

Mass classification in breast DCE-MR images using an artificial neural network trained via a bee colony optimization algorithm

D. Janaki Sathya^{a,*}, K. Geetha^b

^a EEE Department, Karpagam University, Coimbatore, Tamil Nadu, India

^b EEE Department, Karpagam Institute of Technology, Coimbatore, Tamil Nadu, India

*Corresponding author, e-mail: janu_sathya@rediffmail.com

Received 30 Mar 2012

Accepted 28 Mar 2013

ABSTRACT: Breast cancer is becoming the leading cause of cancer deaths among women. The best way to reduce deaths due to breast cancer is early detection and treatment. Dynamic contrast enhanced (DCE) MRI has emerged as a promising new imaging modality for breast cancer screening. Currently, radiologists evaluate breast lesions based on a qualitative description of lesion morphology and contrast uptake profiles. The qualitative description of breast lesions from DCE-MRI introduces a high degree of inter-observer variability. In addition, the high sensitivity of MRI results in good specificity. A computer-assisted evaluation system that can automatically analyse lesion features to differentiate between malignant and benign lesions would be very useful. One of the major characteristics for mass classification is texture. Artificial neural networks exploit this important factor to classify the mass as benign or malignant. The selected texture features were used to classify the mass with a three-layered neural network to predict the outcome of a biopsy. The main objective of this proposed method is to increase the effectiveness, robustness, and efficiency of the classification process in an objective manner to reduce the numbers of false-positive results. The paper presents an intelligent computer assisted mass classification method for breast DCE-MR images. It uses the artificial bee colony algorithm to optimize the a neural network performing benign-malignant classification on the region of interest. A three-layer neural network with seven features was used for classifying the region of interest as benign or malignant. The network was trained and tested using the artificial bee colony algorithm and was found to yield a good diagnostic accuracy.

KEYWORDS: artificial intelligence, dynamic contrast enhanced magnetic resonance images (DCE-MRI), back propagation algorithm, statistical texture features

INTRODUCTION

It is difficult to achieve accurate classifications of diseases, especially on cancers, which are very important in medical science. Accurate classifications allow doctors to select suitable therapies and treatment for diseases. Over the last several decades, cancer classification has advanced, but has limitations caused by traditional method for morphological appearance analysis¹.

The dynamic contrast-enhanced MR imaging is an emerging powerful tool to detect, diagnose, and monitor breast abnormalities. This technique enables radiologists to localize suspicious tissue areas and to further assess their likelihood of malignancy by means of their morphological and kinetic characteristics^{2,3}. The contrast agent which is injected during the MRI scanning typically leads to rapid signal increase and early wash-out for malignant masses and increases for highly vascular tumours, but gradual signal increase

and wash out for benign masses⁴. The ultimate diagnosis of all types of breast disease depends on a biopsy. In most of the cases the decision for a biopsy is based on the mammography findings. Biopsy results indicate that 65–90% of suspected cancers which are detected by mammography turned out to be benign⁵. Masses are one of the important signs of early breast cancer. They are often indistinguishable from the surrounding parenchyma because their features can be obscured or be similar to that of normal inhomogeneous breast tissues. The grey levels of those inhomogeneous tissues in the image could vary with the distribution of breast soft tissues. This makes the automatic mass detection, segmentation, and classification challenging. Hence it would be valuable to develop a computer assisted method for mass classification based on extracted features from the region of interests (ROI) in DCE-MRI. This would reduce the number of unnecessary biopsies in patients with benign disease and thus avoid patient's

physical and mental suffering. DCE-MRI has been one of the most reliable methods for early detection of breast carcinomas. However, retrospective studies have shown that radiologists can miss the detection of a significant proportion of abnormalities having high rates of false positives. The estimated sensitivity of radiologists in breast cancer screening is only 75%.

Computer-assisted evaluation (CAE) systems are useful for supporting radiologists in detecting and assessing suspicious tissue regions in DCE-MRI data. These data are expected to improve the reliability of clinical decisions and potentially decrease the number of unnecessary biopsies. The CAE systems are designed by employing computer algorithms for automated lesion segmentation⁶⁻¹⁴, trailed with proper edge enhancement technique such as unsharp filter as illustrated in Ref. 15, which is used to extract the edges of the tumour very efficiently in segmented MR images, followed by proper threshold for extraction the lesion or region of interest (ROI) from edge enhanced segmented breast MR images. After detection of the ROI, statistical analysis techniques are applied to select an optimal set of features to achieve the highest diagnostic accuracy¹⁶⁻¹⁸. The extracted features are fed as input to the classifier to discriminate whether the lesion is benign or malignant. It is obvious that the performance of CAE systems depends on feature selection, training database and the classifier.

This paper presents an automatic mass classification system, which is developed to classify mass into benign and malignant based on the statistical textural features extracted from mass of the breast region using artificial neural network. It is trained using artificial bee colony optimization algorithm. The selected texture features were used to classify the mass with a three-layered neural network to predict the outcome of biopsy. Three layered artificial neural networks with seven features were proposed for classifying the ROI region into benign and malignant. This work is the first experiment of artificial bee colony algorithm on classifying the ROI of breast DCE-MR imaging.

REVIEW OF THE TECHNIQUES

This section deals with the main techniques involved in mass classification CAE system, which enhances the CAE system performance such as accuracy, specificity, sensitivity and efficiency.

Artificial neural networks

Artificial neural networks (ANNs) are the most novel and powerful artificial tool suitable for solving combinatorial problems such as prediction and classification. Neural networks are attracting more and more

interest for its abilities of parallel operation, self-learning, fault tolerance, associative memory, multifactorial optimization, and extensibility. An ANN is a parallel distributed processor that has a natural tendency to store experiential knowledge. ANNs are being used extensively to solve universal problems intelligently like continuous, discrete, and clustering. ANNs are being applied for different optimization and mathematical problems such as classification, object and image recognition, signal processing, seismic event prediction, temperature, weather forecasting. They can provide suitable solutions for problems, which are generally characterized by nonlinear ties, high dimensionality noise, complex, imprecise and imperfect or error prone sensor data, and lack of a clearly stated mathematical solution or algorithm. A key benefit of neural networks is that a model of the system can be built from the available data¹⁹.

ANNs are invaluable tools in various medical diagnostic systems. The key attributes like distributed representations, local operations, and nonlinear processing make ANN is appropriate for taking few difficult decisions from massive amount of data. Thus when expert knowledge is unavailable in full-fledged sense as for example in case of masses, ANN provides alternative and better solutions. In addition, ANN constructively makes use of trained data set to make complex decisions. It is robust without requiring a rule or explicit expression, and widely applicable²⁰. The ANN has been applied in the diagnosis of breast cancer using subjective impression of different features based on defined criteria²¹. The main objective of this work is to develop a proficient classifier for breast cancer. Detection of masses in breast DCE-MR images using ANN is trained using artificial bee colony optimization algorithm.

Artificial bee colony algorithm

Many complex multi-variable optimization problems cannot be solved exactly within polynomial bounded computation times. The honey bee algorithm is a search algorithm which is capable of locating good solutions efficiently. The algorithm is inspired by the food foraging behaviour of honey bees and could be regarded as belonging to the category of intelligent optimization tools²². In the artificial bee colony (ABC) algorithm, the colony of artificial bees contains three groups of bees: employed bees, onlookers, and scouts. Each cycle of the search consists of three major steps: (1) placing the employed bees onto the food sources and then calculating their nectar amounts; (2) selecting the food sources by the onlookers after sharing the information of employed bees and determining the

nectar amount of the food sources; (3) determining the scout bees and placing them onto the randomly determined food sources. In the ABC, a food source position represents a possible solution to the problem to be optimized and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution^{23–25}.

The ABC algorithm performs a neighbourhood search combined with random search in a way that is indicative of the food foraging behaviour of swarms of honey bees. The algorithm has been successfully applied to different optimization problems including the training of neural networks for control chart pattern recognition²⁶, wood defect identification²⁷. The performance of ABC algorithm is tested on benchmark classification problems of cancer, diabetes, and heart obtained and the classification task is done by neural networks and clustering²⁸, clustering approaches inspired by the collective types of behaviour of bees have been proposed by²⁹, which used the ABC algorithm for clustering of different data sets. Honey bees foraging behaviour model applied to the job scheduling problem³⁰. The clustering approach based on ABC algorithm is used in Ref. 14 to segment breast MR images to extract mass lesion. The results show that the performance of the algorithm is more robust than neural network based segmentation algorithm. The bee's algorithm is used for tuning the parameters of a fuzzy logic controller³¹. Simulation results showed that the use of the algorithm to optimize the membership functions of the fuzzy logic system enhances the controller performance.

Recent studies suggest that ABC algorithm can be used for several real world data clustering and mining problems and it has proved to give a more robust performance than other intelligent optimization methods for a range of complex problems²².

PROPOSED MASS CLASSIFICATION MODEL

The framework of the present work is the development of an efficient and innovative computer-assisted evaluation (CAE) system that can automatically analyse lesion features on region of interest (ROI) that contains mass on breast DCE-MR imaging to differentiate between malignant and benign. The proposed model includes the following phases in its CAE system: feature extraction, feature selection, and mass classification. Such CAE tools present a qualitative and a quantitative description of the disease to the physician.

Feature extraction

Once a lesion is detected in the ROI, characterization is necessary to estimate the pathological nature

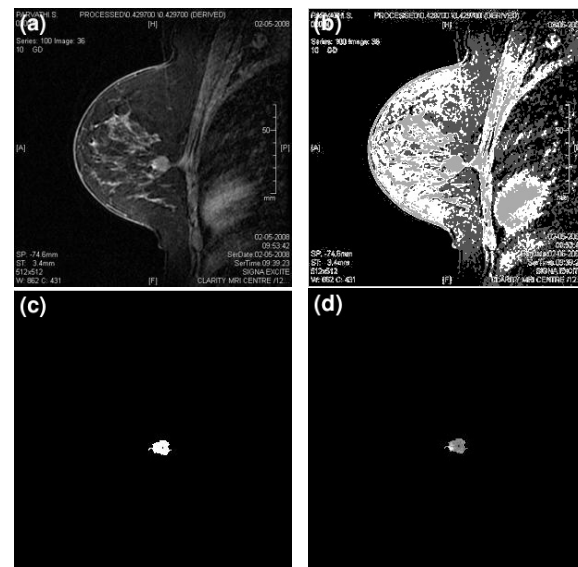


Fig. 1 A representative real breast image with tumour obtained from 1.5 Telsa (Siemens) MRI scanner, the result of segmentation procedure, extraction of the mass separately, and selection of ROI: (a) input image, (b) segmented image, (c) mass extraction, (d) ROI selection.

of the lesion, i.e., whether the lesion is benign or malignant. For the quantitative assessment, many features have been extracted from the masses or ROI of DCE-MR images, detected in segmentation phase of the computer-assisted evaluation system; a detailed procedure of segmentation phase has been reported in Ref. 14. Fig. 1 shows an example of a real image containing a mass lesion, and the results of the segmentation, mass or lesion extraction and ROI selection procedure.

A typical breast DCE-MRI contains a great amount of heterogeneous information that depicts different tissues, vessels, ducts, chest skin, and breast edge characteristics. Texture features have been widely used in breast DCE-MRI mass classification^{7, 18, 32–37}.

The implemented feature extraction procedure relies on the exploration of the textural characteristics of the extracted mass. Textures represent tonal variations in the spatial domain and determine the overall visual smoothness or coarseness of image features. They reveal important information about the structural arrangements of the objects in the image and their relationship to the environment. Consequently, texture analysis provides important discriminatory characteristics related to the variable patterns of digital classifications. Texture analysis³² allows the

classification of the segmented regions. These regions can be used to form a diagnosis. One of the simplest approaches to describe texture is to use moments of the grey-level histogram of an image or region³⁸. The extraction methods of texture feature plays very important role in detecting abnormalities present in breast DCE-MRI. The diagnostic features in texture characteristics were selected to identify those yielding the highest discrimination thus achieving the optimal diagnostic performance. Texture analysis approaches can be summarized into four texture modelling methods: statistical methods, geometrical methods, model based methods, and signal processing methods³⁹. The statistical textures are found to be the best for image classification³³.

A statistical approach sees an image texture as a quantitative measure of the arrangement of intensities in a region. The popular statistical feature extraction method is grey level co-occurrence matrix (GLCM). GLCM features outperformed well in discriminating between malignant and benign^{7,32,36}. In general, this approach is easier to compute and is widely used. The grey level co-occurrence matrix (GLCM), provided by Ref. 40 calculated second-order grey scale statistics using grey-level co-occurrence matrices (GLCM) and defined fourteen statistical measures for texture. Co-occurrence matrices give information about patterning of the texture, and it is possible to calculate textural properties from them. The grey level histogram moments (GLHM) has been proved to be effective for classification of images⁴¹; these features are sensitive to illumination variations⁴²; both of these statistical texture features have been very popular in different texture analysis applications.

In this proposed work, a set of 18 features were extracted from the ROI, fourteen statistical measures for texture (GLCM) as defined in Ref. 40 and four GLHM features. The above described features were extracted using a window size of 16×16 pixels without overlapping. The features extracted are: energy measure, correlation, inertia, entropy, difference moment, inverse difference moment, sum average, sum entropy, difference entropy, sum variance, difference variance, difference average, information measure of correlation, standard deviation, mean, variance, skewness, and kurtosis.

Feature selection

Feature selection has been widely used to improve prediction accuracy of classifiers. The improvement in prediction is related to the redundant features or noisy features in the data which can be eliminated by feature selection. Feature selection is defined as a series of

actions to choose a subset of features that are relevant to correct classification based on specified evaluation and selection criteria⁴³⁻⁴⁵. There are several reasons why feature selection can be an important step in designing CAE systems. A classifier trained on a data set in which there are many more features than there are cases (patients) is less likely to perform adequately on new but supposedly similar cases; this phenomenon is referred to as “overtraining”. Thus reducing the number of features can enable the system designer to build a more robust system and more accurately assess the performance of algorithms under consideration. The goal of feature selection is to choose the optimal feature vector, consisting of the features that minimize the classification error. The time it takes to train a classifier is typically dependent on the number of features, so decreasing the number of features can increase the efficiency of CAE system development. Feature selection helps to reduce the feature space which improves the prediction accuracy and minimizes the computation time.

Statistical methods used to test the null hypothesis are commonly called tests of statistical significance. Hypothesis tests appear to be well suited for dealing with detection problems. In particular they allow for a simple way to estimate and control the percentage of false negatives by appropriate tuning of the confidence level. The statistical hypothesis *t*-test described in Ref. 46 is used to select the set of effective features for the classifier to obtain high accuracy. Hypothesis testing is a common method of drawing inferences about a population based on statistical evidence from a sample. A statistical hypothesis *t*-test is a method of making decisions using data, according to a pre-determined threshold probability. The hypothesis test is applied to decide whether the feature can discriminate the mass type or not. The *t*-test statistic is calculated using:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}, \quad (1)$$

where the sample mean, $\mu = 0$ is the hypothesized population mean, s is the sample standard deviation, and n is the sample size.

Hypothesis testing works by collecting data and measuring how likely the particular set of data is, assuming the null hypothesis is true. It performs a test of the hypothesis that the data in the vector of data set comes from a distribution with $\mu = 0$, and returns the result of the test. If the result equals zero it indicates that the null hypothesis ($\mu = 0$) cannot be rejected at the 20% significance level. If the result equals 1 it indicates that the null hypothesis can be rejected at

the 20% significance level. The critical value for a hypothesis test is a threshold to which the value of the test probability in a sample is compared to determine whether or not the null hypothesis is rejected. The critical value for any hypothesis test depends on the significance level at which the test is carried out. The significance level can be interpreted as the probability of rejecting the null hypothesis when it is actually true. The significance level of a test is a threshold of probability which is to be agreed before the test is conducted.

By providing the hypothesis *t*-test over all the extracted 18 features, the *t*-test results indicate that only 7 features can classify between the two clusters. The features are energy, entropy, mean, variance, skewness, standard deviation and kurtosis. The feature extraction and selection is a key step in mass detection since the performance of CAE depends more on the optimization of the feature selection than the classification method.

Mass type classification

The selection of the correct classifier is a factor that very sensitively affects the performance of the correct classification. Some efforts have already been made to automatically predict lesion^{1, 8, 10, 16, 18, 32–37, 47}. The general approach of breast CAE system is to predict the suspicious areas of breast DCE-MRI into benign or malignant tissue, applying computer algorithms for tumour characterization. The classification technique used in this paper is based on neural network and the bees optimization algorithm is used to train the neural network, based on the work described in Refs. 25, 26.

The classifier chosen is a multilayer feed forward neural network¹⁹. Its nonparametric statistical property is the main reason for choosing it. Unlike the classical statistical classification methods, such as the Bayes classifier, no knowledge of the underlying probability distribution is needed by a neural network. It can learn the free parameters (weights and biases) through training by using examples. This makes it suitable to deal with real problems which are nonlinear, nonstationary, and nonGaussian.

Backpropagation (BP) algorithm is an accepted learning algorithm used for MLP training¹⁹. The main task of BP algorithm is to update the network weights for minimizing output error using backpropagation processing because the accuracy of any approximation depends on the selection of proper weights for the ANNs. It has high success rate in solving many complex problems but it has some drawbacks especially when setting parameter values like initial values of connection weights, value for learning rate,

and momentum. Momentum allows the net to make reasonably large weight adjustments as long as the corrections are in the same general direction for several patterns. Using smaller learning rate prevents a large response to the error from any training pattern. It is well-known that the training speed depends on the choice of the learning rate. If the learning rate is small, the learning process is stable but at the expense of computation time. If the network topology is not carefully selected, the ANNs algorithm can get trapped in local minima or it might lead to slow convergence or even network failure. In order to overcome the disadvantages of standard BP, artificial bee colony optimization based techniques have been proposed for MLP training. ABC algorithm is an easily understandable technique for training MLP on classification problems²⁵.

Architecture of the artificial neural network

Artificial neural networks have been used elsewhere in clinical data modelling; the proposed network architecture used to obtain optimal classifier consists of a three-layer back-propagation neural network, known as multilayer feed forward artificial neural network. The three-layered topology has an input layer, one hidden layer, and an output layer. The number of nodes in the input corresponds to the number of input variables which is seven in this experiment because seven texture features are selected for classification. The linear activation function is used for input layer. The output layer contains one node with values from 0 or 1 indicating level of malignancy, where 0 means absolutely benign and 1 means absolutely malignant. The number of hidden nodes is usually determined by a number of trial-and-error runs. Different neural network architectures with hidden nodes from 3–20 were tested and finally 5 nodes are selected for the hidden layer. The standard method for refining such a neural network is using an error backpropagation algorithm. During the training phase, the feedforward calculation is combined with backward error propagation to adjust the weights. In this work the bee optimization algorithm is used to optimize the weights of the neural network in place of the backpropagation. The bee algorithm can solve a problem without any information apart from that needed to evaluate fitness. The optimal architecture was the chosen one, for which the validation error was the lowest.

Neural network training using the bee algorithm

Basically, training is a process by which the free parameters (i.e., synaptic weights W and bias levels) of an ANN are adapted through a continuous process of

stimulation by the environment in which the network is embedded. The type of learning is determined by the manner in which the parameter changes take place. On the other hand, the learning process may be classified as supervised or unsupervised. In this paper we focus on supervised learning that assumes the availability of a labelled set of training data made up of p input-output samples.

The training of a back propagation neural network involves the minimization of an error function. The error function defines the total difference between the actual output and the desired output of the network over a set of training patterns. Training proceeds by presenting to the network a pattern of known class taken from the training set. The error component associated with that pattern is the sum of the squared differences between the desired and actual outputs of the network corresponding to the presented pattern. The procedure is repeated for all the patterns in the training set and the error components for all the patterns are summed to yield the value of the error function for a back propagation neural network with a given set of basis function centres, spreads and neuron connection weights.

The optimization using the ABC algorithm will involve the “bees” searching for the optimal values of the weights assigned to the connections between the neurons within the network. Each bee represents a neuron in the network with a particular set of basis function centres, spreads and weight vectors. Training an artificial neural network is an optimization task since it is desired to find the optimal set of weights of a neural network in the training process. The aim of the algorithm is to find the bee producing the smallest value of the error function. The algorithm converges to the maximum or minimum without becoming trapped at local optima. The network error function E is given by:

$$E(w(t)) = \frac{1}{n} \sum_{j=1}^n \sum_{k=1}^k (d_k - O_k), \quad (2)$$

where $E(w(t))$ is the error at the t th iteration; $w(t)$ are the weights in the connections at the t th iteration; d_k is the desired output node; O_k is the actual value of the k th output node; k is the number of output nodes; and n is the number of patterns. O is the optimization target to minimize the objective function by optimizing the network weights $w(t)$.

The MLP artificial neural network training procedure using the bee algorithm thus comprises the following steps:

1: Initialise the population of solutions, X_i ,

where $i = 1, \dots, SN$.

2: Evaluate the fitness of the population according to:

$$fit_i = \begin{cases} 1/(1 + f_i) & \text{if } f \geq 0, \\ 1 + |f_i| & \text{if } f < 0. \end{cases} \quad (3)$$

3: Cycle = 1.

4: Repeat from step 5 to step 8.

5: Apply the training data set to determine the value of the error function associated with each bee. This phase is done by the following process:

(i) Produce new solutions V_{ij} in the neighbourhood of X_{ij} for the employed bees using:

$$V_{ij} = X_{ij} + \Phi_{ij}(X_{ij} - X_{kj}), \quad (4)$$

where k is a solution in the neighbourhood of i , Φ is a random number in the range $[-10, 10]$, and evaluate them.

(ii) Apply the Greedy Selection process between processes and calculate the probability values P_i for the solutions X_i as:

$$P_i = \frac{fit_i}{\sum_{i=1}^{SN} fit_i}. \quad (5)$$

(iii) Produce the new solutions V_i for the onlookers from the solutions X_i , selected depending on P_i , and evaluate their fitness.

(iv) Calculate the error value between the target and obtained value.

6: Based on the error value obtained from step 5, create a new population of bees comprising the best bees in the selected neighbourhoods and randomly placed scout bees. This phase is done by the following process:

(i) Apply the Greedy Selection process for the onlookers between X_i and V_i and determine the abandoned solution (source), if exists, replace it with a new randomly produced solution X_i for the scout bees.

(ii) Memorize the best food source position (solution) achieved so far.

7: Stop if the value of the error function has fallen below a predetermined threshold or after completing a number of iterations.

8: Else, return to step 5.

In the proposed ABC algorithm for training neural network, each cycle of the search consists of three steps after initialization of the colony, foods, and three control parameters in the number of food sources, which are equal to the number of employed bees or onlooker bees (n), the value of limit, the maximum

cycle number (R) for MLP-ABC algorithm. The initialization of weights was compared with output and the best weight cycle was selected by scout bee phase. The bees (employed bees, onlooker bees) would continue searching until the last cycle to find the best weights for networks. The food source of which the nectar was neglected by the bees was replaced with a new food source by the scout bees. Every bee (employed bees, onlooker bees) would produce new solution area for the network and the Greedy Selection would decide the best food source position. Suppose that the neglected source is f_j and $j \in \{1, 2, \dots, D\}$, then the scout bees determined a new food source to be replaced with X_i .

The food area was limited in range $[-10, 10]$. It was applied randomly and was initialized for evaluation. Every bee (employed bees, onlooker bees) would produce new evaluated solution area for the network and the greedy selection was decided for the best food source position. If the new food source has equal or better nectar than the old food source, it was replaced with the new food source in the memory. Otherwise, the old food source was retained in the memory. The basic idea of ABC scheme is to use agents of bees to search the best combination of weights for network. The Greedy Selection was applied between two sets of values X_i and V_i while the best scout bees were randomly selected.

The proposed framework can easily train neural network for lesion classification data for prediction task by finding optimal network weights for MLP.

EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the statistical results of applying intelligent computer assisted mass classification method for classifying the region of interest (ROI) that contains mass on real breast DCE-MR images.

Evaluation data and methods

The breast DCE-MR images are used to classify the mass lesions employing the intelligent classification scheme described above. It was tested over real images received from the Radiology Department of Kovai Medical Centre and Hospital, Coimbatore, Tamil Nadu, India. The dataset contains records of 85 patients. The study included 65 malignant and 20 benign histologically-proven lesions. The data is not publicly available and is taken on special request by promising of not to disclose it. The breast DCE-MR images used in this study was acquired with patients prone to 1.5Tesla Siemens MRI scanner with the use of a dedicated surface breast coil array. The

imaging protocol included bilateral fat suppressed T1-weighted images in the sagittal plane of 1 mm slice thickness and a slab interleaved 3D fat suppressed spoiled gradient echo after the injection of contrast.

Results differed by applying different type of classifiers, due to the fact that each classifier has its own method for the formulation of the normal and cancerous clusters upon which it decides whether a tested ROI is considered cancerous or normal. We have evaluated neural network as a potential mechanism for the design of a classifier responsible for delineating between malignant and benign breast lesions from DCE-MRI data. There were a number of motivations for selecting neural network as a classification mechanism. Neural network have been shown to perform well in medical diagnosis applications^{1, 10, 18, 32, 34-37}.

Classifier performance depends greatly on the characteristics of the data to be classified. There is no single classifier that works best for all given problems. Various empirical tests have been performed to compare classifier performance and to find the characteristics of the data that determine classifier performance. Determining a suitable classifier for a given problem is however more an art than a science. Classifier performance is a function of several factors including the statistical distribution of the training and testing data, the internal structure of the classifier and the inherent randomness in the training process. The classification performance can be assessed in terms of the sensitivity, specificity and accuracy of the system. Sensitivity (SN) is the proportion of actual positives which are correctly identified and it is mathematically defined (6) and specificity (SP) is the proportion of negatives which are correctly identified and is mathematically defined in (7):

$$\text{Sensitivity: } SN = \frac{TP}{TP + FN} \quad (6)$$

$$\text{Specificity: } SP = \frac{TN}{TN + FP} \quad (7)$$

$$\text{Accuracy: } ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (8)$$

$$\text{Error in classification: } E = 1 - ACC, \quad (9)$$

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative. It is obvious that the main objective of a classifier is to minimize the false positive and negative rates, similarly, to maximize the true negative and positive rates.

The sensitivity, specificity, accuracy, and error of

the classification technique were evaluated through quantitative measures derived through the comparison of each classified result with its corresponding ground truth. It is defined in (6)–(9). Ground truth is based on the diagnosis of the histopathologist, who analyses the tissue biopsies.

In order to evaluate the performance of the proposed classifier model using cross-validation experimentation, the breast DCE-MR images are divided into the training set and the test set. The training sets of 50 were used to build the classifier model and the test set of 35 are used to verify the trained classifier model. Note that the cases in the test set are not used to train the classifier model.

The use of texture measures in the characterization of the segmented objects presents some advantages in relation to other approaches. Usually, the texture information is enough to describe the mass completely; the use of texture measures discards most of the objects that represent healthy tissues. Features are tested using a hypothesis test to decide whether or not the feature can discriminate between normal and abnormal tissues using a significance level of 20%. The test indicated that only 7 features can discriminate between the two clusters. These seven features are fed as input to the developed ANN based classifier.

In the proposed mass classification method the training of an ANN is achieved using ABC optimization algorithm. The synaptic weights of an ANN are evolved by means of ABC algorithm. Furthermore, the connections among the neurons that belong to the ANN are synthesized. This allows generating easy design of the ANN with a high performance. We have also proved that this novel technique is a good optimization algorithm, because it does not get easily trapped in local minima. In general, the results were satisfactory. The developed mass classifier model allows automatic search of the best values and generates a good solution for a classification problem.

Performance evaluation

Table 1 illustrates the results obtained using cross-validation experimentation of the proposed artificial bee colony optimization algorithm trained neural network based classifier. The impact of sensitivity and specificity may lead to a more automated, objective, and consistent diagnosis. Results showed that the optimal feature set has been extracted by *t*-test in order to achieve high accuracy, sensitivity and high specificity than other available methods. It indicates that more diagnostic information about the lesion is available. The error in classification is 9% which is obtained from (9). It therefore provides a useful

Table 1 Detailed analysis of the benign versus malignant characterization of the proposed mass classifier algorithm using cross validation experimentation.

No. of cases used for classification	35
True positives	24
True negatives	8
False positives	1
False negatives	2
Sensitivity	92%
Specificity	89%
Accuracy	91%
Error in classification	9%
No. of cases misdiagnosed	3

Table 2 Simulation results of mass classification.

Correct classification (%)		Misclassification (%)	
Benign	Malignant	Benign	Malignant
89	92	11	8

method for reducing the number of negative biopsies. Further investigation will be conducted with a larger set of data to determine the generalizability of these results.

The effectiveness of our approach has been demonstrated in Table 2, with a small set of data. The exhibited result indicates that the texture features employed performs an efficient way for discriminating between benign and malignant mass. We can observe that the proposed approach has a significant better performance in detecting malignant masses than benign masses. The classifier could correctly identify a significant fraction of benign cases and malignant cases, which had been recommended for surgical biopsy under current clinical criteria.

Table 3 presents the parameter values used in the proposed artificial ABC implementation. The values were decided empirically.

Table 3 Parameters used in ABC optimization algorithm.

Parameters	Values
No. of scout bees, <i>n</i>	200
No. of sites selected for neighbourhood search, <i>m</i>	20
No. of best elite sites out of <i>m</i> selected sites, <i>e</i>	3
No. of bees recruited for best <i>e</i> sites, <i>nep</i>	70
No. of bees recruited for the other (<i>m</i> – <i>e</i>) selected sites, <i>nsp</i>	30
Maximum number of cycles, <i>R</i>	2000

Table 4 Sample value of statistical features computed for malignant and benign masses.

Features	Benign	Malignant
Entropy	0.138	0.026
Standard deviation	15.5	6.2
Mean	0.665	0.325
Skewness	7.7	20.9
Kurtosis	64.6	42.1
Variance	330	460
Energy	124 800	98 700

Sample values of statistical texture features computed for malignant and benign masses are shown in Table 4. Each value represents the average of the parameter computed over the entire set of data. These preliminary results revealed that the statistical features of the benign are different from the statistical features of the malignant one. This was another indication that the introduced texture features selected using statistical *t*-test were more effective than the previously used texture features for the classification of mass abnormalities.

An objective method is needed to evaluate the performance of the new proposed mass classification algorithm. The most important performance criterion is accuracy. It is the degree to which an algorithm's classification results match histological-proven radiologist's classification results. The accuracy of this algorithm is presented in Table 1. The sensitivity of the algorithm is high because its accuracy is high. This algorithm provides the same classification of images in all runs which makes it more reliable. The efficiency of the proposed algorithm is good, because (i) the time required for training is 38 min, (ii) 30 s for testing phase, and (iii) less space is required. The proposed algorithm represented the advantages of a fast learning speed and high accuracy.

Leave-one-out cross-validation

Because of the limited data set available for this preliminary study, we have also conducted another experiment based on leave-one-out cross-validation method, which keeps the independence between the training and test set. For n samples, the classifier is trained on $n - 1$ samples and tested on the left-out sample. This is repeated n times, and each time a different sample is left out⁴⁸. The leave-one-out cross-validation method was used on a database of eighty five pathologically-proven breast DCE-MR images of patients (20 benign cases and 65 malignant cases) to evaluate the method. The classification results showed an overall accuracy of 96.5%, sensitivity

Table 5 Qualitative performance comparison of the various currently used classifiers.

Algorithm type	ACC (%)	SN (%)	SP [†] (%)
ANN	80	82	72
Logistic regression	80	91.2	64.5
SVM	81	90	75
Linear discriminant analysis	88	90	79
ANN-ABC (Cross-validation)	91	92	89
ANN-ABC (Leave-one-out-cross-validation)	96.5	96.9	95

[†] ACC = accuracy; SN = sensitivity; SP = specificity.

of 96.9%, specificity of 95%, and error of 3.5%. In this experiment, the performance of the classifier was better than that of the cross-validation method. The major drawback of the leave-one-out method is its computational complexity because it requires many repetitions of training.

The difference in performance between the leave-one-out cross-validation method and the cross-validation experiment originates from the shrinkage; here, the shrinkage refers to the difference between the observed training set. Because the size of testing set in the cross-validation method is almost similar to the size of training set than in the leave-one-out method, the shrinkage in the cross-validation method can be considered to be larger than that in the leave-one-out method. In fact, the proposed classifier performance in the leave-one-out method was better than in the cross-validation method. The experimental results of both validation methods declare that our proposed classifier is actually a beneficial tool for the diagnosis of the breast cancer than several published studies reported in Table 5.

ROC analysis of the classifier performance

The accuracy of a model in making predictions is evaluated regularly using an ROC analysis⁴⁹. An ROC curve is generated by combining the true positive fraction (sensitivity) and false positive fraction ($1 - \text{specificity}$) by setting different thresholds. A quantitative measure of the accuracy of the classification technique is obtained by finding the area under the ROC curve termed A_Z , which varies between 0.0, indicating poor classification performance, and 1.0, indicating high classification performance. Generally, a larger A_Z stands for a better predictive performance.

Moreover, the area under the curve (A_Z) is an indication for the overall performance of the observer, and is typically used to analyse the performance of the algorithms. The estimation of the A_Z value

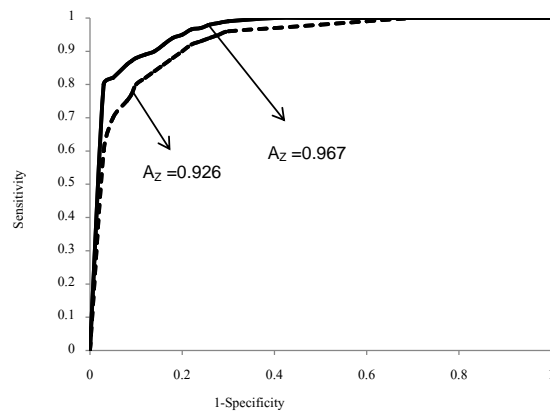


Fig. 2 The ROC curves of classifier by two different evaluation methods, respectively. The dotted line represents the ROC curve from cross-validation method with A_Z of 0.926. The solid line denotes the ROC curve from leave-one-out cross-validation method with A_Z of 0.967.

can be obtained with the trapezoidal rule which can underestimate areas under the curve.

The ROC curve plotted by varying the threshold values of classifier by using two different validation methods is shown in Fig. 2, and the values of A_Z were also calculated to evaluate the ability of the classifier using the cross validation and leave-one-out cross-validation. The results of the experiments suggest that the performance of the leave-one-out cross-validation method with $A_Z = 0.967$ is better than the performance of cross-validation method with $A_Z = 0.926$. Both the A_Z values are high and thus illustrates that the classification performance is optimum.

Comparing performance of different classifiers

The availability of several high-quality predictive models should encourage the clinician to adopt these tools into everyday clinical practice. Arguments favouring such behaviour include standardization of care and of decision-making. Classification performance has been compared in order to determine optimal statistical algorithms for discriminating between disease stages. It is important to notice that an objective comparison of the performance of different CAE methods is very difficult and even impossible due to the use of different databases. The obtained optimal results of the proposed algorithm were compared with those of other existing research papers^{1,37,47,50} in terms of accuracy, sensitivity, and specificity, and are reported in Table 5. It has been seen that the proposed technique shows comparable performance with existing techniques. It should be noted that the

datasets used by techniques in Table 5 are different and the results only show the qualitative insights.

ANNs have advantages such as more tolerance to noisy inputs and representation of Boolean functions, but too many attributes may result in overfitting. In a support vector machine (SVM) overfitting is unlikely to occur. The training speed in the neural networks depends on network structure, momentum rate, learning rate, and converging criteria. In an SVM it depends on the size and class separability of training data. The performance of the SVM depends on the selection of kernel type and kernel parameters. The main drawback of SVM is the high algorithmic complexity and extensive memory requirements. Logistic regression is an attractive alternative to discriminant analysis whenever the dependent variable has only two classes. It is more robust than discriminant analysis to violations of the multivariate normality assumption and the equal variance/covariance assumption. The accuracy of logistic regression classifier increases with the training sample size⁵¹. The key limitations of logistic regression are: (i) it can only be used to predict discrete functions and (ii) it requires much more training data to achieve stability. ANNs provides an alternative to logistic regression. It is the most commonly used method for developing predictive models for dichotomous outcomes in medicine⁵². The ANN algorithm can get trapped in local minima or it might lead to slow convergence or even network failure. Hence neural network alone might not be the best solution. The genetic algorithm (GA) can solve a variety of optimization problems that are not well suited for standard optimization algorithms. GAs can be applied to train artificial neural networks for classification purposes. Its advantage is the ability to search the entire space for the best solution at the expense of speed, though, compared to numerical methods. However, it often tends to converge to local minima. ABC optimization technique have been applied for MLP training, this optimization makes the ANN robust, fast, simple and avoid the problem of ANN trapping in local minima. ABC trained ANN classifier outperforms all other existing classifiers.

Examining the performance evaluation results obtained, it is found that the best results were obtained when using ABC optimization algorithm trained neural network based classifier. The experimental results demonstrate that the proposed algorithm confirms the usefulness of the algorithm as an optimization tool. It shows that the ABC algorithm is very successful on optimization of training neural network. The proposed mass classifier model is implemented using MATLAB 7.5. The computerized texture analysis may

therefore provide useful information for reducing the number of negative biopsies. Further investigation will be conducted with a larger data set to determine the generalizability of these results.

Acknowledgements: The authors would like to thank Dr K. S. Murugan MD, DNB (Rad), Fellow in MRI (Ger) of clarity imaging, and Dr R. Rupa, DMRD, DNB, Consultant Radiologist, Kovai Medical Centre and Hospital, Coimbatore, Tamil Nadu, India, for providing the breast DCE-MR images used for testing the algorithm proposed and her valuable clinical advice, and the anonymous reviewers for their hints as well as comments for improvement of an earlier version of this article, and the editor of the journal for his patience and continued feedback during the revision process.

REFERENCES

1. Nirooei M, Abdolmaleki P, Tavakoli A, Gity M (2009) Feature selection and classification of breast cancer on dynamic magnetic resonance imaging using genetic algorithm and artificial neural networks. *J Electrical Systems* **5**, 1–12.
2. Fischer DR, Wurdinger S, Boettcher J, Malich A, Kaiser WA (2005) Further signs in the evaluation of magnetic resonance mammography: A retrospective study. *Investig Radiol* **40**, 430–5.
3. Kuhl C (2007) The current status of breast MR imaging Part I. Choice of technique, image interpretation, diagnostic accuracy, and transfer to clinical practice. *Radiology* **244**, 356–78.
4. Heywang Kobrunner SH, Beck R (1996) *Contrast Enhanced MRI of the Breast*. Berlin, Heidelberg, New York, Springer verlag, pp 7–56.
5. Cheng HD, Shi XJ, Min R, Hu LM, Cai XP, Du HN (2006) Approaches for automated detection and classification of masses in mammograms. *Pattern Recogn* **39**, 646–68.
6. Chen W, Giger ML, Bick U (2006) A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. *Acad Radiol* **13**, 63–72.
7. Chen W, Giger ML, Bick U, Newstead GM (2006) Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI. *Med Phys* **33**, 2878–87.
8. Chen W, Giger ML, Lan L, Bick U (2004) Computerized interpretation of breast MRI: investigation of enhancement-variance dynamics. *Med Phys* **31**, 1076–82.
9. Liney GP, Sreenivas M, Gibbs P, Garcia-Alvarez R, Turnbull LW (2006) Breast lesion analysis of shape technique: Semiautomated versus manual morphological description. *J Magn Reson Imag* **23**, 493–8.
10. Bahreini L, Fatemizadeh E, Gity M (2010) Gradient vector flow snake segmentation of breast lesions in dynamic contrast-enhanced MR images. In: *Proceedings of the 17th Iranian Conference on Biomedical Engineering (ICBME)*, Nov 3–4; Isfahan, Iran, pp 1–4.
11. Wu Q, Salganicoff M, Krishnan A, Fussell DS, Markey MK (2006) Interactive lesion segmentation on dynamic contrast enhanced breast MRI using a Markov model. *Proc SPIE* **6144**, 1487–94.
12. Cui Y, Tan Y, Zhao B, Liberman L, Parbhu R, Kaplan J, Theodoulou M, Hudis C, Schwartz LH (2009) Malignant lesion segmentation in contrast-enhanced breast MR images based on the marker-controlled watershed. *Med Phys* **36**, 4359–69.
13. Janaki Sathya D, Geetha K (2011) Development of cad system based on enhanced clustering based segmentation algorithm for detection of masses in breast DCE-MRI. *Int J Comput Sci Issues* **8**, 378–87.
14. Janaki Sathya D, Geetha K (2011) Development of intelligent system based on artificial swarm bee colony clustering algorithm for efficient mass extraction from breast DCE-MR images. *Int J Eng Trends Tech* **6**, 82–8.
15. Janaki Sathya D, Geetha K (2011) Comparative study of different edge enhancement filters in spatial domain for magnetic resonance images. *Amse Signal Process Pattern Recogn J* **54**, 30–43.
16. Gilhuijs KGA, Giger ML, Bick U (1998) Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Med Phys* **25**, 1647–54.
17. Christoyianni I, Dermatas E, Kokkinakis G (1999) Neural classification of abnormal tissue in digital mammography using statistical features of the texture. In: *Proceedings of the 6th IEEE International Conference on Electronics, Circuits and Systems*, pp 117–20.
18. Meinel LA, Stolpen AH, Berbaum KS, Fajardo LL, Reinhardt JM (2007) Breast MRI lesion classification: Improved performance of human readers with a back-propagation neural network computer-aided diagnosis (CAD) system. *J Magn Reson Imag* **25**, 89–95.
19. Sivanandam SN, Deepa SN (2006) *Introduction to Neural Networks Using Matlab 6.0*, Tata McGraw-Hill, pp 166–97.
20. Lisboa PJG (2002) A review of evidence of health benefits from artificial neural networks in medical intervention. *Neural Network* **15**, 11–39.
21. Hollingsworth AB, Stough RG (2003) The emerging role of breast magnetic resonance imaging. *J Oklahoma State Med Assoc* **96**, 299–307.
22. Pham DT, Ghanbarzadeh A, Koç E, Otri S, Rahim Zaidi M (2006) The Bees Algorithm—A novel tool for complex optimisation problems. In: *Proc. of the 2nd Virtual International Conference on Intelligent Production Machines and Systems (I*PROMS-06)*, Cardiff, UK, pp 454–9.
23. Karaboga D, Basturk B (2007) A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *J Global Optim* **39**, 459–71.

24. Basturk B, Karaboga D (2006) An artificial bee colony (ABC) algorithm for numerical function optimization. In: *Proceedings of IEEE Swarm Intelligence Symposium*, Indianapolis, IN, USA, pp 1–9.
25. Karaboga D (2005) An idea based on honey bee swarm for numerical optimization; *Technical Report-TR06*; Engineering Faculty, Computer Engineering Department, Erciyes Univ: Kayseri, Turkey, pp 1–10.
26. Pham DT, Ghanbarzadeh A, Koç E, Otri S (2006) Application of the bees algorithm to the training of radial basis function networks for control chart pattern recognition. In: *Proceedings of the 5th CIRP International Seminar on Intelligent Computation in Manufacturing Engineering (ICME-06)*, Ischia, Italy, pp 711–6.
27. Pham DT, Soroka A, Ghanbarzadeh A, Koç E, Otri S, Packianather M (2006) Optimising neural networks for identification of wood defects using the Bees Algorithm. In: *Proceedings of the IEEE International Conference on Industrial Informatics*, Singapore, pp 1346–51.
28. Karaboga D, Ozturk C (2010) Fuzzy clustering with artificial bee colony algorithm. *Sci Res Essays* **5**, 1899–902.
29. Saeedi S, Samadzadegan F, El-Sheimy N (2009) Object extraction from lidar data using an artificial swarm bee colony clustering algorithm. In: Stilla U, Rottensteiner F, Paparoditis N (eds) *Proceedings of CMRT09. IAPRS, Vol. XXXVIII, Part 3/W4*, Paris, France, 3–4 September, pp 133–8.
30. Chong CS, Low MYH, Sivakumar AI, Gay KL (2006) A bee colony optimization algorithm to job scheduling simulation. In: Perrone LF, Wieland FP, Liu J, Lawson BG, Nicol DM, Fujimoto RM (eds) *Proceedings of the Winter Conference*, Washington, DC, pp 1954–61.
31. Pham DT, Haj Darwish A, Eldukhri EE (2009) Optimization of a fuzzy logic controller using the Bees Algorithm. *Int J Comput Aided Eng Tech* **1**, 250–64.
32. Sawant HK, Shinde VD (2010) Breast tumor analysis using texture features and wavelet transform with dynamic neural network based training. *J Inform Knowl Res Comput Eng* **1**, 46–51.
33. Gibbs P, Turnbull LW (2003) Texture analysis of contrast-enhanced MR images of the breast. *Magn Reson Med* **50**, 92–8.
34. Arbach L, Stolpen A, Reinhardt JM (2004) Classification of breast MRI lesions using a backpropagation neural network (BNN). In: *Proceedings of 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano*, Piscataway, NJ, pp 253–6.
35. Szabó BK, Aspelin P, Wiberg MK (2004) Neural network approach to the segmentation and classification of dynamic magnetic resonance images of the breast: Comparison with empiric and quantitative kinetic parameters. *Acad Radiol* **11**, 1344–54.
36. McLaren CE, Chen W-P, Nie K, Su M-Y (2009) Prediction of malignant breast lesions from MRI features: a comparison of artificial neural network and logistic regression techniques. *Acad Radiol* **16**, 842–51.
37. Keyvanfard F, Shoorehdeli MA, Teshnehlab M (2011) Feature selection and classification of breast cancer on dynamic magnetic resonance imaging using ANN and SVM. *Am J Biomed Eng* **1**, 20–5.
38. Gonzalez RC, Woods RE (2005) *Digital Image Processing*, 2nd edn, Prentice Hall of India, pp 644–56.
39. Tuceryan M, Jain AK (1998) Texture analysis. In: *The Handbook of Pattern Recognition and Computer Vision*, 2nd edn, World Scientific Publishing Co, pp 207–48.
40. Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern Syst Hum* **3**, 610–21.
41. Christoyianni I, Koutras A, Dermatas E, Kokkinakis G (2001) Breast tissue classification in mammograms using ICA mixture models. In: *Proceedings of International Conference on Artificial Neural Networks, ICANN 2001*, pp 554–60.
42. Zhang Z, Lu J, Yip J (2008) Computer aided mammography. In: *Proceedings of School of Computing and Engineering Researchers' Conference*, Univ of Huddersfield, pp 125–30.
43. Dash M, Liu H (1997) Feature selection for classification. *Intell Data Anal* **1**, 131–56.
44. Hall M, Holmes G (2003) Benchmarking attribute selection techniques for discrete class data mining. *IEEE Trans Knowl Data Eng* **15**, 1437–47.
45. Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* **3**, 1157–82.
46. Mohamed WA, Kadah YM (2007) Computer aided diagnosis of digital mammograms. In: *Proceedings of International Conference on Computer Engineering and Systems, ICCES 07*, pp 299–303.
47. Sinha S, Lucas-Quesada FA, Debruhl ND, Sayre J, Farria D, Gorczyca DP, Bassett LW (1997) Multifeature analysis of Gd-enhanced MR images of breast lesions. *J Magn Reson Imag* **7**, 1016–26.
48. Tanner C (2005) Registration and lesion classification of contrast-enhanced magnetic resonance breast images. PhD thesis, pp 1–259.
49. Swets JA (1979) ROC analysis applied to the evaluation of medical imaging techniques. *Investig Radiol* **14**, 109–21.
50. Baltzer PAT, Dietzel M, Kaiser WA (2011) Nonmass lesions in magnetic resonance imaging of the breast: additional T2-weighted images improve diagnostic accuracy. *J Comput Assist Tomo* **35**, 361–6.
51. Hosmer DW Jr, Lemeshow S (2000) *Applied Logistic Regression*, 2nd edn, Wiley, New York, USA, pp 1–30.
52. Tu JV (1996) Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* **49**, 1225–31.