

A CAD System for the Detection of Abnormalities in the Mammograms Using the Metaheuristic Algorithm Particle Swarm Optimization (PSO)

Khaoula Belhaj Soulami, Mohamed Nabil Saidi and Ahmed Tamtaoui

Abstract The discovery of a malignant mass in the breast is considered one of the most devastating and depressing health issue women can face. However an early detection can be so helpful and could bring hope to control the disease and even cure it. Nowadays In spite the fact that Digital mammograms have proven to be an efficient tool for the screening of breast cancer, an accurate detection of the abnormalities remains a challenging task for radiologists. In this paper, we propose an effective method for the detection and classification of the suspicious regions. In our proposed approach, we use Entropy thresholding for pectoral muscle removal, and we extract the region of interest (ROI) using the Metaheuristic algorithm Particle Swarm Optimization (PSO). Then we extract Shape and texture features from the abnormalities using Fourier transform and Gray Level Co-Occurrence Matrix (GLCM) respectively. The classification of the detected abnormalities is carried out through the Support Vector Machine, which classifies the segmented region into normal and abnormal based on the extracted features.

1 Introduction

Breast Cancer is the most common worldwide health issue that occurs among middle-aged women, and the leading cause of female cancer deaths. It starts in the tissue of the breast as a group of a dividing cells that forms an abnormal mass known as tumors. They can be cancerous (malignant) tumors or non-cancerous (benign) ones. Early detection plays a fundamental role in cancer prognosis since the death

K.B. Soulami (✉) · M.N. Saidi
National Institute of Posts and Telecommunications
(INPT, CEDOC 2TI, STRS), Rabat, Morocco
e-mail: kbelhajsoulami@gmail.com

M.N. Saidi
e-mail: msaidi@insea.ac.ma

A. Tamtaoui
National Institute of Statistic and Applied Economy (INSEA), Rabat, Morocco
e-mail: tamtaoui@inpt.ac.ma

rate can be significantly reduced. Mammography is currently the most reliable technique for detecting breast abnormalities so the tumor can be treated at an early stage when the cancer would not have been spread yet. However, the identification of the suspicious masses is a tough task, because it is significantly subjective and relies on the radiologists' expertise, and hence can lead to inaccurate predictions. That is why an automated detection using a computer Vision technique is highly recommended to assist radiologists in their diagnosis and give them a second opinion.

Before applying the identification and classification algorithms on the mammograms, a preprocessing task is required, which includes noise reduction, artifacts suppression, and pectoral muscle removal; this step mainly affects both detection and classification of the abnormalities and should be done first. The suppression of the pectoral muscle is highly recommended task in the preprocessing step, it helps in terms of keeping only the breast profile of the mammogram, the removal of this muscle is necessary for the detection of the abnormalities, since it is a high intensity region that has similar features to the abnormal lesions. Many researches were conducted in order to remove the pectoral muscle Yanfeng et al. [1] used homogenous texture and high intensity deviation to identify the edge of the pectoral muscle, then a Kalman Filter was applied to refine the roughness of the edge, the method attains 90% of acceptance. A supervised technique was proposed by Arnau et al. [2], they used a model of three regions in the breast (background, breast and pectoral muscle), and based on intensity, texture, and position information, they applied the training. The approach has shown an overlap between the automated and manual segmentation using 149 mammograms from the Mini-MIAS database. Jawad et al. [3] adopted an approach based on morphological operations and a Seeded Region Growing algorithm to automatically segment the breast profile and remove the pectoral muscle.

After the pectoral muscle removal, comes the step of the detection of abnormalities. There are several types of lesions in the breast, which can indicate cancer, such as microcalcifications, masses, architectural distortions and bilateral asymmetry. Particularly masses are often indistinguishable from the normal breast tissue due to their similar features, thus their detection and classification reveals to be so challenging. Several researchers focused their attention on different techniques to detect and classify abnormal regions. An automated morphological operation based segmentation was proposed by [4] to find the suspicious masses in the breast then the features were extracted from the detected abnormalities using wavelet, and the classification was carried out using Support Vector Machine (SVM). Maitra et al. [5] proposed a Seeded Region Growing Algorithm to isolate normal and abnormal regions in the breast after applying a Divide and Conquer algorithm for mammogram enhancement, followed by an edge detection algorithm, classification was performed using SVM. Anibou et al. [6] used SUSAN algorithm to detect the abnormalities in the high-density breasts, then they applied a Hierarchical watershed transform to detect the edge of the dense regions. They extract the shape features using Fourier Descriptor and an SVM classification was used based on the extracted descriptors and the rate of accuracy using this method achieved 78%.

In this paper, we propose an automated method which detects and classifies the suspicious regions using the metaheuristic algorithm Particle Swarm Optimization,

then we analyze the extracted abnormalities using both shape and Texture features. The content of the paper is organized as follow: Sect. 2 gives an overview of the proposed approach; it describes the preprocessing step, and the techniques used for the segmentation of the breast. This section also illustrates the features extraction methods that we used and describes the procedure of classification. Section 3 presents the details of the image database and gives a highlight of the obtained experimental results using the proposed method. Finally, conclusion is given in the last section.

2 Proposed Method

Our approach is based on a CAD (computer Aided diagnosis) system that takes as an input the mammograms, removes the artifacts and the pectoral muscle in the first place so we can keep only the breast profile, and then we enhance the contrast of the image. We identify the region of interest (ROI) using the Particle swarm Optimization algorithm and we extract both shape and texture features, so we can classify the detected masses into abnormal (cancerous or non-cancerous) or normal ones.

The abnormalities detection in digital mammograms usually consists of the following steps: preprocessing (noise, Artifacts and pectoral muscle removal), segmentation (extraction of the region of interest), features extraction and the classification of the suspicious areas into normal and abnormal. Figure 1 shows the structure of the proposed approach. The following subsections describes each step.

2.1 Preprocessing

Preprocessing methods need to be performed on the mammogram images for the purpose of noise removal, background removal, radiopaque artifacts/label suppression and image contrast adjustment. As the breast profile should optimally be extracted from the background, the pectoral muscle needs also to be removed from the mammograms, since it could bias the process of the identification of abnormalities.

Artifacts and noise removal: This task is so crucial in the preprocessing step, since the radiopaque artifacts are usually sharply defined and bright regions of the mammograms background. It is one of the problems that bias the segmentation of the abnormalities. Generally mammograms contain different types of artifacts which is the case of the Mini-MIAS database images (High intensity labels, low intensity labels, scanning artifacts, Tape Artifacts). We managed to suppress the artifacts using a threshold of 0.16 and then we kept only the largest area which basically includes the breast and the pectoral muscle.

We used Two Dimensional-median filtering in a 3-by-3 connected neighborhood for the purpose of noise removal, since it suppress effectively scratches such as horizontal and vertical lines that tend to appear on most of the mammograms.

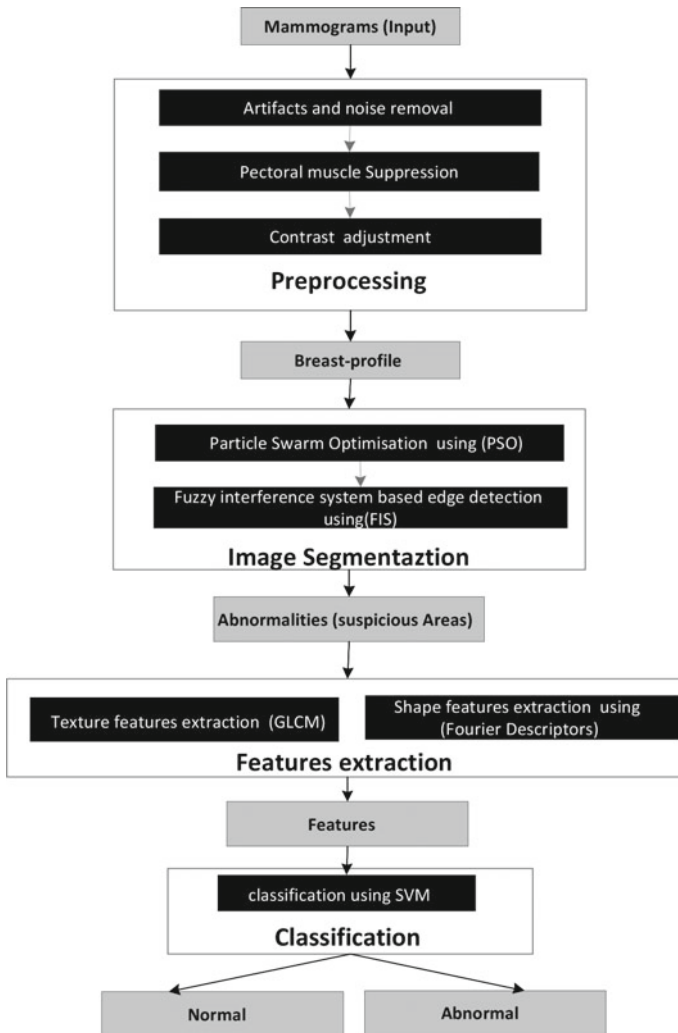


Fig. 1 The proposed CAD for the detection of abnormalities in mammograms

Pectoral muscle suppression: Pectoral muscle is localized in the upper right, left corner of the mammogram, it is a high intensity region that can influence the detection of the suspicious area due to their feature similarity to the abnormalities and hence need to be removed. For this purpose we used a multileveled Minimum Cross Entropy thresholding [8] which has been applied following three levels depending on the density of the mammogram, the more the breast is dense the more it requires a higher level of entropy thresholding because it contains a high intensity region that can be indistinguishable from the pectoral muscle.

Image contrast adjustment: Mammograms adjustment is achieved by performing contrast enhancement. Increasing the contrast of suspicious areas is very essential in mammograms, especially for dense breasts, where the contrast of abnormalities may not be discernable. As a result, differentiating between normal and abnormal regions could be so confusing.

The output of the preprocessing step, consists of the breast part, which will be used in the detection of the suspicious areas (malignant/benign masses).

Remark 1: We applied a morphological operation to refine the rough edges due to the pectoral muscle suppression, especially when it comes to dense breasts.

2.2 Breast Profile Segmentation

Detection of the abnormal masses: The segmentation of the breast profile is a fundamental step that leads to the detection of the lesions; in our method, we used the metaheuristic algorithm Particle Swarm Optimization (PSO).

Particle Swarm Optimization: is a robust stochastic optimization method and a Population-based search procedure that relays on the movement of swarms. It was proposed in 1995 by the social psychologist James Kennedy, from the U.S. Department of Labor Statistics and the electrical engineer and Russell Eberhart from the Purdue University. The particle swarm algorithm applies the concept of social interaction to solve problems, it mimics the principles of social psychology in a way that combines self-experiences with social experience. It was inspired from the simulation of social behavior related to the dynamic movements and communications of insects, birds and fish [9].

PSO uses a number of agents or individuals called particles that constitute a flying around swarm, with a velocity \vec{v}^t , searching the best (optimal) solution in a multi-dimensional search space. Each particle is treated as a point in the space, which adjusts its velocity (1) according to its own flying experience as well as the flying experience of other particles (its neighbors). Which means A PSO system combines local search methods with global search methods, attempting to balance exploration and exploitation, that is why we used it in the detection of the abnormalities which requires both local and global information [10, 11].

$$\vec{v}^{t+1} = \vec{v}^t + c_1 * rand * (\overline{pBest} - \vec{p}^t) + c_2 * rand * (\overline{gBest} - \vec{p}^t) \quad (1)$$

The particle remembers the position where it had its best result. The best solution achieved so far by that particle, known as fitness, and it refers to its personal best (pbest). Particles need help in figuring out where to search, they exchange information about what they have discovered that is why there is another best value that is tracked by the PSO is the best value obtained so far by any particle in the neighborhood. This value is called (gbest) (cf. Algorithm 1). In basic, the co-operation in PSO uses the position of the neighbor with best fitness. This position is simply

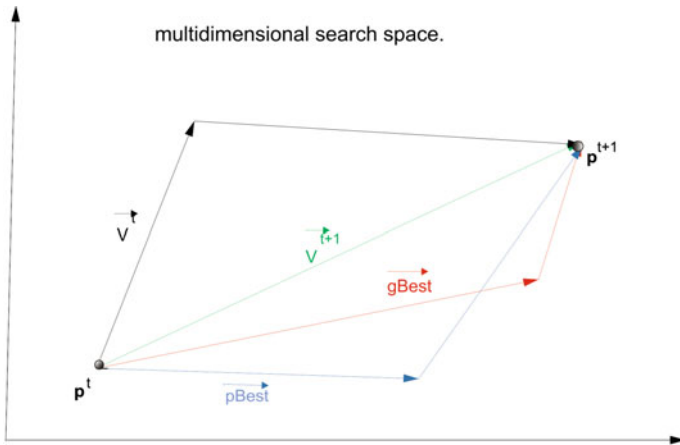


Fig. 2 PSO (particle swarm optimization)

used to adjust the particles velocity. In each iteration, a particle has to move to a new position (2), by adjusting its velocity (1). It relies on random weighted acceleration ($c1, c2$) to accelerate each particle toward its $pBest$ and the $gBest$ locations (Fig. 2).

$$\overrightarrow{p^{t+1}} = \overrightarrow{p^t} + \overrightarrow{v^{t+1}} \tag{2}$$

where p : particles position, v : particle' s velocity, $c1$: weight of local information (importance of personal best), it is the cognition parameter which represent how much the particle trusts its own past experience, $c2$: weight of global information (importance of neighborhood best), it is the social parameter which represents how much the particle trusts the swarm, $pBest$: best position of the particle, $gBest$: best position of the swarm, global best, $rand$: random variable (inertial weight)

```

for each particle do
    Initialize particle;
    Calculate fitness value;
    if the fitness value is better than its peronal best then
        | set current value as the new pBest;
    end
    Choose the particle with the best fitness value of all as gBest;
    Calculate particle velocity(1);
    Update particle position(2)
end
    
```

Algorithm 1: PSO algorithm

Edge detection: consists of finding the boundaries of objects within images. It is used for image segmentation and data extraction. In order to identify the shape of abnormalities, we performed an edge detection algorithm on the extracted Region

of Interest. This task plays an important role in keeping only the important structural properties of the lesions. For this purpose, we have chosen the Fuzzy Interface System based edge detection to detect the profile and shape of the extracted abnormalities. The FIS method was used from MATLAB Fuzzy Logic Image Processing Toolbox.

Fuzzy interference system based edge detection: A Fuzzy Inference System (FIS) is a way of mapping an input space to an output space using fuzzy logic. Instead of Boolean logic, the FIS uses rules and fuzzy membership functions, to reason about data. The membership functions define the degree to which a pixel belongs to an edge or not. The choice of membership function is problem dependent. But the most used function is “Triangular Membership function” (3), which is defined as:

$$f(a, b, c) = \max\left(\min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right), 0\right) \quad (3)$$

where a and c are the feet of the triangle and the parameter b defines the peak.

We have detected the edges of the abnormalities by comparing the gradient of every pixel in the x and y directions. If the gradient for a pixel is not zero, then the pixel belongs to an edge (white). We defined the gradient as zero using Gaussian membership functions for the FIS inputs.

2.3 Features Extraction

During feature extraction, the most important characteristics of the ROIs are studied and analyzed.

Shape feature extraction: The shape of the abnormalities is an important criterion which indicate whether the extract masses is abnormal (cancerous/non-cancerous) or not, so in order to extract the shape information from the abnormalities we used Fourier descriptor which is invariant to translation, rotation.

Fourier Descriptors: Fourier descriptors is a way of encoding the shape of a two-dimensional object by taking the Fourier transform of the boundary, where every point on the boundary is mapped to a complex number. To apply FD on the detected boundaries, two steps needs to be followed:

1. normalisation of the contour: In order to use the fast Fourier transform (FFT) properly we have to normalize the number of data set extracted from the edge, because the contours are different in shape and size.
2. calculation of the shape features using Fourier descriptor (4).

$$DF_n = \frac{1}{N} \sum_{k=0}^{N-1} r(k) \exp\left(\frac{-i2\pi nk}{N}\right), n = 1, 2 \dots N - 1, \quad (4)$$

where N : is the number of normalized points, $r(k)$ is the centroid distance function which represents the distance of the boundary points from the centroid (x_c, y_c) of the shape which is basically the average of the boundary coordinates.

Texture features extraction: The analysis of textures has proven a high efficiency in the detection of breast cancer, since texture is really outstanding when it comes to identifying specific characteristics of breast abnormalities. In our method, the texture-based features are extracted from the ROI region using Gray Level Co-Occurrence Matrices (GLCM).

The Grey-level Co-occurrence Matrix (GLCM): Level Co-occurrence Matrices (GLCMs) is one of the stunning texture analysis techniques. GLCM is a square matrix with dimension N_g (Number of Grey Levels) (5) that contains the occurrence of the combinations of grey level values. It gives an idea about the properties of the spatial distributions of the pixel intensity values in grayscale images. The parameters required for computing the GLCM are:

- Number of Grey Levels: usually it is 256 grey levels.
- Distance between Pixels: the matrix could be computed using non-neighbors pixels. Hence a distance between pixels is defined.
- Angle: the direction of the pair of pixels (0, 45, 90, 135).

$$G = \begin{bmatrix} p(1, 1) & p(1, 2) & \dots & p(1, N_g) \\ p(2, 1) & p(2, 2) & \dots & p(2, N_g) \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ p(N_g, 1) & p(N_g, 2) & \dots & p(N_g, N_g) \end{bmatrix} \tag{5}$$

where $p(i, j)$ is the sum of the occurrence of a pixel “i” in the specified spatial relationship to a pixel “j” in the input image.

In this paper apart from using 11 descriptors texture proposed by Haralick et al. [12], we used other recent texture descriptors [13, 14] and some features from the MATLAB Image Processing Toolbox.

2.4 Classification

Classification is a process related to categorization, the process in which objects are recognized, differentiated, and understood. In the classification step, the dataset is split into two disjoint sets: training and test. The training set is used to train the learning machine and the trained learning machine is then tested on the test set. In this paper the dataset sample was divided into two subsets from which one set was chosen as a training one and the other one was used for test.

In this work, the support vector machine (SVM) was performed using Sigmoid kernel [15]. SVM is basically a linear classification approach based on two classes. It separate individuals from two classes (+1 and -1) using the optimal hyperplane that separate the two sets, and guarantee a large margin between the two classes.

3 Experimental Results

3.1 *Mini-Mias Database*

Digital mammogram images were acquired from the mini-MIAS database [7] which consist of right and left breast images of dense, fatty-glandular and fatty breasts. The acquired mammogram images belongs to three categories: malignant, benign and normal. The abnormalities (benign and malignant) consists of five categories as follows: Ill-defined masses, architecturally distorted masses, Asymmetrical masses, Circumscribed masses and Spiculated masses. The size of the images is 1024 1024 pixels. The images are in grayscale with a pixel intensity of range [0, 255].

3.2 *Preprocessing*

The mammograms of Mini-MIAS database was preprocessed using the techniques described in Sect. 2.1 as the figure (Fig. 3) shows, the preprocess was applied on the three categories of the breast (fatty, fatty glandular, dense), this methods still have some drawbacks when it comes to the removal of pectoral muscle in dense mammograms. To avoid the over segmentation of the breast, we have chosen the level of Entropy thresholding manually, since in this case, the dense tissue of the breast is indistinguishable from the pectoral muscle.

3.3 *Segmentation*

The identification of the region of interest (abnormalities) was carried out using the segmentation methods described in Sect. 2.2. PSO algorithm was first applied on the preprocessed images followed by a fuzzy logic algorithm based edge detection, the figures (cf. Figs. 4, 5 and 6), show the experimental results of this step and it has been performed on the three different categories of the breast (dense Fig. 4, fatty glandular Fig. 5 and fatty Fig. 6). The majority of abnormalities was detected and there was cases where the output image was blank and thats describe a normal breast tissue, which supposed to not contain any abnormalities, this kind of results has fit our expectations.

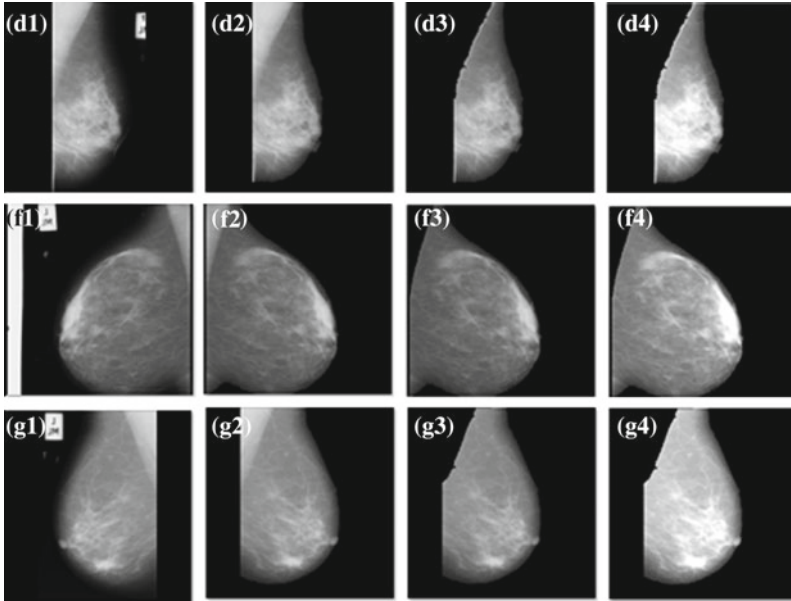


Fig. 3 The preprocessing step was performed on the three categories dense (d), fatty (f) and fatty glandular (g) respectively, **d1, f1, g1** original images **d2, f2, g2** refers to the suppression of artifacts and noise in the three categories. **d3, f3, g3** the removal of pectoral muscle, **d3, f3, g3** the contrast adjustment of the images, **d4, f4, g4**

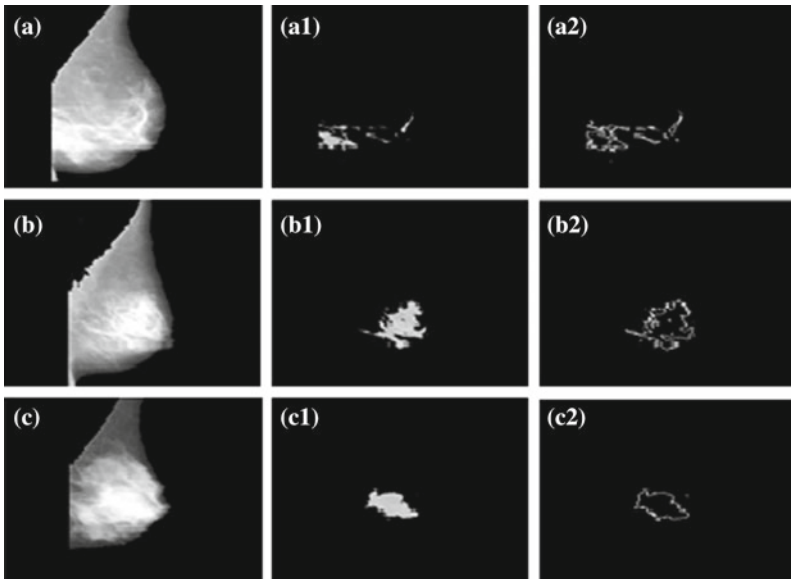


Fig. 4 **a, b, c** the preprocessed images of the dense breast category, which represents the normal malignant and benign cases respectively, **a1, b1, c1** the detected abnormalities using PSO, **a2, b2, c2** the edge of abnormalities using FIS

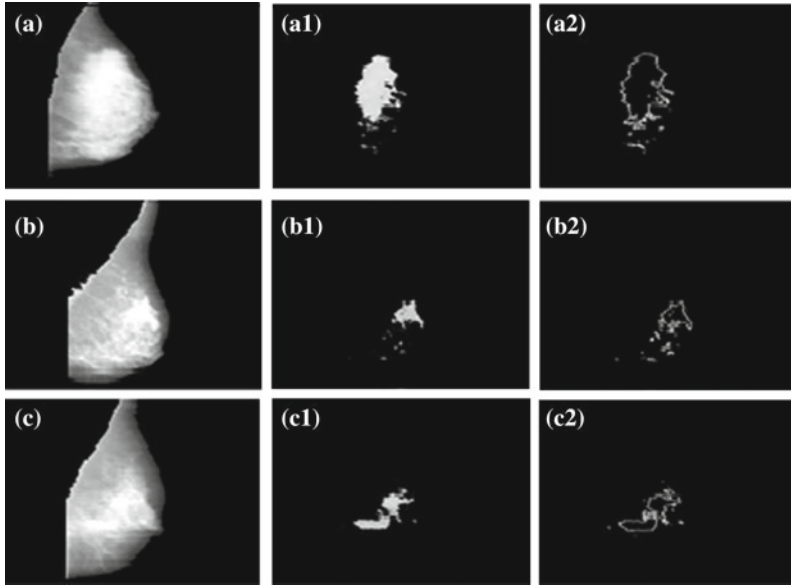


Fig. 5 a, b, c the preprocessed images of the fatty glandular breast category, which represents the normal malignant and benign cases respectively, a1, b1, c1 the detected abnormalities using PSO, a2, b2, c2 the edge of abnormalities using FIS

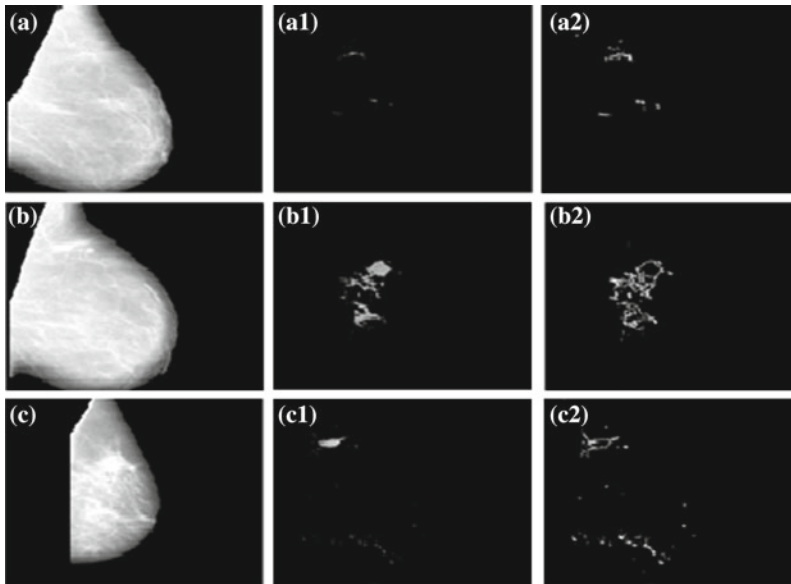


Fig. 6 a, b, c the preprocessed images of the fatty breast category, which represents the normal malignant and benign cases respectively, a1, b1, c1 the detected abnormalities using PSO, a2, b2, c2 the edge of abnormalities using FIS

Table 1 Comparison with other techniques of detection of abnormalities in term of accuracy

Methods	Segmentation	Features	Classifier	Data-Set	Accuracy (%)
[16]	Small patches of 128 × 128 pixels	2DPCA	SVM	IRMA reference DDSM	80.07
[6]	Modified Susan and Watershed Transform	Fourier (centroid distance)	SVM(Sigmoid Kernel)	Mini-MIAS	78.77
The proposed method	Particle Swarm Optimization	Shape (Fourier descriptors) + Texture (GLCM descriptors)	SVM(Sigmoid Kernel)	Mini-MIAS	83.87

3.4 Feature Extraction and Classification

The obtained features from both methods FD and GLCM of the Sect. 2.3, were merged randomly, and normalized so they can fit properly the SVM. All the 107 features out of which 63 are shape features and the remaining describes the texture, were scaled (normalized) in the range between 0 and 1, the Feature normalization has been carried out using the following expression (6):

$$NF(x) = \frac{F(x) - \min(F(x))}{\max(F(x)) - \min(F(x))} \quad (6)$$

where $F(x)$ represents the feature of interest.

The normalized features are divided into two distinct sets, i.e. the training set and the testing set. The total number of ROI samples obtained from the acquired segmented breast data is 306, out of which 195 are normal samples and the remaining 111 are abnormal samples. where 80 % of the sample from both classes (normal/abnormal) was randomly allocated to the training set and the remaining 20 % of the sample from both classes was chosen as a testing set. The Performance of the proposed method is evaluated in terms of Accuracy which attend 83.87 % (Table 1).

4 Conclusion

In this paper, we have proposed a Computer Aided Diagnosis system that detect the abnormalities in digital mammogram and classifies them into normal and abnormal. The acquired images from Mini-MIAS database were preprocessed in order to remove noise, artifacts and pectoral muscle from the breast region so the segmentation algorithms could perform efficiently. Then we have extracted the suspicious regions using PSO algorithm, followed by an edge detection technique based on

FIS. We computed shape descriptors from the edge of abnormalities using Fourier Descriptors, then we extracted the texture-based features from the suspicious regions using GLCM. Both shape-based descriptors and texture-based ones were normalized and stored as feature vector. A support vector machine was carried out to classify suspicious regions into normal or abnormal. The proposed method was tested on Mini-Mias database. For further work, we want to evaluate our method on different private databases, and automate the entropy level thresholding for pectoral muscle removal, so we do not have to interfere manually, we will also try to detect the cancerous regions.

References

1. Lia, Y., Chena, H., Yangb, Y., Yanga, N.: Pectoral muscle segmentation in mammograms based on homogenous texture and intensity deviation. *Pattern Recogn.* **46**(3), 681–691 (2013)
2. Oliver, A., Llado, X., Torrent, A., Mart, J.: One-shot segmentation of breast, pectoral muscle, and background in digitised mammograms. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 912–916
3. Nagi, J., Kareem, S.A., Nagi, F., Ahmed, S.K.: Automated breast profile segmentation for ROI detection using digital mammograms. In: IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), pp. 87–92. Kuala Lumpur, Malaysia (2010)
4. Anitha, J., Peter, J.D.: A wavelet based morphological mass detection and classification in mammograms. In: International Conference on Machine Vision and Image Processing (MVIP), pp. 25–28 (2012)
5. Maitra, I.K., Nag, S., Bandyopadhyay, S.K.: Detection of abnormal masses using divide and conquer algorithm in digital mammogram. *Int. J. Emerg. Sci.* **1**(4), 767–786 (2011)
6. Anibou, C., Saidi, M.N., Aboutajdine, D.: Computer aid diagnostic in mammogram image using susan algorithm and hierarchical watershed transform. In: Lecture Notes in Computer Science, UNet 2015, pp. 355–366 (2016)
7. J. Suckling et al., The Mammographic Image Analysis Society digital mammogram database, *Exerpta Medica* **1069**, 375–378 (1994)
8. Brink, A.D., Pendock, N.E.: Minimum cross-entropy threshold selection. *Pattern Recogn.* **29**, 179–188 (1996)
9. Ait-Aoudia, S., Guerrou, E.-H., Mahiou, R.: Medical image segmentation using particle swarm optimization. In: 18th International Conference on Information Visualisation (IV), pp. 287–291 (2014)
10. Ghamisi, P., Couceiro, M.S., Martins, F.M.L., Benediktsson, A.: Multilevel image segmentation based on fractional-order darwinian particle swarm optimization. *IEEE Trans. Geosci. Remote Sensing* **99**, 1–13 (2013)
11. Raju, N.G., Rao, P.A.N.: Particle swarm optimization methods for image segmentation applied in mammography. *Int. J. Eng. Res. Appl.* **3**(6), 1572–1579 (2013)
12. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features of image classification. *IEEE Trans. Syst. Man Cybern.* **SMC-3**(6) (1973)
13. Soh, L., Tsatsoulis, C.: Texture analysis of SAR sea ice imagery using gray level co-occurrence matrices. *IEEE Trans. Geosci. Remote Sens.* **37**(2) (1999)
14. Clausi, D.A.: An analysis of co-occurrence texture statistics as a function of grey level quantization. *Can. J. Remote Sens.* **28**(1), 45–62 (2002)
15. Sharma, S., Khanna, P.: Computer-aided diagnosis of malignant mammograms using zernike moments and svm. *J. Digit. Imaging* **28**(1), 77–90 (2015)
16. Deserno, T.M., Soiron, M., de Oliveira, J.E.E.: Computer-aided diagnostics of screening mammography using content-based image retrieval. *Proc. SPIE* **8315**, 271–279 (2012)