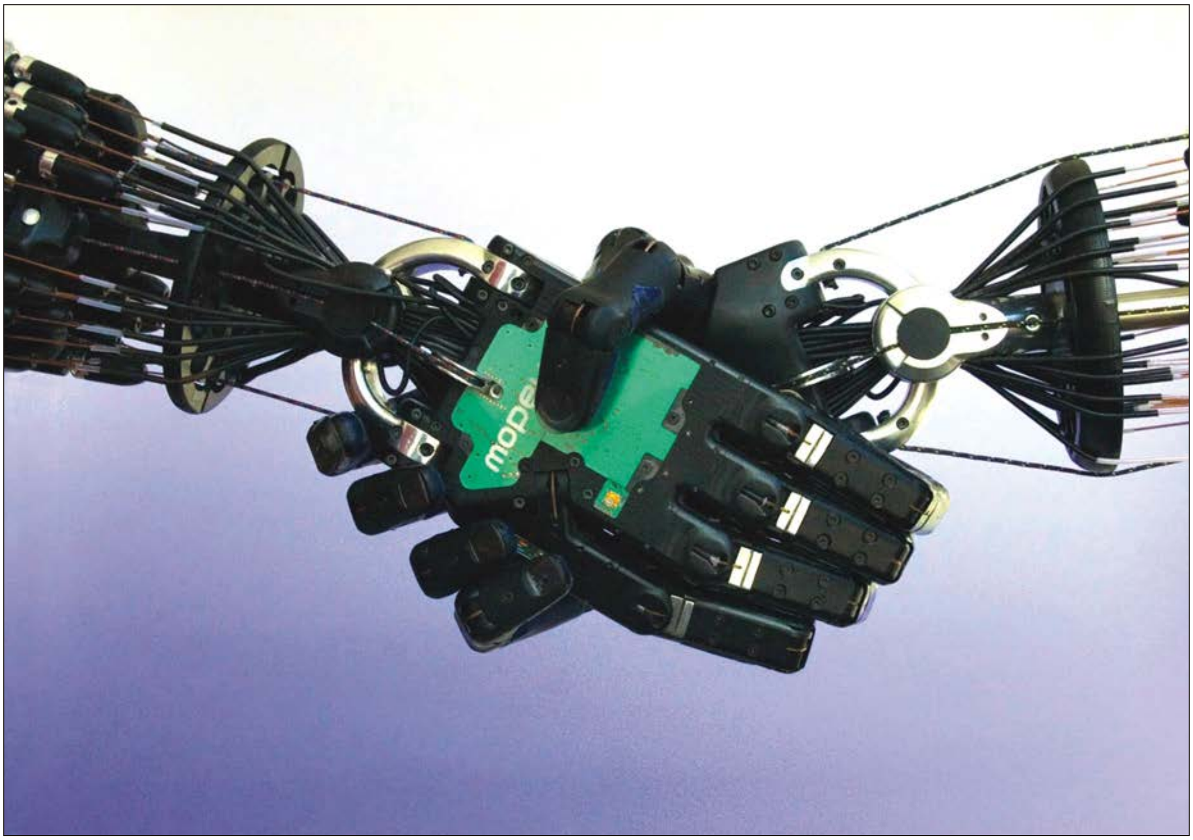

Journal of Artificial Intelligence

ISSN 1994-5450

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Artificial Intelligence 5 (1): 14-25, 2012
ISSN 1994-5450 / DOI: 10.3923/jai.2012.14.25

Automated Clustering of Cancer Cells Using Fuzzy C Means with Repulsions in Ultrasound Images

[1]N. Alamelumangai and [2]J. Devishree
[1]Department of MCA, Karpagam College of Engineering, Coimbatore, India
[2]Department of Electrical and Electronics Engineering, Coimbatore Institute of Technology, Coimbatore, India

Corresponding Author: N. Alamelumangai, Department of MCA, Karpagam College of Engineering, Myleripalayam, Coimbatore, India

ABSTRACT

In the report provided by World Health Organization (WHO), breast cancer is one of the highly deadliest cancers occurred in middle-aged women. Accurate diagnosis and prediction are essential to decrease the high death rate. In the past few years, breast ultrasound images have turn out to be an optional for mammography to help distinguish benign from malignant lesions. Its advantages safety and cost-effectiveness as discussed by various authors have turned ultrasound method into an increasingly significant function in the estimate of breast lesions. Like ultrasound exams in general, breast ultrasound exams are comparatively low-priced and do not utilize X-rays or other kinds of probably dangerous radiation. As a result, breast inspection with the help of ultrasound method has turn out to be a main option to mammography. This study provides a new technique in the field of computer science for diagnosis of breast tumors on ultrasound images. Some preprocessing steps are performed in the obtained ultrasound image in to enhance it for better diagnosis. Finally, the cancerous cells are distinguished by clustering them. This study uses Modified Fuzzy Possibilistic C-Means technique with Repulsion factor for the purpose of clustering the cancerous cells. The proposed computer-aided diagnosis systems are evaluated using the real time breast ultrasound images. The accuracy of predicting the cancer regions in the breast is higher when compared to the conventional technique. Also, the standard deviation resulted for the proposed approach is lesser than conventional technique. The experimental results show that the proposed technique results in better detection of cancer regions as it produces higher accuracy of detecting the cancer regions.

Key words: Ultrasound image, clustering, fuzzy possibilistic c-means, repulsion

INTRODUCTION

The most frequently diagnosed cancer in women aged between 40 and 60 is breast cancer (ACS, 2008; Drukker *et al*., 2004). Based on the report released by World Health Organization there are about 7.6 million deaths occurred worldwide because of cancer every year, among them around 502,000 deaths are resulted by breast cancer alone. By analyzing these data, breast cancer is determined to be the most deadly cancer. For few decades researchers have been involved to determine the best method to diagnosis breast cancer. Triumphant 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. Cancers in their starting stages are susceptible to treatment whereas cancers in their highly developed stages are typically almost impractical to cure.

Even though breast cancer has very high occurrence and death rate, the reason for breast cancer is still not determined. No efficient method to prevent the occurrence of breast cancer subsists. 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 study 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 (Cheng *et al.*, 2003) 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

Initially, the US image (Zhang *et al.*, 2006; Shi *et al.*, 2010) undergoes various preprocessing steps in order to obtain a clear image for diagnosis (Li and Liu, 2007; Adam *et al.*, 2006). Then the obtained image is segmented in order to separate the cancer regions and made it available for diagnosis. Previous well known algorithm for ultrasound image segmentation (Deshmukh and Shinde, 2005; Noble and Boukerroui, 2006; Madabhushi and Metaxas, 2003) is Eliminating Particle Swarm Optimization (EPSO) (Chen and Ye, 2004). This study uses Modified Fuzzy Possibilistic C-Means Technique with repulsion factor (Wachs *et al.*, 2006) for segmentation purpose which overcomes the difficulties of EPSO algorithm and will result in better segmentation.

Oelze *et al.* (2007) 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.* (2003) 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 resampling 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*. (2010) 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.

MATERIALS AND METHODS

With the obtained ultrasound image, preprocessing such as noise removal and image enhancement are performed. After the extraction of mammary gland region using appropriate preprocessing steps, segmentation is performed in order to detect the occurrence of cancer regions. This study 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:

Fuzzy possibilistic clustering algorithm: The fuzzified version of the k-means algorithm is Fuzzy C-Means (FCM). It is a clustering approach which allows one piece of data to correspond to two or more clusters. Dunn in 1973 developed this technique and it was modified by Bezdek in 1981, Li and Liu (2007) and this is widely used in pattern recognition. The algorithm is an iterative clustering approach that brings out an optimal c partition by minimizing the weighted within group sum of squared error objective function J_{FCM}:

$$J_{FCM}(V,U,X)=\sum_{i=1}^{c}\sum_{j=1}^{n}u^{m}_{ij}\,d^{2}(X_j,v_i),1<m<+\infty \tag{1}$$

In the equation $X = \{x_1, x_2,..., x_n\}\subseteq R^p$ the data set in the p-dimensional vector space, the number of data items is represented as p, c represents the number of clusters with 2 = c = n-1. $V = \{v_1, v_2,...,v_c\}$ is the c centers or prototypes of the clusters, v_i represents the p-dimension center of the cluster i and $d^2(x_j, v_i)$ represents a distance measure between object x_j and cluster centre v_i. $U = \{\hat{\imath}_{ij}\}$ represents a fuzzy partition matrix with $u_{ij} = u_i(x_j)$ is the degree of membership of x_j in the ith cluster; x_j is the jth of p-dimensional measured data. The fuzzy partition matrix satisfies:

$$0<\sum_{j=1}^{n}\mu_{ij}<n,\forall i\in\{1,...,n\} \tag{2}$$

$$\sum_{i=1}^{c}\mu_{ij} < n, \forall i \in \{1,...,n\} \tag{3}$$

m is a weighting exponent parameter on each fuzzy membership and establishes the amount of fuzziness of the resulting classification; it is a fixed number greater than one. Under the constraint of U the objective function J_{FCM} can be minimized. Specifically, taking of J_{FCM} with respect to u_{ij} and v_i and zeroing them respectively, is necessary but not sufficient conditions for J_{FCM} to be at its local extrema will be as the following:

$$\mu_{ij} = \left[\sum_{k=1}^{c}\left(\frac{d(X_j v_j)}{d(X_j v_k)}\right)^{2/(m-1)}\right]^{-1}, 1 \le i \le c, 1 \le j \le n \tag{4}$$

$$v_i = \frac{\Sigma_k^n = {}_1\mu_{ik}^m x_k}{\Sigma_k^n = {}_1\mu_{ik}^m}, 1 \le i \le c \tag{5}$$

In noisy environment, the memberships of FCM do not always correspond well to the degree of belonging of the data and may be inaccurate. This is mainly because the real data unavoidably involves some noises. To recover this weakness of FCM, the constrained condition (3) of the fuzzy c-partition is not taken into account to obtain a possibilistic type of membership function and PCM for unsupervised clustering is proposed. The component generated by the PCM belongs to a dense region in the data set; each cluster is independent of the other clusters in the PCM strategy. The following formulation is the objective function of the PCM:

$$J_{PCM}(V,U,X)\sum_{i=1}^{c}\sum_{j=1}^{n}\mu_{ik}^m d^2(Xj, vi) + \sum_{i1}^{c}\eta_i\sum_{j=1}^{n}(1-u_{ij})^m \tag{6}$$

Where:

$$\eta_i = \frac{\Sigma_j^n \; {}_1\mu_{ik}^m \|x_j - v_i\|^2}{\Sigma_j^n = {}_1\mu_{ij}^m} \tag{7}$$

η_i is the scale parameter at the ith cluster, u_{ij}

$$u_{ij}\frac{1}{1+\left[\dfrac{d^2(x_j, v_i)}{\eta_i}\right]}\frac{1}{m-1} \tag{8}$$

u_{ij} represents the possibilistic typicality value of training sample x_j belong to the cluster i.e, m $\varepsilon(1.. \alpha)$ is weighting factor said to be the possibilistic parameter. PCM is also based on initialization typical of other cluster approaches. The clusters do not have a lot of mobility in PCM techniques, as each data point is classified as only one cluster at a time rather than all the clusters simultaneously. Consequently, a suitable initialization is necessary for the algorithms to converge to nearly global minimum.

The characteristics of both fuzzy and possibilistic c-means approaches is incorporated. 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:

$$J_{FCPM}(U,T,V)=\sum_{i=1}^{c}\sum_{j=1}^{n}(\mu_{ij}^{m}+t^{n})\,d^{2}(X_j,v_i) \tag{9}$$

With the following constraints:

$$\sum_{i=1}^{c}\mu_{ij} < n, \forall i \in \{1,\ldots,n\}$$

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

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:

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

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

$$v_i=\frac{\Sigma_{k\ 1}^{n}(\mu_{ik}^{m}+\Sigma_{j}^{n}\,t_{ik}^{n})\,X_K}{\Sigma_{k=1}^{n}(\mu_{ik}^{m}+\Sigma_{j}^{n}\,t_{ik}^{n})}, 1\le i\le c \tag{12}$$

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 (Chen and Ye, 2004). 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:

$$\alpha=\exp\left[-\min_{i\ne k}\frac{\|v_i-v_k\|^2}{\beta}\right] \tag{13}$$

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

$$\beta = \frac{\Sigma_j^n = {}_1 I = | \left| x_j + \bar{x} \right| |^2}{n}$$

Where:

$$\bar{x} = \frac{\Sigma_j^n = x_j}{n}$$

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:

$$w_{ji} = \exp\left[- \frac{\| x_j - v_i \|^2}{[\Sigma_j^n = {}_1 \| x_j - \bar{v} \|^2] * c / n} \right] \quad (14)$$

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:

$$J_{MFPCM} = \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{ij}^m w_{ij}^m d^{2m}(x_j, v) + t_{ij}^n w_{ij}^n d^{2\eta}(x_j, v_i)) \quad (15)$$

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

$$\mu_{ij} = \left[\sum_{k=1}^{c} \left(\frac{d(X_j, v_i)}{d(X_j, v_k)} \right)^{2m/(m-1)} \right]^{-1} \quad (16)$$

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

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

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

$$v_i = \frac{\Sigma_j^n = {}_1(\mu_{ij}^m w_{ji}^m + t_{ji}^\eta) * X_j}{\Sigma_j^n = {}_1(\mu_{ij}^m w_{ji}^m + t_{ji}^\eta w_{ji}^\eta)} \tag{18}$$

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 Eq. 15. 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:

$$\frac{1}{2} v \sum_{x=1}^{n} \sum_{i-1}^{c} (\mu_{x,i}^m \ln a + t_{x,i}^\eta \ln a\beta) \tag{19}$$

Where:

$$\alpha_i = \frac{\Sigma_x^n = {}_1\mu_{x,i}^m}{\Sigma_x^n =_1 \Sigma_i^c = {}_1\mu_{x,i}^m}, i = 1, 2, \ldots\ldots c$$

$$\beta_x = \frac{\Sigma_i^c = {}_1 t_{x,i}^\eta}{\Sigma_x^n =_1 \Sigma_i^c = {}_1 t_{x,i}^\eta}, \ x = 1, 2, \ldots\ldots n$$

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

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

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

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

In the previous expression:

$$\varpi_i = v_i = \frac{\Sigma_k^n = {}_1(\mu_{ik}^m + t_{ik}^\eta) X_K}{\Sigma_k^n = {}_1(\mu_{ik}^m + t_{ik}^\eta)}, 1 \leq i \leq c$$

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

$$\frac{1}{2} t \sum_{x=1}^{n} \sum_{i=1}^{c} (\mu_{x,i}{}^m \tanh \alpha_i + t^\eta{}_{x,i} \tanh \beta_x) \tag{20}$$

where, membership $\mu_{x,i}$ and typicality $t_{x,I}$ for the compenzation is presented below:

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

$$x = 1, 2,..., n, i = 1, 2,...c$$

$$(t_{x,i})_c = \left(\sum_{y=1}^{c} \frac{(\| z_x - \varpi_i \|^2 - \tau \tanh(\beta_i))^{1/(\eta-1)}}{(\|_{zx} - \varpi_i \|^2 - \tau \tanh(\beta^i))^{1/(\eta-1)}} \right)^{-1}$$

$$x = 1, 2,..., n, i = 1, 2,...c$$

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 Eq. 21.

$$\begin{aligned} J_{MFPC} &= \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{x,i}{}^{m} w_{ji}{}^{m} \operatorname{in} \alpha_i + t_{x,i}{}^{\eta} w_{ji}{}^{\eta} d^{2\eta}(x_j, \nu_i)) \\ &- \frac{1}{2} v \sum_{x=1}^{n} \sum_{i=1}^{c} (\mu_{x,i}^{m}) \ln \alpha_i + t_{x,i}^{\eta} \ln \beta_x) \\ &- \frac{1}{2} v \sum_{x=1}^{n} \sum_{i=1}^{c} (\mu_{x,i}^{m}) \tanh \alpha_i + t_{x,i}^{\eta} \tanh \beta_x) \end{aligned} \tag{21}$$

The centroid of ith cluster is calculated in the similar way as the definition in Eq. 18. The final objective function is presented in Eq. 21.

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:

$$\begin{aligned} J_{MFPC} &= \sum_{i=1}^{c} \sum_{j=1}^{n} (\mu_{x,i}^{m}\, w_{x,i}^{m}\, d^2(x_j, \nu) + t_{x,i}^{\eta}\, w_{ji}^{\eta}\, d^{2\eta}(x_j, \nu_i)) \\ &- \frac{1}{2} v \sum_{x=1}^{n} \sum_{i=1}^{c} (\mu_{x,i}^{m} \ln \alpha_i + t_{x,i}^{\eta} \ln \beta_x) \\ &+ \frac{1}{2} v \sum_{x=1}^{n} \sum_{i=1}^{c} (\mu_{x,i}^{m} \tanh \alpha_i + t_{x,i}^{\eta} \tanh \beta_x) \\ &+ \Sigma_{i=1}^{c} \eta_i \Sigma_{k=1}^{n} (1 - u_{ik})^m + \gamma \Sigma_{i=1}^{c} \Sigma_{k=1, k\neq i}^{c} \frac{1}{d^2(v_i, v_k)} \end{aligned} \tag{22}$$

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

$$u_{ik} \in [0,1], \forall_i \tag{23}$$

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:

$$v_i = \frac{\Sigma_j^n{}_{=1} u_{ij} x_j - \gamma \Sigma_{k=1,k\neq i}\ v_k \dfrac{1}{d^2(v_k, v_i)}}{\Sigma_j^n{}_{=1} u_{ij} - \gamma \Sigma_{k=1,k\neq i} \dfrac{1}{d^2(v_k, v_i)}} \tag{24}$$

Singularity occurs when one or more of the distances $d^2(v_k, v_i)$ 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 (22) in order to minimize the objective function is given by:

$$\begin{aligned} J_{MFPCM} = & \sum_{i=1}^{c}\sum_{j=i}^{n}(\mu_{ij}^m w_{ij}^m d^{2m}(x_j v) + t_{ij}^\eta w_{ji}^\eta d^{2\eta}(x_j, v_i)) \\ & -\frac{1}{2} v \sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i}^m \ln \alpha_i + t_{x,i}^\eta \ln \beta_x) \\ & +\frac{1}{2} v \sum_{x=1}^{n}\sum_{i=1}^{c}(\mu_{x,i}^m \tanh \alpha_i + t_{x,i}^\eta \tanh \beta_x) \\ & +\Sigma_{i=1}^c \eta_i \Sigma_{k=1}^n (1-u_{ik})^m \\ & +\gamma \Sigma_{i=1}^c \Sigma_{k=1,k\neq i}^c e^{-d2(v_k, v_i)} \end{aligned} \tag{25}$$

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:

$$bw(u, j) = \begin{cases} 0, & g(i,j) \in C_1 \\ 255, & \text{otherwise} \end{cases} \tag{26}$$

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.

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.

Table 1 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 2 and 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 Fig. 1. Figure 1a indicates the original ultrasound image and Fig. 1b provides the segmented image in which the cancer regions are clearly visible.

Table 1: Segmentation accuracy comparison

	Accuracy (%)	
Ultrasound image	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 2: Standard deviation comparison

	Standard deviation	
Ultrasound image	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	0.0
6	4.6	0.0
7	4.8	2.1
8	4.1	1.2
9	5.6	0.0
10	4.7	3.1

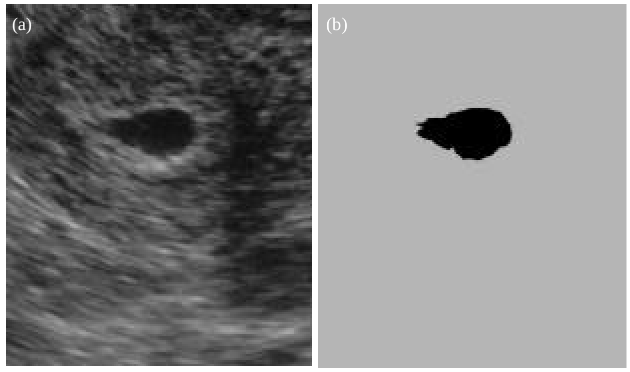

Fig. 1(a-b): (a) Ultrasound image and (b) Segmented ultrasound image

CONCLUSION

Mammography is the traditional screening instrument for breast cancer. It is responsive to early detection of disease but in several situations the results are not satisfactorily precise thus subsequent diagnostic work up is needed. Ultrasound is the chief alternative tool to mammography for diagnosis of suspicious determining. This study uses the Ultrasound images for detecting the cancer occurrence in breast. The noise from the image is removed using some image processing techniques. This will enhance the image for diagnosis. The next process is segmentation, in which the ultrasound images are segmented for separation of cancer regions from the normal region. This user uses modified fuzzy possibilistic c-means algorithm which includes the advantages of fuzzy and c-means clustering algorithm. Finally, the segmentation is enhanced using the repulsion factor. The experimental result indicates that the proposed technique results in better detection of cancer than the conventional techniques.

REFERENCES

ACS, 2008. Breast cancer facts and figures 2007-2008. American Cancer Society, USA. http://www.cancer.org/Research/CancerFactsFigures/breast-cancer-facts--figures-2007-2008.

Adam, D., S. Beilin-Nissan, Z. Friedman and V. Behar, 2006. The combined effect of spatial compounding and nonlinear filtering on the speckle reduction in ultrasound images. Ultrasonics, 44: 166-181.

Chen, C.Y. and F. Ye, 2004. Particle swarm optimization algorithm and its application to clustering analysis. IEEE Int. Conf. Networking Sensing Control, 2: 789-794.

Cheng, H.D., X. Cai, X. Chen, L. Hu and X. Lou, 2003. Computer-aided detection and classification of microcalcifications in mammograms: A survey. Pattern Recognit., 36: 2967-2991.

Deshmukh, K.S. and G.N. Shinde, 2005. An adaptive color image segmentation. Electron. Lett. Comput. Vision Image Anal., 5: 12-23.

Drukker, K., M.L. Giger, C.J. Vyborny and E.B. Mendelson, 2004. Computerized detection and classification of cancer on breast ultrasound. Acad. Radiol., 11: 526-535.

Gefen, S., O.J. Tretiak, C.W. Piccoli, K.D. Donohue and A.P. Petropulu *et al*., 2003. ROC analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis. IEEE Trans. Med. Imaging, 22: 170-177.

Li, X. and D.C. Liu, 2007. Ultrasound image enhancement using dynamic filtering. Proceedings of the 4th International Conference on Image and Graphics, August 22-24, 2007, Institute of Electrical and Electronics Engineers (IEEE), Chengdu, Sichuan, China, pp: 106-109.

Madabhushi, A. and D.N. Metaxas, 2003. Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions. Inst. Electr. Electron. Eng. Trans. Med. Imaging, 22: 155-169.

Noble, J.A. and D. Boukerroui, 2006. Ultrasound image segmentation: A survey. Inst. Electr. Electron. Eng. Trans. Med. Imaging, 25: 987-1010.

Oelze, M.L., W.D. O'Brien and J.F. Zachary, 2007. 11B-4 Quantitative ultrasound assessment of breast cancer using a multiparameter approach. Proceedings of the IEEE Ultrasonics Symposium, October 28-31, 2007, New York, pp: 981-984.

Shi, X., H.D. Cheng, L. Hu, W. Ju and J. Tian, 2010. Detection and classification of masses in breast ultrasound images. Digital Signal Process., 20: 824-836.

Wachs, J., O. Shapira and H. Stern, 2006. A method to enhance the possibilistic C-means with repulsion algorithm based on cluster validity index. Applied Soft Comput. Technol.: Challenge Complexity, 34: 77-87.

Winder, A.A., B. Jadidian and R. Muratore, 2010. Synthetic Structural Imaging (SSI): A new ultrasound method for tracking breast cancer morphology. Proceedings of the 39th Annual Ultrasonic Industry Association Symposium, April 12-14, 2010, Cambridge, MA., pp: 1-4.

Zhang, L.C., E.M.C. Wong, F. Zhang and J. Zhou, 2006. Adaptive pyramid filtering for medical ultrasound image enhancement. Proceedings of the 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, April 6-9, 2006, Arlington, VA., USA., pp: 916-919.