ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

Mammogram Image Analysis for Microcalcification Detection using Bee Colony Optimization

[1]S. Jegadeesan, [2]P. Suresh Babu
[1,2]Dept. of IT, Velammal Institute of Technology, Chennai, India

Abstract
Breast cancer is one of the leading cancers in the female population. Mammography-based screening programs are carried out in many countries. Some of the important signs of breast cancer are masses, clusters of microcalcifications, and architectural distortions. Microcalcifications are quite tiny bits of calcium, and may show up in clusters or in patterns (like circles or lines) and are associated with extra cell activity in breast tissue. In this paper, meta-heuristics algorithm Artificial Bee Colony Optimization technique is implemented to extract the suspicious region based on the asymmetric approach. The ABC technique is a metaheurisitc algorithm for numerical optimization and it is based on the behavior of honey bees of foraging concept. Results obtained with a set of mammograms indicate that this method can improve the sensitivity and reliability of the systems for automated detection of microcalicification clusters. The ROC curve is generated for the mean value of the detection rate for all the 161 pairs of mammograms in MIAS database to calculate the performance of the proposed method.

Keywords
Genetic Algorithm, Mammogram, MIAS Database, Microcalcification, Artificial Bee Colony Optimization.

I. Introduction
Breast Cancer is one of the major causes for the increase in mortality among women, especially in developed countries. Breast cancer is the second most common cancer in women. The World Health Organization's International Agency for Research on Cancer in Lyon, France, estimates that more than 1,50,000 women worldwide die of breast cancer each year. The mortality and the incidence rate of breast cancer were estimated per 1, 00,000 women worldwide. In India, breast cancer accounts for 23% of all the female cancers followed by cervical cancer (17.5%) in major cities such as Mumbai, Calcutta and Bangalore. Although the incidence is lower in India than in the developed countries, the burden of diagnosing and treating of breast cancer in India is alarming. Microcalcifications are quite tiny bits of calcium deposits in the breast. They can be looked like in clusters or in patterns (like circles or lines) and are associated with extra cell activity in breast tissue. Usually the extra cell growth is not cancerous, but sometimes tight clusters of microcalcifications are usually a sign of benign breast tissue.

Medical imaging is the technique and process used to create images of the human body (or parts and function thereof) for clinical purposes or medical science. Medical imaging is undergoing revolution from analogy imaging to digital imaging. This report covers the changing medical imaging field from conventional X-ray to new modalities, such as MRI, Mammograms. This will help the radiologists, able to compare images from different modalities at one time on a workstation.

Three types of mammograms are carried out for the detection of breast cancer. First one, Screening mammograms are done, those who have no symptoms of breast cancer. Second one, Diagnostic mammograms are done when a woman has symptoms of breast cancer or a breast lump. Third one, Digital mammograms take an electric image of the breast and store it directly in a computer for analysis. Digital images are better at finding the cancer than X-ray film images.

Currently screening mammography is advocated for all Indian women. X-ray mammography is the most effective imaging technique for the early-diagnosis of breast cancer. Mammography-based screening programs are carried out in many countries, and their effectiveness has had a great impact on prognoses. A mammographic image is characterized by a high spatial resolution that is adequate enough to detect subtle fine-scale signs such as microcalcifications. The performance of the observer can be degraded by the huge resulting caseload, which is undoubtedly caused by visual fatigue and other psychophysical mechanisms. Double readings, as carried out, for example, by two radiologists, usually improve the quality of diagnostic findings, thus, greatly reducing the probability of misdiagnosis. The Proposed work for mammogram image analysis is designed to help the radiologist in the diagnosis of breast cancer at an early stage and it is also very simple [1, 3, 4].

II. Image Acquisition
The Mammography Image Analysis Society (MIAS), which is an organization of United Kingdom research groups interested in the understanding of mammograms, has produced a Digital mammography database. The data collection used in this experiment was taken from the Mammography Image Analysis Society (MIAS). The X-ray films in the database have been carefully selected from the United Kingdom National Breast Screening Programme and digitized with a Joyce-Lobel scanning microdensitometer to a resolution of $50\mu m \times 50\ \mu m$, 8 bits represent each pixel. The database contains left and right breast images for 161 patients, is used. Its quantity consists of 322 images, which belong to three types such as normal, benign and malign. There are 208 normal images, 63 benign and 51 malign, which are considered abnormal.

The list is arranged in pairs of films, where each pair represents the left (even filename numbers) and right mammograms (odd filename numbers) of a single patient. The size of all the images is 1024 pixels x 1024 pixels. The images have been centered in the matrix. When calcifications are present, centre locations and radii apply to clusters rather than individual calcifications. Coordinate system origin is the bottom-left corner. In some cases calcifications are widely distributed throughout the image rather than concentrated at a single site. In these cases centre locations and radii are inappropriate and have been omitted.

III. Pre-processing & Enhancement
The purpose of pre-processing is to remove the noise and radiopaque artifacts contained within the mammogram image and increase the region homogeneity, with the objective being to improve the algorithm reliability and robustness. Mammograms often contain artifacts in the form of the identification labels, markers, and wedges in the unexposed air-background (non-breast) region. Such artifacts are not transparent radiation and they are removed by

using Weighted Median (WM) filters. These filters have robustness and edge preserving capability of the classical median filter. The weighted median filter is implemented as follows:

$W(x, y) = \text{median } \{w_1 \times x_1 \ldots w_n \times w_n\}$ (1)

$x_1 \ldots x_n$ are the intensity values inside a window centered at (x, y)and w×n denotes replication of x, w times.

Image enhancement improves the quality (clarity) of images for human viewing. Removing blurring and noise, increasing contrast, and revealing details are examples of the enhancement operations. Here we are using the contrast stretching for enhancement purposes. Contrast stretching attempts to improve an image by stretching the range of intensity values it contains to make full use of possible values. The result is less dramatic, but tends to avoid sometimes artificial appearance of equalized images.

IV. Segmentation

Segmentation is the initial step for any image analysis. There are two different stages for segmentation. One is to obtain the locations of suspicious regions to assist the radiologists for diagnose. The other is to classify the abnormalities of the breast cancer into benign or malignant. Image Segmentation has been approached from a variety of approaches. Region based approach, morphological operation, multi scale analysis, fuzzy approaches and stochastic approaches have been used for mammogram image segmentation but with some of the limitations like computational time, memory, size and shape of the structural elements, resolution, etc. Here we are using the MRF model. It is used to deal with the spatial relations between the labels obtained in an iterative segmentation process [7, 8, 13]. Here in our proposed approach, the meta heuristic technique Genetic Algorithm (GA), and Particle Swarm Optimization (PSO) used to segment the suspicious region of microcalcification clusters.

Before going to segment the breast region, we should make some adjustments to be able to detect the cancer spots easier. For this, we are going to extract the pectoral muscle region.

A. Pectoral Muscle Region Extraction

The reliability of the boundary matching may be increased by extracting the pectoral muscle region. The pectoral muscle region appears as a bright rectangular region in the image corner towards the chest. Histogram based thersholding technique is used to separate the pectoral muscle region. The global optimum in the histogram is selected as the threshold value. The intensity value smaller than this threshold are changed to black (0), and the gray values greater than the threshold are changed to white (1). Then the gray level image is converted to binary for segmentation.

B. Nipple Position Identification

Initially the breast is divided into three parts. The middle part contains the nipple region. So we are taking the middle part and take the intensity of the border pixels as the initial population for applying GA. The intensity values of the border points are converted into binary strings and these are assigned as the population strings for Genetic algorithm. Using the GA operators make the new populations. Finally, the breast border pixels, which generates the minimum value of the population. This is considered as the nipple position.

C. Artificial Bee Colony Optimization

ABC is a new member of swarm intelligence (a branch of nature inspired algorithms focused on insect behavior) that tries to model natural behavior of real honey bees in food foraging. The foraging property of bees is used in problem modeling and solution, using several mechanisms of which waggle dance is a key to optimally locate food sources and to search for new ones. The dancing behavior of foraging bees while performing waggle dance is such that the direction of bees indicates the direction of the food source in relation to the sun, the intensity of the waggles indicates how far away it is and the duration of the dance indicates the amount of nectar on related food source.

Today, most scientists in the honeybee behavioral community agree that the communication system of the bees is a language regarding insect capacities. Honeybees use a sophisticated communication system that enables them to share information about the location and nature of resources. If a sugar solution is placed outdoors a long time might elapse before they found the food. Soon after this first visit, however, bees soon begin swarming around the feeder. The communication among bees is performed using what is called the "dance language" as a means of recruitment [9]. The dance language refers to patterned repetitive movements performed by bees that serve to communicate to their nest-mates the location of food sources or nest sites. In this way, the dance is a code that conveys the direction, distance, and desirability of the flower patch, or other discovered resource.

In the ABC algorithm [1], each food source is considered as a possible solution to the problem to be optimized [3]. The nectar amount represents the quality (fitness) of the solution represented by a food source. This algorithm starts by associating all employed bees with randomly generated food source positions. The values of control parameter of the algorithm have to be assigned at this step.

ABC is an iterative algorithm. For each iteration, every employed bee moves to the neighborhood of its currently associated position and evaluates its nectar amount. If the nectar amount of its neighborhood is better than that of its current position, the employed bee leaves the current position and moves to the new position. Otherwise, it just stays in its old current source position. When all employed bees finished their neighborhood search, they fly back into the hive to share their information about nectar amount of the food sources found recently with onlookers which wait in the dance area. Each onlooker selects a food source position according to a probability related to its nectar amount.

After all onlookers have selected their food sources, each of them determines a food source in the neighborhood of its current position and evaluates its fitness. Again, the comparison process of food position is done between neighborhood and current position to determine the new food source. If a solution does not improve during a predetermined number of iterations, that solution represented by a specific food position is abandoned by its employed bee and it becomes a scout. Then, this scout bee will go for searching a new food source position randomly and the new random solution will replace the abandoned one. The whole process is repeated through a predetermined number of cycles called Maximum Cycle Number (MCN) or until a termination criteria is reached.

It can be observed that many parts of the Artificial Bee Colony algorithm can be run in parallel. One way to have a parallel implementation would be to evaluate the fitness for each solution on a different processor. Alternatively, we could distribute the bees among the various processors and allow them to improve the solutions independently. However, this approach would be hindered by the dependencies between the bees that arise in three steps:

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

- Neighbourhood Search
- Probability Calculation
- Roulette Wheel Selection

A parallel implementation of the algorithm is designed for a shared memory architecture, which overcomes these dependencies. The entire colony of bees is divided equally among the available processors. Each processor has a set of solutions in a local memory. A copy of each solution is also maintained in a global shared memory. During each cycle the set of bees at a processor improves the solutions in the local memory. At the end of the cycle, the solutions are copied into the corresponding slots in the shared memory overwriting the previous copies. The solutions are thus made available to all the processors.

This process is maintaining iteratively to find the fittest ones with some threshold criteria. This is summarized in fig. 1.

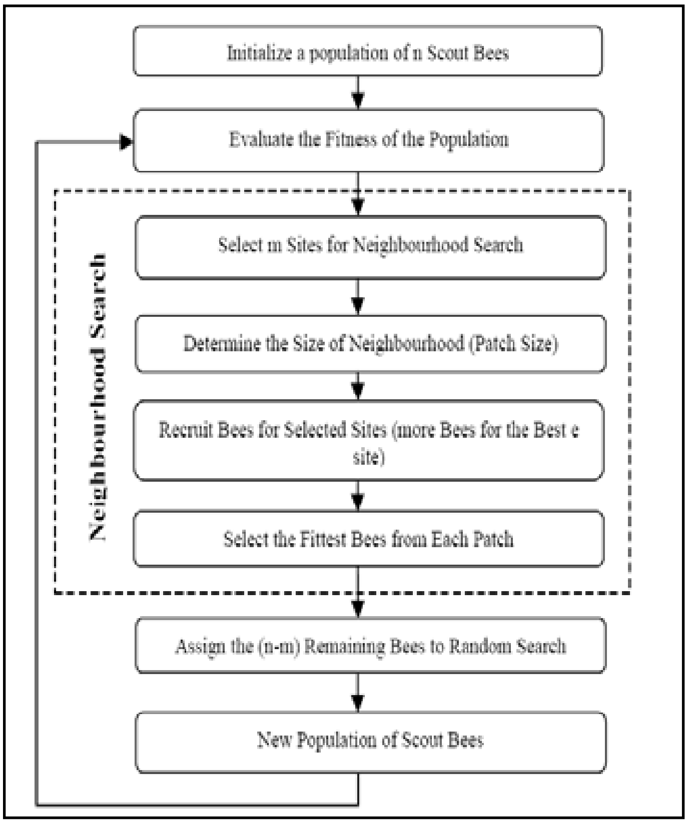
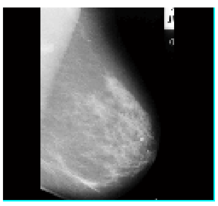


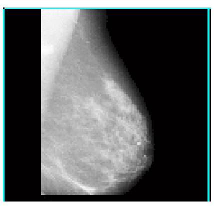
Fig. 1: ABC Optimization Process

V. Experimental Results

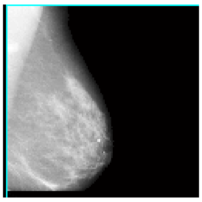
Here, we are using MATLAB 7.1.3 for image analysis and we take the MIAS Database images. The following fig. 2 shows the Original MIAS Database image, Median filtering image, aligned mammogram image, enhanced image using contrast stretching, pectoral muscle region extracted image and segmented mammogram image for microcalcification detection.



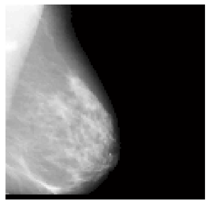
Input Image



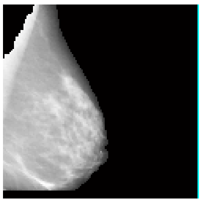
Median Filtering Image



Aligned Image



Enhanced Image



Pectoral Muscle Removal

Segmented Image

Fig. 2: Results of the Segmented Mammographic Image

The following table Table. 1 shows the detection rate of various approaches.

Table 1: Detection Rate of Various Approaches

S.No.	Authors & References	Methods	Detection Rate
1	Ferrair & Rangayyan	Directional Filtering with Gabor Wavelets	74.4%
2	Lau & Bishop	Asymmetry measures	85.0%
3	Sallam & Bowyer	Unwarping Technique	86.6%
4	Evolutionary Approach	Genetic Algorithm	86.6%
5	Meta Heuristic Approach	Ant Colony Optimization	94.8%
6	Meta Heuristic Approach	Particle Swarm Optimization	95.6%
7	Meta Heuristic Approach	Artificial Bee Colony Optimization	97.5%

The back propagation network uses supervised learning, which means that we provide the algorithm with examples of the inputs and outputs we want the network to compute, and then error (difference between actual and expected results) is calculated. The normalized feature values are given as input to a three-layer BPN to classify the microcalcifications into benign, malignant or normal. The network is trained to produce the output value 0.9 for malignant, 0.5 for benign and 0.1 for normal images.

A Receiver Operating Characteristics (ROC) analysis is performed to evaluate the classification performances of the proposed approaches. The performance of intelligent systems is best described in terms of their sensitivity and specificity, quantifying their performance related to false positive and false negative instances. The ROC curve is a plot of the classifier's true positive detection rate versus its false positive rate. The False Positive (FP) is the probability of incorrectly classifying a non-target object (e.g. normal tissue region) as a target object (e.g. tumor region). Similarly, the True Positive (TP) is the probability of correctly classifying a target object as being a target object.

The sensitivity or True Positive Fraction (TPF) of a BPN classifier is defined as the ratio between the number of true positive predictions and the number of positive instances in the test mammogram set.

Sensitivity = TP / (TP + FN) (4)

The specificity or True Negative Fraction (TNF) is defined as the ratio between the number of true negative predictions and the number of negative instances in the test mammogram set. It is defined as follows:

Specificity = TN / (TN + FP) (5)

Overall Accuracy = (Nr / N) * 100 % (6)

Where, Nr is the number of correctly classified mammogram images during the test run and N is the complete number of test mammograms. The TP and FP rates are specified in the interval from 0.0 to 1.0, in the mammogram image.

The ROC is the best suited to analyse the performances of segmentation, feature extraction, selection and classification. Fig. 3 shows the Receiver Operating Characteristics (ROC) curve. The first curve above from the slope is the result of the existing method Particle Swarm Colony Optimization approach to detect the calcification clusters. The Second curve indicates our proposed approach using PSO.

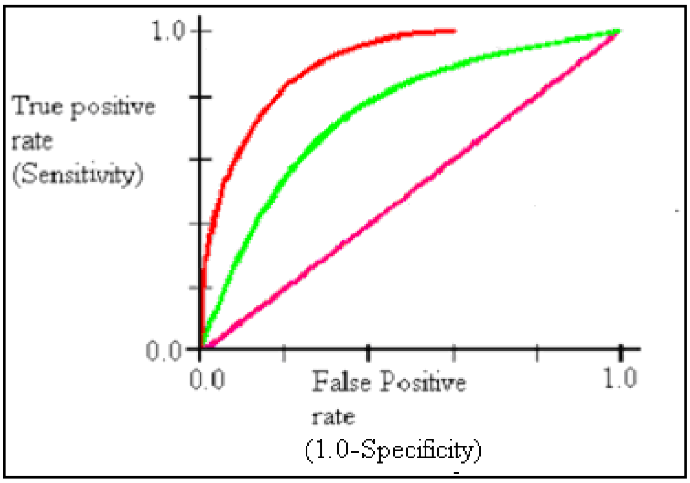


Fig. 3: Receiver Operating Characteristic Curve

VI. Conclusion

Finding an approach capable of segmenting the breast region in mammogram has proven to be difficult task. In this project, we have applied a novel approach for mammogram image segmentation based on the Meta heuristic algorithm ABC. The proposed CAD system is also easier for detection breast cancer. First, the image is pre-processed to remove unwanted labels and other noise factors then the image is replaced with the alignment of left and right breast images. Second, the aligned image is enhanced using normalization technique. Before going for segmentation the pectoral muscles are removed using thresholding the pixel values. Finally we are applying our proposed method Artificial Bee Colony Optimization technique to detect the microcalcification cluster (cancer) spots. The advantage of using the honeybee search algorithm is the robustness against outliers. This algorithm searches the space through a coordinate and intelligent process that removes significantly the outliers. The proposed ABC optimization algorithm can search for multiple thresholds which are very close to the optimal ones examined by the exhaustive search method. The segmentation results using the algorithm are the best and its computation time is relatively low. The microcalicifications on the mammogram images are easily detected using this algorithm. Finally we compare the results with the detection of microcalcification clusters using various techniques. As an extension of the work, this project is implementing with various meta-heuristic algorithms to select the optimal features from the mammogram image.

References

[1] B. Basturk, D. Karaboga,"An artificial bee colony (ABC) algorithm for numeric function optimization", Proc. IEEE Swarm Intelligence Symposium, Indianapolis, IN, USA, May 2006.

[2] Adil Baykaoglu, Lale Ozbakir, Pinar Tpakan,"Artificial Bee Colony Algorithm and its Application to Generalized Assignmnet Problem", Swarm Intelligence: Focus on Ant and Particle Swarm Optimization, Book edited by: Felix T.S Chan and Manoj Kumar Tiwari, pp. 532, December 2007.

[3] H. P. Chan, B. Sahiner, K. L. Lam, Computerized analysis of mammographic microcalcifications in morphological and texture features space, Med. Phys., v. 25, pp: 2007–2019, 1998.

[4] L. P. Cordella, F. Tortorella, M. Vento, Combing experts with different features for classifying clustered microcalcifications in mammograms, Proceedings of 15th International Conference on Patten Recognition, pp. 324–327, 2000.

[5] Dorigo, M., Di Caro, G., Gambardella, L.M.,"Ant algorithms for distributed discrete optimization", Artificial Life. Vol. 5, pp. 137–172, 1999.

[6] Dorigo, M., Gambardella, L.M., "A cooperative learning approach to the traveling salesman problem", IEEE Transactions on Evolutionary Computation, Vol. 1, No. 1, pp. 53-66, 1997.

[7] R. C. Eberhart, J. Kennedy,"A New Optimizer Using Particles Swarm Theory", Proc. Roc. 6th International Symp. Micro Machine and Human Science, pp. 39-43, Oct. 1995.

[8] R. C. Eberhart, Y. Shi,"Comparison between Genetic Algorithms and Particle Swarm Optimization", Proc. 7th international Conference on Evolutionary Programming, pp. 611-616, 1998.

[9] D. Karaboga, B. Basturk,"A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm," Journal of Global Optimization 39, 2007, pp. 459–471.

[10] D. Karaboga, B. Basturk,"On the performance of artificial bee colony (ABC) algorithm", Applied Soft Computing 8, 2008, pp. 687– 697.

[11] D. Karaboga, B. Akay,"A comparative study of artificial bee colony algorithm", Applied Mathematics and Computation, 214, 2009, pp. 108-132.

[12] D. Karaboga, B. Basturk,"Artificial Bee Colony (ABC) optimization algorithm for solving constrained optimization problems", Lecture Notes in Artificial Intelligence 4529, 2007, Springer-Verlag, Berlin, pp. 789–798.

[13] N. Karaboga, A. Kalinli, D. Karaboga,"Designing digital IIR filters using ant colony optimization algorithm", Engineering Applications of Artificial Intelligence, 17, pp. 301–309, 2004.

[14] N. Karssemeijer,"Adaptive noise equalization and image analysis in mammography", Information Processing in Medical Imaging: 13th International Conference, IPMI '93, AZ, USA, pp. 472–486, 1993.

[15] J. Kennedy, R. C. Eberhart,"A Discrete Binary Version of the Particle Swarm Algorithm", Proc. IEEE International Conference on Systems, Man, and Cybernetics, Vol. 5, pp.

ISSN : 0976-8491 (Online) | ISSN : 2229-4333 (Print)

4104-4108, Oct. 1997.
[16] J. Kennedy, R. Eberhart,"Particle swarm optimization", Proc. IEEE International Conference on Neural Networks, pp. 1942-1948, 1995.
[17] Thangavel, K., Karnan, M., Siva Kumar, R., Kaja Mohideen, A.,"Automatic Detection of Microcalcification in Mammograms-A Review", International Journal on Graphics Vision and Image Processing, Vol. 5, No. 5, pp. 31-61, 2005.
[18] Thangavel, K., Karnan, M.,"Computer Aided Diagnosis in Digital Mammograms: Detection of Microcalcifications by Meta Heuristic Algorithms", International Journal on Graphics Vision and Image Processing, Vol. 7, pp. 41-55, 2005.
[19] W. J. H. Veldkamp, N. Karssemeijer,"An improved method for detection of microcalcification clusters in digital mammograms", The SPIE Conference on Image Processing, 3661, pp. 512–522, 1999.

S. Jegadeesan, received B.Tech Information Technology from Anna University in 2005 and M.E Computer Science and Engineering from Anna University in 2010. He is working as an Assistant Professor, Velammal Institute of Technology, Chennai. His research interests include Digital Image Processing, Wireless Communication Networking, Neural Networks and Genetic algorithms.

P. Suresh Babu, received in B.Tech degree Information Technology from AnnaUniversity in 2009 and M.E Master of Engineering degree in Computer and Communication Engineering from Anna University in 2011.He is working as an Assistant Professor, Velammal Institute of Technology, Chennai. His research interests include Digital Image Processing, Genetic Algorithms and Wireless Networking.