

SOM Based Visualization Technique For Detection Of Cancerous Masses In Mammogram

S.Pitchumani Angayarkanni M.C.A,M.Phil,(Ph.d)

Assistant Professor
Department of Computer Science
Lady Doak College
Madurai

Dr. V.Saravanan

H.O.D,Department of M.C.A
Karunya Deemed University
Coimbatore, Tamil Nadu

Abstract— Breast cancer is the most common form of cancer in women. An intelligent computer-aided diagnosis system can be very helpful for radiologist in detecting and diagnosing micro calcifications patterns earlier and faster than typical screening programs. In this paper, we present a system based on gabor filter based enhancement technique and feature extraction techniques using texture based segmentation and SOM(Self Organization Map) which is a form of Artificial Neural Network(ANN) used to analyze the texture features extracted. SOM determines which texture feature has the ability to classify benign, malignant and normal cases. Watershed segmentation technique is used to classify cancerous region from the non cancerous region. We have investigated and analyzed a number of feature extraction techniques and found that a combination of ten features, such as Cor-relation, Cluster Prominence, Energy, Entropy, Homogeneity, Difference variance, Difference Entropy, Information Measure, and Normalized are calculated. These features gives the distribution of tonality information and was found to be the best combination to distinguish a benign micro calcification pattern from one that is malignant and normal. The system was developed on a Windows platform. It is an easy to use intelligent system that gives the user options to diagnose, detect, enlarge, zoom, and measure distances of areas in digital mammograms. Further Using Linear Filtering Technique and the Texture Features as Mask are convolved with the segmented image .The tumor is detected using the above method and using watershed segmentation, a fair segmentation is obtained The artificial neural network with unsupervised learning together with texture based approach leads to the accuracy and positive predictive value of each algorithm were used as the evaluation indicators. 121 records acquired from the breast cancer patients at the MIAS database. The results revealed that the accuracies of texture based unsupervised learning has 0.9534 (sensitivity 0.98716 and specificity 0.9582 which was detected thorough the ROC. The results showed that the gabor based unsupervised learning described in the present study was able to produce accurate results in the classification of breast cancer data and the classification rule identified was more acceptable and comprehensible.

Keywords- Image Enhancement; Gabor Filter; Texture Features; SOM; ROC.

I. INTRODUCTION

Breast cancer is one of the major causes for the increased morality cause many among women especially in developed countries. It is second most common cancer in women. The

World Health Organization's International estimated that more than 1,50,000 women worldwide die of breast cancer in year. In India , breast cancer accounts for 23% of all the female cancer death followed by cervical cancer which accounts to 17.5% in India. Early detection of cancer leads to significant improvements in conservation treatment[1]. However, recent studies have shown that the sensitivity of these systems is significantly decreased as the density of the breast increased while the specificity of the systems remained relatively constant. In this work we have developed automatic neuron genetic algorithmic approach to automatically detect the suspicious regions on digital mammograms based on asymmetries between left and right breast image.

Diagnosing cancer tissues using digital mammograms is a time consuming task even highly skilled radiologists because mammograms contain low signal to noise ratio and a complicated structural background. Therefore in digital mammogram, there is still a need to enhance imaging, where enhancement in medical imaging is the use of computers to make image clearer. This may aid interpretation by humans or computers. Mammography is one of the most promising cancer control strategies since the cause of cancer is still unknown[2].

Radiologist turn to digital mammography as an alternative diagnostic method due to the problems created by conventional screening programs. A digital mammogram is created when conventional mammogram is digitized; through the use of a specific mammogram is digitizer or a camera, so it can be processed by the computer. Image enhancement methods are applied to improve the visual appearance of the mammograms. Initially the mammogram image is read from the dataset and partial filter (Combination of Low and high Pass filter) is applied to remove the noise from the image.

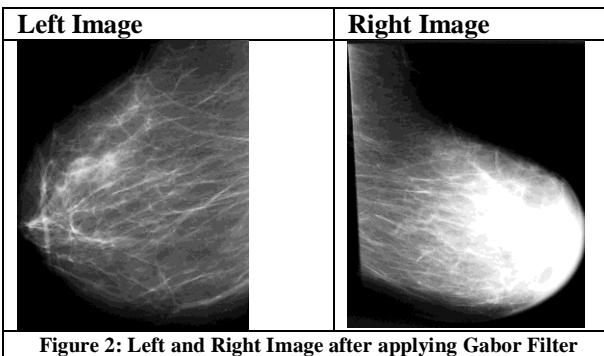
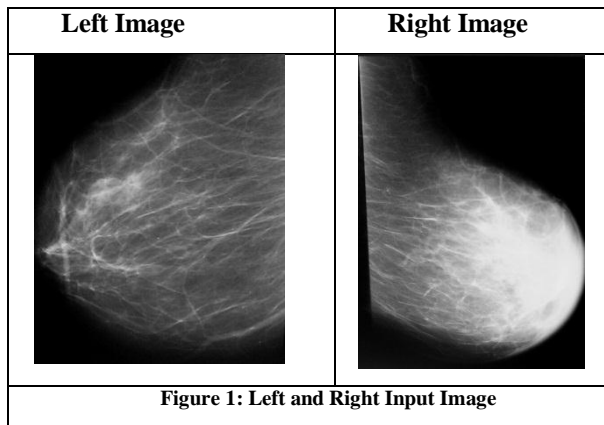
Gabor filter with texture based segmentation and SOM based clustering techniques were used in efficient detection of cancerous masses in mammogram MRI images . The selected features are fed to SOM Network hybrid to detect the texture feature estimation variable which varies highly for normal, benign and malignant tumor[2] . The watershed algorithm is used to segment the cancerous region based on the feature which plays a vital role in classification and the Receiver Operating Characteristic (ROC) analysis is performed to evaluate the performance of the feature selection methods In this paper various steps in detection of microcalcification such as i) Preprocessing and enhancement using Histogram

Equalization and Gabor filter ii) Texture based Segmentation iii) SOM based Visualization iv) Watershed segmentation V) TIC – TOC method to find how efficiency in automatic detection system.

II. IMAGE ACQUISITION

Mammograms used in this research are retrieved from the MIAS(Mammographic Image Analysis Society) database. The databases contains 161 mammogram records including normal and microcalcification cases.

The Mammography Image Analysis Society (MIAS), which is an organization of UK research groups interested in the understanding of mammograms, has produced a digital mammography database (<ftp://peipa.essex.ac.uk>)[2]. The data used in these experiments was taken from the MIAS. The database contains left and right breast images for 151 records, is used. Its quantity consists of 302 images, which belong to three types such as Normal, benign and malignant. There are



200 normal, 44 benign and 58 malignant (abnormal) images.

III. PREPROCESSING & ENHANCEMENT TECHNIQUES

One of the most important problems in image processing is denoising. Usually the procedure used for denoising, is dependent on the features of the image, aim of processing and also post-processing algorithms. Preprocessing is the first phase of image analysis. The purpose of preprocessing is to improve

the quality of the image being processed. It makes the subsequent phases of image processing easier. Median Filter is used for smoothening to retain the edge strengths.

IV. GABOR FILTER

Gabor filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filter are similar to those of human visual system, and it has been found to be particularly appropriate for texture representation and discrimination[3]. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.

Its impulse response is defined by a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function.

The filter has a real and an imaginary component representing orthogonal directions. The two components may be formed into a complex number or used individually.

Complex

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right)$$

Real

Imaginary

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \sin\left(2\pi\frac{x'}{\lambda} + \psi\right)$$

where

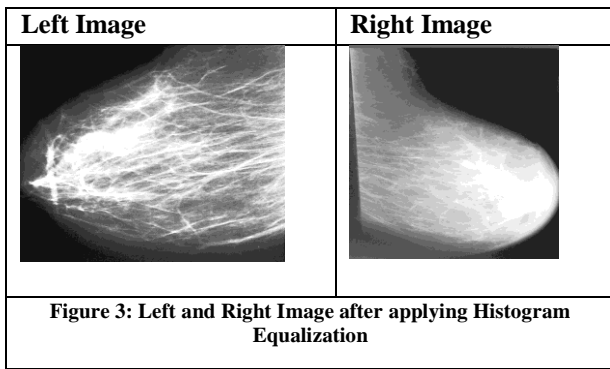
$$\begin{aligned} x' &= x \cos \theta + y \sin \theta \\ y' &= -x \sin \theta + y \cos \theta \end{aligned}$$

Gabor wavelet filters smooth the image by blocking detail information. Mass detection aims to extract the edge of the tumor from surrounding normal tissues and background. PSNR, RMS, MSE, NSD, ENL value calculated for each of 121 pairs of mammogram images clearly shows that gabor wavelet filter when applied to mammogram image leads to best Image Quality[4].

The orientation and scale can be changed in this program to extract texture information. Here 3 scales and 4 orientation was used.

Table 1: Estimated Values

PSNR	RMS	NSD	ENL	MES
87.65	2.97	4.55	89.89	8.83



V. TEXTURE BASED APPROACH

Dhawan et al has specified image structural features for classification as benign and malignant. The textural features are extracted using GLCM(gray level co-occurrence matrix) . They specify specific textural characteristics such as homogeneity, contrast, entropy,energy and regularity of the structure within the image [27,36,37,39,43,86]. The Image is divided into 3*3 pixel values and the following parameters were evaluated .Correlation, Cluster Prominence, Energy, Entropy, Homogeneity, Diff variance,Diff[4]. Entropy and Infinity measure.It was found that the region which is malignant has energy and homogeneity value=0.5 to 1, infinity measure between 0.5 – 1.An image of size 268*100 is divided into 3*3 kernel window and for a total of 289 subimages the texture values were evaluated and clustered based on the above estimated values.



A) Left Image B)Right Image

Figure 4: Texture based Segmentation

Texture Analysis analyzes the textures extracted and finds which aids in detect ability of tumor and it also helps us to find the relationship between the textures extracted. Texture Analysis can be done by using SOM (Self Organization Map)

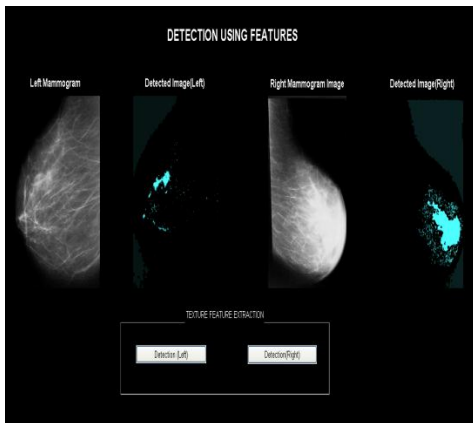


Figure 5 :Feature detection

Correlation	$\frac{\sum \sum (i_x j_x) - \mu_x \mu_y / \sigma_x \sigma_y}{\sigma_x \sigma_y}$ where μ_x, μ_y and σ_x, σ_y are the means and std deviations of p_x and p_y , the partial probability density functions
Entropy	$-\sum_{i,j} p(i,j) \log(p(i,j))$
Difference Entropy	$-\sum_{i=0}^{N_{0-i}} p_{x-y}(i) \log\{p_{x-y}(i)\}$
Difference Variance	Variance of p_{x-y}
Information Measure	$HXY - HXY1 / \max\{HX, HY\}$ Where HX and HY are entropies of p_x and p_y $HXY1 = -\sum_{i,j} p(i,j) \log\{p_x(i) p_y(j)\}$ $HXY2 = -\sum_{i,j} p_x(i) p_y(j) \log\{p_x(i) p_y(j)\}$
Energy	$\sum_{i,j} (\sum (1+(i-j)^2))$
Homogeneity	$\sum_i \left(\sum_j \frac{p(i,j)}{1 + (i-j)^2} \right)$
Normalized	$\sum_i \sum_j \frac{p(i,j)}{1 + i-j }$

VI. SELF ORGANIZING MAP

A self-organized map (SOM)[8]-type of artificial neural network that is trained using unsupervised learning to produce a two-dimensional, representation of the training samples, called a map (Kohonen Map). SOMs operate in two modes: training and mapping. Training builds the map using input examples. It is a competitive process, also called vector quantization. Mapping automatically classifies a new input vector[6].

A self-organizing map consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid.

The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to find the node with the closest weight vector to the vector taken from data space and to assign the map coordinates of this node to our vector.

Illustration of ‘Training’

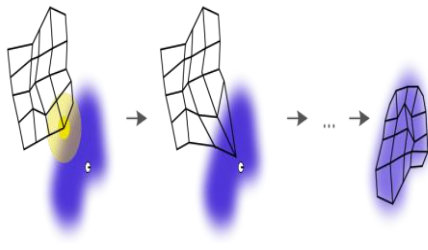


Figure 6 Training Process

The blue blob is the distribution of the training data, and the small white disc is the current training sample drawn from that distribution. At first (left) the SOM nodes are arbitrarily positioned in the data space[5]. The node nearest to the training node (highlighted in yellow) is selected, and is moved towards the training datum, as (to a lesser extent) are its neighbors on the grid. After much iteration the grid tends to approximate the data distribution (right).

Mapping

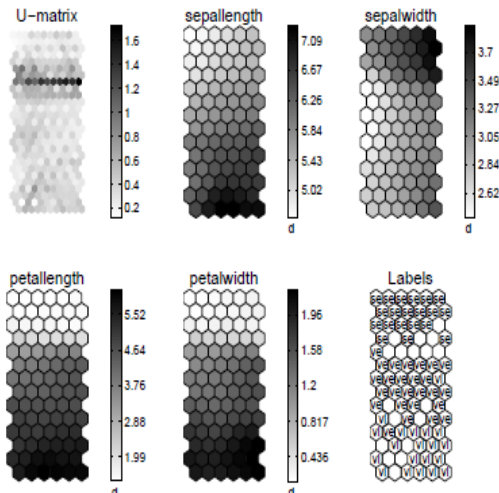


Figure 7 Mapping Visualization

The trained samples are mapped depending upon the values of the samples. Clustering is differentiated using different colors. color map can be HSV, Grayscale, RGB etc.

This approach uses SOM toolbox to perform training and mapping. This toolbox contains GUI based application which is easy to use implementation. Toolbox can be used to preprocess data, initialize and train SOMs using a range of different kinds of topologies, visualize SOMs in various ways, and analyze the properties of the SOM and data.

SOM toolbox in this approach has helped to visualize the relationship between the features and also how the feature varies for different types of cases like (Benign, Malignant and Normal).

From this project it is found that ‘Information Measure related to Cor-relation’ varies for the above specified cases during Mapping and it is also found that Energy and Entropy are oppositely cor-related.

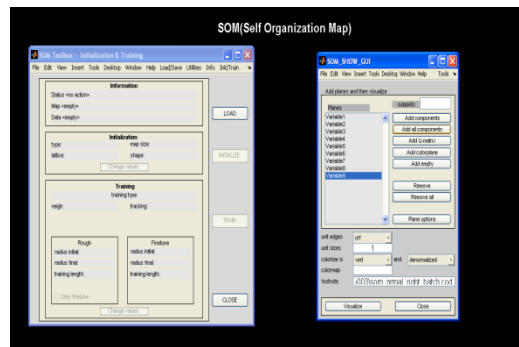


Figure 8: SOM Toolbox

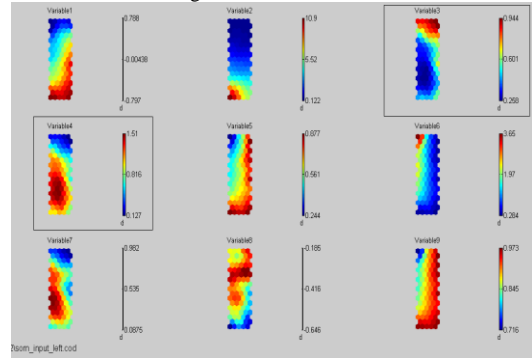


Figure 9: Visualization of texture parameter analysis using SOM Analysis:

- In case of **Normal**, there is no difference in the clustering and we can find that ‘Energy’ and ‘Entropy’ are oppositely related
- In case of **Benign**, there is a difference in the clustering of ‘Information Measure related to correlation’ and we can also find that ‘Energy’ and ‘Entropy’ are oppositely related.
- In case of **Malignant**, there is a difference in the clustering of ‘Information Measure related to correlation’ and we can also find that ‘Energy’ and ‘Entropy’ are oppositely related

VII. WATERSHED SEGMENTATION

This algorithm segments regions of an image. When tumor detected image is passed to the watershed transform, tumor region is shown in a different color. Watershed Transform is usually well suited for bubble or metallographic images but when combined with Linear Filtering Methods, it has been proved that it can segment regions of any shape.

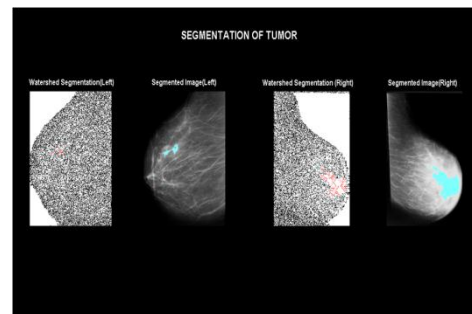


Figure 10: Segmentation of Malignant tumor using watershed transform

VIII. RESULTS AND DISCUSSIONS

From this project, one can analyze that

1. Using SOM, depending on the type of clustering, one can find that ‘Energy’ and ‘Entropy’ are oppositely related.’ Information Measure related to cor-relation’ differs for benign and malignant cases during Mapping.
2. Watershed Transform was usually well suited for bubble or metallographic images but when combined with Linear Filtering Methods, it has been proved in this project that it can segment regions of any shape[7,8].
3. The Time Taken(CPU Time) by the above technique was calculated using Tic Toc method and it was found that it take 3’50’’ms.The efficiency of the technique was compared with other methods and was found that

Table 2 TIC-TOC time estimation

Computation Time for Different Methods		
Methods	Author and References	Computational Time
Morphological Analysis	Wan Mimi Diyana, Julie Larcher, Rosli Besar	3’20’’
Filtering Technique	Proposed Approach	3’50’’
Fractal Dimension Analysis	Wan Mimi Diyana, Julie Larcher, Rosli Besar	7’20’’
Complete HOS Test	Wan Mimi Diyana, Julie Larcher, Rosli Besar	9’20’’

Though the Time Taken is slower than ‘Morphological Analysis’ Method (3’20’’) but the efficiency is better than the other methods.

IX. CONCLUSION

One of the major causes of death among women is due to breast cancer. so early diagnosis through regular screening and timely treatment has been shown to prevent cancer. This paper focuses on the approach to identify the presence of breast cancer mass in mammograms[9,10]. The proposed work utilizes Filtering Techniques for detection and watershed segmentation for segmentation.

Preprocessing with Gabor filter and using the filtering technique with watershed transform is a new approach, using this we have successfully detected the breast cancer masses in mammograms. The results indicate that this system can facilitate the doctor to detect tumor in the early stage of diagnosis process with a fraction of second.

Table 3: Comparison of our method with other methods:

Authors and References	Methods	Detection Rate
------------------------	---------	----------------

Ferrari and Rengayyan[19]	Directional Filtering with Gabor Wavelets	74.4%
Lau and Bischof[20]	Asymmetry measures	85%
Sallam and Bowyer[21]	Unwrapping Technique	86.6%
Thangavel and Karnan[22]	Metaheuristic approach	94.8%
	Redundant Wavelets	84.79%
	Wavelets Packets	87.68%

Proposed Method:

Method	Detection Rate
Histogram based Gabor Wavelet approach	96.67%
Texture based approach	97.78%
SOM Based Clustering	99.78%
Overall Accuracy of Combined approach	98.07%

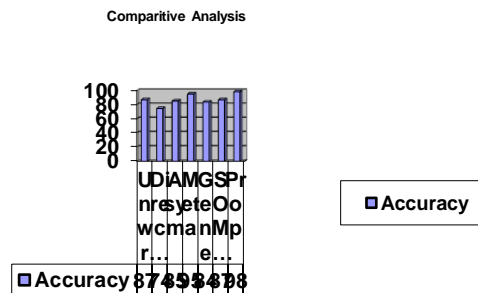


Figure 11: Comparative Study

REFERENCES

- [1] Bosch. A.; Munoz, X.; Oliver.A.; Marti. J., Modeling and Classifying Breast Tissue Density in Mammograms, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on Volume 2, Issue , 2006 Page(s): 1552 – 15582.
- [2] Dar-Ren Chena, Ruey-Feng Changb, Chii-Jen Chenb, Ming-Feng Hob, Shou-Jen Kuo, Shou-Tung Chena, Shin-Jer Hungc, Woo Kyung Moond,Classification of breast ultrasound images using fractal feature, ClinicalImage, Volume 29, Issue4, Pages 234-245.
- [3] Suri, J.S., Rangayyan, R.M.: Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer. 1st edn. SPIE (2006)
- [4] Hoos, A., Cordon-Cardo, C.: Tissue microarray pro.ling of cancer specimens and cell lines: Opportunities and limitations. Mod. Pathol. 81(10), 1331–1338 (2001)
- [5] Lekadir, K., Elson, D.S., Requejo-Isidro, J., Dunsby, C., McGinty, J., Galletly, N., Stamp, G., French, P.M., Yang, G.Z.: Tissue characterization using dimensionality reduction and uorescence

- imaging. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 586–593. Springer, Heidelberg (2006).
- [6] Dr. William H. Wolberg, (no date), Breast Cancer Wisconsin Dataset (online) (<http://www.radwin.org/michael/projects/learning/about-breast-cancer-wisconsin.html>) (1July 2005)
- [7] J.A.M. van Dijck, L.M. Verbeek, J.H.C.L. Hendriks, The current detectability of breast cancer in a mammographic screening program, *Cancer* 72 (1933) 1938–1993.
- [8] I.W. Hutt, S.M. Astley, C.R.M. Boggis, Prompting as an aid to diagnosis in mammography, in: A.G. Gale, S.M. Astley, D.R. Dance, A.Y. Cairns (Eds.), *Digital Mammography*, Elsevier, Amsterdam, 1994, pp.389–398.
- [9] H. Li, Y. Wang, K.J. Ray Liu, Computerized radiographic mass detection. Part II: decision support by featured database visualization and modular neural networks, *IEEE. Trans. Med. Imag.* 20 (4) (2001) 302–313.
- [10] Pfarl G. Breast Imaging Reporting and Data System. 2006; Available from: <http://www.birads.at>, Access at: 2006Oct24.
- [11] Otsu N. A threshold selection method from gray-level histogram. *IEEE Transactions on System Man Cybernetics* 1979; 9(1): 62-66.
- [12] A. R. Abdel-Dayem, M.R. El-Sakka, Fuzzy Entropy Based Detection of Suspicious masses in Digital Mammogram Images. 27th Annual Conference of IEEE, China 2005, September 1-4.
- [13] L. Li, W. Qian, L.P. Clarke, Image feature extraction for mass detection using digital mammography: effects of wavelet analysis, *Proc. SPIE Med. Imag.* 3338 (1998) 168–1176 *Med. Phys.* 26(3) (1999).
- [14] L. Li, W. Qian, L.P. Clarke, R.A. Clark, Improving mass detection by adaptive and multi-scale processing in digitized mammograms, *SPIE (Society of Photographic Scientists and Engineer) Conference on Image Processing*, San Diego, CA, vol. 3661, February 1999, pp. 490–498.
- [15] Silvano DZ, Luigi C, Stefano L. Image thresholding using fuzzy entropies. *IEEE Trans. on systems man and cybernetics, Part B* 1998; 28(1):15-23.
- [16] B. S. Manjunath, W. Y. Ma, Texture features for browsing and retrieval of image data, *IEEE Trans. Pattern Anal. & Machine Intell.*, vol.18, no. 8, pp.837–842, 1996.
- [17] Three-Dimensional Technique for Automatic Brain Segmentation of the Ventricles Based on Optimal Histogram Thresholds of MRI, Danmary Sanchez, Malek Adjouadi, Byron Bernal, Nolan Altman, *WSEAS TRANSACTIONS on COMPUTERS*, Issue 7, Volume 4, July 2005, ISSN 1109-2750.
- [18] Comparison of Three Gaussian Mixture Modeling and Spatial Encoding Methods for Segmenting Human brain MRI, Mahmood Zeydabadi-Nejad, Reza A. Zoroofi, Hamid Soltanian- Zadeh, *WSEAS TRANSACTIONS on ELECTRONICS*, Issue 3, =Volume 1, July 2004, ISSN 1109-9445.
- [19] Fuzzy Rule Based Classifiers from Support Vector Learning, Stergios Papadimitriou, Konstantinos Terzidis, Seferina Mavroudi, Lambros Skarlas , Spiridon Likothanasis, Issue 7, Volume 4, July 2005, ISSN 1109-2750.

AUTHORS PROFILE

S.Pitchumani Angayarkanni M.C.A.,M.Phil.,(Ph.D) working in the department of computer science,Lady Doak College, Madurai for the past ten years. My areas of interest are medical image processing, Neural Network & Data Mining.