Future Computing and Informatics Journal

Volume 3
Issue 2 *Vol 3,Issue 2,2018*

Article 19

2018

Benign and malignant breast cancer segmentation using optimized region growing technique

S. Punitha
Department of Computer Science, Pondicherry University, Pondicherry, India, punitharesearch@gmail.com

A. Amuthan
Department of Computer Science, Pondicherry Engineering College, Pondicherry, India, amuthan1973@gmail.com

K. Suresh Joseph
Department of Computer Science, Pondicherry University, Pondicherry, India, ksjoseph.csc@gmail.com

Available online at www.sciencedirect.com

ScienceDirect

Future Computing and Informatics Journal 3 (2018) 348–358
http://www.journals.elsevier.com/future-computing-and-informatics-journal/

Benign and malignant breast cancer segmentation using optimized region growing technique

S. Punitha [a],*, A. Amuthan [b], K. Suresh Joseph [a]

[a] Department of Computer Science, Pondicherry University, Pondicherry, India
[b] Department of Computer Science, Pondicherry Engineering College, Pondicherry, India

Received 14 June 2018; accepted 24 October 2018
Available online 3 November 2018

Abstract

Breast cancer is one of the dreadful diseases that affect women globally. The occurrences of breast masses in the breast region are the main cause for women to develop a breast cancer. Early detection of breast mass will increase the survival rate of women and hence developing an automated system for detection of the breast masses will support radiologists for accurate diagnosis. In the pre-processing step, the images are pre-processed using Gaussian filtering. An automated detection method of breast masses is proposed using an optimized region growing technique where the initial seed points and thresholds are optimally generated using a swarm optimization technique called Dragon Fly Optimization (DFO). The texture features are extracted using GLCM and GLRLM techniques from the segmented images and fed into a Feed Forward Neural Network (FFNN) classifier trained using back propagation algorithm which classifies the images as benign and malignant. The performance of the proposed detection technique is evaluated using the images obtained from DDSM database. The results achieved by the proposed pixel-based technique are compared to other region growing methods using ROC analysis. The sensitivity of the proposed system reached up to 98.1% and specificity achieved is 97.8% in which 300 images are used for training and testing purposes.

1. Introduction

Breast cancer is the first commonly occurring cancer globally according to the reports of the world health organization. According to Indian Council of Medical Research (ICMR) in 2015–2016, about 150000 women are affected by breast cancers and the death rate was 50 percentage. In 2016–2017 statistics, according to American Cancer Society's 255180 new cases of breast cancer are expected in United States (American Cancer Society, 2016). The lifestyle changes from traditional to modern by the people in developed and developing countries increases the occurrence of breast cancers among women mostly in the age of 35–55. The occurrence of breast cancers can be controlled by identifying the breast cancers in its early stages [1]. The necessity for early and accurate diagnosis of breast masses and micro-calcifications plays a very important role in decreasing the death rate. Since manual methods that are followed by radiologists fail due to the similarity in appearance of breast masses and microcalcifications as of its background segmentation of such abnormalities is a challenging task. The need for early detection requires developing automated systems to assist radiologists in diagnosing the breast cancers accurately and necessary treatments to the patients are further followed.

Screening methods used for breast cancer screening include magnetic resource imaging (MRI), self and clinical breast checks, ultrasound, and mammography [2]. Mammography is

Keywords: DFO, Dragon fly optimization; GLCM, Gray level co-occurrence matrix; GLRLM, Gray level run length matrix; ROC, Receiver operating characteristic; FFNN, Feed forward neural network.

* Corresponding author.
E-mail addresses: punitharesearch@gmail.com (S. Punitha), amuthan1973@gmail.com (A. Amuthan), ksjoseph.csc@gmail.com (K.S. Joseph).

Peer review under responsibility of Faculty of Computers and Information Technology, Future University in Egypt.

https://doi.org/10.1016/j.fcij.2018.10.005

the most reliable and efficient x-ray procedure to identify breast masses. Film mammography is replaced by the digital mammography where special high-quality computerized equipment are used to record images of the breast from the patients and used for further processing like detection and classification. Microcalcifications and masses are the most common abnormality that leads to breast cancer. The image that is produced through mammography is called a mammogram which consists of the background, breast region, fat tissue and the breast masses and microcalcifications with high intensities. Breast masses and microcalcifications occur in the epithelial and connective tissues of breast region [3]. Since the demand for processing the mammograms is increasing, the radiologists may create errors neglecting important clues due to fatigueness [4].

Breast masses appear as a lump in the breast region with different shapes and sizes. The severity of breast masses can be categorized as benign and malignant. Benign breast masses are non-cancerous and not aggressive but they grow and press the surrounding organs that lead to other complications. Malignant breast masses are cancerous and aggressive that has to be treated soon to prevent the death of a patient. Round, oval with smooth and circumscribed margins shape masses is normally benign masses. The irregular masses have a high chance of malignancy. The speculated, rough and blurred masses are categorized as malignant breast masses.

Microcalcifications are calcium deposits which appear as small bright spots in a mammogram. Since breast masses and microcalcifications appear similar to that of the background in the mammogram, the image processing techniques play an important role in early detection of breast masses. Since the detection and classification are difficult in the case of breast masses. Many techniques have been developed by researchers to identify the exact location of breast masses. The evaluation of these techniques can be done based on how the techniques identify the true and false breast masses which can be identified by comparing the attained results with the ground truth markings provided by the radiologists [5].

Since the benefits and challenges of early detection and classification of breast cancer is high developing an automated system to support the skilled radiologists will ensure the high accuracy of the interpretation process. Based on this motivation the paper has the following contributions towards accurate breast cancer detection and classification.

i. The Mammograms collected from the DDSM database are subject to pre-processing where 7 × 7 Gaussian filter is used for noise removal and smoothening the gray level variations in the mammograms.
ii. The mammograms are subject to intensive intensity variations choosing for a fixed threshold for segmentation of abnormal regions will not be sufficient and hence the proposed methodology in the paper aims at achieving the best segmentation that can be done to separate the Regions of Interest (ROI) containing the breast masses and microcalcifications from the digital mammograms by using an RG algorithm based Renyi entropy method to select the optimum seed pixel and threshold.
iii. Texture features plays a vital role in classification of the breast abnormalities we incorporate the texture feature extraction using GLCM and GLRLM techniques from the segmented Regions of Interest (ROI).
iv. Artificial Neural Networks plays an increasingly prominent role in medical diagnosis. The proposed methodology employs a Feed Forward Neural Network with back propagation learning algorithm to classify the abnormal regions as Benign and Malignant masses.

This paper is arranged as follows. The section 2 contains the related works for the proposed methodology. The section 3 contains the proposed methodology in detail depicting the data acquisition, preprocessing, segmentation, feature extraction and classification. Section 4 gives the performance analysis which depicts the segmentation and classification performance of the proposed methodology. Section 5 gives the conclusion of this work.

2. Related works

Rahimeh et al. [6] proposed two techniques for segmentation of breast masses using Region growing where Artificial Neural networks are trained to produce the seeds and thresholds of the segmentation process. The intensity and texture features are extracted and fed in to a Neural classifier to classify the benign and malignant mammograms. The obtained sensitivity, specificity, and accuracy rates are 96.87%, 95.94%, and 96.47%, respectively. Rakoth Kandan et al. [7] proposed a self-adaptive segmentation approach based on dragonfly optimization for multilevel thresholding where optimal thresholds are generated using swarm optimization approach. The real and medical images are used for testing in which the self-adaptive dragon fly optimization is proved to effectively optimize the threshold values. G. Kom et al. [8] proposed a detection algorithm for breast masses which uses a linear transformation enhancement filter is used for the enhancement of local contrast of each pixel and local adaptive threshold technique is used for the binarization of the subtracted images from the original image that contains the masses. The sensitivity of this proposed method reached up to 95.91% when it was tested on a set of 61 mammograms. A. Vadivel et al. [9] proposed a fuzzy system for detection and classification of breast tumors where fuzzy rules are framed using trapezoidal fuzzy membership functions for the shape classification of masses in to round, oval, lobular and irregular. This system proposed some advanced set of shape and margin features where C5.0 decision tree algorithm is used for the generation of rules in the fuzzy inference system. A set of 224 images from DDSM database are used for testing where the maximum classification accuracy was 100% for round and oval masses. Xiaoming Liu et al. [10] proposed an improved method of region growing called multiple concentric layers (MCL) where prior knowledge is used in the phase of training and narrow band based active contour (NBAC) method is used for improving the accuracy. This MCL breast mass detection technique attained a sensitivity of 82.4% and with active contour refinement, feature analysis and classification it attained sensitivity of 78.2% when it is tested on set of 164 mammograms.

Aswini Kumar Mohanty et al. [11] proposed a system which extracts the ROI using the contour provided by the DDSM database and 19 texture features are extracted using gray-level co-occurrence matrix (GLCM) and gray-level run-length matrix (GLRLM) and the classification of breast masses are done based on decision tree C5.0 DT algorithm. The level of accuracy achieved is 93.6% and area under receiver operating curve was 0.995. Weiying Xie et al. [12] proposed a system where the Region of Interest extraction that contains the breast masses is done using Circular Hough Transform (CHT) and a level set method. About 32 features containing the mass, boundary and background features are extracted and 30 features are selected using the scoring provided b SVM and ELM classification methods. The average accuracy obtained is 96.02% when it is tested on the images collected from mini MIAS database and DDSM database. The classification is based on Extreme Machine Learning which outperforms traditional SVM and PSO based SVM. Danilo et al. [13] proposed a system for detection of breast masses where image denoising is done using top hat morphological operations, subtraction operation, Ostu thresholding and multiplication operations. The images are decomposed into different resolutions using 2DWT filters and an adaptive wiener filter is applied on the decomposed images for noise removal. The Hammouche's algorithm which is a combination of wavelet transform, and genetic algorithm are used for segmenting the breast masses. The proposed method was quantitatively evaluated with the images collected from Digital Database for Screening Mammography (DDSM) in CC and MLO views in which the mean ± standard deviation value of area overlap metric was 79.2 ± 8%.

3. Proposed methodology

The proposed methodology consists of different stages involving image processing techniques in each stage. The first stage is the image acquisition where the normal and abnormal mammograms are collected from http://marathon.csee.usf.edu/Mammography/Database.html Digital Database for Screening Mammography (DDSM). The digital mammograms are then preprocessed using Gaussian filters for noise removal. The noise filtered images are further processed to extract the region of interest (ROI) that are targets of breast masses using optimized region growing method. The ROIs are then processed for feature extraction where a set of texture features of the ROIs are extracted using Gray level co-occurrence Matrix (GLCM) and Gray level Run Length Matrix (GLRLM). The extracted texture features are further fed into a Feed Forward Neural Network (FFNN) which is trained using back propagation algorithm called Levenberg–Marquardt to classify the mammograms as normal, benign and malignant. Fig. 1 shows a diagrammatic representation of the proposed methodology.

3.1. Data acquisition

The digital images used in this proposed method are breast images taken from patients collected from) which is used in the work [14,15] is available at http://marathon.csee.usf.edu/Mammography/Database.html. The DDSM database was created by Massachusetts General Hospital in the University of South Florida and Sandia National laboratories. The database contains 2500 cases of digital mammograms. Patient information based on Breast Imaging Reporting and Data System (BI-RADS) and image information are also included in the database. Breast ill-defined and well-defined masses, spiculated masses, microcalcifications and architectural distortions are the severity cases included in the database. A set of 300 digital mammograms containing the severity of normal, benign and malignant are used in this work for training and testing purposes in both MLO and CC views. This set contains 154 benign cases, 146 malignant cases of both left and right breasts. The background tissues in the mammograms are grouped as fatty, granular and dense tissues. The sample images from DDSM database is shown in Fig. 2.

3.2. Pre-processing

The Breast masses that are present in the digital mammograms appear brighter than the background and the preprocessing filter used should be able to retain its natural intensity characteristics while removing the unwanted noise portions. The proposed system uses a 7 × 7 Gaussian filter which is a non-uniform low-pass filter for preprocessing the digital mammograms where the noises are removed and the images are smoothed thereby eliminating its intensity in homogeneities and retain its gray level variations without which the segmentation algorithms may misinterpret to find out the real breast masses. The Gaussian filter for a pixel (i, j) that is used in the proposed system uses a two-dimensional Gaussian distribution function called as point spread function which is shown in equation (1).

$$G(i,j) = \frac{1}{2\pi\sigma^2} e^{\frac{-i^2+j^2}{2\sigma^2}} \tag{1}$$

Where σ is the standard deviation of the distribution. The Gaussian kernel values are used to frame a convolution filter. The convolution filter is applied on the digital mammograms on all the pixels. The matrix multiplication is performed between the kernel values and the intensity values of each pixel of the digital mammograms. Thus the mammograms are subjected to noise removal and smoothening using Gaussian filtering as shown in Fig. 3.

3.3. Segmentation

Segmentation is the process of separating the benign and malignant masses from the background portions by partitioning the digital mammograms into nonoverlapping segments. This segmentation process of finding the breast masses is

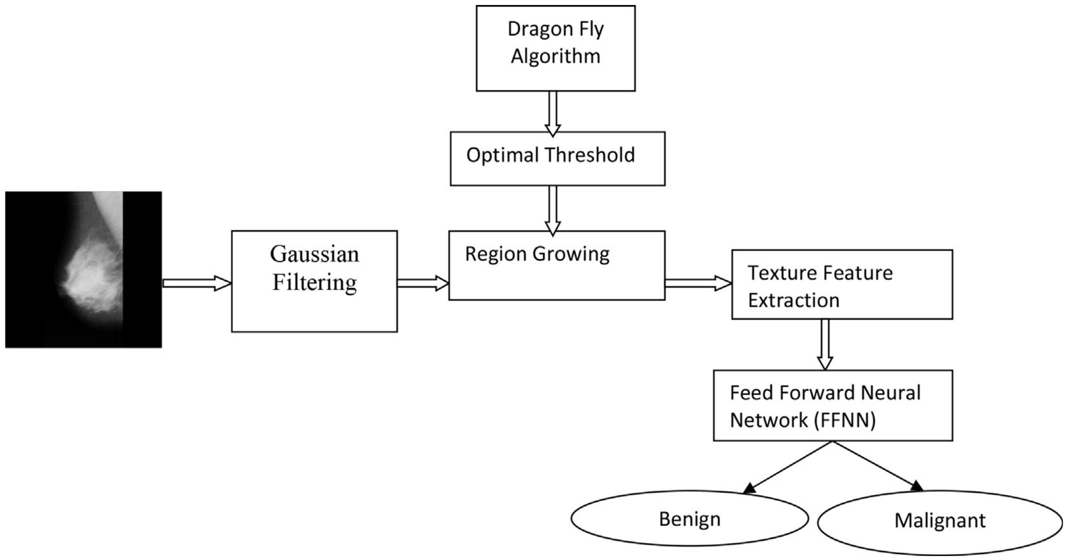


Fig. 1. Architecture of the Proposed System using Optimized Region Growing.

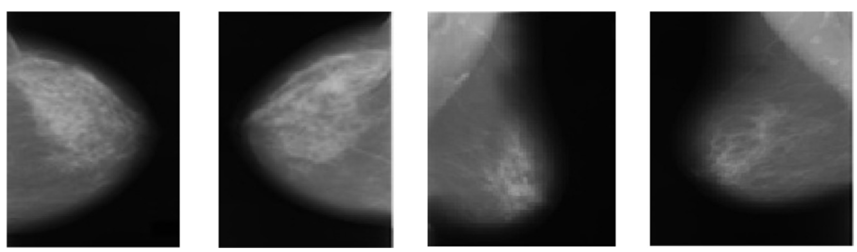
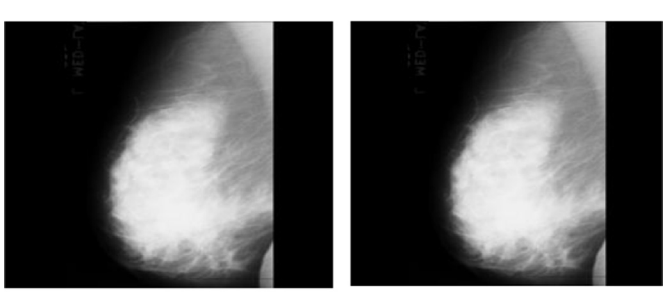


Fig. 2. Sample Images of the left and right breasts in CC and MLO views extracted from DDSM database.

accomplished by various algorithms such as classical approaches that include Global thresholding and local thresholding based on image histograms. Region-based methods, such as region growing where a seed point is used and the region is grown until a homogeneity criterion is met. Markov Random Field (MRF) segmentation method separates the ROIs using maximum a posterior (MAP) functions for the mammograms and clustering methods are also used to segment the ROIs that contain the breast masses present in digital mammograms. The template matching methods are used for segmentation of the abnormal portions of the mammograms where a template is created using the characteristics of the masses and the suspicious regions are matched with the templates. Stochastic relaxation methods partition the image in an unsupervised manner using generic labels. Fuzzy techniques use fuzzy membership function to generate membership values for each and every pixel and for every iteration, an error value is generated and fuzzy rules are updated. Bilateral Image subtraction aligns the left and right breasts and the suspicious regions are detected using the difference between the left and right breast images. This paper proposes an improved version of region growing which comes under pixel based methods where the seed points and the thresholds for homogeneity criteria are generated using Dragon fly optimization approach which is a swarm based optimization technique.



(a) **(b)**

Fig. 3. (a) Original Image. (b) Pre-processed image after Gaussian Filtering.

3.3.1. Optimized region growing

Region growing is a pixel-based segmentation approach where similarity constraints such as intensity, the texture so on are considered for grouping the pixels into regions. Initially, a set of pixels is merged through iteration process based on the similarity constraints. A seed pixel is chosen and the region is grown by grouping the neighboring pixels that are similar where the region size increases. The region growth is stopped when none of the neighboring pixels fulfill the homogeneity criteria and another new seed pixel is chosen. This process continues till all the pixels in the image belong to some region. In region growing segmentation technique, the selection of seed points and thresholds for deciding the homogeneity constraints plays an important role in increasing the segmentation accuracy. Since mammogram suffers from lot of intensity variations choosing a constant threshold will not lead to accurate segmentation. Hence this paper focuses on developing an automated method based on Dragon Fly optimization technique to generate an optimal seed point and threshold.

Steps for Optimized Region Growing Segmentation:

The stepwise procedure for the optimized region growing algorithm is presented below

i. Input the abnormal image.
ii. Let t be the optimized threshold generated using Dragon Fly algorithm.
iii. Set t as the seed point for region growing algorithm.
iv. Add 4 neighborhood pixels.
v. Calculate the distance (d) between the mean intensity of the new neighbor pixel and the mean of the region intensity.
vi. Perform region growing if d $\leq$ t on four neighborhood pixels and include each if they are not included already in the region and save the coordinates of the new pixel.
vii. Save the mean of the new region and return to step 2 and perform region growing until all pixels are grouped.

3.3.2. Dragon fly algorithm

Dragon fly algorithm is a swarm intelligence optimization algorithm. The Dragon fly algorithm is framed based on the static and dynamic behavior of dragonflies for hunting and migration respectively. The static and dynamic behavior of dragonflies can be modeled as two phases of optimization such as exploration and exploitation [16]. The dragonflies are attracted to the direction where the food is available and distracted from the enemies for survival. The behavior of Dragonflies can be modeled using separation, alignment, and cohesion.

Separation can be represented by equation (2)

$$S = \sum_{j=1}^{NI} P_i - P_j \tag{2}$$

where P_i indicates the position of the current dragonfly, P_j represents the position of the jth neighbouring dragonfly and NI represents the number of neighbouring dragonflies.

Alignment can be represented by equation (3)

$$A = \frac{\sum_{j=1}^{NI} V_j}{ND} \tag{3}$$

where V_j represents the velocity of the jth neighbouring dragonflies and ND represents the number of neighbouring dragonflies.

Cohesion can be represented by equation (4)

$$C = \frac{\sum_{j=1}^{ND} P_j}{ND} - P_i \tag{4}$$

where P_i indicates the position of the current dragonfly and P_jP_j represents the position of the jth neighbouring dragonfly and ND represents the number of neighbouring dragonflies.

Attraction to food source is represented by equation (5)

$$FS = Source - P_i \tag{5}$$

where source represents the food source position.

The distraction from the enemies is represented using equation (6)

$$DE = Worst + P_i \tag{6}$$

Where Worst represents the position of the enemy.

The velocity vector of each dragonfly at time t+1 of the dragonfly x is calculated using equation (7)

$$\Delta U_{t+1} = (sS_x + aA_x + cC_x + fFS_x + eDE_x) + w\Delta U_t \tag{7}$$

The position vector U of each dragonfly at a time t+1 can be calculated using equation (8)

$$U_{t+1} = U_t + \Delta U_{t+1} \tag{8}$$

The position of the dragonfly is updated using equation (9)

$$U_{t+1} = U_t + Le'vy\ (\alpha) \times U_t \tag{9}$$

The optimal thresholds for region growing are obtained by maximization based on Renyi's entropy method [17] using the equation (10)

$$F_{obj} = Max\left(\sum H_b^a(x,y) + \sum H_o^a(x,y)\right) \tag{10}$$

X represents the threshold of the gray level of the pixel and y represents the threshold of the average gray level of the neighbors of the pixel and a is the positive real number such that a$\neq$1.

Where $H_b^a(x, y)$is the Renyi entrophy of the background represented by equation (11) and $H_o^a(x,y)$ is the Renyi entrophy of the breast region represented by equation (12).

$$H_b^a(x,y) = \frac{1}{1-a} In \sum_{m=0}^{x} \sum_{n=0}^{y} \left(\frac{p(m,n)^a}{p_2(x,y)^a}\right) \tag{11}$$

$$H_o^a(x,y) = \frac{1}{1-a} In \sum_{m=m+1}^{255} \sum_{n=n+1}^{255} \left(\frac{p(m,n)^a}{1-p_2(x,y)^a}\right) \tag{12}$$

p_2 represents the posteriori class probabilities.

$p(m,n)$ gives the joint probability mass function where m, n = 0, 1, 2 ….255.

The original, preprocessed and the segmented images using the proposed technique is shown in Fig. 4.

Steps for generating optimal seed and thresholds using dragonfly optimization

i. Input Digital Mammogram
ii. Initialize the dragonflies population, step vectors, inertia weight, separation weight, alignment weight, cohesion weight, food weight, and enemy weight, iteration counter c = 0. Initialise the food and the enemies.

iii. Generate the possible dragonflies which is the seed and threshold values for region growing.
iv. Increment the iteration counter c = c+1 until maximum.
v. Calculate the fitness function value using the equation (10) and update the food and enemies.
vi. Obtain the seed and threshold value, Separation, Alignment, Cohesion, Food, Enemies, and neighboring radius of the current dragonfly.
vii. If the current dragonfly has at least one neighboring dragonfly update the velocity and position vectors of the current dragonfly.
viii. Otherwise, update the current dragonfly using equation (7).
ix. Continue till c ≤ Cmax
x. Save the FS value which is the optimal threshold and seed value for region growing.

3.4. Feature extraction

Image features reveal the existing attributions and characteristics of an image. Hence the features taken for classification should be recognizable, efficient and independent [17]. The proposed work follows a second order statistics of gray level pixel values which estimate the occurrences of pairs of pixel values at specific locations in the mammograms. The Gray level co-occurrence matrixes are constructed at different angles such as 0,45,90 and 135 at unit distance. The proposed system uses a set of texture features based on Haralick definitions of texture analysis. The proposed system identifies ten texture features such as Contrast, Correlation, Cluster Prominence, Dissimilarity, Energy, Entropy, Homogeneity, Maximum probability, Sum of squares, Information measure of correlation. This leads to a set of forty texture features. Five set of features based on texture are also extracted using Gray level Run length matrix (GLRLM) which gives the number of runs of each gray level pixel (I,j) in a horizontal direction. These include short Runs Emphasis (SRE), Long Runs Emphasis (LRE), Gray Level Non-Uniformity (GLNU), Run Length Non-Uniformity (RLNU), and Run Percentage (RPERC). Therefore a set of forty five features are extracted for training the classifier. These features are fed into a neural network classifier for classification as benign and malignant masses.

3.5. Classification

Medical diagnosis can be done using pattern classification dilemma which uses a set or subset of input features to identify a person has a disorder [18]. Artificial Neural Networks an information processing paradigm used in proposed system is a Feed Forward Back propagation Neural Network in which the neural network is trained using a training set of 200 ROIs through Levenberg–Marquardt algorithm in which weights are adjusted such that the error between the desired output and actual output is reduced. The training cycles continue such that the Mean Square Deviation is achieved. The texture features extracted in the feature extraction stage is fed into the classifier where the neurons are placed in the input layer which corresponds to a set of texture features extracted in the feature extraction. The Fig. 5 below shows the basic architecture of an artificial neural network with back propagation learning. A two-class classification called benign and malignant classes is performed where only one neuron is placed in the output layer whose output is 0 or 1 to represent the benign and malignant class respectively.

The steps in Back propagation algorithm for classification of breast masses is given as:

i. Initialize the weights randomly.
ii. When the error is large
iii. For each training pattern, apply the input texture features to the ANN.
iv. Calculate the output of the neurons in the ANN using the activation function based on the summation of the input weights and the bias values. For any training pattern at nth neuron, the activation functions is as follows using equation (13).

$$A_n = \sum W_{m,n} y_m + b_n \tag{13}$$

Where A_n is the activation function of the nth neuron, $W_{m,n}$ is the weight between the mth and nth neuron Y_m is the output of the mth neuron and b_n is the bias of the nth neuron.

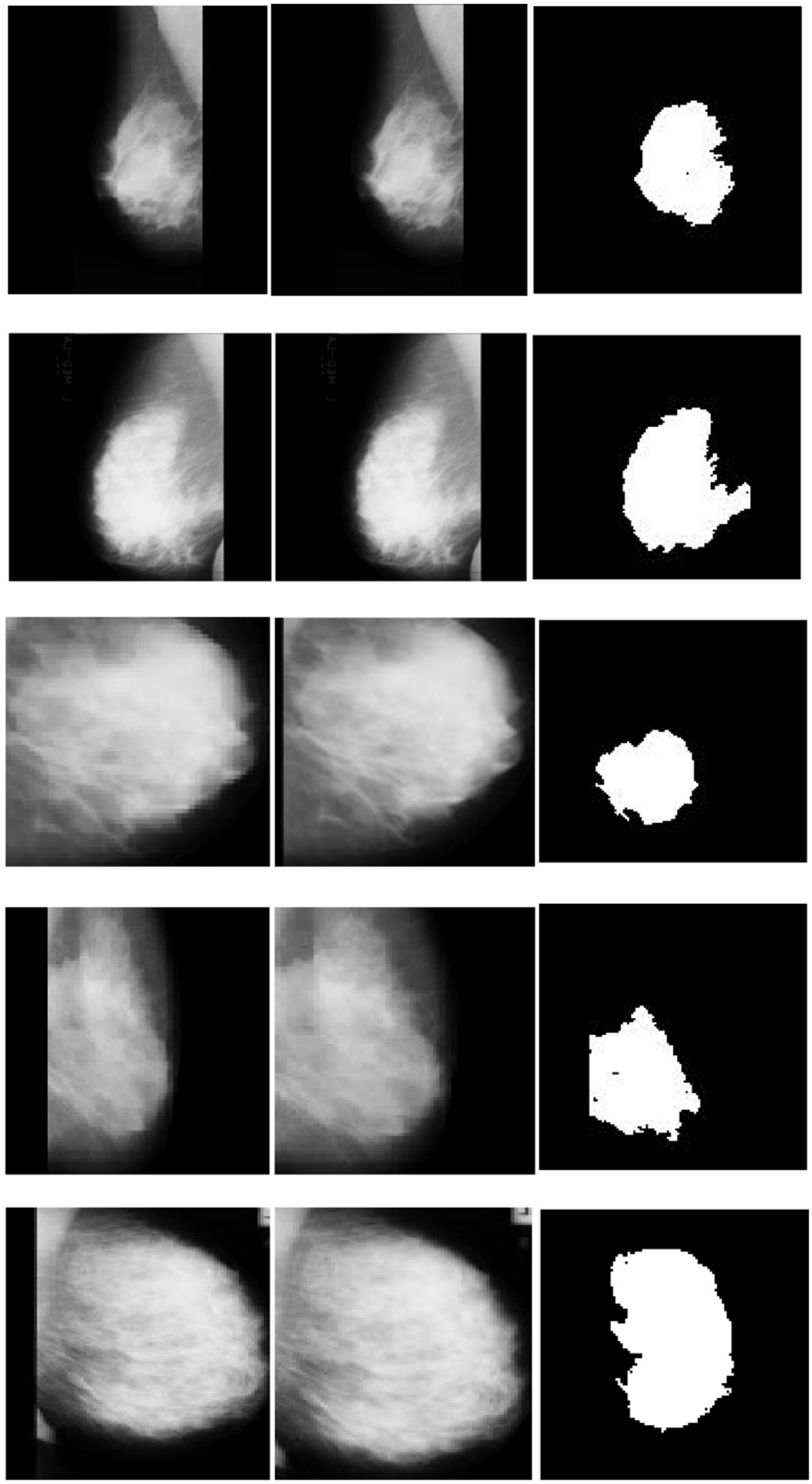
v) Calculate the error at the output neuron using the Mean Squared Deviation (MSD) using equation (14).

$$MSD = \frac{1}{np} \sum_{i=1}^{np} (T_i - O_i)^2 \tag{14}$$

where T is the desired output, O is the actual output and np is the number of predictions.

vi) Compute the error signals based on the output error and compute the weights of the connections for the pre-output layers.
vii) Adjust the weights till error becomes too small.
viii) Apply the testing data.

The Neural Network architecture uses the 300 images of the left and right breasts from MIAS database for the testing and training purposes. Table 1 shows the architecture of the ANN used in the proposed system. The type of the ANN used in the proposed work is the Multilayer Perceptron (MLP). The number of input neurons represents the texture features extracted in the previous stage of feature extraction where 25 neurons are used. The output layer consists of only one neuron which represents 0 or 1 for benign and malignant breast masses respectively. The learning algorithm adapted is the back propagation mechanism where the weights are adjusted based on the mean squared deviation calculation. The activation function



(a) (b) (c)

Fig. 4. (a) Original Images, (b) Preprocessed Images, (c) Segmented Images using proposed method.

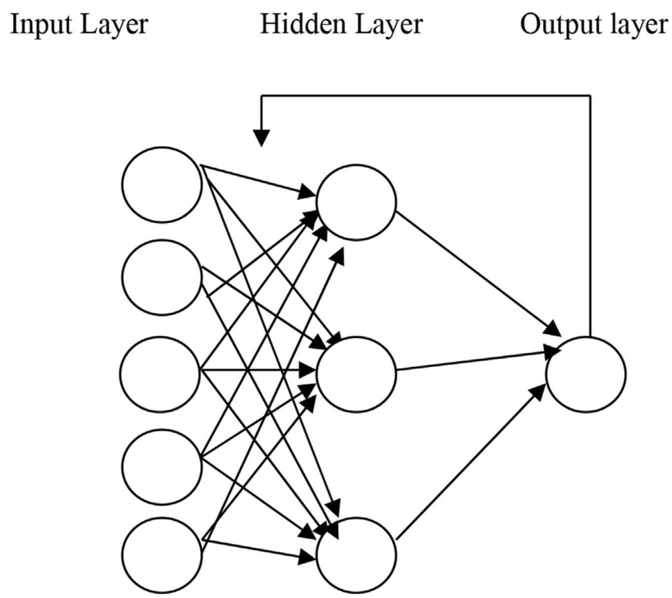


Fig. 5. Architecture of Feed Forward Neural Network with Back propagation.

used for calculating the outputs of the neurons is the summation of the input weights and bias of the neurons.

4. Experimental results

The experimental results are obtained by implementing the proposed algorithm in MATLAB (Matlab version R2013a) on Intel Pentium CORE i5 processor with 4 GB RAM and Windows 7 Operating System. The experimental results are given in the following sub sections which includes the segmentation and classification performance of the proposed system.

4.1. Segmentation performance

The segmentation accuracy of the proposed Dragon fly optimized region growing method can be obtained using the degree of similarity between the ground truth images and the output images produced by the proposed segmentation method. The degree of similarity between two images can be measured using Jaccard index [25] by equation (15).

$$J(G,O) = \frac{|G \cap O|}{G \cup O} \tag{15}$$

Table 1
Architecture of the ANN used in the proposed system.

Database	DDSM
ANN	Multilayer Perceptron
Number of input neurons	45
Number of Hidden Neurons	45
Number of output Neurons	1
Learning rule	Backpropagation
Error Calculated	Mean Squared Deviation
Problem	Classification
Activation function	Summation

where J (G, O) is the Jaccard index of the ground truth image and the corresponding output image. The segmentation accuracy of the proposed method attained using Jaccard index is 90%. The ground truth images and the respective output images generated using the proposed technique are given in the Fig. 6.

4.2. Classification performance

The performance of the proposed system presented in the paper is measured using three metrics called Sensitivity, Specificity and Accuracy which are measured using the equations (16)–(18). To evaluate the performance of the classification while using Dragon Fly optimized region growing in the segmentation stage 300 images are used where 154 benign cases, 146 malignant cases of both left and right breasts. About 200 images of both benign and malignant cases are used for training and the remaining images are used for testing the proposed system. The classification accuracy can be measured [26].

Sensitivity: Sensitivity is a measure which determines the probability of the results that are true positive as ‘that person has the tumor.

$$Sensitivity = \frac{TP}{TP + FN} \tag{16}$$

Specificity: Specificity is a measure which determines the probability of the results that are true negative as ‘that person does not have the tumor.

$$Sensitivity = \frac{TN}{TN + TP} \tag{17}$$

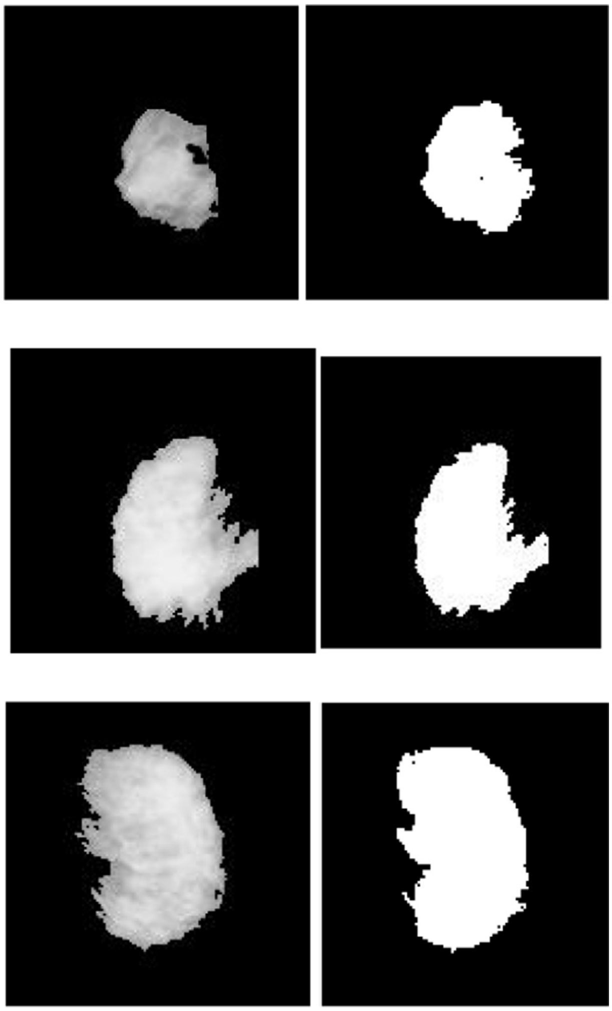
Accuracy: Accuracy is a measure which determines the probability that how many samples are accurately classified.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{18}$$

In the classification process, these evaluations are expressed in terms of various parameters [26] the True Positive (TP), True Negative (TN), False Positive (FP) and the False Negative (FN).

TP: tumor image correctly classified as tumor image
TN: normal image correctly unclassified as tumor image
FP: normal image wrongly classified as tumor image
FN: tumor image wrongly unclassified as tumor image

Mean Squared Deviation (MSD) is used to measure the error of the MLP used in this work according to the equation (14). The classification rate of the proposed system is compared with the classification rate of the same classifier while using traditional region growing, modified region growing, Ant Bee Colony Optimized region growing and Particle Swarm Optimized region growing techniques using Receiver Operating Curve (ROC) which is a standard for measuring the performance of a classification system. A ROC curve is generated in the two-dimensional area where false



(a) (b)

Fig. 6. (a) Ground truth Images, (b) Output Images.

positive rate (1-specificity) is plotted on the x-axis and True Positive rate (Sensitivity) is plotted in the y-axis. A reduced false positive rate shows that the system has the high capacity of diagnosing the breast masses through which the death rate can be decreased. The proposed image segmentation method based on the dragonfly optimization increases the classification performance, where sensitivity reached up to 98.1% which shows that the system can detect the malignant breast masses accurately. The specificity is reached up to 97.8% which shows that the proposed system can detect the benign breast masses accurately. The overall accuracy of the classification system is reached up to 98% making it a more powerful detection and classification system for breast masses. For the comparison of performance of the proposed Dragon Fly Optimized Region Growing (DA-RG)technique is done with other optimized region growing techniques using the same fitness functions based on Reniy'sentropy and the Particle Swarm Optimized Region Growing (PSO-RG). Ant Bee Colony based Region Growing (ABC-RG) is implemented and

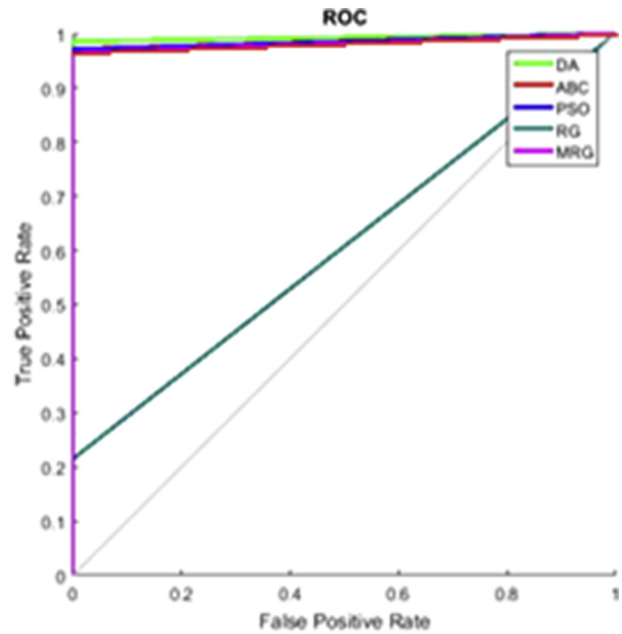


Fig. 7. ROC analysis of the ANN classifier when trained and tested using the ROIs generated by Optimized Region Growing with (DA, ABC, PSO), MRG and RG.

tested using the same training and testing images. The accuracy gained by the PSO-RG and ABC-RG reached up to 96% and 94% respectively. The proposed system is also compared with the traditional region growing, modified region growing (MRG) which uses the orientation constraints along with the intensity values of the pixels as shown in Fig. 7. The traditional region growing techniques get the lowest accuracy when it is tested using the same image set a far as breast mass detection is concerned.

The DDSM database that contains 154 benign cases, 146 malignant cases are divided into two subsets for training and testing purposes in which 200 images are used for training and the 100 images are used as testing samples. The Table 2 shows the performance measures for two mammogram subsets in terms of True Positive Rate(Sensitivity), False Positive Rate (1-Specificity) of the ANN classifier used in the proposed system where the ROIs generated by Optimized Region Growing algorithm using DA, PSO, ABC are used for training and testing purposes. Table 3 shows the comparison of the existing technique with the proposed system.

Table 2
Classification performance measures for ANN classifier in terms of True Positive Rate (TPR), False Positive Rate (FPR).

Segmentation	Image Subset	True Positive Rate (TPR) Sensitivity	False Positive Rate (FPR) 1-Specificity
		Test	Test
DA-RG	1	98.1%	96.8%
	2	97.6%	96.7%
PSO-RG	1	96.2%	94.8%
	2	94.6%	92.3%
ABC-RG	1	94.2%	92.2%
	2	91.6%	89.8%

Table 3
Comparison of the existing technique with the proposed system.

Authors	Preprocessing	Segmentation	Features Extracted	Classification	Accuracy	Dataset
Rahimeh et al. [6]	Median Filtering	Region Growing and CNN optimized using GA	Intensity, shape, and texture features	Random forest, Naïve Bayes, SVM, and KNN	Accuracy 96.47%	MIAS and DDSM
Celia et al. [19]	Iris filter	Adaptive Threshold method	Gray level, texture, contour-related, and morphological features	Backpropagation Neural network	Sensitivity of 88% and 94% at 1.02 false positives per image	Images from hospitals of the health district of Santiago de Compostela, Spain
Danilo et al. [13]	Wiener Filter	Hammouche's Computational algorithm using wavelet and Genetic Algorithm	–	–	AOM 79.2 ± 8%	DDSM-Images in CC and MLO Views
Zhili C et al. [20]	–	Manual segmentation	Topological features	KNN classification	Acc 96%	MIAS and DDSM
Qaisar et al. [21]	CLAHE, Gaussian Filtering	Multiscale feature Fusion and Maximum a posteriori method	–	–	AOM 90%	Mini MIAS and DDSM
Arden et al. [22]	–	Cropping	Laws Texture features using GLCM	Feed-Forward Neural Network	Acc 93.90%	MIAS
Kanchan et al. [23]	Median Filtering	Fuzzy-C-means and thresholding technique	Tamura, shape features	Support Vector Machine using Radial Basis Function (RBF) kernel	Acc 96.92%	Mini-MIAS
Bhagwati et al. [24]	Homomorphic filter	Region Growing	Shape and texture feature features	Multilayer Perceptron Neural Network	Acc 95.6%	DDSM
Proposed System	Manual Cropping and Gaussian Filtering	Dragonfly Optimized Region Growing	Texture features using GLCM and GLRLM	Feed Forward Neural Network using Back propagation learning algorithm	Sensitivity 98.1% Specificity 97.8%	DDSM Images in CC and MLO views

5. Conclusion

The accurate identification of breast masses is an important step in finding out the breast cancer in its early stage. The proposed work can be used as a guide for the radiologists when there is a high demand for processing mammograms and it can also be used as a second opinion in critical cases. The proposed work first attempts to preprocess the images using Gaussian filtering and then extracts the benign and malignant ROIs containing the breast masses using Optimized Region Growing based on Dragon fly optimization. The texture features are extracted using GLCM and GLRLM after which they are fed to Feed-forward Neural network where the learning is done using Levenberg−Marquardt back propagation algorithm and the errors are calculated using Mean Square Deviation. The final output of the classifier is the categorization of the ROIs as benign or malignant. The proposed system uses the images collected from the DDSM database of the both left and the right breasts in MLO and CC views. The segmentation performance of the proposed method is 90% when measured using Jaccard index. The proposed breast mass detection techniques which are based on the optimized region growing prove to provide a nearly and accurate diagnosis of the benign and malignant masses. Its accurate detection of the proposed optimized region growing technique has increased the accuracy of the Feed-forward Neural Network classifier as 98% which is used at the final stage of the proposed system which shows that it is capable of finding the malignant and benign breast masses accurately. Eventhough the proposed method has shown good results in terms of overall accuracy, more emphasis can be given to the feature selection methods in future. Different classification mechanisms with appropriate learning algorithms and tuning of the neural parameters using optimization algorithms can also be done in future. The optimized threshold generated in the region growing technique can also be generated using other evolutionary algorithms in future.

References

[1] Verma Brijesh, Zhang Ping. A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms. Appl Soft Comput 2007;7:612−25.

[2] Dheeba J, Tamil Selvi S. A Swarm optimized neural network system for classification of microcalcification in mammograms. J Med Syst 2012; 36:3051−61.

[3] Cheng HD, Sh XJ, Min R, Hu LM, Cai XP, Du HN. Approaches for automated detection and classification of masses in mammograms. Pattern Recogn 2006;39:646−68.

[4] Pawar Meenakshi M, Talbar Sanjay N. Genetic Fuzzy System (GFS) based wavelet co-occurrence feature selection in mammogram classification for breast cancer diagnosis. Perspect Sci 2016:247−50.

[5] Qian Wei, Fei Mao, Sun Xuejun, Zhang Yan, Song Dansheg, Clarke Robert A. An improved method of region grouping for microcalcification detection in digital mammograms. Comput Med Imag Graph 2002;26:361−8.

[6] Rouhi Rahimeh, Mehdi Jafari, Kasaei Shohreh, Keshavarzian Peiman. Benign and malignant breast tumors classification based on region growing and CNN segmentation. Expert Syst Appl 2015;42: 990−1002.

[7] Kandan Sambandam Rakoth, Jayaraman Sasikala. Self-adaptive dragonfly based optimal thresholding for multilevel segmentation of digital images. J King Saud Univ Comput Inf Sci 2016. https://doi.org/10.1016/j.jksuci.2016.11.002.1319-1578.

[8] Guillaume Kom, Tiedeu Alain, Martin Kom. Automated detection of masses in mammograms by local adaptive thresholding. Comput Biol Med 2007;37:37−48.

[9] Vadivel A, Surendiran B. A fuzzy rule-based approach for characterization of mammogram masses into BI-RADS shape categories. Comput Biol Med 2013;43:259−67.

[10] XiaomingLiu, ZhigangZeng. A new automatic mass detection method for breast cancer with false positive reduction. Neurocomputing 2015;152: 388−402.

[11] Kumar Mohanty Aswini, Ranjan Senapati Manas, Swapnasikta Beberta, Kumar Lenka Saroj. Texture-based features for classification of mammograms using a decision tree. Neural Comput Appl 2013;23:1011−7.

[12] Xie Weiying, Li Yunsong, Yide Ma. Breast mass classification in digital mammography based on extreme learning machine. Neurocomputing 2015;173:930−41.

[13] Pereira Danilo Cesar, Ramos Rodrigo Pereira, do Nascimento MarceloZanchetta. Segmentation and detection of breast cancer in mammograms combining wavelet analysis and genetic algorithm. Comput Methods Progr Biomed 2014;114:88−101.

[14] Magna Gabriele, Casti Paola, Jayaraman SowmyaVelappa, Salmeri Marcello, Mencattini Arianna, Martinelli Eugenio, et al. Identification of mammography anomalies for breast cancer detection by an ensemble of classification models based on the artificial immune system. Knowl Base Syst 2016;101:60−70.

[15] Verma Brijesh, McLeod Peter, Klevansky Alan. A novel soft cluster neural network for the classification of suspicious areas in digital mammograms. Pattern Recogn 2009;42:1845−52.

[16] Mirjalili Seyedali. Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput Appl 2016;27:1053−73.

[17] Renyi Alfred. On measures of entropy and information. In: Proceedings of the fourth berkeley symposium on mathematics ,statistics and probability; 1960. p. 547−61.

[18] Wang Zhiqiong, GeYu, Yan Kang, Zhao Yingjie, Qu Qixun. Breast tumor detection in digital mammography based on extreme learning machine. Neurocomputing 2014;128:175−84.

[19] Varela Celia, Tahoce Pablo G, Méndez Arturo J, Souto M, Vidal JJ. Computerized detection of breast masses in digitized mammograms. Comput Biol Med 2007;37:214−26.

[20] Chen Zhili, Harry Strange, Oliver Arnau, Erika R E Denton, Boggis Caroline, ReyerZwiggelaar. Topological modeling and classification of mammographic microcalcification clusters. IEEE Trans Biomed Eng 2015;62(4):1203−14.

[21] Abbasa Qaisar, EmreCelebic M, Garce Irene Fondon. Breast mass segmentation using region-based and edge-based methods in a 4-stage multiscale system. Biomed Signal Process Control 2013;8:204−14.

[22] Setiawan Arden Sagiterry, Wesley Julian, Yudy Purnama. Mammogram classification using law's texture energy measure and neural networks. Proc Comput Sci 2015;59:92−7.

[23] Kashyap Kanchan Lata, Bajpai Manish Kumar, Khanna Pritee. Breast cancer detection in digital mammograms. In: IEEE international conference on imaging systems and techniques (IST); 2015. p. 16−8.

[24] Bhagwati Charanpate l, Sinha GR. Mammography feature analysis and mass detection in breast cancer images. In: International conference on signal processing and computing technologies (ICESC); 2014. p. 474−8.

[25] Cheetham AH, Hazel JE. Binary(presence−absence)similarity coefficients. J Paleontol 1969;43(5):1130−6.

[26] Metz CE. Basic principles of ROC analysis. Semin Nucl Med 1978;8(4): 283−98.