

Diagnose Breast Cancer through Mammograms Using EABCO Algorithm

R.Sivakumar¹, Marcus Karnan²

¹Research Scholar, Research and Development Centre, Bharathiar University, Coimbatore, India.

¹rksivame@gmail.com

²Prof & Head, Dept. of CSE, Tamilnadu College of Engineering, Coimbatore, India.

²drmkarnan@gmail.com

Abstract—The aim of this research is the development of a reliable tool to detect early signs of breast cancer in mammographic images. Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death of female worldwide. Mammogram is one of the most excellent technologies currently being used for diagnosing breast cancer. In this paper, the Enhanced Artificial Bee Colony Optimization (EABCO) is proposed to automatically detect the breast border and nipple position to identify the suspicious regions on digital mammograms based on bilateral subtraction between left and right breast image. The algorithms are tested on digitized mammograms from MIAS database.

Keyword- Enhanced Artificial Bee Colony Optimization (EABCO), Microcalcifications, Mammograms,

I. INTRODUCTION

Medical image segmentation is the process of labeling each voxel in a medical image dataset to indicate its tissue type or anatomical structure. The labels that result from this process have a wide variety of applications in medical research and visualization. Breast cancer is considered one of the most important health problems in western countries and indeed it is the most common cancer among women. 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.

Breast cancer is a disease of humans and other mammals; while the overwhelming majority of cases in humans are women, men can also develop breast cancer. Mammography is the most used screening tool for abnormality detection, because it allows an easy way to identify the cancer. However, it is widely believed that not all cancers can be detected using this technique. The detected cancer after a negative mammography are called interval cancers, and is one of the goals of the CAD systems to keep low the rate of these cancers. 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. Radiologists interpret the mammograms and attempt to identify areas of potential abnormalities. Therefore, the effectiveness of this screening method is dependent on the radiologist's ability to detect areas of subtle irregular abnormalities. It is estimated that between 10-30% of women diagnosed with breast cancer have false-negative mammograms[3]. Most of the false-negative cases can be attributed to the radiologist's failure to detect a cancer which could be due to misinterpretation, or simply that the radiologist overlooked the area. It has been demonstrated that an independent second reading can significantly improve the detection rate and decrease the number of false positive cases.

Computerized tools and analysis can act as an independent secondary reading. The tools can be described as a supplement or a "second reader" to assist the physician in detecting and diagnosing breast cancer. Detection is the ability to identify abnormal areas in the breast, and diagnosis follows the detection process to identify those regions as being benign or malignant. Before these processes can perform their roles a really important pre-processing step has to take place which is the detection or segmentation of the breast region from the background. 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 [13-16].

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 that was 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 $\mu\text{m} \times$

50 μm , 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. Figure 1 shows the mammogram images from MIAS database [17].

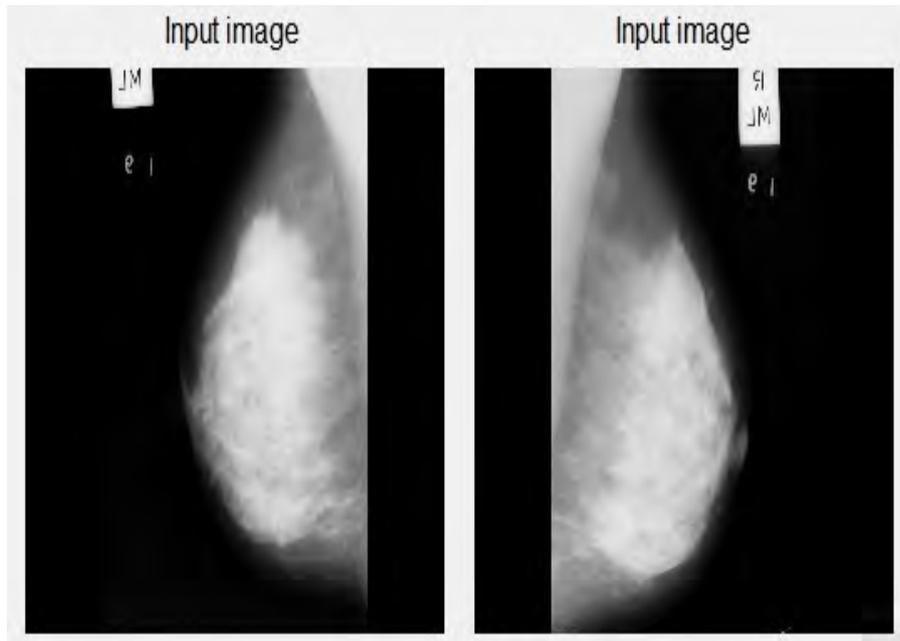


Figure 1: Input mammograms images from MIAS Database

III. ENHANCEMENT

This section deals with pre-processing and enhancement activities such as removal of film artifacts and labels, filtering the image, normalization and removal of pectoral muscle region. The enhancement method consists of four processing steps. In the first step, the given images are identified as left or right breast image and the film artifacts such as labels and X-ray marks are removed from the mammogram. In the second step, the high frequency components are removed using weighted median filtering. In the third step, to avoid the difference in contrast and brightness of the mammograms images caused by the recording procedure are normalized by image processing techniques. In the fourth step, the pectoral muscle region is removed from the breast region by using modified tracking algorithm. Figure 2 shows the enhanced mammogram images

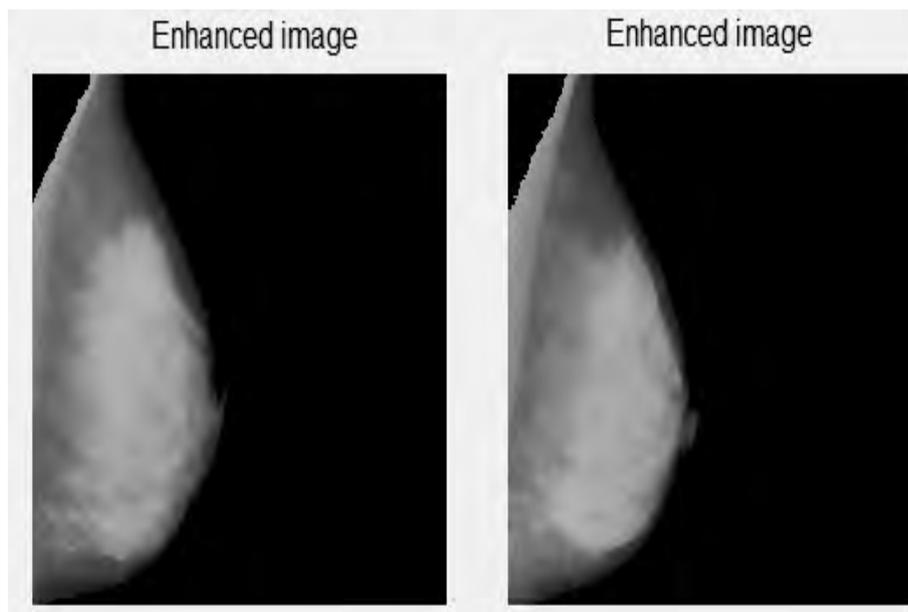


Figure 2: Enhanced mammogram images

IV. SEGMENTATION OF SUSPICIOUS REGION

The suspicious region or microcalcifications is segmented using Bilateral Subtraction for a pair of images. Thangavel et al. and Cheng et al. have presented a study on methods of various stages of automatic detection of microcalcification in digital mammograms [18-20]. In this section, the metaheuristic algorithm such that the Artificial Bee Colony (ABC) is implemented to extract the suspicious region based on the asymmetry approach.

The Artificial Bee Colony algorithm is proposed by Karaboga in 2005, and the performance of Artificial Bee Colony Optimization is analyzed in 2007 [6-12]. The Artificial Bee Colony Optimization algorithm is developed by inspecting the behaviours of the real bees on finding food source, which is called the nectar, and sharing the information of food sources to the 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. Each of them plays different role in the process: the employed bee stays on a food source and provides the neighbourhood of the source in its memory; the onlooker gets the information of food sources from the employed bees in the hive and select one of the food source to gather the nectar; and the scout is responsible for finding new food, the new nectar, sources. The process of the ABC algorithm is presented as follows:

In the ABC algorithm, the number of employed bees is equal to the number of food sources which is also equal to the number of onlooker bees. There is only one employed bee for each food source whose first position is randomly generated. In each iteration of the algorithm, each employed bee determines a new neighboring food source of its currently associated food source by Equation $V_{ij} = Z_{ij} + \theta_{ij} (Z_{ij} - Z_{kj})$, and computes the nectar amount of this new food source: $v_{ij} = Z_{ij} + \theta_{ij} (Z_{ij} - Z_{kj})$ where θ_{ij} is a random number between [0,1]. If the nectar amount of this new food source is higher than that of its currently associated food source, then this employed bee moves to this new food source, otherwise it continues with the old one.

After all employed bees complete the search process; they share the information about their food sources with onlooker bees. An onlooker bee evaluates the nectar information taken from all employed bees and chooses a food source with a probability related to its nectar amount by Equation $P_i = \frac{f_i t_i}{\sum_{n=1}^{SN} f_i t_n}$. This

method, known as roulette wheel selection method, provides better candidates to have a greater chance of being selected: $P_i = \frac{f_i t_i}{\sum_{n=1}^{SN} f_i t_n}$ where $f_i t_i$ is the fitness value of the solution i which is

proportional to the nectar amount of the food source in the position i and SN is the number of food sources which is equal to the number of employed bees. Once all onlookers have selected their food sources, each of them determines a new neighboring food source of its selected food source and computes its nectar amount. Providing that this amount is higher than that of the previous one, and then the bee memorizes the new position and forgets the old one. The employed bee becomes a scout bee when the food source which is exhausted by the employed and onlooker bees is assigned as abandoned. In other words, if any solution cannot be improved further through a predetermined number of cycles which is called limit parameter, the food source is assigned as an abandoned source and employed bee of that source becomes a scout bee. In that position, scout generates randomly a new solution by Equation. $Z_i^j = Z_{min}^j + rand(0,1)(Z_{max}^j - Z_{min}^j)$ Assume that z_i is the abandoned source and $j \in \{1,2,\dots,D\}$ where D is the solution vector, the scout discovers a new food source which will be replaced with z_i : $Z_i^j = Z_{min}^j + rand(0,1)(Z_{max}^j - Z_{min}^j)$ Where j is determined randomly which is different from i .

A. Detection of the Breast Border using EABCO

A histogram-based thresholding technique is used to generate a binary image to separate the breast and the non-breast region. The EABCO enhances the breast border [5]. Border detectors detect the edges in the binary images, where each pixel takes on either the intensity value zero for a non-border pixel or one for a border pixel. Each pixel in the binary map corresponds to an underlying pixel in the original image. In this proposed system, the breast border is detected by the morphological operations.

1. In each row, the border pixel will be considered as the centre and 20 pixels to their right and left are selected for border enhancement.
2. Now a thick border of the breast along with the thin air film region is obtained as a separate image.
3. EABCO is applied to this image to obtain the enhanced border region.
4. The employee bee of the EABCO visits all the given pixels and finds the best pixels with higher intensity values.
5. The onlooker bee consolidates all the results and converges to a single and unique result, thereby identifying the border pixel in each row.
6. The enhanced border by the EABCO will act as reference points for further segmentation process.

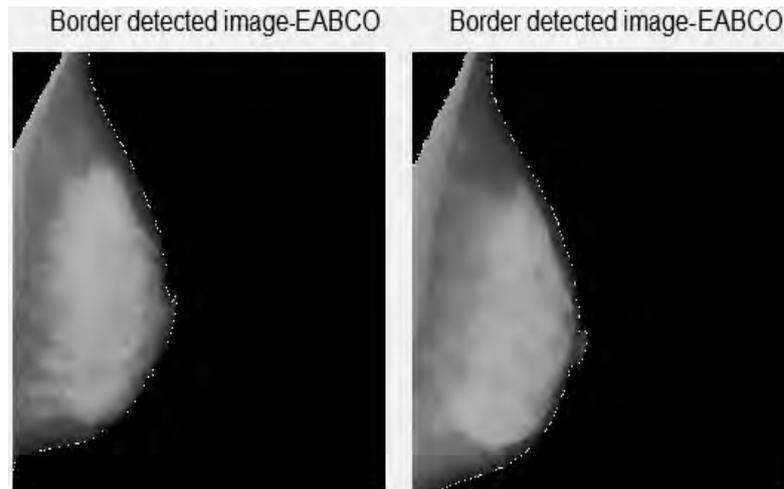


Figure 3: Breast Border detection using EABCO

B. Identification of the Nipple Position EABCO

The nipple may appear either in profile or not in profile. The nipple is located on the mammogram close to where the rate of change among gray levels is larger than in the rest of the breast. This section presents Enhanced Artificial Bee Colony Optimization (EABCO) algorithm to identifying the nipple position. EABCO is a recent heuristic search method whose mechanics are inspired by the swarming or collaborative behaviour of biological populations [1, 2]. The major objective of this paper is to extracting suspicious region from back ground tissue or mammogram. In bilateral subtraction, the asymmetries between corresponding left and right breast images are considered for extracting the suspicious region from the background tissue. The breast border and the nipple position are used as reference points for alignment of mammograms.

Algorithm: Nipple position Identification using EABCO

1. The breast border is considered as the reference region. EABCO is applied only to these reference regions.
2. The pixel values are inverted, to make EABCO as easy process.
3. The employee bee of the EABCO visits each pixel in the given border and shares their best pixels with high intensity with the onlooker bee.
4. The onlooker bee analyse the results of the employee bee.
5. Automatically the results will converge to a single region that defines the nipple position of the given image.
6. The final result of the EABCO will be the nipple position.
7. This enhanced nipple position by the EABCO will act as reference points for further segmentation process.

Figure 4 shows the Identification of the Nipple Position using Enhanced Artificial Bee Colony Optimization algorithm.

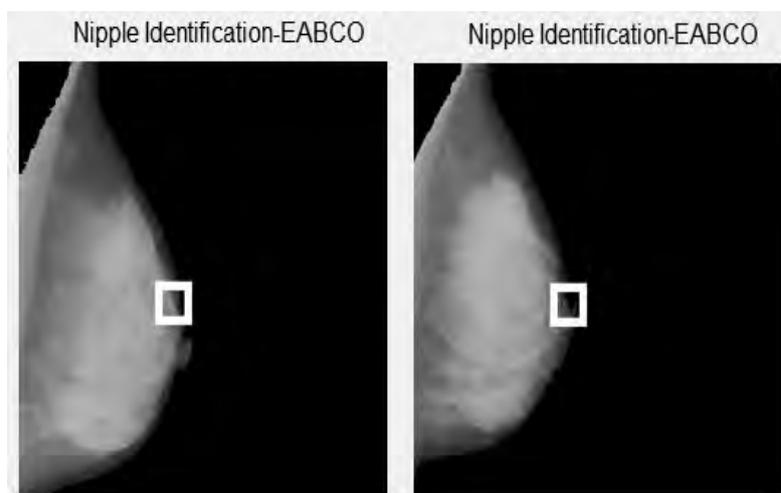


Figure 4: Identification of the Nipple Position using EABCO

C. Alignment of Mammograms

Right and left breast images must be aligned prior to subtraction. Alignment involves the selection of border points and nipple position for transforming the coordinates of one image in order to align the left and the right images. In this section the right breast image is always transformed. The right mammogram is displaced and rotated to align the images and the coordinates of the detected nipples of both the images are used to determine the displacement. Figure 5 shows the transformed image.

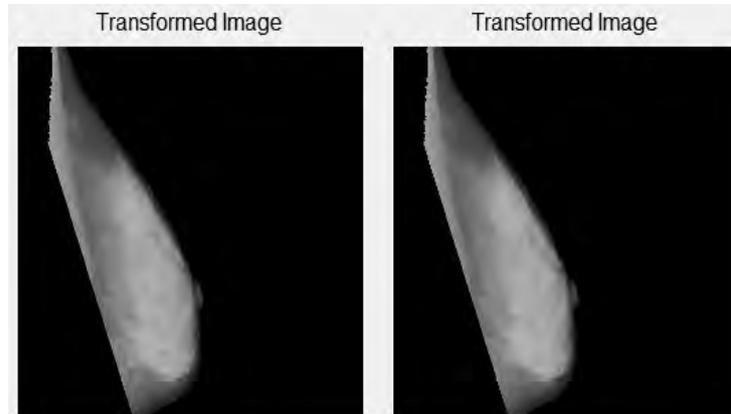


Figure 5: Transformed image

D. Generating the Asymmetry Image

After the images are aligned, bilateral subtraction will be performed by subtracting the digital matrix of the left and right breast image. 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 value is zero, which indicates no difference between the left and the right images. The asymmetry image can be threshold to extract suspicious regions. Figure 6 shows the segmented image using bilateral subtraction

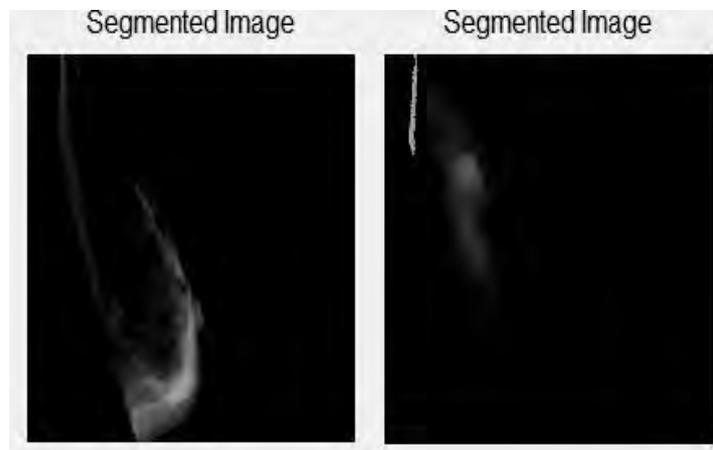


Figure 6: Segmented image using bilateral subtraction

V. EXPERIMENTS AND RESULTS

The effectiveness of the proposed technique is determined by extracting the suspicious region from the mammogram image using bilateral subtraction. The true positive detection rate and the number of false positive detection rate at various thresholds of the asymmetry images are used to measure the algorithm's performance. These rates are represented using receiver operating characteristic curves (ROC). True positive (TP) and false positive (FP) rates are calculated and selected on the segmented image to generate the ROC curve. A region extracted in the asymmetry image that overlaps with a true abnormality as provided on the ground of truth image is called a true positive detection. The ground truth image is taken from the MIAS database. The white region indicates the suspicious region. An overlap means that at least 60% of the region extracted lies within the circle indicating the true abnormality as determined by the MIAS database.

VI. PERFORMANCE STUDY

The ROC curve is a plot of the classifier's true positive detection rate versus its false positive rate [2]. 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. suspicious tumor region). Similarly, the true positive (TP) detection rate is the probability of correctly classifying a target object as being a target object. The ROC is the best suited to analyze the performances of segmentation. The area under the ROC curve (Az value) is an important criterion for evaluating diagnostic performance. The ROC curve is in the range between zero and one. The value of Az is 1.0 when the diagnostic detection has perfect performance, which means that TP rate is 100% and FP rate is 0%. The Az value is computed by trapezoidal rule. The Az value for the proposed algorithm is 0.964.

TABLE I
The Experimental Results and Comparison

Authors	Methods	Detection Rate
Ferrari and Rangayyan, 2001 [4]	Directional Filtering with Gabor wavelets	74.40%
Lau and Bischof, 1991 [13]	Asymmetry Measures	85.00%
The proposed Approach	EABCO and Bilateral Subtraction	96.40%

V. CONCLUSION

The EABCO was used to detect the border and to find the nipple position. The experimental results and comparison shows in table 1. Based on the table, the bilateral approach produces better result. It was observed that the EABCO has performed well. The effectiveness of the proposed technique is determined by extracting the suspicious region from the mammogram image using bilateral subtraction. The true positive detection rate and the number of false positive detection rate at various thresholds of the asymmetry images are used to measure the algorithm's performance.

REFERENCES

- [1] A.E. Eiben, J.E. Smith, Introduction to Evolutionary Computing, *Springer*, 2003.
- [2] R.C. Eberhart, Y. Shi, J. Kennedy, Swarm Intelligence, Morgan Kaufmann, 2001.
- [3] Cheng, H.D., Cai, X., Chen, X.W., Hu, L., and Lou, X.: "Computer Aided Detection and Classification of Microcalcifications in Mammograms: A Survey," *Pattern Recognition*, vol. 36, pp: 2967-2991, 2003
- [4] Ferrari, R.J., Rangayyan, R.M., Desautels, J.E.L., Borges, R.A., and Frere, A.F.: "Analysis of Asymmetry in Mammograms via Directional Filtering With Gabor Wavelets," *IEEE Transactions on Medical Imaging*, vol. 20, no. 9, pp: 953-964, 2001
- [5] D. Karaboga, *An Idea Based On Honey Bee Swarm For Numerical Optimization*, Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.
- [6] Dervis Karaboga, Bahriye Akay," A comparative study of Artificial Bee Colony algorithm" *Journal of Applied Mathematics and Computation*, Elsevier Inc, 214 (2009) 108-132
- [7] D. Karaboga, B. Basturk, A. powerful and efficient algorithm for numerical function optimization: artificial bee colony (abc) algorithm, *Journal of Global Optimization* 39 (3) (2007) 459-471.
- [8] D. Karaboga, B. Basturk, On the performance of artificial bee colony (ABC) algorithm, *Applied Soft Computing* 8 (1) (2008) 687-697.
- [9] D. Karaboga, B. Basturk, in: *Advances in Soft Computing: Foundations of Fuzzy Logic and Soft Computing*, LNCS, vol. 4529/2007, Springer-Verlag, 2007, pp. 789-798.
- [10] D. Karaboga, B. Basturk Akay, C. Ozturk, in: *Modeling Decisions for Artificial Intelligence*, LNCS, vol. 4617/2007, Springer-Verlag, 2007, pp. 318-329.
- [11] D. Karaboga, B. Basturk Akay, An artificial bee colony (ABC) algorithm on training artificial neural networks, in: *15th IEEE Signal Processing and Communications Applications*, SIU 2007, Eskisehir, Turkiye, June, pp. 1-4.
- [12] Karaboga D, Ozturk C, Fuzzy clustering with artificial bee colony algorithm, *Scientific Research and Essays*, Volume: 5 Issue: 14 Pages: 1899-1902, 2010.
- [13] Lau, T.K., and Bischof, W.F.: "Automated detection of breast tumors using the asymmetry approach," *Computers and Biomedical Research*, vol. 24, pp: 273-295, 1991.
- [14] Mendez, A.J., Tahocesb, P.G., Lado, M. J., Souto, M., Correa, J.L., and 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.
- [15] Muttarak, M., Peh, G., and Chaiwun, B.: "Malignant germ cell tumors of undescended testes: imaging features with pathological correlation," *Clinical Radiology*, vol. 59, pp: 198-204, 2004.
- [16] Sallam, M.Y., and Bowyer, K.W.: "Registration and difference analysis of corresponding mammogram images," *Medical Image Analysis*, vol. 3, no. 2, pp: 103-118, 1999.
- [17] 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.
- [18] 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, no.3, pp: 29-40, 2005.
- [19] 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. 5, no. 7 , pp: 41-55, 2005.
- [20] Thangavel, K., and Karnan, M.: "Automatic Detection of Asymmetries in Mammograms Using Genetic Algorithm," *International Journal on Artificial Intelligence and Machine Learning*, vol.5, no.3, pp:55-62, 2005.