Combining Clustering, Morphology and Metaheuristic Optimization Technique for Segmentation of Breast Ultrasound Images to Detect Tumors

Prabusankarlal.K.M
Research and Development Centre
Bharathiar University
Coimbatore, India

Thirumoorthy.P
Department of Electronics and Communication
Government Arts College
Dharmapuri, India

Manavalan.R
Department of Computer Applications
K.S.R.College of Arts and Science, Tiruchengode, India

ABSTRACT
A framework which combines morphological operations and metaheuristic optimization technique with clustering method for the precise segmentation of breast tumours using ultrasound images is proposed in this study. Malignant tumours are pernicious when neglected to detect and treat at the earliest. Women with dense breasts are more prone to this malady and ultrasonagraphy is the suitable screening cum diagnosis method to aid the physician to estimate the amount of malignancy. This method is exclusively proposed for segmenting B- mode breast ultrasound images, characterized by low contrast and critically affected by speckle noise which hinders the finer details. The images are median filtered initially, in order to suppress the speckle noise and they are enhanced by a sticks algorithm based filter. The clustering is performed by FCM algorithm which is optimized by Particle swarm optimization. Automated morphological operations are performed on the clustered image as post processing procedure to improve the accuracy. To evaluate the proposed method, a database of 32 pathologically proven breast lesion images including 18 benign cysts and 14 malignant tumours is used. The segmented contours are compared with manually delineated contours and obtained MR of 93.24%, OF of 0.903 and EF of 0.1017. Moreover, the quantitative results are compared and analyzed with other existing methods and the values evidenced that the proposed method distinctly outperforms other methods.

General Terms
Medical Ultrasound, Image Segmentation, Tumor Detection.

Keywords
Breast Ultrasound, Segmentation, Clustering, Morphology, Particle swarm optimization.

1. INTRODUCTION
Breast cancer is the most common form of cancer among women worldwide. Early detection and diagnosis of breast cancer increases the treatment options for the medical practitioner which reduces the mortality rate and increase the survivability of patients [1][2].Early detection requires an accurate and reliable diagnosis method which could distinguish between benign and malignant tumors. A good detection approach should produce both low false positive (FP) rate and false negative (FN) rate [3]. Although mammography is currently the most effective tool for early detection of breast cancer, during screening, noncancerous lesions can be misinterpreted as a cancer (false-positive value), while cancers may be missed (false-negative value)[4] and this leads to unnecessary painful biopsy operations [5]. In addition to this, the dense breasts have the probability of cancer is four to six times higher than in nonsense breasts.

Mammography can hardly detect breast cancer in adolescent women with dense breasts. Moreover, the ionizing radiation of mammography is a health risk for the patients and radiologists. An important alternative to mammography is ultra- sound (US) imaging, and it shows an increasing interest in the use of ultrasound images for breast cancer detection and diagnosis [6]. Ultrasound is cheaper and faster than mammography suitable for the low-resource countries [7] and it is more sensitive than mammography for detecting abnormalities in dense breasts [8]. Ultrasound imaging becomes an important diagnostic tool for computer aided detection and diagnosis (CAD) systems have been developed to reduce the operator dependence and to increase the diagnosis sensitivity and specificity in breast cancer diagnosis [9]. The Breast ultrasound CAD system requires an efficient segmentation method in order to increase the efficiency, accuracy and productivity.

In the proposed method, a set of preprocessing techniques are used to remove speckle noise and to enhance the breast ultrasound images. The Median filtering which suppresses the speckle noise and sticks algorithm which enhances the contrast of lesion edges make the images more suitable for segmentation. The segmentation process employ FCM algorithm for clustering, the PSO algorithm optimize the clustering process. This is followed by a set of automated morphological operations as post processing which finally generate precise contour of the breast tumor.

2. MATERIALS AND METHODS
A database of 32 B-mode breast ultrasound images are used in this study, among them 18 are Benign cases including 5 Non mass lesions, 5 Infected cysts, 4 Fibro adenomas and 4 Ductal lesions. Malignant cases used are totally 14 which includes 4 carcinoma ,5 Inflammatory carcinoma and 5 Metastases images. All images are acquired using high end scanners over different periods of time [10] and have the dimension of 256X 256. The benign images are labeled as BEN_1 to BEN_18 and malignant images are tagged as MAL_1 to MAL_14.

2.1 Clustering based Image segmentation
Segmentation is a critical and essential component in processing breast ultrasound images and it is the most difficult task since in the BUS images number of characteristic artifacts such as attenuation, speckle, shadows and signal dropout which makes the segmentation task complicated[11]. The speckle noise [12] makes BUS images to have low signal/noise ratio (SNR), blurry boundaries and poor quality and also the shapes, sizes and locations of breast lesions have great variations [13]. This may result in missing

boundaries during segmentation and make the detection and diagnosis task less accurate.

Segmentation [3] is a partition of the image I into non-overlapping regions

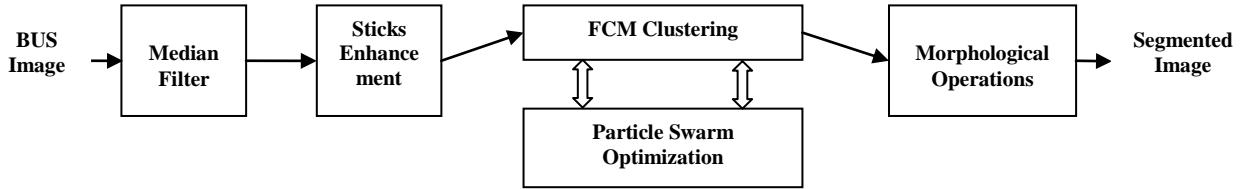$$\cup I_i = I \ and \ I_i \cap I_j = \emptyset, i \neq j \qquad (1)$$

The principle objective of any segmentation method is to locate the suspicious areas in the ultrasound image in order to assist physician in diagnoses.

The K means algorithm was formulated by Hartigan and Wong has been proven as a robust algorithm for clustering data. Also known as hard clustering, algorithm, the k-means finds a locally minimal solution through a simple iterative scheme [14]. Further, each element can belongs to only one cluster and if no data point is equidistant from two centers, the algorithm finally converge in to a local minimum for the distortion and it need not be a global minimum. The fuzzy theory was introduced by Lotfi Zadeh for clustering which assign data objects partially into multiple clusters and any

element need not present in only one. The closeness of data objects with their cluster centers defines the degree of membership in the fuzzy clusters [15].

The Fuzzy C-Means (FCM), introduced by [16] was an efficient tool for solving all fuzzy clustering related problems. It was proved as a steepest descent algorithm by [17] with global and local Convergence with variable step length. But, the problem is a combinatorial optimization problem [18] and if the fuzzy data sets contain severe noise points such as the image from sonography, the FCM often fails to find the global optimum and settle with the local optimum. To increase and ensure the highest probability of finding the global optimum, the evolutionary algorithms, inspired by the behaviors of natural systems such as swarm intelligence were developed [19]. Particle Swarm Optimization (PSO) is a population-based stochastic optimization technique, developed by Kennedy and Eberhart is highly dependent on stochastic processes based on swarm of particles flying through the search space [20].

The proposed framework for segmentation of breast ultrasound images is depicted in Fig 1.

Fig 1: The proposed segmented method for breast ultrasound image segmentation.

2.2 Preprocessing

The preprocessing of breast ultrasound image requires speckle removal as well as enhancement of images. A median filter with a window size of [8 X 8] is applied to smooth out speckle noise [21]. Since any method used for enhance the images enhance the speckle as well, a sticks filter [22] is used to reduce the speckles further and to improve the edge information which contains finer details of the contour of the tumor. The boundaries between tissue layers in US images will appear as all sorts of lines, the stick, itself a set of short line segments of variable orientation, is able to locally approximate the boundaries to suppress speckles as well as to enhance and highlight the edge information.

2.3 Clustering with Optimization:

The FCM clustering method has been applied with the preprocessed images this work, the hybrid PSO-FCM approach is used to improve the performance of FCM algorithm in order to avoid non convergence and to find the global optimum with medical breast ultrasound images which is normally affected by speckle noise and contain low contrast details of tumors. Features extracted from BUS images with inadequate segmentation and fed to classifiers will not provide robust and accurate diagnosis.

A swarm consists of M individuals [23], called particles, which change their position over time. Each particle represents a potential solution to the problem of optimization. In FCM, the problem of optimization is to minimize the value of the objective function. Let the particle k in a D-dimension space ($D = nc$) be represented as [23],

$$\mu k = \begin{pmatrix} \mu k11, \mu k12, \dots, \mu k1n, \mu k21, \mu k22, \dots, \\ \mu k2n, \dots, \mu kc1, \mu kc2, \dots, \mu kcn \end{pmatrix} \quad (2)$$

where k=1,2,…m

Let the objective function of FCM be the fitness function as follows:

$$J_{FCM}^m(U, A, X) = \sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^m \, d_{ij}^2 \quad (3)$$

$$\sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^m \, \|x_{j-}a_i\|^2 \quad (4)$$

Where, $a_i = \left(\sum_{j=1}^{n}[\mu_{ij}]^m\right)^{-1}\left(\sum_{j=1}^{n}[\mu_{ij}]^m x_j\right)$, $i = 1,2,\dots,c$.

The best previous position which possesses the best fitness value of particle k was denoted by $Pk = (Pk_1, Pk_2, \dots, Pk_D)$ which is also called P_{best}. The index of the best P_{best} among all the particles is denoted by the symbol g.

The best fitness value of the position $Pk = (Pk_1, Pk_2, \dots, Pk_D)$ is defined as g_{best}. The velocity for the particle k is represented as, $Vk = (Vk_1, Vk_2, \dots, Vk_D)$.

The P_{best} and g_{best} location for iteration t according to following two formulas,

$$vk(t + 1) = wvk(t) + c1r1\big(Pbest(t) - \mu k(t)\big) \\ + c2r2\big(gbest(t) - \mu k(t)\big) \quad (5)$$

$$\mu k(t+1) = normalized\left[\mu k(t)\right.$$
$$\left. + vk(t+1)\right] satisfying$$

$$\mu_{kij}(t+1) = \frac{\mu'_{kij}(t+1)}{\sum_{i=1}^{c}\mu'_{kij}(t+1)} \quad (6)$$

$$\forall i = 1,2,\dots,c, j = 1,2,..,n, k = 1,2,..m$$

where, $\quad \mu'_{kij}(t+1) = \frac{U_{kij}(t) - \min_{1\le j\le n}U_{kij}(t)}{\max_{1\le j\le n}U_{kij}(t) - \min_{1\le j\le n}U_{kij}(t)}$

$$U_{kij}(t) = \mu_{kij}(t) + v_{kij}(t+1) \quad (7)$$

where w is the inertia coefficient which is a constant in the interval [0,1], and can be adjusted in the direction of linear decrease, ($w=0.75$); $c1$ and $c2$ are learning rates which are nonnegative constants.(In our method, $c1 =2$ and $c2 =2$);$r1$ and $r2$ are generated randomly in the interval [0,1].

The termination criterion for iterations is calculated according to whether the maximum generation or a assigned value of the fitness is reached. In this paper, the given converging error is ε =0.001.

$$\max_{1\le j\le n}\|a_i(t+1) - a_i(t)\| < \varepsilon \quad (8)$$

where,

$$a_i(t+1) = \frac{\sum_{j=1}^{n}[U_{ij}(t)]^m x_j}{\sum_{j=1}^{n}[U_{ij}(t)]^m} , i = 1,2,\dots,c$$

The segmented images show the precise contour of the tumor and in this work, the image is treated with automated morphological operations as post processing task in order to clear the boundaries of the BUS images.

2.4 Morphological Operations

The standard mathematical morphological operators are dilation and erosion and the other morphological operations are derived from the combination of these two operations [24].

Let $F(x, y)$ is a grey-scale 2D image and B is the structuring element SE, the dilation of an image is given by,

$$F \oplus B(x, y) = \max\{F(x - s, y - t) + B(s, t)\} \quad (9)$$

where, $B(s, t)$,is the SE. The erosion operation is given by,

$$F \ominus B(x, y) = \min\{F(x + s, y + t) - B(s, t)\} \quad (10)$$

The combination of opening and closing operations is given by,

$$F \circ B = (F \ominus B) \oplus B \quad (11)$$

$$F \cdot B = (F \oplus B) \ominus B \quad (12)$$

The combination of morphological opening and closing is used to clear the boundary of the images and preserve the contour of the segmented tumour from any possible noises present in the image.

3. RESULTS AND DISCUSSION

The proposed is tested on a database of 32 breast ultrasound images. In order to validate the experimental results, the lesions are manually delineated by an experienced radiologist. The manually delineated lesion boundaries are compared with the automatically generated lesion boundaries qualitatively and quantitatively.

3.1 Qualitative evaluation

The method is applied on the breast ultrasound images of the entire database. The images shown in the Fig 2 and 3 are obtained at each stage of the method. The Fig 2 shows a benign image (BEN_4) at different stage which is characterized by smooth contours with regular shape, whereas a malignant image (MAL_7) with irregular shape with rough contour is shown in Fig 3. The Fig (a) shows original speckle affected image, (b) shows the speckle suppressed, edge enhanced version of the same image obtained through our preprocessing method. The segmented image is shown in (c), the area other than the tumor is cleared using morphological operations in (d). The manually delineated image by the radiologist is shown in (e) and automatically segmented image with outline is shown in (f).
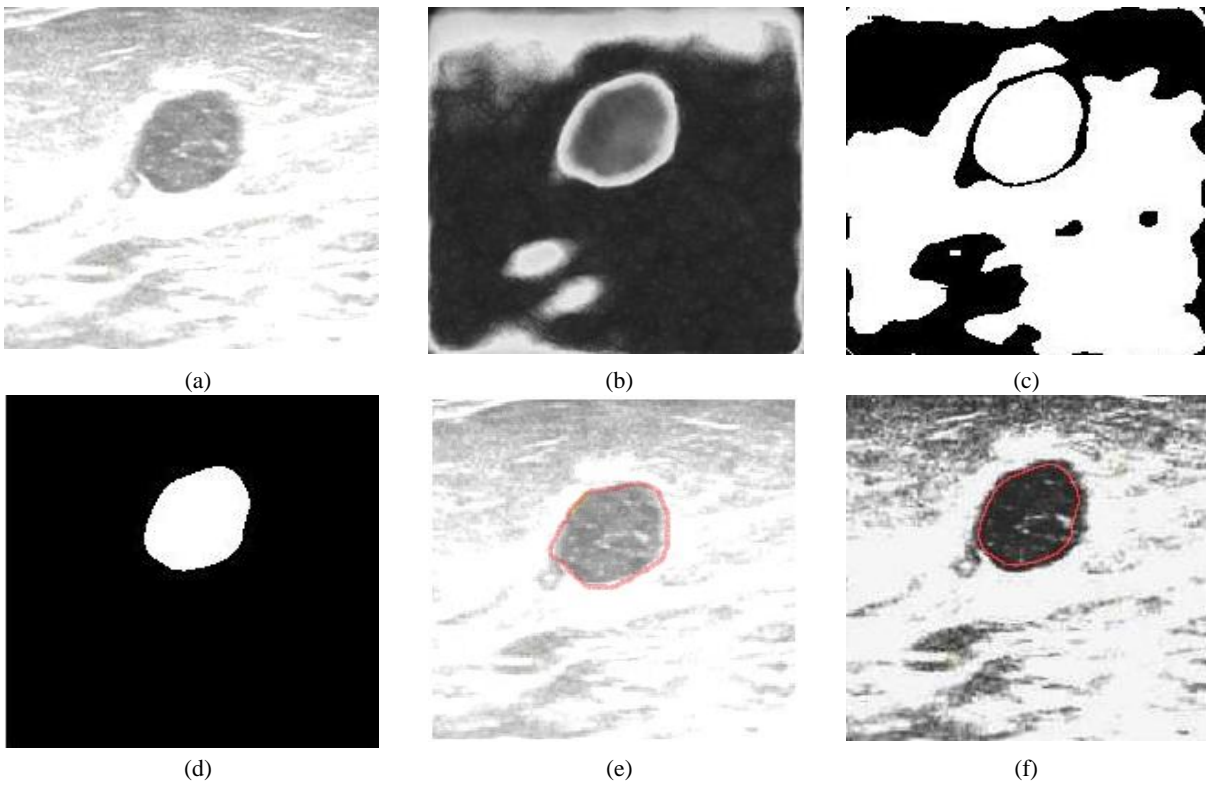
The qualitative comparison between the manually delineated and computer generated segmented images demonstrated that the proposed method produced precise tumor boundaries for both benign and malignant cases.
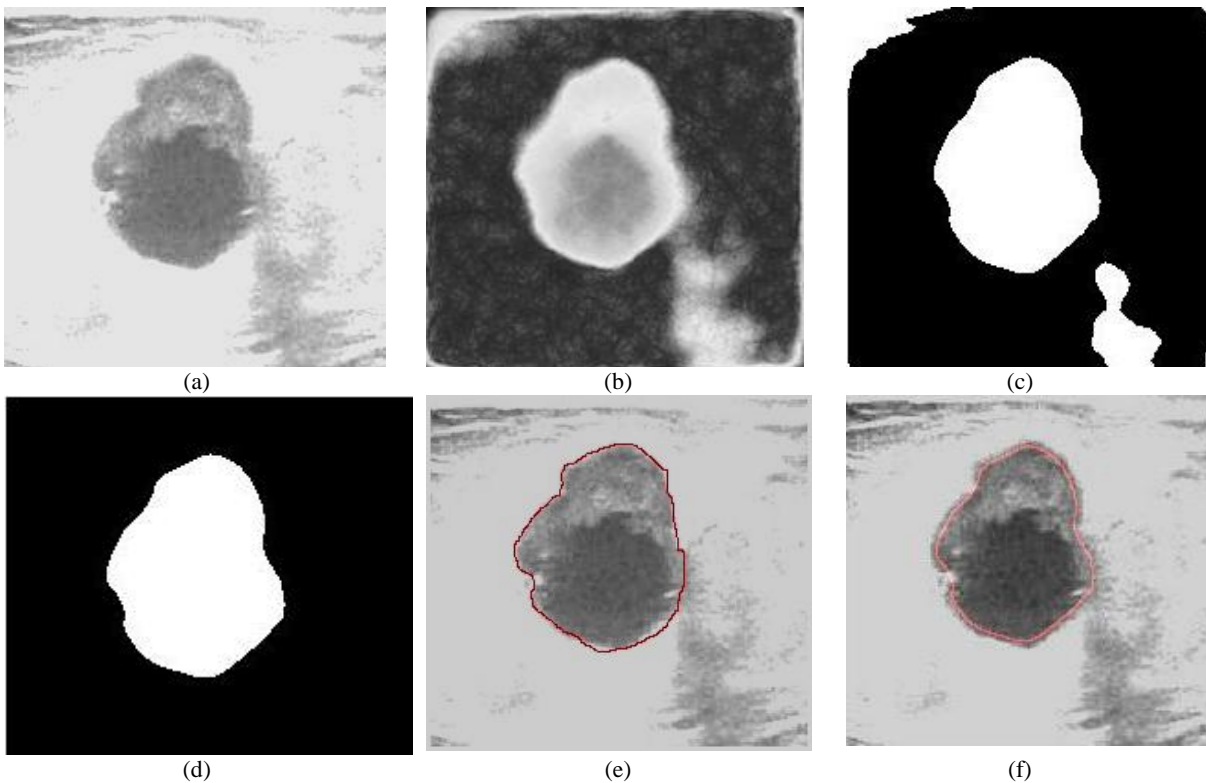
3.2 Quantitative evaluation

Lack of a standard database with gold standard breast ultrasonic images is not yet available anywhere and this makes for quantitative evaluation under a common measure, very difficult [3].

The manual segmentations obtained from expert observers cannot be considered as gold standard when observer bias and interobserver as well as intraobserver variability are factored [13].

However, in most state of the art methods in the literature, the segmentation performance is evaluated by comparing radiologist's delineated boundary with computer generated boundary. Area or boundary error metrics are utilized as quantitative evaluation measure for segmentation. They numerically indicate the similarity or dissimilarity between the manually segmented lesions and automatically segmented lesions.

(a)

(b)

(c)

(d)

(e)

(f)

Fig 2. Segmented results for benign image (BEN_4). (a) Original image.(b) Preprocessed image (with median filter[8X8] and sticks method(n=3 , k=1)).(c) Image obtained through segmentation method.(d) Final segmented image.(e) Delineated image by radiologist.(f) Automatic outlined image by the method

(a)

(b)

(c)

(d)

(e)

(f)

Fig 3. Segmented results for malignant image (MAL_7). (a) Original image.(b) Preprocessed image (with median filter[8X8] and sticks method(n=3 , k=1)).(c) Image obtained through segmentation method.(d) Final segmented image.(e) Delineated image by radiologist.(f) Automatic outlined image by the method.

3.3 Evaluation Metrics:

We have used three area error metrics namely, Match Rate (MR) [9], Overlap Fraction (OF) and Extra Fraction (EF) [25]. The Match rate, Overlap Fraction and Extra fraction is given as,

$$MR = \left(1 - \frac{|A_M - A_W|}{A_M}\right) X\ 100\% \quad (13)$$

$$OF = \frac{(A_M \cap A_W)}{A_M} \quad (14)$$

$$EF = \frac{(\overline{A_M} \cap A_W)}{A_M} \quad (15)$$

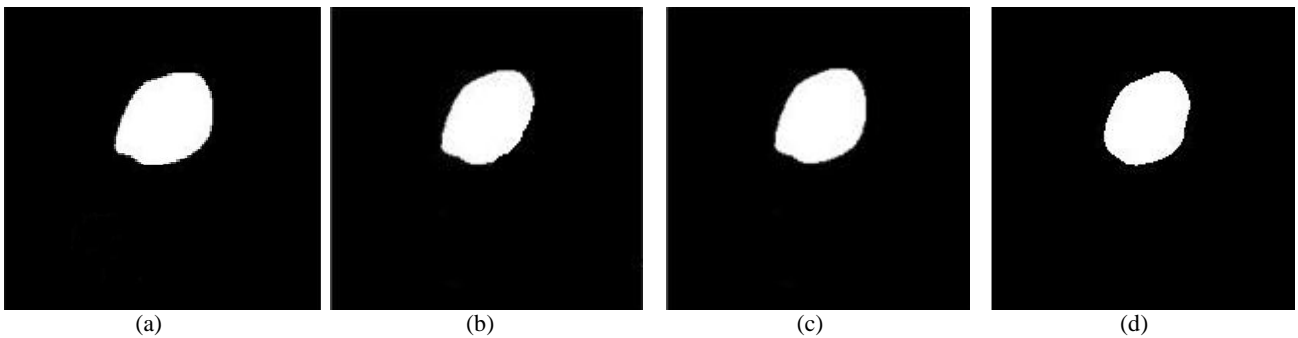Where, A_M is the area of manual delineated contour and A_W is the area of automatically generated contour.
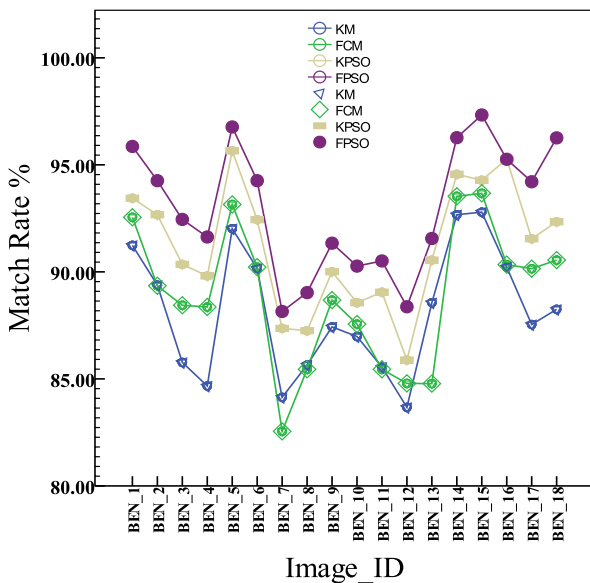
3.4 Comparative analysis:

A Quantitative comparative analysis is also performed by comparing the performance of our proposed framework with some other methods such as k- Means clustering, FCM clustering and PSO Optimized k- Means Clustering. The Fig 4 shows the segmented results of the same images by the three methods and our method. The k-Means segmentation is shown in (a), FCM segmentation is shown in (b), the (c) shows optimized k-Means and segmented image using our method is shown in (d). The Table 1 shows the evaluation results produced by the methods compared. As shown, the algorithms have been run for almost 20 runs and the values given are mean values. The MR measures the similarity between the two contours. The average MR value obtained is 93.24%, which is higher than the values produced by other methods. The OF is a measure of how correctly the segmentation is done with respect to reference image. A value close to 1 corresponds to degree of closeness between computer generated contour and manually delineated contour. The average value obtained by our method for the entire database is 0.903. Finally, the EF indicates the falsely segmented area as lesion in relative to the reference area. A very low average EF value is obtained by our method (0.1017).
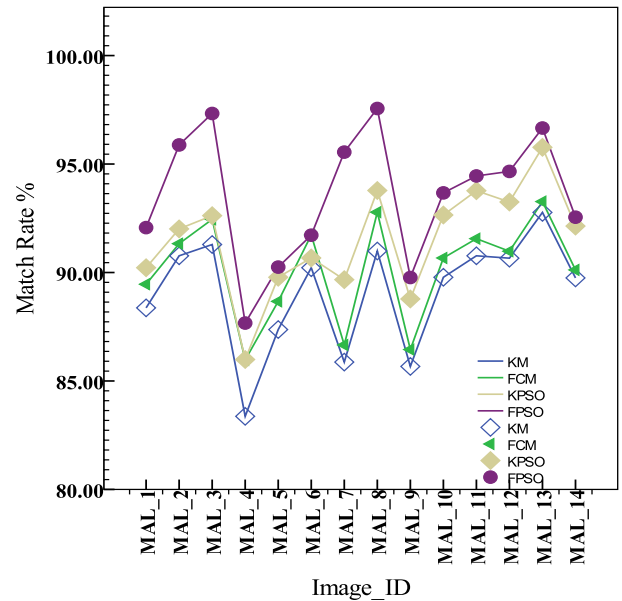
The curves in Fig 5 and Fig 6 depict the comparative Match Rate (MR) values for all individual benign and malignant images in the database. The Fig 7 shows Overflow Fraction (OF) and Extra Fraction (EF) values obtained by the proposed method for benign cases which ranges from BEN_1 to BEN_18. The OF and EF values of cases MAL_1 to MAL_14 are graphically depicted in Fig 8.

(a)   (b)   (c)   (d)

Fig 4. Comparison of segmentation methods.(a) k-Means.(b) Fuzzy C Means.(c) Optimized k-Means.(d) Proposed Method.
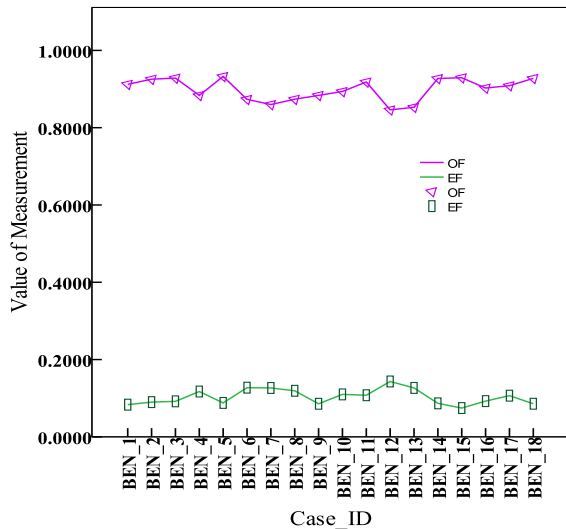

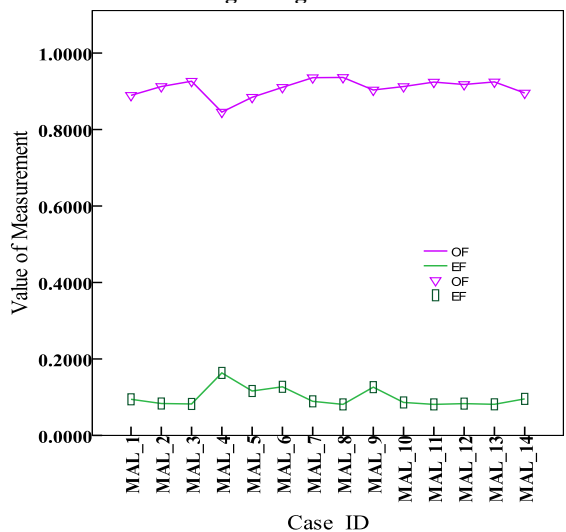**Fig 5. Comparison of Match Rate(MR) for Benign images**

Fig 6. Comparison of Match Rate(MR) for Malignant images

Table 1. Comparative analysis of segmentation methods(values shown are mean values)

Method	MR	OF	EF
k-Means Clustering	88.57 %	0.8422	0.1752
FCM Clustering	89.43 %	0.8562	0.1532
PSO Optimized k-Means	91.32 %	0.8823	0.1246
Proposed Method	93.24 %	0.9030	0.1017

Fig 7. Overlap Fraction (OF) and Extra Fraction (EF) for Benign images

Fig 8. Overlap Fraction (OF) and Extra Fraction (EF) for Malignant images

4. CONCLUSION

A method for precise segmentation of breast ultrasound images which combines FCM clustering with Particle swarm optimization and mathematical morphology is presented in this paper. A database of 32 breast ultrasound images is used in this study for validating the performance of the proposed method. Area metrics namely, Match Rate, Overlap Fraction and Extra Fraction are utilized to evaluate the method quantitatively. The results are also comparatively analyzed with three other clustering based segmentation methods, such as k-Means, FCM and PSO optimized k-Means. The results manifested that the proposed algorithm outperformed other algorithms and it is more appropriate for computer aided breast cancer detection and diagnosis systems.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Lee.C,(2002),Screening mammography: proven benefit, continued controversy. Radiologic Clinics of North America. 40(3), 395-407.

[2] Jalalian.A, (2013) Computer-aided detection/ diagnosis of breast cancer in mammography and ultrasound: a review, Clinical Imaging 37 (2013) 420–426

[3] Cheng, H. D., Shan, J., Ju, W., Guo, Y., Zhang, L. (2010). Automated breast cancer detection and classification using ultrasound images: A survey. Pattern Recognition, 43(1), 299-317.

[4] Bird RE, Wallace T, Yankaskas B. 1992.Analysis of cancers missed at screening mammography. Radiology 184(3):613-7.

[5] Jesneck.J, Lo.J, Baker.J, (2007) .Breast mass lesions: computer-aided diagnosis models with mammographic and sonographic descriptors, Radiology 244 (2) 390–398.

[6] Shankar.P, Piccoli.C, Reid.C, Forsberg.J, Goldberg.B, (2005) Application of the compound probability density function for characterization of breast masses in ultrasound B scans, Physics in Medicine and Biology 50 (10) 2241–2248.

[7] Anderson. B, Shyyan. R, Eniu. A, Smith. R, Yip. C, (2006) Breast cancer in limited- resource countries: an overview of the breast health global initiative 2005 guidelines, The Breast Journal 12 S3–15.

[8] Drukker.K, Giger.M.L, Horsch.K, Kupinski.M.A, Vyborny.C.J, Mendelson.E.B, (2002) Computerized lesion detection on breast ultrasound, Medical Physics 29 (7) 1438–1446.

[9] Huang.Y.L, ChenD.R, Liu.Y.K, (2004), Breast cancer diagnosis using image retrieval for different ultrasonic systems, in: International Conference on Image Processing, vol. 5, 2598–2960.

[10] 'Ultrasoundcases',http://ultrasoundcases.info/category.aspx?cat=67 , Accessed October 2013

[11] Bridal.S.L, Correas.J.M, Saied.A, Laugier.P, (2003), Milestones on the road to higher resolution, quantitative,

and functional ultrasonic imaging, Proceedings of the IEEE, vol. 91, no. 10, 1543–1561.

[12] Wagner.R.F, Smith. S.W, Sandrik. J.M, Lopez. H, (1983) Statistics of speckle in ultrasound B-scans, IEEE Trans. Sonics Ultrason. 30 ,156–163.

[13] Madabhushi.A, Metaxas.D.N, (2003) Combining low-, high-level and empirical domain knowledge for automated segmentation of ultrasonic breast lesions, IEEE Trans. Med. Imag. 22 ,155–169.

[14] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C., Silverman, R., Wu, A. Y. (2000). The analysis of a simple k-means clustering algorithm. Proc. of sixteenth ACM annual symposium on Computational geometry(pp. 100-109).

[15] Izakian, H., Abraham, A., Snasel, V. (2009). Fuzzy clustering using hybrid Fuzzy C-Means and fuzzy particle swarm optimization. Proc. of IEEE World Congress on Nature & Biologically Inspired Computing, pp. 1690-1694.

[16] Bezdek, J. C., Ehrlich, R., Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. Computers & Geosciences, 10(2), 191-203.

[17] Groll, L., Jakel, J. (2005). A new convergence proof of Fuzzy C-Means. Fuzzy Systems, IEEE Transactions on, 13(5), 717-720.

[18] Mehdizadeh, E., Sadi-Nezhad, S., Tavakkoli-Moghaddam, R. (2008). Optimization of fuzzy clustering criteria by a hybrid PSO and Fuzzy C-Means

[19] Kennedy.J.F, Kennedy.J, Eberhart.R.C,(2001) Swarm intelligence, Acdemic Press.

[20] Bilal .M. Zahran , Kanaan.G,(2009), Text Feature Selection using Particle Swarm Optimization Algorithm, World Applied Sciences Journal 7 (Special Issue of Computer & IT): 69-74.

[21] Eng, H. L., . Ma, K. K. (2001). Noise adaptive soft-switching median filter. IEEE Transactions on Image Processing, , 10(2), 242-251.

[22] Czerwinski.R.N, Jones.D.L, O'Brien.W.D, (1999),Detection of lines and boundaries in speckle images—Application to medical ultrasound, IEEE Trans. Med. Imag., vol. 18, no. 2, pp.126–136.

[23] Yih, J. M., Lin, Y. H., Liu, H. C. (2007). Clustering Analysis Method based on Fuzzy C-Means Algorithm of PSO and PPSO with Application in Real Data.International Journal of Geology, 4(1).

[24] Qian Z. Y, Hua G. W, Cheng C. Z, Tian T. J, Yun L. L, (2005).Medical Images Edge Detection Based on Mathematical Morphology, Proc of IEEE Engineering in Medicine and Biology 27th Annual Conference, 6492-6495.

[25] Huang, Y. L., Jiang, Y. R., Chen, D. R., & Moon, W. K. (2007). Level set contouring for breast tumor in sonography. Journal of digital imaging, 20(3), 238-247.

clustering algorithm. Iranian Journal of Fuzzy Systems, 5(3).