Optimized Segmentation of Breast Images Using FF-EM Framework

Amuthadevi C.
Department of CSE, Misrimal Navajee Munoth Jain Engineering College, Chennai, Tamilnadu, INDIA
camuthadevi@gmail.com

Abstract

Background: Many medical surveys report that the number of breast cancer death has been increasing 5% to 15% in every 10 years, since the year 1940. One of the benign states of breast cancer is micro-calcifications, but all the micro calcifications may not develop the breast cancer. Earlier diagnosis is helpful to give the proper treatment and reduce the fatality rate and Computer Aided Diagnosis will be helpful to the radiologists to enhance the visualization the micro-calcifications and masses.

Method: In this work, Mammogram image segmentation with global histogram was used. The parameters for segmentation are selected based on Expectation Maximization (EM) algorithm. Multi Resolution Wavelet Analysis is used for feature extraction. Optimized EM is used for segmentation. Firefly Algorithm is used for statistical parameter optimization of EM.

Conclusion: The proposed method's efficiency is evaluated by the parameters such as total correct fraction, sensitivity, specificity, dice coefficient and total volume error. The firefly-optimized EM segmentation results are compared with simple EM segmentation. Quantitative measures of proposed method proved that firefly optimization is efficient than simple EM algorithm.

Keywords: Breast Cancer, Mammograms, Micro-calcifications, Multiresolution Wavelet Analysis, Expectation Maximization, Firefly Optimization.

Introduction

Breast cancer is one of the most leading cause of death of women in the age of above 40 years and the rate of affected people are increasing in every day. In worldwide, when comparing with all the types of cancers, breast cancer contributes 22.9% in women [1]. Risk factors define the reasons which develop the cancer. Even though many of the risk factors are identified, there are no specific proven factors for breast cancer. Some of factors are lifestyle changes in food habits, lack of doing exercise, smoking, drinking habits, obesity, age, late pregnancy and family history. Among these factors the reason of family history and age are not avoidable. The Measurement of Quality of Life (QOL) for survivors of breast cancer is dependent on the clinical practice, current and future researches by assessing the outcomes of the treatment. This is because of rising number of breast cancer survivors. Earlier diagnosis is essential to increase the life time of cancer survivors. In clinics, the cancer is diagnosed based on the physical examination by clinicians, and non-invasive scanning such as mammograms, ultrasound and Magnetic Resonance Image (MRI) are used.

For the earliest diagnosis of breast cancer, mammogram is widely utilized to diagnose the abnormality. At the initial stage of breast cancer, the affected breast part of patient gets lump which is painless and in most of the cases the affected-person/clinician may not notice it. Next stage of symptoms is irritation on skin, pain, nipple discharge and retraction. This stage is benign stage and the first identifiable symptom of breast cancer[2]. The breast cancer has the properties of radiodense and radiolucent fat which are useful to identify the small cancers on the excellent background. In mammograms, as per radiodense property the cancer tissues are shown as white and as per radiolucent property cancer tissues are shown as dark-grey black. Sensitivity of the breast mammography can be determined based on the density of the breast tissue of the patients. The two major abnormalities found in mammograms are masses and the calcifications and their radiological appearance plays a crucial role to find the likelihood of malignancy.

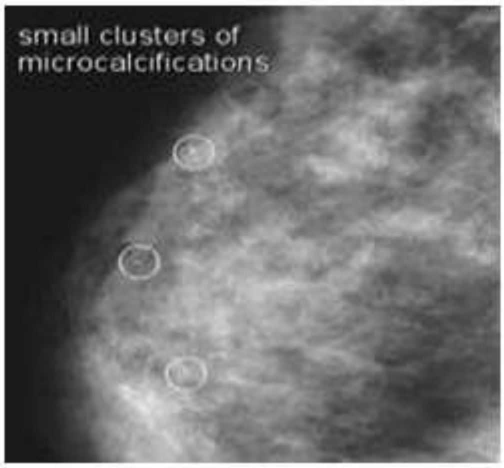

Figure 1: Mammogram with macro-calcification

Breast calcifications are calcium deposits developed on the breast tissues and their size may be macro or micro. They are mostly common in the breast and sometimes certain types of these calcifications lead to brain cancer. In case of suspicious calcifications, biopsy is taken and test is done at clinical labs, because it is the earlier sign of breast cancer. Micro-calcification size is usually not greater than 0.1 mm. So it is a tiny white dot and localization and analysis will be useful for earlier screening of breast cancer. In this work an optimized approach is proposed for segmentation of micro calcifications based on Expectation Maximization algorithm on breast mammograms. Multiresolution Wavelet Analysis is used for extracting contextual features. To set the parameters for segmentation Firefly optimization is used.

The following sections of this article are organized in this manner: Section 2 explains related work done in this domain, section 3 explains about the methodologies used in this proposed work, section 4 discusses about the results and section 5 gives the conclusion as well as directions for future research.

Related Works

Chen and Lee proposed a feature extraction method that was implemented using a Gaussian Markov random field (non-stationary) and on the basis of a multi-resolution wavelet model. This method provided very effectual features for both the Artificial Neural Networks and also for fuzzy c-means (FCM) classification [3]. Natural textures and mammography (digitized) were used. After segmentation each region was labeled based on extracted features from multi resolution wavelet transformation.

Maitra et al. [4] proposed Binary Homogeneity Enhancement Algorithm (BHEA), combined with Seeded Region Growing Algorithm (SRGA) to distinguish from normal tissues and abnormal tissues. In this method region growing protocol was utilized for segmenting the breast image into different regioareasns and then Anatomical Segmentation of Breast (ASB) was used to find abnormal tissues.

The wavelet based micro-calcification method of detection was used by Hashemi et al.[5] This method proved that the features extracted from wavelet transformation were having the properties which will be useful for image reconstruction from mother wavelets. Reconstruction was done by using common values of "skewness as well as kurtosis‟ at the intersection point of rows as well as columns. Therefore, the authors concluded that the rows and columns which have top values of "skewness and kurtosis‟ is the good candidate to identify micro-calcification clusters.

Suhail et al, used energy computation method on each pixel of the mammograms by using two different window functions. The energy function was computed for both small window (size 3 x 3) as well as large window (size 11 x 11) and ratio of energy is also calculated. The threshold is selected as 80% in ratio to detect the abnormal region. Then intensity based threshold methods were applied and morphological operations were also used to confirm the abnormal tissues [6].

Diaplaros et al, proposed a constrained generative model using spatial information combined with Expectation-Maximization (EM) to segment the images [7]. The generative model considered that the unnoticed category labels of pixels in the image were computed by using prior distributions by means of similar variables. The similarity was determined using the entropic quantities that relates to the nearest priors. The parameters of EM were derived iteratively by maximizing the lower-bound data logarithmic likelihood. This method is data dependent. Experiments were done on both synthetic as well as real images and the outcomes were contrasted with Markov-based models. This proposed method gave competitive segmentation results than markov-based models. TS vector quantization (TSVQ) method was proposed by Jong et al., for finding the initial values for EM for segmenting MRI brain images. The method was not used any assumption about the data and it was completely "data-driven‟ [8]. Ruslan et al., proposed Expectation-Conjugate-Gradient (ECG) algorithm to estimate the maximum likelihood for the latent variable models to prove that ECG was superior to simple EM model by means of convergence only in few instances[9]. Newton-like convergence and first-order convergence properties were compared. Then a hybrid EM-ECG method was also implemented with the synthetic and real data values.

Xin-She Yang developed Firefly algorithm and it is used in the applications such as digital image processing for compression, feature selection for classification algorithms, designing of extremely nonlinear, multimodal problems and for NP-hard scheduling problems [10]. Saibal et al., proved that Firefly algorithm suits for optimization problems even the data is so noisy[11] when comparing to other optimization algorithms.

Material and Methods

In this work, Multiresolution wavelet analysis is used for extracting features from the mammograms. Then a global thresholding based model with EM algorithm is used for segmenting normal and microcalcifiaction tissues. The outcome of the EM algorithm deepens on selection of initial parameters. In this work Firefly based optimization is used.

Multiresolution Wavelet Feature Extraction: In wavelet transformation each function f(n) is represented as a superposition of basis mother wavelet. By using the dilation and erosion operations the transformations are done. The mother wavelet is represented as $\psi(x)$. Then the orthogonal family of base functions are formed by:

$$\psi_{m,n}(x) = 2^{-\frac{m}{2}}(2^{-m}x - n)$$

By changing the values of m and n different vector values are obtained. Changing the value of n gives the distinct vector space in spatial orientations and changing the value of m does the dilation operation. The dilation transforms the actual function f into a vector space having a different resolution. If m is the resolution, then the following relationship is true, V_m is the proper subset of V_{m+1} such that $f(x) \epsilon V_0$ that implies $f(2^m m) \epsilon V_m$

A function f(x) is decomposed by:

$$f(x) = \sum_{m,n} c_{m,n}\, \psi_{m,n}(x)$$

Where $C_{m,n}$ is the coefficient which defines resemblance of the function f(x) at a specific resolution m as well as translation n. Therefore selecting the wavelet function gives the spatial and orientation details which will be used for defining shapes and their orientations.

The multiresolution wavelet analysis [12] transforms the actual image into a decomposed hierarchy of localized subimages and each one is in different spatial frequencies. A single image X is represented in 2D frequency spectrum and decomposed into one low-pass sub-band image X_{L0} and a collection of many band-pass sub-images X^i_j. Here $i = 1, . . . , L$, then $j = 0, 1, 2, 3$. The value i denotes the total number of resolution levels and the value j denotes the total number of orientations.

Therefore, X^i_j‟s for all the values of j (j=1, 2 and 3) give the detail about the sub images which are obtained by the output of the 2D wavelet decomposition. They are denoted as $H1$ and $H2$ respectively.

This transformation is used to visualize the coarse features in lower frequency components or resolutions and finer details in the high frequency components or resolutions. This spectral and spatial multiresolution aspect of the wavelet analysis is similar to the visualization by Human visual system. Therefore extracted features are useful for identifying mirocalcifications in mammograms.

EM for segmentation

EM algorithm: The EM (expectation-maximization) uses much iteration for calculating the maximum-likelihood estimates especially in a situation where the perceptions are incomplete. EM uses density mixures and the observable component/sample origin may be un-observable [13]. Density judgment of data points is done by EM in an unsupervised setting. There are two steps in EM algorithm and they are performed iteratively until convergence is reached. The EM method is widely used for the searching the value of the parameter by meeting the maximum likelihood. The steps are:

i) E (Expectation) step: It performs the expectation of the likelihood with the use of latent parameters as if those were noted.

ii) Maximization (M) step: This calculates the maximum likelihood estimates of the variables using the maximization of the expected likelihood identified in the previous E step. Then, the variables identified on current M step will be utilized to start another E step. This procedure is iterated till the convergence is achieved.

The stopping criteria of the EM algorithm may be any one of the following

i) Quantity of iterations is reached the maximum quantity of iterations.

ii) The error value is lower than some predefined value

iii) Calculation time reached to a predefined limit.

EM algorithm for image segmentation: EM algorithm can be efficiently used for binary segmentation of images. It assumes the distribution of intensity as Gaussian mixture. For Gaussian distribution, the initial parameters are identity matrix as well as k- means vectors. These are based on the centers extracted from the Gaussian mixture. In this work, k mixtures are assumed. Therefore the probability of density is denoted as:

$$f\left(\frac{x}{\theta}\right) = \sum_{i=1}^{K} \alpha_i\, f\left(\frac{x}{\theta_i}\right)$$

Where f defines the density of Gaussian model, α_i is the weight given to mixture and θ_i are the parameters defined by mean as well as standard deviation μ_i and σ_i respectively.

Then the probability of conditional densities may be expressed by:

$$\left(p\, {}^{x}/_{\theta_i}\right) = \frac{1}{\sqrt{2\pi\sigma_i}} \exp\left(-\frac{(x-\mu_i)^2)}{2\sigma_i^2}\right)$$

Optimization of EM parameters by Firefly Algorithm

The convergence for EM algorithm completely depends on the statistical parameter initialization. The statistical parameters used in EM are mean and standard deviation. The objective function is designed based on maximizing the likelihood of prior probabilities of two classes. They are background and micro calcification tissues.

Firefly Algorithm: The firefly algorithm is developed by Zhin-she and it uses three important principles [14]:

1. The fireflies are unisex in nature, so that elements available in the population able to attract one another.

2. The attractiveness measured between fireflies is directly relative to the brightness. The firefly having less brightnes will shift towards the more bright one. If none is more bright than a specific firefly, it shifts in random.
The "attractiveness‟ is directly proportional to the brightness that reduces with increasing distance between fireflies.

The brightness of an artificial-firefly is designed using the function that is to be optimized. In the algorithm, brightness of every firefly must be "directly proportional" to the value of the proposed objective function

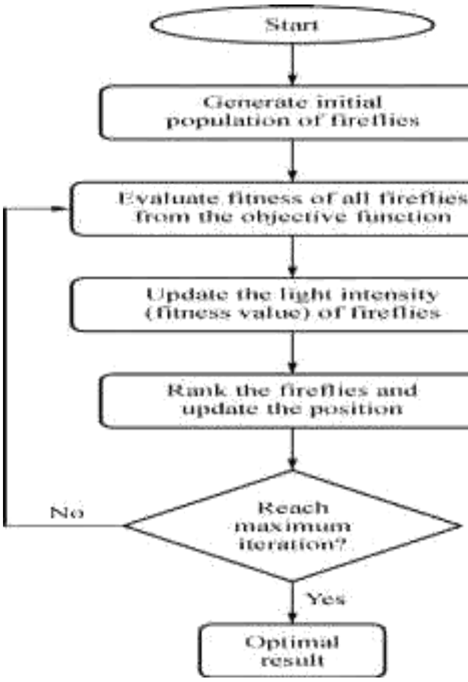

Figure 2: Firefly Algorithm

Proposed Optimization: The objective function is given by max(h(x) log P(x)/ θ) where x= 1 to L-1 wherein L is the maximum gray scale intensity of the mammogram.

The initial values of two classes P_1, P_2, μ_i and σ_i are selected in random and iteratively refined/tuned based on the objective function. The algorithm converges either maximum number of iterations is reached or correct fraction rate is reached to

Results and Discussion

For the experiments 10 mammogram images are taken and proposed algorithm is used. The results are compared by the parameters such as total correction fraction, sensitivity, specificity, dice coefficient and total volume error.

Correct fraction rate= correctly segmented pixels/ total number of pixels.

To find the total volume error, pixels are counted in 2D mammogram and added to form number of voxels. The volume error is found by using the number of wrongly marked voxels by an expert and the proposed algorithm.

From the Figure 3, it is observed that the total correct fraction reached 0.902 using the proposed method.

From the figure 4 it is observed that the sensitivity of Firefly-EM increases sensitivity from minimum of 1.45 % to maximum of 4.18%.

From the above figure 5 it is observed that the specificity of Firefly-EM increases sensitivity from minimum of 0.25 % to maximum of 3.11%.

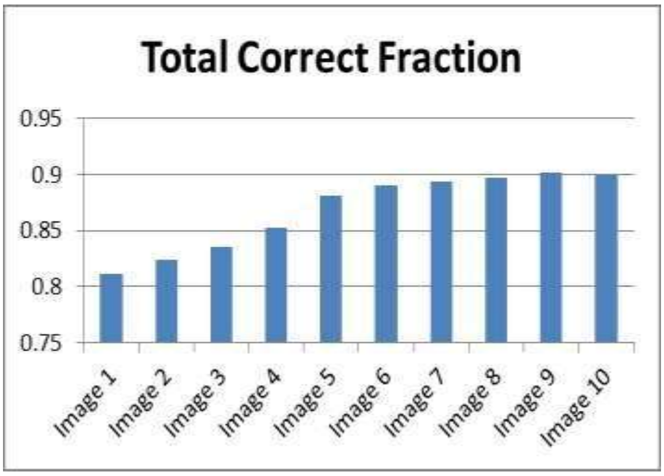

Figure 3: Total Correct Fraction

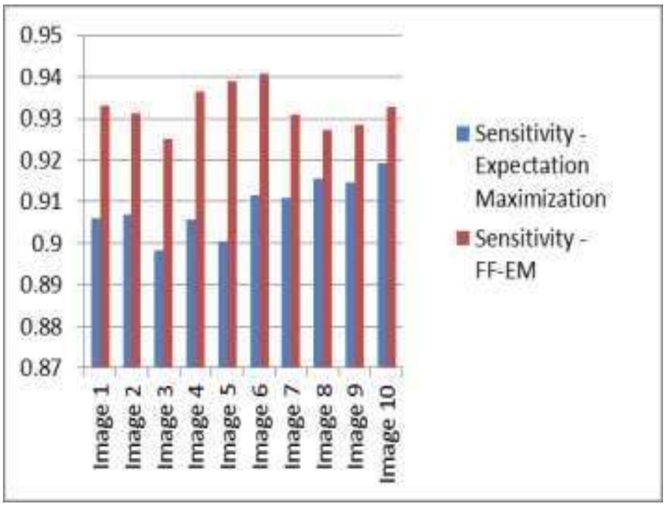

Figure 4: Sensitivity

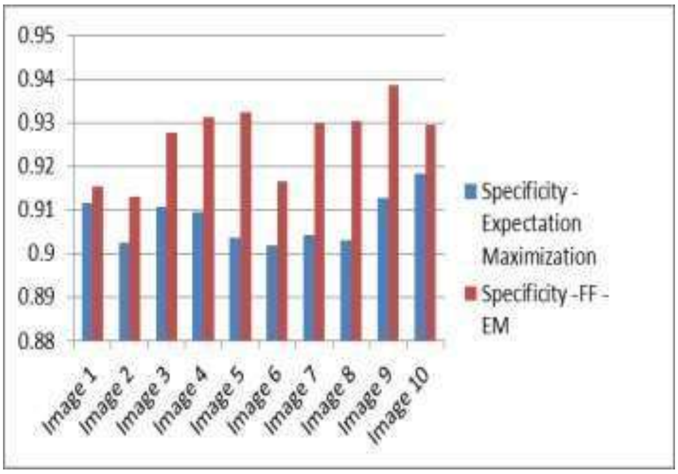

Figure 5: Specificity

From the Figure 6, it is observed that the Dice coefficien reached above 90% using the proposed method.

From the Figure 7, it is observed that the total volume error is less than 20 voxels when comparing the number of micro calcified voxels marked by an expert and the proposed method.

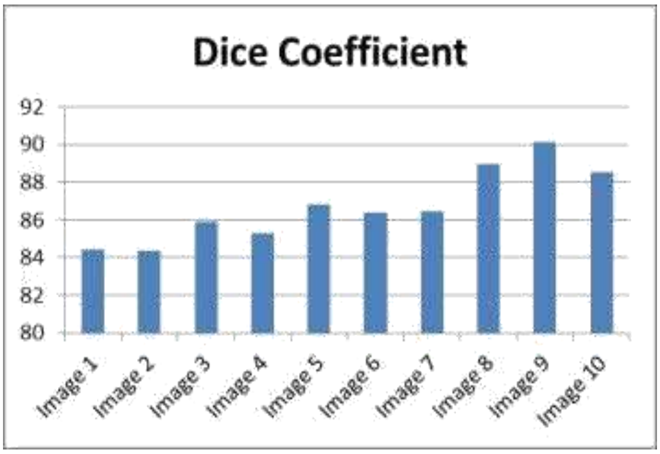

Figure 6: Dice coefficient

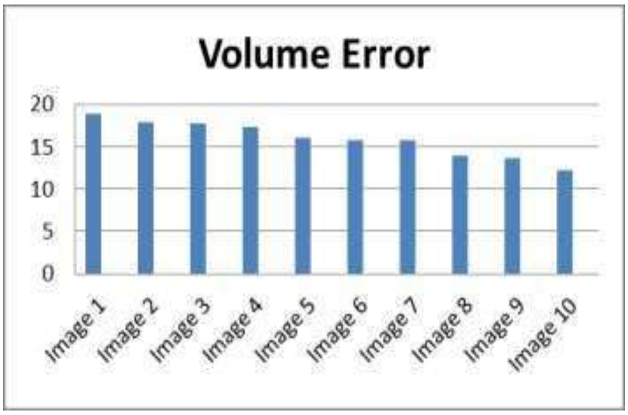

Figure 7: Volume Error

Conclusion

The Measurement of Quality of Life (QOL) for survivors of breast cancer is dependent on the clinical practice, current and future researches by assessing the outcomes of the treatment. This is because of rising number of breast cancer survivors. Earlier diagnosis is essential to increase the life time of cancer survivors. The modified EM algorithm using Firefly optimization gave better diagnostics help than simple EM algorithm.

References

1. Lakshmi R., Raju Athira, Joy Teena Mary and Vijayalakshmi S., Breast Cancer Risk Factors: Preventable and Non-Preventable, *International Research Journal of Pharmacy*, **3(4)**, 49-53 **(2012)**

2. Sachin Prasad N. and Dana Houserkova, The Role of Various Modalities in Breast Imaging, *Biomed Pap Med Fac. Univ. Palacky Olomouc Czech Repub*., **151(2)**, 209–218 **(2007)**

3. Chen C.H. and Lee G.G., Multiresolution wavelet analysis based feature extraction for neural network classification, In IEEE International Conference on Neural Networks, **3**, 1416-1421 **(1996)**

4. Maitra I.K., Sanjay N. and Bandyopadhyay S., Automated digital mammogram segmentation for detection of abnormal masses using Binary homogeneity enhancement algorithm, *International Journal of Computer Science, Engineering*, **2(3)**, 416–27 **(2011)**

5. Hashemi E., Asadzadeh M. and Kozakevicius A., Microcalcification detection in mammography using wavelet transform and statistical parameters, In the Conference of National Applied and Computational Mathematics Congress (CNMAC), 367–368 **(2012)**

6. Zobia Suhail, Mansoor Sarwar and Kashit Murtaza, Automatic detection of abnormalities in mammograms, *BMC Medical Engineering*, **15(53)**, DOI: 10.1186/s12880-015-0094-8 **(2015)**

7. Aristedidis Diploras, Nikos Vlassis and Theo Gevers, A Spatially Constrained Generative Model and an EM Algorithm for Image Segmentation, *IEEE Transactions on Neural Networks*, **18(3)**, 798-808 **(2007)**

8. Fwu Jong-Kae and Djuric Petar M., EM Algorithm for Image Segmentation Initialized by a Tree Structure Scheme, *IEEE Transactions on Neural Networks*, **6(2)**, 349-552 **(1997)**

9. Ruslan Salakhutdinov, Sam Roweis and Zoubin Ghahramani, Optimization with EM and Expectation-Conjugate-Gradient, Proceedings of the Twentieth International Conference on Machine Learning **(2013)**

10. Xin-She Yang and Xingshi He, Firefly Algorithm: Recent Advances and Applications, *Int. J. Swarm Intelligence*, **1(1)**, 36-50 **(2013)**

11. Pa Saibal K. and Rai C.S. Comparative Study of Firefly Algorithm and Particle Swarm Optimization for Noisy Non Linear Optimization Problems Amrit Pal Singh, *International Journal Intelligent Systems and Applications*, **10**, 50-57 **(2012)**

12. Mallat S.G., Qian W., Clark L.P., Li H.D., Clark R. and Silbiger M.L., A theory for multiresolution signal decomposition: The Digital wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intelligence,* **63(11)**, 674–693 **(1989)**

13. Mohamed Ali Mahjoub and Karim Kalti, Image segmentation by adaptive distance based on EM algorithm, *International Journal of Advanced Computer Science and Applications*, Special Issue on Image Processing and Analysis **(2011)**

14. Francisco Rogério B., Costa M. Fernanda P. and Rocha Ana Maria A.C., Experiments with Firefly Algorithm, *Computational Science and Its Applications*, **8580**, 227-236 **(2014)**.