Designing an Algorithm for Cancerous Tissue Segmentation Using Adaptive K-means Cluttering and Discrete Wavelet Transform

Rezaee Kh[1], Haddadnia J[2]*

[1]Biomedical Engineering Student, Biomedical Engineering Group, Department of Electrical and Computer Engineering, Hakim Sabzevari University of Sabzevar, Sabzevar, Iran.
[2]Biomedical Engineering Group, Department of Electrical and Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran.

ABSTRACT

Background: Breast cancer is currently one of the leading causes of death among women worldwide. The diagnosis and separation of cancerous tumors in mammographic images require accuracy, experience and time, and it has always posed itself as a major challenge to the radiologists and physicians.

Objective: This paper proposes a new algorithm which draws on discrete wavelet transform and adaptive K-means techniques to transmute the medical images implement the tumor estimation and detect breast cancer tumors in mammograms in early stages. It also allows the rapid processing of the input data.

Method: In the first step, after designing a filter, the discrete wavelet transform is applied to the input images and the approximate coefficients of scaling components are constructed. Then, the different parts of image are classified in continuous spectrum. In the next step, by using adaptive K-means algorithm for initializing and smart choice of clusters' number, the appropriate threshold is selected. Finally, the suspicious cancerous mass is separated by implementing the image processing techniques.

Results: We Received 120 mammographic images in LJPEG format, which had been scanned in Gray-Scale with 50 microns size, 3% noise and 20% INU from clinical data taken from two medical databases (mini-MIAS and DDSM). The proposed algorithm detected tumors at an acceptable level with an average accuracy of 92.32% and sensitivity of 90.24%. Also, the Kappa coefficient was approximately 0.85, which proved the suitable reliability of the system performance.

Conclusion: The exact positioning of the cancerous tumors allows the radiologist to determine the stage of disease progression and suggest an appropriate treatment in accordance with the tumor growth. The low PPV and high NPV of the system is a warranty of the system and both clinical specialists and patients can trust its output.

Keywords

Breast cancer, Winner filter, Discrete wavelet transform, K-means clustering, Edge detection

*Corresponding author: J Haddadnia Biomedical Engineering Group, Department of Electrical and Computer Engineering, Hakim Sabzevari University, Sabzevar, Iran. Khorasan Razavi-Sabzevar-Center for Research of Medical Technologies Sabzevar University of Medical Sciences, Sabzevar, Iran. E-mail: haddadnia@hsu.ac.ir

Introduction

Medical images are the perfect tools for viewing tissue problems. However these images involved with problems such as complex shape structures, weakness details, non-homogeneous in brightness and poor contrast. After lung cancer, breast cancer is the second cause of death in the world especially among women and one out of seven women loses their life due to this disease [1]. In Iran, this type of cancer is the main cause of death among women and one fifth of women

mortality is related to breast cancer [2]. The number of women suffering from breast cancer in Iran has been on rise, and compared to the Western countries, the disease age has dropped by 10 years [3]. Breast cancer is the type of cancer in which cell growth and change goes out of control. The benign part of breast tissue is made of milk-producing glands, the pores attached to it and the nipple. The other parts of breast are made of lymphoid tissue and fat [4]. Mammography is currently one of best imaging methods to detect cancerous tumors in breasts. The advantages of digital mammography over the classical method are obvious including the higher chance of detecting cancerous tumor [5]. Given the complexity of the mammographic images and similarities between cancerous tissues and normal breast tissues, sometimes even radiologists or experienced physicians may have various interpretations about the cancerous tumors [6,7]. For example, mammographic method fails to detect 30% of breast cancer due to the imprecise positioning of tumor [8]. Given the high percentage of false positive in mammographic images, which is due to contrast dynamics and light intensity, the need for an accurate and sensitive system to detect the position, size and form of cancerous tumors is urgently felt.

Related Works

With the increasing detection of abnormal tissues by means of computers, researchers, relying on features such as shape, size and position of the tumor and its segmentation, seek to assist radiologists. Shen et al. [9] used morphological techniques and algorithm for detection of cancerous tumors to process mammogram image. Moreover, Woods et al. [10] used KNN algorithm for classification of different parts of mammographic images. In 2000, Verma and Zakos proposed CAD system for detection and classification of masses and tumors in digital images [11]. Sharma et al. [12] classified different parts of mammographic

images using watershed technique, which is commonly used for topography contour. Cascio et al. were also able to locate cancerous masses by means of shape structure searching [13]. In their method, neural network was also used to classify the type of malignancy and lesion in breast cancer. In 2008, Robottino et al. [14] using topography contour, were able to separate various masses of breast cancer in mammographic images. The extraction of tissue of breast cancer by means of the algorithm used for dividing different areas of the image based on contrast and intensity of mammographic images was carried out by Bankman et al. [15]. However, few methods have been proposed for segmentation of cancerous tumors based on different image processing techniques and K-means clustering so far. In most methods, including clustering and the methods based on pixel property of the images, the detection system is slow and time consuming. Clustering is one of the simplest, and has been widely used in segmentation of grey level images. Clustering techniques have received attention in many fields of study such as engineering, medicine, biology and data mining [16]. Also researchers use feature extraction method based on wavelet analysis. Wavelet transforms have been shown to be good tools for extracting information about texture [17]. Computer-aided classification system was used to help in diagnosing abnormalities faster than traditional screening program without the drawback attribute to human factors [18].

Some methods such as ROI are imprecise in extraction of cancerous masses. Despite having an acceptable performance, other methods suffer from shortcomings such as inappropriate selection of threshold, low sensitivity, high FP errors, loss of image data during edge detection, variability in position, size and shape of masses and unreliability in detection of cancerous tumors.

Material And Methods

In this paper, an algorithm based on adaptive K-means and discrete wavelet transform, the selection of appropriate threshold for partial segmentation of the image and edge detection technique has been used. Figure 1 shows the entire process. Using 120 sample mammographic images taken from mini-MIAS and DDSM databases [19,20] we tested three factors of accuracy (AC), specificity (SP) and sensitivity (SE). The images taken from DDSM database were in LJPEG format, scanned in Gray-Scale with 50 microns size. The resolution of these images was 200 micrometers. Moreover, the images taken from the mini-MIAS database were characterized by 8-bit depth, 200 mi-
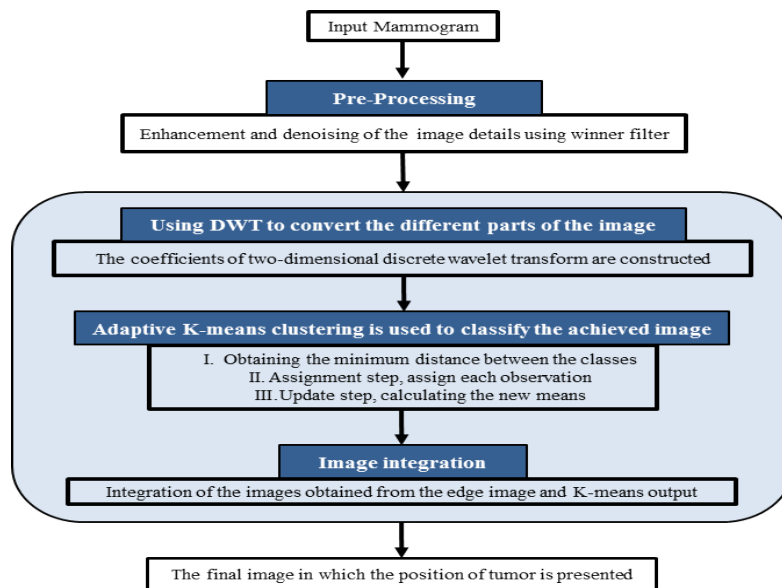

Figure 1: Implementation of the proposed algorithm

crometers resolution and 1024×1024 dimensions. The image dimension was changed into 256×256 pixels.

Pre-Processing Step

Medical images provide valuable information about masses, especially when the mass is not centralized in contour tissue. In many image-processing techniques, De-noising is used as a preprocessing step, and analysis tasks such as registration or segmentation are used to reduce the random noise produced in the acquisition process [21]. Images from various modalities need to be De-noising as a pre-processing step for many planning, navi-gation, detection, data-fusion and visualization tasks in medical applications. CT slices, for example, are often corrupted by the noise produced during acquisition or transmission [22]. Radiologists need a noise free image to detect the place and position of the masses. Median filter is a non-linear filter that is used to eliminate impulse noises.

Salt-and-pepper noise appears when there is an inappropriate function of glass sensor cells, memory cells errors and the synchronized errors at the point of digitization or data transition. The main function of this filter is to use the neighbor pixels of the target pixel during which a mask pads all pixels. While a pixel

is processed, all neighbor pixels are organized based on intensity of pixels ascend. Then, the intensity of median element of the organized numbers is selected as a new pixel. Wiener fil-

ter is used to remove the blurring and noise from images with high level of obscurity. Frequency response of this filter is calculated according to (1):
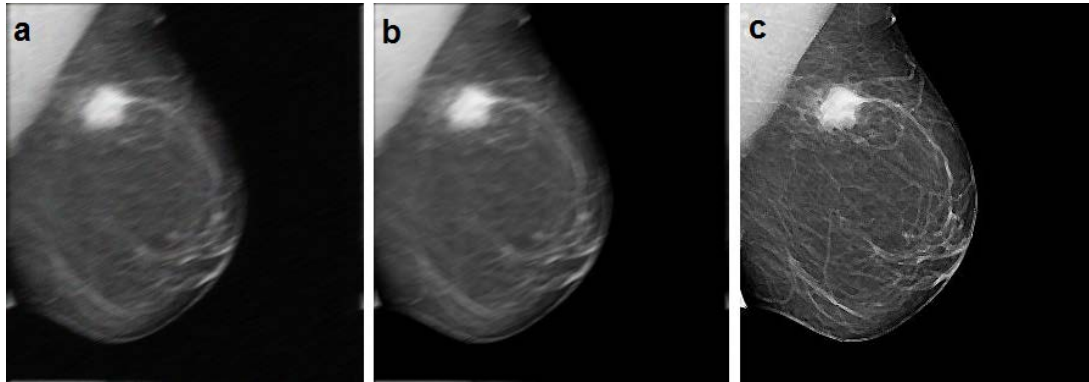

Figure 2: (a) Original image with 8% noise and 36% blurring, (b) the output of winner filter without noise and (c) the blurred image after reducing the noise.

$$\tilde{F}(u,v) = \frac{\tilde{B}*(u,v)}{\left|\tilde{B}(u,v)\right|^2 + \eta} \qquad (1)$$

Where η is the noise power, N is the white noise and $\tilde{B}*(u,v)$ is the complex conjugate of $\tilde{B}(u,v)$. Thus, the component of the Discrete Fourier Transform (DFT) of the B variable is $B(u,v)$. In figure 2 the pre-processing step is applied on mammogram sample which has 8% noise and 36% blurring.

Discrete Wavelet Transform

Wavelet transform is defined both for discrete Fourier transform and continuous transform. Having numerous applications, these transforms, occur in one-dimensional and two-dimensional spaces.

In the two-dimensional discrete wavelet transform, an image treated as a matrix, each row or column is considered as a signal that its amplitude is the brightness of pixels in that certain row or column. With the application of one-dimensional discrete wavelet to each row, and fixation of columns, parallel signals will

be achieved. Then, assuming the constancy of rows, this procedure is applied to all columns again and as a result, the rows with a rate of 2 are sampled [23]. With the break of two-dimensional discrete wavelet transform into four components of approximate coefficients at j level, it can be inferred that the approximation is transformed into j+1 level and three other levels, i.e. horizontal, vertical and diagonal levels. To obtain the two-dimensional discrete wavelet transform, we first need to build the wavelet series expansion function and the scaling function according to (2) [24]:

$$f(n) = \sum_k c_{j_0}(k)\varphi_{j_{0,k}}(x) + \sum_{j=j_0}^{\infty}\sum d_j(k)\psi_{j,k}(x) \quad (2)$$

Where is j_0 the arbitrary starting scale, $Cj_0(k)$ is scaling coefficients or approximation and $d_j(k)$ is the partial wavelet coefficients of the wavelet. The expanded coefficients are calculated according to (3) and (4) as follows:

$$C_{j_0}(k) = \left\langle f(x), \tilde{\varphi}_{j_{0,k}}(x)\right\rangle = \int f(x)\tilde{\varphi}_{j_{0,k}}(x)dx \quad (3)$$

$$d_j(k) = \left\langle f(x), \tilde{\psi}_{j,k}(x)\right\rangle = \int f(x)\tilde{\psi}_{j,k}(x)dx \quad (4)$$

If the function is expanded to a series of numbers, then it would resemble the $f(x)$ continuous form. The resulted discrete wavelet transfo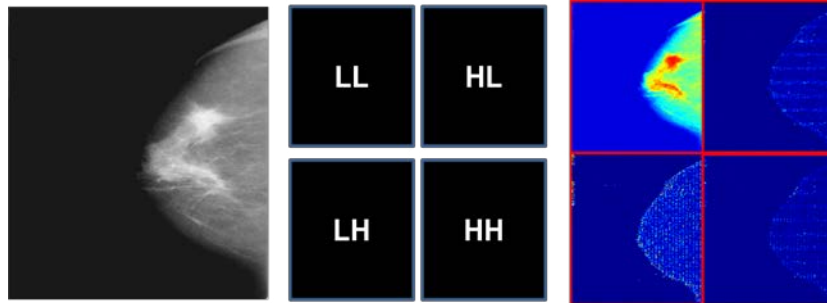rm coefficients are called $f(x)$ .Thus, with the expansion of this series, the final transform of discrete wavelet using (5), (6) and (7) would be:

Figure 3: The two-dimensional discrete wavelet transform in four directions for mammographic image.

$$W_{\varphi}\left(j_{0,k}\right) = \frac{1}{\sqrt{M}} \sum_{x=0}^{M-1} f(x)\tilde{\varphi}_{j_{0,k}}(x) \qquad (5)$$

$$W_{\psi}(j,k) = \frac{1}{\sqrt{M}} \sum f(x)\tilde{\psi}_{j,k}(x) \qquad (6)$$

And for $j>j_0$

$$f(x) = \frac{1}{\sqrt{M}} \sum W_{\phi}\left(j_{0,k}\right)\phi_{0,k}(x) + \frac{1}{\sqrt{M}} \sum_{j=j_0}^{\infty} \sum_k W_{\psi}(j,k)\psi_{j,k}(x) \quad (7)$$

In two-dimensional space, the two-dimensional scaling wavelet function $\varphi(x,y)$ and three two-dimensional wavelet functions including $\psi^v(x,y)$ and $\psi^h(x,y)$ are required. If an image is decomposed into n levels of the basic functions, in that case, its stage 1 involves high-frequency components while its n stage involves low frequency component. Discrete wavelet transform coefficients will be decomposed according to the frequency range of each basic function, carrying the local data of the original image. In figure 3, the two-dimensional discrete wavelet transform in four directions for mammographic image is shown.

Adaptive K-means clustering

Unsupervised learning is closely associated with the problem of density estimation in statistics [25]. The clustering process is classified among unsupervised techniques based on the fact that samples in the data space lack any information regarding their belonging to a specific cluster [26]. The K-means algorithm is an iterative algorithm that is used to partition an image into K clusters. The objective function was to obtain the minimum distance between the classes, or basically between the image pixels. Let $X_n = (x_1, x_2, ..., x_n)$ where each component is a d-dimensional real vector and this algorithm attempts to partition the n components (or observations) into k sets $S = \{S_1, S_2, ..., S_k\}$ with k<n, so as to minimize the within cluster sum of squares in (8):

$$\underset{S}{\arg\min} \sum_{i=1}^{k} \sum_{X_j \in S_j} \|X_i - \mu_i\|^2 \qquad (8)$$

where μ_i is the mean points in S_i. Given an initial set of k means $m_1^{(1)}, m_2^{(1)}, ..., m_k^{(1)}$, which may be specified randomly, and according to Hamerly et al [27], the Random Partition method is generally preferable for algorithms such as the k-harmonic means and fuzzy k-means to initialize the parameters. For expectation maximization and standard k-means al-

gorithms, the Forgy method of initialization is preferable. The algorithm proceeds by alternating between two steps [28]:

Assignment step: we assign each observation to the cluster with the closest mean and define it as (9):

$$S_i^{(t)} = \{x_p : \left\| x_p - m_i^{(t)} \right\| \leq \left\| x_p - m_j^{(t)} \right\| \forall 1 \leq j \leq k\} \quad (9)$$

Where each x_p is assigned to exactly one $S^{(t)}$.

Update step: the new means to be the centroids of the observations in the new clusters is calculated as follows (10):

$$m_i^{(t+1)} = \frac{1}{\left| S_i^{(t)} \right|} \sum_{x_j \in S_i^{(t)}} x_j \quad (10)$$

Image pixels are classified based on the brightness intensity. The frequency of brightness intensity of images (histogram) has been used to select the appropriate number of secretions. Based on this classification, pixels can be divided into a maximum of 255 clusters. Here, based on results obtained from Ray and Salvador, the number of proposed clusters was between 4 and 7. In image segmentation, the choice of K is not critical because the adaptation of the intensity functions allows the same region type to have different intensities in dif-

ferent parts of the image. We used Ray [29] and Salvador [30] techniques to determine the optimum number of the clusters in proposed *k*-means clustering in color images segmentation. The distance of intra-cluster, which is simply the distance between a point and its cluster center (DBPC), and inter-cluster or the distance between clusters (DBC), are measured according to (11) and (12) respectively:

$$DBPC = \frac{1}{N} \sum_{i=1}^{K} \sum_{x \in C_i} \left\| x - z_i \right\|^2 \quad (11)$$

$$DBC = \min(\left\| z_i - z_j \right\|^2) \; i = 1,2,...,K\text{-}1, \; j = i+1...,K \; (12)$$

Where in (16), N is the number of pixels in the image, K is the number of clusters, and z_i is the cluster center of cluster C_i and in (12) z_j is the cluster that is compared to z_i. We obviously desire to minimize (11) and take the minimum of (12) value. Finally, we define the ratio as *Validity=DBPC/DBC* in which the goal is to minimize the intra-cluster distance, and consequently the *Validity* measure, as well as the inter-cluster distance measure. We compare the results obtained from the proposed method based on adaptive *K*-means to select the number of clusters and handy choice in figure 4.
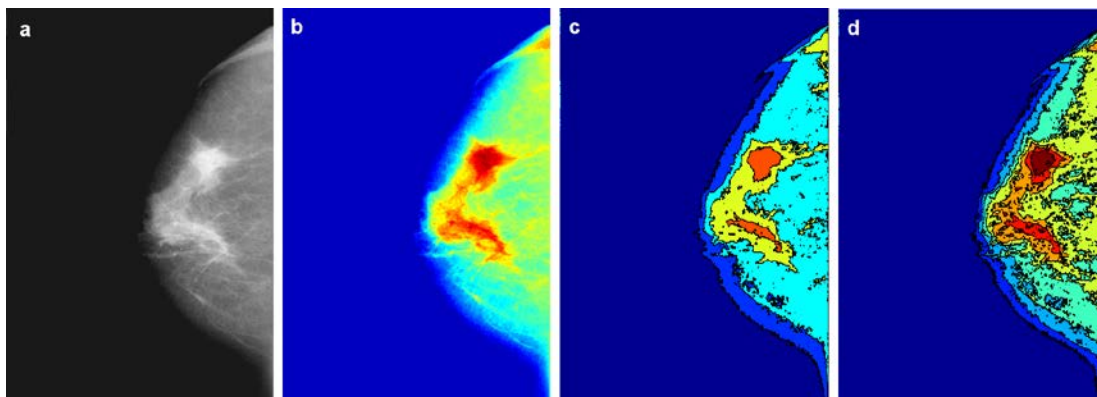

Figure 4: (a) Pre-processing image, (b) output of DWT, (c) applying adaptive K-means (K=6) and (d) handy choice of K (K=9).

Table 1: Implementation of the system and the results of the evaluation

Databases		No. Images	Tumor		Like to Tumors		F-Measure	Accuracy	A_z
			N_{TP}	N_{FN}	N_{TN}	N_{FP}			
Mini-Mias	Age>40	53	13	1	36	2	0.8965 (±0.07)	0.9423 (±0.02)	0.9131 (±0.01)
	Age<40	17	7	0	10	1	0.9333 (±0.01)	0.9444 (±0.06)	0.9268 (±0.06)
DDSM	Age>40	39	12	2	23	2	0.8571 (±0.04)	0.8974 (±0.02)	0.9011 (±0.04)
	Age<40	11	5	1	5	0	0.9090 (±0.03)	0.9090 (±0.09)	0.9085 (±0.01)
Total		44.3±6.2	120	37	4	74	5	0.8915 (±0.04)	0.9233 (±0.048)	0.9123 (±0.03)

Combining the white range of edge image with the output of adaptive K-means

By concerting the intensity of edge color range into the central rang of tumor growth, the location of disease is revealed to the physician. Then, using the matrix obtained from the mass, which falls in logical or binary matrix class, we will turn to the image after pre-processing or applying K-means.

If this image is called I and the matrix of the segmented colored tumors is t, then, according to (13) we will have:

$$I (t) = 0 \qquad (13)$$

Thus, all the elements of I matrix, which are corresponding to the zero elements of t, remain and the rest of elements become zero.

Experimental Results

Three assessment factors proposed to evaluate the accuracy of tumor detection are calculated according to the following equation. Traditionally, accuracy has been one of the key measures for evaluating the tumor segmentation results. For a binary classification, this measure is defined as:

$$accuracy = \left(\frac{t_p + t_n}{t_p + t_n + f_p + f_n}\right) \times 100 \qquad (14)$$

Where t_p and f_p indicate true and false positives respectively, while t_n and f_n show true and false negatives. The unbiased F-measure, on the other hand, is given by:

$$F_1 = 2.\frac{precision \times recall}{precision + recall} \qquad (15)$$

Where precision and recall are defined as:

$$precision = \frac{t_p}{t_p + f_p} \ , \ recall = \frac{t_p}{t_p + f_n} \qquad (16)$$

Table 1 includes the mean F-measure, accuracy and the area under the ROC curve (A_z) for each database with the corresponding standard deviation given in parentheses (Age>40 and Age<40). The patients had already been diagnosed with the breast cancer by radiologist and specialists, thus the presence of malignant tissues in images was obvious.

Of these people, 92 were over 40 and the rest were under 40 with the means age of 52.5±9.3 and 33.5±5.7 respectively. Of total 120 mammographic images, no disease was detected in 79 images and the existence of tumor was confirmed in 41 patients. Of 79 images of the healthy tissues, the proposed system was wrong in 5 cases. Of 41 images, which revealed disease and contained malignant tumors, the proposed system was wrong in 4 cases. In the rest of images, the position of cancerous tissues was accurately detected. Evaluating factors were implemented separately on both databases and the average accuracy, sensitivity and specificity were calculated. Three final factors were calculated for each database in average. Overall, the accuracy, sensitivity and specificity were respectively 92.33%, 90.24% and 93.67%. The proposed

system was quite successful in segmentation of tumors. All methods mentioned in Table 2 used the images taken from mini-Mias database. The proposed method, however, was implemented on DDSM.

Figure 5 presents ROC curves for the proposed system for 60 randomly selected mammography images (30 containing cancerous tumors and 30 normal tissues) in two ranges (Age>40 and Age<40 conditions). High accuracy and low false positive rate in segmentation distinguish this method from other techniques used for detection of the breast cancer in mammograms. Figure 6 presents the overlapping percentage of the proposed method with the diagnosis of the physicians and radiologists, which is approximately 78%, while both CQ and SRGA have overlapping per-

Table 2: Comparison of the proposed system with valid segmentation methods

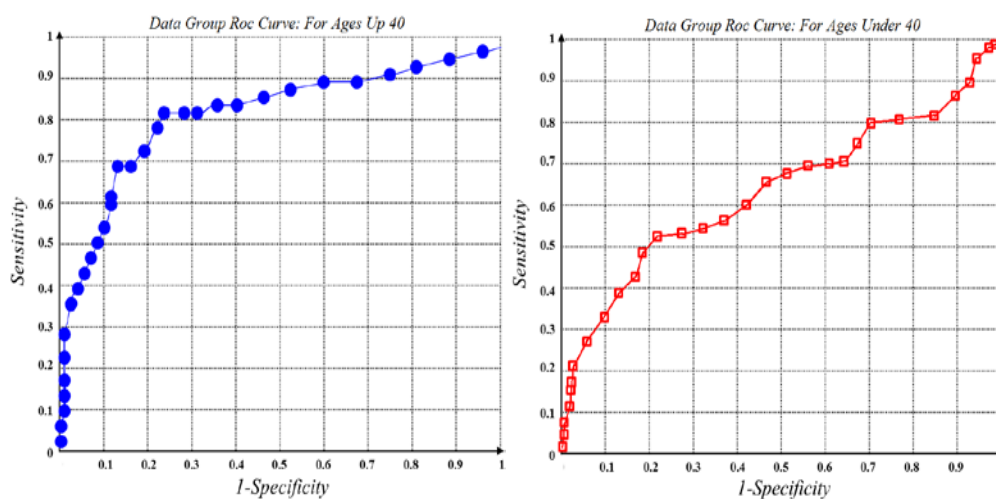| Techniques	AC	SE	SP
Rabottino [14]	91.32%	88.34%	93.38%
B. Zheng [31]	90.68%	85.00%	91.96%
Maitra [34]	92.32%	96.00%	94.06%
Sheng [35]	91.76%	84.20%	88.36%
Yang [36]	91.41%	90.12%	89.25%
Patel [37]	91.00%	89.69%	88.16%
Proposed Algorithm	92.33%	90.24%	93.67%

Figure 5: ROC curve in two ranges (left- Age>40), (right- Age<40).
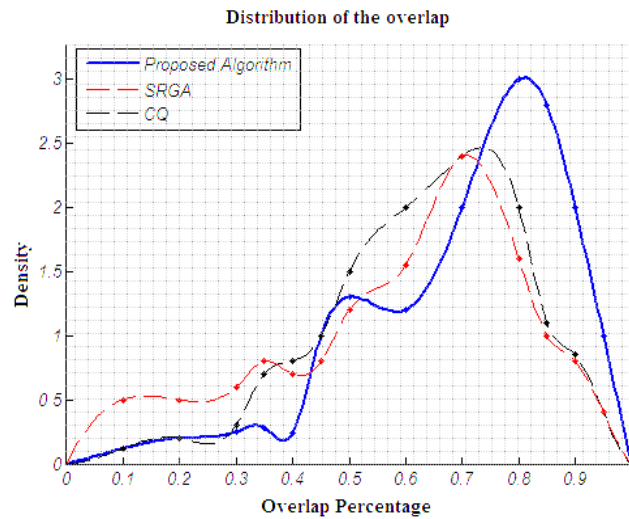
Figure 6: The distribution of the overlapping percentage for CQ, SRGA and the proposed algorithm.

centage of 50% and 57% respectively. Kappa coefficient shows the reliability of the system performance which is introduced in (17).

$$K = \frac{2(N_{TP}N_{TN} + N_{FN}N_{FP})}{(N_{TP} + N_{FN})(N_{TN} + N_{FN}) + (N_{TN} + N_{FP})(N_{TP} + N_{FP})} \quad (17)$$

The results indicates Kappa=84.65% which is suitable for system performance. The low PPV and high NPV of the system (reliability coefficient for clinical specialists and the pa-

tient) is a guarantee of the system reliability and both clinical specialist and patients can trust its software and output. Such statistical difference is highly significant (P <0.01). As a practical example, the three-dimensional images below contain detailed information of the masses. Figure 7 shows two examples with three-dimensional disturbed input data. This technique is based on topography contour that could produce better results and improve
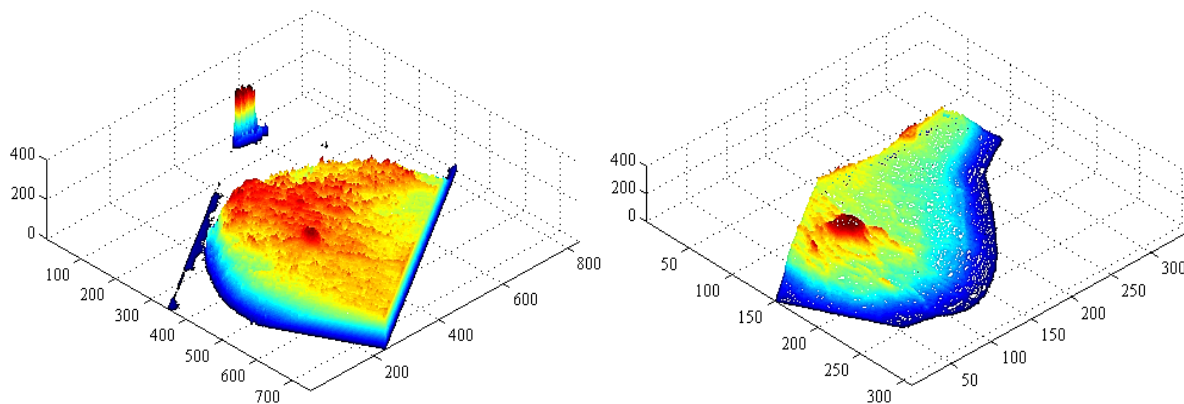

Figure 7: Three-dimensional mammograms based on K-means and DWT

segmentation. In figure 8, SRGA [34] and Color Quantization [38] methods have been displayed, though researchers usually prefer mammographic images for the segmentation of malignant tumors.

Conclusion

In this paper, a new method for segmentation of cancerous masses in mammographic images based on adaptive K-means clustering technique and discrete wavelet transform was proposed. Compared to other methods used for segmentation and detection of cancerous tumor such as thresholding, formal classification or K-means clustering, the proposed algorithm has the advantage of retaining the data and details of the original image. The high accuracy in detecting small cancerous masses also distinguishes this method from commonly used techniques. After testing 120 mammographic images taken from two distinct databases, 93% accuracy and 91% sensitivity was achieved. Improving the accuracy, sensitivity and specificity in future works and also implementing an efficient system suitable for all medical images is the goal pursued by the authors of this paper. The employment of this method in mammographic images broadens horizons for cancer detection and minimization of human errors in medicine.
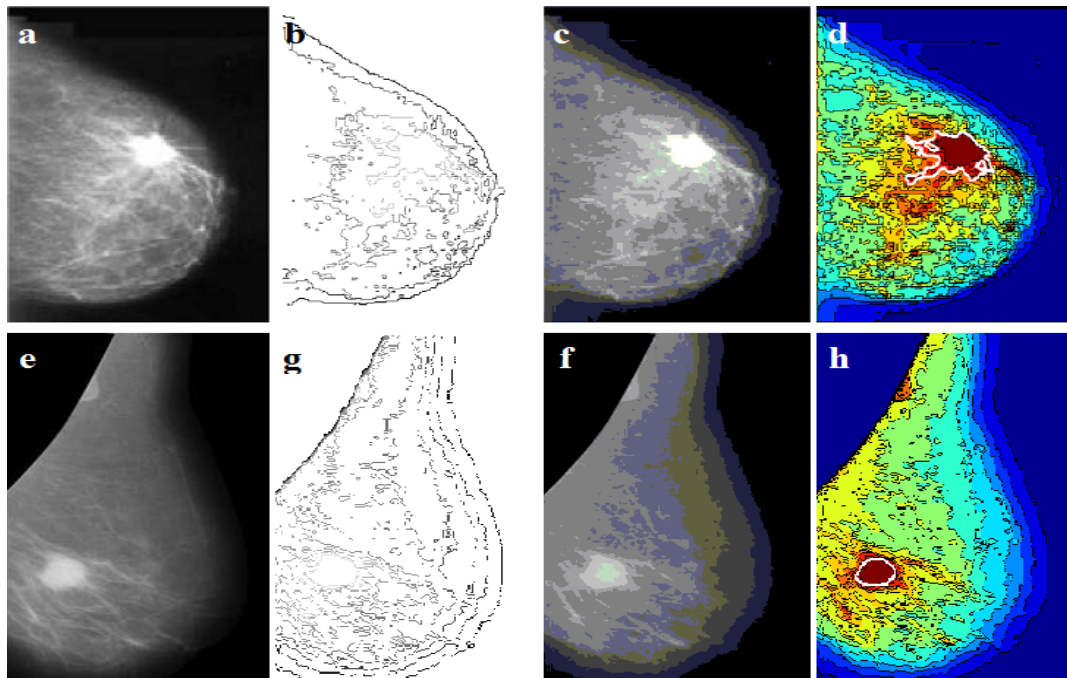
Conflict of Interest

None

Figure 8: (a) and (e) input mammographic images, (b) and (g) implementation of SRGA, (c) and (f) implementation of Color Quantization technique and (d) and (h) implementation of the proposed system.

References

1. American Cancer Society. Cancer Facts & Figures 2008. Atlanta: American Cancer Society; 2008.

2. Anderson WF, Pfeiffer RM, Dores GM, Sherman ME. Comparison of age distribution patterns for different histopathologic types of breast carcinoma. *Cancer epidemiology, biomarkers & prevention.* 2006;**15**:1899-905. doi: 10.1158/1055-9965. epi-06-0191. PubMed PMID: 17035397.

3. Harirchi I, Ebrahimi M, Zamani N, Jarvandi S, Montazeri A. Breast cancer in Iran: a review of 903 case records. *Public health.* 2000;**114**:143-5. doi: 10.1038/sj.ph.1900623. PubMed PMID: 10800155.

4. *American Cancer Society.* Breast Cancer Facts & Figures 2009-2010. Atlanta: American Cancer Society; 2010.

5. Mammography [Internet]. Oak Brook: Radiological Society of North America, Inc.; c2009. Available from: http://www.radiologyinfo.org/en/info.cfm?pg=mammo

6. Mangasarian O, Street N, Wolberg W. Breast cancer diagnosis and prognosis via linear programming. *J Oper Res.* 1995;**43**:570-7.

7. Beam CA, Layde PM, Sullivan DC. Variability in the interpretation of screening mammograms by US radiologists, Findings from a national sample. *Archives of internal medicine.* 1996;**156**:209-13. PubMed PMID: 8546556.

8. Majid AS, de Paredes ES, Doherty RD, Sharma NR, Salvador X. Missed breast carcinoma: pitfalls and pearls. *Radiographics.* 2003;**23**:881-95. doi: 10.1148/rg.234025083. PubMed PMID: 12853663.

9. Shen L, Rangayyan RM, Desaultels J. Detection and Classification of Mammographic Calcifications. *Int J Pattern Recognition and Artificial Intelligence.* 1993;**07**:1403-16. doi: doi:10.1142/S0218001493000686.

10. Woods KS, Solka JS, Priebe CE, Doss CC, Bowyer KW, Clarke LP, et al. Comparative Evaluation of Pattern Recognition Techniques for Detection of Microcalcifications in Mammography. *Int J Pattern Recognition and Artificial Intelligence.* 1993;**7**:1417-36. doi: 10.1142/S0218001493000698.

11. Verma B, Zakos J. A computer-aided diagnosis system for digital mammograms based on fuzzy-neural and feature extraction techniques. *IEEE transactions on information technology in biomedicine.* 2001;**5**:46-54. PubMed PMID: 11300216.

12. Sharma J, Sharma S. Mammogram Image Segmentation Using Watershed. *Int J Info Tech and Knowledge Management.* 2011;**4**:423-5.

13. Cascio D, Fauci F, Magro R, *et al.* Mammogram Segmentation by Contour Searching and Mass Lesions Classification With Neural Network. *IEEE Trans Nucl Sci.* 2006;**53**:2827-33. doi: 10.1109/tns.2006.878003.

14. Rabottino G, Mencattini A, Salmeri M, Caselli F, Lojacono R. Mass Contour Extraction in Mammographic Images for Breast Cancer Identification. In: International Measurement Confederation, editor. 16th IMEKO TC4 Symposium, Exploring New Frontiers of Instrumentation and Methods for Electrical and Electronic Measurements; 2008 Sept 22-4; Florence, Italy: Universitat de Florència; 2008.

15. Bankman IN, Nizialek T, Simon I, Gatewood OB, Weinberg IN, Brody WR. Algorithms for segmenting small low–contrast objects in images. In: Suri JS, Rangayyan RM, editors. Recent advances in breast imaging, mammography, and computer–aided diagnosis of breast cancer. Bellingham: SPIE Press; 2006. p. 723-38.

16. Niknam T, Taherian Fard E, Pourjafarian N, Rousta A. An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering. *Engineering Applications of Artificial Intelligence.* 2011;**24**:306-17. doi: http://dx.doi.org/10.1016/j.engappai.2010.10.001.

17. Issac Niwas S, Palanisamy P, Sujathan K, Bengtsson E. Analysis of nuclei textures of fine needle aspirated cytology images for breast cancer diagnosis using Complex Daubechies wavelets. *Signal Processing.* 2013;**93**:2828-37. doi: http://dx.doi.org/10.1016/j.sigpro.2012.06.029.

18. Mohanty A, Senapati M, Lenka S. An improved data mining technique for classification and detection of breast cancer from mammograms. *Neural Computing and Applications.* 2013;**22**:303-10. doi: 10.1007/s00521-012-0834-4.

19. The mini-Mias database for mammograms [Internet]. Brussels: Information Society Technologies. c2012. Available from: http://peipa.essex.ac.uk/info/mias.html

20. Heath M, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K, et al. Current Status of the Digital Database for Screening Mammography. In: Karssemeijer N, Thijssen M, Hendriks J, Erning

L, editors. Digital Mammography. Computational Imaging and Vision. Netherlands: Springer; 1998. p. 457-60.

21. Manjon JV, Coupe P, Buades A, Collins D, Robles M. New methods for MRI denoising based on sparseness and self-similarity. *Medical image analysis.* 2012;**16**:18-27. doi: 10.1016/j.media.2011.04.003. PubMed PMID: 21570894.

22. Barzigar N, Roozgard A, Verma P, Cheng S. Removing Mixture Noise from Medical Images Using Block Matching Filtering and Low-Rank Matrix Completion. In: IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB); 2012 Sep 27-28; San Diego, California, USA. Washington DC: IEEE Computer Society; 2012.

23. Kim JK, Song J, Lee S, Park IC. Tiled interleaving for multi-level 2-D Discrete wavelet Transform. In: The Institute of Electrical and Electronics Engineers, editor. IEEE International Symposium on Circuits and Systems; 2007 May 27-30; New Orleans. Washington DC: IEEE Computer Society; 2007. p. 3984-7.

24. Gonzalez R, Wintz P. *Digital image processing.* 5th ed. New York: Addison-Wesley Publishing Co; 2008. p. 139-203.

25. Jordan MI, Bishop CM. Neural Networks. In: Tucker AB, editor. Computer Science Handbook. 2nd ed. Boca Raton, FL: Chapman & Hall/CRC Press LLC; 2004.

26. Kaufman L, Rousseeuw PJ. *Finding Groups in Data: an Introduction to Cluster Analysis.* New Jersey: John Wiley & Sons; 1990.

27. Hamerly G, Elkan C. Alternatives to the k-means algorithm that find better clusterings. In: Proceedings of the eleventh international conference on Information and knowledge management; 2002; McLean, USA. VA: ACM; 2002. p. 600-7.

28. MacKay D. Information Theory, Inference and Learning Algorithms. Cambridge: Cambridge University Press; 2003. Chapter 20, An Example Inference Task: Clustering; p. 284-92.

29. Ray S, Turi RH. Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation. In: Proceedings of the 4th International Conference on Advances

in Pattern Recognition and Digital Techniques (ICAPRDT'99); 1999 Dec 27-29; Calcutta, India. New Delhi, India: Narosa Publishing House; 1999. p. 137-43.

30. Salvador S, Chan P. Determining the Number of Clusters/Segments in Hierarchical Clustering/Segmentation Algorithms. In ICTAI '04 Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence; 2004 Nov 15-17; Boca Raton, FL. Washington DC: IEEE Computer Society; 2004. p. 576-84.

31. Zhang Noriko Y, Tomuro T, Furst J, Stan Raicu D. A Contour-based Mass Segmentation in Mammograms. In: DePaul CDM School of Computing Research Symposium (SOCRS-09); 2009 May 9; DePaul Center, Chicago IL.

32. Gonzales RC, Woods RE. *Digital Image Processing.* 2nd ed. New Jersey: Prentice Hall; 2002.

33. Zheng B, Jiantao P, Park SC, Zuley M, Gur D. Assessment of the Relationship between Lesion Segmentation Accuracy and Computer-aided Diagnosis Scheme Performance. In: Giger ML, Karssemeijer N, editors. Proceeding of SPIE, Medical Imaging 2008: Computer-Aided Diagnosis; 2008 March 17. p. 1530-9.

34. Maitra I, Nag S, Bandyopadhyay SK. Detection of Abnormal Masses using Divide and Conquer Algorithm in Digital Mammogram. *Int J Emerg Sci.* 2011;**1**:767-86.

35. Sheng L, Babbs CF, Delp EJ. Multiresolution detection of speculated lesions in digital mammograms. *IIEEE Trans Image Processing.* 2001;**10**:874-84. doi: 10.1109/83.923284.

36. Yang SK, Moon WK, Cho N, Park JS, Cha JH, Kim SM, et al. Screening mammography-detected cancers: sensitivity of a computer-aided detection system applied to full-field digital mammograms. *Radiology.* 2007;**244**:104-11. doi: 10.1148/radiol.2441060756. PubMed PMID: 17507722.

37. Patel BC, Sinha GR. An Adaptive K-means Clustering Algorithm for Breast Image Segmentation. *Int J Comput Applic.* 2010;**10**:35-8.

38. Kekre HB, Sarode TK, Gharge SM. Tumor Detection in Mammography Images using Vector Quantization Technique. *Int J Intellig Infor Tech Applic.* 2009;**2**:237-42.