part of the data. Here what the methodology we have used is to analyze the different types of MRS data and store the corresponding results in the database, and then whatever new MRS data we get can be analyzed and compared with the existing results in the database to classify the new MRS data to its category its type and status can be easily determined

REFERENCES

[1] Arizmendi, C., Vellido, A., & Romero, E. (2009). Frequency selection for the diagnostic characterization of human brain tumours. Frontiers in Artificial Intelligence and Applications, 202, 391–398.

[2] Foresee, F., & Hagan, M. (1997). Gauss-Newton approximation to Bayesian regularization. In Proceedings of the IEEE international conference on neural networks (ICNN) pp. 1930–1935.

[3] Marosi C, Hassler M, Roessler K, Reni M, Sant M, Mazza E, et al. Meningioma. Crit Rev Oncol Hematol 2008;67:153–71..

[4] Kecman V. Learning and Soft Computing, Support Vector Machines, Neural Networks, and Fuzzy Logic Models. Cambridge: MIT Press; 2001.

[5] Muller KR, Mika S, Ratsch G, Tsuda K, Scholkopf B. An introduction to kernel-based learning algorithms. IEEE Trans Neural Netw 2001;12:181–201

[6] Yi-Min H, Shu-Xin D. Weighted support vector machine for classification with uneven training class sizes. Guangzhou: Proceedings of the Fourth International Conference on Machine Learning andCybernetics; 2005; p. 4365–9.

[7] O. Chapelle, V. Vladimir, O. Bousquet, and S. Mukherjee, “Choosing multiple parameters for support vector machines,” Journal of Machine Learning Research, 2002.

[8] A. Frome, Y. Singer, and J. Malik, “Image retrieval and classification using local distance functions,” in Proceedings of Neural Inforamtion and Processing Systems, 2006.

[9] M.-E. Nilsback and A. Zisserman, “A visual vocabulary for flower classification,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2006, pp. 1447–1454.

[10] A. Berg, T. Berg, and J. Malik, “Shape matching and recognition using low distortion correspondence,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005.

[11] T. Hastie and R. Tibshirani, “Discriminative adaptive nearest neighbor classification,” IEEE transactions on Pattern Analysis and Machine Intelligence, 1996

[12] Luts et al. (2007). A combined MRI and MRSI based multiclass system for brain tumour recognition using LS-SVMs with class probabilities and feature selection. Artificial Intelligence in Medicine, 40(2), 87–102.

[13] Lisboa, P.J.G., Mehri-Dehnavi, R., “Sensitivity Methods for Variable Selection Using the MLP”, *International Workshop on Neural Networks for Identification, Control, Robotics and Signal/Image*, 1996, pp. 330-338.

[14] Lin, T.-S., Meador, J., “Statistical Feature Extraction and Selection for IC Test Pattern Analysis”, *Circuits and systems*, 1992, pp.391-394 vol 1.