

A Novel Evolutionary Approach to Detect Microcalcifications in Mammogram Image

R. Sivakumar

Research Scholar,
Research and Development Centre, Bharathiyar
University, Coimbatore, India

M. Karnan

Professor, Department of Computer Science and
Engineering, Tamilnadu College of Engineering,
Coimbatore, India

ABSTRACT

This paper describes a new approach for detection of Microcalcification using Evolutionary algorithms. The proposed system consists of two steps: First, the mammogram images are enhanced using median filter, normalized the image, pectoral muscle region is removed and the border of the mammogram is detected for both left and right images. Second, using the border points and nipple position as the reference the mammogram images are aligned and subtracted to extract the suspicious region. The Artificial Bee Colony Optimization Algorithm (ABC) algorithm is used to detect breast border and nipple position. In bilateral subtraction, the asymmetries between corresponding left and right breast images are considered for extracting the suspicious region from the background tissue. The textural features are extracted from the segmented mammogram image to classify the microcalcifications into benign, malignant or normal. Textural analysis methods such as Spatial Gray Level Dependency Matrix (SGLDM) and Gray-Level Run-Length Method (GLRLM) are used to extract the fourteen Haralick features from the segmented image. The normalized feature values are given as input to a three-layer BPN to classify the microcalcifications into benign, malignant or normal. The BPN classifier is validated using Jack-Knife Method.

Keywords

Microcalcification, Artificial Bee Colony Optimization Algorithm, Swarm Intelligence, Haralick features easy location of the manuscript using any search engines.

1. INTRODUCTION

Breast cancer is a very common disease. It is the most common cancer for females and the second most common cancer for males and females combined. Breast Cancer (malignant breast neoplasm) is cancer originating from breast tissue, most commonly from the inner lining of milk ducts or the lobules that supply the ducts with milk. Cancers originating from ducts are known as ductal carcinomas; those originating from lobules are known as lobular carcinomas. Currently screening mammography is advocated for all Indian women Mammography is the process of using low-energy-X-rays to examine the human breast and is used as a diagnostic and a screening tool. The goal of mammography is the early detection of breast cancer, typically through detection of characteristic masses and/or microcalcifications. Mammography-based screening programs are carried out in many countries, and their effectiveness has had a great impact on prognoses. The proposed intelligent system for mammogram image analysis is designed to help radiologists in the diagnosis of cancer at an early stage and it is shown to be effective [1-4].

2. SWARM INTELLIGENCE

Swarm Intelligence is a collection of nature-inspired algorithms. These algorithms are population based algorithms. A population of individuals cooperating among themselves and statistically becoming better and better over generations and eventually finding good enough solution. Researchers are investigated each swarm intelligence algorithm has been studied and modified to improve its performance from different perspectives such as the convergence, solution accuracy, and algorithm efficiency. Generally, the algorithms in swarm intelligence are applied to solve optimization problems.

The evolutionary computing is used to solve the optimization problems. Many swarm intelligence algorithms are solving problems with proposed optimization techniques such as the Cat Swarm Optimization (CSO), the Parallel Cat Swarm Optimization (PCSO), the Artificial Bee Colony (ABC), the Particle Swarm Optimization (PSO), and the Fast Particle Swarm Optimization (FPSO).

2.1 The Artificial Bee Colony Optimization Algorithm (ABC).

The ABC algorithm is proposed by Karaboga in 2005, and the performance of ABC is analyzed in 2007 [7-9]. The ABC algorithm is developed by inspecting the behaviors of the real bees. The real bees are finding food source, which is called the nectar, and sharing the information of food sources to another bees in the nest. In the ABC, the artificial agents are defined and classified into three types, namely, the employed bee, the onlooker bee, and the scout bee. The employed bee stays on a food source and provides the neighborhood of the source in its memory. The onlooker gets the information of food sources from the employed bees in the hive. Then they select one of the food sources to gather the nectar. The scout bees are responsible for finding new food, the new nectar, and sources. The ABC algorithm is presented as follows:

Step 1. Initialization: Spray N percentage of the populations into the image randomly, and then calculate their fitness values, which are called the nectar amounts, where N represents the ratio of employed bees to the total population. Once these populations are positioned into the solution space, they are called the employed bees.

Step 2. Onlookers activity: Calculate the probability of selecting a food source by the equation

$$p_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n}$$

a food source to move to by roulette wheel selection for every onlooker bees and then determine the nectar amounts of them. The movement of the onlookers follows the equation $v_{ij} = x_{ij} + \Phi_{ij} (x_{ij} - x_{jk})$ where fit_i is the fitness value of the solution

i which is proportional to the nectar amount of the food source in the position i , SN is the number of food sources which is equal to the number of employed bees (BN), $k \in \{1, 2, \dots, SN\}$ and $j \in \{1, 2, \dots, D\}$ are randomly chosen indexes. Although k is determined randomly, it has to be different from i . $\phi_{i,j}$ is a random number between $[-1, 1]$. It controls the production of neighbour food sources around $x_{i,j}$ and represents the comparison of two food positions visually by a bee.

Step 3. Scouts activity: If the fitness values of the employed bees do not be improved by a continuous predetermined number of iterations, which is called "Limit", those food sources are abandoned, and these employed bees become the scouts. The scouts are moved by the equation

$$x_i^j = x_{\min}^j + rand(0,1)(x_{\max}^j - x_{\min}^j).$$

Step 4. Update the Best Food Source: Memorize the best fitness value and the position, which are found by the bees.

Step 5. Termination Checking: Check if the amount of the iterations satisfies the termination condition. If the termination condition is satisfied, terminate the program and output the results i.e., the spatial co-ordinate of the center pixel is considered as nipple position; otherwise go back to the Step 2.

3. OVER VIEW OF THE WORK

Digitized mammograms are used in various stages of Computer Aided Detection systems [10-12]. The MIAS, which is an organization of UK-based research groups interested in the understanding of mammograms, has produced a digital mammography database. Initially the tracking algorithm is proposed to remove film artifacts such as labels and X-ray marks from the mammogram. The median filtering technique is applied to remove the high frequency components in the mammogram image. The advantage of using the median filter is that it removes the noise without disturbing the edges.

The global appearance (brightness, contrast, etc.) of the breasts may differ, due to variations in the recording procedure and this can be avoided by using the normalization method. The pectoral muscle appears as a bright triangular region in the image corner towards the chest wall and the top of the breast region. The gray level mammogram image is converted into a binary image to segment the pectoral muscle region. A histogram-based thresholding technique is used to separate the pectoral muscle region. The global optimum in the histogram is selected as the threshold value. The intensity values smaller than this threshold are changed to black (zero), and the gray values greater than the threshold are changed to white (one). Erosion and dilation operators are used to preserve the pectoral muscle region. The white pixels in the upper left corner of the binary mammogram image indicate the pectoral muscle region. The corresponding spatial coordinate of the white pixels in the gray level image is changed into black (zero) to remove the pectoral muscle region from the breast region. In the bilateral subtraction, the asymmetries between the corresponding region in the left and the right breast images are considered for segmentation. The breast border points and the nipple position are used as reference points for alignment of mammograms.

In this paper the detection of breast border and the nipple position is identified using a novel method called Artificial Bee Colony algorithm. The right mammogram is displaced and rotated in order to align the image. The coordinates of the detected border points and the nipple position of the left and

the right images are used to determine the displacement and the angle of rotation to align the images. Then the digital matrix of the left breast image is subtracted from the digital matrix of the right breast image to extract the suspicious region.

Microcalcifications in the right breast image have positive pixel values in the image obtained after subtraction, while microcalcifications in the left breast image have negative pixel values in the subtracted image. As a result, two new images are generated: one with positive values and the other with negative values. The most common gray level values are zero, which indicates that there is no difference between the left and the right images. The asymmetry image can be threshold to extract suspicious regions and further details can be found out.

Figure 1 shows the mammogram images from MIAS database.

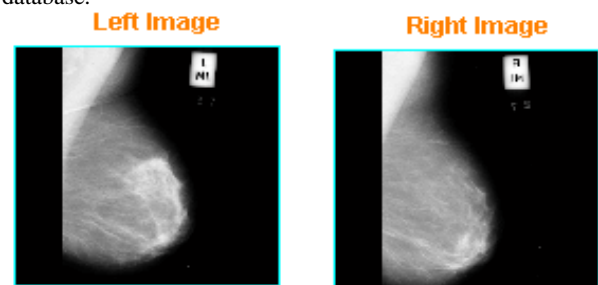


Figure 1: Input mammograms images from MIAS Database

Figure 2 shows the breast border detection using ABC.

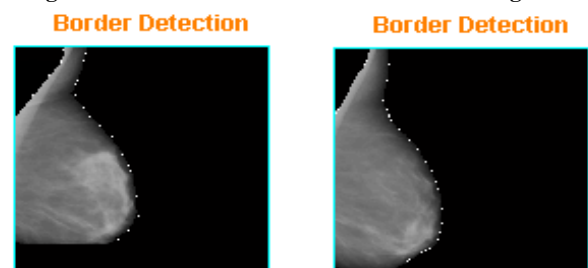


Figure 2: Breast Border detection using ABC
 Figure 3 shows the Identification of the Nipple Position using Artificial Bee Colony algorithm

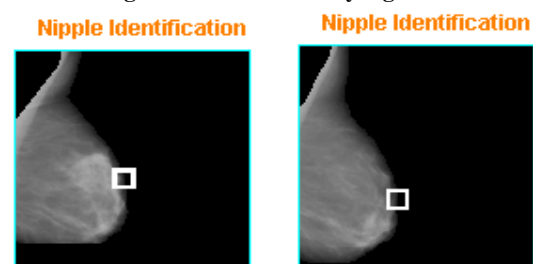


Figure 3: Identification of the Nipple Position using ABC
 Figure 4 shows the segmented image using bilateral subtraction

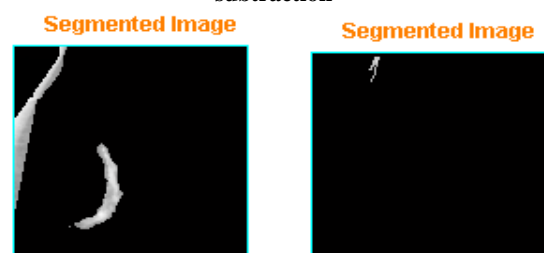


Figure 4: Segmented image using bilateral subtraction

4. FEATURE EXTRACTION

Textural features can be extracted to classify the segmented image into benign and malignant. This section presents the texture analysis methods such as Spatial Gray-Level Dependence Method (SGLDM), and the Gray-Level Run-Length Method (GLRLM) performance are compared.

4.1 Spatial Gray-Level Dependence Method (SGLDM)

In this method, co-occurrence matrix is generated to extract the texture features from the segmented mammogram image [13, 14]. The co-occurrence matrix is a technique that allows for the extraction of statistical information from the image regarding the distribution of pairs of pixels. It is computed by defining a direction (θ) and a distance (d), and pairs of pixels separated by this distance, computed across the defined direction, are analyzed. A count is then made of the number of pairs of pixels that possess a given distribution of gray-level values. Each entry of the matrix thus corresponds to one such gray-level distribution. Normally a set of 20 co-occurrence matrices are computed, for five different distances, in the horizontal, vertical, and two diagonal directions i.e., the distances are 1, 3, 5, 7 and 9, and the four angles 0° , 45° , 90° , and 135° are defined for calculating the matrix for each of the five distances.

4.2 Gray-Level Run-Length Method (GLRLM)

In run-length matrix, it counts the number of pixels having same gray level values with various distances in a specified direction (θ). Normally 4 different run-length matrixes can be generated for each image in horizontal, vertical and two diagonal directions.

The 14 Haralick features such as Angular Second Moment (ASM), Contrast (CON), Correlation (COR), Variance (VAR), Inverse Difference Moment (IDM), Sum Average (SA), Sum Variance (SV), Sum Entropy (SE), Entropy (ENT), Difference Variance (DV), Difference Entropy (DE), Information Measure of Correlation I (IMC1), Information Measure of Correlation II (IMC2) and Maximal Correlation Coefficient (MCC) are calculated for each texture analysis method [5]. The next section describes the classification of extracted features from the textural analysis methods are classified using BPN classifier.

5. BACKPROPAGATION NETWORK CLASSIFIER

The classifier employed in this paper is a three-layer Backpropagation neural network. The Backpropagation neural network optimizes the net for correct responses to the training input data set. More than one hidden layer may be beneficial for some applications, but one hidden layer is sufficient if enough hidden neurons are used.

Initially the features extracted from the textural analysis method, are normalized between [0,1]. That is each value in the feature set is divided by the maximum value from the set. These normalized values are assigned to the input neurons. The number of hidden neurons is equal to the number of input neurons. And only one output neuron. Initial weights are assigned randomly between [-0.5 to 0.5]. The output from the each hidden neuron is calculated using the sigmoid function,

$S_1 = 1/(1 + e^{-\lambda x})$, where $\lambda=1$, and $X = \sum_i w_{ih} k_i$ where w_{ih} is the weight assigned between input and hidden layer, and k is the input value. The output from the output layer is calculated

using the sigmoid function, $S_2 = 1/(1 + e^{-\lambda x})$ where $\lambda=1$ and $X = \sum_i w_{ho} S_i$ where w_{ho} is the weight assigned between hidden and output layer, and S_i is the output value from hidden neurons. S_2 is subtracted from the desired output. Using this error (d) value, the weight change is calculated as: $\Delta = d * S_2 * (1 - S_2)$ And the weights assigned between input and hidden layer and hidden and output layer are updated as: $w_{ho} = w_{ho} + (n * \Delta * S_1)$, $w_{ih} = w_{ih} + (n * \Delta * K)$ where n is the learning rate, k is the input values. Again calculate the output from hidden and output neurons. Then check the error (d) value, and update the weights. This procedure is repeated till the target output is equal to the desired output. The network is trained to produce a 0.9 output value for malignant and 0.1 output value for benign images.

6. RECEIVER OPERATING CHARACTERISTIC (ROC)

The Receiver Operating Characteristic ROC analysis is based on statistical decision theory, developed in the context of electronic signal detection, and has been applied extensively to diagnostic systems in clinical medicine. The ROC curve is a plot of the classifier's true positive detection rate versus its false positive rate. The False Positive (FP) rate 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) detection rate is the probability of correctly classifying a target object as being a target object. The TP and FP rates are specified in the interval from 0.0 to 1.0, in the mammogram image. The TP rate is referred to as sensitivity and 1.0 minus FP rate is called specificity. The input to the BPN classifier has parameters such the distance, theta and hidden neurons for SGLDM and theta and hidden neurons for GLRLM that can be varied to alter the TP and FP rates. Using these parameters ROC curve can be generated.

Table 1 shows the classification performance of SGLDM and GLRLM.

Table 1. Classification Performance of SGLDM and GLRLM

SGLDM		GLRLM	
Distance	Bilateral subtraction	Angles	Bilateral subtraction
	Az values		Az values
d=1	0.859	$\theta = 0, 180$	0.706
d=3	0.838	$\theta = 45, 225$	0.613
d=5	0.826	$\theta = 90, 270$	0.619
d=7	0.859	$\theta = 135, 315$	0.678

7. RESULTS

The images used in this work were taken from the Mammography Image Analysis Society (MIAS). The database consisting of 320 images, which belong to three normal categories: normal, benign and malign. There are 206 normal images, 63 benign and 51 malignant. In this paper, only the benign and malign images are considered for feature extraction. The classification performance is tested using jack-knife method, round-robin method. The results were analyzed by using ROC curve.

8. JACK KNIFE METHOD

The classification performance is studied using the Jack Knife method and ROC analysis. For the Jack Knife method, one half of the segmented images are selected randomly from the database for training of the neural network; subsequently, the other halves of the segmented images are used for testing the trained neural network. The training set is used to train the BPN algorithm. In this experiment, training is completed once the value of the error is less than 0.1. In this paper, ten combinations of training and testing pairs are used to generate the ROC curves. The table 3 shows the classification results based on the Jack Knife method

Table 3. Classification Performance Based on the Jack Knife Method

Textural Matrices	Bilateral subtraction
	Az values
SGLDM	0.859
GLRLM	0.706

9. CONCLUSION

The proposed system consists of two steps: First, the mammogram images are enhanced using median filter, normalized the image, pectoral muscle region is removed and the border of the mammogram is detected for both left and right images. Second, using the border points and nipple position as the reference the mammogram images are aligned and subtracted to extract the suspicious region. Textural analysis methods such as Spatial Gray Level Dependency Matrix (SGLDM) and GLRLM are used to extract the fourteen Haralick features from the segmented image. The normalized feature values are given as an input to the three-layer BPN to classify the microcalcifications into benign, malignant or normal. The BPN classifier is validated using Jack-Knife Method. In this paper, a novel approach is implemented for mammogram segmentation and classification based on the combination of ABC and Backpropagation Neural Network with ROC Analysis. The experiments and results show that ABC performs better than other existing algorithms.

10. REFERENCES

- [1] Cordella.L.P, Tortorella.F and Vento.M, Combing experts with different features for classifying clustered microcalcifications in mammograms, Proceedings of 15th International Conference on Patten Recognition, pp: 324–327, 2000.
- [2] Chan.H.P, Sahiner.B and Lam.K.L, Computerized analysis of mammographic microcalcifications in morphological and texture features space, Med. Phys., v. 25, pp: 2007–2019, 1998.
- [3] Dorigo, M., Di Caro, G., and Gambardella, L.M.: “Ant algorithms for distributed discrete optimization,” *Artificial Life*. vol.5, 137–172, 1999.
- [4] Dorigo, M., and 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.
- [5] Haralick, R.M., Shanmugan, K., and Dinstein, I.: “Textural features for image classification,” *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp: 610–621, 1973.
- [6] Jean, H., David, M.C., Charles, F.B., Zygmunt, P., Edward, J.D.: “Preclinical ROC studies of digital stereo mammography,” *IEEE Trans. Med. Imag.*, vol. 14, no. 2, pp: 318–327,1995.
- [7] Karaboga.D, Bahriye Akay,” A comparative study of Artificial Bee Colony algorithm” *Journal of Applied Mathematics and Computation*, Elsevier Inc, 214 (2009) 108–132.
- [8] Karaboga, D, Basturk,.B, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm, *Journal of Global Optimization* 39 (3) (2007) 459–471.
- [9] Karaboga,.D, Basturk.B, On the performance of artificial bee colony (abc) algorithm, *Applied Soft Computing* 8 (1) (2008) 687–697.
- [10] Mendez, A.J., Tahocesb, P.G., Lado, M. J., Souto, M., Correa, J.L., Vidal, J.J.: “Automatic Detection of Breast Border and Nipple in Digital Mammograms,” *Computer Methods and Programs in Biomedicine*, vol. 49, pp: 253–262, 1996.
- [11] Thangavel, K., Karnan, M., Siva Kumar, R., and 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.
- [12] Thangavel, K., and 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.
- [13] Thangavel, K., Karnan, M., Siva Kumar, R., and Kaja Mohideen, A.: “Segmentation and Classification of Microcalcification in Mammograms Using the Ant Colony System,” *International Journal on Artificial Intelligence and Machine Learning*, vol. 5, pp: 29-40, 2005.
- [14] H. Yoshida, W. Zhang, W. Cai, K. Doi, R. M. Nishikawa and M. L. Giger, Optimizing wavelet transform based on supervised learning for detection of microcalcifications in digital mammograms, Proceedings of the IEEE International Conference on Image Processing, v. 3, pp: 152–155, 1995.