VOL. 6, NO. 9, SEPTEMBER 2011 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

A NOVEL CAD SYSTEM FOR BREAST CANCER SEGMENTATION IN SONOGRAMS

Alamelumangai N.[1] and Devishree J.[2]
[1]Department of MCA, Karpagam College of Engineering, Coimbatore, India
[2]Department of EEE, Coimbatore Institute of Technology, Coimbatore, India
E-Mail: alamelumangai.n@gmail.com

ABSTRACT

Breast cancer has turned out to be the most important health concern in the world over the past years. Early diagnosis is an efficient method to detect and supervise breast cancer. Computer diagnosis system can act as a major function in the early detection of breast cancer and can decrease the death rate among women with breast cancer. This paper provides a better system which detects cancer in early stages. The proposed system involves three phases such as speckle noise reduction, image enhancement and image segmentation. To remove speckle noise, we have used Neuro-Fuzzy based Memetic algorithm. Image enhancement is performed using Hough transform. Later, the enhanced image is segmented using clustering technique called Modified Fuzzy Possibilistic C-Means technique with Repulsion factor to identify the cancer affected region. The experimental result suggests that the proposed system results in better detection of cancer regions when compared to the existing technique.

Keywords: breast cancer, ultrasound image, memetic algorithm, hough transform, modified fuzzy possibilistic C-means, repulsion.

1. INTRODUCTION

The most frequently diagnosed cancer in women aged between 40 and 60 is Breast cancer [1, 2]. Based on the report released by WHO there are about 502,000 deaths resulted due to breast cancer. On analyzing these data, breast cancer is determined to be the most deadly cancer. In recent decades researchers have been involved to determine the best method to diagnosis breast cancer. Successful healing is a means to decrease the high death rate. To effectively treat a patient with breast cancer it is required to diagnose it as soon as possible. Cancer in their earlier stages is susceptible to treatment whereas cancers in their developed stages are typically almost impractical to cure. Though breast cancer has very high occurrence and death rate, the reason for breast cancer is still un-determined. So, early detection is the primary vital process towards diagnosing breast cancer. It acts as a main function in breast cancer diagnosis and treatment.

There are some factors which are utilized by the physicians to identify whether a breast nodule is benign or malignant. This paper utilizes the computational methods for the study and classification of shapes and textures to assist the physicians for detecting the occurrence of cancer with the help of ultrasound images.

The usage of images instead of mammography [3, 4] Ultrasound (US) images possesses the following advantages:

- Breast ultrasound examinations can produce any section image of breast, and examine the breast tissues in real-time and dynamically.
- Ultrasound imaging can depict small, early-stage malignancies of dense breasts, which is complex for mammography to attain.
- Sonographic equipment is portable and relatively low cost, and has no ionizing radiation and side effects.

The proposed CAD system for detection of breast cancer from the ultrasound image involves the following phases:

- Speckle Noise Reduction
- Image Enhancement
- Segmentation

In the first phase, the obtained US image [5, 6] is initially treated to remove the speckle noise [7] with the help of Neuro-Fuzzy Memetic algorithm. In the second phase, the image is enhanced [8] using Fuzzy Set and Hough transform for better diagnosis. Finally, segmentation is performed in the enhanced image with the help of clustering technique called Modified Fuzzy Possibilistic C-Means technique with Repulsion factor [9]. Previous well known algorithm for ultrasound image segmentation [10, 11, 12, 13] is Eliminating Particle Swarm Optimization (EPSO) [14]. The segmented image obtained from proposed segmentation technique will provide the cancer affected regions clearly.

2. RELATED WORKS

Oelze *et al.*, [15] proposed a Quantitative Ultrasound Assessment of Breast Cancer Using a Multiparameter Approach. Detection and diagnosis of breast cancer in starting stage drags to better prediction. Quantitative ultrasound (QUS) methods uses a multiparameter set have been created for categorizing breast cancer. The enhancement in detection and diagnosis of breast cancer with the help of QUS will have considerable medical importance. Two types of mammary tumors, carcinoma and sarcoma, were analyzed in mice with the help of QUS imaging Ten tumors for every types of cancer were scanned with a 20-MHz single-element transducer (f/3). The formation of the tumors was also featured by a clustering factor and the uncertainty of the scattered positions by contrasting the envelope statistics of

the backscatter to a homodyned-K distribution. F-tests performed on the backscattered power spectra from the two varieties of tumors exposed statistically considerable variations for frequencies above 16 MHz. QUS images of the tumors using the ASD, AAC, beta, and S parameter approximations from the new model and the envelope statistics were constructed. High-frequency QUS uses a multiparameter characteristic set enhanced the diagnostic prospective of ultrasound for breast cancer identification.

Gefen *et al*., [16] suggested ROC analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis. Breast cancer identification by means of ultrasound tissue categorization was examined with the help of Receiver Operating Characteristic (ROC) analysis of groupings of acoustic features, patient age, and radiological results. A characteristic fusion technique was developed that performs well even if only fractional diagnostic data are offered. The ROC technique utilizes ordinal dominance theory and bootstrap re-sampling to estimate Az and confidence intervals in simple and also paired data analyses. The merged diagnostic characteristic had an Az of 0.96 with a confidence interval of [0.93, 0.99] at a implication level of 0.05. The combined features indicate statistically considerable enhancement over prebiopsy radiological findings. These outcomes represent that ultrasound tissue characterization, in combination with patient record and clinical findings, may significantly decrease the requirement to carry out biopsies of benign breast lesions.

Winder *et al*., [17] suggested Synthetic Structural Imaging (SSI): A new ultrasound method for tracking breast cancer morphology. A novel signal interrogation concept according to Synthetic Structural Imaging (SSI) physics has been designed to guide therapies. The SSI technique was formerly triumphant in radar and sonar imaging; here it is established for acoustic scattering from penetrable biological targets. Operating at ultrasonic frequencies of several hundred kilohertz, SSI trades the higher resolution of typical B-mode ultrasound imaging for a significantly stronger correlation to target shape and volume, which are among the primary tissue classifiers.

3. METHODOLOGY

Phase 1: Speckle noise reduction

The problem with the existing Neuro-Fuzzy or Fuzzy-Neuro methods is that it fails to determine the number of rules or the membership functions (msf) [18]. While integrating evolutionary approach into these systems it optimizes the structure and parameters of the fuzzy rules. This paper discuss about optimizing the parameters.

Structure of the system

In the proposed system, the capabilities of the Evolutionary Algorithm are added to the NFS: Memetic Algorithm. Memetic algorithm is used to optimize and set the neuro-fuzzy parameters. This acts as a filter to despeckle the ultrasound image. The filter which is suggested is a self organized adaptive neuro-fuzzy filter which is based on neuro-fuzzy and memetic learning. The system uses the neural network ability to learn and knowledge in the system is in fuzzy form. The network used is a feed-forward 5 layered network Figure-1 [19], where the first layer executes a fuzzification process, the second layer executes the fuzzy AND of the antecedent part of the fuzzy rules, the third layer normalizes the membership functions (MFs), the fourth layer executes the consequent part of the fuzzy rules, and finally the last layer computes the output of fuzzy system by summing up the outputs of fourth layer [20]. The noise is detected based on local statistical features and uses the fuzzy knowledge. Here, the memetic algorithm is used to decide and optimize the parameters to the network [21].

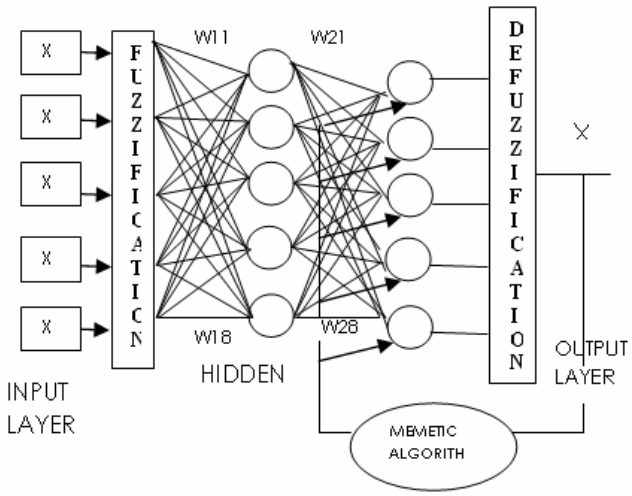


Figure-1. Structure of the neuro-fuzzy MA system.

The basic Sugeno type fuzzy model is used which is a single network to filter the speckle noises [18]. The input parameters are the fuzzy values based on the difference between the main pixel and its neighboring window pixels. The window size is number of input layers nodes in the system. In this 5×5 window sizes is used and hence there are five input nodes. Each node of the input layer is coupled with its neighboring window pixel and therefore the data used in this layer are fuzzy data. The hidden layers in the system provide knowledge to the system based on the fuzzy rules and their implications [22]. The inferences to the system are based on the fuzzy IF-THEN rules, which involve the parameters of the system. The weights are added to the network between the input layers and the hidden layers which are binary values. To improve the efficiency of the encoding system a set of five binary weights which identify the pattern of pixels is used. With this encoding mechanism a five bit substring is evolved. These 5-bit substrings result in three patterns of 90, 180 and 270 degree rotations. Optimization is done to the non-zero elements in weight-sets which identify a pattern in the neighboring window [23]. Binary weights in the genetic string are optimized in training steps. Estimation of the noise amplitude in the neighboring

patterns is applied the same manner as applying the local statistics [24].

Optimized parameter learning

MA is chosen, as it is a class of algorithm for maximization of functions [8]. It exploits the features of the error functions and does not rely on the parameter space. MA applies the mechanism of natural selection and genetics to its population of solutions. Hence, it is more suitable to train the NFS than the GAs. The features of MA which make it suitable for the probability of selection of operators are given in Figure-2.

- Global optimization,
- Stochastic and
- Selection is based on good features.

An algorithm with one individual string is chosen which is defined as the 'queen' string. The generation is generated by performing mutation on this string. This string contains the msf width parameter P and the weights which have to be applied between the input and hidden layers. The steps of the algorithm are:

a) Randomly generate the individual string (queen string)
b) Initialize it to the bit string of 0's and 1's
c) Child string is generated using the mutation operator
d) Mutation points are chosen randomly
e) For each string
   - Decode the parameter values
   - Assign to the neuro-fuzzy filter
f) Stop the process after 100 generations

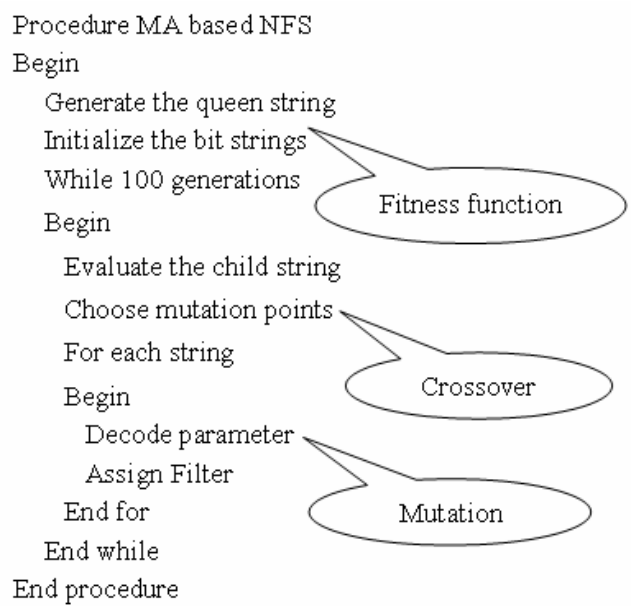


Figure-2. Application of MA parameters.

After the speckle noise is removed, the enhancement process is carried out.

Phase 2: Image enhancement

The process included in image enhancement phase is shown in Figure-3.

Normalization

It is complicated to simplify the processes for the ultrasound images which huge variations in the gray. Normalization can be utilized to overcome this difficulty and map various gray levels into the similar range [0, 1] with the help of the membership function μ_A

$$\mu_A(g) = 255 \times \frac{g(i,j) - g_{min}}{g_{max} - g_{min}} \quad (1)$$

where g_{min} and g_{max} represents the minimum and maximum gray levels of the original ultrasound image, respectively and g (i, j) represents the gray levels of pixel (i, j).

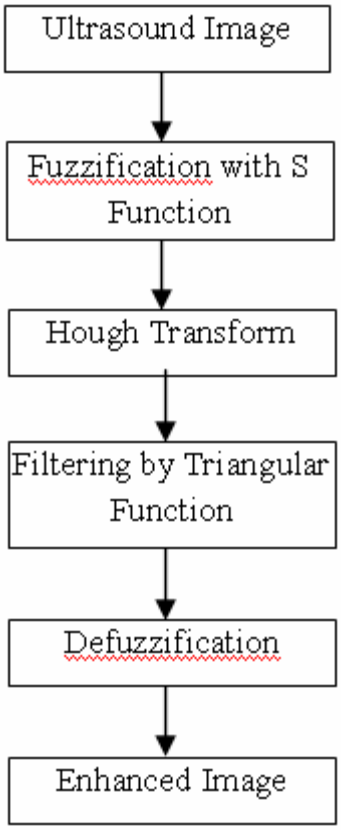


Figure-3. Image enhancement process.

Fuzzification

According to the information of fuzzy theory, the ultrasound images can record every pixels of a set into the real numbers in [0, 1] by selecting suitable fuzzy set for more opinion. In this technique, the fuzzy set of S function is utilized not only to map the image to the fuzzy domain, but to make use of the membership values of its gray levels for further operations. The S function is defined as below:

$$S(\mu_A;a,b,c)=\begin{cases}0, & if\ \mu_A \le a,\\ \dfrac{(\mu_A(g)-a)^2}{(b-a)(c-a)}, & if\ a<\mu_A\le b,\\ 1-\dfrac{(\mu_A(g)-c)^2}{(c-a)(c-b)}, & if\ b<\mu_A\le c,\\ 1, & if\ \mu_A>a\end{cases} \quad (2)$$

where a, b and c indicates the parameters that decides the shape of S function. In this computation, the maximum entropy is utilized to find out specific numbers of the parameters and the formula can be diven by

$$Entropy_{max}=\underset{i=1}{\overset{N}{Argmax}}\{H_l(t)+H_g(t)\} \quad (3)$$

where $H_l(t)$ and $H_g(t)$ indicates probability distributions of gray level below and above the threshold gray level t correspondingly, and the optimal threshold, t_{max}, can be determined by maximizing the sum of the $H_l(t)$ and $H_g(t)$.

After the fuzzification process, the ultrasound image is transformed from the gray level into the fuzzification domain by

$$\mu(i,j)=S(\mu_A;a,t_{max},c) \quad (4)$$

The parameters of a and c are nothing but the gray levels from the search of the first peak and last peak of image histogram, respectively.

Hough transform

Hough transform (HT) is a method for detecting features of a definite shape inside an image. In an application of HT, it permits detection of lines in an image. In realistic implementation of Hough transform all the line is indicated by two parameters, such as r and θ, which are the distance and the angle of the normal to the to-be-detected line from the center of the co-ordinate system as indicated in Figure-4.

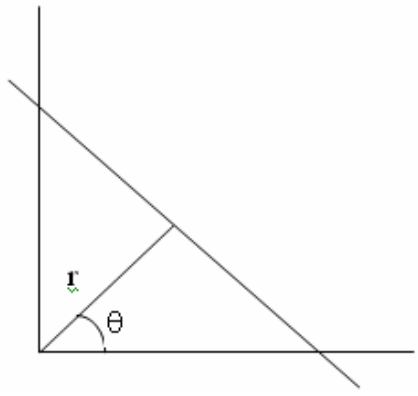


Figure-4. A Typical line showing the parameters for Hough transform.

With these parameters the equation of the straight line is given by

$$x\cos\theta + y\cos\theta = d \quad (5)$$

Note that a line can be indicated by a single r and θ value. A point is indulgenced as meeting points of infinite number of lines. By evaluating the value of r for all probable θ value for all these lines through a point, a sinusoid curve is generated in the r-θ plane which is also called as Hough plane. The offerings from all such points are collected in the Hough space. Two points occurring on the same line should have a common r and θ values. Therefore the two sinusoids equivalent to these points belonging to a line intersect at a common (r_i, θ_i) in the parameter space. Therefore every straight line with a fixed (r_i, θ_i) value will emerge to have a high peak in the accumulated Hough space that will be proportional to the number of pixels occurred in the line. Identifying the peak in the Hough space, identify the straight lines with the equivalent (r_i, θ_i) value. When lines indicate edges of objects are to be determined, generally an edge detector is utilized to decrease the quantity of information and the background noise. It can be noted that matched filtering based method will not be suitable here because the features of all the corners are not same. It may require a large number of templates to encompass all the differences.

Filtering

After performing the Hough transform, a filter of triangular function that is created to filter the needless signal is defined in equation 6 and is represented in Figure-5. The intention of the process is to safeguard the most low frequency component and take away the noise occurred in the images by the high frequency component. Hence, if there exist more artifacts on the image of a lesion, this needless content will be removed or suppressed.

$$F(f)=\begin{cases}1+\dfrac{f}{f'}, & if\ f<0\\ 1-\dfrac{f}{f'}, & if\ f\ge 0\end{cases} \quad (6)$$

where F represents a function of the spatial frequencies transferred from the image amplitude at particular frequency *f*, where the upper and lower limit of frequency range are [*-f'*,*f'*].

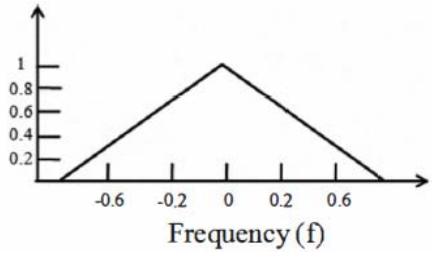


Figure-5. The demonstration of the suggested triangular filter.

Defuzzification

The adjustment intensity of images can be derived with the help of the inverse function

$$S(\mu_A; a, t_{max}, c) = \begin{cases} 0, & if\ \mu_A(g) \le a, \\ \dfrac{(\mu_A(g)-a)^2}{(t_{max}-a)(c-a)}, & if\ a < \mu_A(g) \le t_{max}, \\ 1-\dfrac{(\mu_A(g)-c)^2}{(c-t_{max})(c-a)}, & if\ b < \mu_A(g) \le c, \\ 1, & if\ \mu_A(g) > a \end{cases} \quad (7)$$

After the defuzzification step is carried out, the reconstructed image will show the enhancement efficacy by replacement of the original image.

Phase 3: Image segmentation

After the extraction of mammary gland region using appropriate preprocessing steps, segmentation is performed in order to detect the occurrence of cancer regions. This paper uses a new segmentation algorithm called Modified Fuzzy Possibilistic C-Means with Repulsion. This clustering technique includes the advantages of both Fuzzy Possibilistic Clustering Algorithm and C-Means clustering algorithm. After that, the weight factor is included in the clustering algorithm, so that the objective function gets enhanced. Finally, a repulsion term is introduced in the objective function in order to increase the intra cluster distance in the cluster. This will help in better segmentation result. The methodology is discussed below:

The characteristics of both fuzzy and possibilistic c-means approaches is incorporated in fuzzy and possibilistic c-means algorithm. Memberships and typicalities are very important factors for the correct feature of data substructure in clustering problem. Consequently, an objective function in the FPCM depending on both memberships and typicalities can be represented as below:

$$J_{FPCM}(U,T,V) = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij}^{m} + t^{n})\, d^2(X_j, v_i) \quad (8)$$

with the following constraints:

$$\sum_{i=1}^{c} \mu_{ij} = 1, \forall j \in \{1, \ldots, n\} \quad (9)$$

$$\sum_{j=1}^{n} t_{ij} = 1, \forall i \in \{1, \ldots, c\} \quad (10)$$

A solution of the objective function can be obtained through an iterative process where the degrees of membership, typicality and the cluster centers are update with the equations as follows.

$$\mu_{ij} = \left[\sum_{k=1}^{c}\left(\frac{d(X_j, v_i)}{d(X_j, v_k)}\right)^{2/(m-1)}\right]^{-1}, 1 \le i \le c, 1 \le j \le n. \quad (11)$$

$$t_{ij} = \left[\sum_{k=1}^{n}\left(\frac{d(X_j, v_i)}{d(X_j, v_k)}\right)^{2/(\eta-1)}\right]^{-1}, 1 \le i \le c, 1 \le j \le n. \quad (12)$$

$$v_i = \frac{\sum_{k=1}^{n}(\mu_{ik}^{m} + t_{ik}^{\eta})X_K}{\sum_{k=1}^{n}(\mu_{ik}^{m} + t_{ik}^{\eta})}, 1 \le i \le c. \quad (13)$$

PFCM constructs memberships and possibilities simultaneously, along with the usual point prototypes or cluster centers for each cluster. Hybridization of possibilistic c-means (PCM) and fuzzy c-means (FCM) is the PFCM that often avoids various problems of PCM, FCM and FPCM. The noise sensitivity defect of FCM is solved by PFCM, which overcomes the coincident clusters problem of PCM. But the estimation of centroids is influenced by the noise data.

Modified Fuzzy Possibilistic C-Means Technique (FPCM)

Objective function is very much necessary to enhance the quality of the clustering results. Wen-Liang Hung presented a new approach called Modified Suppressed Fuzzy C-Means (MS-FCM), which significantly improves the performance of FCM due to a prototype-driven learning of parameter α. Exponential separation strength between clusters is the base for the learning process of α and is updated at each of the iteration. The parameter α can be computed as

$$\alpha = exp\left[-\min_{i \ne k}\frac{\|v_i - v_k\|^2}{\beta}\right] \quad (14)$$

In the above equation β is a normalized term so that β is chosen as a sample variance. That is, β is defined:

$$\beta = \frac{\sum_{j=1}^{n}\|x_j - \bar{x}\|^2}{n} \; where \; \bar{x} = \frac{\sum_{j=n}^{n} x_j}{n}$$

But the remark which must be pointed out here is the common value used for this parameter by all the data at each of the iteration, which may induce in error. A new parameter is added with this which suppresses this common value of α and replaces it by a new parameter like a weight to each vector. Or every point of the data set possesses a weight in relation to every cluster. Consequently this weight permits to have a better classification especially in the case of noise data. The following equation is used to calculate the weight.

$$w_{ji} = exp\left[-\frac{\|x_j - v_i\|^2}{\left[\sum_{j=1}^{n}\|x_j - \bar{v}\|^2\right] * c/n}\right] \quad (15)$$

In the previous equation w_{ji} represents weight of the point j in relation to the class i. In order to alter the fuzzy and typical partition, this weight is used. The objective function is composed of two expressions: the first is the fuzzy function and uses a fuzziness weighting exponent, the second is possibililstic function and uses a typical weighting exponent; but the two coefficients in the

objective function are only used as exhibitor of membership and typicality. A new relation, lightly different, enabling a more rapid decrease in the function and increase in the membership and the typicality when they tend toward 1 and decrease this degree when they tend toward 0. This relation is to add Weighting exponent as exhibitor of distance in the two under objective functions. The objective function of the MFPCM can be given as follows:

$$J_{MFPCM} = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij}^{m} w_{ij}^{m} d^{2m}(x_j,v) + t_{ij}^{\eta} w_{ij}^{\eta} d^{2\eta}(x_j,v_i)) \quad (16)$$

U = $\{\mu_{ij}\}$ represents a fuzzy partition matrix, is defined as:

$$\mu_{ij} = \left[\sum_{k=1}^{c}\left(\frac{d(X_j,v_i)}{d(X_j,v_k)}\right)^{2m/(m-1)}\right]^{-1} \quad (17)$$

T = $\{t_{ij}\}$ represents a typical partition matrix, is defined as:

$$t_{ij} = \left[\sum_{k=1}^{n}\left(\frac{d(X_j,v_i)}{d(X_j,v_k)}\right)^{2\eta/(\eta-1)}\right]^{-1} \quad (18)$$

V = $\{v_i\}$ represents c centers of the clusters, is defined as:

$$v_i = \frac{\sum_{j=1}^{n}(\mu_{ij}^{m} w_{ji}^{m} + t_{ij}^{\eta} w_{ji}^{\eta}) * X_j}{\sum_{j=1}^{n}(\mu_{ik}^{m} w_{ji}^{m} + t_{ik}^{\eta} w_{ij}^{\eta})} \quad (19)$$

Penalized and compensated constraints based modified fuzzy possibilistic C-Means (PCMFPCM)

The Penalized and compensated constraints are embedded with the previously discussed Modified Fuzzy Possibilistic C-Means algorithm. The objective function of the FPCM is given in equation (16). In the proposed approach the penalized and compensated terms are added to the objective function of FPCM to construct the objective function of PCMFPCM. The penalized constraint can be represented as follows:

$$\frac{1}{2}v\sum_{x=1}^{n}\sum_{i-1}^{c}(\mu_{x,i}^{m} ln\alpha_i + t_{x,i}^{\eta} ln\beta_x) \quad (20)$$

Where

$$\alpha_i = \frac{\sum_{x=1}^{n}\mu_{x,i}^{m}}{\sum_{x=1}^{n}\sum_{i=1}^{c}\mu_{x,i}^{m}}, \quad i = 1,2,\ldots\ldots c,$$

$$\beta_x = \frac{\sum_{i=1}^{c}t_{x,i}^{\eta}}{\sum_{x=1}^{n}\sum_{i=1}^{c}t_{x,i}^{\eta}} \quad x = 1,2,\ldots,n$$

where α_i is a proportional constant of class i; β_x is a proportional constant of training vector z_x, and v (v≥0); τ (τ≥0) are also constants. In these functions, α_i and β_x are defined in equations above. Membership $\mu_{x,i}$ and typicality $t_{x,i}$ for the penalize is presented below.

$$(\mu_{x,i})_P = \left(\sum_{i=1}^{c}\frac{(\|z_x-\varpi_i\|^2 - v\,ln\alpha_i)^{1/(m-1)}}{(\|z_x-\varpi_i\|^2 - v\,ln\alpha_i)^{1/(m-1)}}\right)^{-1}$$

$$x = 1,2,\ldots,n, i = 1,2,\ldots c,$$

$$(t_{x,i})_P = \left(\sum_{y=1}^{n}\frac{(\|z_x-\varpi_i\|^2 - v\,ln\beta_x)^{1/(\eta-1)}}{(\|z_x-\varpi_i\|^2 - v\,ln\beta_y)^{1/(\eta-1)}}\right)^{-1}$$

$$x = 1,2,\ldots,n, i = 1,2,\ldots c,$$

In the previous expression

$$\varpi_i = v_i = \frac{\sum_{k=1}^{n}(\mu_{ik}^{m}+t_{ik}^{\eta})x_k}{\sum_{k=1}^{n}(\mu_{ik}^{m}+t_{ik}^{\eta})}, 1 \le i \le c.$$

which is the centroid. The compensated constraints can represented as follows

$$\frac{1}{2}\tau\sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i}^{m} tanh\,\alpha_i + t_{x,i}^{\eta} tanh\,\beta_x) \quad (21)$$

Where Membership $\mu_{x,i}$ and typicality $t_{x,i}$ for the compenzation is presented below

$$(\mu_{x,i})_C = \left(\sum_{i=1}^{c}\frac{(\|z_x-\varpi_i\|^2 - \tau\tanh(\alpha_i))^{1/(m-1)}}{(\|z_x-\varpi_i\|^2 - \tau\,tanh(\alpha_i))^{1/(m-1)}}\right)^{-1}$$

$$x = 1,2,\ldots,n, i = 1,2,\ldots c,$$

$$(t_{x,i})_C = \left(\sum_{y=1}^{n}\frac{(\|z_x-\varpi_i\|^2 - \tau\,tanh(\beta_x))^{1/(\eta-1)}}{(\|z_x-\varpi_i\|^2 - \tau\tanh(\beta_y))^{1/(\eta-1)}}\right)^{-1}$$

$$x = 1,2,\ldots,n, i = 1,2,\ldots c,$$

To obtain an efficient clustering the penalization term must be removed and the compensation term must be added to the basic objective function of the existing FPCM. This brings out the objective function of PCFPCM and it is given in equation (22).

$$J_{MFPCM} = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij}^{m} w_{ji}^{m} d^{2m}(x_j,v) + t_{ij}^{\eta} w_{ji}^{\eta} d^{2\eta}(x_j,v_i)) - \frac{1}{2}v\sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i}^{m}\ln\alpha_i + t_{x,i}^{\eta}\ln\beta_x) + \frac{1}{2}v\sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i}^{m}\tanh\alpha_i + t_{x,i}^{\eta}\tanh\beta_x) \quad (22)$$

The centroid of *i*th cluster is calculated in the similar way as the definition in equation (19). The final objective function is presented in equation (22).

Clustering enhancement using repulsion

In the above described clustering technique, objective function is truly minimized only if all the centroids are identical (coincident), since the typicality of a point to a cluster, depends only on the distance between the point to that cluster.

The usage of repulsion aims to minimize the intracluster distances, while maximizing the intercluster distances, without using implicitly the restriction, but by adding a cluster repulsion term to the objective function.

$$J_{MFPCM} = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij}^{m} w_{ji}^{m} d^{2m}(x_j v) + t_{ij}^{\eta} w_{ji}^{\eta} d^{2\eta}(x_j, v_i)) - \frac{1}{2}\nu\sum_{x=1}^{n}\sum_{i=1}^{c}\left(\mu_{x,i}^{m}\ln\alpha_i + t_{x,i}^{\eta}\ln\beta_x\right) + \frac{1}{2}\nu\sum_{x=1}^{n}\sum_{i=1}^{c}\left(\mu_{x,i}^{m}\tanh\alpha_i + t_{x,i}^{\eta}\tanh\beta_x\right) + \sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-u_{ik})^m + \gamma\sum_{i=1}^{c}\sum_{k=1,k\neq i}^{c}\frac{1}{d^2(v_i,v_k)} \quad (23)$$

Where γ is a weighting factor, and u_{ik} satisfies:

$$u_{ik} \in [0,1], \forall i \quad (24)$$

The repulsion term is relevant if the clusters are close enough. With growing distance it becomes smaller until it is compensated by the attraction of the clusters. On the other hand, if the clusters are sufficiently spread out and the intercluster distance decreases, the attraction of the cluster can be compensated only by the repulsion term.

Minimization of objective function with respect to cluster prototypes leads to:

$$v_i = \frac{\sum_{j=1}^{n} u_{ij}x_j - \gamma\sum_{k=1,k\neq i} v_k \frac{1}{d^2(v_k,v_i)}}{\sum_{j=1}^{n} u_{ij} - \gamma\sum_{k=1,k\neq i}\frac{1}{d^2(v_k,v_i)}} \quad (25)$$

Singularity occurs when one or more of the distances $d^2(v_k,v_i) = 0$ at any iteration. In such a case, v_i cannot be calculated. When this happens, assign zeros to each nonsingular class (all the classes except i) and assign 1 to class i, in the membership matrix U.

An alternative repulsion term for (23) in order to minimize the objective function is given by

$$J_{MFPCM} = \sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij}^{m} w_{ji}^{m} d^{2m}(x_j v) + t_{ij}^{\eta} w_{ji}^{\eta} d^{2\eta}(x_j, v_i)) - \frac{1}{2}\nu\sum_{x=1}^{n}\sum_{i=1}^{c}\left(\mu_{x,i}^{m}\ln\alpha_i + t_{x,i}^{\eta}\ln\beta_x\right) + \frac{1}{2}\nu\sum_{x=1}^{n}\sum_{i=1}^{c}\left(\mu_{x,i}^{m}\tanh\alpha_i + t_{x,i}^{\eta}\tanh\beta_x\right) + \sum_{i=1}^{c}\eta_i\sum_{k=1}^{n}(1-u_{ik})^m + \gamma\sum_{i=1}^{c}\sum_{k=1,k\neq i}^{c} e^{-d^2(v_k,v_i)} \quad (26)$$

The weighting factor γ is used to balance the attraction and repulsion forces, i.e., minimizing the intradistances inside clusters and maximizing the interdistances between clusters.

The proposed segmentation technique is applied to mammary gland image segmentation. The pixel values are the inputs of the clustering algorithm, and the pixels are clustered based on the optimum centers of clustering. The values of the pixels contained in the lesion are very low, the cluster of pixels with the lesser intensities can be considered as the lesion-like pixels. The mammary gland region is determined by the following formula:

$$bw(i,j) = \begin{cases} 0, & g(i,j) \in C_1 \\ 255, & otherwise \end{cases} \quad (27)$$

Where,

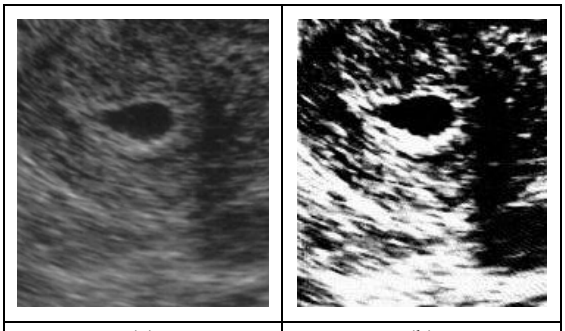
g(i, j) = pixel in mammary gland region at the location (i,j)
C_1 = cluster with the lesser intensities.
bw - binary mammary gland image after segmentation.

After the mammary gland is segmented, the round-like regions are kept as the lesion-like regions and the others are rejected.

4. EXPERIMENTAL RESULTS

The experiments are conducted on the proposed computer-aided diagnosis systems with the help of real time breast ultrasound images. This experimentation data consists of 10 ultrasound images. Those 10 ultrasound images are passed to the proposed system. The noise from the images is removed and it is enhanced for better diagnosis. Then the proposed segmentation algorithm is applied to the gathered image. This segmentation algorithm will cluster the ultrasound image according to its intensity. This will help in identifying the cancer affected regions and finally it will detect whether the supplied lung image is with cancer or not.

Figure-6 represents the resulted enhancement in image using the proposed technique.



(a) (b)

Figure-6. (a) Original image (b) enhanced image using Hough transform.

Table-1 and Figure-7 represents the resulted Mean Square Error (MSE) for the used ultrasound images. From the data, it can be observed that the proposed technique results in lesser MSE values which indicate the better enhancement when compared to the conventional technique.

Table-1. Resulted mean square error (MSE).

Ultrasound image	MSE	
	Adaptive mean filter	Proposed
1	0.5012	0.4854
2	0.5214	0.4912
3	0.6148	0.5135
4	0.5987	0.5221
5	0.5824	0.5187
6	0.4991	0.4518
7	0.5354	0.4821
8	0.5864	0.4997
9	0.5648	0.5102
10	0.6058	0.5861

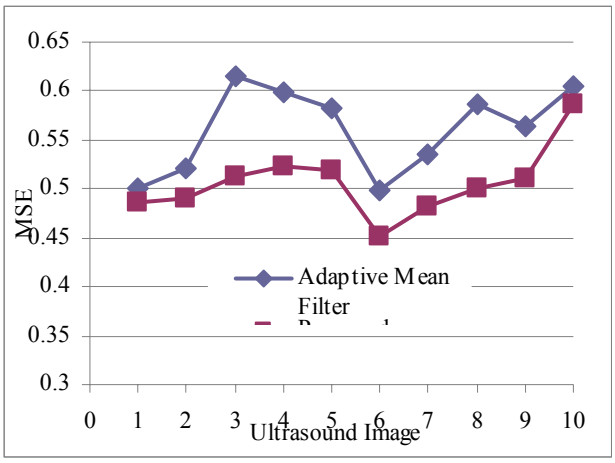


Figure-7. Comparison of mean square error (MSE).

Table-2 and Figure-8 represents the resulted accuracy for segmentation of the used ultrasound images. From the data, it can be observed that the proposed segmentation algorithm results in better accuracy for segmentation when compared to the conventional technique.

Table-3 and Figure-9 represents the resulted standard deviation for segmentation of the used ultrasound images. From the data, it can be observed that the proposed segmentation algorithm results in lesser standard deviation for segmentation when compared to the conventional technique. This suggests that the proposed technique is resulted in better segmentation.

The segmentation result is shown in Figure-10. Figure-10(a) indicates the original ultrasound image and Figure-10(b) provides the segmented.

Table-2. Segmentation accuracy comparison.

Ultrasound image	Accuracy	
	EPSO	MFPCM
1	97.31	98.49
2	97.21	98.05
3	97.38	98.15
4	98.45	99.12
5	97.98	98.31
6	97.81	98.32
7	98.15	99.04
8	98.45	99.55
9	98.52	99.08
10	98.75	99.51

Table-3. Standard deviation comparison.

Ultrasound image	Standard deviation	
	EPSO	MFPCM
1	4.4	3.4
2	5.3	2.2
3	5.8	2.5
4	6.3	2.3
5	5.1	0
6	4.6	0
7	4.8	2.1
8	4.1	1.2
9	5.6	0
10	4.7	3.1

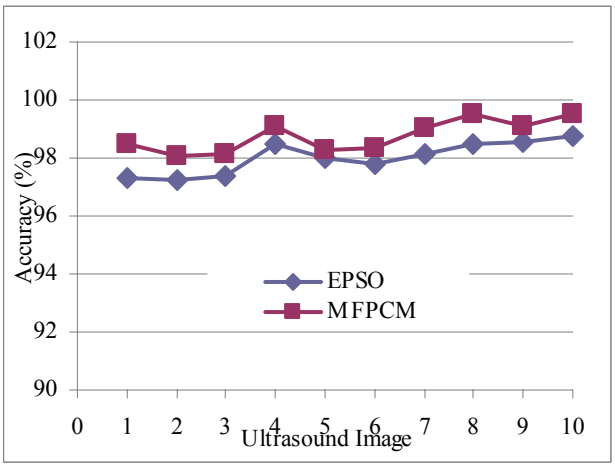


Figure-8. Resulted accuracy for segmentation.

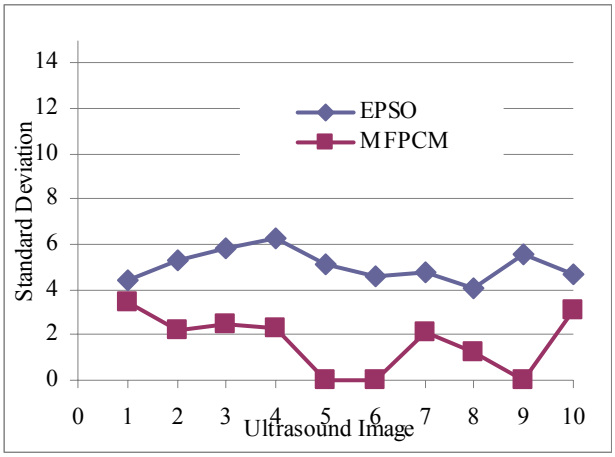


Figure-9. Resulted standard deviation for segmentation.

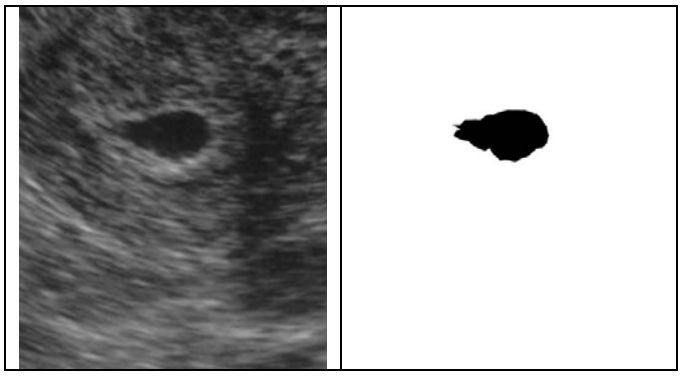


Figure-10. (a) Ultrasound image (b) segmented image.

5. CONCLUSIONS

The difficulties faced by the existing system for breast cancer detection is solved in this paper. This paper suggests a CAD system for detection of breast cancer from the ultrasound image which involves three phases such as speckle noise reduction, image enhancement and segmentation. For removing the speckle noise, this paper uses Memetic algorithm. Image enhancement is performed using Hough transform. Finally, the enhanced image is segmented using clustering technique called Modified Fuzzy Possibilistic C-Means technique with Repulsion factor to identify the cancer affected region. The proposed system is evaluated using the real time dataset collected from various patients. The experimental result shows that the proposed CAD performs better when compared to the existing system.

REFERENCES

[1] 2008. American Cancer Society, Breast Cancer Facts and Figures 2007-2008, American Cancer Society.

[2] Drukker K., Giger M.L., Vyborny C.J. and Mendelson E. B. 2004. Computerized detection and classification of cancer on breast ultrasound. Academic Radiology. 11(5): 526-535.

[3] Cheng H.D., Cai X., Chen X., Hu L. and Lou X. 2003. Computer-aided detection and classification of micro calcifications in mammograms: a survey. Pattern Recognition. 36(12): 2967-2991.

[4] Breastcancer.org. Mammograms. 2009. http://www.breastcancer.org/symptoms/testing/types/mammograms/. June.

[5] Zhang L.C., Wong E.M.C., Zhang F. and Zhou J. 2006. Adaptive pyramid filtering for medical ultrasound image enhancement. In 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro. pp. 916-919.

[6] Shi X., Cheng H.D., Hu L., Ju W. and Tian J. 2010. Detection and classification of masses in breast ultrasound images. Digital Signal Processing. 20(3): 824- 836.

[7] Adam D., Beilin-Nissan S., Friedman Z. and Behar V. 2006. The combined effect of spatial compounding and nonlinear filtering on the speckle reduction in ultrasound images. Ultrasonics. 44(2): 166.

[8] Li X. and Liu D.C. 2007. Ultrasound image enhancement using dynamic filtering. In 4th International Conference on Image and Graphics. pp. 106-109.

[9] Juan Wachs, Oren Shapira and Helman Stern. 2006. A Method to Enhance the Possibilistic C-Means with Repulsion Algorithm based on Cluster Validity Index. Advances in Intelligent and Soft Computing, Springerlink. 34: 77-87.

[10] Deshmukh K.S. and Shinde G.N. 2005. An adaptive color image segmentation. Electronic Letters on Computer Vision and Image Analysis. 5(4): 12-23.

[11] Noble J.A. and D. Boukerroui D. 2006. Ultrasound image segmentation: A survey. IEEE Transactions on Medical Imaging. 259(8): 987-1010.

[12] Madabhushi A. and Metaxas D.N. 2003. Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. IEEE Transactions on Medical Imaging. 22(2): 155-169.

[13] Noble J.A. and Boukerroui D. 2006. Ultrasound image segmentation: A survey. IEEE Transactions on Medical Imaging. 25(8): 987.

[14] Chen C.-Y. and Ye F. 2004. Particle swarm optimization algorithm and its application to

clustering analysis. In IEEE International Conference on Networking, Sensing and Control. pp. 789-794.

[15] Oelze M.L., O'Brien W.D., Zachary J.F. 2007. 11B-4 Quantitative Ultrasound Assessment of Breast Cancer Using a Multiparameter Approach. IEEE Ultrasonics Symposium. pp. 981-984.

[16] Gefen S., Tretiak O.J., Piccoli C.W., Donohue K.D., Petropulu A.P., Shankar P.M., Dumane V.A., Lexun Huang, Kutay M.A., Genis V., Forsberg F., Reid J.M., Goldberg B.B. 2003. ROC analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis. IEEE Transactions on Medical Imaging. 22(2): 170-177.

[17] Winder A.A., Jadidian B., Muratore R. 2010. Synthetic Structural Imaging (SSI): A new ultrasound method for tracking breast cancer morphology. 39th Annual Ultrasonic Industry Association Symposium (UIA). pp. 1-4.

[18] J.-S.R. Jang. 1993. ANFIS: adaptive network-based fuzzy inference system. IEEE Trans. Systems, Man, Cybern. 23(03): 665-685.

[19] S. Haykin. 1998. Neural Networks, Prentice-Hall, Englewood Cliffs, NJ.

[20] Elif Derya Ubeyli, Inan Guler. 2005. Teaching Automated Diagnostic systems for Doppler ultrasound blood flow signals to biomedical engineering students using Matlab. International Journal of Engineering Education. 21(4): 649-667.

[21] I. Kalaykov, G. Tolt. 2003. Real-time image noise cancellation based on fuzzy similarity, in: M. Nachtegael *et al.* (Eds.), Fuzzy Filters for Image Processing, Springer, Berlin, Heidelberg, New York. pp. 54-71.

[22] B. Kosko. 1992. Neural Networks and Fuzzy System, Prentice-Hall, Englewood Cliffs, NJ.

[23] Y. S. Ong, M.H. Lim, N. Zhu, K.W. Wong. 2006. Classification of adaptive memetic algorithms: a comparative study. IEEE Trans. Syst. Man Cybern. Part B. 36(1): 141-152.

[24] A. Rafiee Kerachi, M.H. Moradi, M.R. Farzaneh. 2003. Speckle noise reduction in sonography images by using online genetic neuro fuzzy filters. Proceedings of the 4th Seminars on Fuzzy Sets and its Applications- Iran. pp. 70-77.