Asian Journal of Computer Science and Technology (AJCST)
Vol.2.No.1 2014 pp 58-65
available at: www.goniv.com
Paper Received :05-03-2014
Paper Published:28-03-2014
Paper Reviewed by: 1. John Arhter 2. Hendry Goyal
Editor : Prof. P.Muthukumar

CLASSIFICATION OF COMPRESSED MAMMOGRAM IMAGE USING IMPROVED WATERSHED TRANSFORM

D. Kavitha,
S.K.P Engineering College
kavitha33it@gmail.com

ABSTRACT

A mammogram image segmentation and compression technique is proposed for classifying and storing information about breast cancer tissue. Initially a preprocessing is done in the mammogram images with Contrast limited adaptive histogram equalization (CLAHE). The features are extracted from the images. Then improved watershed transform is applied to the images for segmentation. Genetic training is applied to the images. Neural network classifier (genetic algorithm) is used for classification of breast cancer tissue in normal, benign, and malignant. Derivative based feature saliency technique is applied to the images for feature extraction. vector quantization (VQ) with Artificial neural network is used in the three regions with different compression rate based on the importance of the region and by this we also retain important tumor characteristics and minimize the size of mammogram images and thus reducing the storage space and increase its efficiency. At least the images undergo morphological process and micro classification area detection. This is specially done for benign and malignant regions. Two types of compression namely lossy and lossless compression are done in the segments based on the importance of those regions. Results show that this method gives accurate results in diagnosis of breast cancer and gives efficient result in the compression applications. The proposed method gives accurate results in differentiating malignant and benign tumor. It reduces the storage space when compared with other existing compression mechanisms. Thus the proposed system is more accurate in detecting tumor tissue and efficient in storing the images.

Keywords: Breast cancer, the Competitive Hopfield neural network (CHNN), Contrast limited adaptive histogram equalization (CLAHE), Derivative based feature saliency technique, neural network classifier (genetic algorithm), Preprocessing, vector quantization (VQ).

1. INTRODUCTION

Breast cancer is a type cancer that occurs in the milk ducts of our breast. Breast cancer usually occurs in the inner layers of milk ducts or the labels that secrete milk. A malignant tumor is the type of tumor spreads to other parts of the body. The vast majority of breast cancer cases occurs in females. Breast cancer is the most common occurring cancer type especially in women. It accounts for 16.3% of all female cancers which is of invasive type in women. About 18.25% of deaths due to cancer in the whole world for both men and women are from breast cancer. This type of cancer is very high in developing nations such as U.S and U.K compared to developing ones such as India and China. There are several reasons for this breast cancer, including life span of a person is one of the key factors. Breast cancer is now more common in women who are above the age of 30.Women in the rich countries have a life span longer than women in the poorest countries. The different lifestyles and food habit of females in rich and poor countries are also taken as factors for breast cancer by experts.

A mammogram is obtained by projecting low energy x-ray through the human breast cancer tissue. It is done by applying low energy X-ray through the breast. The mammograms look black and white. In the human breast, most of the fibrous tissue is transformed into fatty tissue. Generally mammogram images appear black or gray .General segmentation is the process of dividing the image into disjointed segments so as to the characteristics of each region are homogeneous [6]. A large variety of image segmentation methods have been presented [4, 5, 8, 11, and 13]. Many techniques in watersheds have been presented in recent decades [9, 15]. The main disadvantage of watershed algorithm is that it produces over-segmented results. That is, when the watershed obtains catchment depressions from the slope of a particular image segment, so the results of watershed contain too many small regions [2, 7] which require a large space to store it. Moreover, it shows sensitivity to noise; local changes in the images can significantly change the results [2, 7]. It is due to the usage of high pass filter. In addition, it shows poor detection of important regions or segments with low variable ratio. Accordingly, the improved version of the watershed algorithm may be able overcome the intrinsic problems. Various kinds of preprocessing have been developed to solve the problems of over-segmentation, such as median filter and anisotropic diffusion filters [17] .Here we use the Contrast limited adaptive histogram [19, 20, and 21] for preprocessing.

Moreover it's important for medical images to have compression including lossy and lossless to preserve information on various regions based on its importance. The Canny edge detector [3] is used to know the edge between breast and background. Then the original mammogram image is divided into the tumor, the breast without tumor, and background [7], (Improved watershed transform for tumor segmentation: Application to mammogram image compression. Wei-Yen Hsu. Expert Syst. Apple 01/2012; 39:3950-3955). Since the two regions, the breast without tumor and background, are not important, the vector quantization with neural network is applied to lossy compressed. The Hopfield neural network [7] (Hopfield & Tank, 1985) is a technique for solving optimization problems based on the Lyapunov energy function. The CHNN is constructed as a two-dimensional fully interconnected array with the rows as the training vectors and columns for the code vectors in the codebook. The tumor region is classified as benign, malignant and normal. Normally the cell in human breast and other parts of our body grow and break to form new cells which are needed. When normal cells become old or get wounded, they die, and new cells are formed in that place.

Benign is a mass of cells that don't invade its neighboring tissues. It grows slowly. They are not harmful but rarely invade the tissues around them. They won't spread to other parts of the breast .They can be destroyed and usually don't grow again. Malignant tumor in the breast can even lead to death and they can easily spread to nearer parts and tissue in the chest. This type of tumor can easily spread to its neighboring tissue or parts of our body.

2. METHODS

The Improved Watershed Transforms

The watershed transform is one of the segmentation methods for segmenting medical images such as that of brain, breast etc... It has some drawbacks such as over segmentation, sensitivity to noise, and poor detection of important areas, areas with low contrast features, and poor detection of thin structures [2, 7] .It has some advantage also such as high speed, ability to parallelize [12]. An improved watershed transforms based on intrinsic prior information [7] is used to obtain tumor outline. We first introduce important definitions of the watershed transform about the lower slope and lower neighbors. Let f be a gray image. The lower slope off at a pixel p, LS (p) is defined,

$$\boldsymbol{L\ (p) =} \quad {}_{(\ )\cup} \quad \frac{(\ )\quad(\ )}{(\ ,\ )} \qquad \mathbf{(1)}$$

Where N (p) is the set of neighbors of p, and d (p, q) is the distance between pixels p and q. To define steepest slope relation between pixels is necessary for the lower slope to compute the watershed transform of the image. The lower neighbors are directly derived from the lower slope.

$$(\ ) = \quad ' \quad (\ )\ \frac{(\ )\quad(\ )'}{{}_{,}\quad '} \qquad \mathbf{(2)}$$

In real time applications, the watershed transformation cannot be directly calculated. Thus, the lower slope is taken as

$$\boldsymbol{LS\ (p)\ =} \quad {}_{(\ )\cup} \quad \frac{|\,(\ )\quad(\ )|}{(\ ,\ )} \qquad \mathit{(3)}$$

Also a set of lower functions is used, it's given by

$$\boldsymbol{LS_e\ (p) =} \quad {}_{(\ )\cup} \quad \frac{(\ ,\ )}{(\ ,\ )} \qquad \mathit{(4)}$$

A function is used to measure the difference In class probability between neighboring pixels for generic image segmentation. Posterior probabilities for each class I at each pixel p is calculated using Bayes' rule as

$$P = \frac{(\ )}{\Sigma \qquad (\ )} \qquad (5)$$

Marko random fields [1] (Besag, 1986; Geman & Geman, 1984) provide a way to model local correlations between pixels. To estimate the model, iterative conditional modes [1] are used to iteratively
Solve the equation.

Contrast Limited Adaptive Histogram Equalization (CLAHE)

Normal histogram equalization makes use of the transformation is obtained from the histogram transformation of pixels. So this method works well when the arrangement of pixels is uniform throughout the image. When the image contains pixels or regions that are blurred when compared to other image segments, those regions will not be enhanced . Adaptive histogram equalization (AHE) improves the normal histogram equalization by transforming each pixel with transformation function obtained from a nearby region [19, 20]. It was initially developed for aircraft display function. Usually each pixel is transformed based on histogram of all the square values around it. The derivation of transforming a function from histograms is same for ordinary histogram equalization.

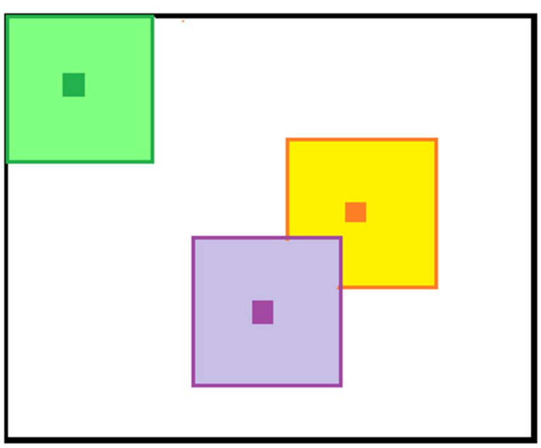


Figure.1. Cumulative Distribution Function (CDF) of pixels in mammogram images.

Pixel values near the tumor boundary have to be treated in a special way as some of the pixels in the neighborhood wouldn't lie entirely within the image. This problem can be solved by increasing the size of the image segments by duplicating lines and columns of pixel values in order to image boundary. Simply duplicating the pixel lines on the tumor border will not be the right decision to take as it will result in highly peaked neighborhood histogram images. The size of the neighborhood will vary based on the image segments.
Due to the characteristics of histogram equalization the resulting value of pixel with respect to AHE will vary based on the pixels in its neighborhood. It is the better way to implementing on hardware that compares the given central pixel value with pixels in its neighborhood. An abnormal result can be calculated by adding a value of 2 for each pixel with a value less than the value of central pixel and adding 1 for all the pixels with the same value.
When the image having a pixel's neighborhood is of the same type, its histogram will have strong peak value. This transform function will plot pixel values irregularly with respect to the whole range of a resulting image.This makes the AHE to over amplify the small values [19] of noise in a homogeneous region of images.
Contrast Limited Adaptive Histogram Equalization (CLAHE) [19, 20, and 21] varies from normal Adaptive Histogram Equalization (AHE) by its contrast limiting value. Contrast Limited Histogram Equalization (CLHE) is obtained by applying this feature to normal histogram equalization. But CLHE is rarely in practice.
CLAHE was developed for the purpose of reducing the over amplification of noise which is obtained due to AHE. This is done by reducing the contrast value of AHE.CLAHE reduces the noise before computing the cumulative distribution function (CDF) [19, 21]. This reduces the slope of both cumulative distribution functions (CDF) and the transformation function. The clip limit of the histogram is proportional to normalization of histogram equalization and also on the neighborhood region size.
The value where the histogram is clipped is called clip-limit. It depends on size of neighborhood segments and histogram normalization. Common values reduce the resulting amplification to approximately 3.5 histogram mean values. So this region of histograms that goes beyond the original clip value limit is not discarded and the values are redistributed equally to all histogram bins.
The redistribution process will make the histogram bins to have values that exceed the clip limit depending on the image segments. If this process is not required. The redistributed process is done recursively until the excess values are negligible.

Vector Quantization

Vector quantization is one of the effective techniques for digital image compression. It is similar to k-means and clustering algorithms. It divides the

image segments based on the density.The Hopfield neural network which is a simple method in many fields [10, 16] (Hsu & Sun, 2009; Wang& Zhou, 2009). It consists of a single layer of processing elements where each neuron is connected to every other neuron in the network. The Hopfield neural network is a well-known technique used f0or solving optimization problems based on the Lyapuno energy function. In this study, we use a discrete artificial neural network, where the winner-takes-all method is adopted to learn weighting factors in the energy function, for vector quantization. That is, the CHNN [14, 16] is used to quantize the vectors from segmented regions, and compress and reconstruct the regions. Suppose an image is divided into n blocks (vectors of pixels) and each block occupies $\ell\times\ell$ pixels. A vector quantize maps the Euclidean into the $\ell\times\ell$ dimensional space R into a Set
{ $_j$, j=1, 2,…., c } of points in $\mathbf{R}^{\ell\times\ell}$, called a codebook. The simplified object function of the competitive Hopfield neural network (CHNN) based on cumulative learning can be modified as

Hopfield Neural Network Classifier

A **Hopfield network(***"Hopfield Neural network classifier"*, Available at:<http://en.wikipedia.org/wiki/Hopfield_network>**)**is a form of recurrent artificial neural network invented by John Hopfield. Hopfield nets [7] serve as content-addressable memory systems with binary threshold nodes. They are guaranteed to converge to a local minimum, but convergence in a false data rather than the stored data can occur.

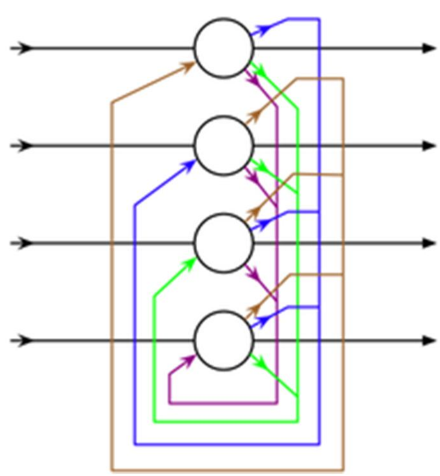


Figure.2. Hopfield Neural network classifier

"Hopfield Neural network classifier", Availableat:<http://en.wikipedia.org/wiki/Hopfield_network>

Derivative-Based Feature Saliency Techniques

Based on the values of the features of mammogram images, the breast tissue is classified into four basic categories like fatty, uncompressed fatty, dense and high density.

Derivative-based feature saliency techniques [18] were used to define the best of 25 Laws texture features for the classification of 101 malignant mass and benign mass regions. Statisticians and derivative-based saliency techniques were used to select the best size, shape, contrast, and Laws texture features of the mass model. Many features were chosen to explain the model, in that four have been used. Using this model, the regions were classified using a multilayered perception with false-positive rates of 1.9/image. It also classifies the malignant tissues from the benign tissues. It gives excellent results when compared to other techniques. In this proposed method the input image is first preprocessed and the tumor segmentation is done then feature extraction is done to obtain the feature data set of mammogram images.

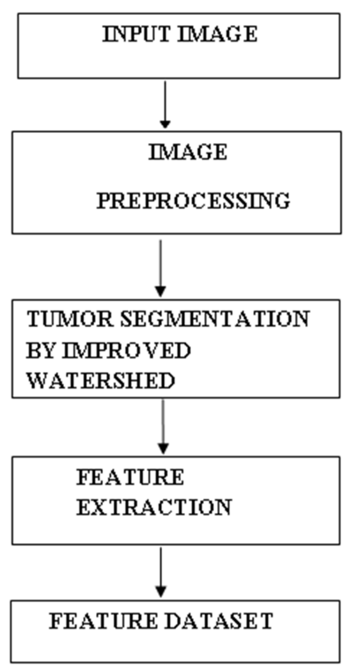


Figure.3. Feature Extraction on mammogram images

3. RESULT

Image Preprocessing

Image preprocessing is done to remove noisy and dirty data from a data set. The main aim of pre-processing is removing noise, surplus, duplicate data from a data set. It is mainly used for maintaining

the quality of images. In mammograms it is essential to maintain the quality of the images in detecting the tumor. So it is necessary to preprocess the input mammogram images. Preprocessing is done in mammogram images to remove the dirty pixels and high frequency data.
In this study CLAHE method is used to avoid the over amplification of noise in the mammogram images before segmentation.

Feature Extraction

After preprocessing of the mammogram images , feature extraction is done .The texture features, statistical features and structural features have been calculated accurately for the segmented tumour from the given input mammogram image. All the values of the features extracted are accurate. Then the features are extracted from mammogram image segments. A feature data set has been obtained for further segmentation. Feature extraction is done for differentiating malignant and benign tissues and is seen to be more accurate.

Tumor Segmentation

The mammogram images are segmented using improved watershed transform with the use of different prior information based functions for different monograms, instead of the usual calculation. To evaluate the performance of the proposed segmentation method, the standard Jaccard similarity indexes [14] (Shattuck, Sandor-Leahy,
Schaper, Rottenberg, & Leahy, 2001) are adopted. The index is calculated comparing the region I segmented by improved watershed transform with region E outlined by the experts.

3.4 CLASSIFICATION

Classification is done after extracting the features of the tumor in the mammogram images of the breast, and then the dataset is constructed in the proper format. ANNs have proven suitable for satisfactory diagnosis of various diseases. In addition, their use makes the diagnosis more reliable and therefore increases patient satisfaction. However, despite their wide application in modern diagnosis, they must be considered only as a tool to facilitate the final decision of a clinician, who is ultimately responsible for critical evaluation of the ANN output. Methods of summarizing and elaborating on informative and intelligent data are continuously improving and can contribute greatly to effective, precise, and swift medical diagnosis.

3.5 IMAGE COMPRESSION

After mammogram images are segmented into the tumor, the breast without the tumor, and the background, the vector quantization
With CHNN is applied to all the separated parts with different compression rate according to their importance by simultaneously reserving important region information and reducing the size of mammogram images for more efficient storage or transmission.

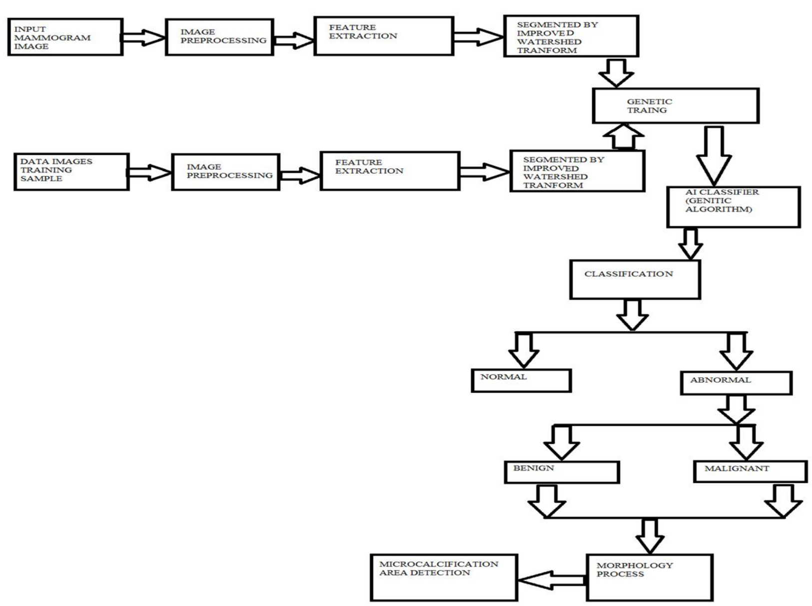


Figure.4. The proposed system diagram

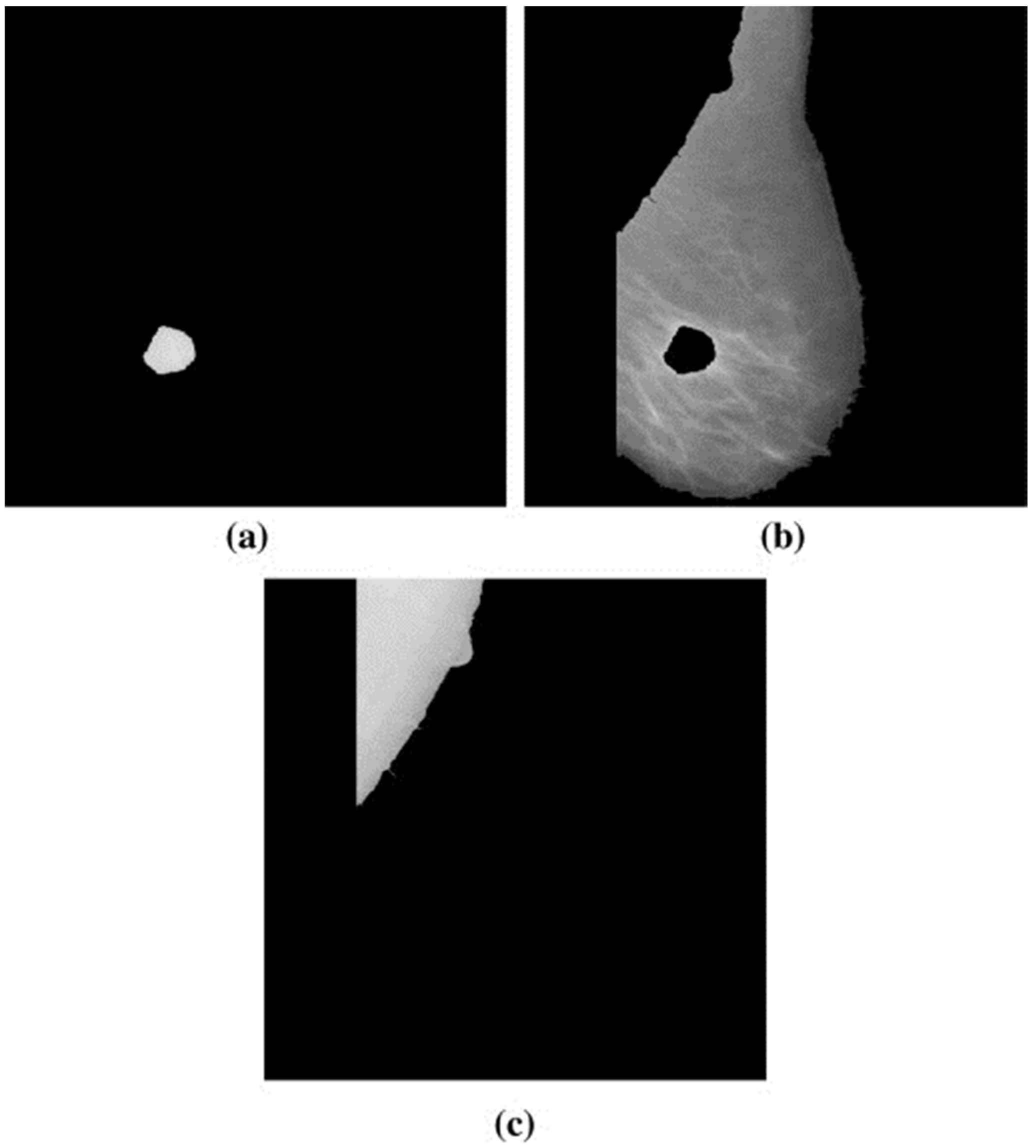
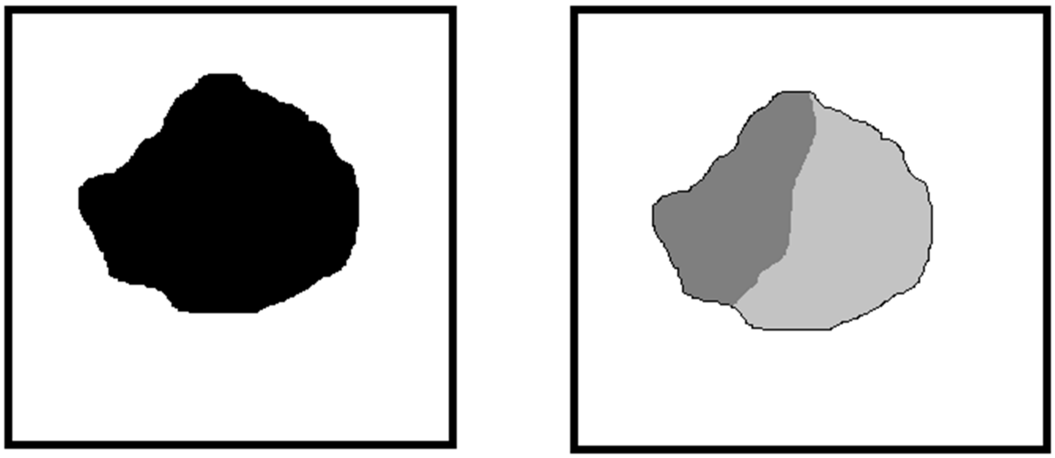


Figure.5. Segmented results. (a) The tumour. (b) The breast without the tumour. (c) The background



(a) **(b)**

Figure.6. Classification result (a) The tumour (b) Benign and Malignant tumour

4. CONCLUSION

We have proposed region segmentation, feature extraction, classification, and compression method in this study for breast cancer diagnosis. The mammogram images are segmented into tumour tissue, breast without tumour and the background of the breast automatically by improved watershed transform using prior information. The individual segmented regions are then applied to loss or lossless compression for minimizing the storage space. The experimental results show that it is an effective method for tumour segmentation while comparing to those outlined by an expert. After segmentation, original mammogram images are separated into tumour, breast without tumour. The vector Quantization with CHNN is applied to all the separated parts with different compression rate according their importance to simultaneously reserve important details and reduces the size of the storage or transmission. The experiment results also show the presented method can reconstruct very well in the applications of mammogram image compression. We then apply classification techniques for that tumour region. Then the tumour is classified into benign and malignant region. The experimental results show that this presented methodology is observed as more accurate in detecting the benign and malignant tumour regions.

REFERENCES

[1]. Besag, (1986), *"On the statistical analysis of dirty pictures"*, Journal of the Royal Statistical Society B, 48, 259–302.

[2]. Beucher, S., & Meyer, F. (1993*)," The morphological approach to segmentation:*

[3]. *The watershed transforms"*, In E. R. Dougherty (Ed.). Mathematical morphology in image processing (Vol. 12, pp. 433–481)", New York: Marcel Dekker.

[4]. Canny, J, 1986,"*A computational approach to edge detection"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 8, 679–698.

[5]. Dokur, Z, 2008,"*A unified framework for image compression and segmentation by using an incremental neural network*", Expert Systems with Applications, 34 (1), 611–619.

[6]. Dokur, Z., &Ölmez, T, 2008,"*Tissue segmentation in ultrasound images by using Geneticalgorithms."*, Expert Systems with Applications, 34 (4), 2739–2746.

[7]. Geman, S., & Geman. D, 1984,"*Stochastic relaxation, Gibbs distributions, and the*
[8]. *Bayesian restoration of images*", IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721–741.

[9]. Grau, V., Mewes, A. U. J., Alcañiz, M., Kikinis, R., & Warfield, S. K., 2004,"*Improvedwatershed transform for medical image segmentation using prior information"*, IEEE Transactions on Medical Imaging, 23 (4), 447–458.

[10]. Harlick, R. M., & Shapiro, L. G, 1985,"*Image segmentation techniques"*, CVGIP, 29,100–132.

[11]. Hopfield, J. J., & Tank, D. W., 1985,"*Neural computation of decisions in optimization*
[12]. *Problems*", Biological Cybernetics, 52, 141–152.

[13]. Hsu, W. Y., 2010,"*EEG-based motor imagery classification using neural-fuzzy predictionand wavelet fractal features"*, Journal of Neuroscience Methods, 189 (2), 295–302.

[14]. Hsu, W. Y., Lin, C. C., Ju, M. S., & Sun, Y. N. , 2007," *Wavelet-based fractal featureswith active segment selection: Application to single-trial EEG data"*, Journal ofNeuroscience Methods, 163 (1), 145–160.

[15]. Hsu, W. Y., & Sun, 2009,*"EEG-based motor imagery analysis using weighted wavelet transform features"*, Journal of Neuroscience Methods, 167 (2), 310–318.

[16]. Lai, C. C., & Chang, C.Y, 2009," *A hierarchical evolutionary algorithm for automatic medical image segmentation"*,Expert Systems with Applications, 36 (1), 248–259.

[17]. Moga, A. N., & Gabbouj, M., 1997," *Parallel image component labelling withwatershed transformation"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 19, 441–450.

[18]. Pal, N. R., & Pal, S. K, 1993,*"A review of image segmentation techniques"*, Pattern Recognition, 26, 1277–1294.

[19]. Shattuck, D. W., Sandor-Leahy, S. R., Schaper, L. A., Rottenberg, D. A., & Leahy, R. M., 2001, *" Magnetic resonance image tissue classification using a partial volume model"*, Neuro Image, 13, 856–876.

[20]. Vincent, L., & Soille, P, 1991," *Watersheds in digital spaces: An efficient algorithm based on immersion simulations"*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13 (6), 583–593.

[21]. Wang, J., & Zhou, Y. , 2009,"*Stochastic optimal competitive Hopfield network for partitional clustering* "Expert Systems with Applications, 36 (2), 2072–2080.

[22]. Weickert, J. , 1998,"*Anisotropic diffusion in image processing"*, Stuttgart, Germany, Teubner-Verlag.

[23]. William E. Polakowski ; Steven K. Rogers ; Dennis W. Ruck ; Richard A. Raines; Jeffrey W. Hoffmeister,1996,*"Derivative-based feature saliency for computer-aided breast cancer detection and diagnosis", Proc.SPIE* 2760, A, applications and Science of Artificial Neural Networks II, 312 (March 22, 1996); doi: 10.1117/12.235922.

[24]. Pizer, S.M. ; North Carolina Univ., Chapel Hill, NC, USA; Johnston, R.E. ; Ericksen, J.P. ,1990,
[25]. "*Contrast-limited adaptive histogram equalization: speed and effectiveness*", Visualization in biomedical computing.

[26]. Mohan, Shelda ; Mahesh, T.R.,2013," *Particle Swarm Optimization based Contrast Limited enhancement for mammogram images"*, Intelligent Systems and Control (ISCO), 2013 IEEE , Conference , 10.1109/ISCO.2013.6481185.

[27]. Josephus, C.S. ; Dept. of Comput. Sci., Univ. of Kerala, Trivandrum, India ; Remya, S.,2011," *Multilayered Contrast Limited Adaptive Histogram Equalization Using Frost Filter"*, Recent Advances in Intelligent Computational Systems (RAICS), 2011 IEEE, 638 – 641, 978-1-4244-9478-1,10.1109/RAICS.2011.6069388.